



REVIEW

A Review of Next-Generation Smart Manufacturing Enabled by Engineering Systems, Materials Modeling, and High-Performance Computing: A System-Oriented Perspective for Semiconductor Manufacturing

Hsiao-Chun Han¹, Der-Chen Huang^{2,*} and Chin-Ling Chen^{3,*}

¹Department of Applied Artificial Intelligence, Ming Chuan University, Taoyuan City, Taiwan

²Department of Computer Science and Engineering, National Chung Hsing University, Taichung City, Taiwan

³Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung City, Taiwan

*Corresponding Authors: Der-Chen Huang. Email: huangdc@cs.nchu.edu.tw; Chin-Ling Chen. Email: clc@mail.cyut.edu.tw

Received: 18 April 2026; Accepted: 22 June 2026; Published: 13 August 2026

ABSTRACT: Semiconductors represent the most complex production activities and stand at the forefront of smart manufacturing. In particular, the yield of advanced processes is strongly influenced by coupling among engineering systems, material behavior, and computational infrastructure. Consequently, the integration of deep learning (DL) and digital twins has become essential for driving the next-generation transformation of smart manufacturing. However, existing reviews predominantly organize literature through algorithm-oriented taxonomies, while isolated AI paradigms alone remain insufficient to effectively capture system-level interactions and industry-driven technological evolution. Therefore, this study proposes a system-oriented and industry-driven review framework, termed the System-under-Industry Guided Literature Review (S_{IG}LR), to reorganize existing literature. Specifically, 103 survey papers and the 13,377 studies they reference are systematically analyzed. In addition, a triadic ESA–MM–HPC framework is introduced as a system-oriented conceptual representation for analyzing next-generation semiconductor smart manufacturing systems, capturing the coupling among engineering system analysis (ESA), material modeling (MM), and high-performance computing (HPC) in complex manufacturing environments. Furthermore, a conceptual modeling perspective is proposed for predictive digital twin (DT) analysis. Overall, this study uses the semiconductor industry as a representative domain to extend existing smart manufacturing research from algorithm-centered approaches toward more system- and structure-oriented perspectives. The perspectives provide a unified, scalable foundation for AI-driven industrial systems research, advancing manufacturing technologies and discussing challenges associated with globally distributed semiconductor manufacturing and industrial reshoring.

KEYWORDS: Engineering system analysis; material modeling; high performance computation; digital twin; smart manufacturing; semiconductor; TSMC; automatic control systems; robotic manipulators; automated guided vehicles

1 Introduction

1.1 Motivation

Semiconductor manufacturing is regarded as the most complex cyber-physical production system, characterized by high automation density, multi-scale coupling, and material-physics constraints in advanced process nodes. Consequently, production efficiency and yield increasingly depend on integrating engineering, materials, and computational infrastructures rather than relying solely on localized optimization approaches. (Fig. 1). However, the highly coupled nature also dramatically increases system complexity

and makes analysis more difficult using conventional frameworks alone. Therefore, beyond DL, DT's synchronization, simulation, and predictive capabilities is necessary to support real-time responses to dynamic process variations and disturbances. These developments point the way to next-generation smart manufacturing, not only to improve production efficiency and yield but also to support the continued advancement of semiconductor process technologies. In addition, the complexity intensifies the challenges associated with distributed production. Accordingly, this study considers this is important direction for addressing issues of globalization and industrial reshoring.

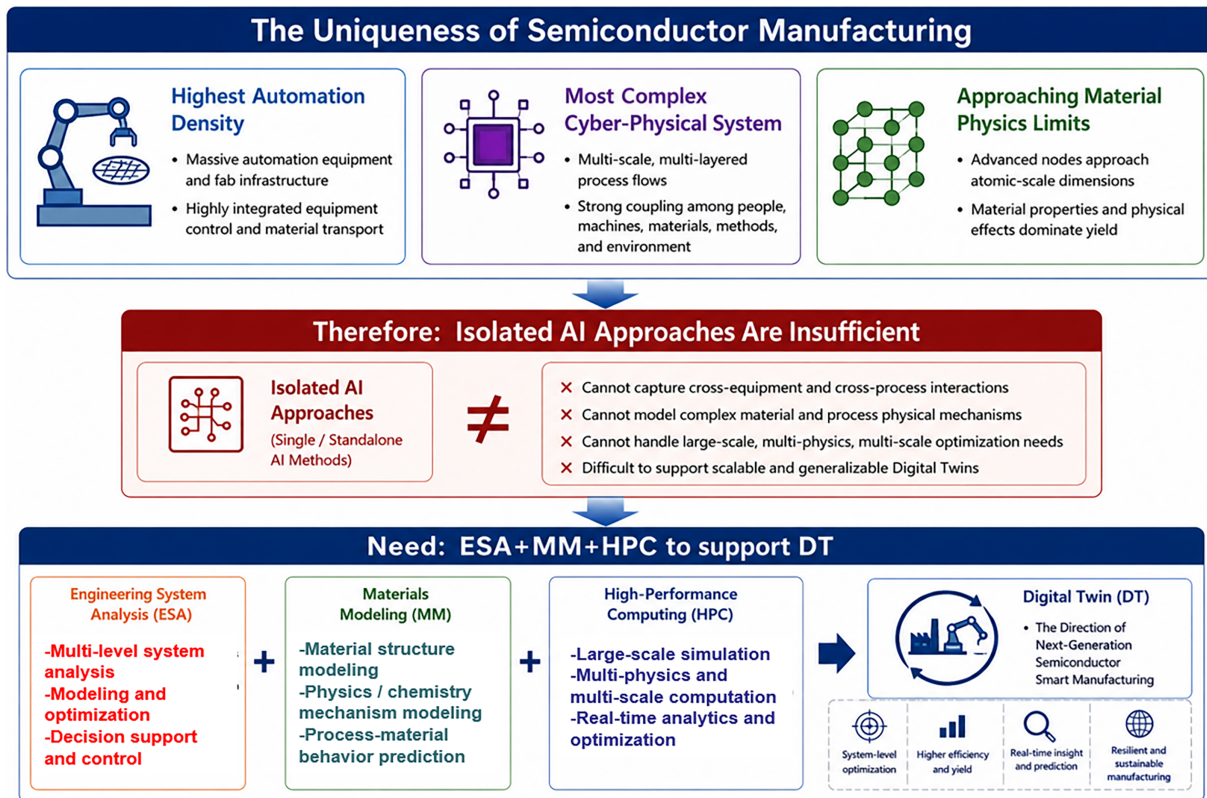


Figure 1: Process complexity drives the need for ESA–MM–HPC-integrated DT to support next-generation semiconductor manufacturing.

Such a context implies that semiconductor smart manufacturing cannot be understood solely through isolated AI algorithms, but instead through innovations in industrial systems and engineering science. For example, ESA, MM, and HPC represent three important and increasingly interconnected domains in engineering science and advanced manufacturing (as evidenced by the extensive body of literature summarized in Fig. 2 and Table 1). And ESA encompasses domains such as automatic control systems, robotics, vehicle and suspension systems. For complex production, engineering analysis has evolved toward high-fidelity dynamic synchronization between cyber and physical systems. Concurrently, materials science has advanced to the angstrom scale and developed multiscale predictive modeling to overcome physical constraints. Therefore, traditional analytical frameworks based on partial differential equations (PDEs) and numerical methods face challenges due to exponential growth in computational complexity and resource bottlenecks. These issues are especially critical in semiconductor manufacturing, where multiscale physical phenomena must be addressed alongside stringent real-time control requirements, motivating researchers to explore more scalable, computationally efficient surrogate models and data-driven mechanisms. The

objective is to improve simulation efficiency while preserving physical fidelity in increasingly complex manufacturing environments. Such a transition—from compute-intensive modeling to data-enhanced engineering approaches—may provide a promising direction for addressing large-scale, complex system evolution, materials lifetime prediction, and heterogeneous computing resource orchestration.

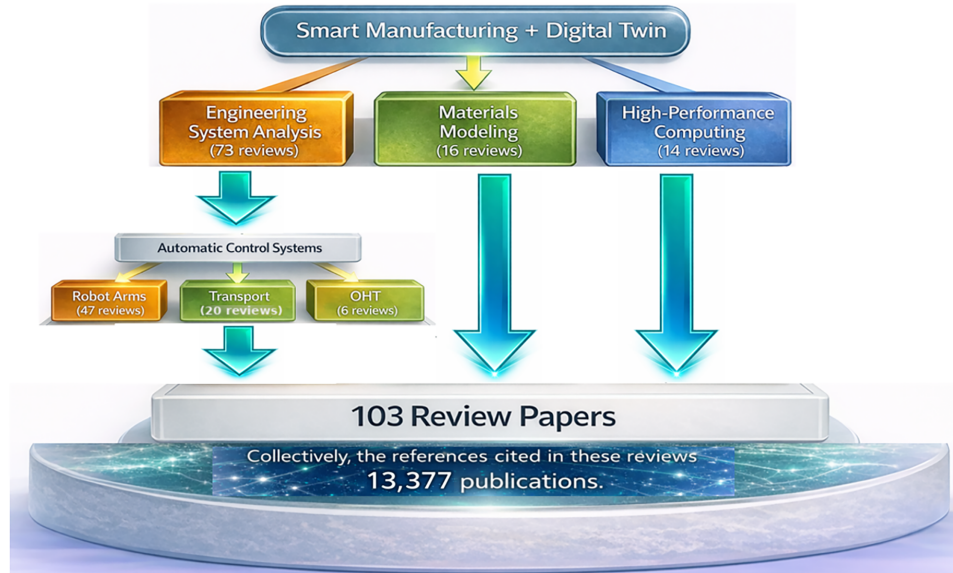


Figure 2: To illustrate the research landscape of next generation semiconductor smart manufacturing enabled by engineering system analysis, materials modeling, and high-performance computing. A total of 103 review articles indicate the growing importance and interconnections among these domains in modern intelligent manufacturing research (see Table 1). Collectively, the references cited in these reviews are 13,377 publications, highlighting the rapid expansion and interdisciplinary nature of this research field.

Table 1: System-oriented classification of 103 review papers within engineering, materials, and HPC domains.

Category	Subcategory	Reviewed Literature
Control–Cyber Axis (73 reviews)	Smart Manufacturing and Automatic Control Systems (18 reviews)	[1–18]
	Robotic Manipulators (29 reviews)	[19–47]
	Transport Systems (20 reviews)	[48–67]
	Suspension Systems (6 reviews)	[68–73]
Material–Physics Axis (16 reviews)		[74–89]
HPC–Compute Axis (14 reviews)		[90–103]

Recently, DL has increasingly influenced research in ESA, MM, and HPC. By leveraging physics-informed neural networks (PINN), surrogate models, and simulation acceleration techniques, DL has improved the efficiency of engineering simulation and analysis and enabled more adaptive prediction and optimization under uncertainty. These developments are increasingly reflected in semiconductor smart manufacturing. Taiwan Semiconductor Manufacturing Company (TSMC) is often cited as an example of advanced semiconductor manufacturing. With the explosive growth in global demand for AI chips, wafer fabrication is increasingly emphasizing sustainability, intelligence, predictability, and scalability. Meanwhile, recent developments in semiconductor manufacturing have increasingly integrated digital twins, data-driven automation, and cyber-physical coordination.

Amid geopolitical tensions, the United States (US) promotes reshoring manufacturing and encourages enterprises to produce domestically. However, these initiatives face challenges related to skilled labor shortages and cultural, management, and regulatory disparities. In this context, TSMC's investment in Arizona is widely regarded as a flagship example. These developments highlight the importance of automation, digital infrastructure, and system integration in geographically distributed environments. Meanwhile, process technologies, automation, and digital capabilities are emphasized. The coordination among ESA, MM, and HPC within DT-enabled cyber-physical environments becomes an important research topic for smart semiconductor manufacturing.

Beyond relocating fabs, advanced economies can further strengthen engineering research, materials science, AI innovation, and computing infrastructure. Meanwhile, semiconductor smart manufacturing increasingly integrates engineering systems, materials modeling, and computational infrastructures to highlight the role of cyber-physical integration, automation, and digital infrastructure in supporting distributed, coordinated manufacturing environments. This trend enables advanced economies to leverage their research and innovation strengths to support both global manufacturing deployment and reshoring initiatives. It also highlights the importance of foundational scientific research in sustaining long-term advances in semiconductor technologies. Within this paradigm, ESA, MM, and HPC can be viewed as closely interacting components in next-generation smart manufacturing research. This study proposes a system-oriented perspective to provide a conceptual framework for examining their interactions.

1.2 Contributions and Objectives

Based on the substantial body of review literature on industrial intelligence, this study systematically synthesizes 103 reviews and their 13,377 references to identify limitations in the current research landscape when addressing complex engineering systems, as follows:

1. Most reviews emphasize technology-centric applications with limited discussion of cross-domain integration [88], or focus narrowly on individual technical components with limited consideration of system interoperability and cross-platform integration challenges [47];
2. Current data-driven research is largely confined to algorithmic comparisons based on static models while neglecting the behavior of dynamic systems under real-world physical constraints [38]. Moreover, discussions on cyber-physical systems (CPS) are often restricted to isolated case studies, lacking architectural comparisons and system-level generalization [33];
3. Although the role of the Industrial Internet of Things (IIoT) in smart manufacturing has been extensively discussed, few review studies provide an integrated analysis spanning the full pipeline from data acquisition and system-level analysis to decision-making processes [62].

Therefore, this study makes the following key contributions to the field:

1. This study refines the Systematic Literature Review (SLR) methodologies. SLR implicitly assumes that research topics and technological developments primarily evolve within the academic community and organizes literature using an algorithm-oriented taxonomy. However, in highly engineered domains such as semiconductor manufacturing, blockchain, and AI, technological advancement is shaped by industrial dynamics, rendering SLR insufficient for capturing system-level interdependencies of technologies. Therefore, this study proposes the S_{IGLR} framework, a system-centric literature analysis approach grounded in industrial sectors. S_{IGLR} establishes correspondences between technological developments and industrial system functionalities, thereby providing a system-oriented perspective for analyzing industry-driven technological evolution.
2. This study proposes a triadic ESA–MM–HPC architecture (refer to Fig. 2) as a system-oriented conceptual framework for next-generation semiconductor smart manufacturing systems. This architecture describes the interactions among materials, system-level decisions, and computational capabilities in complex engineering systems. By integrating this architecture with S_{IGLR} , this study provides a system-oriented perspective for examining the relationships between industrial requirements and technological evolution.

This study proposes S_{IGLR} , approaching the literature from the perspective of ESA in next-generation semiconductor smart manufacturing. It systematically organizes and interprets studies to examine cross-scale and cross-domain technological interdependencies in relation to industrial functionalities. The study pursues the following three primary objectives:

1. To establish a tri-axis analytical framework, comprising the Control–Cyber Axis, Material–Physics Axis, and HPC–Compute Axis for examining major functional domains in semiconductor smart manufacturing systems;
2. To systematically map the selected literature onto system modules and cross-scale interaction structures, thereby analyzing the roles and coupling relationships of key technologies—such as DL, DT, and HPC—within industrial systems;
3. To identify structural gaps in existing research and propose a system-oriented future research roadmap to support the development of AI-driven semiconductor smart manufacturing and digital twin systems.

1.3 Organization and Reading Map

This study adopts a system-oriented perspective to discuss the interactions among ESA, MM, and HPC in next-generation smart manufacturing and DT systems. Accordingly, the structure is organized as follows (refer to Fig. 3):

- Section 2 reviews the technological background and research foundations of semiconductor smart manufacturing, including processes, equipment architectures, and roles of ESA, MM, HPC, and DT in supporting real-time intelligent manufacturing;
- Section 3 presents the research methodology, S_{IGLR} , adopted in this study and details the database search strategies, Boolean query formulations, inclusion and exclusion criteria, and the resulting classification axes and processing workflow;
- Section 4 surveys the three research domains—ESA, MM, and HPC—within the context of semiconductor smart manufacturing, and covers topics including automatic control systems, robotic control, transportation dynamics analysis, multiscale materials modeling, and HPC infrastructures;

- [Section 5](#) discusses open research challenges and practical limitations that remain unresolved in current studies, and proposes future research directions and development considerations. For example, a modeling approach that integrates Laplacian flow and sheaf theory [104,105] can be adopted to reveal the mechanisms of anomaly propagation under multiple dynamic constraints;
- [Section 6](#) concludes the paper by summarizing the main contributions and key limitations of this study.

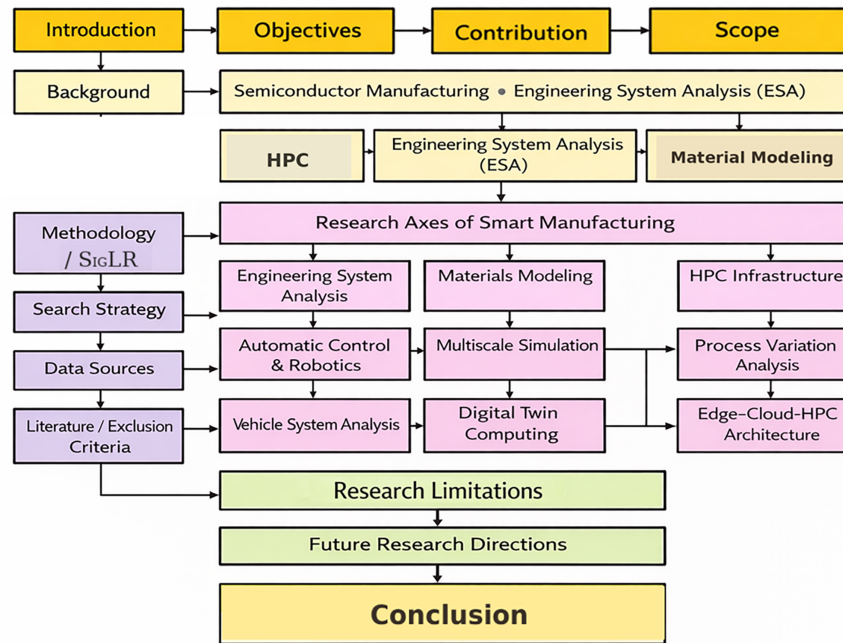


Figure 3: A system-oriented literature review framework for semiconductor smart manufacturing based on SIGLR and the ESA-MM-HPC perspective.

2 Background

Smart manufacturing has become a direction of semiconductor industrial upgrading. It supports yield management, process coordination, and large-scale semiconductor manufacturing operations. In this context, the ESA, MM, and HPC research domains are receiving increasing attention. Among these, DT is an approach for improving cyber-physical coordination and system-level integration in semiconductor smart manufacturing.

2.1 Semiconductor Processes, Equipment, and Globalization

Semiconductor manufacturing is a complex, automated process in which integrated circuit (IC) structures are formed on silicon wafers through multiple steps, including lithography, thin-film deposition, etching, and ion implantation. During production, wafers are stored in Front Opening Unified Pods (FOUP) to prevent contamination (e.g., [Fig. 4](#)). In addition, FOUPs are transported to equipment via automated transport systems, such as Overhead Hoist Transport (OHT) systems. Upon arrival, the Equipment Front End Module (EFEM) interfaces with the FOUP to enable wafer access and transfer. Individual wafers are then transferred into downstream processing equipment for subsequent processing steps.

A cluster tool consists of a vacuum isolation mechanism, multiple process chambers, robotic manipulators, and a wafer transport module (e.g., [Fig. 5](#)). This architecture enables automated wafer transfer across multiple steps, supports parallel processing, reduces contamination risks, and enhances efficiency, thereby

making it one of the core equipment configurations in the fab. Throughout the process, a large number of sensors continuously monitor equipment conditions, including temperature, gas flow rates, pressure, motion velocity, and parameters. These data are transmitted in real time to the Manufacturing Execution System (MES) and the Equipment Engineering System (EES), enabling real-time monitoring, anomaly detection, and process optimization [106].

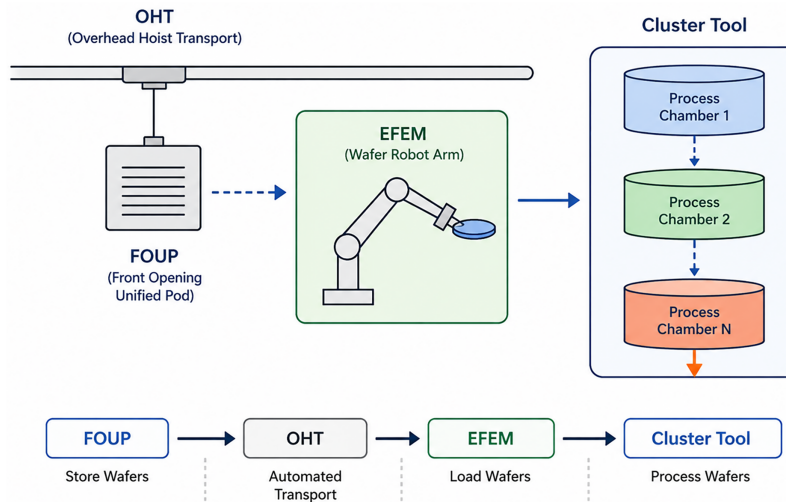


Figure 4: Automated wafer transfer workflow via FOUP, OHT, and EFEM.

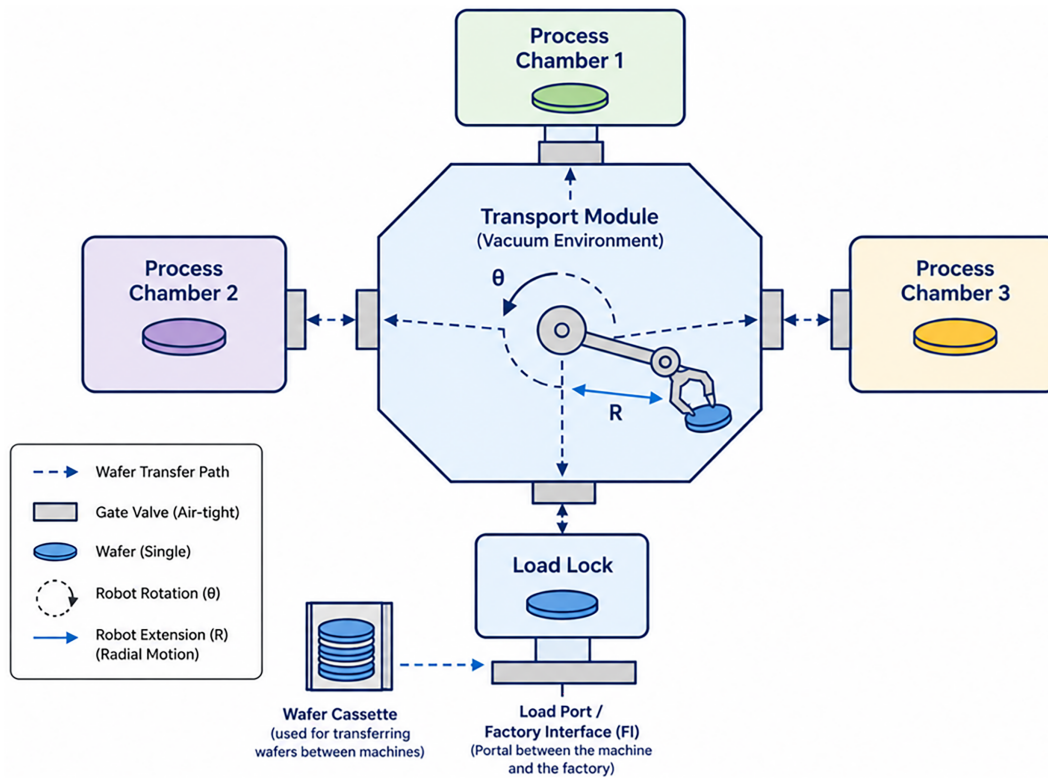


Figure 5: Internal architecture of a semiconductor cluster tool.

Leading semiconductor manufacturers, including TSMC, have increasingly adopted smart manufacturing technologies such as DL-assisted process control, intelligent equipment diagnostics, automated platforms, and DT. These technologies are further integrated with IIoT, edge computing, and cloud computing to enable process variation monitoring and sustain high-precision control [107].

Recently, due to geopolitical tensions, TSMC has expanded its global footprint by establishing fabs in the US and Japan. To support this strategy, TSMC accelerates the integration and diffusion of smart manufacturing technologies across distributed fabs, while advancing DT and CPS integration. As a result, sensor networks and data-driven analytics become important components of semiconductor smart manufacturing, further motivating research on ESA, MM, and HPC [10].

2.2 ESA and Semiconductor Manufacturing

ESA is a system-level analytical perspective encompassing systems engineering, engineering systems, and systems analysis to examine interactions, dynamic behaviors, and overall system performance. Current engineering research often focuses on individual equipment or isolated physical phenomena. In contrast, ESA integrates MM, control theory, and dynamic analysis to investigate system behavior and optimization strategies from multi-level and interdisciplinary perspectives. From the ESA viewpoint, semiconductor manufacturing constitutes a large-scale, multi-scale, and tightly coupled dynamic system, comprising hundreds of pieces of equipment, thousands of steps, complex transport systems, and hierarchical control architectures. It encompasses automatic control systems, robotic manipulators, and vehicle dynamics [10]. Furthermore, due to re-entrant characteristics, it is widely regarded as one of the most challenging scheduling problems.

Automatic control systems provide a framework for utilizing sensor feedback and control algorithms—such as model predictive control (MPC) and adaptive control—to regulate physical entities (e.g., robotic manipulators and vehicles) in real time [60,95]. It is responsible for translating control objectives derived from ESA into precise physical actions, while enabling automatic correction in response to deviations or environmental disturbances (e.g., vision-guided path compensation in robotic systems) [13,34,39]. Accordingly, it contributes to real-time responsiveness and system stability within smart manufacturing architectures [6].

Robotic manipulators and automated transport systems constitute the physical execution layer of ESA and automatic control, translating electrical signals into high-precision motions. During operation, they continuously transmit real-time data through IoT sensors to the ESA framework for model validation and system-level analysis [10], and receive control commands to execute physical actions. Within the IIoT and CPS architectures, digital systems are tightly integrated with physical equipment, enabling coordinated operation and real-time interaction between cyber and physical domains [95].

Furthermore, beyond the interactions among equipment, robotic manipulators, transport vehicles, and scheduling information, the incorporation of materials physics, equipment states, process, and environmental variability introduces multiple feedback loops and nonlinear dynamics, significantly increasing control complexity. As a result, extensive research has investigated challenges in mathematical modeling, real-time decision-making, and system-level optimization in automatic control. Recently, with the advancement of IIoT, cloud-based architectures, and edge computing, ESA has been discussed as a system-oriented analytical perspective for smart manufacturing. It serves as a critical bridge linking process engineering, MM, and HPC, and, when integrated with DT, further drives the evolution of smart manufacturing toward the next generation [106,107].

2.3 Material Modeling and ESA

Materials modeling uses physics-based models, numerical simulations, and multiscale computational methods to characterize the physical behavior of materials under various processing conditions. Common approaches include Density Functional Theory (DFT), Molecular Dynamics (MD), the Finite Element Method (FEM), and phase-field modeling. These methods enable the analysis of material properties across angstrom- to microscale regimes under processes such as deposition, etching, thermal treatments, and mechanical loading. Recently, materials modeling has been integrated with DL and surrogate models to accelerate material behavior prediction and support process optimization [79,82].

Many process variabilities in semiconductor manufacturing originate from the physical behavior of materials. For example, microstructural evolution during thin-film growth, lattice defects, interface roughness, electromigration, stress migration, and thermally induced migration can affect process stability and yield by coupled mechanisms, thereby becoming considerations within ESA. Process control parameters—such as temperature, pressure, gas flow rates, and power settings—govern the evolution of materials during fabrication, while the resulting material physical properties in turn influence process outcomes and system states. Consequently, MM represents an analytical component of ESA: material behavior influences the physical limits of the process, and these constraints determine the feasible range of control parameters, which ultimately shape the state space and transition dynamics of system-level models [82,87].

In this context, MM provides physical insights into ESA, as it governs how microscopic physical mechanisms are translated into macroscopic system behavior. The physical insights not only enhance the interpretability of process models but also support DT systems, improving the representation of real-world dynamics and thereby support real-time monitoring, prediction, and process optimization for smart manufacturing [79,81].

2.4 High Performance Computation and ESA

HPC utilizes large-scale parallel architectures—such as supercomputers, GPU clusters, and cloud platforms—to solve computationally intensive problems in a reasonable timeframe. It is widely applied in climate modeling, materials science, fluid dynamics, and engineering optimization, and has become a component for smart manufacturing with the advancement of cloud and GPU technologies [90,91] (e.g., Fig. 6).

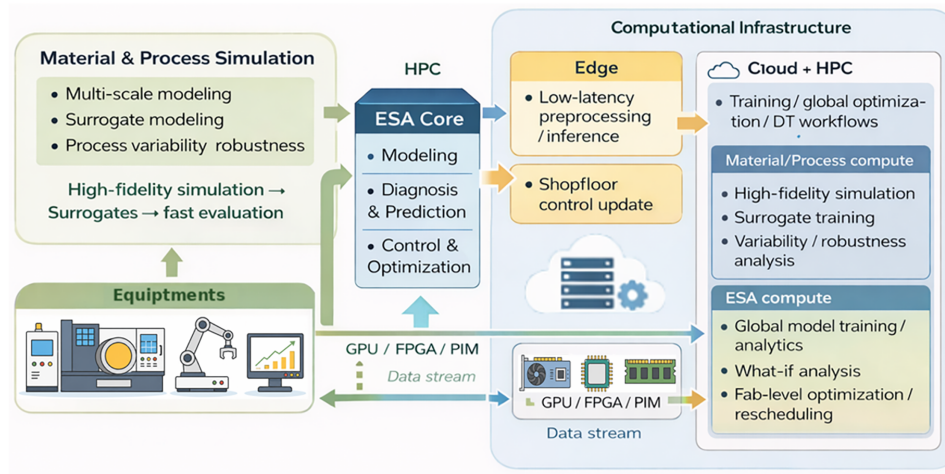


Figure 6: Conceptual HPC-supported ESA perspective for semiconductor smart manufacturing.

Semiconductor manufacturing involves extensive multiscale physical phenomena and challenges in equipment control, which are analyzed using physics-based models such as DFT, MD, and FEM. These simulations are computationally intensive. For instance, DFT and *ab initio* molecular dynamics are limited to temporal scales of femtoseconds to picoseconds, whereas equipment control and process adjustments require decision-making at the millisecond scale, resulting in a significant scale mismatch and corresponding computational bottlenecks. Consequently, in smart manufacturing systems, HPC is integrated with surrogate models or DL to accelerate computation and prediction [85]. By leveraging HPC for large-scale simulations and model training, surrogate models that capture physical behaviors can be constructed, enabling ESA to rapidly predict process outcomes and support real-time control decisions [94,95]. Furthermore, with the advancement of cloud, GPU, and edge computing, HPC has been integrated into DT systems, supporting system analysis and decision-making [90].

2.5 AI, ESA, and Smart Manufacturing

Current AI research in smart manufacturing focuses on integrating CPS, IIoT, big data, and cloud computing to support adaptive system simulation. Among these approaches, Deep Reinforcement Learning (DRL) has been widely applied to intelligent control, predictive maintenance, and smart production to support more adaptive and automated systems [6]. In addition, recent studies leverage DT to construct virtual training environments, addressing the challenges of data scarcity and poor data quality in AI model training. This enables AI agents to be trained and optimized in safe and controlled environments [7].

In ESA, AI research focuses on accurately identifying and virtually mapping complex processes. For instance, in semiconductor manufacturing, the large-scale, high-dimensional data generated during production can be analyzed using DL to extract latent knowledge [8]. Moreover, polynomial neural networks are employed to predict quality metrics—such as the material removal rate (MRR)—without direct physical measurements, thereby enabling real-time process monitoring and the development of adaptive Virtual Metrology (VM) systems [8].

In Automatic control systems, AI research overcomes the limitations of physical models and enables closed-loop control. For example, DL is employed to integrate visual perception modules, localization, path planning, and intelligent control, in which visual feedback triggers control actions, enabling systems to respond to disturbances and autonomously avoid obstacles [34]. In addition, DT enables real-time simulation in virtual environments to implement compensation strategies prior to defect occurrence, thereby reducing equipment downtime and process disruptions [13,16]. Furthermore, model predictive control (MPC) is applied to address nonlinear constraints to optimize control trajectories in precision transport systems and process equipment [17,60].

In robotic manipulators, AI research focuses on adaptive systems and on integrating kinematics, actuation principles, and human–robot interaction. Deep convolutional neural networks (DCNN) are employed to enhance scene understanding and environmental perception for robotic systems operating in unstructured environments [30]. Moreover, DL is applied to quality monitoring and defect detection in electronic manufacturing and materials inspection [95]. In addition, recent studies analyze joint vibration signals and electrical current signatures of robotic manipulators to apply DL to prognostics and health management (PHM). This enables the prediction of mechanical fatigue and the remaining useful life of critical components [11].

2.6 Digital Twin and Semiconductor Manufacturing

In semiconductor manufacturing, DT does not initially achieve a fully synchronized cyber–physical system but rather develops progressively as data infrastructure, modeling capabilities, and system integration mature, ultimately evolving into intelligent systems capable of supporting closed-loop decision-making (e.g., Fig. 7) [106].

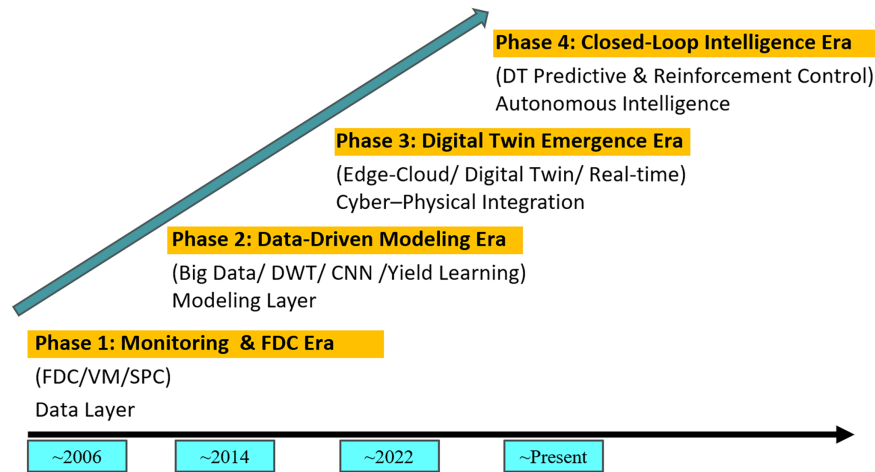


Figure 7: Evolution of digital twin in semiconductor manufacturing.

In the initial phase, the Data Monitoring Era (prior to 2006), the primary focus of fabs remained on Fault Detection and Classification (FDC), VM, and Statistical Process Control (SPC). Sensor data monitoring and feature extraction were employed to enhance equipment health and process stability. For example, Ref. [108] used the wavelet transform for dimensionality reduction and fault classification of high-frequency process signals, highlighting the issues of data availability and the identifiability of anomaly signals. Meanwhile, Ref. [109] developed automated recognition techniques for spatial patterns of wafer defects, highlighting the issue of defect monitoring and anomaly localization.

In the second phase, the Data-Driven Modeling Era (approximately 2006–2014), the deployment of IIoT led to a substantial increase in both data volume and dimensionality. Consequently, research focuses on leveraging big data and DL to enhance modeling capabilities for diagnostics and advanced process control (APC). Specifically, Ref. [110] indicated that big data can support applications such as APC, VM, and predictive maintenance (PdM) and emphasized the importance of distributed computing architectures. Ref. [111] proposed a retrospective data mining framework to extract interpretable diagnostic rules. In this stage, despite advancements in data-driven analytics, most approaches remain limited to offline or near-online decision support.

In the third phase, the DT Emergence Era (approximately 2014–2022), the concepts of DT became well-defined, and research emphasizes the integration of models, data, and decision-making along with CPS. Ref. [112] reviewed the foundations, models, and issues of DT, establishing its role as a core methodology for smart manufacturing. Furthermore, Ref. [113] applied DT to scheduling and decision support, and validated strategies through cyber–physical synchronization.

In the final phase, the Closed-Loop Intelligence Era (post-2022), the Digital Twin becomes integrated with reinforcement learning (RL) and closed-loop optimization, advancing from visualization and prediction to control and strategy generation. For example, Ref. [114] applied RL to optimize dynamic dispatching and scheduling in a fab under constraints, demonstrating the decision value of closed-loop intelligence. In addition, Ref. [115] combines DT, surrogate models, and closed-loop optimization to provide insights for future adoption of closed-loop DT frameworks in semiconductor manufacturing, including run-to-run (R2R) control, predictive control, and cross-scale integration [90,106,107].

Overall, the evolution of DT in semiconductor manufacturing can be viewed as a continuous spectrum, progressing from data monitoring to modeling to frameworks and ultimately to closed-loop intelligence. Future research should focus on integrating real-time capability, model interpretability, and control feasibility into deployable, fab-scale closed-loop architectures. The DT-related literature reviewed in this study represents varying levels of integration maturity and evolving development trends in semiconductor intelligent manufacturing, rather than treating all DT-related studies as fully closed-loop DT implementations.

3 Survey Methodologies and Taxonomy

Traditional SLR methodology (e.g., [116]) faces challenges from rapid technological advancement in industry. Existing reviews in robotics, autonomous vehicles, MM, and smart manufacturing typically organize the literature by algorithm types or platform categories. This approach weakens their ability to guide future research directions, as it overlooks technological trajectories driven by industry. This limitation arises from an implicit assumption underlying SLR that research topics and the evolution of the literature emerge endogenously within the academic community. However, in semiconductors, blockchain, AI, robotics, materials science, and HPC, the industry's influence is substantial. As a result, SLR struggles to explain why certain technological directions rapidly become focal points of academic research within relatively short time frames. To address these limitations, this study proposes $S_{IG}LR$, an industry-driven framework that extends SLR by structuring the literature according to industrial architecture rather than algorithm categories. However, $S_{IG}LR$ still follows the same procedures, including search strategy, inclusion/exclusion criteria, screening, data extraction, and synthesis.

Semiconductor manufacturing is the most robotics-intensive, highly automated, and materials-intensive industrial domain. In particular, leading-edge process nodes (sub-3 nm) have driven process complexity, capital intensity, and system integration, thereby shaping the trajectory of research in related fields. Under such conditions, applying SLR to categorize the literature based on algorithmic taxonomy can fragment dependencies among technologies, failing to accurately capture their interactions within industrial systems. This classification bias is pronounced in highly coupled engineering systems, where it may lead to misleading interpretations of research directions and reduce the practical relevance of review studies. Therefore, $S_{IG}LR$ should not be regarded merely as an alternative methodology, but as a necessary framework to avoid structural mischaracterization in the analysis of industry-driven technological evolution.

3.1 Data Sources

Generative AI, autonomous driving, and robotics have significantly increased demand for advanced semiconductor devices and accelerated the development of leading-edge process technologies. In principle, such industrial demand should stimulate corresponding academic research. However, technological advances are primarily documented in equipment vendor reports, in-fab technical documentation, or patents, while academic publications often refer to semiconductor manufacturing only implicitly through broader terms such as electronics manufacturing, high-precision manufacturing, or cleanroom automation. As a result, existing review studies, lacking an explicit relational framework, struggle to establish clear

mappings between these research efforts and the semiconductor industry. Moreover, robotic systems and automated transport equipment in semiconductor manufacturing are embedded within a highly coupled system involving material variability, complex processes, multi-layer logistics, and hierarchical control architectures. Therefore, this study introduces ESA as a cross-layer modeling and organizational framework to define system boundaries, functional roles, and interdependencies across hierarchical levels. This study uses “ESA + semiconductor/TSMC + smart manufacturing + digital twin” as the basis for literature classification and search. This framework is further integrated with key modules supporting wafer fabrication, including robotic manipulators, transport systems, MM, and HPC. Corresponding keyword combinations are constructed according to the core technologies of each domain for literature retrieval, screening, and reorganization. Here, MM and HPC are not treated as research methods per se, but as critical indicators of system complexity in advanced semiconductor manufacturing, supporting ESA in defining system-level requirements.

3.1.1 Keywords and Boolean Logic

This study adopts “ESA + semiconductor/TSMC + smart manufacturing + digital twin” as the foundational framework, which is further integrated with key enabling components of wafer fabrication—namely robotic manipulators, transport systems, MM, and HPC—to construct keyword sets and Boolean search queries. The framework is structured as follows:

- Application Layer: ESA + semiconductor/TSMC + Smart Manufacturing + Digital Twin
- Manufacturing Enabling Layer: ESA, MM, and HPC
- ESA Supporting Layer: Automatic control systems, robotic manipulators, transport systems, and suspension systems
- Foundational Technology Layer: dynamics, variability control and adaptation, closed-loop control, synchronization and coordination, multi-objective and multi-scale modeling, surrogate modeling, reinforcement learning models, resource scheduling, simulation, hybrid modeling, and edge computing

The use of “TSMC” as a keyword stems from a subset of academic studies that explicitly investigate TSMC’s manufacturing systems. Moreover, in academic databases, publications indexed under the keyword “TSMC” are typically associated with semiconductor manufacturing or wafer fabrication. To ensure comprehensive coverage and avoid omission of relevant studies, the keyword “semiconductor” is incorporated into the search strategy to effectively identify literature within targeted research domains.

Subsequently, this study adopts a system-driven keyword generation approach. Each keyword combination is constructed by pairing system modules with enabling technologies to identify cross-layer research relevant to semiconductor smart manufacturing. First, the core Boolean query is defined as follows:

- Core query = (“Engineering System Analysis”) AND (“TSMC” OR “Semiconductor”) AND (“Smart Manufacturing”) AND (“Digital Twin”)

Based on this core query, 26 extended Boolean search queries are constructed by incorporating additional keywords.

