International Journal of Innovative Research in Engineering and Management (IJIREM)
ISSN (Online): 2350-0557, Volume 12, Issue 6, December 2025
DOI: https://doi.org/10.55524/ijirem.2025.12.6.20
Article ID IJIR3146, Pages 111-115
www.ijirem.org

A Review of Scheduling Research for Reentrant Process Segments with Special Constraints in Semiconductor FABs

Jongmin Lee

Ph.D. Scholar, Department of e-Business, Graduate School, Kyonggi University, Suwon, Republic of Korea

Correspondence should be addressed to Jongmin Lee; smart7391@naver.com

Received: 8 November 2025 Revised: 25 November 2025 Accepted: 9 December 2025

ABSTRACT- Reentrant flow shops differ fundamentally from traditional linear production lines in which each job or lot visits a machine only once. In a reentrant structure, identical equipment must be revisited multiple times before the manufacturing process is completed, thereby breaking the assumption of a strictly unidirectional production flow. As a result, job precedence relationships become increasingly complex and the load on bottleneck tools grows in a nonlinear manner. In semiconductor fabrication, additional practical constraints—such as recipe changeover minimization in photolithography and batch processing requirements in wet cleaning—further amplify the difficulty of scheduling. These factors lead to the dispersion of waiting times, accumulation of work-in-process (WIP), and a decline in equipment utilization, ultimately lengthening overall production lead time. Such effects are particularly pronounced in high-tech manufacturing industries, including semiconductors, displays, and printed circuit boards, where a single product typically passes through hundreds of processing steps and repeatedly returns to specific stages. This study conducts an in-depth review of five representative research works that focus on scheduling in process segments with reentrant structures and special operational constraints in semiconductor FABs. From an integrated perspective, the paper analyzes how these studies define their scheduling problems, design algorithms and mathematical models to capture structural complexity, and what implications their results offer in terms of both theoretical advancement and practical performance improvement.

KEYWORDS- Cleaning, Photolithography, Reentrant, Scheduling, Semiconductor

I. INTRODUCTION

Modern manufacturing operates under rapid technological advancement and shortened product life cycles, making continuous efficiency improvement and precise operational control indispensable. Within this context, production scheduling has evolved beyond a simple task of arranging job sequences and has become a strategic decision-making domain directly influencing productivity, on-time delivery, and operating costs. In industries such as semiconductors, displays, and electronic components, where process stages can number in the hundreds and equipment interdependencies are strong, any deviation from the planned job flow can destabilize the entire system, thereby undermining a firm's competitiveness in the market [2]. Reentrant flow shops are a representative structure that dramatically exposes these challenges. While a conventional flow shop assumes that each job moves forward once through the sequence of machines, a reentrant flow shop is designed so that the same job revisits specific equipment or stages multiple times. This fundamentally breaks the linear flow assumption and transforms job interactions into a nonlinear and highly entangled structure. Consequently, traditional scheduling methods and simple rule-based dispatching quickly reach structural limitations when applied to such environments [1]. In reentrant systems, merely optimizing the order of equipment usage is insufficient to relieve bottlenecks; instead, the repeated visits of jobs to shared tools cause a sharp increase in waiting times and escalate the complexity of the scheduling problem. Semiconductor fabrication (FAB) is a typical environment in which reentrant structures occur frequently. In photolithography, etching, and deposition processes, products repeatedly visit the same type of equipment in different stages, and these repeated visits intensify contention among tools and generate unexpected waiting times and imbalances. Furthermore, process constraints are far from simple: variability in lot arrival times, setup costs due to mask changes, the need for batch processing, and differences in processing times across recipes must all be considered simultaneously. As this complexity accumulates, conventional scheduling approaches are unable to reflect real-world constraints adequately, and more precise, structure-based algorithmic approaches become necessary. The five studies reviewed in this paper were motivated by this recognition. Each of them starts from the practical difficulties of reentrant and constrained process segments in semiconductor manufacturing, diagnoses the underlying structural characteristics of the process, and proposes scheduling algorithms tailored to these characteristics. In reentrant flexible flow shops, a new perspective based on order-level scheduling was introduced to overcome the limitations of lot-centric approaches [3]. In two-stage reentrant flow shops, branch and bound optimization techniques were employed to analyze the structural properties of NP-hard scheduling problems from a mathematical standpoint [4], [9]. Research on two-machine reentrant flow shops demonstrated how incorporating job weights alters the structure of optimal schedules [5], while a study on semiconductor wet cleaning captured the special nature of batch processing and proved its practical usefulness [6]. Finally, a study on photolithography proposed an integrated scheduling algorithm that simultaneously considers multiple complex constraints and achieves significantly better performance than conventional rule-based operation [7]. Taken together, these studies show that the complexities of reentrant flow shops are not merely technical obstacles but core structural issues that must be explicitly modeled and resolved. By conducting an in-depth analysis of these works, this paper aims to provide a comprehensive understanding of how scheduling research in reentrant segments of semiconductor FABs has evolved and to offer a foundation for discussing future research directions in this field.