- Core query
- (“automatic control systems”) AND Core query
- (“robotic manipulators”) AND Core query
- (“motion control”) AND (“robotic manipulators”) AND Core query
- (“force control”) AND (“robotic manipulators”) AND Core query
- (“visual servoing control”) AND (“robotic manipulators”) AND Core query

- (“closed-loop”) AND (“robotic manipulators”) AND Core query
- (“cleanroom”) AND Core query
- (“synchronization and coordination”) AND (“robotic manipulators”) AND Core query
- (“process integration”) AND (“robotic manipulators”) AND Core query
- (“vehicle dynamic systems”) AND Core query
- (“dynamics”) AND (“vehicles”) AND Core query
- (“adaptive control”) AND (“vehicles”) AND Core query
- (“multi-objective control”) AND (“vehicles”) AND Core query
- (“reinforcement learning”) AND (“vehicles”) AND Core query
- (“digital twin”) AND (“vehicles”) AND Core query
- (“suspension systems”) AND Core query
- (“materials modeling”) AND Core query
- (“multiscale simulation”) AND (“IC”) AND Core query
- (“surrogate modeling”) AND (“IC”) AND Core query
- (“simulation under variability”) AND (“IC”) AND Core query
- (“reinforcement learning”) AND (“IC”) AND Core query
- (“high-performance computing”) AND Core query
- (“cloud computing”) AND (“HPC”) AND (“resource scheduling”) AND Core query
- (“numerical simulation”) AND (“hybrid models”) AND Core query
- (“edge-HPC”) AND Core query

3.1.2 Literature Sources and Databases

Traditional SLR approaches typically begin by selecting several academic databases and conducting keyword searches within each. In this study, survey literature retrieval and preliminary screening were also conducted in IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and arXiv (including preprints), based on the 26 sets of Keywords and Boolean Logic defined in [Section 3.1.1](#). However, during the literature screening process, significant cross-database indexing inconsistencies were identified. Specifically:

1. When a publication was available in authoritative databases such as IEEE Xplore, the official database source was preferentially adopted for subsequent analysis. Nevertheless, some studies indexed in IEEE Xplore or other databases could not always be effectively retrieved through direct database queries alone. Therefore, Google Search was employed as a supplementary retrieval mechanism to facilitate the identification of official publication sources and accessible versions during the practical literature screening process.
2. The same publication frequently appeared across multiple platforms, while inconsistencies existed in metadata, document versions, downloadable content, and reference accessibility.

Accordingly, the proposed S_{IGLR} framework further incorporates Google Search as a supplementary retrieval mechanism. For the preliminarily screened survey papers, cross-searching was performed by combining paper titles, author information, and the corresponding Boolean Logic expressions. This approach was intended to reduce database-dependent retrieval bias, identify sources providing the most complete information (e.g., PDF availability, full-text access, and reference accessibility). The practical execution procedure was as follows:

1. The 26 sets of Keywords and Boolean Logic expressions were applied to IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and arXiv for literature retrieval.
2. Preliminary screening was conducted according to paper titles, abstracts, research relevance, and the inclusion/exclusion criteria defined in [Section 3.2](#).
3. For the preliminarily screened survey papers, cross-searches were further performed in Google Search by combining paper titles, author information, and the corresponding Boolean Logic expressions.
4. Literature records and accessible information from different sources were cross-compared.
5. Among multiple available sources, the source providing the most complete metadata, PDF access, full-text availability, and reference accessibility was selected for subsequent analysis.

Subsequently, the reference lists of the 103 selected survey papers were collected, yielding a total of 13,377 referenced studies that served as the foundation for the subsequent analysis.

3.2 Inclusion and Exclusion Criteria

The inclusion and exclusion criteria adopted in this study are defined as follows:

- Inclusion Criteria:

References cited in relevant review papers are considered primary sources for candidate literature. Eligible publications include peer-reviewed journal articles, international conference proceedings, book chapters, and survey/review papers.

- Exclusion Criteria:

1. Document Type: Non-peer-reviewed materials, such as technical reports, corporate white papers, internal documents, news articles, and unreviewed online resources, were generally excluded. However, certain preprints were included after manual verification if they had been cited by selected review papers, demonstrated sufficient citation relevance, or had subsequent formally published versions that were less accessible.
2. Research Topic: Studies that do not fall within the following relevant technical domains are excluded: ESA, semiconductor, smart manufacturing, digital twins, automatic control systems, robotic manipulators, transport system, motion control, force control, visual servoing, closed-loop control, cleanroom automation, synchronization and coordinated control, process integration, dynamical systems, dynamics, adaptive control, MPC, multi-objective control, RL, suspension systems, MM, multiscale simulation, surrogate modeling, cloud, resource scheduling, numerical simulation, hybrid models, edge.
3. Relevance to Semiconductor Manufacturing: Even if a study belongs to the aforementioned technical domains, it is excluded if its applicability to semiconductor manufacturing scenarios is limited or if it lacks clear relevance to semiconductor smart manufacturing.

3.3 Data Processing and Classification Methodology

This study adopts a System-under-Industry Guided Literature Review (S_{IG}LR) approach, an industry-oriented framework for literature review. Boolean-logic-based keyword combinations are first used to perform systematic searches across major international academic databases. The retrieved literature is then filtered and refined through the following procedures (e.g., [Fig. 8](#)):

1. Search: Keyword queries are executed, yielding a total of 103 review papers.
2. Reference Extraction: The reference lists of the 103 review papers are aggregated, resulting in 13,377 candidate publications.

3. **Screening (Preliminary Semantic Organization):** To process 13,377 publications, this study adopted a multi-stage screening procedure. First, the literature was organized into 26 groups of Keywords and Boolean Logic, followed by an initial keyword-based screening. Subsequently, LLM-based semantic assistance (ChatGPT 5.3) was employed for preliminary semantic grouping and relevance prioritization. The LLM-based semantic assistance was not the final decision-making tool for inclusion or exclusion of literature. Final inclusion and exclusion decisions were manually verified by the author against predefined screening criteria, title relevance, abstract content, and overall research-topic relevance. During the process, studies demonstrating limited relevance to semiconductor smart manufacturing, DT, ESA, MM, or HPC, or lacking sufficient contextual relevance, were excluded. Through the above multi-stage screening process, a total of 680 potential candidate papers (including duplicate records) were retained as the basis for subsequent analysis.
4. **Manual Review:** For the 680 candidate papers, this study examines the title, abstract, introduction, and conclusion to evaluate their relevance to the research scope and the appropriateness of their technical content. Ultimately, 259 publications were selected for final analysis. Duplicate records were identified and removed through manual cross-verification based on paper titles, DOI information, publication metadata, and cross-database records.

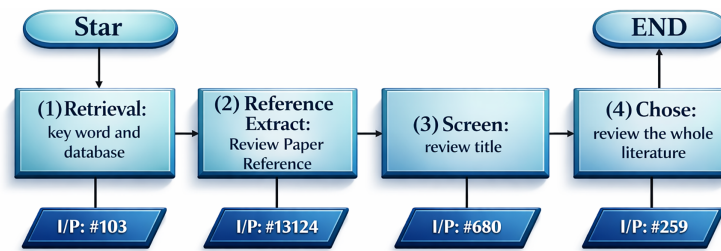


Figure 8: Literature retrieval and multi-stage screening framework under the S_{IGLR} approach.

All literature retrieval, screening, semantic organization, and verification procedures were conducted manually and verified by the author against predefined screening criteria. Therefore, no multi-reviewer voting or conflict-resolution mechanisms were applied. The literature retrieval and screening process was conducted primarily between 20 November 2025, and 15 December 2025.

A total of 259 publications were ultimately selected. The number of studies retained at each screening stage was quantitatively tracked. The final corpus was reorganized according to Control–Cyber Axis, Material–Physics Axis, and HPC–Compute Axis, which correspond to the system pillars of next-generation semiconductor smart manufacturing. These axes are further decomposed based on the employed keyword sets, aligning with the 26 subsections presented in [Section 4](#). This structured approach enables a transparent and traceable quantitative characterization of the literature filtering process, from the initial candidate pool to the final curated research dataset.

Among the initial 13,377 publications ([Table 2](#)), the Control–Cyber Axis accounts for the largest proportion, including Automatic control, robotic, transportation, and suspension systems in fab automation. The Material–Physics Axis encompasses MM, multiscale simulation, and variation analysis.

Meanwhile, HPC–Compute Axis includes HPC, numerical simulation, hybrid models, and Edge–Cloud architectures. [Table 2](#) provides a quantitative overview of the screening process and demonstrates how S_{IGLR} reorganizes the literature by industrial modules rather than conventional algorithm-centric taxonomies to depict the technological landscape in semiconductor smart manufacturing.

Table 2: Multi-stage filtering and system-driven distribution of literature across research axes and subsections.

Categories	Search Keywords	Corresponding Subsection	Retrieve	Filter	Refine
Control–Cyber Axis	ESA, Semiconductor, TSMC, Smart Manufacturing, DT	4.1 ESA for Semiconductor Smart Manufacturing	1696	47	28
	Automatic control systems, Core query	4.2.1 Automatic Control Systems	983	43	8
	Robotic manipulators, Core query	4.2.2 Robotic Manipulators	772	50	10
	Motion control, robotic manipulators, Core query	4.2.3 Motion Control	666	37	15
	Force control, robotic manipulators, Core query	4.2.4 Force Control	402	14	13
	Visual serving control, robotic manipulators, Core query	4.2.5 Visual Serving Control	322	20	11
	Closed-loop, robotic manipulators, Core query	4.2.6 Closed-Loop Control	196	13	9
	Cleanroom, Core query	4.2.7 Contamination Free	109	37	12
	Synchronization and coordination, robotic manipulators, Core query	4.2.8 Synchronized and Cooperative	682	21	12
	Process integration, robotic manipulators, Core query	4.2.9 Integration of Robot Control with Processes	241	27	11
	Vehicle dynamic systems, Core query	4.3 Transport Systems for Smart Manufacturing	113	8	8
	Dynamics, vehicles, Core query	4.3.1 Dynamics Modeling	1039	10	9
	Adaptive control, vehicles, Core query	4.3.2 Vehicle Adaptive Control and MPC	751	21	5
	Multi-objective control, vehicles, Core query	4.3.3 Multi-Objective Control	144	26	5
	Reinforcement learning, vehicles, Core query	4.3.4 Reinforcement Learning-Based Control	408	22	10
	Digital twin, vehicles, Core query	4.3.5 Digital Twin Simulation	80	23	7
Material–Physics Axis	Suspension systems, Core query	4.3.6 Suspension Systems	767	17	8
	Materials modeling, Core query	4.4.1 Materials Modeling	901	25	16
	Multiscale simulation, IC, Core query	4.4.2 Multiscale Simulation	270	35	11
	Surrogate modeling, IC, Core query	4.4.3 Surrogate Model	312	26	7
	Simulation under variability, IC, Core query	4.4.4 Materials Simulation under Variations	239	32	11
HPC–Compute Axis	Reinforcement learning, IC, Core query	4.4.5 Data-Driven Reinforcement Learning	397	28	2
	HPC, Core query	4.5.1 High-Performance Computing for DT	750	22	6
	Cloud computing, HPC, resource scheduling, Core query	4.5.2 Cloud HPC and Resource Scheduling	254	12	6
	Numerical simulation, hybrid models, Core query	4.5.3 Hybrid Models	485	38	11
	edge–HPC, Core query	4.5.4 Edge–HPC Integration	398	26	8

4 Digital Twin for Next Generation Semiconductor Smart Manufacturing

The ESA, MM, and HPC disciplines will become increasingly critical to industrial development as AI advances rapidly. Due to geopolitical tensions, the US promotes manufacturing reshoring initiatives but

faces challenges of skilled labor shortages, cultural differences, and management conflicts. In this context, TSMC has successfully deployed an advanced fab and achieved its yield target through advanced smart manufacturing technologies. Such capabilities continue to strengthen through advances in engineering and materials research. Semiconductor manufacturing processes are highly complex and require ESA to integrate materials, equipment, data, and control strategies to sustain efficiency and yield. Advanced process nodes operate near physics limits, and fabs deploy extensive equipment to execute thousands of steps across multiple concurrent production orders. Even minute variations may be amplified by coupled dynamics, thereby impacting yield. Consequently, AI is employed to support scheduling optimization, fault prediction, and quality control, all of which involve large-scale, high-dimensional data. For example, production automation relies on extensive infrastructure, including OHT, automated guided vehicles (AGV), and autonomous mobile robots (AMR). These function as cyber-physical nodes, with their operational states monitored by sensors and transmitted as high-frequency real-time data to centralized platforms. DT leverages such data to perform simulation analysis, enabling optimization of processes across the fab.

In this context, DT is emerging as a key enabler for next-generation smart manufacturing. By synchronizing the real-time states of physical entities via IIoT and constructing high-fidelity virtual replicas in the digital domain, DT systems enable real-time monitoring, predictive analytics, and closed-loop control optimization. These require extensive mathematical modeling across ESA, MM, and HPC domains, encompassing automatic control, robotic manipulation, autonomous transport systems, and suspension systems. In this context, TSMC represents a leading industrial implementation of such highly complex, multi-scale, and nonlinearly coupled cyber-physical systems. To provide an overview, Table 3 categorizes the 259 reviewed studies according to physical layers and corresponding ESA functions, including modeling, diagnosis, and prediction, and control and optimization. The subsequent subsections further elaborate on each research domain under this framework.

Table 3: Classification of ESA research in semiconductor smart manufacturing across physical layers and system functions.

Physical/ESA	Modeling	Diagnosis & Prediction	Control & Optimization
Materials	[117–132]	[133–148]	[149–153]
Equipment	[154–163]	[164–187]	[188–206]
Robots/AMR	[207–218]	[219–228]	[229–269]
OHT/AGV/AMR	[270–288]	[289–294]	[295–312]
Sensor/IIoT	[313–316]	[317–329]	[330–336]
Infrastructure (HPC/Edge/Cloud)	[337–353]	[354–365]	[366–375]

4.1 ESA for Semiconductor Smart Manufacturing

This study is based on existing reviews to argue that TSMC applies ESA to transform process data into an advantage, thereby widening its lead in advanced process technologies. First, yield diagnosis serves as a critical indicator of process performance and acts as the trigger for intelligent control. Second, VM focuses on real-time quality prediction to ensure product consistency. Third, FDC reduces downtime risk and process variability. Finally, the scheduling engine dynamically adjusts resource allocation and production flow to maximize efficiency. The core of this framework is DT models built from extensive process datasets and powered by AI servers, serving as the backbone of semiconductor manufacturing systems. When yield diagnosis identifies anomalies, the DT system dynamically adjusts parameters; equipment monitoring ensures tool stability; and the scheduling system reconfigures optimal production routes. Each stage is

supported by advanced research in ESA. To establish a direct linkage between research in ESA, MM, and HPC and real-world processes, this study departs from conventional classifications and instead adopts a fab practice perspective. The reviewed studies summarized in Table 4 mainly focus on three functional dimensions: (i) to support real-time factory operation (scheduling), (ii) to ensure yield stability (FDC/VM), and (iii) to enable intelligent decision-making (DT). This reframing not only aligns with the key drivers of process intelligence but also highlights ESA's research potential in the semiconductor industry.

Table 4: ESA research supporting semiconductor smart manufacturing in Section 4.1.

Physical/ESA	Modeling	Diagnosis & Prediction	Control & Optimization
Materials			
Equipment		[164–175]	[188–199]
Robots/AMR			
OHT/AGV/AMR			
Sensor/IIoT			
Infrastructure (HPC/Edge/Cloud)	[337–339]	[354]	

4.1.1 Yield Diagnosis

Yield is the core ESA metric in fabs. When smart manufacturing advances toward the next generation, yield management requires a systematic “effect-to-cause” analysis, starting from defects and tracing backward to identify potential associations with process flows, equipment states, and material properties, thereby providing a quantitative basis for subsequent process tuning and equipment control. In practice, transforming yield loss into a tractable ESA problem hinges on selecting appropriate observables. Compared with scalar yield metrics, the spatial distribution of wafer defects, represented as Wafer Bin Maps (WBM), preserves structural information arising from the interactions among processes, equipment, and materials. Accordingly, academic research focuses on WBM and applies pattern recognition to automatically identify system-level defect signatures. For example, Ref. [170] proposes an automated framework integrating spatial filtering and spectral clustering to identify both convex and non-convex defect patterns. The framework first performs denoising via spatial filtering, then uses spectral clustering and decision tree classification to detect systematic defects in equipment. Validation using real data from a DRAM fab in Taiwan demonstrates that the framework can effectively identify more than four types of complex mixed defect patterns. Although the study does not explicitly use the term WBM, its core objective—analyzing the spatial distribution of wafer defects—is methodologically aligned with subsequent WBM-based pattern recognition studies.

In addition, Ref. [169] addresses the clustered nature of defective dies on wafers and proposes a DBSCAN (Density-Based Spatial Clustering of Applications with Noise)-based framework to automatically detect arbitrarily shaped defect patterns (e.g., ring-shaped, stripe-like, or arc-shaped) in WBM. Evaluations on the real WM-811K dataset demonstrate its effectiveness in identifying complex and mixed defect structures. Furthermore, Ref. [173] targets mixed defect patterns by first applying spatial statistical tests to filter maps exhibiting spatial correlation, followed by unsupervised clustering using Adaptive Resonance Theory 1 (ART1) to classify and label failure patterns. This approach establishes associations between defect patterns and process causes, enabling automated WBM to extract features and significantly reducing fault diagnosis time. Its feasibility has been validated in a fab in Taiwan and integrated into expert systems.

Collectively, these studies suggest that next-generation smart manufacturing will be centered on closed-loop DT architectures. By integrating the DBSCAN-based framework in [169] with the spatial filtering and spectral clustering techniques in [170], discrete defective dies in WBM can be transformed into physically interpretable defect patterns, enabling precise identification of complex anomaly geometries, including convex and non-convex structures. This defect-driven analytical paradigm enables deep integration of spatial product quality data with temporal sensor streams, including Status Variable ID (SVID) signals and FDC time-series data, thereby facilitating a transition from point-based monitoring to Spatio-Temporal analysis. Such capabilities not only support wafer acceptance test (WAT) anomaly localization and cross-tool comparison but also establish an autonomous diagnostic loop spanning detection, root cause analysis, parameter compensation, and scheduling optimization. This framework exemplifies the strategic potential of ESA in enabling the autonomous operation of a “four-in-one” smart manufacturing system that integrates sensing, analysis, decision-making, and control into a unified cyber-physical architecture.

4.1.2 VM for Quality Prediction

In fabs, VM plays a dual role in real-time monitoring and preventive compensation, addressing three limitations of physical metrology tools—particularly in advanced nodes below 5 nm:

- High cost: Physical metrology equipment (e.g., optical critical dimension (OCD), scanning electron microscopy (SEM)) is extremely expensive.
- Metrology latency: Wafers often experience delays ranging from hours to days from queueing to measurement completion, introducing significant risk.
- Sampling limitations: To mitigate latency, physical metrology typically relies on sampling, which fails to capture wafer-to-wafer variability within equipment.

Therefore, VM has emerged. In processes such as chemical mechanical planarization (CMP) and etching, VM leverages real-time equipment data (SVID)—such as gas flow, power, and chamber pressure—to predict wafer processing quality. This enables a transition from offline sampling inspection to full-wafer, dynamic monitoring. Furthermore, VM integrates equipment-level sensing, process control logic, and quality prediction, reflecting a system design paradigm for multi-layer information fusion. By ESA, VM can operate robustly in environments with high variability and stringent precision, becoming an indispensable pillar of quality assurance. For example, a feature is the critical link between low-level sensing and high-level decision-making. By transforming raw sensor signals into highly discriminative inputs, feature engineering enhances the accuracy and robustness of VM models. For instance, Ref. [172] proposes a deep autoencoder architecture tailored for complex, time-dependent, multi-step process signals in wafer fabrication. By incorporating clipping-fusion regularization, the method effectively captures nonlinear deviations while preserving the features and physical meanings of the original data. Experimental validation on etching processes demonstrates its capability to support predictive quality control in next-generation smart manufacturing.

In addition, VM and FDC must be highly coordinated within ESA. For example, Ref. [164] extracts yield-relevant features from large-scale equipment sensor data and applies VM to predict quality and identify equipment anomalies. This approach not only improves prediction accuracy but also enhances equipment utilization, ensures production stability, mitigates yield loss, and prevents unexpected downtime. More importantly, VM predictions must support dynamic compensation, as all processes (e.g., CMP, etching, deposition) are subject to equipment aging and material variability. Compensation is the core of closed-loop control in ESA to ensure system dynamic stability. For example, in multi-layer fabrication, CMP is used to planarize surfaces prior to interconnect formation. However, polishing pad wear induces MRR drift and shift, degrading planarity and affecting downstream lithography focus. To address this, Ref. [165] proposes a deep RL (DRL) R2R controller that uses VM-estimated pad conditions to perform

parameter compensation for subsequent runs. Experimental validation of CMP processes demonstrates that the controller accurately tracks MRR targets and effectively handles process variability. This approach advances DT from static models to adaptive, intelligent control systems, enabling autonomous operation in next-generation smart manufacturing.

As process nodes shrink below 5 nm, fab-scale expansion and IIoT deployment led to exponential growth in data volume and complexity. From an ESA perspective, constructing dynamic models faces challenges, including strong nonlinearity, time-varying behavior, high-dimensional input spaces, reduced modeling efficiency, and diminished interpretability. These impact the accuracy of APC. To address these challenges, Ref. [337] asserts that big data analytics can enhance PDM and VM performance. By leveraging distributed computing architectures and deep learning algorithms, the approach reduces the cost of ownership (CoO) and improves computational efficiency. This advancement optimizes ESA's computational architecture, mitigates the curse of dimensionality, and accelerates dynamic processing, thereby strengthening real-time DT mapping and simulation of physical systems. Ultimately, it delineates a technological pathway toward predictive, optimized, and self-correcting next-generation smart manufacturing systems.

4.1.3 Fault Detection and Classification

FDC ensures uninterrupted operation and is a state identification problem in complex dynamical systems within ESA. Massive data in processes poses challenges for real-time processing. To address this, DT technologies bridge physical equipment and virtual intelligence to enable ultra-low-latency responses from sensing to inference. This capability drives APC toward PdM, realizing next-generation smart manufacturing that prevents failures before they occur. Relevant studies span data processing, model design, edge computing, and DT implementation, illustrating ESA's multi-layer technological evolution in the FDC domain. Photolithography is the most critical and costly step, with advanced high-NA EUV tools costing hundreds of millions of dollars. Operation involves frequent mask changes (often thousands of reticles), and improper management can create bottlenecks, reduce utilization, and degrade work-in-process (WIP) quality due to excessive queue time. Existing heuristic policies are insufficient to handle unexpected events such as tool failures or temporary mask shortages. To address this, Ref. [190] employs a DT-based photolithography simulator to predict the impact of decisions on equipment performance. Real-time production information is incorporated, and a rollout algorithm is used to trial candidate mask-switching policies in the virtual environment; the selected optimal policy is then dispatched to the equipment automation/control system. Benchmarking and statistical validation in a fab show a 13.45% reduction in cycle time and up to a 20% increase in scanner utilization.

During processing in tools such as etchers and Rapid Thermal Chemical Vapor Deposition (RTCVD) systems, sensors record pressure, gas flow, and power at millisecond resolution. These high-frequency data streams are fundamental for DT equipment modeling; however, DL models may suffer from information overload and response latency, hindering real-time decision-making. To address this, Ref. [166] applies the discrete wavelet transform (DWT) to reduce input data volume by over 90% while maintaining approximately 99% accuracy, thereby reducing synchronization latency between DT models and physical equipment. Moreover, for microsecond-scale anomalies (e.g., radio-frequency (RF) sparks), conventional Fourier transforms tend to smooth transient events, whereas wavelet transforms—through “time–frequency localization”—capture such instantaneous signatures effectively, which is critical for FDC as in [164]. Besides, Ref. [166] using tree classification to demonstrate that wavelet features improve both efficiency and accuracy of fault classification compared with raw data and exhibit strong robustness in etching processes. This exemplifies how ESA leverages advanced signal processing to convert high-frequency, high-dimensional data into actionable features for next-generation AI-driven decision systems.

Another challenge in process data is the lack of standardization. Even within the same equipment, processing different wafers leads to variable-length sensor sequences. Traditional FDC methods may lose fine-grained information during compression, and many DL models assume fixed-length inputs, making them unsuitable for SVID data with variable durations. To overcome this, Ref. [167] proposes a model that combines self-attention mechanisms with convolutional neural networks (CNNs) to learn directly from raw SVID data. The approach enables DT models to capture physical phenomena such as RF sparks and transient pressure fluctuations, enhancing simulation fidelity. Experiments show strong robustness across sequences of varying lengths without the need for padding or truncation. This work relaxes ESA constraints on input formats, allowing flexible handling of variable-length, high-frequency signals and improving sensitivity to transient anomalies, thereby strengthening DT capabilities for real-time monitoring and dynamic simulation in highly variable processes.

Semiconductor manufacturing is an ideal testbed for integrating advanced research. For example, in etching, sensor signals exhibit abrupt transitions between process steps. If FDC systems use improperly smoothed features, they may fail to detect subtle anomalies such as RF sparks. Therefore, Ref. [167] uses SVID data to identify RF power module sparking events and indicate equipment fault states. Ref. [172] uses clipping fusion regularization to predict that “the etch depth of the current wafer batch will be approximately 2 nm lower,” prompting downstream parameter adjustment. The integration of these approaches represents a dual-monitoring paradigm—linking product quality and equipment health—and marks a key milestone in advancing DT from predictive simulation to diagnostic closed-loop control in next-generation smart manufacturing.

4.1.3.1 Edge DT and Real-Time Detection

From the ESA perspective, next-generation smart manufacturing requires not only accurate modeling but also multi-level, distributed diagnostic architectures. In the highly dynamic environment, system stability depends on the synchronization quality between physical entities and virtual counterparts. When IIoT drives rapid system-scale expansion, communication bandwidth and computational latency can delay decisions. This poses challenges because DT must remain tightly synchronized with physical systems to accurately simulate equipment behavior for FDC. For example, transmitting high-frequency SVID data (e.g., [167]) to cloud platforms will affect real-time monitoring due to network latency. To address this issue, Ref. [354] presents an equipment-level FDC system for wet-cleaning equipment that leverages edge computing to enable real-time anomaly detection. Through context mapping, equipment logs are transformed into analyzable data for integration with deep neural networks (DNN) to support PdM. This architecture ensures that DT can achieve predictive capability and real-time control with minimal latency.

Furthermore, when equipment-level FDC systems (e.g., [354]) detect concurrent minor drifts across multiple tools, it is difficult to identify yield-impacting root causes. To address this, Ref. [171] proposes a retrospective design of experiments (DOE) to enhance diagnostic capability. Retrospective DOE is a highly automated AI-driven data mining system. It employs pattern-matching and search algorithms to reconstruct large-scale data and extract samples corresponding to specific factor combinations (e.g., high temperature with low pressure, or *vice versa*), followed by feature selection to eliminate irrelevant noise. Subsequently, association rule mining is used to analyze higher-order interactions among variables, and decision trees are employed to generate interpretable diagnostic rules. This approach enables DT to perform spatio-temporal analysis to accurately identify latent parameter combinations responsible for yield deviations within large-scale data.

While edge computing enhances the timeliness of FDC, detected equipment anomalies must be validated to determine whether they lead to actual product defects to avoid unnecessary downtime caused by false alarms. Moreover, individual machines may operate within acceptable tolerances, yet cumulative tolerance effects can result in systematic defect patterns observed in WBM. This transition marks a critical step from FDC toward advanced equipment control (AEC). To this end, Ref. [168] extends the scope of FDC to encompass full-process data, including WIP, metrology, defect inspection, WAT, and chip probing (CP). A hybrid decision tree approach is proposed to automatically identify key influencing factors to improve diagnostic efficiency. This enables DT to simulate not only equipment behavior but also final product performance. The method has been adopted in a fab. Ref. [168] demonstrates a transition from single-point equipment monitoring to full-process system integration and next-generation smart decision-making, thereby exemplifying the practical realization of ESA in complex manufacturing systems.

4.1.3.2 Dynamic DT and Predictive FDC

In highly dynamic environments, ESA faces the challenge of collecting sufficiently comprehensive fault samples for model training. Therefore, Ref. [174] proposes a two-stage diagnostic framework. First, DT is employed for front-running to simulate diverse strategies and potential fault scenarios, thereby training DNN. Second, deep transfer learning (DTL) is utilized to transfer the DT-trained models to the physical domain for real-time monitoring. This approach was validated in complex automated production lines to effectively detect faults. In addition, PdM can manage the equipment lifecycle to mitigate scheduling disruptions caused by unexpected failures. To this end, Ref. [175] proposes the SAFE method (Supervised Aggregative Feature Extraction), which extracts key health indicators from large-scale time-series data in processes to quantify the remaining useful life (RUL) of equipment prior to failure. When degradation in indicators is detected, the system integrates DT to assess risk and trigger maintenance strategies in real time, achieving overall fab efficiency (OFE) objectives and ensuring uninterrupted production.

Collectively, Refs. [174,175] demonstrate a progression from deep simulation to predictive maintenance to address the challenge of insufficient fault samples and advance DT from post hoc diagnostics toward proactive fault prediction. It highlights the role of ESA in maintaining system stability under high-variability processes and enables next-generation smart manufacturing with enhanced autonomous decision-making capabilities and operational resilience.

4.1.4 Scheduling Optimization Engine

The goal of smart manufacturing is to achieve real-time responsiveness, dynamic scheduling, and global optimality. To this end, ESA provides a framework spanning parameter identification, behavioral modeling, and strategy optimization. Confronted with the high complexity and re-entrant nature of semiconductor manufacturing, ESA must address the high-dimensional parameters of individual tools and dynamically integrate equipment operation, resource allocation, and decision strategies to support real-time scheduling and adaptive responses. Recently, studies combining DT, RL, and distributed decision intelligence have begun to redefine scheduling logic, spanning from micro-level equipment parameter tuning to macro-level capacity planning. At the micro-level, cluster tools enable wafer transfer between steps without leaving mini-environments. For example, lithography bundle systems, as discussed in [190], are critical to performance and yield. However, variations often render preset parameters suboptimal. Therefore, Ref. [188] employs DT to simulate real equipment and process environments, integrating physics engines and event-triggering mechanisms to accurately reproduce wafer movements. And Ref. [188] iteratively applies a hill-climbing algorithm to search for optimal parameter configurations, achieving a 41.14% reduction in cycle time.

Similarly, scheduling parameter optimization for etch cluster tools becomes complex due to process diversification. For example, JIT1 (Just-In-Time 1) governs the coordination between robot movement and process initiation. Therefore, Ref. [191] proposes a KNN-based automated modeling and parameter-optimization approach to address time fluctuations in production data. Simulation results show that the method improves equipment performance. Furthermore, the method integrates a chatbot for real-time decision support.

Collectively, Ref. [188] uses physics engines to reconstruct real systems to provide interpretability for parameter optimization. Ref. [191] uses machine learning (ML) to capture time fluctuation and improve robustness. Ref. [199] uses DRL to learn action policies and enable intelligent control. These methods represent different ESA approaches to extend smart manufacturing from efficiency optimization to intelligent decision-making.

In addition, the multi-layer stacking of a chip leads to a re-entry characteristic. An advanced chip can exceed 100 layers, and each layer must be processed by the same equipment. This results in strong nonlinear coupling across layers, which is widely recognized as the most challenging scheduling optimization problem [194]. Therefore, Ref. [194] proposes a fuzzy hierarchical reinforcement learning (FHRL) framework. It employs recurrent reinforcement learning (Recurrent RL) units to precisely control the sub-cycle time (Sub-CT) associated with each layer of the chip, thereby capturing temporal dependencies across layers. A fuzzy reward mechanism is introduced to mitigate schedule fluctuations caused by batch rematching. Experimental results show that the mean deviation (MD) between actual and expected completion times is reduced by 30% compared with conventional methods.

Re-entry also requires the same equipment to simultaneously process wafers at different stages of the product lifecycle. When short-term equipment faults occur (e.g., etch tool anomalies identified in [167]), the system faces decisions about equipment repair and wafer processing priorities. Although ERP/MES systems generate production plans based on order due dates, shop-floor scheduling often deviates from these plans due to real-time equipment conditions and defect risk. This may delay urgent orders and create a mismatch between planned and actual production. Even under full capacity utilization, order delays can still occur. In this context, traditional heuristic scheduling rules are insufficient to resolve resource competition across thousands of steps. Therefore, Ref. [193] applies a deep Q-network (DQN) to optimize scheduling policies, and experimental results show clear improvement over conventional dispatching rules. Ref. [197] further proposes a decentralized framework that combines discrete-event simulation and DRL. Each piece of equipment is modeled as an agent responsible for its own dispatching decisions, with an independent DQN trained using a discrete-event simulation model. Experimental results show that collaborative DRL scheduling improves both local resource utilization and global fab performance.

Re-entry-induced scheduling challenges also arise in advanced packaging processes. For example, die attach and wire bonding reuse the same equipment across different process stages. Similar characteristics are observed in heterogeneous integration technologies (e.g., CoWoS), where multi-component matching and multi-stage path dependencies are required. Conventional RL methods face scalability issues when the number of machines increases and job types become more diverse, leading to significant challenges in function approximation and policy exploration. To address this, Ref. [195] proposes a centralized DRL approach executed by an AI agent. By introducing a novel state representation and continuous action space, the model adapts to varying machine configurations and dynamic production demands. Experimental results show that this method reduces the make span compared with rule-based and conventional RL methods. The AI agent learns scheduling policies directly from system states and dynamically adjusts decisions, enabling efficient scheduling under changing production conditions. This represents a concrete implementation of efficiency maximization in smart manufacturing.

These studies implement ESA from multiple perspectives. Ref. [194] uses hierarchical modeling to mitigate cascading effects caused by local disturbances. Ref. [193] addresses scheduling imbalance caused by plan–execution mismatch. Refs. [195,197] demonstrate, respectively, decentralized and centralized AI agent decision-making in dynamic environments. Overall, scheduling evolves from rule-based methods to learning-based intelligent scheduling.

4.1.5 Fab Macro Scheduling and Dynamic Resilient Decision

From the ESA perspective, fabs are highly complex dynamic systems with strong nonlinear coupling. When the scope shifts from single-tool optimization to fab macro scheduling, ESA's challenge is to establish a production release and real-time scheduling framework to achieve flow balancing and lead-time control under large-scale resource competition and stochastic disturbances. Wafer fabrication involves thousands of steps with repeated visits, which can easily cause tool congestion (WIP accumulation) or resource idling. Production release planning is therefore required to regulate the production rate. Therefore, Ref. [189] develops a DT framework to collect real-time physical process parameters and update the simulation model. Piecewise linear regression is used to fit clearing functions (CF). The CF is used by the simulation model for optimization. Experimental results show that the optimized release plan reduces cost by at least 26% compared with traditional release control methods.

When demand changes rapidly (e.g., urgent orders), the ERP system may issue dense production plans to meet due dates. If shop-floor scheduling cannot absorb these plans while the fab continues releasing lots, delays can propagate to all orders. Long queue time may also degrade product quality (e.g., film oxidation or contamination). Therefore, Ref. [196] uses CNN to extract fab-level production features and applies an asynchronous advantage actor–critic (A3C) algorithm to learn composite dispatching rules (CDR) with multi-objective optimization. The method adjusts release rates and scheduling priorities in response to dynamic events. Experimental results show reduced mean completion time (MCT) and improved on-time delivery rate (ODR) for urgent orders, enabling resilient scheduling. Besides, Ref. [198] focuses on high-mix, low-volume IoT orders and trains scheduling agents using DQN. The system consists of multiple cooperative DQN agents. Experimental results show that the agents optimize the first-come, first-served scheduling for different objectives (e.g., cycle time or due date).

Within ESA, these studies address fab-level control from different aspects. Ref. [189] establishes the basis for flow control through CF and release optimization. Ref. [196] integrates CNN and A3C for real-time scheduling under dynamic events, thereby improving system resilience. Ref. [198] uses multi-agent reinforcement learning to handle diversified order scheduling under high variability. Together, they link structural modeling, data-driven state estimation, and decision optimization to support real-time response and global balance in next-generation smart manufacturing.

4.1.6 Hardware Security and ESA

Counterfeit chips and hardware Trojans pose significant threats to defense supply chains and are a major driver for semiconductor sovereignty [116]. These counterfeit components cause annual losses exceeding USD 100 billion and compromise the security of critical infrastructure, including military, aerospace, and energy systems [376]. Existing inspection standards and procedures are insufficient to detect all types of counterfeit components. In addition to applying advanced blockchain for origin fraud detection and forensics, next-generation smart manufacturing requires more advanced physical and electrical inspection mechanisms to counter increasingly sophisticated counterfeiting techniques.

From the ESA perspective, developing hardware Trojan detection methods beyond traditional heuristic or static analysis requires standardized attack models for validation. To this end, Ref. [338] establishes a benchmarking framework for hardware Trojans and maliciously compromised circuits. Due to variations in process, trigger mechanisms, and design scale, current metrics, such as detection rate, false-positive rate, and test coverage, are insufficient for a fair comparison across methods. Therefore, Ref. [338] models threat across stages of the IC supply chain from register-transfer level to netlist and layout, and implements the framework on platforms such as microcontrollers and cryptographic IP cores. Through vulnerability analysis—ranging from structural analysis to test pattern generation—the method identifies regions in circuits that are most susceptible to Trojan insertion. It further defines quantitative evaluation metrics based on the activation efficiency of test patterns and resistance to side-channel analysis. These results provide the underlying data required for developing physical and electrical detection mechanisms and enable the identification of uncontrolled variables within the supply chain.

Heterogeneous systems-on-chip (SoC) are widely deployed in critical domains such as aerospace, automotive, finance, and healthcare, where security is essential. Current protection methods are often limited to processor cores or rely on manual implementation, leading to inefficiency and potential vulnerabilities. Therefore, Ref. [339] proposes to treat hardware security as a primary objective in high-level synthesis (HLS) and advocates a secure-by-construction (SBC) design paradigm rather than post hoc detection. For example, design flaws in accelerator controllers (e.g., CWE-691) may allow attackers to manipulate inputs, alter execution behavior, or even access sensitive data through shared memory or output interfaces. Under the SBC approach, accelerators can enforce mechanisms such as fixed execution time to mitigate certain classes of physical attacks. Future HLS tools should support the automated synthesis of multiple protection mechanisms to enable heterogeneous architectures that achieve both performance and security.

4.2 ESA for Automated Control and Robotic Manipulators

From the ESA perspective, automatic control systems in next-generation semiconductor smart manufacturing serve as the core interface for cyber–physical integration within a DT architecture. They use sensor feedback and advanced algorithms to regulate physical processes in real time, translating ESA-enabled system-level decisions into precise physical actions. Through high-fidelity virtual replicas, DT supports real-time monitoring, predictive analysis, and control optimization, ensuring tight coupling between control logic and physical phenomena and process stability. They operate continuously with high repeatability and cleanliness, avoiding contamination and human errors. Robotic systems are widely used across operations, including automated material handling systems (AMHS), atmospheric and vacuum robots, and precision tasks in back-end processes such as wire bonding and pick-and-place operations. These applications ensure precise control from process operations to material flow. As primary actuators in automatic control systems, robotic manipulators function as execution and feedback nodes in wafer fabs. Their control logic integrates closed-loop feedback and visual serving to achieve micrometer positioning accuracy. In addition, force and contact control enable manipulators to perform fine operations under constrained conditions and in the presence of uncertain disturbances. IoT sensors continuously stream joint-level data to support DT model calibration and prognostics. To improve efficiency, robotic systems are evolving toward multi-arm coordination and arm-to-machine integration, enabling obstacle avoidance and real-time error correction. Within ESA, scheduling systems apply RL to adjust dispatching policies based on low-level anomaly signals from robotic systems, supporting autonomous operation in smart manufacturing.

The reviewed studies in this section are summarized in Table 5, which categorizes the literature by physical system levels and ESA functional roles (modeling, diagnosis and prediction, control, and optimization), and illustrates how ESA supports automatic control and robotic systems.

Table 5: Automatic control and robotic manipulators research supporting ESA in semiconductor smart manufacturing.

Physical/ESA	Modeling	Diagnosis & Prediction	Control & Optimization
Materials			
Equipment	[154,155]	[176–182]	[200–206]
Robots/AMR	[207–215]	[219–228]	[229–269]
OHT/AGV/AMR			[295–299]
Sensor/IIoT		[317–322]	[330–332]
Infrastructure (HPC/Edge/Cloud)	[340–343]	[355–360]	[366]

4.2.1 Automatic Control Systems

From the ESA perspective, semiconductor manufacturing constitutes a complex scheduling problem characterized by strong coupling, multi-level constraints, and resource competition. In this context, automatic control systems have evolved into control networks capable of distributed sensing, local decision-making, and cross-layer coordination in smart manufacturing. As advanced processes below 2 nm impose stringent requirements on stability and accuracy, conventional control strategies based on static rules or fixed parameters are insufficient to address high-dimensional, nonlinear, and time-varying process variations. Recent studies have therefore shifted toward control architectures that integrate Industrial AI and DT to enhance adaptability to dynamic uncertainty and overall resilience. Ref. [330] indicates that wafer fabrication is the most advanced and complex technology, where automatic control is a fundamental component. In addition to highly automated material-handling systems that transfer wafers through hundreds of steps, wafer selection and placement are also performed by robotic manipulators. Both robotic and material-handling systems require extremely high precision, and any deviation or delay can affect operational goals.

4.2.1.1 Adaptive Control (Single-Tool Intelligence Layer)

While Ref. [330] ensures precise execution at the equipment level, maintaining such precision in dynamic environments remains a challenge. For example, long-term operation under vacuum conditions may alter joint friction characteristics; handling FOUP of different specifications may introduce abrupt changes in inertial load; and minor bearing wear may cause positioning drift. To compensate for these dynamic deviations, adaptive control must be incorporated to enable autonomous error correction in robotic manipulators, ensuring that equipment maintains high precision under nonlinear variations and accurately executes control strategies derived from ESA. In processes, wafers traverse hundreds of nodes, including photolithography, etching, and cleaning tools, forming highly complex discrete routes. This is analogous to job shop machining environments characterized by high uncertainty and complex routing. In this context, Ref. [200] identifies those unpredictable events—such as urgent order insertion, material shortages, and equipment failures—that disrupt predefined schedules, and proposes a function-block approach to generate adaptive adjustments. The method demonstrates strong responsiveness and fault tolerance under uncertainty, enabling production equipment to perform real-time autonomous control based on shop-floor conditions. Ref. [201] further proposes a future smart manufacturing framework incorporating intelligent machine agents, self-organizing models, and adaptive models. Each machine is modeled as an intelligent agent with capabilities for data acquisition and autonomous control. The adaptive mechanism monitors process execution and autonomously handles anomalies when occurring. Prototype validation shows that dynamic machine reconfiguration can achieve optimal resource utilization.

4.2.1.2 System-Level Control and Edge Computing

Fabs deploy extensive transport systems and robotic manipulators. Beyond single-tool automation, large-scale coordination across hundreds of tools must be achieved through ESA. Edge computing must be tightly integrated into system-level control to enhance distributed decisions of IIoT. This integration improves real-time responsiveness and adaptation efficiency while handling high-dimensional, nonlinear variations. In this context, data flow and response among controllers, edge nodes, and equipment become critical. Robotic manipulators and tools serve as key nodes for integrating control with edge intelligence. To enhance responsiveness, Ref. [340] embedded ML at edges to rapidly extract critical information, and employed DRL to realize scenario-aware Dynamic Adaptive Planning (DAP). This approach enables edge devices to autonomously respond to high-dimensional nonlinear variations and to perform optimized scheduling and control adjustments. Experimental results show that DAP improves IIoT's responsiveness to dynamic edge environments.

In addition, automated control logic (e.g., PLC/DCS programs) uses Function Blocks (FB) as the fundamental units for encapsulating control and decision logic. ESA configures, coordinates, and integrates FB to support complex process planning and adaptive control. This allows individual tools to adjust control actions based on real-time feedback data to correct execution deviations. For example, Ref. [154] indicates that FB enables distributed control systems to support distributed dynamic decision-making, thereby enhancing the autonomy and intelligence of robots. This study proposes a dual architecture comprising Supervisory and Operation Planning to enable bidirectional information flow between the centralized control and individual nodes.

4.2.1.3 Automatic Control and Digital Twin

The objective of automatic control systems is to establish a robust, safe global protection framework. As process scale complexity increases, minor errors or transient interruptions may lead to large-scale coordination failures. Therefore, integrated automation is required to enhance security and stability. Therefore, Ref. [341] uses DT to construct Cyber-Physical Production Systems (CPPS) to minimize latency between data acquisition and model generation. CPPS integrates isolated data-collection schemes into a unified system to enable near-real-time production control for large-scale coordination. In addition, Ref. [341] introduces the concept of inherent data security to protect data by separating volatile data from master data, thereby preventing the leakage of intellectual property (IP) associated with processes. Moreover, Ref. [342] indicates that DT enables the simulation, prediction, and optimization of physical systems, thereby enabling proactive protection before failures occur (e.g., predicting robot trajectory conflicts). When ESA generates global control strategies, DT validates safety and dynamically detects anomalies during execution to enhance control integration and enhance process safety. Ref. [202] proposes Digital Engine to leverage high-fidelity simulation to address the high trial-and-error cost and the lack of physical data when training AI agents on production lines. It trains DRL algorithms within a digital environment to achieve near-synchronous control through a communication interface network. Digital Engine acquires process knowledge and predicts faults, enabling AI agents to identify optimal actions and scheduling policies prior to deployment, thereby improving control robustness and autonomous decision-making under dynamic conditions.

4.2.2 Robotic Manipulators

Robots must adapt in real time; otherwise, small local errors may propagate and amplify into system losses due to complexity and dynamic variations of processes. Therefore, robots can adjust their behavior through imitation learning and RL, enabled by advances in AI and sensing. In addition, robots are

constrained by global objectives and resources and must coordinate with process planning, scheduling, equipment states, and logistics. Therefore, robots must operate as perceptive, adaptive, and cooperative AI agents. This defines the direction of next-generation smart manufacturing. DL and DT are integrated to provide validation and pre-evaluation of system behavior for ESA decisions. This shifts automatic control from conventional feedback-level regulation to task-level autonomous decisions to maintain process stability under local disturbances. This is the basis for the literature organization in this study.

4.2.2.1 Robotic Manipulators in High-Precision Processes

Many studies do not explicitly target wafer fabrication, but they exhibit functional equivalence in engineering conditions, including cleanliness, precision, consistency, and system integration. For example, Ref. [229] proposes a generic, self-assertive robotic skill model for highly automated manufacturing, along with executable code for mobile manipulators. Experiments show that analyzing standard operating procedures (SOPs) can extract a set of robot skills during new equipment deployment. These skills are modular, parameterizable for different environments, and support cyber-physical validation after execution. In addition, robots can handle environmental uncertainty by integrating sensing and acting. Complex control logic is encapsulated within skills, reducing the need for expert intervention. This skill-based architecture is considered an enabler for transformable factories.

In addition, Ref. [230] proposes a learning-based auto programming method to construct robot skills through observational learning. The approach uses workspace modeling and simulated retargeting to plan robot actions, and applies RL to pre-train skills for handling uncertainty and variation. The method is validated on production lines. Ref. [219] develops an online fault-diagnosis system for robots based on transfer learning (TL). A hybrid model combining 1D multi-channel CNN and recurrent neural networks accurately detects joint conditions. This enables self-awareness and predictive maintenance in collaborative robots and supports future self-healing architectures. Experimental results show that this method achieves 99.03% accuracy in joint fault detection. From an ESA perspective, such online monitoring and diagnosis mechanisms enhance system stability and predictability, preventing local failures from propagating into system disruptions. Ref. [203] analyzes the evolution of control architectures in automated production lines. The study shows that robots operate within highly interconnected control networks composed of controllers, sensors, and actuators, and are constrained by global objectives and resources. Modular, reconfigurable control architectures enable cross-station coordination and system optimization, thereby enhancing flexibility, stability, and performance.

Refs. [231,232] focus on flexible production lines different from wafer manufacturing. However, Ref. [231] proposes a reconfigurable autonomous mobile manipulator (AMM) architecture that integrates manipulation capabilities and edge computing for real-time processing and decision-making in complex environments. Virtual simulation and real-world experiments validate stability and accuracy, reducing deployment risk. From the ESA perspective, such cyber-physical validation and system risk pre-evaluation align with the requirements of cross-station coordination and system risk control in advanced wafer fabrication. Ref. [232] designs a sixth-generation autonomous industrial mobile manipulator (AIMM) and proposes a skill-oriented control framework integrated with CPS. This approach transforms robots from isolated control units into system-configurable and verifiable engineering entities. Using skill-based structures, robots execute modular tasks such as pick-and-place operations. A unified network connects sensors and robots to acquire and analyze cross-equipment data, thereby optimizing processes. Experimental results demonstrate high stability and precise execution in automated material handling tasks.

4.2.2.2 Digital Twin and Real-Time Scheduling

Ref. [19] reviews robot learning in smart manufacturing, including imitation learning, policy gradient, value-based learning, actor–critic methods, and model-based learning. These methods are applied to robotic grasping, assembly, process control, and human–robot collaboration. Learning-based approaches enable robots to identify unknown states in complex environments and improve robustness under stochastic conditions. The study also discusses the use of DT to reduce the simulation-to-real gap.

Refs. [295–298] focus on AGV scheduling in dynamic and complex environments and address real-time decision, energy constraints, and multi-agent coordination. Ref. [296] proposes a DT-driven DRL framework for robot scheduling and path planning under frequent process changes. It uses a Markov decision process (MDP) to formulate the AGV scheduling problem and combines an improved double deep Q-network (ID3QN) for discrete scheduling. Meanwhile, deep deterministic policy gradient (DDPG) supports continuous control for obstacle avoidance and path adjustment. The DT provides a synchronized virtual–physical environment for policy learning and optimization. Experimental results show reduced delay and energy consumption. Ref. [295] proposes a DT decision-support method that accounts for path conflicts, system failures, and battery constraints. This method uses Q-learning to generate scheduling plans and virtual–physical synchronization to enable rapid validation. Experimental results show improved battery utilization and reduced operation time. Ref. [297] proposes an adaptive DRL scheduling method using an MDP to model task volume, waiting time, transport distance, and AGV state. Hybrid decision rules are introduced. Experimental results show that this method outperforms genetic algorithms and rule-based scheduling. Ref. [298] proposes a multi-agent deep deterministic policy gradient (DDPG) algorithm based on an MDP formulation and obstacle avoidance. Experimental results show that this method outperforms GA, DQN, and single-agent DDPG. Although these studies are conducted in container terminal environments, their key characteristics—high task concurrency, strict timing constraints, and system-level coordination—closely align with AMHS and OHT scheduling problems in wafer fabs.

4.2.3 Motion Control

Robotic motion control is grounded in mathematical theories, including forward and inverse kinematics, dynamic modeling, motion planning, feedback control, Jacobian-based control, and planning algorithms such as PRM and RRT. Recent advances integrate DL to support motion control strategies, enabling robots to address inverse kinematics and obstacle avoidance in dynamic, uncertain environments. In addition, the cascading effects of motion control strategies in large-scale heterogeneous systems are becoming increasingly important. In advanced wafer fabrication, local errors may couple and propagate, affecting overall stability and yield. Therefore, this study examines the role of robotic motion control in next-generation semiconductor smart manufacturing. Within ESA, motion control is a critical subsystem that influences cross-station coordination, process continuity, and the propagation of system risk. Inverse kinematics, motion planning, and learning control form the engineering foundation for advanced wafer fabrication.

4.2.3.1 Contact Tasks and High-Precision Motion Control

In advanced processes, including back-end advanced packaging, precision assembly, and material handling, there are challenges such as ensuring compliance and reducing damage risk. These tasks require consideration of mechanical coupling, manipulator impedance, physical constraints in dynamic interaction, and unified motion planning. Therefore, Ref. [233] proposes an impedance control framework with a conceptual mathematical formulation that achieves Cartesian end-effector impedance for nonlinear manipulators. The control law shapes desired impedance behavior, enabling compliant, bio-inspired interaction and

preventing excessive contact forces under rigid constraints (e.g., door opening). This is applicable to wafer alignment, probe contact, and precision bonding. During wafer handling, robot–equipment interaction requires spatial decomposition to enforce constant force along constrained directions while maintaining precise position control along unconstrained directions. Therefore, Ref. [234] proposes a hybrid position/force control scheme that integrates force/torque and position feedback to satisfy trajectory requirements under constrained motion. The task space is decomposed into orthogonal subspaces assigned to position and force control. The method requires real-time sensing of environmental interaction. Experimental results demonstrate stable performance under step and ramp inputs.

For probe stations and wafer probers in wafer electrical testing (wafer sort/chip probing), operations such as probe contact, wafer alignment, and precision bonding require low damage and high consistency. Therefore, Ref. [235] decomposes robot motion into compliant motion and guarded motion (contact transition). Compliance is achieved through mechanical design or control algorithms. Force control guides motion by using task geometry to define force, direction, and position constraints, and by applying the orthogonal constraint principle to address model uncertainty. With advances in AI, this work builds on the impedance control framework in Refs. [233,236] and uses RL to dynamically adjust impedance parameters to address uncertainties in contact, model errors, and precision alignment issues. This study proposes a framework that integrates RL with an operational space controller to enable robots to exhibit adaptive, compliant behavior in contact-rich assembly scenarios that support contact exploration. Experimental results show robustness under small environmental variations (e.g., positional deviations) in contact-rich assembly tasks.

4.2.3.2 Contact Tasks and High-Precision Motion Control

Recent research on trajectory planning and obstacle avoidance in dynamic environments integrates real-time perception, model constraints, and learning-based decision-making. These frameworks enable robots to resolve path conflicts, equipment occupancy, and state uncertainty without trial-and-error or process interruption. Such methods align with trajectory planning and collision avoidance in fab automation. In advanced processes, AMHS, OHT, and cleanroom transport involve dynamic equipment states, human–robot coexistence, cross-station logistics, and continuous operation. Motion control must generate feasible, smooth trajectories under high density, strong constraints, and low fault tolerance, without disrupting the overall process. For example, Ref. [207] proposes a novel dynamics modeling approach that combines physical priors with data-driven learning. The method learns accurate manipulator dynamics directly from trajectory data using an enhanced Deep Lagrangian Network with a Feedforward Neural Network. This hybrid modeling framework enforces physical consistency in the learned dynamics (e.g., energy conservation) and compensates for non-ideal effects not captured by the explicit physical model, such as friction and unmodeled disturbances. Experimental results show that the method applies to multibody robots, legged robots, and autonomous vehicles, achieving high accuracy and good generalization, and demonstrating its potential as a physics-consistent learning model for complex dynamical systems.

Ref. [237] proposes a robotic manipulator motion-planning method that combines recurrent neural networks (RNN) and virtual fencing for obstacle avoidance and target tracking. Level set functions are used to define virtual fences, converting complex obstacles into geometric inequality constraints. Task requirements, physical limits (e.g., joint angles and velocities), and spatial constraints are formulated as a quadratic programming problem. An RNN-based dynamical model with parallel computation capability is used as a real-time solver to enable continuous-time convergence to constraint-satisfying feasible solutions. Experimental results show that the method ensures obstacle avoidance while maintaining joint states (angles and velocities) within safe limits. For path planning of robotic manipulators in complex, cluttered

environments, Ref. [238] proposes an improved Rapidly-Exploring Random Tree (RRT) algorithm. The method replaces random sampling with hybrid constrained sampling and integrates an Artificial Potential Field (APF) to guide tree expansion. Bidirectional trees are constructed in Cartesian and configuration spaces to effectively support local path planning and replanning. Experiments on real robotic manipulator platforms show that it can serve as a baseline path planning module for non-contact motion phases in cleanroom environments or within equipment. For redundant-degree-of-freedom manipulators (e.g., wafer handling robots in EFEM systems) operating in constrained spaces and dynamic environments (e.g., dynamic obstacle avoidance within equipment), Ref. [239] proposes a framework that integrates offline path planning with online motion control. Potential fields are used to generate an initial feasible path that avoids static obstacles and adapts to dynamic obstacles. Bézier curves are then applied to smooth the planned trajectory, reducing velocity and acceleration discontinuities during motion. Simulation results show that joint angular acceleration can be reduced by approximately 90%, effectively extending hardware lifetime. In the online phase, redundancy resolution and a damped least-squares method are used to compute the inverse Jacobian and avoid kinematic singularities. Experimental results demonstrate that the method achieves highly smooth motion and effective dynamic obstacle avoidance. Traditional sampling motion planning algorithms (e.g., RRT) may suffer from slow convergence and insufficient real-time performance in high-dimensional configuration spaces. Therefore, Ref. [240] proposes goal-oriented sampling to reduce the time required to find an initial feasible path. Experiments on a six-degree-of-freedom (6-DOF) robotic manipulator demonstrate that the method achieves faster path generation in complex environments.

4.2.3.3 Pose Error Compensation and Nanometer Precision Correction

Wafer and reticle stages are parallel or quasi-parallel mechanisms (e.g., Stewart-like or flexure-based structures). In metrology systems such as CD-SEM, overlay metrology, AFM, and e-beam inspection, the relative position between the probe or measurement head and the wafer must remain statically ultra-stable. However, error modeling of parallel manipulators is highly complex and difficult to describe accurately using analytical models. Therefore, Ref. [208] proposes a CNN-BiLSTM-SE model for pose error prediction and compensation. CNNs are used to extract spatial error features across different robot configurations. Bidirectional long short-term memory is used to capture bidirectional dependencies in pose sequences, improving prediction continuity and accuracy. A squeeze-and-excitation module is introduced to weight feature importance, enabling the model to focus on dominant error factors. Experiments on a 6-UPS parallel manipulator show an 88.6% reduction in pose deviation. In addition, Ref. [209] proposes a hybrid algorithm combining particle swarm optimization (PSO) and ant colony optimization (ACO) to compensate for geometric parameter errors caused by link deformation in industrial robots, thereby overcoming the tendency of single algorithms to become trapped in local optima. The method is built on the classical Denavit-Hartenberg model to formulate the robot error model. A PSO variant with linearly decreasing inertia weight and stochastic perturbation enhances population diversity. ACO provides a global search capability. Experimental results show that the absolute positioning error is reduced from 45 to 0.45 mm.

4.2.3.4 Inverse Kinematics and High-DOF Manipulator Control

In wafer fabrication, inverse kinematics (IK) and high-DOF motion control are fundamental for precision subsystem stability, such as wafer alignment, probe contact, and in-chamber positioning. Ref. [241] proposes a framework that combines neural networks (NNs) with numerical error minimization to address the challenges of obtaining accurate, efficient solutions to the IK problem for general six-axis robots. The joint space is partitioned into multiple subspaces for piecewise fitting. Forward kinematics (FK) is used to generate training data for each subspace, and multiple multilayer perceptron (MLP) networks are trained

to improve approximation accuracy. A Newton–Raphson iterative refinement improves the initial solution. Simulation results show position errors below 0.001 mm and orientation errors below 0.001 rad, with an average computation time of 0.0218 s per target pose. Ref. [242] proposes an LSTM method for solving IK in a 5-DOF collaborative robot. An accurate analytical FK model is established based on the robot's physical structure. The FK model generates large-scale pose data for training, and the LSTM learns a mapping from end-effector poses to joint angles. Experimental results show higher efficiency than conventional numerical IK methods. Ref. [243] uses CNN to solve the IK problem for a 7-DOF manipulator with a linear guide. D–H parameters are obtained to define the workspace, and MATLAB is used to generate kinematic datasets. These data are used to train the CNN model, demonstrating that neural networks can effectively handle intensive matrix inversion operations.

4.2.3.5 Robotic Control-Level DT

In high-precision manufacturing, applications involving real-time control of robotic manipulators, contact tasks, and precision motion are often affected by difficulties in controller parameter tuning, model uncertainty, and the high cost of physical trial-and-error, thereby degrading system performance. ESA concerns how such local control decisions propagate through system coupling to affect overall stability and risk transmission. Accordingly, recent studies introduce DT at the control level and motion level to enable virtual models to synchronize with the real-time dynamic states of robotic manipulators. These DT serve as core platforms for controller parameter optimization, motion behavior validation, and simulation-to-real transfer. They allow pre-evaluation of control strategies without disrupting physical production, reducing the risk that local errors propagate into system failures. For example, Ref. [244] applies DT at the controller level to address overshoot and tracking error in integer-order Proportional-Integral-Derivative (PID) control for high-precision manufacturing. The framework integrates DT with fractional-order control (FOC). DT is used to build a 3D visual model to enable cyber–physical monitoring, while real-time data analytics ensure accurate state alignment between virtual and physical systems. Fractional-order calculus is used to enhance control flexibility, and an improved particle swarm optimization algorithm is used to identify optimal FOC parameters. Experimental results show that overshoot is reduced by 42%, elevating DT from a passive monitoring tool to an active engineering component for control decision-making and system optimization.

4.2.4 Force Control

In high-precision manufacturing and cleanroom automation, robotic manipulators must operate under uncertain contact conditions and satisfy system constraints, including low damage, minimal error amplification, and stable cycle time. Force control, compliant control, and contact-task learning are subsystems at the local control level in ESA. Their stability and error compensation propagate through system coupling, affecting overall reliability and yield. Recent studies integrate compliance/impedance control with RL, imitation learning, and multimodal sensing to enhance the adaptability and robustness of the control system. Although most approaches are validated in industrial assembly or electronics manufacturing rather than in wafer fabs, their control logic is well-suited to cleanroom environments and high-precision processes, demonstrating strong cross-system transferability.

4.2.4.1 Force and Compliance Control

Compliance (admittance) control computes a desired displacement from measured interaction forces and executes it via a position controller; it can be regarded as a variant of force control. For example, Ref. [245] proposes an adaptive admittance control scheme integrated with RL for time-varying environments. Critic

learning is used to estimate the value function and update admittance parameters. A cost function that combines interaction forces and trajectory-tracking error is defined. RL is used to implicitly approximate the optimal control law associated with the discrete-time Algebraic Riccati Equation to achieve optimal compliant motion of the robot. A Radial Basis Function Neural Network (RBFNN) adaptive controller is further employed to compensate for dynamic uncertainties. Experimental results show compliant interaction with the environment and effective suppression of excessive contact forces. When input saturation (e.g., actuator force limits) occurs, the control performance of n-DOF manipulators will degrade, leading to increased overshoot and tracking errors. Therefore, Ref. [246] develops an adaptive neural network impedance control scheme to regulate the impedance between the end-effector and the environment. A RBFNN is used to approximate and compensate for unknown disturbances. Both state-feedback and output-feedback controllers are designed. Experiments demonstrate robust tracking performance under saturation constraints. To meet stringent requirements for motion speed and impact safety in advanced processes, Ref. [247] proposes an algorithm that combines variable admittance with adaptive fixed-time sliding-mode control. The outer loop implements the admittance controller to adjust compliance parameters online, providing flexibility and reducing impact forces during contact. The inner loop employs adaptive control to ensure fast, accurate tracking of the modified trajectory, and a nonsingular sliding-mode manifold is used to avoid controlling singularities. Experimental results show rapid convergence of errors and improved dynamic response in dynamic environments.

4.2.4.2 AI for Force Feedback

In advanced processes, wafer probers and advanced packaging systems operate as closed-loop vision-force systems, and model error compensation is inherent. Accordingly, Ref. [248] proposes a residual RL strategy that fuses visual and force-sensing information. Using Red, Green, and Blue (RGB) images as input, DRL is used to determine the robot's Cartesian position increments. A modified parallel force/position controller is adopted to ensure stable contact operations. The approach is validated through simulation and on real robotic platforms to demonstrate robustness under model uncertainty and strong transferability of the residual compensation. In addition, for force estimation and compensation, Ref. [249] uses generalized momentum theory and joint motor torque to construct a disturbance observer to estimate end-effector contact forces. A compensation scheme is proposed to improve estimation accuracy. CNN is applied for data-driven calibration and compensation of the estimated forces. The method is validated on a physical robotic system.

4.2.4.3 Imitation Learning Frameworks

For contact-rich tasks, Ref. [250] adopts a Generative Adversarial Imitation Learning (GAIL) framework to analyze force sensor recordings and emulate production-line operational behaviors. To improve adaptability in contact-rich tasks, Ref. [251] propose a residual Learning from Demonstration framework that learns correction policies to incorporate as compensation terms in Dynamic Movement Primitive base trajectories to handle uncertainty. Experimental results show a significant reduction in robot learning time. To meet stringent reliability and cycle-time requirements, Ref. [252] proposes Super-Human Insertion using the Learning from Demonstration framework. It employs Deep Deterministic Policy Gradient from Demonstrations for deep optimization and asynchronous training, enabling on-policy correction during execution to improve learning efficiency. The framework achieves a 99.8% success rate on the National Institute of Standards and Technology Assembly Benchmark and outperforms rule-based methods in both speed and reliability. For high-precision assembly, Ref. [253] proposes an RL framework to use trajectory optimization as a semi-supervisor. Q-values derived from optimal control are used to accelerate Critic

network training and reduce convergence time. The method is validated in simulations and real robotic platforms. For position-controlled robots, Ref. [254] uses DRL to dynamically adjust robot compliance to actively tune force control gains along predefined task trajectories. Experimental results show that active DRL control outperforms conventional passive Remote Center Compliance mechanisms in high-precision tasks.

4.2.4.4 Multimodal Sensing and Predictive Learning

For contact-rich manipulation of deformable objects, Ref. [255] proposes a deep predictive learning approach integrating vision, tactile sensing, and action to address manipulation challenges. Convolutional architectures are used to process high-dimensional tactile data to enhance dynamic response. A SoftMax transformation is applied for action prediction to make manipulation behavior smoother and more natural. Experimental results show improved task success rates. In addition, Ref. [204] proposes a learning-based force control method using a neural network inverse model to address nonlinear control challenges in Twisted String Actuators. DNN's inverse model maps desired forces to motor voltage commands and automatically compensates for unmodeled dynamics such as non-ideal helical deformation, friction, and limited string stiffness. Experimental results demonstrate superior performance compared to conventional feedback linearization methods. Ref. [256] reviews DRL applications in robotic manipulation and analyzes how to combine DL with RL to address high-dimensionality and data scalability challenges in robotic manipulation tasks. The study highlights that DRL can be integrated with imitation learning, generative adversarial networks (GAN), and meta-learning. However, improving sample efficiency remains a key challenge for deploying DRL in real-world robotic systems. Although most learning-based force control and imitation learning methods are validated in assembly rather than wafer fabs, the control logic—such as contact stability, uncertainty compensation, and error-sensitive motion adaptation—is highly compatible with cleanroom and advanced processes.

4.2.5 Visual Servoing Control

In high-precision manufacturing, visual servoing control is the key sensing and outer-loop mechanism for precision alignment, error correction, and pose compensation. In semiconductor processes, typical applications include wafer pre-alignment and notch detection, pose measurement of wafers, reticles, and substrates, and micron- to submicron-level alignment correction in probe stations and bonding equipment. In ESA, because measurement errors, latency, and uncertainty can propagate through system coupling to affect overall stability and yield, visual servoing control forms a hierarchical closed-loop architecture of “sensing–correction–control.” In addition, Deep learning is introduced into a sensing and observation enhancement layer, which encompasses image feature extraction, landmark detection, pose and error estimation, and compensation for environmental and sensor drift. These approaches improve the robustness of visual servoing control against occlusion, noise, and illumination variations and enhance the accuracy and stability of visual measurements. This aligns with the stringent requirements for stability, verifiability, and process auditing in wafer fabs and cleanrooms.

Ref. [220] proposes a DL pose prediction method guided by image similarity to improve the accuracy of robotic visual servoing. The method combines an L1 loss function (absolute error between the predicted pose and the ground truth) with Mean Similarity Image Measurement. In addition, a Residual Neural Network (ResNet-152) is adopted as the backbone to achieve high-precision, low-dimensional pose regression. The CNN-predicted pose is applied to position-based visual servoing (PBVS) to achieve decoupled control of the manipulator. Experimental results demonstrate strong robustness to occlusion disturbances. Ref. [221] proposes an image visual servoing method integrated with DL. CNN are used to construct a two-stream network as the main framework for visual alignment and feature extraction. A neural network is employed

to model the complex nonlinear mapping from image space to task space. Experimental results demonstrate accurate visual servoing prediction for a 4-DoF robotic manipulator.

Ref. [223] proposes an architecture combining DRL, GAN, and visual servoing to address the sim-to-real “reality gap” in DRL for robotic grasping. Proximal Policy Optimization is used in simulation to train grasping policies to avoid potential hardware damage during real-world training. Visual servoing is introduced to correct end-effector pose estimation errors from DL models, thereby mitigating the positioning inaccuracies observed in end-to-end networks. Comparative results against pure DRL + GAN and pure VS approaches demonstrate the superior robustness of the integrated framework. Ref. [224] employs DRL to adaptively adjust controller gains to maintain feature visibility to address feature loss caused by limited camera field of view (FOV) in classical image-based visual servoing (C-IBVS). This study employs the Deep Deterministic Policy Gradient (DDPG) algorithm to adaptively adjust controller gains in a continuous action space. Visibility constraints are incorporated into RL as penalty terms to guide the robot away from feature-loss regions. Experimental results show superior performance in high-risk scenarios where features approach image boundaries. This method is applicable to logistics robots and autonomous systems with visual navigation. Ref. [225] redesigns a pretrained CNN for direct visual servoing, suffering from limited convergence domains. The classification layer is replaced with a regression layer to estimate the relative pose between the current and target images. The network output is directly used as input to a PBVS controller to guide robot motion. Experiments on a 6-DoF manipulator demonstrate sub-millimeter positioning accuracy.

For human–robot collaboration, Ref. [226] develops a system integrating IBVS, human hand prediction, and Bayesian repulsive pose estimation to ensure optimal grasping efficiency and safety. Monte Carlo Dropout is used to treat ResNet as a Bayesian approximation, enabling uncertainty estimation. The Bayesian DNN reduces the risk that robots will take potentially unsafe actions when encountering unseen states not covered by the training data. By modeling weight uncertainty, the system can assess the reliability of predictions and prevent the robot from moving to invalid positions outside its reachable workspace. Experimental results show prediction interval coverage probabilities of 0.84, 0.94, and 0.95 in the x -, y -, and z -directions, respectively. Ref. [299] uses the DDPG algorithm to develop an RL image-based visual servoing controller for visual tracking and guidance tasks in multirotor UAVs. Effective transfer from simulation to real flight platforms is demonstrated. Ref. [258] develops a vision control method for an eye-in-hand configuration. Bounding box parameters predicted by a CNN are used as control inputs. An LSTM network is employed for temporal smoothing to mitigate chattering in bounding box outputs. A Lyapunov-like approach is adopted for rigorous stability analysis of the closed-loop system. Experiments demonstrate strong tracking performance of the CNN controller.

Ref. [259] proposes an end-to-end visual servoing strategy based on RL. A neural network extracts features from target detection outputs. The distance between the current state and the cluster center is computed to define the state’s novelty, which is then converted into an intrinsic reward to encourage the agent to explore unknown regions of the state space, and is finally mapped into robot control commands. The Soft Actor-Critic algorithm is used to ensure stability and diversity of the control policy. Experimental results show a 0.35 improvement in reward and an 8.0% increase in task success rate compared to the baseline algorithm. Ref. [257] proposes an autonomous mobile manipulation system integrating coordinated control, visual/force servoing, and reactive task control. The combination of visual servoing and force control significantly improves success rates in precision insertion tasks. A high-level executive is combined to enable error recovery when constraint conflicts are detected. Experimental results demonstrate high reliability and robustness under tight tolerance requirements. For occluded and space-constrained environments, Ref. [222] proposes a computer vision–assisted system. Efficient 3D object registration methods and visual algorithms

are employed for precise geometric alignment and gap measurement. Ant colony optimization and virtual reality are employed for assembly path planning and validation. A dedicated machine vision inspection system is developed to provide high-precision feedback without interfering with components. Case studies demonstrate the approach's feasibility.

4.2.6 Closed-Loop Control

Industry 4.0 transforms manufacturing systems from linear, automation-oriented paradigms into highly interconnected, data-driven, and closed-loop architectures [355]. For example, advanced semiconductor manufacturing is a highly coupled, cross-layer, closed-loop engineering system. From wafer processes, equipment states, and measurement feedback to process decision-making, the physical and virtual domains form a continuous feedback loop (for example, in real-time process control). This enables semiconductor fabs to become leading implementations of closed-loop manufacturing. As process technologies advance, system coupling effects are further amplified, necessitating CPS centered on DT and AI as the foundation of semiconductor smart manufacturing to enable process monitoring, anomaly detection, and quality prediction. From the ESA perspective, the challenges of closed-loop smart manufacturing include cross-layer data integration, system stability, latency, uncertainty propagation, and verifiability and governability. These constraints drive the semiconductor industry to adopt closed-loop control, AI, and DT to advance toward next-generation smart manufacturing.

To achieve energy efficiency and a low-carbon economy, Ref. [359] indicates that enterprises' primary challenges lie in incomplete data acquisition, communication, and infrastructure. In the context of Industry 4.0, deploying self-aware machines can reduce processing costs by enabling energy savings. Energy-efficient operation encompasses physical systems, unit models for virtual space, virtual production models, and active energy-saving decision models. Among these, DT establishes dynamic mappings between physical processes and virtual models. DT is combined with communication infrastructures to transform fragmented data into actionable energy-saving indicators. Therefore, Ref. [359] proposes the DT architecture to provide the essential prerequisites for the "sensing–analysis–decision–feedback" process in energy closed-loop operations. In addition, Ref. [358] investigates the impact of AI across manufacturing domains, including robotics, automation, supply chains, predictive maintenance, and quality control. By leveraging ML and real-time data analytics, manufacturing systems can predict equipment failures before they occur, enabling predictive maintenance to reduce unplanned downtime, optimize resource allocation, and extend equipment lifespans. This forms a decision loop from "sensing" and "prediction" to feedback and corrective actions. The study emphasizes AI's capability for real-time monitoring and data analytics of production metrics, enabling continuous perception and evaluation of process states and adjustment of operations and decisions. This constitutes a critical information layer and supports control and decisions for closed-loop manufacturing.

Traditional closed-loop control cannot effectively predict temporal disturbances such as thermal accumulation. Therefore, Ref. [177] develops an intelligent closed-loop control system integrating DT and real-time MPC to mitigate quality instability caused by thermal accumulation in additive manufacturing. DT serves as a computational and decision-making platform for integrating physical sensing and model inference. LSTM is used to construct a surrogate model to overcome the high computational cost of physical models and to capture the spatiotemporal dynamics and future temperature trends in Directed Energy Deposition processes. Experimental results show that the system dynamically adjusts laser power over time to accommodate variations in cooling rates at different spatial locations.

In hybrid production environments, Ref. [176] integrates digital threads, multimodal sensor networks, and CNN to support the management framework for zero-defect manufacturing. The framework can accurately classify 14 defect types in high-precision continuous fiber composite materials, achieving 97%

accuracy. AI insights enhance the closed-loop feedback system, reducing response time to production anomalies to 45 s and the proportion of potential defects propagating to downstream processes by 78%. These results demonstrate its feasibility for real-time quality monitoring and rapid feedback. In addition, Ref. [317] integrates ML, sensor data, and IIoT to continuously monitor production parameters and establish a feedback decision process to reduce production interruptions and defect risks. The study analyzes sensor data to enable early fault prediction and anomaly detection, supporting proactive quality management and adaptive process optimization. The results indicate that AI technologies have strong application potential in real-time quality monitoring, predictive maintenance, and process decision support.

Material flow in complex processes is prone to physical disturbances, such as parts tipping on conveyor systems, which can lead to unexpected incidents. Therefore, Ref. [356] uses a physics engine to simulate interactions between parts and material-handling systems and develops a DT to improve the predictability of material flow behavior under variations in transport paths and material properties. In addition, Ref. [356] applies a model-based systems engineering approach to manage system complexity, covering requirement definition, structural design, and interaction modeling. The framework is integrated with industrial environments via the Open Platform Communications Unified Architecture protocol to enable bidirectional connectivity between DT and the physical system.

Ref. [357] proposes a DT architecture enhanced by generative AI for intelligent virtualized network management. The DT serves as a platform for network state modeling and strategy evaluation. It simulates network behavior, abstracts feature representations, and supports management decisions. GANs are used to improve DT's simulation accuracy. Generative AI generates high-fidelity simulation data to complement physical measurements and reduce the frequency of physical data acquisition under controlled error conditions. The system establishes an external closed loop between the physical network and the DT. It uses an error-discrimination mechanism to dynamically adjust the measurement frequency and reduce resource consumption. An internal closed loop is formed within the DT for strategy evaluation, comparing the performance of generative AI and conventional models to select the optimal network management strategy. The results show that the generative AI-enhanced DT effectively supports closed-loop network management while maintaining model accuracy and system resource efficiency.

The lack of bidirectional feedback between physical and virtual systems prevents factories from responding to unexpected events and, in severe cases, may lead to process interruptions. Therefore, Ref. [318] integrates sensors, data analytics, and DT to transmit physical data to the virtual model for analysis. The resulting decisions are fed back to the operational and execution layers of the physical system. This establishes a decision feedback loop based on near-real-time alignment of system state. The study uses sensors and IIoT devices to capture behavior and performance data from the physical system as inputs to the DT. The DT performs virtual experiments and parameter optimization without interrupting actual production. The analytical results are translated into actionable operational decisions, completing a closed-loop cycle from data perception to execution and correction. Experimental results demonstrate that the near real-time decision feedback loop improves resource utilization.

4.2.7 Contamination Free

In wafer fabs, robotic manipulators handle wafers and perform process integration tasks, operating within cleanrooms. This makes contamination-free control of cleanrooms a typical problem in ESA. The cleanroom system is dynamic and highly coupled with equipment behavior, process conditions, and contamination control strategies. Any motion or behavior of robotic manipulators can alter local airflow structures and particle distributions, thereby affecting process stability and yield. Refs. [178,179] show that contamination risk in cleanrooms exhibits systemic characteristics. It does not originate from a single source

but is closely related to human behavior, automated equipment operation, and environmental control. In addition, systematic contamination control and real-time environmental monitoring, when combined with automated monitoring, reduce human intervention and contamination uncertainty. With sensor networks, real-time monitoring, and data analytics, manufacturing systems can monitor the environment and provide feedback to equipment for control and decisions to form a closed-loop architecture [178]. In this context, control of robotic manipulators in contamination-free cleanrooms becomes a smart manufacturing problem that involves environmental perception, system coupling, and cross-layer feedback adaptation.

Excessive particle concentration in cleanrooms can degrade yield. However, when conventional sensors detect an increase in particle concentration, contamination has dispersed within the environment. In this context, Ref. [180] assumes that human occupants are the dominant dynamic source. Ref. [181] also reports that human occupants account for approximately 80% of total particle sources in bio-cleanrooms. Particle concentration reaches its peak within 1 min of entry and then gradually decreases over the next 2 min, stabilizing at approximately one-fifth of the peak value. Ref. [181] further indicates that activity intensity is the primary factor affecting particle emission rates, and particle concentration exhibits dynamic fluctuations with changes in activity type. The analysis shows that optimizing the fan system in bio-cleanrooms can achieve energy savings exceeding 50%. Ref. [180] proposes that predictive occupancy is necessary for demand-driven ventilation to compensate for the response delay of physical systems, such as ventilation timescales. The study demonstrates a strong correlation between occupancy and particle concentration, which can serve as a surrogate indicator for environmental control. These findings indicate that, in high-cleanliness cleanrooms, human behavior, airflow control, and contamination risk form a coupled engineering system with inherent time-delay characteristics. This characteristic is directly relevant to wafer fabrication.

Ref. [366] proposes a five-layer Industry 4.0 CPS architecture. The first layer (Connection) and the second layer (Conversion) emphasize the acquisition of seamless, accurate, real-time data from equipment and sensors, and ensuring that the state of the physical world is reflected in the virtual system in real time. The system must transform raw data acquired from the environment into information with management relevance. Real-time integration enables operators or control systems to observe current states, eliminating information asymmetry in decision-making. These principles are helpful for constructing the environmental sensing data integration layer and system state representation layer of cleanroom systems within ESA.

Ref. [320] proposes a monitoring system integrating wireless sensors, embedded gateways, and an IIoT server to enable real-time data flow and prevent over-ventilation, reducing energy waste. Wireless sensors enable simultaneous, real-time monitoring across multiple target locations. System parameters can be flexibly adjusted to account for specific contamination sources in different environments. This architecture can support environmental state measurement in ESA for cleanrooms. Ref. [321] develops a real-time air quality monitoring system integrating sensors through IIoT. The system uses an ESP32 microcontroller to acquire sensing data and transmit it to a server platform. When pollutant concentrations exceed predefined safety thresholds, the system automatically sends real-time notifications. The monitoring results can be used to improve laboratory ventilation systems and to support the installation of pollutant-capture devices. In addition, thermal anomalies in equipment are indicators of accidents. Therefore, Ref. [360] develops an autonomous monitoring system based on an Edge-Fog-Cloud architecture. The system uses the Internet of Robotic Things as mobile edge monitoring nodes and uses DL to enable autonomous navigation. Unmanned aerial vehicles (UAV) are deployed as fog nodes to transmit hotspot data from target locations. Thermal imaging data are uploaded to the cloud and analyzed by CNN. Visual images are analyzed to identify hotspot locations and associated equipment. The system automatically generates anomaly reports for corrective actions. Ref. [319] develops an IIoT-based air quality monitoring system that provides intuitive, real-time

composite indicators. The system continuously monitors air quality and aggregates data to a server by integrating sensors and networks.

Pharmaceutical manufacturing facilities and cleanrooms generate large volumes of monitoring data. However, these data are not used to effectively reduce contamination risks because they are not fed back to real-time control. Therefore, Ref. [182] uses analytics tools to transform heterogeneous and distributed monitoring data into visualized insights. A rolling period method is introduced to process sparse and highly fluctuating raw data and to identify trends of improvement or degradation in environmental quality. The study emphasizes that data must be contextualized and interpreted before they can be translated into effective control actions.

Ref. [331] develops a real-time indoor air quality monitoring and automation system that integrates IIoT and LoRaWAN communication technology. The system uses a communication module and a multifunctional sensor module to perform sensor calibration, data acquisition, cloud visualization, and actuation. Experimental results demonstrate high accuracy and efficiency in both data acquisition and automated actuation. This real-time measurement–threshold evaluation–actuation response architecture provides a reference framework for environmental state sensing and control in cleanrooms. In addition, Ref. [322] develops a real-time air quality forecasting system based on IIoT and ML. The system adopts a microcontroller architecture and continuously monitors particulate matter. The data acquisition pipeline includes device-level adjustment, local smoothing, and cloud transmission. By integrating long-term data storage and feature engineering, the system generates accurate short-term air quality forecasts.

4.2.8 Synchronized and Cooperative

Semiconductor manufacturing is a multi-agent system, including multiple robotic manipulators, equipment modules, and human operators. These systems are influenced by multiple state variables and their interactions, and their synchronized and cooperative control, including multi-arm coordination and arm-to-machine coordination, can be formulated as multi-input multi-output (MIMO) dynamic control problems under multiple constraints. Studies indicate that control strategies for collaborative robots have evolved from independent motion planning to real-time coordinated feedback control, which considers the states of other robots, equipment, and human operators simultaneously. The system is highly coupled and time-sensitive. Synchronization errors, control delays, and interaction disturbances can be amplified through system dynamics and affect stability.