II. LITERATURE REVIEW

Research on scheduling in reentrant flow shops has developed steadily as part of a broader effort to understand complex manufacturing systems. Early studies concentrated on elucidating how the structural characteristic of reentrancy affects production flow. Graves and Zipkin argued that when jobs revisit the same machine multiple times, the load on bottleneck resources increases nonlinearly and cannot be adequately explained by classical scheduling models [1]. As jobs return to equipment across multiple passes, precedence constraints between passes become intricately intertwined; the resulting sequence conflicts and accumulation of waiting times emerge as major sources of system instability. These insights suggest that reentrant processes are not just “complicated” but contain fundamental difficulties that lie beyond the reach of conventional scheduling approaches. The work of Garey and Johnson, which systematically analyzed NP-hard problems, provided theoretical support for these difficulties, clarifying that scheduling in reentrant settings is computationally intractable in general [9]. Building on this recognition, many researchers have sought to interpret reentrant structures from various perspectives. Among these contributions, the heuristic proposed by Nawaz, Enscore, and Ham (NEH) has become one of the most influential methods in flow shop scheduling [8]. The NEH algorithm constructs partial schedules based on total processing time and iteratively improves them by job insertion, exhibiting excellent performance in standard flow shops. However, in reentrant structures, simple insertion rules fail to fully capture the repeated-visit characteristics of the process, motivating numerous variants and extensions of NEH that attempt to better reflect reentrancy [3]. In the semiconductor manufacturing domain, Wein’s study represents a milestone [2]. Wein empirically demonstrated the critical importance of bottleneck tool management for overall productivity and formalized the concept of bottleneck-centered scheduling. Because semiconductor processes are characterized by large variation in processing times across tools and a strong tendency for certain machines to dominate overall throughput, scheduling bottleneck tools is not merely a matter of efficiency but a central determinant of system stability. This bottleneck perspective has since become a fundamental paradigm in semiconductor scheduling research. Korean researchers have also actively contributed to this field, and the five papers reviewed in this article are representative examples of their efforts. Choi and Kim [3], [5] and Choi and Shim [4], [7] mathematically analyzed the structure of reentrant flow shop problems, systematically derived dominance properties and lower bounds, and designed branch and bound algorithms capable of finding optimal solutions for medium-sized problems. These studies helped shift the focus of flow shop scheduling research from purely heuristic approaches to more structure-based analytical frameworks. In addition, the study on semiconductor wet cleaning with batch processing by Choi and Kim [6] meticulously modeled practical constraints and achieved substantial performance improvements over conventional rule-based scheduling. The photolithography study by Choi and Shim [7] proposed a modular scheduling algorithm that simultaneously considers mask constraints, recipe differences, and setup costs, while significantly enhancing scheduling performance. In summary, the existing literature has analyzed the structural complexity of reentrant flow shops and semiconductor manufacturing processes and attempted to integrate these insights into the design of scheduling algorithms. These efforts have extended scheduling research from a computational problem to the broader domains of process structure understanding and system optimization. The five studies reviewed in this paper inherit this stream of research while tackling distinct problems in their own ways, thereby occupying important positions in the development of reentrant scheduling research.