4.2.8.1 Multi-Agent and Multi-Arm Synchronization and Coordination

This study considers hierarchical co-simulation as a system methodology for analyzing, validating, and designing synchronization and cooperative control in dynamically coupled multi-arm systems. Ref. [210] proposes a framework to integrate multiple heterogeneous simulation models within CPS to describe how spatiotemporal interactions among components at the micro level evolve into system behavior at the macro level. The hierarchical design decomposes complex CPS into manageable layers. Co-simulation ensures that coupling relationships across layers are accurately captured. Fast dynamic models at the micro level, such as motor control of robotic manipulators, are synchronized with discrete-event models at the macro level, such as production line scheduling. This achieves spatiotemporal synchronization, immediately reflecting parameter adjustments in system performance. Ref. [155] indicates that intelligent manufacturing networks are characterized by the integration of digital technologies such as the IIoT, big data, and AI. These technologies blur the boundary between low-level equipment control and high-level operational management. Human decision-making and automated systems form an inseparable whole. Therefore, the

interactions among physical equipment, digital algorithms, and human behavior create a complex and dynamically coupled system. Future research should focus on data-driven decision support, cyber-physical integrated real-time control, and system adaptability under human-machine collaboration.

Ref. [211] proposes an adaptive, nonsingular terminal sliding-mode control method to ensure accurate trajectory synchronization and stable tracking performance in multi-agent systems. The purpose is to address parameter uncertainties, friction, external disturbances, and stochastic time-varying delays introduced by communication networks in robotic manipulators. The method mitigates the drawbacks of conventional sliding mode control, including excessive control energy, chattering, and oscillations caused by constant high gains. This study introduces a novel time-varying gain based on a logistic function to improve convergence smoothness and steady-state tracking accuracy and to prevent significant overshoot under large initial position deviations. A hybrid feedback signal-processing mechanism is employed to reduce phase-shift effects introduced by network transmission. Experimental results show that the system maintains accurate position synchronization under weakly connected topologies and time-varying network delays, and the tracking error remains bounded. In addition, Ref. [260] develops a nonsingular terminal sliding-mode control scheme to address tracking delays caused by the leader's variable-speed motion via pose synchronization. The method provides full control of the end-effector, including both position and orientation, rather than only joint space or translational motion. Unit quaternions are used to achieve decoupled control of position and orientation dynamics. The nonsingular design avoids numerical singularities during computation. A mixed-type feedback mechanism is adopted to reduce follower signal deviations caused by network delays and to optimize performance under weakly connected directed graphs. The study provides rigorous stability analysis and proves that the tracking error converges to a bounded region within finite time.

Ref. [261] investigates the coupled dynamics and control of two mobile manipulators manipulating a common object with relative motion under parameter uncertainties and external disturbances. The purpose is to address geometric constraints, reduced degrees of freedom, and internal force regulation arising from closed-chain kinematics in multi-robot cooperation. A centralized control scheme is proposed to ensure that both the motion trajectory and force trajectory of the constrained object converge to desired bounds. Lyapunov stability theory is employed to prove the closed-loop system's stability and the finite-time boundedness of the tracking errors. Simulation results demonstrate robustness against disturbances induced by relative motion and environmental noise, ensuring coordinated operation of the platforms and manipulators. In addition, Ref. [212] proposes a control algorithm for cooperative transportation of a single object by multiple mobile manipulators. A leader-follower control strategy based on impedance dynamics is adopted to achieve synchronized cooperation. Each mobile manipulator is modeled as an impedance system connected to the object. Compliant physical interaction is emulated by software. A compensation mechanism is introduced to reduce motion deviations caused by model uncertainties. Experimental results validate the algorithm's feasibility and extend conventional multi-robot cooperative manipulation methods to mobile manipulator systems with mobile bases.

For multiple mobile manipulators with nonholonomic constraints, Ref. [262] proposes a cooperative grasping control framework with integrated obstacle avoidance to employ Nonholonomic Passive Decomposition to decompose complex robot motion into three independent dimensions. According to obstacle size, the system selects either global formation transformation or internal configuration reconfiguration for obstacle avoidance. It explicitly handles nonholonomic constraints, such as no lateral slip of mobile platform wheels, to ensure dynamic coordination between platforms and manipulators. Simulation results verify the theoretical feasibility.

4.2.8.2 Dual-Arm Cooperative Control

In multi-arm cooperative systems, local-level interaction control plays a critical role alongside system synchronization and coordination strategies. In dual-arm collaborative robots, studies integrate admittance control with hierarchical control architectures to regulate contact forces and internal force stability in arm–object interaction. This means that hierarchical design enables the incorporation of local interaction control into the overall cooperative framework. Ref. [332] investigates dual-arm cooperative manipulation of a common object using 6-DOF impedance control to regulate motion and interaction forces. A centralized external impedance control scheme is adopted to provide compliant behavior of the manipulated object during contact with the environment. Distributed internal impedance control is implemented at each end-effector to prevent excessive internal stress caused by asynchronous motion between manipulators. Force measurements from wrist-mounted sensors are used to adjust impedance in real time. Experimental results validate system stability. Ref. [264] develops a control framework for dual-arm collaborative robots in unstructured environments. Admittance control in a cooperative task space is employed to regulate internal and external forces. The dual-arm task is decomposed into absolute motion and relative motion components to achieve precise coordination and control. Tracking of relative motion is assigned the highest priority to enhance safety and stability in cooperative manipulation. The framework is implemented and validated on a physical experimental platform.

4.2.8.3 Human–Robot Collaboration

Ref. [265] examines human–machine cooperation in assembly lines by combining the high productivity and fatigue-free operation of robots with human flexibility and sensory capabilities in handling complex or deformable components. Robots act as intelligent assistants in collaborative tasks, supporting operations that are difficult, repetitive, or physically demanding. In addition, Ref. [266] investigates symbiotic human–robot collaboration by integrating robots' strength, endurance, and precision with human intuition, flexibility, and perception-driven problem-solving. The approach establishes a shared workspace in which humans and robots operate side by side. The study identifies 12 future research directions, including advanced human–robot symbiosis mechanisms and intelligent control, for example the use of DL for gesture classification, context awareness, and task recognition, as well as the integration of augmented reality and intelligent sensing systems to provide real-time decision support for operators.

4.2.9 Integration of Robot Control with Processes

From the perspective of ESA, the challenge in smart manufacturing is to achieve end-to-end integration of robotic manipulation control, process parameter adjustment, and information systems to ensure consistency and real-time controllability. Robotic control must respond in real time to equipment states, process variations, and scheduling changes. This requires decision mechanisms based on cross-system data integration. If data semantics, timing, and consistency are not aligned at the system level, robot control quality and process stability are directly degraded. This makes data integration a critical ESA issue. In addition, CPS in smart manufacturing still face challenges in cross-layer interoperability and standardization. Without a unified integration framework for process and management workflows, information silos can reduce overall efficiency. Recent studies reposition robot control from an isolated equipment problem to a coordination problem across process, control, and management systems.

Ref. [45] integrates CPS, system automation, and the process execution layer to propose that robotic manipulators (e.g., wafer handling robots) should evolve from devices executing predefined commands into

adaptive execution cores embedded in an end-to-end system architecture. This transformation ensures transparency of process parameters and consistency of the decision chain to support autonomous production. In addition, Ref. [46] proposes a three-layer framework integrating process, control, and management. The study emphasizes that without consistent integration of process and management workflows, high levels of automation at the control layer remain ineffective. The framework redefines robot control as a core systems engineering integration challenge. For tightly coupled and high-precision wafer fabrication, this framework provides a practical pathway for coordinating distributed decision logic and enabling system autonomy.

Ref. [47] demonstrates a logical chain from data integration to decision support, which influences automation and process control behaviors. Process control quality and system adaptability are no longer limited to motor speed or motion trajectory, but depend on the data–decision linkage between robotic manipulators and the overall production system. A representative scenario is semiconductor material handling. If process parameters transmitted by ERP systems, such as wafer type and thickness, are semantically inconsistent with robot-side sensor data, pick-and-place operations may fail or cause damage even when positioning accuracy is high. Furthermore, temporal misalignment degrades control quality due to data latency or misregistration. In ESA, robot control becomes a complex data engineering problem, where alignment of data flows is a prerequisite for achieving operational resilience and system autonomy. Within the overall process and decision chain, resource virtualization enables proper embedding, coordination, and scheduling of robot control. Therefore, Ref. [343] proposes a practical test-driven resource virtualization model. Through sensors, wireless networks, and cloud storage, smart factories can monitor the real-time status of all manufacturing resources, including personnel, robotic manipulators, equipment, and materials. Data analytics provide predictive insights into machine conditions and factory capacity for management. Furthermore, CPS virtualizes physical assets to support the development of resilient, intelligent, and self-aware machine nodes.

4.2.9.1 Manipulator Kinematics and Control Laws

Ref. [227] proposes an adaptive control law employing an online parameter estimation mechanism to automatically update control parameters of robotic manipulators. Lyapunov stability theory is used to prove that the trajectory tracking error will converge to zero despite parameter uncertainties. This result has important implications for semiconductor manufacturing. The study also expresses the nonlinear robot dynamics as a linear combination of parameters (linear parameterization property). This property enables computational tractability and serves as a standard paradigm in adaptive control design for manipulators.

Ref. [213] systematically presents robot kinematics, dynamics, and control theory, with emphasis on mathematical modeling for precise positioning of wafer handling robots. The study introduces state-space representations and covers modeling approaches for sensor feedback and vision feedback control in robotic systems. It integrates adaptive control into a framework for robot modeling and control analysis. The study also defines the physical limits of a single machine, which later become a prerequisite for integrating robots into higher-level manufacturing systems and management architectures.

4.2.9.2 Reinforcement Learning

Ref. [214] develops a DQN agent to learn control policies from high-dimensional sensory inputs. An end-to-end RL framework is adopted to map sensory inputs to action commands. This approach bridges the gap between high-dimensional perception and executable actions. Under high-dimensional sensing and partially observable conditions, conventional manipulator control methods based on explicit models and

control laws exhibit limitations at both the perception and policy levels. Therefore, the perception–policy learning paradigm is consistent with recent trends in learning-based control for robotics.

Ref. [267] employs a DDPG algorithm for continuous control. The method avoids the curse of dimensionality caused by action space discretization in high-DOF manipulators. DDPG preserves the continuous structure of the action space, which is critical for precise physical control. The approach has been successfully applied to more than 20 physical tasks, including vehicle driving and articulated motion control. Ref. [215] performs end-to-end joint training of perception and control to develop a DL policy that maps raw image observations to robot motor torques. The model consists of a seven-layer CNN with 92,000 parameters, designed for real-time robotic control. The algorithm enables policy training within tens of minutes of real-world interaction. Experimental results demonstrate the capability of the system in visual tracking, object localization, and complex contact dynamics in real-world tasks.

4.2.9.3 Vision-Guided Control

Visual serving control uses computer vision data as feedback in the servoing loop to achieve precise control of robot motion. Ref. [268] summarizes image-based visual serving (IBVS) and position-based visual serving (PBVS). PBVS treats the vision sensor as a three-dimensional (3D) sensor and defines errors and control laws in Cartesian space. IBVS treats the camera as a two-dimensional (2D) sensor and drives the robot using feature point errors in the image plane. These frameworks provide a common basis for subsequent research on vision-guided control, weak-model approaches, and learning-based visual serving. Ref. [228] proposes a method based on large-scale CNN that uses monocular images to predict gripper motion in the workspace and perform real-time serving control for precise grasping. Experimental results demonstrate that data-driven DL methods enable continuous, real-time visual serving for grasping under monocular vision, supporting the feasibility of learning-based visual serving in practical robotic grasping tasks. Ref. [269] develops a robotic pick-and-place system to grasp and recognize both known and novel objects. An affordance-based prediction algorithm is used to identify graspable regions directly from RGB-Depth images. After grasping, a cross-domain image matching framework is used to match captured images with online product images for object recognition, while CNN are used to learn visual features. Experimental results demonstrate that the system maintains high grasp success rates and recognition accuracy under severe occlusion and high object variability.

4.2.9.4 Single-Machine (Cell-Level) Optimization

When each single-machine is optimized independently, inherent variability affects system performance. Mass individualization aims to make cell-level optimization applicable within highly dynamic systems. Therefore, Ref. [205] analyzes challenges and future research directions, highlighting a paradigm shift in smart manufacturing from centralized control to distributed autonomy. Agent computation transforms intelligent and decentralized manufacturing systems. Machines (cells) can be modeled as autonomous agents with local goals. Ref. [206] indicates that multi-agent systems (MAS) typically employ coordination or negotiation mechanisms to resolve conflicts arising from distributed decision-making among autonomous entities, such as locally optimized cells. In ESA, MAS provides a promising approach to mitigate system conflicts induced by local optimization behavior.

4.3 *Transport Systems for Smart Manufacturing*

Wafer fabrication depends on the coordination of material, information, and control flows for system integration, real-time decision-making, and predictability. The vehicle dynamic systems here include

AMHS, AGV, AMR, and OHT. The powertrain characteristics, control architecture, and multi-controller coordination of these systems are similar to those of electric vehicles (EV), particularly in real-time control, multi-controller interaction, energy management, and reliability validation. In addition, DT, AI-driven monitoring, human-machine collaboration, and advanced logistics are key enabling technologies for smart manufacturing. DT establishes high-fidelity models of dynamic systems, enabling multi-controller interaction testing, fault injection, and extreme condition simulation, thereby improving system observability and predictability. This study examines the functional roles, control architectures, and digital twin modeling methods of automated mobile equipment in highly complex manufacturing environments within ESA. It further clarifies how these systems support operation in wafer fabs (as shown in Fig. 9).

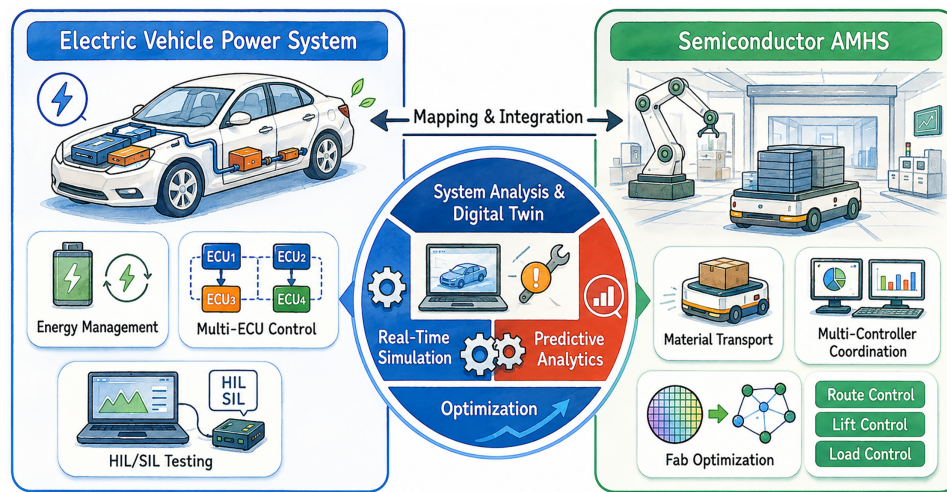


Figure 9: Conceptual framework for mapping electric vehicle power systems to semiconductor AMHS via digital twin-based system analysis.

This subsection summarizes the reviewed studies in Table 6. The table categorizes the literature according to physical system levels and ESA functional roles (modeling, diagnosis and prediction, control and optimization), to illustrate how ESA supports transport systems in semiconductor smart manufacturing.

Table 6: Transport systems research supporting ESA in semiconductor smart manufacturing.

Physical/ESA	Modeling	Diagnosis & Prediction	Control & Optimization
Materials			
Equipment	[156–159]		
Robots/AMR	[216–218]		
OHT/AGV/AMR	[270–288]	[289–294]	[300–312]
Sensor/IIoT	[313,314]	[323–325]	
Infrastructure (HPC/Edge/Cloud)	[344,345]		

DT integrates multi-physics, multi-scale, and stochastic simulations to emphasize bidirectional feedback between virtual and physical models. IIoT generates large volumes of real-time sensing data that DT uses for simulation, analysis, and manufacturing optimization. EV integrates mechanical, electrical, and software subsystems. DT can fully replicate these complex interdisciplinary systems. Therefore, Ref. [270] proposes a DT architecture for EV power systems to integrate sensing, control, and energy management.

Deep perception and fuzzy adaptive control can be applied to path planning. The DT modeling logic for mobile CPS with multi-controller coordination and energy-flow constraints can be mapped to AMHS/AGV systems in wafer fabs to analyze material-flow stability. In addition, Ref. [301] reviews research on AMR planning and control in intralogistics and distinguishes them from traditional AGV. Existing studies mainly focus on machine digitalization and IT infrastructure, while overlooking material handling systems in supporting cloud manufacturing. Therefore, Ref. [272] replaces fixed conveyor systems with AMR and integrates cloud platforms for real-time data sharing. The system distributes decision points and allocates materials based on real-time capacity, enabling dynamic responses to supply–demand variations. The study develops and validates a mathematical model demonstrating that conventional material-handling equipment leads to rigid factory layouts and process routing. Contrarily, AI-enabled AMR support more flexible transportation strategies.

The mathematical models for intralogistics, AMR scheduling, and control can be applied to intralogistics systems and AMHS path planning in wafer fabs. For example, Ref. [156] employs CPS to link Industry 4.0 and technical theories and proposes a framework to integrate advanced sensing, robotics, IIoT, and AI algorithms to support smart workshops. The study introduces a quantitative model to assess the state of smart manufacturing and identify weaknesses. The CPS design logic of system layering and feedback coordination can be mapped to multi-level coordinated control architectures in semiconductor material flow, including AMHS vehicles, path planning, and station-level control. Ref. [157] proposes a smart manufacturing system (SMIS) and a roadmap to integrate the physical, information, network, data, and visualization domains. It is implemented in several national projects. It adopts an ESA perspective to provide valuable guidance on defining the role and system boundaries of AMHS as a key material-flow subsystem in fabs.

Integration of vehicle powertrain models requires a detailed understanding of interactions among engines, transmissions, batteries, and motors. Hardware-in-the-loop (HIL) connects physical components to virtual environments to enable controlled and repeatable testing. Accordingly, Ref. [271] proposes a modeling technique that integrates vehicle energy systems into existing HIL platforms. The study employs three case studies to predict generator voltage variations, battery and generator currents, and generator efficiency and input torque. The results demonstrate that virtual testing systems reduce the risk and cost of physical testing while enabling performance evaluation of powertrain systems under realistic conditions. The HIL co-validation framework, which embeds physical controllers into virtual systems, can be extended to AMHS and AGV systems.

Control of permanent magnet synchronous motors (PMSM) is a critical component in EV subsystems. Accordingly, Ref. [300] integrates robust nonlinear model predictive control with space vector modulation (SVM) to establish a cascaded control architecture that ensures zero steady-state error under various operating conditions and exhibits strong robustness against parameter variations and unknown disturbances. In experiments, the method reduces electromagnetic torque ripple by 66% and current ripple by 54% compared with conventional strategies. The accurate reference tracking and fast dynamic response are useful for testing AMHS systems under extreme operating conditions, such as high load and high-frequency scheduling.

4.3.1 Dynamics Modeling

The combination of dynamics and smart manufacturing integrates the dynamics of transport equipment into an ESA framework for wafer fabs that is monitorable, predictable, and optimizable via closed-loop control. It is built on CPS to integrate sensing, communication, computation, and control into closed loops to become a distributed, self-adaptive, and self-optimizing smart factory architecture [51,52]. However, bottlenecks in wafer fab transport chains—such as FOUP transport, robotic transfer, vehicle-based transport, and dynamic routing—lie in the fidelity and real-time capability of dynamic models. Purely data-driven

approaches enable rapid approximation but often lack physical consistency and generalization capability. A more practical approach is to use rigid-body dynamics based on Euler–Lagrange or Newton–Euler formulations as the backbone, and then apply ML to model residuals, including friction, load variations, and unmodeled disturbances. This hybrid approach supports safe, robust, and interpretable motion control and optimization [53].

In cleanrooms, the fluid dynamics of airflow, pressure differentials, temperature control, and contaminant transport directly affect yield and process stability. Therefore, multi-physics simulations based on computational fluid dynamics (CFD), when accelerated by cloud/HPC resources, IIoT boundary conditions, and ML surrogate models, can evolve from offline design tools into near real-time decision support systems and serve as key solvers in DT frameworks [54,55]. Accordingly, this study treats DT as an engineering platform that integrates high-fidelity dynamic models, real-time data connectivity, and service-oriented decision-making. DT provides reusable capabilities for monitoring, prediction, diagnosis, and control, while CPS offers system closed-loop interaction [217]. In this context, the value of DT extends beyond visualization. DT incorporates dynamic uncertainties, data latency, and cross-system interoperability into a unified ESA optimization problem, enabling the development of next-generation smart manufacturing that is evaluable, verifiable, and deployable [57,67].

4.3.1.1 Rigid-Body Dynamics

The inverse dynamics model computes the required torques or forces based on desired robot motion, including position, velocity, and acceleration. It is built on a rigid-body model that assumes all robot links are rigid, and it is parameterized using kinematic parameters, such as the Denavit–Hartenberg (DH) model, and inertial parameters, including mass and inertia tensors. Existing approaches are sensitive to calibration errors and suffer from high computational dimensionality, often requiring large training datasets to achieve high accuracy. Ref. [216] proposes structured kernels that embed robot dynamics into a reproducing kernel Hilbert space (RKHS). This approach enables convergence to rigid-body dynamics with fewer training samples by incorporating structural priors. It is suitable for semiconductor manufacturing applications, such as Front Opening Unified Pod (FOUP) robotic manipulators and transfer equipment.

Data-driven identification of dynamic parameters may produce physically implausible results. To address this issue, Ref. [217] introduces physical constraints into the recursive Newton–Euler algorithm (RNEA) during parameter learning to ensure physical consistency. This improves training efficiency and model generalization. Validation is conducted using a 7-degree-of-freedom robotic manipulator in both simulation and real-world experiments, demonstrating improved accuracy. The differentiable RNEA proposed in [217] provides a practical approach for embedding dynamics into intelligent control and DT systems. It ensures that virtual models remain physically consistent during data-driven learning and serves as a bridge between physical systems and intelligent algorithms.

4.3.1.2 Computational Fluid Dynamics

Contaminant particles in cleanrooms originate from human activities and mechanical friction during equipment operation. Ventilation systems are required to remove particles and maintain a uniform temperature distribution. Accordingly, Ref. [158] employs CFD to determine ventilation control strategies to achieve target air conditions and to validate performance through coupled simulations of fluid flow and thermal fields. In addition, Ref. [159] integrates CFD simulation with *in-situ* measurements to propose an improvement strategy. CFD is used to simulate airflow distribution and particle dispersion within the factory, while measured data are used to provide boundary conditions and improve simulation accuracy.

Direct numerical simulation (DNS) exhibits computational cost that scales cubically with the Reynolds number, making large-scale turbulent flow simulation using conventional CFD methods, such as finite difference or finite volume methods, impractical. Therefore, Ref. [344] proposes a hybrid approach that embeds ML models within traditional numerical solvers to correct low-resolution simulation errors. The method replaces components most affected by resolution loss, such as convective fluxes and residual terms, with ML models. Differentiable programming with JAX enables end-to-end gradient-based optimization of the entire algorithm. In experiments, the method achieves accuracy comparable to high-resolution DNS.

Furthermore, Ref. [345] uses CNN to develop a turbulence model to improve low-resolution simulations of the incompressible Navier–Stokes equations. The method introduces physics-informed loss functions derived from turbulence theory and employs a differentiable numerical solver to propagate gradients across multiple solution steps. In experiments, this hybrid approach achieves better physical consistency than purely data-driven predictive models. In addition, Ref. [344] presents a decision-acceleration framework that integrates CFD and ML and is suitable for incorporation into DT systems. Ref. [345] proposes a method that ensures physical consistency and long-term stability in transient turbulence prediction through physics-informed loss functions and long-horizon differentiable training.

4.3.1.3 Vehicle System Dynamics

Mathematical models of vehicle dynamics are central to control loops and predictive learning systems in autonomous driving. Physical models provide interpretability, but parameter estimation is difficult and computationally intensive. Accordingly, Ref. [274] compares standard nonlinear physical models, linear state-space models, and neural network models. The results indicate that the latter two data-driven models achieve higher prediction accuracy than conventional nonlinear physical models. In addition, physical parameters estimated by structured neural networks, such as aerodynamic drag and rolling resistance, exhibit better physical consistency than those obtained from traditional models, thereby providing technical support for accurate DT modeling.

Ref. [273] proposes a method that integrates a component-based reduced-order model library with Bayesian state estimation to construct data-driven yet physics-informed digital twins. Physics-based reduced-order models provide computationally efficient, sufficiently accurate representations that meet DT prediction requirements. The component-based approach enables scalable modeling of large, complex systems and supports flexible model adaptation. Furthermore, model adaptation and uncertainty quantification in DT are formulated as a Bayesian state estimation problem, which infers the most suitable candidate model from the model library. Experimental results validate the feasibility of the proposed framework and provide an engineering methodology for modularizing, real-time calibration, and uncertainty management in dynamic models of AMHS subsystems.

4.3.1.4 Cyber–Physical Production Systems

Ref. [275] defines CPS as the confluence of embedded systems, real-time systems, distributed sensing systems, and control technologies. CPS is characterized as an emerging discipline that requires close collaboration among computer scientists and engineers from mechanical, civil, and biological domains. From the perspective of computational science and dynamics, CPS must integrate discrete logic with continuous dynamics while accurately handling noise, uncertainty, and imperfect spatiotemporal synchronization in physical systems.

In addition, Ref. [51] summarizes ESA research by the International Academy for Production Engineering (CIRP) on cyber–physical production systems (CPPS). This includes the technological evolution

from embedded systems to CPPS and the computational mechanisms required to achieve self-regulation and adaptation. Ref. [51] also defines the application boundaries of CPPS. Technologies such as multi-agent systems, RFID, and cloud computing are identified as key enablers of CPS, and CPPS is recognized as a core driver of Industry 4.0. For wafer fabrication, CPPS's ability to perceive and reason about the dynamic states of physical assets in real time provides a theoretical foundation for transport subsystems such as FOUP and AMHS. From a dynamics perspective, semiconductor equipment and material-handling systems are continuous-time systems. Their dynamic states must be incorporated into CPPS closed-loop sensing and control architectures that provide real-time responsiveness and self-regulation, supporting high-frequency and low-tolerance requirements.

4.3.2 Vehicle Adaptive Control and MPC

Transport chains in wafer fabs, including AMHS, OHT, and AGV systems, are dynamic systems governed by multiple constraints, including safety distances, path conflicts, acceleration and deceleration limits, load variations, and network delays. MPC explicitly handles state and input constraints in closed-loop control through receding-horizon optimization. It can be extended from single-vehicle control to fleet coordination under both centralized and distributed architectures. MPC is therefore well suited for high-frequency and low fault-tolerance requirements [58]. For stable long-term deployment in fabs, control models must address uncertainties arising from friction, load variations, aging, and environmental disturbances. Recent research trends integrate MPC with adaptive control, state estimation, and multivariable coordinated control. These approaches are applied to fleet coordination tasks, such as platooning, cooperative lane changing, and intersection coordination, to improve efficiency and safety margins. This control paradigm, characterized by coordination, constraint handling, and uncertainty management, is structurally analogous to system optimization problems in wafer fab transport networks [59].

In addition, DT integrates equipment/vehicle dynamic models, sensor data fusion, and real-time control decisions into a closed-loop framework that is calibratable, traceable, and verifiable. For example, DT-assisted localization and data registration improve the state reliability of transport equipment in complex environments. This provides more reliable online state and parameter inputs for MPC, thereby reducing conservatism and improving overall throughput [67]. ML complements this framework by enhancing models in a data-driven manner, estimating uncertainties, and reducing online computational burden. This supports MPC requirements for accurate predictive models and real-time solvability, enabling constrained optimal control to approach deployable engineering scales. It also establishes a scalable technical interface for DT-MPC closed-loop integration in semiconductor transport systems [60].

Ref. [58] reviews the extensive application of MPC in both single and multiple AGV. MPC utilizes predicted future information to optimize control actions while explicitly handling system constraints. AGV can respond autonomously to its environment, including acquiring sensor inputs, identifying system states, and planning actions. In multi-AGV systems, coordination requires information exchange to avoid path conflicts and collisions, which necessitates distributed MPC. This distributed architecture is analogous to the coordination requirements of multiple OHT and AGV systems in wafer fabs, particularly at intersections and merging segments where path conflict avoidance is critical. The control framework must maintain real-time responsiveness and safety margins while avoiding centralized computational bottlenecks. This corresponds to the core optimization problem addressed by MPC.

Demand-driven mobility systems actively allocate transport resources to regions with predicted demand, significantly improving transport performance. Accordingly, Ref. [290] proposes a data-driven distributionally robust optimization (DRO) approach. The method employs quad-tree dynamic region partitioning to capture high-resolution spatiotemporal demand dynamics. It further utilizes stochastic demand

covariance to design efficient algorithms for constructing demand probability distributions. Experimental results show superior average system performance. This is analogous to AMHS operation in wafer fabs, where uncertainties such as load variations, path congestion, and equipment state changes require both system stability and real-time responsiveness. DRO provides a useful framework for analyzing resource allocation strategies and imbalance risks in AMHS systems.

When GPS signals are unavailable, vehicle position and path curvature cannot be obtained. Vision sensors, such as cameras, can be used to detect lane markings and estimate path-following states. Accordingly, Ref. [291] proposes a composite nonlinear feedback (CNF) controller based on a fuzzy observer. The approach integrates a Takagi–Sugeno fuzzy dynamic model to transform nonlinear system characteristics into a convex structure. The CNF controller reduces convergence time and minimizes overshoot. A parallel distributed compensation control scheme is incorporated to handle actuator saturation constraints. The control performance is validated through software simulations. In addition, the CNF framework provides feasible path estimation under sensor failure conditions. This approach can enhance the reliability of MPC inputs in OHT and AGV systems and support high-performance control under narrow-path conflict scenarios. In addition, wireless channel uncertainty in vehicle-to-vehicle (V2V) communication introduces system delays. Accordingly, Ref. [302] proposes a communication–control framework to enhance the stability of vehicle control systems. The framework models V2V link performance under complex environments with 5G cellular network interference. The framework provides a useful engineering methodology for stability design in large-scale distributed AMHS control architectures under network uncertainty.

Ref. [303] employs ML and MPC to reduce pollutant emissions from compression-ignition engines without increasing computational burden. The study develops a support vector machine (SVM) linear parameter-varying (LPV) model as the embedded MPC model. To address the high computational cost of MPC, an imitation controller based on a DNN is introduced to replicate the LPV-MPC's control policy. The proposed “offline training and online imitation” strategy does not replace control decisions with ML. Instead, it learns the optimal control policy embedded in MPC. This approach significantly reduces real-time computational cost while preserving constraint handling and stability properties. This methodology can be extended to OHT and AMHS systems, where high-frequency MPC obstacle avoidance and scheduling problems can be reformulated as neural network-based real-time decision approximations.

To enhance safety control under driving limit conditions, Ref. [304] proposes an integrated controller based on tube-based MPC. The approach incorporates control vector constraints and lateral stability constraints into the control formulation. The method is validated through HIL simulations and achieves multiple objectives, including path-following accuracy, handling stability, and rollover prevention. At the methodological level, the maintenance of feasible sets and robustness characteristics emphasized by tube-based MPC provides a useful reference for analyzing safety boundaries in real-time obstacle avoidance and speed control for OHT and AGV systems in wafer fabs.

4.3.3 Multi-Objective Control

Operations of AMHS, OHT, and AGV systems in wafer fabs must simultaneously satisfy multiple objectives, including real-time responsiveness, collision avoidance, congestion reduction, minimizing empty loads, and maintaining global throughput stability. This constitutes a multi-objective, strongly coupled, and uncertain dynamic optimization problem, analogous to multi-objective control in autonomous driving. Therefore, analytical frameworks that integrate multibody dynamics models, CAD/CAE simulations, and multi-objective optimization in autonomous driving can be directly adapted. DT provides a system integration platform for this type of analysis. By integrating dynamic models, sensor data, and optimal control strategies within a CPS, DT supports multi-objective trade-off analysis and real-time decisions. In wafer fabs,

this enables dynamic adjustment of efficiency and resource allocation, allowing multi-objective optimization to become a sustainable capability in smart manufacturing.

Autonomous Mobility-on-Demand systems are designed to handle dynamic variations in the demand and supply of electric vehicles. Ref. [276] proposes an uncertainty set construction algorithm that ensures model parameters lie within a confidence interval with a specified probability. An equivalent convex optimization formulation is introduced to improve computational efficiency. This approach dispatches empty vehicles to predicted demand regions and guides low-battery vehicles to charging stations. This distributionally robust optimization framework can be applied to AMHS in wafer fabs to address uncertainties arising from time-varying demand (e.g., batch arrivals and machine waiting), constrained supply (e.g., vehicle availability and charging or maintenance states), and network congestion.

Distributed-drive electric vehicles are equipped with four in-wheel motors. Ref. [305] proposes a hierarchical control method to address independent drive control and energy efficiency. The upper-level motion controller uses sliding-mode control to compute the required total driving force and yaw moment. The lower-level allocation controller performs optimal torque distribution among the four in-wheel motors to maximize energy efficiency. A motor efficiency model integrates the motor efficiency map with a current closed-loop model. Experimental results show simultaneous trajectory-tracking performance and energy savings. The hierarchical control framework provides a practical and transferable control architecture for OHT systems in wafer fabs. It ensures transport stability and safety constraints while extending battery endurance and reducing energy consumption. In addition, Ref. [306] proposes a longitudinal and lateral trajectory planning and tracking control algorithm in V2V. The upper-level trajectory planner generates a set of collision-free trajectory clusters for different lane-change durations. The trajectories are represented using polynomial functions, and a single tuning factor is adopted to simultaneously optimize both longitudinal and lateral paths. The lower-level controller employs a multi-input multi-output (MIMO) triple-step nonlinear control approach. It incorporates an error feedback structure similar to PID control and establishes closed-loop stability using a Lyapunov function. The hierarchical framework for multi-vehicle collision avoidance and lane-changing planning with tracking control can be mapped to coordinated scheduling strategies for OHT systems in wafer fabs. This mapping is applicable to scenarios involving multi-track switching and dynamic conflict avoidance among vehicles.

Parameter uncertainty, modeling errors, and external disturbances degrade trajectory tracking performance in autonomous driving systems. Accordingly, Ref. [307] proposes a gain-scheduled controller that integrates LPV modeling, MPC, and H_∞ robust control theory. The LPV framework captures time-varying characteristics of vehicle speed and tire cornering stiffness. H_∞ robust control is applied to handle system uncertainties and external disturbances. MPC is employed to enforce state and input constraints. Experimental results demonstrate that the proposed controller improves trajectory tracking accuracy and ensures both lateral and rollover stability of the vehicle. At the control methodology level, this controller provides a reference framework for precise positioning control of OHT systems in wafer fabs under varying load conditions, such as fully loaded and empty wafer carriers.

Cooperative driving automation (CDA) lacks convenient validation tools. Accordingly, Ref. [292] proposes OpenCDA, an open-source architecture that integrates co-simulation, including a co-simulation platform, a full-stack cooperative driving system, and a scenario manager. The architecture adopts a highly modular design and supports testing with user-defined algorithms. At the system architecture level, this design paradigm provides an important reference for constructing DT platforms for AMHS in wafer fabs. By modularizing OHT path planning, re-planning, and cooperative collision avoidance logic, and integrating scenario management with co-simulation mechanisms, the framework enables evaluation of the coordinated behavior and system stability of hundreds of OHT vehicles in a complex network under a CPS setting.

4.3.4 Reinforcement Learning-Based Control

In autonomous driving systems, motion planning and control address multi-DOF dynamics, strict safety constraints, real-time requirements, and coupled interactions among multiple agents. Conventional real-time optimal control methods based on analytical models are limited by computational complexity, which constrains system performance. Accordingly, RL/DRL does not rely on online solving of high-dimensional models [66]. It learns control policies through an offline state–action–reward loop, where complex system dynamics are implicitly encoded in model parameters. During online operation, control commands are generated via low-computational-cost inference, satisfying real-time requirements. In addition, DT provides high-fidelity simulation, parameterized system models, and multi-scenario testing spaces. These support training, stress testing, and performance benchmarking of RL models, and enable system transfer from simulation to real-world deployment. In semiconductor smart manufacturing, integrating RL and DT provides an engineering methodology for exploring complex motion and scheduling strategies without compromising safety or verifiability. This approach provides clear, consistent engineering insights for system control and the optimization of mobile equipment.

4.3.4.1 Reinforcement Learning

Path following is a core task in autonomous driving and involves time-varying and nonlinear vehicle dynamics. Ref. [308] adopts proximal policy optimization as the core DRL algorithm and combines it with the conventional pure pursuit method. Simulation results show improved path-following accuracy and performance in low-speed driving scenarios. Applying RL to adaptive path-following control introduces several challenges. These include how to represent environmental states using features such as vehicle speed, lateral deviation, and road friction, and how to perform real-time fine-tuning of control parameters using limited feedback. Accordingly, Ref. [278] proposes an online RL algorithm based on multi-kernel dual heuristic programming. The method employs recursive least-squares temporal difference learning to iteratively update the weights of kernel functions and uses an approximate linear dependency mechanism to construct a kernel dictionary. This approach effectively builds a feature space suitable for multi-kernel representations. Simulation results show that the proposed controller achieves high tracking accuracy and smoothness. RL for lane-changing control typically constrains control actions to discrete spaces. Ref. [309] applies the DDPG algorithm to learn control policies through trial-and-error and constructs a continuous action space for lane-changing control in dynamic driving environments. Simulation results demonstrate a high lane-change success rate across diverse driving scenarios.

DNN has demonstrated strong performance in robotic applications, but it lacks formal robustness guarantees. This limitation poses risks in safety-critical scenarios such as pedestrian avoidance and sensor disturbances. Accordingly, Ref. [279] develops an online verifiable robust defense algorithm for DRL, referred to as certified adversarial robustness for DRL. Under worst-case perturbations, the algorithm identifies and selects the most robust discrete control actions. Experimental results show improved safety performance under adversarial disturbances and sensor noise.

Ref. [311] proposes a supervisor agent that leverages deep distributional RL to enhance existing driver assistance systems. The agent adopts an end-to-end approach that maps camera images and LiDAR data to maneuver decisions. Quantile regression DQN is employed as the core distributional RL algorithm. Experimental results show superior performance. This architecture resembles the monitoring and scheduling logic used in equipment control in wafer fab automation. In addition, Ref. [312] applies transfer RL to high-level decisions in autonomous driving. The method first trains in a simplified environment and then transfers the learned policy to a simulator that accurately captures real vehicle dynamics. This approach

reduces exploration in hazardous environments and accelerates model training. Experimental results show a reduced driving failure rate due to perception errors, from 10% to 2.75%. Ref. [277] develops a photorealistic simulation and training engine to address trajectory deviation or near-collision conditions. The system generates novel training data by rendering new trajectories from limited driving data and constructs continuous paths using virtual agents. This enables RL agents to learn robust control policies without relying on manual annotation. Experimental results demonstrate improved capability in correcting vehicle motion direction. At the ESA level, this approach provides a reference for designing robust motion control and fault-tolerant training processes for AMHS in wafer fabs under constrained safety boundaries and abnormal disturbance conditions.

4.3.4.2 Digital Twin

Conventional RL is costly and risky to train in real-world environments. Accordingly, Ref. [293] applies RL to path-following control in autonomous driving to minimize the transfer gap. A digital counterpart is constructed to model vehicle dynamics, traffic scenarios, and sensing systems. The simulation environment incorporates a large number of critical and hazardous scenarios, thereby reducing costs and risks. Parameter randomization is employed during training to enhance the robustness of AI agents against uncertainties in different paths and vehicle models. This approach can be extended to wafer fabs. By developing a DT model of AMHS, various scenarios such as production load variations, traffic congestion, and system expansion can be reproduced in simulation. RL can then be used to train robust scheduling and routing strategies.

Dynamic variations in complex engineering systems are often accompanied by unpredictable stochastic disturbances and nonlinear deviations. Therefore, Ref. [294] proposes a time-evolving DT architecture that integrates physics-based and data-driven models via a weighting mechanism to dynamically capture the behavior of physical entities. The architecture is validated by a two-stage cascade tank system. At the ESA level, this methodology can be extended to semiconductor process modeling. For example, it can integrate chemical reaction models based on reaction kinetics with data-driven models derived from sensor data to capture variations in process behavior caused by equipment degradation or sensor drift.

4.3.5 Digital Twin Simulation

DT bridges physical systems and data-driven models. It constructs a virtual representation that co-evolves with the physical system, enabling real-time simulation, monitoring, and analysis of system states. In intelligent vehicle systems, DT supports vehicle localization, sensor data registration, and system state estimation. This capability improves stability and reliability in high-noise, uncertain environments. This provides a transferable methodology for wafer fabrication, including process monitoring, virtual metrology, and equipment management. DT is employed to overcome limitations in onboard sensing coverage and computational resources. Simulation results indicate reduced travel time and congestion caused by traffic incidents. This method can be mapped to AMHS in wafer fabs. For example, facility-sensing data and scheduling information can be integrated to construct a vehicle-state map that supports real-time routing and scheduling decisions.

Ref. [281] proposes an end-to-end DT system for autonomous driving. Roadside units are used to capture traffic information. The data are preprocessed through edge computing and then transmitted to the cloud to construct a synchronized DT model. Experimental results indicate that the path planning services achieve a reliability of 99%. This sensor and state map sharing mechanism can be mapped to wafer fabrication. It can be interpreted as process-state sharing among multiple machines to support coordinated decision-making and scheduling under shared process-state conditions. In addition, Ref. [282] proposes a

DT co-simulation framework that integrates connected vehicles with pedestrians in the loop to address the limitation of conventional simulation environments, which lack interaction with real traffic participants, thereby reducing test validity. A virtual cave environment enables real pedestrians to participate in the DT simulation. Experimental results indicate a better performance of the warning system.

Ref. [325] proposes a vision-based DT framework for personnel localization and safety monitoring. The study replaces conventional Computer-Aided Design (CAD) or manual modeling with 3D scanned models and applies DL to detect personnel in video streams. The detected states are mapped in real time to the DT to enable region-wise occupancy counting, trajectory tracking, and scheduling optimization. The experimental results demonstrate that the method meets management requirements. This approach can be applied to real-time safety monitoring in human-machine collaboration zones, maintenance pathways, and hazardous areas in wafer fabs, and supports system observability in smart manufacturing. In addition, Ref. [313] proposes a centralized DT localization system that does not rely on high-precision sensors on individual robots. The framework utilizes onboard robot sensors and surveillance camera images, and applies an extended Kalman filter for state estimation. The DT system integrates 3D simulation to achieve accurate state synchronization with the physical twin. The architecture supports the integration of robots with heterogeneous technologies and specifications, is easy to deploy, and incurs low maintenance costs. This architecture can be extended to fab AMHS, for example, by using existing facility surveillance cameras as centralized localization sources to manage AGV systems with different brands and technical specifications, thereby reducing reliance on high-end onboard sensors and lowering overall system upgrade cost.

The first step of DT is to transform the physical world into an accurate digital representation. Ref. [323] proposes a data-driven collaborative sensing architecture employing a message passing neural network to enable information sharing among vehicles, thereby improving localization accuracy and providing a geometric model of the environment state for DT. Data are aggregated in a mobile edge cloud to achieve optimal localization performance. Experimental results demonstrate better performance. Similar collaborative sensing and implicit information fusion mechanisms can be extended to AMHS localization in wafer fabs. For example, shared track reference-point information can improve navigation accuracy in confined spaces and address sensing blind spots. In addition, to mitigate cumulative drift in visual-inertial odometry, Ref. [324] proposes matching sparse 3D point clouds with a DT map using a point-to-plane approach. This method performs geometric alignment and provides 6-DoF global measurements. Experimental results show that the method achieves higher robustness in environments with significant viewpoint variations.

4.3.6 Suspension Systems

The suspension system is simultaneously influenced by structural mechanics, actuator dynamics, sensing uncertainty, and control coordination. It is a highly coupled cross-scale dynamic system and exhibits strong system-level similarity to the integrated control of precision motion platforms, vibration isolation modules, and equipment in wafer fabs. Recent studies indicate that suspension system research has rapidly shifted from passive and semi-active designs to active actuation, intelligent control, and DT approaches. This transition is consistent with the evolution in semiconductor smart manufacturing from static recipe control to real-time model updating and adaptive optimization. Ref. [283] proposes an interval-based multi-objective robust design optimization algorithm and applies it to a 15-DOF full-vehicle model under double lane change maneuvers. A non-probabilistic α -cut interval analysis is adopted. Experimental results indicate that the optimized suspension parameters effectively mitigate the impact of uncertainty propagation on vehicle dynamic performance. At the methodological level, this algorithm can be extended to the design and vibration suppression of AMHS suspension systems in wafer fabs under non-probabilistic uncertainties such as load variations and track wear.

Ref. [284] proposes a numerical optimization framework for multi-objective optimization of passive suspension systems operating on stochastic road profiles. The method integrates the NSGA-II algorithm with vertical dynamic analysis. For an 8-DOF vertical dynamic system, the approach generates a Pareto-optimal front that balances ride comfort and safety objectives. This multi-objective Pareto optimization method can be regarded as a transferable systems engineering tool to support load balancing and vibration suppression in AMHS systems in wafer fabs. In addition, Ref. [285] proposes a controllable electrically interconnected suspension system composed of a controllable electrical network and two independent electromagnetic suspensions. The independent electromagnetic suspensions convert vibration energy into electrical energy. A 2-DOF model is used to analyze the mechanical characteristics of the electromagnetic suspensions, including damping, stiffness, and inertia. Experimental results confirm that adjusting the resistance in the electrical network enables real-time tuning of suspension characteristics and improves vertical vibration performance. This energy conversion and control architecture provides a reference for the design of virtual stiffness and damping, for example, in vibration isolation and parameter tuning of precision motion equipment in wafer fabrication, where high-speed, high-precision requirements are required.

Ref. [286] proposes a serial active variable geometry suspension (SAVGS) scheme in which a controlled single-link active component is connected in series to the upper end of the spring-damper strut in a conventional double A-arm suspension. The design attenuates high-frequency vertical acceleration and has the potential to replace conventional anti-roll bars. The approach suppresses high-frequency vertical vibration through controllable kinematic DOF. This concept can inform the design of vibration-isolation mechanisms for mobile robotic platforms used in wafer fabrication. In such systems, motion-induced disturbances can directly affect localization accuracy and overall system reliability.

Continuous damping control (CDC) is an active damper with an adjustable damping coefficient and is widely used in robotic and manufacturing systems. However, its application in DT systems is limited by high computational cost and insufficient accuracy. Therefore, Ref. [218] proposes a hybrid model that integrates a lumped-parameter model with a neural network. The former represents the CDC output damping force using simplified physical component combinations, while the latter compensates for the nonlinear hysteretic force that the lumped-parameter model cannot accurately capture. Experimental results indicate that the output force computed by the hybrid model under random excitation closely matches the measured results. This modeling approach is consistent with the concept in wafer fabrication of monitoring equipment states and predicting process outcomes through virtual metrology. In addition, Ref. [287] proposes a semi-active suspension control system based on a 7-DOF full-vehicle model with four electrorheological dampers. An LPV fast MPC scheme is adopted to simplify the computation of dissipative constraints and to achieve balanced control objectives within a very short control period (5 ms). This system exhibits similarity to multi-DOF vibration isolation platforms used in precision manufacturing equipment, and can serve as a reference for mitigating multi-directional micro-vibrations in systems such as wafer scanners.

Hub motor electric vehicles experience increased vertical vibration due to the added unsprung mass. Accordingly, Ref. [288] proposes a self-powered electromagnetic vibration suppression and absorption system that integrates a magnetorheological damper (MRD) with a linear motor. The MRD is used as the suspension damper, while the linear motor acts as a linear electromagnetic dynamic vibration absorber. The system converts vibration energy into electrical energy to power the MRD, enabling self-powered operation. Particle swarm optimization is used to optimize system parameters. This concept can inform the design of self-powered condition-monitoring systems for wafer material-handling environments, particularly in elevated OHT track scenarios where wiring is difficult. The vibration compensation strategy for unsprung mass can be extended to AMR systems carrying heavy modules for vertical vibration suppression and transport stability design. In addition, Ref. [314] integrates heterogeneous sensing modalities, including

stereo vision, GPS, and an inertial measurement unit (IMU), to improve control accuracy, stability, and safety of active suspension systems. Ant Colony Optimization is applied to finely tune suspension parameters. HIL simulation results indicate that suspension dynamic deflection is reduced by more than 4%. This control logic, with state recognition and mode-switching capabilities, can be extended to vibration-suppression strategies for OHT systems in wafer fabs when traversing tracks with different turning radii, for example, by switching between speed and damping control modes based on curvature and dynamic state.

4.4 Materials Modeling in Semiconductor Smart Manufacturing

In addition, material problems in advanced processes commonly exhibit multiscale (micro–meso–macro) and multiphysics coupling. Examples include microstructure evolution, interfacial behavior, and thermo–mechanical–electrical interactions, which influence process outcomes across different temporal and spatial scales. However, materials simulations are computationally expensive and are difficult to apply directly to real-time process control or large-scale parameter exploration. Therefore, recent developments have shifted MM toward DT and surrogate or reduced-order models, enabling physical consistency while improving computational efficiency and system integration. The studies discussed in this subsection are summarized in Table 7. The table classifies the literature according to physical system levels and ESA functional roles (modeling, diagnosis and prediction, and control and optimization) to illustrate how MM research supports engineering system analysis in semiconductor smart manufacturing.

Table 7: Materials modeling research supporting ESA in semiconductor smart manufacturing.

Physical/ESA	Modeling	Diagnosis & Prediction	Control & Optimization
Materials	[117–132]	[133–148]	[149–152]
Equipment	[160–163]	[183–187]	
Robots/AMR			
OHT/AGV/AMR			
Sensor/IIoT			
Infrastructure (HPC/Edge/Cloud)	[346,347]		

4.4.1 Materials Modeling

Data-driven and graph-based modeling approaches are important for linking material microstructures with macroscopic properties. By representing materials as structural systems composed of nodes and edges, graph neural networks (GNN) can effectively capture structure–property relationships and demonstrate strong performance in various material behavior prediction tasks. Compared with purely physics-based models, these approaches are better suited for integration with DT, as they provide fast, updatable predictions with limited computational resources, meeting the requirements of real-time operation and scalability in smart manufacturing.

4.4.1.1 Material Modeling and ESA

Ref. [117] reviews research on the nonlinear mechanics of materials and identifies gaps between material models used in industry and those in academic laboratories. Ref. [118] indicates that the crystal plasticity finite element method is powerful for mechanical problems in materials science. It integrates micromechanical models into a unified plasticity framework to capture multiphysics interactions at the mesoscale (grain clusters) and microscale (intragranular level). It can address the influence of boundary

conditions on mechanical response and microstructure prediction. Refs. [117,118,298,299] indicate that the key challenge of MM is to embed cross-scale material models into engineering decision frameworks that are predictive and verifiable.

4.4.1.2 Material Microstructure

Synthetic microstructures provide stochastic samples for numerical MM, but they incur high computational cost, low reproducibility, and limited control over sample distributions. Ref. [119] proposes a DL framework that combines GAN with actor–critic RL to generate three-dimensional microstructures conditioned on quantities of interest. Experimental results indicate that the error between predicted physical properties and target values is within 5%. In advanced packaging and interconnect processes, low-k materials typically exhibit porous microstructures. The approach in [119] can be used to synthesize and optimize such porous microstructures and to evaluate their structural stability during CMP. In addition, photoresist stochasticity is a key factor affecting line edge roughness and yield in advanced lithography. The controllable generation of three-dimensional microstructures demonstrated in [119] can be extended to numerical modeling and statistical analysis of stochastic photoresist microstructures, providing a tool to assess the impact of microstructure evolution on process stability.

In addition, Ref. [120] applies next-generation computational materials engineering (CME) to optimize the physical and thermodynamic properties of Al–Si alloys before actual casting, thereby replacing costly, time-consuming trial-and-error approaches. CME performs atomistic-scale modeling through simulation, enabling the integration of experimental and virtual environments. It supports the construction of DT for production chains and products and enables the embedding of MM into manufacturing decision processes.

4.4.1.3 DT, MM, and Smart Manufacturing

Ref. [74] indicates that DT serves as a dynamic virtual replica of physical materials or processes for real-time prediction and optimization of material behavior. The virtual model integrates finite element analysis, phase-field modeling, and molecular dynamics, and is constructed using both physics simulations and ML models. Furthermore, DT captures complex process–structure–property relationships and monitors microstructural evolution under loading. For example, the properties and service life of metal additive manufacturing components are strongly influenced by geometry, microstructure, and defects. Optimization of process variables relies on costly and time-consuming trial-and-error approaches. Accordingly, Ref. [121] proposes a first-generation DT that employs transient three-dimensional models to compute the spatial and temporal evolution of metallurgical parameters. Experimental results demonstrate that DT can transform expensive physical experiments into fast and low-cost numerical experiments, thereby reducing development time.

Only a few commercial alloys can be fabricated by 3D printing into defect-free and structurally stable components. Component qualification relies on costly and time-consuming trial-and-error approaches. Accordingly, Ref. [122] integrates mechanistic models, control models, and statistical models to construct a DT of the printing system to reduce trial iterations and shorten qualification time. In addition, the performance of a DT depends on high-fidelity digital models and surrogate models. To address the complex thermophysical evolution in additive manufacturing, Ref. [123] investigates a hybrid DT architecture that combines physics and AI approaches, including multiscale–multiphysics modeling, surrogate modeling, and bidirectional communication with closed-loop control. Using laser powder bed fusion (LPBF) as an example, the study shows that melt pool instability can be predicted by the digital model and mitigated by adjusting laser power to eliminate defects. Ref. [124] proposes a methodology for constructing DT for

the LPBF process. The approach uses physics mechanistic models to establish a virtual representation of the LPBF process as the foundation of DT, and incorporates model uncertainty and process variability. The performance of DT is validated using data collected from the manufacturing process. The material-centric DT approaches in [123,124] provide a systematic reference for handling material variability, yield control, and process optimization in semiconductor manufacturing.

4.4.1.4 GNN Applications in Materials Science

Ref. [76] indicates that atomistic materials science is increasingly adopting advanced ML techniques. Conventional ML faces limitations when handling complex crystal structures due to incompatibility with non-Euclidean systems. In contrast, GNN provides strong capabilities in feature extraction, relational reasoning, and compositional generalization. Accordingly, Ref. [76] presents several GNN-based models, applied to crystal structure stability prediction, rapid materials screening, and the computation of diverse optoelectronic and mechanical properties. In addition, atomistic GNN provides high-accuracy and low-computational-cost property prediction modules that can serve as surrogate models in semiconductor DT. These modules can be embedded into ESA to enable real-time parameter feedback and serve as a foundation for metallurgical and electronic properties.

Conventional ML requires manual feature engineering or complex transformations of atomic coordinates. Accordingly, Ref. [133] proposes a crystal graph convolutional neural network (CGCNN) framework that learns material properties directly from atomic connectivity. The crystal structure is represented as a crystal graph encoding atomic features and chemical bond interactions. Convolutional layers iteratively update atomic features by aggregating information from local environments, including neighboring atoms and bonds, through nonlinear transformations. The model is trained using data from the Materials Project, covering more than 46,000 diverse inorganic crystals. Experimental results show that CGCNN achieves high prediction accuracy for eight different crystal properties, with performance comparable to density functional theory (DFT) calculations.

Ref. [134] proposes a materials graph network (MEGNet) model that enables accurate property prediction for both molecules and crystals. On the QM9 molecular dataset, MEGNet outperforms previous ML models on 11 out of 13 property prediction tasks. For 60,000 crystal structures in the Materials Project dataset, MEGNet achieves higher accuracy than DFT in predicting formation energy, band gap, and elastic modulus. In wafer fabrication such as chemical vapor deposition and dry etching, material behavior is strongly influenced by the reaction chamber environment. MEGNet incorporates these physical variables as global inputs and enables accurate prediction of material stability and electronic properties under specific process conditions. It can serve as a property prediction module for constructing process DT systems.

Ref. [77] indicates that GNN can process structured representations of chemical molecules. For example, message-passing neural network architectures convert molecular structures into computational models by propagating node features (atoms) along edges (chemical bonds) and iteratively aggregating information from local neighborhoods. This is applied to property prediction, database screening, inverse materials design, simulation acceleration, and synthesis pathway prediction. Compared with conventional DFT calculations, GNN can handle compositionally disordered systems, such as high-entropy alloys, and predict their properties. In wafer fabrication, the chemical structures of photoresists and deposition precursors directly determine film quality. GNN can automatically learn chemical features and predict thermal stability, chemical reactivity, and reaction kinetics under specific temperature and pressure conditions, enabling early-stage evaluation and reducing experimental and computational costs.

Ref. [135] proposes an extended neural message passing model for predicting the formation energy of molecules. The model employs an edge-update network, in which the information exchanged between atoms depends on the receiving atom's state. The model is evaluated on the QM9, Materials Project, and OQMD datasets and demonstrates superior prediction accuracy compared to existing state-of-the-art methods. In addition, existing atomistic GNN models rely on interatomic distances and neglect bond-angle effects. Many material properties, such as band gaps, are highly sensitive to small variations in bond angles. Accordingly, Ref. [136] proposes an atomistic line graph neural network that performs message passing on both the atomic bond graph and its corresponding line graph, which encodes bond-angle information. In the line graph, nodes represent chemical bonds and edges represent the angles between bonds. By alternating convolution operations on the bond graph and the line graph, angular information is efficiently incorporated into atomic representations. In semiconductor materials, small geometric distortions (strain) induced by processes can directly alter the band structure. ALIGNN can predict band-gap shifts induced by process-related stress and serve as a reference for DT device performance modeling and automated parameter optimization in wafer fabrication.

4.4.1.5 Microscopic Materials and Polycrystalline Silicon

Ref. [78] indicates that AI and ML play central roles in property prediction, materials design, and large-scale data analysis, and emphasizes that physics mechanistic models are important for developing AI surrogate models. Porous materials, such as metal–organic frameworks, piezoelectric poly (vinylidene fluoride) (PVDF), and 3D-printed metamaterials, are representative cases of materials informatics applied to the design of micro- and angstrom-scale structures. For example, in MOF, ML is used to screen structures with target adsorption or catalytic properties from large candidate spaces. This materials informatics paradigm, combining physics models with AI surrogate models, can serve as a framework for material structure search, property prediction, and rapid evaluation of process parameters in wafer fabrication. For polycrystalline silicon, the high degrees of freedom in the microstructure and the wide range of length scales make pure physics simulations insufficient for exploration under process-iteration constraints. Therefore, materials informatics also employs GNN to perform relational reasoning and interpretable feature extraction for polycrystalline microstructure characterization.

Within the materials informatics framework, ML is used to establish structure–property relationships between microstructures and effective properties, providing a foundation for handling polycrystalline systems with higher DOF. For example, Ref. [137] employs a three-dimensional CNN to link material microstructures with their effective properties. The 3D CNN automatically learns salient features from microstructures and uses them to estimate higher-order spatial correlations. Experimental results demonstrate that this approach learns interpretable structural features and improves prediction accuracy for unseen microstructures.

Properties of polycrystalline materials are often described using statistical representations or image methods, which do not account for physical interactions between neighboring grains. Accordingly, Ref. [127] proposes a GNN model that transforms polycrystalline microstructures into graph embeddings. The embeddings are connected to target properties via a feedforward neural network (FNN) that incorporates intergranular interactions. The model provides high interpretability and enables quantitative analysis of the contribution of grain-level features to the overall material properties.

Grain boundaries have five macroscopic geometric DOF. Together with temperature and composition, they form a seven-dimensional space that is difficult to model. Ref. [125] proposes a data-driven approach combining a genetic algorithm (GA) with a DNN to construct grain boundary property maps. The GA is used for variable selection to identify the most influential geometric features among many candidates.

The DNN evaluates grain boundary properties over millions of geometric configurations. Experimental results indicate that the model accurately predicts the distribution of grain boundary properties in the seven-dimensional space. In wafer fabrication, conductive and structural materials, such as metal interconnects and polycrystalline thin films, are also polycrystalline systems strongly governed by grain boundary behavior. Models that describe grain boundary property distributions in the space of temperature, composition, and geometric DOF are structurally consistent with the materials–process–reliability analysis framework commonly used in semiconductor engineering.

4.4.1.6 High-Performance Materials

High-performance materials, such as titanium alloys, stainless steels, and specialized synthetic materials (e.g., Teflon), require specific heat treatment, pressure, and processing parameters and are more difficult to process than conventional materials. Ref. [79] indicates that smart manufacturing can rapidly adapt to the unique processing requirements of advanced materials. Conventional robotic systems require manual error correction and reprogramming, whereas smart manufacturing enables automatic implementation of design changes through interconnected systems. In addition, intelligent technologies accelerate simulation and experimental validation, reducing the development-to-testing cycle. In wafer fabrication, materials must withstand extreme chemical corrosion, temperature, and pressure conditions. The real-time feedback and microstructure adjustment of smart manufacturing ensure stable material performance under complex operating environments.

In smart manufacturing, the processing of high-performance materials involves highly complex, high-dimensional process and material parameter spaces, making real-time decision-making and rapid convergence critical challenges. Although process conditions can be dynamically adjusted, decision performance depends on a computable description of material microstructure–property relationships. For this purpose, Ref. [126] proposes a predictive ML framework that reduces high-dimensional microstructure search spaces using data-driven methods, providing a computational basis for material design and process decision-making in smart manufacturing. Due to the high dimensionality of microstructure spaces, conventional optimization methods are inefficient and often fail to achieve optimal solutions. Ref. [126] uses the magneto strictive Fe–Ga alloy (Galfenol) to demonstrate the application of predictive ML to microstructure optimization and materials design. This framework includes random data generation, feature selection, and classification algorithms. Experimental results indicate highly efficient search space reduction, with the search range of some variables reduced by up to 99%.

4.4.2 Multiscale Simulation

In wafer fabrication, reliability and performance degrade due to cross-scale coupling effects. For example, in electronic packaging interconnect materials, such as solder, intermetallic compounds, and interfaces, reliability is jointly governed by angstrom-scale defects (vacancies, dislocations, and grain boundaries), microstructural evolution (grain growth, phase transformation, and accumulated plasticity), and macroscopic temperature and stress fields in the package structure. Under temperature fluctuations and power cycling, these effects manifest as thermomechanical fatigue. Therefore, the value of multiscale modeling lies in propagating microscale material properties to macroscale structural responses and establishing a computable and transferable information chain. For example, microscale models can be used to derive effective constitutive parameters or fatigue damage parameters, which are then incorporated into package-level FEM simulations.

In addition, multiscale simulation results can be used for engineering decisions and risk management. For example, multiscale multiphysics simulation can be embedded into DT frameworks and integrated with sensing and process data. AI surrogate models can accelerate microscale computations, enabling real-time lifetime monitoring, predictive maintenance, and parameter optimization. This approach transforms the cross-scale causal chain of materials–structure–process interactions into a continuously operating closed-loop decision engine, thereby supporting reliability, yield, and life-cycle resilience in semiconductor smart manufacturing.

4.4.2.1 Angstrom-Scale Defects, Diffusion, and Sintering

As process technology advances below the 5 nm node, local material behavior in solder joints and intermetallic compounds used in advanced 3D packaging and chip stacking is increasingly governed by angstrom-scale physical effects, including size-dependent discontinuities and microscopic degradation mechanisms, posing major challenges to system reliability. Conventional models based on macroscopic material parameters exhibit reduced predictive accuracy at the angstrom-scale. Accordingly, Ref. [129] employs MD simulations to analyze the thermomechanical properties of Sn–Ag–Cu (SAC) lead-free solder at the angstrom-scale. The simulations use a modified embedded-atom method (EAM) potential and investigate SAC105, SAC205, SAC305, and SAC405 alloys with varying silver content. The results indicate that increasing silver content enhances the uniaxial tensile strength and elastic modulus of SAC solder at the angstrom-scale. The coefficient of thermal expansion also increases, while fracture strain (ductility) and thermal conductivity decrease. The accuracy of the modified EAM potential parameters is validated. These results provide a physical basis for the design and manufacturing of angstrom-scale interconnect components in electronic packaging.

Ref. [128] employs MD simulations based on EAM to investigate the pressure less sintering kinetics of silver nanoparticles with sizes ranging from 4 to 20 nm. The sintering process can be divided into three main stages. The study indicates that pore geometry significantly affects the onset of densification and the densification rate in the first two stages. These effects are further influenced by sintering temperature, surface diffusion, grain boundary diffusion, and plastic deformation. Accordingly, the study develops a linearly scalable parallel MD code for simulation. The results resolve limitations of conventional microscale sintering theories in accurately predicting angstrom-scale behavior. The angstrom-scale sintering kinetics model established in this study provides a foundation for subsequent multiscale investigations that link process parameters with joint structure evolution.

Nanostructured copper is an ideal interconnect material for wide bandgap (WBG) semiconductors due to its high melting point, low cost, and high thermal conductivity. It enables low-temperature packaging and high-temperature operation. Accordingly, Ref. [130] employs MD simulations to study the sintering mechanisms of nanostructured copper and the subsequent mechanical behavior. The simulations use an EAM potential to accurately describe interatomic interactions in copper. Five models combining nano spherical and nanoflake structures are constructed to analyze the effect of geometry on sintering. Pressure less sintering is simulated in the temperature range of 500–650 K. The results indicate that mechanical properties increase with sintering temperature. Strength can be further enhanced by adjusting material composition, such as incorporating nanoflake structures.

WBG semiconductor devices operating at high temperatures can induce ratcheting in conventional solder joints, leading to packaging failure. Accordingly, Ref. [138] employs two viscoelastic constitutive models to predict ratcheting behavior in sintered nano silver lap shear joints. Experimental results indicate that the viscoelastic model with the Ohno–Wang and Armstrong–Frederick (OW–AF) hardening rules outperforms

the Anand model in predicting ratcheting behavior, particularly in capturing complex stress–strain responses at elevated temperatures.

Conventional shear or tensile tests for nano silver thin films suffer from limited alignment accuracy, difficulty in controlling thickness due to solvent evaporation, and a large strain-rate mismatch between nanoindentation and uniaxial testing. Accordingly, Ref. [131] proposes a nanoindentation analysis method and introduces a rate factor to bridge the data gap between indentation tests and uniaxial tensile tests. Experimental results indicate that the rate factor can convert indentation strain rates into corresponding uniaxial strain rates, effectively reducing the discrepancy between the two by up to two orders of magnitude.

Silicon carbide (SiC) devices can experience a mismatch in thermal expansion coefficient between the chip and the substrate. Under power cycling loading, this mismatch induces fatigue cracking in sintered silver interconnect layers. Accordingly, Ref. [139] combines FEM with a fracture phase-field method to simulate crack initiation and propagation. The simulations indicate that fatigue cracks initiate at the chip-side corners and propagate toward the center until forming through-thickness cracks. These results support reliability optimization of die-attachment structures.

Random microporous structures generated during sintering of nano silver paste affect the mechanical strength, thermal conductivity, and fatigue life of nano silver joints. Conventional homogenization models cannot capture these effects. Accordingly, Ref. [140] uses a Gaussian random field to automatically generate stochastic microporous geometries and applies a phase-field method to simulate crack initiation and evolution under thermomechanical conditions. The simulation results are highly consistent with experimentally observed cracking behavior and enable the prediction of fatigue failure in the sintered layer under thermal cycling.

High-temperature cycling in WBG semiconductors can induce thermomechanical fatigue failure in sintered silver interconnect materials. Accordingly, Ref. [141] integrates insulated-gate bipolar transistor (IGBT) chips using sintered silver onto multiple substrates, including printed circuit boards, insulated metal substrates (IMS), copper lead frames, and direct-bonded copper. The assemblies are evaluated using a combination of active power cycling tests, passive temperature cycling tests, and FEM. Experimental results indicate that the copper lead frame substrate exhibits the highest risk of delamination in the sintered layer, consistent with simulation results. Two failure mechanisms are observed, including wire bond lift-off and die-attach degradation.

4.4.2.2 Multiscale Coupling and Multiphysics

MD simulations exhibit low computational efficiency when modeling large spatial domains and long-time scales in fluid systems. Conventional continuum mechanics also becomes invalid at the angstrom-scale due to molecular effects. Multiscale approaches that couple MD with the Navier–Stokes equations are complex and may introduce interfacial noise or violate conservation laws. Accordingly, Ref. [132] proposes dissipative particle dynamics as a mesoscale simulation method that bridges microscopic MD and macroscopic continuum fluid dynamics. The method is validated by reproducing known polymer scaling laws and by accurately simulating polymer chain behavior in dilute solutions. In addition, Ref. [346] proposes an open-source, highly modular, and future-proof software architecture named Cactus. The framework emphasizes abstract mechanisms for parallelization, data exchange, and module management to support cross-platform execution and co-simulation of multiphysics modules. Although the system is not specifically designed for materials or semiconductor process simulation, its modular design and multiphysics integration principles provide an important systems engineering reference for the development of multiscale and multiphysics simulation frameworks.

Developing multiphysics and multiscale simulations faces challenges in coupling heterogeneous computational models, high replacement costs of sub models, and frequent interface modifications across different spatial and temporal scales. Accordingly, Ref. [347] proposes a coupling framework, the distributed spatiotemporal coupling library, based on the complex automata paradigm and agent technologies. Each sub-simulation model is encapsulated as an independent software agent using Java Agent Development Framework middleware. Experimental results demonstrate that the platform enables the coordinated operation of multidisciplinary models spanning physics, biology, and biochemistry across different spatial and temporal scales. This software architecture-centric coupling approach provides a reusable system design paradigm for subsequent multiscale and multiphysics simulation systems, including materials and process simulations.

4.4.3 Surrogate Model

In ESA, materials surrogate models serve as computational interfaces between high-fidelity material physics and system decision, enabling the integration of material behavior into closed-loop DT systems. From a DT perspective, real-time process-side decisions, such as reliability risk assessment of packaging and interconnects, parameter adjustment, and maintenance scheduling, are often constrained by the computational latency of multiscale and Multiphysics material models that incorporate microstructure evolution, thermo-Elastoplast city, and damage. Therefore, surrogate models are required to provide near real-time predictive feedback. In addition, the training pipeline of surrogate models must be engineered. This includes data acquisition (experimental design and simulation sampling), model training, validation, and deployment, forming a traceable, maintainable, and modular workflow. Such a framework enables continuous updating and reliable operation across different material systems, process windows, and even varying equipment or batch conditions.

Further, treating surrogate model training as an integral part of a digitalized workflow, rather than as a one-time research task, is critical for smart manufacturing. Conventional finite element analysis (FEA) is computationally expensive when modeling complex, randomly distributed two-phase microstructures, due to the need for large representative volume elements. Accordingly, Ref. [142] proposes a hybrid approach combining supervised ML with FEA. A microstructure generation algorithm is used to produce random two-phase structures with varying volume fractions, size distributions, and morphologies. FEA is performed to compute elastic modulus, Poisson's ratio, and phase stress as ground truth data. After extracting key geometric features, ANN and ensemble forest methods are trained. The results indicate accurate prediction of homogenized linear elastic properties.

Full-field simulations can capture the linkage between microstructure and material properties but incur high computational cost. Accordingly, Ref. [143] proposes a method that integrates Gaussian process regression from ML to predict unknown responses, including their uncertainty, and to enable selective execution of full-field simulations in regions with high model uncertainty, thereby improving computational efficiency. The method is evaluated on a dataset of 1100 synthetic two-phase microstructures. The results indicate that the approach can identify microstructures that maximize target properties within the dataset using only a small number of full-field simulations.

Polycrystalline materials, such as steels, incur extremely high computational costs when high-resolution simulations of mechanical behavior are performed using conventional numerical methods, such as spectral methods or FEM. Accordingly, Ref. [144] proposes a CNN based on a U-Net architecture. Numerical methods are used to solve boundary-value problems for polycrystalline microstructures under quasi-static uniaxial loading, and the resulting von Mises stress fields serve as reference models for training the CNN.

Experimental results demonstrate that the trained CNN predicts stress fields approximately 500 times faster than conventional spectral methods.

Existing CNN surrogate models for microstructure analysis often cannot simultaneously incorporate structural geometry and diverse material property information as inputs. Accordingly, Ref. [145] proposes a method that encodes material properties into microstructure images. The properties of each phase are transformed into higher-order tensors and combined with grayscale microstructure images, enabling the model to capture the spatial distribution of material properties. Evaluation on 5000 test samples within the training domain indicates that 96% of predictions achieve an absolute percentage error below 5%.

Conventional high-resolution simulations are computationally inefficient for evaluating the effective mechanical properties of periodic composite materials. Accordingly, Ref. [146] proposes a CNN surrogate model that represents periodic unit-cell structures using binary bitmap images. A multilayer neural network is used to establish a nonlinear mapping from structural pixel data to mechanical properties. The trained surrogate model enables rapid evaluation of mechanical properties and can serve as an alternative to conventional full-field simulations. Experimental results indicate that the optimal structures identified by the surrogate model exhibit mechanical performance highly consistent with solutions obtained from traditional topology optimization methods, such as the solid isotropic material with penalization approach. This study demonstrates that CNN surrogate models can provide a fast alternative to high-resolution simulations. The image structural representation and homogenization prediction framework offer a reference methodology for developing material surrogate models in semiconductor packaging and process DT systems.

ANNs in MM often lack physical consistency, leading to poor extrapolation and predictions that may violate physical laws. Accordingly, Ref. [147] proposes an elastoplastic neural network as a surrogate model for classical elastoplastic constitutive relations. This method replaces traditional yield functions, plastic potentials, and flow rules with more flexible algorithmic representations and incorporates physical constraints into the loss function during training. Experimental results demonstrate that this method outperforms a CNN in predicting sand behavior with different initial densities under unseen strain-controlled loading paths.

In many complex systems, spatiotemporal data of pressure or temperature can be observed, but the underlying governing physical laws remain unknown. In addition, differentiation of spatial and temporal observational data is highly sensitive to noise, with small measurement errors being significantly amplified during differentiation, leading to model instability. Accordingly, Ref. [148] proposes Deep Hidden Physics Models that directly identify unknown nonlinear partial differential equations as the governing dynamics from sparse, potentially noisy spatiotemporal observations. Experimental results indicate that the models can accurately reconstruct the underlying dynamics and forecast future system states.

4.4.4 Materials Simulation under Variations

Process variations affect chip performance, power consumption, yield, and reliability. Variations typically manifest at multiple scales, including within-wafer, wafer-to-wafer, within-lot, and lot-to-lot. These variations originate from stochastic deviations in key process parameters, such as lithography, etching, thin-film deposition, and doping, and directly induce shifts in electrical parameters, including threshold voltage, channel length, and channel width. In addition, re-entry, multi-product, large-scale equipment fleets, and long process sequences introduce coupled effects from equipment heterogeneity, drift, failures, measurement errors, and scheduling variations. These factors impact system stability, traceability, and productivity. In ESA, material and process models must address interactions among variations, spatiotemporal correlations, and multi-source uncertainties to serve as closed-loop decision components in DT systems. Material

uncertainty and equipment variability must be integrated into a unified system framework. Besides, real-time variation management strategies must be supported by quantifiable robustness metrics, including sensitivity to variation, risk boundaries, and feasible process windows. These conditions can advance wafer fabrication toward next-generation smart manufacturing.

4.4.4.1 Incorporating Materials Physics into Process Simulation

When accurate and timely prediction of wafer surface geometry and composition evolution is lacking, chip and process development cycles are significantly prolonged. Accordingly, Ref. [160] proposes a virtual fabrication framework that integrates physically based models with topography simulation frameworks. The framework employs ballistic transport and reaction models to simulate transport and reaction kinetics at submicron feature scales, covering deposition, etching, flow, and planarization processes. Feature-scale simulation serves as a natural coupling point between atomistic and equipment-scale models, improving process predictability and enabling realistic simulation for process development and complex device manufacturing. Modern virtual fabrication systems are evolving toward comprehensive DT architectures, and topography simulation is a key technology for mapping equipment parameters to physical wafer states.

In lithography processes, stochastic disturbances and drift arising from equipment, materials, and environmental variations lead to deviations in the critical dimension and sidewall angle of photoresist patterns from their target values, as well as print bias between isolated and dense lines. Accordingly, Ref. [149] proposes three run-to-run (R2R) feedback controllers combined with a Kalman filtering scheme to estimate and compensate for process disturbances. Focus is introduced as an additional control input, enabling a multi-input multi-output control architecture. Simulation results demonstrate that these controllers effectively reduce the adverse impact of disturbances and print bias. By incorporating geometric quality metrics from lithography simulation into R2R control, this approach represents an early paradigm of a simulation–control closed loop.

Wafer fabrication faces challenges due to cross-wafer critical dimension non-uniformity, leading to variations in device performance, power consumption, and yield. Accordingly, Ref. [150] proposes a method that compensates for systematic critical dimension variations by adjusting the temperature distribution in the post-exposure bake step of the lithography process. The method leverages the higher spatial controllability of the lithography step to compensate for deviations introduced by downstream plasma etching processes, where spatial control is limited, such as etch-loading effects. Experimental results show that the temperature-to-offset model has higher fidelity and transferability and is better suited for critical dimension uniformity control.

In wafer fabrication, the interactions among etching, deposition, and CMP steps are difficult to predict. Accordingly, Ref. [161] proposes a feature-scale 2D physics simulator based on contact mechanics and integrates it into a process topography simulation environment. The model accounts for pad surface roughness and elastic deformation to compute pressure distribution and incorporates the Preston equation to determine the local material removal rate. The simulator supports multi-line two-step polishing simulations using different pads, slurries, and process parameters. The approach is validated by simulating barrier-layer deposition in oxide trenches and copper overfill processes, demonstrating the integration of the topography simulation module with the CMP module.

In wafer fabrication, plasma etching process development relies on time-consuming, costly manual trial-and-error design of experiments, while conventional technology computer-aided design simulations incur high computational costs. Accordingly, Ref. [162] proposes a high-efficiency process workflow that integrates 3D plasma physics simulation with ML for mechanism calibration and etch profile optimization.

Neural networks are used to generate samples and predict etch depth, critical dimension, and profile under different process conditions. Optimal physical parameters are rapidly identified based on target etching outcomes, enabling accurate model calibration. In addition, a 3D plasma etching physical model for shallow trench isolation is developed. Experimental results indicate a reduction in prediction time, with calibrated model achieving relative errors below 7%.

4.4.4.2 Incorporating Materials Physics into Process Simulation

In wafer fabrication, high production volume and high cost prevent full wafer or die-level quality measurement for every unit. Accordingly, Ref. [163] proposes a probabilistic virtual metrology model that integrates CNN with Gaussian process regression (GPR). CNN is used to extract key features from large-scale raw sensor data, which are then fed into the GPR model. The Bayesian formulation of GPR captures process stochasticity and its relationship with quality variables, providing both predictive mean and quantified uncertainty. Experimental results demonstrate that the proposed model effectively handles large-scale datasets and achieves a better balance between efficiency and prediction accuracy.

In CMP, the complex interaction between chemical and mechanical effects makes accurate prediction of the MRR challenging. Accordingly, Ref. [183] proposes a physics-informed multi-task learning (PI-MTL) model that integrates physical constraints into a CNN. This model is used to learn shared representations in the network's early layers. By learning multiple targets, the model captures interactions in the feature space. Domain-specific physical laws are incorporated into the DL model as constraints to guide learning. Experimental results indicate that auxiliary tasks improve the accuracy of the primary task. In addition, in plasma etching processes, optical emission spectroscopy (OES) data are high-dimensional and large-scale, causing conventional methods to lose critical physical information and reduce prediction accuracy. Accordingly, Ref. [184] proposes a novel VM method based on singular value decomposition (SVD). The wafer OES matrix signals are flattened into a single long vector, and SVD is applied to the training set to obtain low-rank projection features, which are used as inputs to ridge regression. The method is evaluated using real plasma etching datasets with five-fold cross-validation. The results show that SVD-based OES features consistently outperform baseline VM models that use selected wavelengths or spectral bands across all evaluation metrics.

Conventional wafer geometric quality metrics, such as total thickness variation, bow, and warp, cannot fully capture intra-wafer local variations and provide limited support for process improvement. Accordingly, Ref. [185] proposes a hierarchical statistical model that uses 2D data maps to characterize wafer geometry and thickness variation. The model decomposes wafer thickness variation into macro-scale and micro-scale components. The macro-scale mean trend is modeled using a cubic function. The micro-scale local variation is modeled using a first-order intrinsic Gaussian Markov random field. The effectiveness of the method is demonstrated through a real-world case study in a wafer fab.

In wafer fabrication, defect clustering leads to inaccurate predictions in conventional yield models. Accordingly, Ref. [186] proposes a yield model based on hierarchical Bayesian modeling that incorporates the spatial distribution of defects on wafers. Spatial coordinates on the wafer are used as covariates to construct four regression models, including Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial. A hierarchical Bayesian framework is employed to introduce random effects, enabling the model to capture both intra-wafer spatial characteristics and wafer-to-wafer variations in defect distributions. The model is validated using real wafer map datasets. The results show that these regression models provide more reliable and accurate yield estimation and prediction.

Conventional semiconductor yield models often suffer from degraded prediction performance due to violations of model assumptions, such as normality and independence, arising from non-normal data distributions and hierarchical correlations. Accordingly, Ref. [187] proposes the generalized linear mixed models for yield prediction. The study compares two approaches, batch-specific and population-averaged, and evaluates the effects of different link functions, sample sizes, and levels of data aggregation, including chip-level and wafer-level models, on prediction performance. Validation using yield datasets indicates superior predictive performance compared with conventional generalized linear models and previously reported models.

4.4.5 Data-Driven Reinforcement Learning

In wafer fabrication, materials evolve as dynamic systems under external process conditions. The role of data-driven RL is not to construct predictive models of material properties, but to couple material evolution with process decision and formulate the problem as a Markov decision process. Under the influence of temperature, pressure, plasma power, gas flow rate, and time, the system state interacts iteratively with control actions to learn an optimal policy that guides material states along desired evolution trajectories. Data-driven RL transforms material evolution from a static property-prediction problem into a dynamic optimal-control problem and enables policy learning under multi-step decision-making and delayed feedback. When material models are integrated with process DT systems to form real-time feedback loops, RL can further serve as the decision core for cross-scale process control. Material behavior is analyzed and actively optimized through data-driven control policies as an engineered system. However, the application of RL in materials and process systems requires that material models can be evaluated rapidly and repeatedly. Conventional DFT provides high accuracy but incurs prohibitive computational cost, making it unsuitable for multi-step decision and policy search. Accordingly, Ref. [152] proposes a neural network representation for constructing DFT potential energy surfaces, where the energy and atomic forces are expressed as functions of atomic positions. The model is validated using bulk silicon, demonstrating computational speeds several orders of magnitude faster than conventional DFT, while achieving higher accuracy than empirical potential models and maintaining close agreement with DFT results. This acceleration provides the computational foundation for repeated policy evaluation and closed-loop decision.