III. REVIEW OF SELECTED STUDIES

Scheduling problems in reentrant flow shops and semiconductor fabrication have long been recognized as representative examples of high-complexity problems. However, unlike earlier literature that often described the difficulties of reentrancy—such as sequencing complexity, nonlinear amplification of bottlenecks, and crossing precedence relations—only as obstacles to be overcome, the five studies examined here systematically and integratively analyze these structural issues and propose algorithmic solutions tailored to specific objectives and process characteristics. Rather than simplifying the inherent uncertainty and complexity within a single scheduling problem, these studies confront them directly and clarify how the unique structure of reentrant processes affects performance in actual manufacturing systems. This section analyzes each of the five studies as an independent unit, focusing on problem definition, structural characteristics, algorithm design logic, mathematical foundations, experimental validation, and their academic and industrial contributions to the field of reentrant flow shop scheduling.

A. Order-Based Scheduling in Reentrant Flexible Flow Shops

The first study, by Choi and Kim [3], addresses a flexible flow shop environment with reentrancy and highlights the structural limitations of conventional lot-based scheduling. A flexible flow shop provides multiple parallel machines at each stage, and when combined with reentrant routing—where jobs revisit stages—the dynamic behavior of the system becomes highly intricate. If multiple lots belonging to the same customer order are assigned to different routes or experience divergent waiting times, order-level delivery performance deteriorates significantly, directly affecting customer satisfaction. Recognizing this issue, the authors propose a new perspective that reconstructs schedules at the order level rather than the lot level. They extend the NEH algorithm into an "Order-Based Extended NEH" heuristic, which designs schedules in such a way that all lots belonging to an order progress consistently across stages. In doing so, they shift the scheduling objective from minimizing lot-level metrics to achieving balanced order-level flow. The algorithm also incorporates dispatching rules for parallel machine selection that prevent disruption of order-level flow, thereby mitigating excessive concentration of reentrant jobs at bottleneck stages. Computational experiments show that as the degree of reentrancy increases, the proposed order-based algorithm outperforms conventional NEH variants by a significant margin [3]. This confirms that performance gains are driven not merely by heuristic tuning but by a fundamental alignment between scheduling logic and the underlying production concept—namely, order-level fulfillment in reentrant flexible flow shops.

B. Branch and Bound in Two-Stage Reentrant Flow Shops

The second study, by Choi and Shim [4], considers a two-stage reentrant flow shop in which each job visits two machines twice in the sequence 1–2–1–2, and the objective is to minimize total flow time. Although the problem appears simple, the interleaving of first and second passes and the resulting precedence relations generate an extremely complex combinatorial structure. The relative ordering of jobs across passes leads to numerous potential conflicts, making naive enumeration infeasible and confirming the NP-hard nature of the problem [9]. Departing from heuristic-only approaches, the authors design a branch and bound algorithm that exploits structural properties specific to the reentrant setting. They derive dominance rules that capture conditions under which one partial sequence can be proven inferior to another, thus enabling aggressive pruning of the search tree. Furthermore, they develop sophisticated lower bounds that quantify the impact of interactions between the remaining passes and the current partial schedule on total flow time. The key contribution of this study is that it does not merely solve the problem computationally but translates structural insights into mathematical arguments that guide the search strategy. As a result, the algorithm can obtain optimal solutions for medium-sized problem instances within practical time limits, demonstrating that branch and bound can be effective in reentrant flow shop scheduling when supported by well-designed dominance rules and bounds [4].

C. Total Weighted Flow Time in Two-Machine Reentrant Flow Shops

The third study, also by Choi and Kim [5], examines a two-machine reentrant flow shop but introduces job weights into the objective function. In real manufacturing environments, jobs differ in importance depending on production costs, delivery urgency, and customer priority; thus, minimizing total weighted flow time is a highly realistic objective. Applying this objective to a reentrant flow shop provides insight into how economic priorities interact with structural complexity. The authors analyze the relationship between the first and second passes of each job and show that, once weights are introduced, the completion times of second passes become the dominant determinants of the objective function. In other words, delays in the second pass of heavily weighted jobs disproportionately affect total weighted flow time. This observation is then used to construct lower bounds that explicitly account for the impact of job weights on the remaining schedule. To obtain high-quality upper bounds, the study employs a weighted NEH heuristic as an initial solution generator. This combination of weighted heuristic initialization and structurally informed bounds enables the branch and bound procedure to explore the search space efficiently. The resulting methodology offers practically relevant scheduling guidelines for manufacturers who must prioritize critical jobs while operating under reentrant constraints [5].