Conventional materials development relies on intuition and trial-and-error. It is inefficient when dealing with complex chemical compositions and large combinatorial design spaces. Accordingly, Ref. [151] proposes an adaptive design strategy tightly coupled with experiments to reduce the search space through sequential guidance. The study focuses on nickel–titanium-based shape memory alloys and establishes a closed-loop iterative system that integrates ML regression models, uncertainty estimation, and experimental feedback. Alloys selected during the experimental process outperform randomly selected candidates. This framework shows that materials discovery can be formulated as a sequential closed-loop iteration driven by uncertainty-guided exploration and experimental feedback, establishing a prototype of materials systems as interactive decisions. When these models are embedded in process DT systems with real-time feedback loops, RL can optimize policies across cross-scale material–process systems. Material behavior can be actively controlled through data-driven strategies as an engineered system.

4.5 High Performance Computing and Smart Manufacturing

In wafer fabs, equipment and sensors generate massive, high-frequency time-series data every second. This imposes bandwidth limitations and communication latency on conventional cloud computing for real-time quality control and robotic coordination. Edge computing addresses this limitation by performing preprocessing and real-time inference at the data source, shifting computation to the network edge, and

enabling millisecond-level response times for critical process parameters and predictive maintenance. This transition drives smart manufacturing from centralized control systems toward distributed architectures composed of HPC, cloud manufacturing, and edge intelligence. However, edge computing alone is insufficient for large-scale optimization across multiple fabs, which requires a cloud–edge–end architecture. End devices are responsible for data acquisition; the edge layer performs real-time decision-making and model compression optimization; and the cloud leverages HPC for large-scale data analytics and global model training. The deep integration of HPC with AI/ML relies on specialized hardware accelerators to overcome the memory wall. This integration ensures efficient execution and energy sustainability of complex DL models. The studies discussed in this subsection are summarized in Table 8. The table classifies the literature based on physical system levels and ESA functional roles, including modeling, diagnosis, and prediction, and control and optimization, to illustrate how HPC, edge computing, and cloud-based infrastructures support ESA in smart manufacturing.

Table 8: HPC, edge, and cloud computing research supporting ESA in semiconductor smart manufacturing.

Physical/ESA	Modeling	Diagnosis & Prediction	Control & Optimization
Materials			[153]
Equipment			
Robots/AMR			
OHT/AGV/AMR			
Sensor/IIoT	[315, 316]	[326–329]	[333–336]
Infrastructure (HPC/Edge/Cloud)	[348–353]	[361–365]	[367–375]

4.5.1 High Performance Computing for DT

In smart manufacturing, DT and cloud–edge architectures serve as surrogate models that connect physical systems with virtual representations. These systems rely on HPC for large-scale parallel computation to support real-time simulation and complex what-if scenario analysis. In wafer fabrication, HPC-enabled DT architectures allow engineering systems to identify subtle defects in virtual environments and autonomously optimize process parameters. This enables a transition from reactive operations to proactive, intelligent decision-making. With the adoption of extreme ultraviolet (EUV) lithography, feature sizes have scaled to the nm regime. Conventional scanning electron microscopy inspection requires comparison with large-scale design layout data and suffers from a shortage of defect samples for training ML models. Accordingly, Ref. [315] proposes an unsupervised learning method based on autoencoders. These images are decomposed into multiple sub-patterns, compressed into low-dimensional latent vectors using an autoencoder, and then reconstructed. A discrepancy index between the input and reconstructed images is used as the detection criterion, where large discrepancies indicate anomalies. The method effectively handles large-scale image data generated by high-throughput, high-speed inspection and captures subtle variations in process conditions.

DL imposes excessive computational demands on embedded hardware, resulting in inefficient deployment in production lines. Accordingly, Ref. [367] proposes a cloud–edge computing architecture. The framework employs CNN to integrate defect localization and classification within a unified network, enabling robust detection under complex backgrounds and of micro-scale defects. Edge computing provides high-throughput processing capability near the data source to mitigate latency constraints. The cloud layer is responsible for large-scale data storage and continuous model training and updating. Experimental results indicate high-accuracy detection within a very short response time and outperform conventional image classification methods. In addition, Ref. [368] integrates cloud computing and DL to establish a

DT framework for structural health monitoring. The architecture integrates physical structures, device-level measurements, and digital models to enable real-time monitoring. Cloud computing infrastructure and web-based applications are employed to enhance data interaction among physical structures, digital models, and human intervention. A fog computing layer is deployed to reduce the digital model's data volume and computational demand. Experimental results indicate a detection accuracy of 92% in damage identification. Besides, Ref. [348] proposes a high-precision, HPC-driven DT architecture that transforms HPC from offline simulation to a dynamic workflow. Tasks are dynamically allocated to edge, cloud, or HPC resources based on urgency and computational requirements. Experimental results show that the architecture achieves near-ideal strong scaling performance within practical resource limits when processing compute-intensive workflows, enabling DT to become more intelligent, responsive, and capable of handling highly complex systems.

Modern cloud data centers must handle increasingly complex, compute-intensive workloads, including DL, scientific computing, and big data analytics. Accordingly, Ref. [369] introduces the NVIDIA A100 GPU, based on the Ampere architecture, as a flagship hardware solution that delivers extreme performance and scalable flexibility. Validation results demonstrate that, in molecular dynamics, physics, and engineering sciences, the A100 achieves performance improvements of $1.5\times$ to over $2\times$ compared to the V100 GPU, and accelerates DL models for electronic design automation (EDA), photolithography simulation, and wafer defect inspection. In addition, HPC, AI, and data analytics workloads impose increasing demands on computational throughput, memory bandwidth, and data locality. Accordingly, Ref. [370] introduces the H100 Tensor Core GPU based on the new Hopper architecture. A dedicated DPX instruction set is introduced to accelerate inner-loop computations in complex algorithms. Within the existing CUDA hierarchy of threads, thread blocks, and grids, a new Thread Block Cluster level is added to improve data locality. Through hardware–software co-design innovations, including the introduction of a cluster-level hierarchy and asynchronous execution units, the H100 effectively addresses locality optimization challenges associated with scaling hardware resources.

4.5.2 Cloud HPC and Resource Scheduling

Smart manufacturing is a highly coupled engineering system spanning materials, equipment, and production lines. It involves multi-level decision-making, heterogeneous resource allocation, and dynamic data-flow control. Cloud computing and HPC provide the core capabilities to ensure system computability and schedulability. Manufacturing resources, including equipment, capabilities, knowledge, and data, can be encapsulated as callable cloud services through virtualization and service mechanisms, enabling dynamic allocation and sharing under the Everything-as-a-Service (XaaS) paradigm. HPC supports large-scale process simulation, production scheduling optimization, and multi-scenario what-if analysis, thereby determining system response time and decision quality. Meanwhile, IIoT connects equipment and sensors, generating high-frequency, large-volume time-series data streams. If processed only at edge nodes, data are constrained by computational capacity and memory. If fully transmitted to the cloud, they are subject to bandwidth constraints and latency. Therefore, distributed computing architectures and cloud resource scheduling are introduced. Wafer fabrication involves thousands of highly coupled steps, where any scheduling or parameter adjustment affects the entire system. Cloud HPC and resource scheduling enable timely global optimization, including complex rescheduling, process window simulation, yield prediction model training, and multi-scenario what-if analysis. By leveraging cloud HPC, wafer fabs can achieve real-time updates and dynamic simulation of DT, advancing smart manufacturing to the next level.

With the expansion of global IC foundry outsourcing and increasing product complexity, brand companies require robust platforms to manage dynamic variations. Although TSMC has developed the

eFoundry service, some companies lack the capability to build such large-scale systems and therefore rely on cloud manufacturing as an alternative. Accordingly, Ref. [371] proposes a Software-as-a-Service (SaaS) cloud architecture for the semiconductor industry, which virtualizes and services resources to enable more flexible global coordination. A case study demonstrates that cloud manufacturing can be effectively applied to outsourced processes, including wafer fabrication, packaging, and testing.

Wafer fabrication generates large volumes of heterogeneous data with diverse formats and high complexity, making integration and analysis challenging. Accordingly, Ref. [372] proposes a cloud-based big data lake architecture that includes a standardized pipeline for automated ingestion, processing, and storage from multiple data sources. The architecture stores large-scale, multi-source data in raw and fine-grained formats. It supports advanced analytics, process optimization, and predictive maintenance. In addition, cloud manufacturing cannot provide services effectively if it fails to sense and connect underlying physical resources. Accordingly, Ref. [334] proposes an IIoT cloud architecture that leverages Radio Frequency Identification (RFID), embedded systems, and various intelligent sensors to transform physical equipment and software systems into identifiable and connected digital nodes. The architecture adopts a five-layer design comprising the resource layer, perception layer, network layer, service layer, and application layer, covering hardware, computational, and intelligent resources. A case study on machine tools demonstrates that IIoT can connect underlying resources with cloud platforms and enable automated equipment management and control.

In cloud manufacturing, a granularity gap exists between distributed, heterogeneous physical resources and complex requirements. Resource providers and users describe functional characteristics at different levels of granularity, ranging from fine-grained parameter control of equipment to coarse-grained scheduling. Accordingly, Ref. [349] proposes a multi-granularity resource virtualization and sharing strategy that includes resource aggregation functions and clustering algorithms. Experimental results indicate that the strategy improves resource utilization and provides a more robust foundation for resource discovery and selection. In addition, many cloud systems adopt proprietary architectures tied to specific vendors, resulting in high switching costs and complex migration processes when enterprises operate across different business stages or environments. Accordingly, Ref. [373] proposes a Hybrid Manufacturing Cloud architecture that supports three deployment models: private, community, and public clouds. The architecture allows enterprises to define customized access control policies for each resource, preventing unauthorized access. A feasibility study conducted on a group of enterprises with complex resource access requirements demonstrates that the system can effectively handle complex authorization constraints.

4.5.3 Hybrid Models

Wafer fabrication spans plasma physics, heat transfer, and material phase transitions, as well as scheduling, quality control, and equipment health management. It constitutes a highly complex, multi-scale, multi-physics coupled dynamic engineering system with stringent requirements on modeling methodologies. Traditional physics-based numerical simulation models, such as FEA, provide strong interpretability and physical consistency in process design and analysis. However, they incur high computational costs, require difficult parameter identification, and are not suitable for real-time decision support. Besides, large volumes of process data highlight ML/DL in predictive maintenance and quality prediction. However, their black-box nature and limited extrapolation capability restrict their application. Therefore, hybrid models emerge to achieve a balance among accuracy, computational efficiency, and interpretability by embedding physical knowledge into ML surrogate models for numerical simulations and applying residual learning to correct errors in physics models. Besides, HPC-driven numerical simulations are difficult to apply to real-time decisions and dynamic optimization. Hybrid models construct high-fidelity surrogate models

that compress simulation time from hours to milliseconds. They also provide the theoretical and technical foundation for developing adaptive DT systems.

In CMP, strong assumptions in physics models and a lack of interpretability affect MRR predictions. Accordingly, Ref. [153] introduces a physics-informed ML framework. The physical model captures the microscopic contact behavior between the polishing pad and abrasives, as well as between abrasives and the wafer. ML is used to predict the asperity radius and density of the polishing pad, and predictions are fed back into the physical model to estimate MRR. Experimental results indicate high-accuracy prediction of MRR under various operating conditions. In addition, in full-chip CMP simulation, virtual dummy filling synthesis is inefficient due to the high computational cost of numerical gradient evaluation. Accordingly, Ref. [352] proposes a model to migrate the conventional full-chip CMP simulator to a neural network architecture and to integrate a multi-start sequential quadratic programming optimizer. Experimental results indicate significant acceleration in both objective function evaluation and gradient computation.

Data acquisition in complex physical, biological, and engineering systems is costly, leading to small-data learning that is difficult to converge to physically consistent solutions. Accordingly, Ref. [350] proposes a DL framework embedding physical laws into the loss function. The framework includes continuous-time and discrete-time models and automatically differentiates neural network outputs to compute physical residuals. The effectiveness is validated through fluid mechanics, quantum mechanics, reaction–diffusion systems, and nonlinear shallow water wave propagation. In addition, Ref. [361] proposes a framework incorporating simulation results from physics models as input features to the neural network and uses a penalty term to enforce consistency with governing physical equations and to reduce model variance to improve generalization performance on unseen data. In the lake water temperature modeling, experimental results indicate superior performance to purely data-driven models in both prediction accuracy and physical consistency. Ref. [362] proposes a model based on a cross physics–data fusion scheme. The model leverages physical cutting processes to extract latent feature information from unlabeled samples. Experimental results indicate improved prediction safety and reliability.

For discontinuous solutions of the Euler equations in gas dynamics, similar scenarios arise in wafer fabrication, such as high-speed gas deposition in CVD, plasma etching under sensor-limited conditions for flow dynamics simulation, and fluid simulation in precision nozzles with complex geometries. Accordingly, Ref. [351] employs a physics-informed neural network (PINN) framework that integrates the Euler equations (conservation laws) with initial and boundary conditions to construct a DL surrogate model for nonlinear partial differential equations. The method uses a small number of randomly distributed sampling points and applies automatic differentiation to compute the residuals of the Euler equations, enabling the solution of 1D and 2D problems involving contact discontinuities and oblique shock waves. The model is validated using Sod and Lax Riemann problems, and the results indicate strong agreement with benchmark solutions reported in the literature.

DL relies on large-scale labeled datasets but high-fidelity CFD simulation data are costly and difficult to obtain. Accordingly, Ref. [363] proposes a data-free, physics-constrained DL surrogate model driven by physical laws. The governing Navier–Stokes equations are incorporated into the loss function of DNN to guide model training. Labeled CFD data are not required, and model parameters are optimized by minimizing the residuals of the governing equations. Experimental results indicate strong agreement with conventional first-principles CFD simulations. In addition, Ref. [364] proposes a hybrid modeling framework integrating physics simulation data and experimental measurement data. High-fidelity CFD simulations and optical imaging experiments are combined to generate an aggregated dataset of melt pool width in laser powder bed fusion. Experimental results indicate the highest prediction accuracy among all compared models.

Computer Numerical Control machine prediction methods do not account for the diversity and consistency of equipment states across the lifecycle, resulting in inaccurate and untimely predictions. Accordingly, Ref. [326] proposes a hybrid predictive maintenance approach that integrates DT and data-driven components to reflect actual operating conditions. Data collected from various sensors are used to drive a remaining useful life prediction model. Experimental results indicate higher prediction accuracy than purely data-driven or physical approaches. In addition, functional component failures lead to reliability degradation. Accordingly, Ref. [328] proposes a DT real-time remaining useful life prediction method, adopting a sensing data acquisition approach to collect real-time operational data from the shop floor. A DT model of the shop floor is constructed, along with a prediction model based on nonlinear Brownian motion. The effectiveness of the method is validated through case studies. Besides, Ref. [327] proposes a ML framework integrating physical model predictions with *in-situ* sensing signals for real-time process monitoring. The framework is evaluated in laser powder bed fusion and directed energy deposition additive processes, achieving defect detection F-scores close to 90%.

4.5.4 Edge-HPC Integration

HPC is deployed in cloud or data center environments for high-fidelity numerical simulation, fab-wide optimization, and long-term model training. As IIoT is adopted, high-frequency data streams introduce latency and cause a bottleneck for centralized cloud architectures. Edge computing is used to shift computation closer to the data source to perform real-time preprocessing, anomaly detection, and preliminary inference, while reducing reliance on the cloud. To support simulation and model training, a Cloud-Edge-End architecture emerges to integrate real-time decisions at the edge with cloud HPC and to enable model deployment at the edge through model compression, hierarchical inference, and collaborative training. Besides, federated learning and collaborative inference allow multiple devices to update models without sharing raw data to enhance data privacy and system robustness. As a result, the Edge-HPC collaborative architecture is the enabler for DT systems.

Cloud systems cannot identify production line anomalies in real time, such as random machine failures and urgent tasks, nor support timely decisions. Accordingly, Ref. [336] proposes an adaptive anomaly control method driven by edge-cloud cooperation that leverages IIoT and embedded edge computing to achieve self-sensing, self-analysis, and self-optimization. A fuzzy Bayesian network is used to diagnose anomaly causes and extract key production indicators. Experimental results indicate improvement in anomaly detection, efficiency, and stability. In addition, Ref. [316] proposes an edge CNN model to transmit only the resulting parameters to a cloud server. An optimization mechanism is employed to reduce computational requirements. A testbed implemented on Google Cloud indicates reduced network traffic in IIoT environments. Besides, Ref. [329] proposes a computing framework that integrates fog computing, cloud computing, wireless sensor networks, and ML. ML algorithms are deployed on Amazon Elastic Compute Cloud to build predictive models. Wireless sensors are deployed to monitor vibration and energy consumption of power plant pumps and machines. Experimental results demonstrate that real-time machine state data can be collected to predict milling tool wear.

Ref. [374] proposes an end-to-end architecture that integrates 5G non-public networks, time-sensitive networking, and edge computing to enhance determinism and reliability within local networks and to reduce application latency. The architecture is applicable to AMHS in fabs, such as AGV and OHT. 5G provides stable wireless connectivity, while the edge architecture ensures that machine control commands are not delayed due to network congestion. In addition, frequent migration between edge and cloud environments increases the load on underlying networks. Ref. [375] proposes a software-defined networking edge-cloud interplay mechanism. The approach employs a multi-objective optimization algorithm and applies

the Tchebycheff decomposition method for traffic scheduling and routing. Experimental results indicate improved system performance in IIoT environments. Besides, Ref. [353] proposes a set of design guidelines for DT architectures based on IIoT and emerging protocols. Experimental implementations indicate that IIoT and communication protocols facilitate DT development and enable secure interaction.

Manufacturing systems must support real-time upgrading, downgrading, and modification of production plans to meet customized demands. Ref. [335] employs edge computing to integrate interconnected nodes into a distributed control architecture. Each node is equipped with holons, digital agents, and local DT. Experimental results demonstrate that the architecture enables local control, autonomous decision-making, and process optimization through information sharing and collaboration. In addition, Ref. [365] proposes a hybrid edge–cloud framework that integrates edge computing and cloud computing. Edge nodes execute latency-sensitive fault detection to reduce communication delay and data loss. Under normal operation, only low-resolution data are transmitted to save bandwidth. When anomalies are detected at the edge, full-resolution measurement data are transmitted to the control center. Experimental results indicate high reliability and automated fault localization for power grid monitoring.

5 Future Research

This study proposes the S_{IGLR} methodology and the ESA–MM–HPC perspectives to discuss relationships between industrial and technological evolution (as illustrated in Fig. 10) and provides system-driven directions for future research.

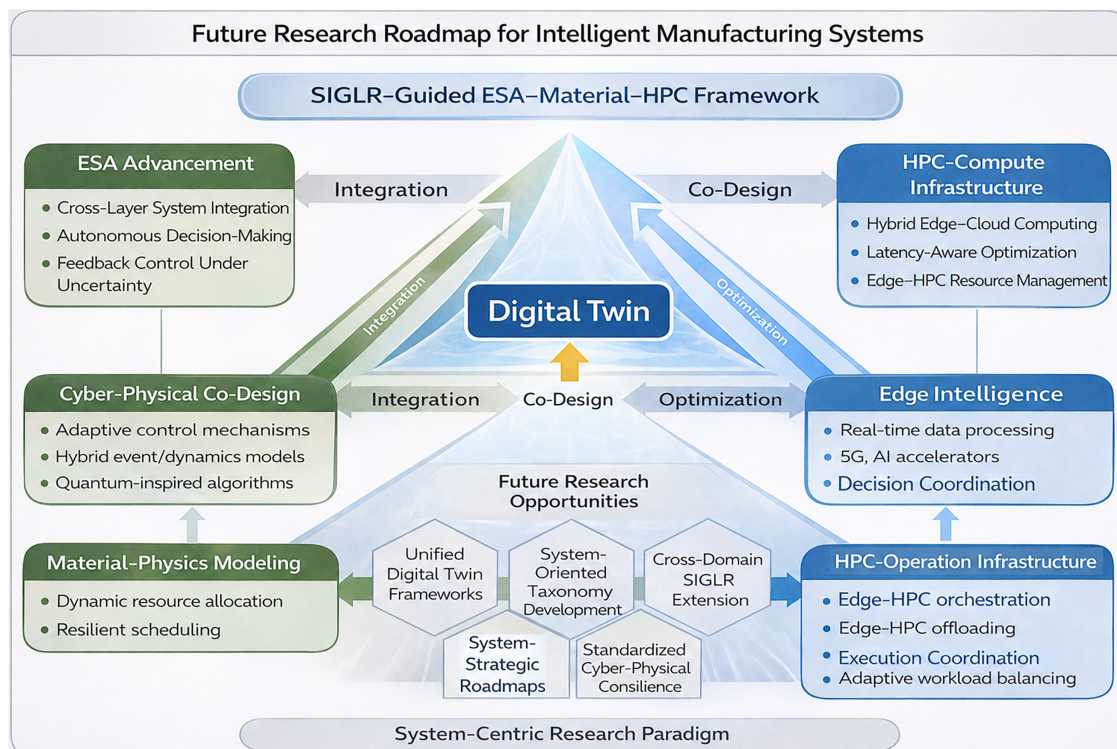


Figure 10: A S_{IGLR} -driven system-oriented future research roadmap based on the ESA–MM–HPC perspectives.

First, from ESA's perspective, future research should focus on cross-layer system integration to support real-time decision-making under multiple dynamic constraints. This direction includes the development

of adaptive control mechanisms, system-level optimization under uncertainty, and the integration of heterogeneous subsystems, such as robotic manipulators, autonomous transport systems, and process control modules. In semiconductor manufacturing environments, interactions between cyber–physical systems require new modeling approaches that simultaneously capture discrete events and continuous dynamics. Future research should therefore move toward topology-consistent system perspectives based on graph-theoretic formulations, rather than treating system components as independent units. Through distributed peer-to-peer network modeling approaches [104], physical flows, information propagation, and operational constraints can be examined within integrated modeling perspectives. On this basis, conservation laws (e.g., Laplacian flow) and topological consistency principles (e.g., sheaf theory) provide useful analytical perspectives for examining multidimensional dependencies and ordering structures across system layers. This perspective may help analyze relationships between local behaviors and global consistency to detect anomalies. Such structural representations can be further integrated with machine learning models for structured anomaly detection and predictive analytics [105], supporting predictive analysis within DT systems to explore how local variations propagate and amplify across the system, thereby affecting yield and production stability. This perspective suggests a direction for integrating structure-aware analysis with data-driven anomaly detection. The paradigm is naturally aligned with the ESA–MM–HPC triadic framework, in which structural modeling (ESA), physical consistency (MM), and computational scalability (HPC) are intrinsically coupled.

Second, in the Material–Physics domain, future research should further integrate multi-scale modeling and data-driven approaches. Existing studies typically treat material behavior and process parameters separately. Establishing a computable mapping between microstructural evolution and system-level performance is a critical requirement for next-generation semiconductor smart manufacturing. Integrating surrogate modeling and RL can help bridge this gap.

Third, the integration of cloud computing, edge computing, and HPC represents a direction for next-generation semiconductor smart manufacturing. Future research should investigate hybrid Edge–HPC architectures that enable real-time data processing at the edge while leveraging large-scale cloud computing resources. Key challenges include resource allocation, latency-aware scheduling, and the co-design of algorithms and computing architectures.

Fourth, coupling mechanisms among ESA, MM, and HPC represent an important research topic for achieving closed-loop optimization across the design, manufacturing, and operation stages in cross-layer systems. This is reflected in the interaction between data flow and control decisions, as well as in the interdependent relationships among physical constraints, system behavior, and computational resources. In this context, DT serves as a virtual representation and a cross-layer coupling mechanism that integrates physical modeling, system control, and computational optimization processes. Through bidirectional information feedback and constraint propagation, the system can achieve closed-loop decisions and dynamic adaptation within ESA. This integrates cyber–physical semiconductor manufacturing and support improved system coordination and operational stability.

Fifth, the proposed $S_{IG}LR$ framework provides a systematic approach to reorganizing the literature in complex engineering domains. By restructuring literature based on industrial system architectures, $S_{IG}LR$ is intended to examine relationships between technological developments and system functions and establishes a structured foundation for cross-domain knowledge integration. Based on this methodology, the study demonstrates that $S_{IG}LR$ has strong potential for cross-domain extension and can be applied to engineering domains such as robotics, autonomous vehicles, and smart energy systems, where technological evolution is driven by industrial development. This further indicates that a system-oriented taxonomy is not only feasible but can also provide a system-oriented perspective for cross-domain knowledge organization and

comparative analysis, thereby promoting a comprehensive understanding of complex engineering systems and the analysis of technological evolution.

Finally, this study indicates that future research may explore system-level co-design approaches, which integrate the optimization of material properties, control strategies, and computing architectures. This transition is critical for advancing next-generation smart manufacturing systems and reflects a broader shift in research paradigms from technology-driven approaches to system-oriented methodologies.

6 Conclusion and Limitations

6.1 Conclusion

This study discusses challenges of conventional algorithm-oriented literature review methods in complex engineering systems. It proposes a system-driven, industry-guided literature review framework, termed the System-under-Industry Guided Literature Review (S_{IGLR}). From the perspective of ESA for next-generation semiconductor smart manufacturing, the framework reorganizes and interprets cross-domain research. Overall, the study achieves predefined research objectives and analyzes the methodological value of a system-oriented literature review approach.

First, this study proposes a perspective consisting of the Control–Cyber Axis, Material–Physics Axis, and HPC–Compute Axis to characterize the structure of semiconductor smart manufacturing systems. This perspective complements conventional classifications based on individual technologies or algorithms. It integrates key domains, including automatic control, robotic manipulators, material modeling, high-performance computing, and digital twins, at the engineering system level. It provides an analytical approach for cross-layer and cross-domain integration.

Second, this study applies the S_{IGLR} methodology to systematically map the selected literature to system modules and cross-layer interactions. It integrates 103 survey papers and their extended references to construct a comprehensive knowledge structure for smart manufacturing research. Compared with conventional literature reviews, which often yield fragmented insights, this approach examines the functional roles and coupling relationships of key technologies, including deep learning, digital twins, and high-performance computing, within industrial systems. It further supports system-level discussion of large-scale complex engineering systems.

Finally, this study adopts an industrial perspective on semiconductor smart manufacturing and discusses challenges in current research to identify structural gaps in existing research, including system integration, physics-based modeling, computational resource allocation, and cross-layer decisions. Based on these findings, a system-oriented future research roadmap is proposed that emphasizes the co-evolution of physical modeling capabilities, system-level decision mechanisms, and high-performance computing infrastructures.

In addition, the proposed S_{IGLR} provides a system-oriented approach for organizing literature in complex industrial contexts. By structuring the literature around industrial architectures, S_{IGLR} is intended to examine relationships among technologies and their functional roles in real-world systems. Furthermore, the ESA–MM–HPC is a system-oriented conceptual perspective for examining semiconductor smart manufacturing systems. It highlights the importance of considering interactions among physical models, system control, and computational capabilities, rather than from improvements in any single technology.

Overall, this study presents the value of incorporating system-oriented perspectives into next-generation smart manufacturing research. The proposed methodology provide useful insights for future studies in other complex engineering domains. They can be extended to other complex engineering systems,

such as electric vehicles, unmanned aerial vehicles, and smart supply chains. The approach also supports broader discussion between academic research and industrial technology development.

6.2 Limitation

Although this study discusses cross-domain integration and in-depth system-level analysis, several limitations should be noted. First, this study integrates 103 survey papers and their extended references, totaling 13,377. However, literature selection is still influenced by the search strategy and inclusion/exclusion criteria defined in the S_{IGLR} methodology. Because S_{IGLR} emphasizes literature screening based on semiconductor industrial system architectures, certain domain-specific or emerging research topics may not be fully covered.

Second, this study adopts TSMC as an industrial proxy for semiconductor smart manufacturing. While it represents advanced process technologies and highly automated manufacturing environments, it may introduce industry-specific bias. System characteristics in other manufacturing contexts, such as small and medium-sized fabs or different industrial domains, may not be fully captured by the proposed analytical framework. In addition, the operational characteristics, automation strategies, and infrastructure conditions observed in advanced-node semiconductor fabs may not generalize to legacy-node manufacturing environments, or non-Taiwanese semiconductor ecosystems.

Third, ESA-MM-HPC is a perspective derived from a system-level concept. It has not yet been developed into a quantifiable or formally verifiable model. Its applicability in specific application scenarios requires further validation through empirical studies. In addition, the proposed perspective primarily focuses on engineering, materials, and computational infrastructures, while other dimensions such as supply chain constraints, economics, cybersecurity, and operational policies remain outside the current scope. Operational systems such as MES, APC, FDC, VM, and metrology are treated as application-level implementations within broader ESA-enabled manufacturing environments rather than as independent dimensions of the proposed perspective.

Finally, although the S_{IGLR} methodology provides a system-oriented approach to literature review, its cross-domain applicability requires further evaluation. The current analysis focuses on semiconductor smart manufacturing. Future research should extend this methodology to other complex engineering systems to assess the feasibility of applying the proposed taxonomy.

Acknowledgement: The author used ChatGPT 5.3 as a semantic-assistance tool during the preliminary literature screening process. ChatGPT 5.3 was used to support preliminary semantic grouping and relevance prioritization. All final decisions regarding literature inclusion and exclusion were made solely by the author through manual verification based on predefined screening criteria.

Funding Statement: The author received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Hsiao-Chun Han, Der-Chen Huang and Chin-Ling Chen made substantial contributions to the conception and design. Hsiao-Chun Han was involved in drafting the manuscript. Der-Chen Huang acquired the data and conducted the analysis and interpretation. The critically important intellectual contents of this manuscript were revised by Chin-Ling Chen. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data used in the study are from IEEE Xplore, ScienceDirect, SpringerLink, ACM Digital Library, and arXiv.

Ethics Approval: This manuscript is non-biological or non-medical.

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

Abbreviations

The following abbreviations are used in this manuscript

AEC	Advanced equipment control
ANN	Artificial neural networks
ALIGNN	Atomistic line graph neural network
AMHS	Automated material handling systems
AMR	Autonomous mobile robots
CVD	Chemical Vapor Deposition
CF	Clearing functions
MCT	Completion time
CFD	Computational fluid dynamics
CAD	Computer-Aided Design
CNN	Convolutional neural networks
CoO	Cost of ownership
CPPS	Cyber-Physical Production Systems
DCNN	Deep convolutional neural networks
DL	Deep learning
DQN	Deep Q-network
DTL	Deep transfer learning
D–H	Denavit–Hartenberg
DOE	Design of experiments
DNS	Direct numerical simulation
DAP	Dynamic Adaptive Planning
EDA	Electronic design automation
ESA	Engineering System Analysis
EFEM	Equipment Front End Module
FDC	Fault Detection and Classification
FEM	Finite Element Method
FOUP	Front Opening Unified Pods
FB	Function Blocks
GAN	Generative adversarial networks
GNN	Graph neural networks
HPC	High performance computation
IBVS	Image-based visual servoing
IGBT	Insulated gate bipolar transistor
IP	Intellectual property
IK	Inverse kinematics
LPBF	Laser powder bed fusion
LSTM	Long Short-Term Memory
MRD	Magnetorheological damper
MDP	Markov decision process
MRR	Material removal rate
MOF	Metal–organic frameworks
MD	Molecular Dynamics
MLP	Multiple multilayer perceptron
ODR	On-time delivery rate
OES	Optical emission spectroscopy
OHT	Overhead Hoist Transport
PSO	Particle swarm optimization

PINN	Physics-informed neural networks
PdM	Predictive maintenance
PID	Proportional Integral Derivative
RF spark	Radio-frequency sparks
RNN	Recurrent neural networks
RL	Reinforcement learning
R2R	Run-to-run
SBC	Secure-by-construction
SiC	Silicon carbide
SAC	Sn–Ag–Cu
SOP	Standard operating procedures
SVID	Status Variable ID
SoC	Systems-on-chip
US	The United States
3D	Three-dimensional
2D	Two-dimensional
V2V	Vehicle-to-vehicle
WAT	Wafer acceptance test
WBG	Wide bandgap
APC	Advanced process control
APF	Artificial Potential Field
AGV	Automated guided vehicles
AIMM	Autonomous industrial mobile manipulator
CMP	Chemical mechanical planarization
CP	Chip probing
ACO	Colony optimization
CDR	Composite dispatching rules
CME	Computational materials engineering
(CDC)	Continuous damping control
CDA	Cooperative driving automation
CGCNN	Crystal graph convolutional neural network
CPS	Cyber–physical systems
DDPG	Deep Deterministic Policy Gradient
DNN	Deep neural networks
DRL	Deep RL
DOF	Degree-of-freedom
DFT	Density functional theory
DT	Digital twin
DRO	Distributionally robust optimization
EV	Electric vehicle
EAM	Embedded atom method
EES	Equipment Engineering System
EUV	Extreme ultraviolet
FNN	Feedforward neural network
FEA	Finite element analysis
FOC	Fractional-order control
GAIL	Generative Adversarial Imitation Learning
GA	Genetic algorithm
HIL	Hardware-in-the-loop

HLS	High-level synthesis
IIoT	Industrial Internet of Things
IC	Integrated circuit
CIRP	International Academy for Production Engineering
KNN	k-Nearest Neighbors
LPV	Linear parameter-varying
ML	Machine learning
MES	Manufacturing Execution System
MM	Material modeling
MEGNet	Materials graph network
MPC	Model predictive control
MAS	Multi-agent system
CNF	Nonlinear feedback
OCD	Optical critical dimension
OFE	Overall fab efficiency
PDE	Partial differential equations
PMSM	Permanent magnet synchronous motors
PBVS	Position-based visual servoing
PHM	Prognostics and health management
RFID	Radio Frequency Identification
RRT	Rapidly-exploring Random Tree
RGB	Red, Green, and Blue
RUL	Remaining useful life
SEM	Scanning electron microscopy
SAVGS	Serial active variable geometry suspension
SVD	Singular value decomposition
SVM	Space vector modulation
SPC	Statistical Process Control
SLR	Systematic Literature Review
S _{IG} LR	System-under Industry Guided Literature Review
TSMC	Taiwan Semiconductor Manufacturing Company
TL	Transfer learning
UAV	Unmanned Aerial Vehicles
VM	Virtual Metrology
WBM	Wafer Bin Maps
WIP	Work-in-process

References

1. Wang Y, Su Z, Guo S, Dai M, Luan TH, Liu Y. A survey on digital twins: architecture, enabling technologies, security and privacy, and future prospects. arXiv:2301.13350. 2023.
2. Saeedi S. Digital twins and their applications in modeling different levels of manufacturing systems: a review. arXiv:2511.06119. 2025.
3. Al Shaikh H, Bin Monjil M, Chen S, Asadizanjani N, Farahmandi F, Tehranipoor M, et al. Digital twin for secure semiconductor lifecycle management: prospects and applications. arXiv:2205.10962. 2022.
4. Wang K, Wang Y, Li Y, Fan X, Xiao S, Hu L. A review of the technology standards for enabling digital twin. Digit Twin. 2024;1(1):4. doi:10.12688/digitaltwin.17549.2.
5. Wang CC, Chien CH. Machine learning for industrial optimization and predictive control: a patent-based perspective with a focus on Taiwan's high-tech manufacturing. Processes. 2025;13(7):2256. doi:10.3390/pr13072256.

6. Li C, Zheng P, Yin Y, Wang B, Wang L. Deep reinforcement learning in smart manufacturing: a review and prospects. *CIRP J Manuf Sci Technol.* 2023;40(6):75–101. doi:10.1016/j.cirpj.2022.11.003.
7. Liu X, David I. AI simulation by digital twins: systematic survey of the state of the art and a reference framework. In: *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*; 2024 Sep 22–27; Linz, Austria. p. 401–12. doi:10.1145/3652620.3688253.
8. Espadinha-Cruz P, Godina R, Rodrigues EMG. A review of data mining applications in semiconductor manufacturing. *Processes.* 2021;9(2):305. doi:10.3390/pr9020305.
9. Chowdhury H. Semiconductor manufacturing process improvement using data-driven methodologies. Preprints. 2023. doi:10.20944/preprints202310.0056.v1.
10. Khakifirooz M, Fathi M, Wu K. Development of smart semiconductor manufacturing: operations research and data science perspectives. *IEEE Access.* 2019;7:108419–30. doi:10.1109/ACCESS.2019.2933167.
11. Huang AC, Meng SH, Huang TJ. A survey on machine and deep learning in semiconductor industry: methods, opportunities, and challenges. *Cluster Comput.* 2023;26(6):3437–72. doi:10.1007/s10586-023-04115-6.
12. Lee J, Su H, Macchi M, Polenghi A, Wu W, Zhao Z, et al. 2026 roadmap on artificial intelligence and machine learning for smart manufacturing. *Mach Learn Eng.* 2026;2(2):022001. doi:10.1088/3049-4761/ae5967.
13. Soori M, Arezoo B, Dastres R. Digital twin for smart manufacturing: a review. *Sustain Manuf Serv Econ.* 2023;2:100017. doi:10.1016/j.smse.2023.100017.
14. Zemskov AD, Fu Y, Li R, Wang X, Karkaria V, Tsai YK, et al. Security and privacy of digital twins for advanced manufacturing: a survey. *arXiv:2412.13939.* 2024.
15. Zhong RY, Xu X, Klotz E, Newman ST. Intelligent manufacturing in the context of industry 4.0: a review. *Engineering.* 2017;3(5):616–30. doi:10.1016/J.ENG.2017.05.015.
16. Huang Z, Shen Y, Li J, Fey M, Brecher C. A survey on AI-driven digital twins in industry 4.0: smart manufacturing and advanced robotics. *Sensors.* 2021;21(19):6340. doi:10.3390/s21196340.
17. Goecks LS, Habekost AF, Coruzzolo AM, Sellitto MA. Industry 4.0 and smart systems in manufacturing: guidelines for the implementation of a smart statistical process control. *Appl Syst Innov.* 2024;7(2):24. doi:10.3390/asi7020024.
18. Wang Y, Kang X, Chen Z. A survey of Digital Twin techniques in smart manufacturing and management of energy applications. *Green Energy Intell Transp.* 2022;1(2):100014. doi:10.1016/j.geits.2022.100014.
19. Liu Z, Liu Q, Xu W, Wang L, Zhou Z. Robot learning towards smart robotic manufacturing: a review. *Robot Comput Integr Manuf.* 2022;77:102360. doi:10.1016/j.rcim.2022.102360.
20. Wang Y. Flexible robots in the smart manufacturing industry: review. In: *Proceedings of the 2025 2nd International Conference on Electrical Engineering and Intelligent Control (EEIC 2025)*; 2025 Jul 3–4; Auckland, New Zealand. p. 716–27. doi:10.2991/978-94-6463-864-6_60.
21. Pantano M, Eiband T, Lee D. Capability-based frameworks for industrial robot skills: a survey. In: *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*; 2022 Aug 20–24; Mexico City, México. p. 2355–62. doi:10.1109/CASE49997.2022.9926648.
22. Daneshmand M, Noroozi F, Corneanu C, Mafakheri F, Fiorini P. Industry 4.0 and prospects of circular economy: a survey of robotic assembly and disassembly. *arXiv:2106.07270.* 2021.
23. Adebayo RA, Obiuto NC, Olajiga OK, Festus-Ikhuoria IC. AI-enhanced manufacturing robotics: a review of applications and trends. *World J Adv Res Rev.* 2024;21(3):2060–72. doi:10.30574/wjarr.2024.21.3.0924.
24. Sameh A, Fanni M, Rashad M. Advances in intelligent industrial manipulators for smart manufacturing and standardized automation technologies. *Discover Robot.* 2025;1(1):12. doi:10.1007/s44430-025-00012-2.
25. Urrea C, Kern J. Recent advances and challenges in industrial robotics: a systematic review of technological trends and emerging applications. *Processes.* 2025;13(3):832. doi:10.3390/pr13030832.
26. Jin S, Ke K, Gao B, Fu L, Huang X. A comprehensive review of key technologies for robot motion planning in contact tasks in industrial automation scenarios. *Chin J Mech Eng.* 2025;38(1):198. doi:10.1186/s10033-025-01356-x.
27. Luo S, Zhang M, Zhuang Y, Ma C, Li Q. A survey of path planning of industrial robots based on rapidly exploring random trees. *Front Neurorobot.* 2023;17:1268447. doi:10.3389/fnbot.2023.1268447.