D. Minimizing Total Tardiness in Semiconductor Wet Cleaning

The fourth study, by Choi and Kim [6], targets the wet cleaning segment in semiconductor FABs, where batch processing is performed. In this process, a single piece of equipment must handle multiple lots simultaneously, and the quality of batch formation strongly influences overall performance. Simple dispatching rules cannot guarantee good batch composition, often resulting in increased waiting times and severe delivery delays. The authors develop a batch scheduling algorithm that integrates lot arrival times, recipe and process-condition compatibility, and batch feasibility into a unified decision framework. They critically evaluate the limitations of widely used rules such as EMRF, showing that these heuristics do not adequately account for the structural characteristics of batch operations. Simulation experiments reveal that the proposed algorithm significantly reduces total tardiness compared with existing rules, demonstrating that structure-based scheduling can deliver substantial performance gains in highly constrained batch processes [6]. Given that wet cleaning is a critical segment in semiconductor FABs, this study has strong practical implications and offers one of the most immediately applicable algorithmic contributions among the five works reviewed.

E. Modular Scheduling in Semiconductor Photolithography

Photolithography is the most complex and heavily loaded stage in semiconductor manufacturing. It involves high-cost setup operations due to mask changes, heterogeneous parallel tools, recipe-dependent processing times, and uncertainty in lot arrivals. These factors collectively create one of the most challenging scheduling environments in the FAB. In their study, Choi and Shim [7] acknowledge the extreme complexity of photolithography and propose a modular scheduling algorithm that decomposes the problem into structural components. Specifically, they separate the scheduling logic into a "recipe selection module" and a "job selection module," thereby constructing the schedule in stages. This modular design enables the algorithm to simultaneously minimize mask changes and reduce job waiting times—something that conventional rule-based systems, such as EMRF, cannot achieve. Experimental results show that the proposed modular algorithm dramatically improves total weighted flow time relative to EMRF, maintaining robust performance even under high load conditions [7]. This provides empirical evidence that integrated, structure-aware scheduling algorithms can outperform traditional experiential rules, especially in complex, constraint-rich environments like photolithography. Overall, the five studies reviewed in this section analyze the unique structural and operational characteristics of reentrant and constrained processes in semiconductor manufacturing and incorporate them into scheduling algorithms with a high degree of fidelity. Their contributions go beyond incremental performance improvements and instead reshape how complex manufacturing systems are understood and optimized.

IV. DISCUSSION AND FUTURE RESEARCH

The review of research on reentrant flow shops and semiconductor manufacturing environments highlights a clear conclusion: these problems are too complex and volatile to be adequately addressed within traditional scheduling frameworks. Reentrancy amplifies interactions among machines in a nonlinear manner, transforming scheduling from a simple sequencing problem into a structural challenge. In particular, the explosive growth of waiting times around bottleneck tools severely distorts overall process flow and adversely affects key operational indicators such as due-date performance and productivity. The studies examined here illuminate the limitations of conventional rule-based scheduling and point toward a new direction in which complex process structures are systematically analyzed and scheduling algorithms are designed to address the resulting structural issues. Scheduling is thus reframed not as a purely computational exercise but as a strategic tool for interpreting process behavior and alleviating structural bottlenecks. The first direction for future research is multi-objective optimization. Many existing scheduling studies focus on a single objective, such as minimizing total flow time or tardiness. However, in real manufacturing environments, multiple criteria—including energy efficiency, maintenance schedules, workforce allocation, and quality stability—must be considered simultaneously. Multi-objective optimization techniques can help clarify trade-offs among these objectives and identify balanced operating points for the overall system. This would provide a more comprehensive decision support foundation for manufacturing enterprises. The second direction is AI-based scheduling. Recent advances in reinforcement learning, deep neural networks, and metaheuristics have shown strong potential for high-dimensional decision-making problems. Reinforcement learning, in particular, shares a structural similarity with scheduling, as it learns policies through repeated interaction with an environment. In reentrant and highly uncertain settings such as semiconductor FABs, AI-based methods could offer substantial benefits, especially for real-time decision-making under stochastic disturbances. The third direction involves digital twin technology. A digital twin is a high-fidelity simulation model that mirrors the physical manufacturing system. It provides a testbed where scheduling algorithms can be evaluated and tuned under realistic conditions. Digital-twin-based scheduling makes it possible not only to generate static plans but also to monitor system status in real time and dynamically adapt scheduling policies. This can be viewed as an extension of the structure-aware approaches proposed in the five reviewed studies, expanding them into a real-time, closed-loop context. Finally, tight integration between MES and scheduling algorithms will be essential in future manufacturing systems. In real FABs, unexpected events such as equipment failures and rush orders occur frequently and can instantly invalidate static schedules. Therefore, scheduling systems must be capable of responding quickly to real-time data, shifting from fixed plans to dynamically updated schedules. The studies reviewed in this paper provide theoretical foundations for such systems by modeling structural constraints and bottlenecks in detail. Future research should build on these foundations to implement real-time, MES-integrated scheduling solutions.