28. Song X, Gao H, Ding T, Gu Y, Liu J, Tian K. A review of the motion planning and control methods for automated vehicles. *Sensors*. 2023;23(13):6140. doi:10.3390/s23136140.
29. Liu J, Yap HJ, Khairuddin ASM. Review on motion planning of robotic manipulator in dynamic environments. *J Sens*. 2024;2024(1):5969512. doi:10.1155/2024/5969512.
30. Ghodsian N, Benfriha K, Olabi A, Gopinath V, Arnou A. Mobile manipulators in industry 4.0: a review of developments for industrial applications. *Sensors*. 2023;23(19):8026. doi:10.3390/s23198026.
31. Calzada-Garcia A, Victores JG, Naranjo-Campos FJ, Balaguer C. A review on inverse kinematics, control and planning for robotic manipulators with and without obstacles via deep neural networks. *Algorithms*. 2025;18(1):23. doi:10.3390/a18010023.
32. Zhu M, Gong D, Zhao Y, Chen J, Qi J, Song S. Compliant force control for robots: a survey. *Mathematics*. 2025;13(13):2204. doi:10.3390/math13132204.
33. Tsuji T, Kato Y, Solak G, Zhang H, Petrić T, Nori F, et al. A survey on imitation learning for contact-rich tasks in robotics. *arXiv:2506.13498*. 2025.
34. Singh A, Kalaichelvi V, Karthikeyan R. A survey on vision guided robotic systems with intelligent control strategies for autonomous tasks. *Cogent Eng*. 2022;9(1):2050020. doi:10.1080/23311916.2022.2050020.
35. Zhang Y. Vision servo of industrial robot: a review. *AIP Conf Proc*. 2018;1955:040125. doi:10.1063/1.5033789.
36. Botezatu AP, Burlacu A. A short review of deep learning methods in visual servoing systems. *Electr Eng Power Eng Electron Sect*. 2023;69(3):113–36. doi:10.2478/bipie-2023-0018.
37. Phatthamolrat N, Tongloy T, Boonsang S, Chuwongin S. Deep learning based visual servo for autonomous aircraft refueling. *Eng Rep*. 2025;7(3):e70055. doi:10.1002/eng2.70055.
38. Besigomwe K. Closed-loop manufacturing with AI-enabled digital twin systems. *Cognizance J Multidiscip Stud*. 2025;5(1):18–38. doi:10.47760/cognizance.2025.v05i01.002.
39. Guha B, Moore S, Huyghe JM. Conceptualizing data-driven closed loop production systems for lean manufacturing of complex biomedical devices—a cyber-physical system approach. *J Eng Appl Sci*. 2023;70(1):50. doi:10.1186/s44147-023-00222-8.
40. Santosh T, Sahu PK, SR Parthasarathy P. Advancing cleanroom contamination control strategies with automation and AI: current status and future perspectives in the manufacturing of parenterals. *Curr Trends Biotechnol Pharm*. 2025;19(2):2330–47. doi:10.5530/ctbp.2025.2.20.
41. Dobrilovic D, Brtka V, Stojanov Z, Jotanovic G, Perakovic D, Jausevac G. A model for working environment monitoring in smart manufacturing. *Appl Sci*. 2021;11(6):2850. doi:10.3390/app11062850.
42. Montini E, Daniele F, Agbomemewa L, Confalonieri M, Cutrona V, Bettoni A, et al. Collaborative robotics: a survey from literature and practitioners perspectives. *J Intell Rob Syst*. 2024;110(3):117. doi:10.1007/s10846-024-02141-z.
43. Pan YJ, Buchanan S, Chen Q, Wan L, Chen N, Forbrigger S, et al. Survey on recent advances in planning and control for collaborative robotics. *IEEE J Ind Appl*. 2025;14(2):139–51. doi:10.1541/ieejia.24005652.
44. Dhanda M, Rogers BA, Hall S, Dekoninck E, Dhokia V. Reviewing human-robot collaboration in manufacturing: opportunities and challenges in the context of industry 5.0. *Robot Comput Integr Manuf*. 2025;93(5):102937. doi:10.1016/j.rcim.2024.102937.
45. Lu Y, Xu X, Wang L. Smart manufacturing process and system automation-a critical review of the standards and envisioned scenarios. *J Manuf Syst*. 2020;56:312–25. doi:10.1016/j.jmsy.2020.06.010.
46. Pereira MÂ, Vieira G, Varela L, Putnik G, Cruz-Cunha M, Santos A, et al. Manufacturing management processes integration framework. *Appl Sci*. 2025;15(16):9165. doi:10.3390/app15169165.
47. Balaha F, Albinali H, Alrabiah H, Ali M, Bahroun Z. An analytical review of data integration for decision support in smart manufacturing. *Decis Anal J*. 2025;17(2):100647. doi:10.1016/j.dajour.2025.100647.
48. Li C, Lei J, Yang L, Xu W, You Y. Research on electric vehicle powertrain systems based on digital twin technology. *Electronics*. 2024;13(20):4103. doi:10.3390/electronics13204103.
49. Lee J, Chua PC, Chen L, Ng PHN, Kim Y, Wu Q, et al. Key enabling technologies for smart factory in automotive industry: status and applications. *Int J Precis Eng Manuf-Smart Tech*. 2023;1(1):93–105. doi:10.57062/ijpem-st.2022.0017.

50. Tao F, Qi Q, Wang L, Nee AYC. Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: correlation and comparison. *Engineering*. 2019;5(4):653–61. doi:10.1016/j.eng.2019.01.014.
51. Monostori L, Kádár B, Bauernhansl T, Kondoh S, Kumara S, Reinhart G, et al. Cyber-physical systems in manufacturing. *CIRP Ann*. 2016;65(2):621–41. doi:10.1016/j.cirp.2016.06.005.
52. Oks SJ, Jalowski M, Lechner M, Mirschberger S, Merklein M, Vogel-Heuser B, et al. Cyber-physical systems in the context of industry 4.0: a review, categorization and outlook. *Inf Syst Front*. 2024;26(5):1731–72. doi:10.1007/s10796-022-10252-x.
53. Geist AR, Trimpe S. Structured learning of rigid-body dynamics: a survey and unified view from a robotics perspective. *GAMM-Mitt*. 2021;44(2):e202100009. doi:10.1002/gamm.202100009.
54. Silvestri L. CFD modeling in Industry 4.0: new perspectives for smart factories. *Procedia Comput Sci*. 2021;180:381–7. doi:10.1016/j.procs.2021.01.359.
55. Wang H, Cao Y, Huang Z, Liu Y, Hu P, Luo X, et al. Recent advances on machine learning for computational fluid dynamics: a survey. *arXiv:2408.12171*. 2024.
56. Wu D, Zheng A, Yu W, Cao H, Ling Q, Liu J, et al. Digital twin technology in transportation infrastructure: a comprehensive survey of current applications, challenges, and future directions. *Appl Sci*. 2025;15(4):1911. doi:10.3390/app15041911.
57. Spiryagin M, Edelmann J, Klinger F, Cole C. Vehicle system dynamics in digital twin studies in rail and road domains. *Veh Syst Dyn*. 2023;61(7):1737–86. doi:10.1080/00423114.2023.2188228.
58. Yu S, Hirche M, Huang Y, Chen H, Allgöwer F. Model predictive control for autonomous ground vehicles: a review. *Auton Intell Syst*. 2021;1(1):4. doi:10.1007/s43684-021-00005-z.
59. Liu W, Hua M, Deng Z, Meng Z, Huang Y, Hu C, et al. A systematic survey of control techniques and applications in connected and automated vehicles. *arXiv:2303.05665*. 2023.
60. Norouzi A, Heidarifar H, Borhan H, Shahbakhti M, Koch CR. Integrating machine learning and model predictive control for automotive applications: a review and future directions. *Eng Appl Artif Intell*. 2023;120(2):105878. doi:10.1016/j.engappai.2023.105878.
61. Gu X, Duan W, Zhang G, Hou J, Peng L, Wen M, et al. Digital twin technology for intelligent vehicles and transportation systems: a survey on applications, challenges and future directions. *IEEE Commun Surv Tutor*. 2026;28:3235–71. doi:10.1109/COMST.2025.3581152.
62. Arshad MW, Lodi S, Liu DQ. Multi-objective optimization of independent automotive suspension by AI and quantum approaches: a systematic review. *Machines*. 2025;13(3):204. doi:10.3390/machines13030204.
63. Llopis-Albert C, Rubio F, Devecé C, Zeng S. Multiobjective optimization framework for designing a steering system considering structural features and full vehicle dynamics. *Sci Rep*. 2023;13(1):19537. doi:10.1038/s41598-023-45349-z.
64. Ye F, Zhang S, Wang P, Chan CY. A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles. *arXiv:2105.14218*. 2021.
65. Hu X, Li S, Huang T, Tang B, Huai R, Chen L. How simulation helps autonomous driving: a survey of Sim2real, digital twins, and parallel intelligence. *arXiv:2305.01263*. 2023.
66. Li Z, Jin G, Yu R, Chen Z, Li N, Han W, et al. A survey of reinforcement learning-based motion planning for autonomous driving: lessons learned from a driving task perspective. *arXiv:2503.23650*. 2025.
67. Sarker MAL, Sarker MOF, Han DS. A survey on digital twin-assisted intelligent vehicle localization. *ICT Express*. 2025;11(5):981–92. doi:10.1016/j.icte.2025.08.008.
68. Yu M, Evangelou SA, Dini D. Advances in active suspension systems for road vehicles. *Engineering*. 2024;33(2):160–77. doi:10.1016/j.eng.2023.06.014.
69. Mei Y, Wang R, Ding R, Jiang Y. Classification evolution, control strategy innovation, and future challenges of vehicle suspension systems: a review. *Actuators*. 2025;14(10):485. doi:10.3390/act14100485.
70. Zhang N, Yang S, Wu G, Ding H, Zhang Z, Guo K. Fast distributed model predictive control method for active suspension systems. *Sensors*. 2023;23(6):3357. doi:10.3390/s23063357.
71. Li Q, Chen Z, Song H, Dong Y. Model predictive control for speed-dependent active suspension system with road preview information. *Sensors*. 2024;24(7):2255. doi:10.3390/s24072255.

72. Kimball JB, DeBoer B, Bubbar K. Adaptive control and reinforcement learning for vehicle suspension control: a review. *Annu Rev Control*. 2024;58(3):100974. doi:10.1016/j.arcontrol.2024.100974.
73. Jiang Z, Xue H, Yue H, Bao X, Zhu J, Wang X, et al. A review of artificial intelligence-driven active vibration and noise control. *Machines*. 2025;13(10):946. doi:10.3390/machines13100946.
74. Vergara D, del Bosque A, Fernández-Arias P. A decade of digital twins in materials science and engineering. *Comput Mater Contin*. 2025;85(1):41–64. doi:10.32604/cmc.2025.067881.
75. Tudorache L, Babur Ö, Lucas SS, van den Brand M. Current approaches to digital twins in additive manufacturing: a systematic literature review. *Prog Addit Manuf*. 2025;10(12):10819–53. doi:10.1007/s40964-025-01262-7.
76. Shi X, Zhou L, Huang Y, Wu Y, Hong Z. A review on the applications of graph neural networks in materials science at the atomic scale. *Mater Genome Eng Adv*. 2024;2(2):e50. doi:10.1002/mgea.50.
77. Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C, et al. Graph neural networks for materials science and chemistry. *Commun Mater*. 2022;3(1):93. doi:10.1038/s43246-022-00315-6.
78. Zivic F, Malisic AK, Grujovic N, Stojanovic B, Ivanovic M. Materials informatics: a review of AI and machine learning tools, platforms, data repositories, and applications to architected porous materials. *Mater Today Commun*. 2025;48:113525. doi:10.1016/j.mtcomm.2025.113525.
79. Antony Jose S, Tonner A, Feliciano M, Roy T, Shackelford A, Menezes PL. Smart manufacturing for high-performance materials: advances, challenges, and future directions. *Materials*. 2025;18(10):2255. doi:10.3390/ma18102255.
80. Zhou Z, Lang F, Farlim V, Zhang Z, Li S, Dong R. Review on multi-scale mechanics fundamentals and numerical methods for electronics packaging interconnect materials. *Front Mater*. 2024;11:1532859. doi:10.3389/fmats.2024.1532859.
81. Tsang CI, Pu H, Chen J. Multiscale simulation and machine learning facilitated design of two-dimensional nanomaterials-based tunnel field-effect transistors: a review. *arXiv:2409.18965*. 2024.
82. Groen D, Zasada SJ, Coveney PV. Survey of multiscale and multiphysics applications and communities. *arXiv:1208.6444*. 2012.
83. Nguyen BD, Potapenko P, Demirci A, Govind K, Bompas S, Sandfeld S. Efficient surrogate models for materials science simulations: machine learning-based prediction of microstructure properties. *Mach Learn Appl*. 2024;16(4):100544. doi:10.1016/j.mlwa.2024.100544.
84. Scotti L, Basoalto H, Moffat J, Cogswell D. Review of material modeling and digitalization in industry: barriers and perspectives. *Integr Mater Manuf Innov*. 2023;12(4):397–420. doi:10.1007/s40192-023-00318-2.
85. Bárkányi Á, Chován T, Németh S, Abonyi J. Modelling for digital twins—potential role of surrogate models. *Processes*. 2021;9(3):476. doi:10.3390/pr9030476.
86. Dequeant K, Viallette P, Lemaire P, Espinouse ML. A literature review on variability in semiconductor manufacturing: the next forward leap to Industry 4.0. In: *Proceedings of the 2016 Winter Simulation Conference; 2016 Dec 11–14; Arlington, VA, USA*. p. 2598–609. doi:10.5555/3042094.3042419.
87. Chen HY, Chen C. Review of applications of regression and predictive modeling in wafer manufacturing. *Electronics*. 2025;14(20):4083. doi:10.3390/electronics14204083.
88. Deenen PC. Data-driven modeling and optimization of semiconductor manufacturing [dissertation]. Eindhoven, The Netherlands: Eindhoven University of Technology; 2023. 150 p.
89. Nematov D, Hojamberdiev M. Machine learning-driven materials discovery: unlocking next-generation functional materials—a review. *arXiv:2503.18975*. 2025.
90. Weingram A, Cui C, Lin S, Munoz S, Jacob T, Viers J, et al. A definition and taxonomy of digital twins: case studies with machine learning and scientific applications. *Front High Perform Comput*. 2025;3:1536501. doi:10.3389/fhpcp.2025.1536501.
91. Farahani MA, McCormick MR, Gianinny R, Hudacheck F, Harik R, Liu Z, et al. Time-series pattern recognition in smart manufacturing systems: a literature review and ontology. *J Manuf Syst*. 2023;69(3):208–41. doi:10.1016/j.jmsy.2023.05.025.

92. Rahman MA, Shahrir MF, Iqbal K, Abushaiba AA. Enabling intelligent industrial automation: a review of machine learning applications with digital twin and edge AI integration. *Automation*. 2025;6(3):37. doi:10.3390/automation6030037.
93. Silvano C, Ielmini D, Ferrandi F, Fiorin L, Curzel S, Benini L, et al. A survey on deep learning hardware accelerators for heterogeneous HPC platforms. *arXiv:2306.15552*. 2023.
94. Henzel R, Herzwurm G. Cloud Manufacturing: a state-of-the-art survey of current issues. *Procedia CIRP*. 2018;72:947–52. doi:10.1016/j.procir.2018.03.055.
95. Farooq MS, Abdullah M, Riaz S, Alvi A, Rustam F, Flores MAL, et al. A survey on the role of industrial IoT in manufacturing for implementation of smart industry. *Sensors*. 2023;23(21):8958. doi:10.3390/s23218958.
96. Wang J, Li Y, Gao RX, Zhang F. Hybrid physics-based and data-driven models for smart manufacturing: modelling, simulation, and explainability. *J Manuf Syst*. 2022;63(6):381–91. doi:10.1016/j.jmsy.2022.04.004.
97. Schweidtmann AM, Zhang D, von Stosch M. A review and perspective on hybrid modeling methodologies. *Digit Chem Eng*. 2024;10(3):100136. doi:10.1016/j.dche.2023.100136.
98. Kasilingam S, Yang R, Singh SK, Farahani MA, Rai R, Wuest T. Physics-based and data-driven hybrid modeling in manufacturing: a review. *Prod Manuf Res*. 2024;12(1):2305358. doi:10.1080/21693277.2024.2305358.
99. Yang S, Navarathna P, Ghosh S, Bequette BW. Hybrid modeling in the era of smart manufacturing. *Comput Chem Eng*. 2020;140:106874. doi:10.1016/j.compchemeng.2020.106874.
100. Argungu JI, Idina MM, Chalawa UA, Ummar M, Bello SB, Arzika I, et al. A survey of edge computing approaches in smart factory. *Int J Adv Res Comput Commun Eng*. 2023;12(9):13–8. doi:10.17148/IJARCCCE.2023.12903.
101. Subramanian S. IoT and edge computing for smart manufacturing: architecture and future trends. *Int Jour Eng Com Sci*. 2024;13(10):26504–22. doi:10.18535/ijecs/v13i10.4922.
102. Bayar A, Şener U, Kayabay K, Eren PE. Edge computing applications in industrial IoT: a literature review. In: *Economics of grids, clouds, systems, and services*. Cham, Switzerland: Springer; 2023. p. 124–31. doi:10.1007/978-3-031-29315-3_11.
103. Wang X, Han Y, Leung VC, Niyato D, Yan X, Chen X. Convergence of edge computing and deep learning: a comprehensive survey. *IEEE Commun Surv Tutor*. 2020;22(2):869–904. doi:10.1109/comst.2020.2970550.
104. Han HC, Huang DC. Graph-theoretic detection of anomalies in supply chains: a PoR-based approach using Laplacian flow and sheaf theory. *Mathematics*. 2025;13(11):1795. doi:10.3390/math13111795.
105. Han HC, Huang DC. The blockchain neural network superior to deep learning for improving the trust of supply chain. *Comput Model Eng Sci*. 2025;143(3):3921–41. doi:10.32604/cmes.2025.065627.
106. Wang J, Pan C, Hu H, Li L, Zhou Y. A cyclic scheduling approach to single-arm cluster tools with multiple wafer types and residency time constraints. *IEEE Trans Autom Sci Eng*. 2019;16(3):1373–86. doi:10.1109/TASE.2018.2878063.
107. Taiwan Semiconductor Manufacturing Company. An innovation pioneer. TSMC 2024 sustainability report [Internet]. 2024 [cited 2025 Nov 25]. Available from: https://esg.tsmc.com/file/public/e-AnInnovationPioneer_1.pdf.
108. Jin J, Shi J. Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. In: *SMC'98 Conference Proceedings, 1998 IEEE International Conference on Systems, Man, and Cybernetics*; 1998 Oct 14; San Diego, CA, USA. p. 4716–21. doi:10.1109/ICSMC.1998.727597.
109. Huang CJ, Wu CF, Wang CC. Image processing techniques for wafer defect cluster identification. *IEEE Des Test Comput*. 2002;19(2):44–8. doi:10.1109/54.990441.
110. Moyne J, Schulze B, Iskandar J, Armacost M. Next generation advanced process control: leveraging big data and prediction. In: *Proceedings of the 2016 27th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*; 2016 May 16–19; Saratoga Springs, NY, USA. p. 191–6. doi:10.1109/ASMC.2016.7491123.
111. Braha D, Shmilovici A. On the use of decision tree induction for discovery of interactions in a photolithographic process. *IEEE Trans Semicond Manuf*. 2003;16(4):644–52. doi:10.1109/TSM.2003.818959.
112. Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F. Digital twin-driven product design, manufacturing and service with big data. *Int J Adv Manuf Technol*. 2018;94(9):3563–76. doi:10.1007/s00170-017-0233-1.

113. Gao Y, Chang D, Chen CH. A digital twin-based approach for optimizing operation energy consumption at automated container terminals. *J Clean Prod.* 2023;385:135782. doi:10.1016/j.jclepro.2022.135782.
114. Stricker N, Kuhnle A, Sturm R, Friess S. Reinforcement learning for adaptive order dispatching in the semiconductor industry. *CIRP Ann.* 2018;67(1):511–4. doi:10.1016/j.cirp.2018.04.041.
115. Liu C, Le Roux L, Körner C, Tabaste O, Lacan F, Bigot S. Digital twin-enabled collaborative data management for metal additive manufacturing systems. *J Manuf Syst.* 2022;62:857–74. doi:10.1016/j.jmsy.2020.05.010.
116. Han HC, Huang DC, Chen CL. Applications of AI and blockchain in origin traceability and forensics: a review of ICs, pharmaceuticals, EVs, UAVs, and robotics. *Comput Model Eng Sci.* 2025;145(1):67–126. doi:10.32604/cmesci.2025.070944.
117. Besson J, Cailletaud G, Chaboche JL, Forest S. Non-linear mechanics of materials. Dordrecht, The Netherlands: Springer; 2010. doi:10.1007/978-90-481-3356-7.
118. Roters F, Eisenlohr P, Bieler TR, Raabe D. Crystal plasticity finite element methods. In: *Materials science and engineering*. New York, NY, USA: John Wiley & Sons; 2010. doi:10.1002/9783527631483.
119. Nguyen PCH, Vlassis NN, Bahmani B, Sun W, Udaykumar HS, Baek SS. Synthesizing controlled microstructures of porous media using generative adversarial networks and reinforcement learning. *Sci Rep.* 2022;12(1):9034. doi:10.1038/s41598-022-12845-7.
120. Yağci T, Cöcen U, Çulha O. Aluminum alloy development for wheel production by low pressure die casting with new generation computational materials engineering approaches. *Arch Foundry Eng.* 2021;21(4):35–46. doi:10.24425/afe.2021.138677.
121. Knapp GL, Mukherjee T, Zuback JS, Wei HL, Palmer TA, De A, et al. Building blocks for a digital twin of additive manufacturing. *Acta Mater.* 2017;135(3):390–9. doi:10.1016/j.actamat.2017.06.039.
122. Mukherjee T, DebRoy T. A digital twin for rapid qualification of 3D printed metallic components. *Appl Mater Today.* 2019;14:59–65. doi:10.1016/j.apmt.2018.11.003.
123. Gunasegaram DR, Murphy AB, Barnard A, DebRoy T, Matthews MJ, Ladani L, et al. Towards developing multiscale-multiphysics models and their surrogates for digital twins of metal additive manufacturing. *Addit Manuf.* 2021;46(9):102089. doi:10.1016/j.addma.2021.102089.
124. Nath P, Mahadevan S. Probabilistic digital twin for additive manufacturing process design and control. *J Mech Des.* 2022;144(9):091704. doi:10.1115/1.4054521.
125. Hu C, Zuo Y, Chen C, Ping Ong S, Luo J. Genetic algorithm-guided deep learning of grain boundary diagrams: addressing the challenge of five degrees of freedom. *Mater Today.* 2020;38:49–57. doi:10.1016/j.mattod.2020.03.004.
126. Liu R, Kumar A, Chen Z, Agrawal A, Sundararaghavan V, Choudhary A. A predictive machine learning approach for microstructure optimization and materials design. *Sci Rep.* 2015;5(1):11551. doi:10.1038/srep11551.
127. Dai M, Demirel MF, Liang Y, Hu JM. Graph neural networks for an accurate and interpretable prediction of the properties of polycrystalline materials. *npj Comput Mater.* 2021;7(1):103. doi:10.1038/s41524-021-00574-w.
128. Alarifi HA, Atis M, Özdoğan C, Hu A, Yavuz M, Zhou Y. Molecular dynamics simulation of sintering and surface premelting of silver nanoparticles. *Mater Trans.* 2013;54(6):884–9. doi:10.2320/matertrans.md201225.
129. Motalab M, Paul R, Saha S, Mojumder S, Ahmed T, Suhling JC. Atomistic analysis of the thermomechanical properties of Sn-Ag-Cu solder materials at the nanoscale with the MEAM potential. *J Mol Model.* 2019;25(3):59. doi:10.1007/s00894-019-3939-1.
130. Luo R, Hu D, Qian C, Liu X, Fan X, Zhang G, et al. Molecular dynamics simulations on mechanical behaviors of sintered nanocopper in power electronics packaging. *Microelectron Reliab.* 2024;152(5):115284. doi:10.1016/j.microrel.2023.115284.
131. Long X, Jia Q, Shen Z, Liu M, Guan C. Strain rate shift for constitutive behaviour of sintered silver nanoparticles under nanoindentation. *Mech Mater.* 2021;158(27):103881. doi:10.1016/j.mechmat.2021.103881.
132. Symeonidis V, Karniadakis GE, Caswell B. A seamless approach to multiscale complex fluid simulation. *Comput Sci Eng.* 2005;7(3):39–46. doi:10.1109/MCSE.2005.44.
133. Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett.* 2018;120(14):145301. doi:10.1103/PhysRevLett.120.145301.

134. Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater*. 2019;31(9):3564–72. doi:10.1021/acs.chemmater.9b01294.
135. Jørgensen PB, Jacobsen KW, Schmidt MN. Neural message passing with edge updates for predicting properties of molecules and materials. *arXiv:1806.03146*. 2018.
136. Choudhary K, DeCost B. Atomistic Line Graph Neural Network for improved materials property predictions. *npj Comput Mater*. 2021;7(1):185. doi:10.1038/s41524-021-00650-1.
137. Cecen A, Dai H, Yabansu YC, Kalidindi SR, Song L. Material structure-property linkages using three-dimensional convolutional neural networks. *Acta Mater*. 2018;146(17):76–84. doi:10.1016/j.actamat.2017.11.053.
138. Chen G, Zhang ZS, Mei YH, Li X, Yu DJ, Wang L, et al. Applying viscoplastic constitutive models to predict ratcheting behavior of sintered nanosilver lap-shear joint. *Mech Mater*. 2014;72:61–71. doi:10.1016/j.mechmat.2014.02.001.
139. Su Y, Fu G, Liu C, Liu C, Long X. Fatigue crack evolution and effect analysis of Ag sintering die-attachment in SiC power devices under power cycling based on phase-field simulation. *Microelectron Reliab*. 2021;126:114244. doi:10.1016/j.microrel.2021.114244.
140. Su Y, Fu G, Liu C, Zhang K, Zhao L, Liu C, et al. Thermo-elasto-plastic phase-field modelling of mechanical behaviours of sintered nano-silver with randomly distributed micro-pores. *Comput Meth Appl Mech Eng*. 2021;378:113729. doi:10.1016/j.cma.2021.113729.
141. Mathew A, Dudek R, Otto A, Scherf C, Rzepka S, Subbiah N, et al. Lifetime modelling of sintered silver interconnected power devices by FEM and experiment. In: *Proceedings of the 2021 22nd International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*; 2021 Apr 19–21; St. Julian, Malta. doi:10.1109/EuroSimE52062.2021.9410877.
142. Ford E, Maneparambil K, Rajan S, Neithalath N. Machine learning-based accelerated property prediction of two-phase materials using microstructural descriptors and finite element analysis. *Comput Mater Sci*. 2021;191(7):110328. doi:10.1016/j.commatsci.2021.110328.
143. Jung J, Yoon JI, Park HK, Kim JY, Kim HS. An efficient machine learning approach to establish structure-property linkages. *Comput Mater Sci*. 2019;156:17–25. doi:10.1016/j.commatsci.2018.09.034.
144. Khorrami MS, Mianroodi JR, Siboni NH, Goyal P, Svendsen B, Benner P, et al. An artificial neural network for surrogate modeling of stress fields in viscoplastic polycrystalline materials. *npj Comput Mater*. 2023;9(1):37. doi:10.1038/s41524-023-00991-z.
145. Nakka R, Harursampath D, Ponnusami SA. A generalised deep learning-based surrogate model for homogenisation utilising material property encoding and physics-based bounds. *Sci Rep*. 2023;13(1):9079. doi:10.1038/s41598-023-34823-3.
146. Messner MC. Convolutional neural network surrogate models for the mechanical properties of periodic structures. *J Mech Des*. 2020;142(2):024503. doi:10.1115/1.4045040.
147. Eghbalian M, Pouragha M, Wan R. A physics-informed deep neural network for surrogate modeling in classical elasto-plasticity. *Comput Geotech*. 2023;159:105472. doi:10.1016/j.compgeo.2023.105472.
148. Raissi M. Deep hidden physics models: deep learning of nonlinear partial differential equations. *J Mach Learn Res*. 2018;19(1):932–55. doi:10.1016/j.jcp.2017.11.039.
149. Chemali CE, Freudenberg J, Hankinson M, Bendik JJ. Run-to-run critical dimension and sidewall angle lithography control using the PROLITH simulator. *IEEE Trans Semicond Manuf*. 2004;17(3):388–401. doi:10.1109/TSM.2004.831530.
150. Zhang Q, Poolla K, Spanos CJ. Across wafer critical dimension uniformity enhancement through lithography and etch process sequence: concept, approach, modeling, and experiment. *IEEE Trans Semicond Manuf*. 2007;20(4):488–505. doi:10.1109/TSM.2007.907627.
151. Xue D, Balachandran PV, Hogden J, Theiler J, Xue D, Lookman T. Accelerated search for materials with targeted properties by adaptive design. *Nat Commun*. 2016;7(1):11241. doi:10.1038/ncomms11241.
152. Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys Rev Lett*. 2007;98(14):146401. doi:10.1103/PhysRevLett.98.146401.

153. Yu T, Li Z, Wu D. Predictive modeling of material removal rate in chemical mechanical planarization with physics-informed machine learning. *Wear*. 2019;426(2):1430–8. doi:10.1016/j.wear.2019.02.012.
154. Wang L, Adamson G, Holm M, Moore P. A review of function blocks for process planning and control of manufacturing equipment. *J Manuf Syst*. 2012;31(3):269–79. doi:10.1016/j.jmsy.2012.02.004.
155. Frazzon EM, Agostino IRS, Broda E, Freitag M. Manufacturing networks in the era of digital production and operations: a socio-cyber-physical perspective. *Annu Rev Control*. 2020;49(11):288–94. doi:10.1016/j.arcontrol.2020.04.008.
156. Zheng M, Ming X. Construction of cyber-physical system-integrated smart manufacturing workshops: a case study in automobile industry. *Adv Mech Eng*. 2017;9(10):1687814017733246. doi:10.1177/1687814017733246.
157. Zhang X, Ming X, Liu Z, Qu Y, Yin D. An overall framework and subsystems for smart manufacturing integrated system (SMIS) from multi-layers based on multi-perspectives. *Int J Adv Manuf Technol*. 2019;103(1):703–22. doi:10.1007/s00170-019-03593-6.
158. Khaokom A, Thongsri J. Feasibility study for installing machine in production line to avoid particle contamination based on CFD simulation. In: *Proceedings of the 5th Asia Conference on Mechanical and Materials Engineering (ACMME 2017)*; 2017 Jun 9–11; Tokyo, Japan.
159. Zhang J, Long Z, Liu W, Chen Q. Strategy for studying ventilation performance in factories. *Aerosol Air Qual Res*. 2016;16(2):442–52. doi:10.4209/aaqr.2014.09.0210.
160. Cale TS, Merchant TP, Borucki LJ, Labun AH. Topography simulation for the virtual wafer fab. *Thin Solid Films*. 2000;365(2):152–75. doi:10.1016/S0040-6090(00)00756-2.
161. Nguyen PH, Bär E, Lorenz J, Ryssel H. Modeling of chemical-mechanical polishing on patterned wafers as part of integrated topography process simulation. *Microelectron Eng*. 2004;76(1–4):89–94. doi:10.1016/j.mee.2004.07.018.
162. Guo J, Geng M, Ren K, Ni D, Gao D. Optimizing plasma etching: integrating precise three-dimensional etching simulation and machine learning for multi-objective optimization. *IEEE Access*. 2024;12:127065–73. doi:10.1109/ACCESS.2024.3444454.
163. Wu X, Chen J, Xie L, Chan LLT, Chen CI. Development of convolutional neural network based Gaussian process regression to construct a novel probabilistic virtual metrology in multi-stage semiconductor processes. *Control Eng Pract*. 2020;96:104262. doi:10.1016/j.conengprac.2019.104262.
164. Tsuda T, Inoue S, Kayahara A, Imai SI, Tanaka T, Sato N, et al. Advanced semiconductor manufacturing using big data. *IEEE Trans Semicond Manuf*. 2015;28(3):229–35. doi:10.1109/TSM.2015.2445320.
165. Yu J, Guo P. Run-to-run control of chemical mechanical polishing process based on deep reinforcement learning. *IEEE Trans Semicond Manuf*. 2020;33(3):454–65. doi:10.1109/TSM.2020.3002896.
166. Jeong MK, Lu JC, Huo X, Vidakovic B, Di C. Wavelet-based data reduction techniques for process fault detection. *Technometrics*. 2006;48(1):26–40. doi:10.1198/004017005000000553.
167. Kim E, Cho S, Lee B, Cho M. Fault detection and diagnosis using self-attentive convolutional neural networks for variable-length sensor data in semiconductor manufacturing. *IEEE Trans Semicond Manuf*. 2019;32(3):302–9. doi:10.1109/TSM.2019.2917521.
168. Hsu CY, Chien CF, Chen PN. Manufacturing intelligence for early warning of key equipment excursion for advanced equipment control in semiconductor manufacturing. *J Chin Inst Ind Eng*. 2012;29(5):303–13. doi:10.1080/10170669.2012.702135.
169. Jin CH, Na HJ, Piao M, Pok G, Ryu KH. A novel DBSCAN-based defect pattern detection and classification framework for wafer Bin map. *IEEE Trans Semicond Manuf*. 2019;32(3):286–92. doi:10.1109/TSM.2019.2916835.
170. Wang CH. Recognition of semiconductor defect patterns using spatial filtering and spectral clustering. *Expert Syst Appl*. 2008;34(3):1914–23. doi:10.1016/j.eswa.2007.02.014.
171. Chien CF, Chang KH, Wang WC. An empirical study of design-of-experiment data mining for yield-loss diagnosis for semiconductor manufacturing. *J Intell Manuf*. 2014;25(5):961–72. doi:10.1007/s10845-013-0791-5.
172. Choi J, Jeong MK. Deep autoencoder with clipping fusion regularization on multistep process signals for virtual metrology. *IEEE Sens Lett*. 2019;3(1):7101804. doi:10.1109/LENS.2018.2884735.
173. Hsu SC, Chien CF. Hybrid data mining approach for pattern extraction from wafer Bin map to improve yield in semiconductor manufacturing. *Int J Prod Econ*. 2007;107(1):88–103. doi:10.1016/j.ijpe.2006.05.015.

174. Xu Y, Sun Y, Liu X, Zheng Y. A digital-twin-assisted fault diagnosis using deep transfer learning. *IEEE Access*. 2019;7:19990–9. doi:10.1109/ACCESS.2018.2890566.
175. Susto GA, Beghi A. Dealing with time-series data in predictive maintenance problems. In: *Proceedings of the 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*; 2016 Sep 6–9; Berlin, Germany. doi:10.1109/ETFA.2016.7733659.
176. Chowdary MK, Selvi M, Kandavalli SR, Ramesh JVN. Digital thread weaves reality in a closed-loop AI system for zero-defect hybrid production. *Int J Adv Manuf Technol*. 2024. doi:10.1007/s00170-024-14738-7.
177. Karkaria V, Goeckner A, Zha R, Chen J, Zhang J, Zhu Q, et al. Towards a digital twin framework in additive manufacturing: machine learning and Bayesian optimization for time series process optimization. *J Manuf Syst*. 2024;75(6):322–32. doi:10.1016/j.jmsy.2024.04.023.
178. Tao F, Qi Q. Make more digital twins. *Nature*. 2019;573(7775):490–1. doi:10.1038/d41586-019-02849-1.
179. Raval JS, Koch E, Donnenberg AD. Real-time monitoring of non-viable airborne particles correlates with airborne colonies and represents an acceptable surrogate for daily assessment of cell-processing cleanroom performance. *Cytotherapy*. 2012;14(9):1144–50. doi:10.3109/14653249.2012.698728.
180. Strauss L, Larkin J, Zhang KM. The use of occupancy as a surrogate for particle concentrations in recirculating, zoned cleanrooms. *Energy Build*. 2011;43(11):3258–62. doi:10.1016/j.enbuild.2011.08.027.
181. Zhang F, Shiue A, Fan Y, Liu J, Meng H, Zhang J, et al. Dynamic emission rates of human activity in biological cleanrooms. *Build Environ*. 2022;226(3):109777. doi:10.1016/j.buildenv.2022.109777.
182. Walid EA, Shady EA. Advanced data analysis as an enabler to near real-time contamination control strategy evaluation. *GMP J*. 2023;37:14–23.
183. Rahman MW, Vogl GW, Jia X, Qu Y. Physics-informed multi-task learning for material removal rate prediction in semiconductor chemical mechanical planarization. In: *Proceedings of the 2024 IEEE International Conference on Prognostics and Health Management (ICPHM)*; 2024 Jun 17–19; Spokane, WA, USA. p. 385–92. doi:10.1109/ICPHM61352.2024.10627679.
184. Dailey R, Bertelson S, Kim J, Djurdjanovic D. Virtual metrology of critical dimensions in plasma etch processes using entire optical emission spectrum. *IEEE Trans Semicond Manuf*. 2024;37(3):363–72. doi:10.1109/TSM.2024.3416844.
185. Bao L, Wang K, Jin R. A hierarchical model for characterising spatial wafer variations. *Int J Prod Res*. 2014;52(6):1827–42. doi:10.1080/00207543.2013.849389.
186. Yuan T, Ramadan SZ, Bae SJ. Yield prediction for integrated circuits manufacturing through hierarchical Bayesian modeling of spatial defects. *IEEE Trans Reliab*. 2011;60(4):729–41. doi:10.1109/TR.2011.2161698.
187. Krueger DC, Montgomery DC. Modeling and analyzing semiconductor yield with generalized linear mixed models. *Appl Stoch Models Bus Ind*. 2014;30(6):691–707. doi:10.1002/asmb.2074.
188. Hwang J, do Noh S. Digital twin-based optimization of operational parameters for cluster tools in semiconductor manufacturing. *IEEE Access*. 2024;12(4):122078–100. doi:10.1109/ACCESS.2024.3450869.
189. Zhang Z, Gong Y, Guan Z. Clearing function-based simulation optimization for release planning under digital twin wafer fabs. *IFAC-PapersOnLine*. 2022;55(10):2539–44. doi:10.1016/j.ifacol.2022.10.091.
190. Sivasubramanian CK, Dodge R, Ramani A, Bayba D, Janakiram M, Butcher E, et al. DTFab: a digital twin based approach for optimal reticle management in semiconductor photolithography. *J Syst Sci Syst Eng*. 2023;32(3):320–51. doi:10.1007/s11518-023-5564-x.
191. Kim Y, Lee G, Jeong J. ML-based JIT1 optimization for throughput maximization in cluster tool automation. *Appl Sci*. 2022;12(15):7519. doi:10.3390/app12157519.
192. Leachman RC. Closed-loop measurement of equipment efficiency and equipment capacity. *IEEE Trans Semicond Manuf*. 1997;10(1):84–97. doi:10.1109/66.554490.
193. Lee YH, Lee S. Deep reinforcement learning based scheduling within production plan in semiconductor fabrication. *Expert Syst Appl*. 2022;191(1):116222. doi:10.1016/j.eswa.2021.116222.
194. Wang J, Gao P, Zheng P, Zhang J, Ip WH. A fuzzy hierarchical reinforcement learning based scheduling method for semiconductor wafer manufacturing systems. *J Manuf Syst*. 2021;61(3):239–48. doi:10.1016/j.jmsy.2021.08.008.

195. Park IB, Park J. Scalable scheduling of semiconductor packaging facilities using deep reinforcement learning. *IEEE Trans Cybern.* 2023;53(6):3518–31. doi:10.1109/TCYB.2021.3128075.
196. Liu J, Qiao F, Zou M, Zinn J, Ma Y, Vogel-Heuser B. Dynamic scheduling for semiconductor manufacturing systems with uncertainties using convolutional neural networks and reinforcement learning. *Complex Intell Syst.* 2022;8(6):4641–62. doi:10.1007/s40747-022-00844-0.
197. Sakr AH, Aboelhassan A, Yacout S, Bassetto S. Simulation and deep reinforcement learning for adaptive dispatching in semiconductor manufacturing systems. *J Intell Manuf.* 2023;34(3):1311–24. doi:10.1007/s10845-021-01851-7.
198. Waschneck B, Reichstaller A, Belzner L, Altenmüller T, Bauernhansl T, Knapp A, et al. Deep reinforcement learning for semiconductor production scheduling. In: *Proceedings of the 2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*; 2018 Apr 30–May 3; Saratoga Springs, NY, USA. p. 301–6. doi:10.1109/ASMC.2018.8373191.
199. Lee C, Lee S. A practical deep reinforcement learning approach to semiconductor equipment scheduling. In: *Proceedings of the 2021 22nd IEEE International Conference on Industrial Technology (ICIT)*; 2021 Mar 10–12; Valencia, Spain. p. 979–85. doi:10.1109/ICIT46573.2021.9453533.
200. Wang L, Feng HY, Song C, Jin W. Function block design for adaptive execution control of job shop machining operations. *Int J Prod Res.* 2009;47(12):3413–34. doi:10.1080/00207540701666212.
201. Zhang Y, Qian C, Lv J, Liu Y. Agent and cyber-physical system based self-organizing and self-adaptive intelligent shopfloor. *IEEE Trans Ind Inform.* 2017;13(2):737–47. doi:10.1109/TII.2016.2618892.
202. Xia K, Sacco C, Kirkpatrick M, Saidy C, Nguyen L, Kircaliali A, et al. A digital twin to train deep reinforcement learning agent for smart manufacturing plants: environment, interfaces and intelligence. *J Manuf Syst.* 2021;58:210–30. doi:10.1016/j.jmsy.2020.06.012.
203. Krüger J, Wang L, Verl A, Bauernhansl T, Carpanzano E, Makris S, et al. Innovative control of assembly systems and lines. *CIRP Ann.* 2017;66(2):707–30. doi:10.1016/j.cirp.2017.05.010.
204. Kwon H, Kim SW, Joe HM. Learning-based force control of twisted string actuators using a neural network-based inverse model. *IEEE Robot Autom Lett.* 2024;9(9):8170–7. doi:10.1109/LRA.2024.3428909.
205. Gu X, Koren Y. Mass-individualisation—the twenty first century manufacturing paradigm. *Int J Prod Res.* 2022;60(24):7572–87. doi:10.1080/00207543.2021.2013565.
206. Monostori L, Váncza J, Kumara SRT. Agent-based systems for manufacturing. *CIRP Ann.* 2006;55(2):697–720. doi:10.1016/j.cirp.2006.10.004.
207. Wu S, Li Z, Chen W, Sun F. Dynamic modeling of robotic manipulator via an augmented deep Lagrangian network. *Tsinghua Sci Technol.* 2024;29(5):1604–14. doi:10.26599/TST.2024.9010011.
208. Zhu X, Liu Z, Cai C, Yang M, Zhang H, Fu L, et al. Deep learning-based predicting and compensating method for the pose deviations of parallel robots. *Comput Ind Eng.* 2024;191(26):110179. doi:10.1016/j.cie.2024.110179.
209. Mustafa AG, Eman M. A hybrid PSO-ACO algorithm for precise localization and geometric error reduction in industrial robots. *Instrumentation.* 2025;12(1):70–6.
210. Mittal S, Tolk A. Complexity challenges in cyber physical systems: using modeling and simulation (M&S) to support intelligence, adaptation and autonomy. New York, NY, USA: John Wiley & Sons; 2019. doi:10.1002/9781119552482.
211. Shen H, Pan YJ. Tracking synchronization improvement of networked manipulators using novel adaptive nonsingular terminal sliding mode control. *IEEE Trans Ind Electron.* 2021;68(5):4279–87. doi:10.1109/TIE.2020.2984984.
212. Kume Y, Hirata Y, Kosuge K. Object handling by coordinated multiple mobile manipulators without force/torque sensors. *J Robot Mechatron.* 2008;20(3):394–402. doi:10.20965/jrm.2008.p0394.
213. Spong MW, Hutchinson S, Vidyasagar M. Robot modeling and control. Hoboken, NJ, USA: John Wiley & Sons; 2006.
214. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature.* 2015;518(7540):529–33. doi:10.1038/nature14236.