V. CONCLUSION

This paper has conducted an in-depth review of five representative studies that address scheduling problems arising in reentrant flow shops and semiconductor manufacturing processes. Although these studies target different process structures and objective functions, they share a common characteristic: each accurately captures the complexity of real manufacturing systems and proposes structural solutions capable of alleviating bottlenecks and inefficiencies. By demonstrating that reentrant structures, batch operations, and bottleneck-centered architectures are not merely local features but fundamental determinants of system performance, these works occupy an important position in the scheduling literature. A key feature that unites the five studies is their rigorous understanding of problem complexity and clear explanation of how this complexity affects process performance. Their algorithms are not designed solely to enhance computational efficiency; instead, they incorporate practical operating conditions and constraints with considerable detail, thereby bridging the gap between theory and practice. The photolithography and wet cleaning studies, in particular, show that by precisely modeling real constraints, it is possible to develop algorithms that can be applied directly in industrial settings [6], [7]. Meanwhile, the reentrant flexible flow shop and two-machine reentrant flow shop studies demonstrate how structural analysis can be translated into dominance rules, lower bounds, and algorithmic frameworks that address the intrinsic complexity of scheduling problems [3] [4] [5]. Ultimately, the research reviewed here confirms that scheduling in reentrant environments is not merely a technical issue but a strategic problem that directly influences system-wide efficiency and corporate competitiveness. The theoretical insights and methodological innovations they provide will remain relevant in future digital manufacturing environments and will serve as important milestones in the evolution toward intelligent scheduling. Korean researchers have produced these studies with a high level of originality and quality, and they can be regarded as meaningful academic contributions in the global literature on reentrant process scheduling.

REFERENCES

[1] S. C. Graves and P. Zipkin, "A queueing-theoretic perspective on reentrant manufacturing systems," *Operations Research*, vol. 41, no. 4, pp. 731–746, 1993.

[2] L. M. Wein, "Scheduling semiconductor wafer fabrication," *IEEE Transactions on Semiconductor Manufacturing*, vol. 1,

no. 3, pp. 115–130, 1988. Available from: https://ieeexplore.ieee.org/abstract/document/4384

[3] S.-W. Choi, "Scheduling algorithm for reentrant flexible flow shop with the objective function of minimizing orders' total flow time," *Journal of the Korean Production and Operations Management Society*, vol. 34, no. 1, pp. 129–143, 2023.

[4] S.-W. Choi and S. Shim, "Branch and bound algorithm for two-machine reentrant flowshop with the objective of minimizing total flowtime," *Journal of the Korean Society for Industrial and Systems Engineering*, vol. 33, no. 4, pp. 1–9, 2010.

[5] S.-W. Choi and H. Kim, "Minimizing total weighted flow time in two-machine reentrant flowshops," *Journal of the Korean Institute of Industrial Engineers*, vol. 40, no. 4, pp. 295–306, 2014.

[6] S.-W. Choi, "Scheduling algorithms for minimizing total weighted flowtime in photolithography workstation of FAB," *Journal of the Korean Society for Industrial and Systems Engineering*, vol. 35, no. 1, pp. 79–86, 2012.

[7] S.-W. Choi and D.-K. Kim, "Scheduling algorithms for minimizing total tardiness on clean process in a semiconductor wafer fabrication facility," *Journal of Korean Management Engineering Society*, vol. 15, no. 3, pp. 101–114, 2010.

[8] M. Nawaz, E. E. Enscore, and I. Ham, "A heuristic algorithm for the m-machine, n-job flow-shop sequencing problem," *Omega*, vol. 11, no. 1, pp. 91–95, 1983. Available from: https://doi.org/10.1016/0305-0483(83)90088-9

[9] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA, USA: W. H. Freeman, 1979.