215. Levine S, Finn C, Darrell T, Abbeel P. End-to-end training of deep visuomotor policies. *J Mach Learn Res.* 2016;17(1):1334–73. doi:10.5555/2946645.2946684.
216. Cheng CA, Huang HP, Hsu HK, Lai WZ, Cheng CC. Learning the inverse dynamics of robotic manipulators in structured reproducing kernel Hilbert space. *IEEE Trans Cybern.* 2016;46(7):1691–703. doi:10.1109/TCYB.2015.2454334.
217. Sutanto G, Wang AS, Lin Y, Mukadam M, Sukhatme GS, Rai A, et al. Encoding physical constraints in differentiable Newton-Euler algorithm. *arXiv:2001.08861.* 2020.
218. Li T, Liao Z, Cai Y. Physical modeling for digital twin of continuous damping control damper. *J Manuf Process.* 2023;99(3):96–104. doi:10.1016/j.jmapro.2023.05.019.
219. Bilal H, Obaidat MS, Shamrooz Aslam M, Zhang J, Yin B, Mahmood K. Online fault diagnosis of industrial robot using IoRT and hybrid deep learning techniques: an experimental approach. *IEEE Internet Things J.* 2024;11(19):31422–37. doi:10.1109/JIOT.2024.3418352.
220. He Y, Gao J, Chen Y. Deep learning-based pose prediction for visual servoing of robotic manipulators using image similarity. *Neurocomputing.* 2022;491(1):343–52. doi:10.1016/j.neucom.2022.03.045.
221. Liu J, Li Y. An image based visual servo approach with deep learning for robotic manipulation. *arXiv:1909.07727.* 2019.
222. Liu Y, Li S, Wang J, Zeng H, Lu J. A computer vision-based assistant system for the assembly of narrow cabin products. *Int J Adv Manuf Technol.* 2015;76(1):281–93. doi:10.1007/s00170-014-6274-9.
223. Pedersen OM, Misimi E, Chaumette F. Grasping unknown objects by coupling deep reinforcement learning, generative adversarial networks, and visual servoing. In: *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA); 2020 May 31–Aug 31; Paris, France.* p. 5655–62. doi:10.1109/ICRA40945.2020.9197196.
224. Jin Z, Wu J, Liu A, Zhang WA, Yu L. Policy-based deep reinforcement learning for visual servoing control of mobile robots with visibility constraints. *IEEE Trans Ind Electron.* 2022;69(2):1898–908. doi:10.1109/TIE.2021.3057005.
225. Bateux Q, Marchand E, Leitner J, Chaumette F, Corke P. Visual servoing from deep neural networks. *arXiv:1705.08940.* 2017.
226. Shi L, Copot C, Vanlanduit S. A Bayesian deep neural network for safe visual servoing in human-robot interaction. *Front Robot AI.* 2021;8:687031. doi:10.3389/frobt.2021.687031.
227. Slotine JJE, Li W. On the adaptive control of robot manipulators. *Int J Robot Res.* 1987;6(3):49–59. doi:10.1177/027836498700600303.
228. Levine S, Pastor P, Krizhevsky A, Quillen D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *arXiv:1603.02199.* 2016.
229. Pedersen MR, Nalpantidis L, Andersen RS, Schou C, Bøgh S, Krüger V, et al. Robot skills for manufacturing: from concept to industrial deployment. *Robot Comput Integr Manuf.* 2016;37:282–91. doi:10.1016/j.rcim.2015.04.002.
230. Ji S, Lee S, Yoo S, Suh I, Kwon I, Park FC, et al. Learning-based automation of robotic assembly for smart manufacturing. *Proc IEEE.* 2021;109(4):423–40. doi:10.1109/JPROC.2021.3063154.
231. Lee O, Kim D, Kim Y, Kim W, Lee S, Kim DY, et al. Development and validation of reconfigurable autonomous mobile manipulator for flexible manufacturing process. In: *Proceedings of the 2023 23rd International Conference on Control, Automation and Systems (ICCAS); 2023 Oct 17–20; Yeosu, Republic of Korea.* p. 914–8.
232. Andersen RE, Hansen EB, Cerny D, Madsen S, Pulendralingam B, Bøgh S, et al. Integration of a skill-based collaborative mobile robot in a smart cyber-physical environment. *Procedia Manuf.* 2017;11:114–23. doi:10.1016/j.promfg.2017.07.209.
233. Hogan N. Impedance control: an approach to manipulation. *J Dyn Syst Meas Control.* 1985;107(1):1–24. doi:10.1115/1.3140701.
234. Raibert MH, Craig JJ. Hybrid position/force control of manipulators. *J Dyn Syst Meas Control.* 1981;102(2):126–33. doi:10.1115/1.3139652.
235. Mason MT. Compliance and force control for computer controlled manipulators. *IEEE Trans Syst Man Cybern.* 1981;11(6):418–32. doi:10.1109/TSMC.1981.4308708.

236. Luo J, Solowjow E, Wen C, Ojea JA, Agogino AM, Tamar A, et al. Reinforcement learning on variable impedance controller for high-precision robotic assembly. In: Proceedings of the 2019 International Conference on Robotics and Automation (ICRA); 2019 May 20–24; Montreal, QC, Canada. p. 3080–7. doi:10.1109/ICRA.2019.8793506.
237. Xu Z, Zhou X, Wu H, Li X, Li S. Motion planning of manipulators for simultaneous obstacle avoidance and target tracking: an RNN approach with guaranteed performance. *IEEE Trans Ind Electron*. 2022;69(4):3887–97. doi:10.1109/TIE.2021.3073305.
238. Jiang L, Liu S, Cui Y, Jiang H. Path planning for robotic manipulator in complex multi-obstacle environment based on Improved_RRT. *IEEE/ASME Trans Mechatron*. 2022;27(6):4774–85. doi:10.1109/TMECH.2022.3165845.
239. Palmieri G, Scoccia C. Motion planning and control of redundant manipulators for dynamical obstacle avoidance. *Machines*. 2021;9(6):121. doi:10.3390/machines9060121.
240. Kang G, Kim YB, Lee YH, Oh HS, You WS, Choi HR. Sampling-based motion planning of manipulator with goal-oriented sampling. *Intell Serv Robot*. 2019;12(3):265–73. doi:10.1007/s11370-019-00281-y.
241. Lu J, Zou T, Jiang X. A neural network based approach to inverse kinematics problem for general six-axis robots. *Sensors*. 2022;22(22):8909. doi:10.3390/s22228909.
242. Wang S, Zhang Y, Chen S, Xu M, Yu Y, Liu P. Inverse kinematics analysis of 5-DOF cooperative robot based on long short-term memory network. In: Proceedings of the 2023 IEEE 3rd International Conference on Software Engineering and Artificial Intelligence (SEAI); 2023 Jun 16–18; Xiamen, China. p. 245–9. doi:10.1109/SEAI59139.2023.10217721.
243. Elkholy HA, Shahin AS, Shaarawy AW, Marzouk H, Elsamanty M. Solving inverse kinematics of a 7-DOF manipulator using convolutional neural network. In: Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020); 2020 Apr 8–10; Cairo, Egypt. p. 343–52. doi:10.1007/978-3-030-44289-7_32.
244. Liu X, Gan H, Luo Y, Chen Y, Gao L. Digital-twin-based real-time optimization for a fractional order controller for industrial robots. *Fractal Fract*. 2023;7(2):167. doi:10.3390/fractalfract7020167.
245. Peng G, Chen C, Yang C. Neural networks enhanced optimal admittance control of robot-environment interaction using reinforcement learning. *IEEE Trans Neural Netw Learn Syst*. 2022;33(9):4551–61. doi:10.1109/TNNLS.2021.3057958.
246. He W, Dong Y, Sun C. Adaptive neural impedance control of a robotic manipulator with input saturation. *IEEE Trans Syst Man Cybern Syst*. 2016;46(3):334–44. doi:10.1109/TSMC.2015.2429555.
247. Gao H, Ma C, Zhang X, Zhou C. Compliant variable admittance adaptive fixed-time sliding mode control for trajectory tracking of robotic manipulators. *Robotica*. 2024;42(6):1731–60. doi:10.1017/s0263574724000468.
248. Zhang Z, Wang Y, Zhang Z, Wang L, Huang H, Cao Q. A residual reinforcement learning method for robotic assembly using visual and force information. *J Manuf Syst*. 2024;72(54):245–62. doi:10.1016/j.jmsy.2023.11.008.
249. Zhang T, Liang X, Zou Y. Robot peg-in-hole assembly based on contact force estimation compensated by convolutional neural network. *Control Eng Pract*. 2022;120(2):105012. doi:10.1016/j.conengprac.2021.105012.
250. Gubbi S, Kolathaya S, Amrutur B. Imitation learning for high precision peg-in-hole tasks. In: Proceedings of the 2020 6th International Conference on Control, Automation and Robotics (ICCAR); 2020 Apr 20–23; Singapore. p. 368–72. doi:10.1109/ICCAR49639.2020.9108072.
251. Davchev T, Luck KS, Burke M, Meier F, Schaal S, Ramamoorthy S. Residual learning from demonstration: adapting DMPs for contact-rich manipulation. *IEEE Robot Autom Lett*. 2022;7(2):4488–95. doi:10.1109/LRA.2022.3150024.
252. Luo J, Sushkov O, Pevceviute R, Lian W, Su C, Vecerik M, et al. Robust multi-modal policies for industrial assembly via reinforcement learning and demonstrations: a large-scale study. *arXiv:2103.11512*. 2021.
253. Fan Y, Luo J, Tomizuka M. A learning framework for high precision industrial assembly. In: Proceedings of the 2019 International Conference on Robotics and Automation (ICRA); 2019 May 20–24; Montreal, QC, Canada. p. 811–7. doi:10.1109/ICRA.2019.8793659.
254. Beltran-Hernandez CC, Petit D, Ramirez-Alpizar IG, Harada K. Variable compliance control for robotic peg-in-hole assembly: a deep-reinforcement-learning approach. *Appl Sci*. 2020;10(19):6923. doi:10.3390/app10196923.

255. Ichiwara H, Ito H, Yamamoto K, Mori H, Ogata T. Contact-rich manipulation of a flexible object based on deep predictive learning using vision and tactility. In: Proceedings of the 2022 International Conference on Robotics and Automation (ICRA); 2022 May 23–27; Philadelphia, PA, USA. p. 5375–81. doi:10.1109/ICRA46639.2022.9811940.
256. Liu R, Nageotte F, Zanne P, de Mathelin M, Dresch-Langley B. Deep reinforcement learning for the control of robotic manipulation: a focussed mini-review. *Robotics*. 2021;10(1):22. doi:10.3390/robotics10010022.
257. Hamner B, Koterba S, Shi J, Simmons R, Singh S. An autonomous mobile manipulator for assembly tasks. *Auton Rob*. 2010;28(1):131–49. doi:10.1007/s10514-009-9142-y.
258. Guo J, Nguyen HT, Liu C, Cheah CC. Convolutional neural network-based robot control for an eye-in-hand camera. *IEEE Trans Syst Man Cybern Syst*. 2023;53(8):4764–75. doi:10.1109/TSMC.2023.3257416.
259. Gao J, He Y, Chen Y, Li Y. Learning end-to-end visual servoing using an improved soft actor-critic approach with centralized novelty measurement. *IEEE Trans Instrum Meas*. 2023;72:2514512. doi:10.1109/TIM.2023.3273687.
260. Shen H, Pan YJ, Ahmad U, He B. Pose synchronization of multiple networked manipulators using nonsingular terminal sliding mode control. *IEEE Trans Syst Man Cybern Syst*. 2021;51(12):7497–509. doi:10.1109/TSMC.2020.2971850.
261. Li Z, Tao PY, Ge SS, Adams M, Wijesoma WS. Robust adaptive control of cooperating mobile manipulators with relative motion. *IEEE Trans Syst Man Cybern Part B Cybern*. 2009;39(1):103–16. doi:10.1109/TSMCB.2008.2002853.
262. Yang H, Lee D. Cooperative grasping control of multiple mobile manipulators with obstacle avoidance. In: Proceedings of the 2013 IEEE International Conference on Robotics and Automation; 2013 May 6–10; Karlsruhe, Germany. p. 836–41. doi:10.1109/ICRA.2013.6630670.
263. Ahmad U, Pan YJ, Shen H. Robust control design for teleoperation of multiple mobile manipulators under time delays. *Int J Robust Nonlinear Control*. 2020;30(16):6454–72. doi:10.1002/rnc.5106.
264. Tarbouriech S, Navarro B, Fraisse P, Crosnier A, Cherubini A, Sallé D. An admittance based hierarchical control framework for dual-arm cobots. *Mechatronics*. 2022;86(4):102814. doi:10.1016/j.mechatronics.2022.102814.
265. Krüger J, Lien TK, Verl A. Cooperation of human and machines in assembly lines. *CIRP Ann*. 2009;58(2):628–46. doi:10.1016/j.cirp.2009.09.009.
266. Wang L, Gao R, Váncza J, Krüger J, Wang XV, Makris S, et al. Symbiotic human-robot collaborative assembly. *CIRP Ann*. 2019;68(2):701–26. doi:10.1016/j.cirp.2019.05.002.
267. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. *arXiv:1509.02971*. 2015.
268. Chaumette F, Hutchinson S. Visual servo control. I. Basic approaches. *IEEE Robot Autom Mag*. 2006;13(4):82–90. doi:10.1109/MRA.2006.250573.
269. Zeng A, Song S, Yu KT, Donlon E, Hogan FR, Bauza M, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. In: Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA); 2018 May 21–25; Brisbane, QLD, Australia. p. 3750–7. doi:10.1109/ICRA.2018.8461044.
270. Bhatti G, Mohan H, Raja Singh R. Towards the future of smart electric vehicles: digital twin technology. *Renew Sustain Energy Rev*. 2021;141(3):110801. doi:10.1016/j.rser.2021.110801.
271. Soeiro LGG, Filho BJC. Vehicle power system modeling and integration in hardware-in-the-loop (HIL) simulations. *Machines*. 2023;11(6):605. doi:10.3390/machines11060605.
272. Fragapane G, Ivanov D, Peron M, Sgarbossa F, Strandhagen JO. Increasing flexibility and productivity in Industry 4.0 production networks with autonomous mobile robots and smart intralogistics. *Ann Oper Res*. 2022;308(1):125–43. doi:10.1007/s10479-020-03526-7.
273. Kapteyn MG, Knezevic DJ, Huynh DBP, Tran M, Willcox KE. Data-driven physics-based digital twins via a library of component-based reduced-order models. *Int J Numer Meth Eng*. 2022;123(13):2986–3003. doi:10.1002/nme.6423.
274. James SS, Anderson SR, Da Lio M. Longitudinal vehicle dynamics: a comparison of physical and data-driven models under large-scale real-world driving conditions. *IEEE Access*. 2020;8:73714–29. doi:10.1109/ACCESS.2020.2988592.

275. Rajkumar RR, Lee I, Sha L, Stankovic J. Cyber-physical systems: the next computing revolution. In: Proceedings of the 47th Design Automation Conference (DAC'10); 2010 Jun 13–18; Anaheim, CA, USA. p. 731–6. doi:10.1145/1837274.1837461.
276. He S, Zhang Z, Han S, Pepin L, Wang G, Zhang D, et al. Data-driven distributionally robust electric vehicle balancing for autonomous mobility-on-demand systems under demand and supply uncertainties. *IEEE Trans Intell Transp Syst.* 2023;24(5):5199–215. doi:10.1109/TITS.2023.3237804.
277. Amini A, Gilitschenski I, Phillips J, Moseyko J, Banerjee R, Karaman S, et al. Learning robust control policies for end-to-end autonomous driving from data-driven simulation. *IEEE Robot Autom Lett.* 2020;5(2):1143–50. doi:10.1109/LRA.2020.2966414.
278. Liu J, Huang Z, Xu X, Zhang X, Sun S, Li D. Multi-kernel online reinforcement learning for path tracking control of intelligent vehicles. *IEEE Trans Syst Man Cybern Syst.* 2021;51(11):6962–75. doi:10.1109/TSMC.2020.2966631.
279. Everett M, Lütjens B, How JP. Certifiable robustness to adversarial state uncertainty in deep reinforcement learning. *IEEE Trans Neural Netw Learn Syst.* 2022;33(9):4184–98. doi:10.1109/TNNLS.2021.3056046.
280. Wang K, Li Z, Nonomura K, Yu T, Sakaguchi K, Hashash O, et al. Smart mobility digital twin based automated vehicle navigation system: a proof of concept. *IEEE Trans Intell Veh.* 2024;9(3):4348–61. doi:10.1109/TIV.2024.3368109.
281. Wang K, Yu T, Li Z, Sakaguchi K, Hashash O, Saad W. Digital twins for autonomous driving: a comprehensive implementation and demonstration. In: Proceedings of the 2024 International Conference on Information Networking (ICOIN); 2024 Jan 17–19; Ho Chi Minh City, Vietnam. p. 452–7. doi:10.1109/ICOIN59985.2024.10572108.
282. Wang Z, Zheng O, Li L, Abdel-Aty M, Cruz-Neira C, Islam Z. Towards next generation of pedestrian and connected vehicle in-the-loop research: a digital twin co-simulation framework. *IEEE Trans Intell Veh.* 2023;8(4):2674–83. doi:10.1109/TIV.2023.3250353.
283. Drehmer LRC, Gomes HM, Paucar Casas WJ. An interval-based multi-objective robust design optimization for vehicle dynamics. *Mech Based Des Struct Mach.* 2023;51(12):7076–101. doi:10.1080/15397734.2022.2088557.
284. Fossati GG, Miguel LFF, Casas WJP. Multi-objective optimization of the suspension system parameters of a full vehicle model. *Optim Eng.* 2019;20(1):151–77. doi:10.1007/s11081-018-9403-8.
285. Ning D, Du H, Zhang N, Sun S, Li W. Controllable electrically interconnected suspension system for improving vehicle vibration performance. *IEEE/ASME Trans Mechatron.* 2020;25(2):859–71. doi:10.1109/TMECH.2020.2965573.
286. Yu M, Cheng C, Evangelou SA, Dini D. Series active variable geometry suspension: full-car prototyping and road testing. *IEEE/ASME Trans Mechatron.* 2022;27(3):1332–44. doi:10.1109/TMECH.2021.3097153.
287. Morato MM, Nguyen MQ, Sename O, Dugard L. Design of a fast real-time LPV model predictive control system for semi-active suspension control of a full vehicle. *J Frankl Inst.* 2019;356(3):1196–224. doi:10.1016/j.jfranklin.2018.11.016.
288. Wang R, Jiang Y, Ding R, Liu W, Meng X, Sun Z. Design and experimental verification of self-powered electromagnetic vibration suppression and absorption system for in-wheel motor electric vehicles. *J Vibr Control.* 2022;28(19–20):2544–55. doi:10.1177/10775463211014419.
289. Chung Y, Yang YP. Hardware-in-the-loop simulation of self-driving electric vehicles by dynamic path planning and model predictive control. *Electronics.* 2021;10(19):2447. doi:10.3390/electronics10192447.
290. Miao F, He S, Pepin L, Han S, Hendawi A, Khalefa ME, et al. Data-driven distributionally robust optimization for vehicle balancing of mobility-on-demand systems. *ACM Trans Cyber Phys Syst.* 2021;5(2):1–27. doi:10.1145/3418287.
291. Hu C, Chen Y, Wang J. Fuzzy observer-based transitional path-tracking control for autonomous vehicles. *IEEE Trans Intell Transp Syst.* 2021;22(5):3078–88. doi:10.1109/TITS.2020.2979431.
292. Xu R, Guo Y, Han X, Xia X, Xiang H, Ma J. OpenCDA: an open cooperative driving automation framework integrated with co-simulation. In: Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC); 2021 Sep 19–22; Indianapolis, IN, USA. p. 1155–62. doi:10.1109/ITSC48978.2021.9564825.

293. Voogd KL, Allamaa JP, Alonso-Mora J, Son TD. Reinforcement learning from simulation to real world autonomous driving using digital twin. *IFAC-PapersOnLine*. 2023;56(2):1510–5. doi:10.1016/j.ifacol.2023.10.1846.
294. Edington L, Dervilis N, Ben Abdesslem A, Wagg D. A time-evolving digital twin tool for engineering dynamics applications. *Mech Syst Signal Process*. 2023;188(10):109971. doi:10.1016/j.ymssp.2022.109971.
295. Gao Y, Chang D, Chen CH, Sha M. A digital twin-based decision support approach for AGV scheduling. *Eng Appl Artif Intell*. 2024;130:107687. doi:10.1016/j.engappai.2023.107687.
296. Zhang L, Yang C, Yan Y, Cai Z, Hu Y. Automated guided vehicle dispatching and routing integration via digital twin with deep reinforcement learning. *J Manuf Syst*. 2024;72:492–503. doi:10.1016/j.jmsy.2023.12.008.
297. Zheng X, Liang C, Wang Y, Shi J, Lim G. Multi-AGV dynamic scheduling in an automated container terminal: a deep reinforcement learning approach. *Mathematics*. 2022;10(23):4575. doi:10.3390/math10234575.
298. Gong L, Huang Z, Xiang X, Liu X. Real-time AGV scheduling optimisation method with deep reinforcement learning for energy-efficiency in the container terminal yard. *Int J Prod Res*. 2024;62(21):7722–42. doi:10.1080/00207543.2024.2325583.
299. Sampedro C, Rodriguez-Ramos A, Gil I, Mejias L, Campoy P. Image-based visual servoing controller for multirotor aerial robots using deep reinforcement learning. In: *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; 2018 Oct 1–5; Madrid, Spain. p. 979–86. doi:10.1109/IROS.2018.8594249.
300. Kasri A, Ouari K, Belkhier Y, Oubelaid A, Bajaj M, Berhanu Tuka M. Real-time and hardware in the loop validation of electric vehicle performance: robust nonlinear predictive speed and currents control based on space vector modulation for PMSM. *Results Eng*. 2024;22:102223. doi:10.1016/j.rineng.2024.102223.
301. Fragapane G, de Koster R, Sgarbossa F, Strandhagen JO. Planning and control of autonomous mobile robots for intralogistics: literature review and research agenda. *Eur J Oper Res*. 2021;294(2):405–26. doi:10.1016/j.ejor.2021.01.019.
302. Zeng T, Semiari O, Saad W, Bennis M. Joint communication and control for wireless autonomous vehicular platoon systems. *IEEE Trans Commun*. 2019;67(11):7907–22. doi:10.1109/TCOMM.2019.2931583.
303. Norouzi A, Shahpour S, Gordon D, Winkler A, Nuss E, Abel D, et al. Machine learning integrated with model predictive control for imitative optimal control of compression ignition engines. *IFAC-PapersOnLine*. 2022;55(24):19–26. doi:10.1016/j.ifacol.2022.10.256.
304. Hang P, Xia X, Chen G, Chen X. Active safety control of automated electric vehicles at driving limits: a tube-based MPC approach. *IEEE Trans Transp Electr*. 2022;8(1):1338–49. doi:10.1109/TTE.2021.3100843.
305. Hua M, Chen G, Zhang B, Huang Y. A hierarchical energy efficiency optimization control strategy for distributed drive electric vehicles. *Proc Inst Mech Eng Part D J Automob Eng*. 2019;233(3):605–21. doi:10.1177/0954407017751788.
306. Wang Y, Shao Q, Zhou J, Zheng H, Chen H. Longitudinal and lateral control of autonomous vehicles in multi-vehicle driving environments. *IET Intell Transp Syst*. 2020;14(8):924–35. doi:10.1049/iet-its.2019.0846.
307. Tian Y, Yao Q, Hang P, Wang S. A gain-scheduled robust controller for autonomous vehicles path tracking based on LPV system with MPC and H_∞ . *IEEE Trans Veh Technol*. 2022;71(9):9350–62. doi:10.1109/tvt.2022.3176384.
308. Chen IM, Chan CY. Deep reinforcement learning based path tracking controller for autonomous vehicle. *Proc Inst Mech Eng Part D J Automob Eng*. 2021;235(2–3):541–51. doi:10.1177/0954407020954591.
309. Wang P, Li H, Chan CY. Continuous control for automated lane change behavior based on deep deterministic policy gradient algorithm. In: *2019 IEEE Intelligent Vehicles Symposium (IV)*; 2019 Jun 9–12; Paris, France. p. 1454–60. doi:10.1109/IVS.2019.8813903.
310. Ma X, Driggs-Campbell K, Kochenderfer MJ. Improved robustness and safety for autonomous vehicle control with adversarial reinforcement learning. In: *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*; 2018 Jun 26–30; Changshu, China. p. 1665–71. doi:10.1109/IVS.2018.8500450.
311. Min K, Kim H, Huh K. Deep distributional reinforcement learning based high-level driving policy determination. *IEEE Trans Intell Veh*. 2019;4(3):416–24. doi:10.1109/TIV.2019.2919467.
312. Balakrishnan A, Lee J, Gaurav A, Czarnecki K, Sedwards S. Transfer reinforcement learning for autonomous driving: from WiseMove to WiseSim. *ACM Trans Model Comput Simul*. 2021;31(3):1–26. doi:10.1145/3449356.

313. Ren J, Nalpantidis L, Andersen NA, Ravn O. Building digital twin of mobile robotics testbed using centralized localization system. In: Proceedings of the 2023 11th International Conference on Control, Mechatronics and Automation (ICCMA); 2023 Nov 1–3; Grimstad, Norway. p. 139–45. doi:10.1109/ICCMA59762.2023.10375050.
314. Huang C, Liu Y, Sun X, Wang Y. Intelligent active suspension control method based on hierarchical multi-sensor perception fusion. *Sensors*. 2025;25(15):4723. doi:10.3390/s25154723.
315. Fukuda H, Kondo T. Anomaly detection in random circuit patterns using autoencoder. *J Micro Nanopattern Mater Metrol*. 2021;20(4):044001. doi:10.1117/1.JMM.20.4.044001.
316. Liang F, Yu W, Liu X, Griffith D, Golmie N. Towards edge-based deep learning in industrial Internet of Things. *IEEE Internet Things J*. 2020;7(5):4329–41. doi:10.1109/jiot.2019.2963635.
317. Okuyelu O, Adaji O. AI-driven real-time quality monitoring and process optimization for enhanced manufacturing performance. *J Adv Math Comput Sci*. 2024;39(4):81–9. doi:10.9734/jamcs/2024/v39i41883.
318. Santos R, Piqueiro H, Dias R, Rocha CD. Transitioning trends into action: a simulation-based Digital Twin architecture for enhanced strategic and operational decision-making. *Comput Ind Eng*. 2024;198(1):110616. doi:10.1016/j.cie.2024.110616.
319. Kang J, Hwang KI. A comprehensive real-time indoor air-quality level indicator. *Sustainability*. 2016;8(9):881. doi:10.3390/su8090881.
320. Benammar M, Abdaoui A, Ahmad SHM, Touati F, Kadri A. A modular IoT platform for real-time indoor air quality monitoring. *Sensors*. 2018;18(2):581. doi:10.3390/s18020581.
321. Rahmadani AA, Syaifudin YW, Setiawan B, Panduman YYF, Funabiki N. Enhancing campus environment: real-time air quality monitoring through IoT and web technologies. *J Sens Actuator Netw*. 2025;14(1):2. doi:10.3390/jsan14010002.
322. Yildiz O, Sucuoglu HS. Development of real-time IoT-based air quality forecasting system using machine learning approach. *Sustainability*. 2025;17(19):8531. doi:10.3390/su17198531.
323. Barbieri L, Tedeschini BC, Brambilla M, Nicoli M. Deep learning-based cooperative LiDAR sensing for improved vehicle positioning. *IEEE Trans Signal Process*. 2024;72:1666–82. doi:10.1109/TSP.2024.3377375.
324. Merat R, Cioffi G, Bauersfeld L, Scaramuzza D. Drift-free visual SLAM using digital twins. *IEEE Robot Autom Lett*. 2025;10(2):1633–40. doi:10.1109/LRA.2024.3518840.
325. Shi K, Qian K, Yu H. Visual human localization and safety monitoring in a digital twin of workspace. In: Proceedings of the 2022 41st Chinese Control Conference (CCC); 2022 Jul 25–27; Hefei, China. p. 6117–21.
326. Luo W, Hu T, Ye Y, Zhang C, Wei Y. A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin. *Robot Comput Integr Manuf*. 2020;65(1):101974. doi:10.1016/j.rcim.2020.101974.
327. Gaikwad A, Yavari R, Montazeri M, Cole K, Bian L, Rao P. Toward the digital twin of additive manufacturing: integrating thermal simulations, sensing, and analytics to detect process faults. *IIEE Trans*. 2020;52(11):1204–17. doi:10.1080/24725854.2019.1701753.
328. Guo J, Yang Z, Chen C, Luo W, Hu W. Real-time prediction of remaining useful life and preventive maintenance strategy based on digital twin. *J Comput Inf Sci Eng*. 2021;21(3):031003. doi:10.1115/1.4049153.
329. Wu D, Liu S, Zhang L, Terpenney J, Gao RX, Kurfess T, et al. A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing. *J Manuf Syst*. 2017;43(4):25–34. doi:10.1016/j.jmsy.2017.02.011.
330. Gentner T, Breitenbach J, Neitzel T, Schulze J, Buettner R. A systematic literature review of machine learning applications for process monitoring and control in semiconductor manufacturing. In: Proceedings of the 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC); 2022 Jun 27–Jul 1; Los Alamitos, CA, USA. p. 1081–6. doi:10.1109/COMPSAC54236.2022.00169.
331. Abubeker KM, Prajitha C, Rinesh S, Arun M, Senthil Kumar AP. Internet of Things and LoRaWAN-assisted real-time indoor air quality monitoring and automation system for enhancing passenger safety. *Int J Low Carbon Tech*. 2025;20:1427–39. doi:10.1093/ijlct/ctaf083.
332. Caccavale F, Chiacchio P, Marino A, Villani L. Six-DOF impedance control of dual-arm cooperative manipulators. *IEEE/ASME Trans Mechatron*. 2008;13(5):576–86. doi:10.1109/TMECH.2008.2002816.
333. Tao F, Cheng Y, Xu LD, Zhang L, Li BH. CCIoT-CMfg: cloud computing and Internet of Things-based cloud manufacturing service system. *IEEE Trans Ind Inform*. 2014;10(2):1435–42. doi:10.1109/TII.2014.2306383.

334. Tao F, Zuo Y, Xu LD, Zhang L. IoT-based intelligent perception and access of manufacturing resource toward cloud manufacturing. *IEEE Trans Ind Inform.* 2014;10(2):1547–57. doi:10.1109/TII.2014.2306397.
335. Protner J, Pipan M, Zupan H, Resman M, Simic M, Herakovic N. Edge computing and digital twin based smart manufacturing. *IFAC-PapersOnLine.* 2021;54(1):831–6. doi:10.1016/j.ifacol.2021.08.098.
336. Wang W, Hu T, Gu J. Edge-cloud cooperation driven self-adaptive exception control method for the smart factory. *Adv Eng Inform.* 2022;51(11):101493. doi:10.1016/j.aei.2021.101493.
337. Moyne J, Samantaray J, Armacost M. Big data emergence in semiconductor manufacturing advanced process control. In: *Proceedings of the 2015 26th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*; 2015 May 3–6; Saratoga Springs, NY, USA. p. 130–5. doi:10.1109/ASMC.2015.7164483.
338. Shakya B, He T, Salmani H, Forte D, Bhunia S, Tehranipoor M. Benchmarking of hardware trojans and maliciously affected circuits. *J Hardw Syst Secur.* 2017;1(1):85–102. doi:10.1007/s41635-017-0001-6.
339. Pilato C, Garg S, Wu K, Karri R, Regazzoni F. Securing hardware accelerators: a new challenge for high-level synthesis. *IEEE Embed Syst Lett.* 2018;10(3):77–80. doi:10.1109/LES.2017.2774800.
340. Chen B, Wan J, Lan Y, Imran M, Li D, Guizani N. Improving cognitive ability of edge intelligent IIoT through machine learning. *IEEE Netw.* 2019;33(5):61–7. doi:10.1109/MNET.001.1800505.
341. Uhlemann THJ, Lehmann C, Steinhilper R. The digital twin: realizing the cyber-physical production system for industry 4.0. *Procedia CIRP.* 2017;61:335–40. doi:10.1016/j.procir.2016.11.152.
342. Lu Y, Liu C, Wang KI, Huang H, Xu X. Digital Twin-driven smart manufacturing: connotation, reference model, applications and research issues. *Robot Comput Integr Manuf.* 2020;61:101837. doi:10.1016/j.rcim.2019.101837.
343. Lu Y, Xu X. Resource virtualization: a core technology for developing cyber-physical production systems. *J Manuf Syst.* 2018;47:128–40. doi:10.1016/j.jmsy.2018.05.003.
344. Kochkov D, Smith JA, Alieva A, Wang Q, Brenner MP, Hoyer S. Machine learning-accelerated computational fluid dynamics. *Proc Natl Acad Sci U S A.* 2021;118(21):e2101784118. doi:10.1073/pnas.2101784118.
345. List B, Chen LW, Thuerey N. Learned turbulence modelling with differentiable fluid solvers: physics-based loss functions and optimisation horizons. *J Fluid Mech.* 2022;949:A25. doi:10.1017/jfm.2022.738.
346. Goodale T, Allen G, Lanfermann G, Massó J, Radke T, Seidel E, et al. The Cactus framework and toolkit: design and applications. In: *Proceedings of the 5th International Conference on High Performance Computing for Computational Science (VECPAR 2002)*; 2002 Jun 26–28; Porto, Portugal. p. 197–227. doi:10.1007/3-540-36569-9_13.
347. Hegewald J, Krafczyk M, Tölke J, Hoekstra A, Chopard B. An agent-based coupling platform for complex automata. In: *Proceedings of the 8th International Conference on Computational Science, Part II (ICCS 2008)*; 2008 Jun 23–25; Berlin, Germany. p. 227–33. doi:10.1007/978-3-540-69387-1_25.
348. Iraola E, García-Lorenzo M, Lordan-Gomis F, Rossi F, Prieto-Araujo E, Badia RM. HP2C-DT: high-precision high-performance computer-enabled digital twin. *Future Gener Comput Syst.* 2026;180(1):108333. doi:10.1016/j.future.2025.108333.
349. Liu N, Li X, Shen W. Multi-granularity resource virtualization and sharing strategies in cloud manufacturing. *J Netw Comput Appl.* 2014;46(4):72–82. doi:10.1016/j.jnca.2014.08.007.
350. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys.* 2019;378:686–707. doi:10.1016/j.jcp.2018.10.045.
351. Mao Z, Jagtap AD, Karniadakis GE. Physics-informed neural networks for high-speed flows. *Comput Meth Appl Mech Eng.* 2020;360(5):112789. doi:10.1016/j.cma.2019.112789.
352. Cai J, Yan C, Ma Y, Yu B, Zhou D, Zeng X. NeurFill: migrating full-chip CMP simulators to neural networks for model-based dummy filling synthesis. In: *Proceedings of the 2021 58th ACM/IEEE Design Automation Conference (DAC)*; 2021 Dec 5–9; San Francisco, CA, USA. p. 187–92. doi:10.1109/DAC18074.2021.9586325.
353. Souza V, Cruz R, Silva W, Lins S, Lucena V. A digital twin architecture based on the industrial Internet of Things technologies. In: *Proceedings of the 2019 IEEE International Conference on Consumer Electronics (ICCE)*; 2019 Jan 11–13; Las Vegas, NV, USA. doi:10.1109/ICCE.2019.8662081.

354. Kim N, Choi H, Chun J, Jeong J. Introduction of equipment level FDC system for semiconductor wet-cleaning equipment optimization and real-time fault detection. In: Proceedings of the 2022 33rd Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC); 2022 May 2–5; Saratoga Springs, NY, USA. p. 1–4. doi:10.1109/ASMC54647.2022.9792476.
355. Boppana VR. Industry 4.0: revolutionizing the future of manufacturing and automation. *Innov Comput Sci J*. 2024 [cited 2025 Nov 25]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5135027.
356. Glatt M, Sinnwell C, Yi L, Donohoe S, Ravani B, Aurich JC. Modeling and implementation of a digital twin of material flows based on physics simulation. *J Manuf Syst*. 2021;58(2):231–45. doi:10.1016/j.jmsy.2020.04.015.
357. Huang X, Yang H, Zhou C, He M, Shen X, Zhuang W. When digital twin meets generative AI: intelligent closed-loop network management. *IEEE Netw*. 2025;39(5):272–9. doi:10.1109/MNET.2024.3524474.
358. Lodhi SK, Gill AY, Hussain I. AI-powered innovations in contemporary manufacturing procedures: an extensive analysis. *Int J Multidiscip Sci Arts*. 2024;3(4):15–25. doi:10.47709/ijmdsa.v3i4.4616.
359. Teng SY, Touš M, Leong WD, How BS, Lam HL, Máša V. Recent advances on industrial data-driven energy savings: digital twins and infrastructures. *Renew Sustain Energy Rev*. 2021;135(3):110208. doi:10.1016/j.rser.2020.110208.
360. Ghazal M, Basmaji T, Yaghi M, Alkhedher M, Mahmoud M, El-Baz AS. Cloud-based monitoring of thermal anomalies in industrial environments using AI and the Internet of robotic things. *Sensors*. 2020;20(21):6348. doi:10.3390/s20216348.
361. Daw A, Karpatne A, Watkins W, Read J, Kumar V. Physics-guided neural networks (PGNN): an application in lake temperature modeling. *arXiv:1710.11431*. 2017.
362. Wang J, Li Y, Zhao R, Gao RX. Physics guided neural network for machining tool wear prediction. *J Manuf Syst*. 2020;57(1):298–310. doi:10.1016/j.jmsy.2020.09.005.
363. Sun L, Gao H, Pan S, Wang JX. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Comput Meth Appl Mech Eng*. 2020;361(4):112732. doi:10.1016/j.cma.2019.112732.
364. Moges T, Yang Z, Jones K, Feng S, Witherell P, Lu Y. Hybrid modeling approach for melt-pool prediction in laser powder bed fusion additive manufacturing. *J Comput Inf Sci Eng*. 2021;21(5):050902. doi:10.1115/1.4050044.
365. Sodin D, Rudež U, Mihelin M, Smolnikar M, Čampa A. Advanced edge-cloud computing framework for automated PMU-based fault localization in distribution networks. *Appl Sci*. 2021;11(7):3100. doi:10.3390/app11073100.
366. Lee J, Bagheri B, Kao HA. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf Lett*. 2015;3:18–23. doi:10.1016/j.mfglet.2014.12.001.
367. Wang Y, Liu M, Zheng P, Yang H, Zou J. A smart surface inspection system using faster R-CNN in cloud-edge computing environment. *Adv Eng Inform*. 2020;43(1):101037. doi:10.1016/j.aei.2020.101037.
368. Dang HV, Tatipamula M, Nguyen HX. Cloud-based digital twinning for structural health monitoring using deep learning. *IEEE Trans Ind Inform*. 2022;18(6):3820–30. doi:10.1109/TII.2021.3115119.
369. Choquette J, Gandhi W, Giroux O, Stam N, Krashinsky R. NVIDIA A100 tensor core GPU: performance and innovation. *IEEE Micro*. 2021;41(2):29–35. doi:10.1109/MM.2021.3061394.
370. Choquette J. NVIDIA hopper H100 GPU: scaling performance. *IEEE Micro*. 2023;43(3):9–17. doi:10.1109/MM.2023.3256796.
371. Wu X, Qiao F, Poon K. Cloud manufacturing application in semiconductor industry. In: Proceedings of the Winter Simulation Conference 2014; 2014 Dec 7–10; Savannah, GA, USA. p. 2376–83. doi:10.1109/WSC.2014.7020081.
372. Sun S, Ye J, Schwarthoff H, Rosin J, Vakkalagadda V, Chang J, et al. Cloud big data lake for advanced analytics in semiconductor manufacturing. In: Proceedings of the 2024 35th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC); 2024 May 13–16; Albany, NY, USA. p. 1–5. doi:10.1109/ASMC61125.2024.10545365.
373. Lu Y, Xu X, Xu J. Development of a hybrid manufacturing cloud. *J Manuf Syst*. 2014;33(4):551–66. doi:10.1016/j.jmsy.2014.05.003.
374. Harmatos J, Maliosz M. Architecture integration of 5G networks and time-sensitive networking with edge computing for smart manufacturing. *Electronics*. 2021;10(24):3085. doi:10.3390/electronics10243085.

375. Kaur K, Garg S, Aujla GS, Kumar N, Rodrigues JJPC, Guizani M. Edge computing in the industrial Internet of Things environment: software-defined-networks-based edge-cloud interplay. *IEEE Commun Mag.* 2018;56(2):44–51. doi:10.1109/MCOM.2018.1700622.
376. Guin U, Huang K, DiMase D, Carulli JM, Tehranipoor M, Makris Y. Counterfeit integrated circuits: a rising threat in the global semiconductor supply chain. *Proc IEEE.* 2014;102(8):1207–28. doi:10.1109/JPROC.2014.2332291.