

SIMULATION OF A SEMICONDUCTOR MANUFACTURING RESEARCH AND DEVELOPMENT CLEANROOM

Baptiste Loriferne^{1,3}, Gaëlle Berthoux¹, Valeria Borodin², Vincent Fischer¹, and Agnès Roussy³

¹Technology Platforms Dept., Univ. Grenoble Alpes, CEA, LETI, 17 Rue des Martyrs, 38054 Grenoble Cedex 9, FRANCE

²IMT Atlantique, LS2N-CNRS, La Chantrerie, 4 rue Alfred Kastler, 44307 Nantes Cedex 3, FRANCE

³Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, 13541 Gardanne, FRANCE

ABSTRACT

This paper focuses on a Research and Development (R&D) semiconductor manufacturing system. By virtue of their vocation, R&D facilities tolerate much more variability in processes and outcomes than industrial-scale ones. In such environments, operating under conditions characterized by high uncertainty and occurrences of (un)knowns corresponds to normal operating conditions rather than abnormal ones. This paper characterizes the key entities and operational aspects of a semiconductor R&D cleanroom and introduces a discrete-event simulation model that captures these elements. The simulation model is grounded in empirical data and reflects real-life operations management practices observed in actual R&D cleanroom settings. Preliminary computational results based on real-life instances are presented, and future research directions are outlined to support resilient decision-making in environments where high levels of uncertainty are part of normal operating conditions.

1 CONTEXT, MOTIVATIONS, AND RELATED BACKGROUND

While industrial-scale facilities (*fabs*, for short) are focused on efficiently manufacturing large volumes of chips with high yield and low cost, R&D semiconductor manufacturing systems are driven by innovation and the development of new technologies, materials, and processes (see Table 1). These systems are specifically designed to support experimentation and rapid iteration, enabling the creation and refinement of next-generation manufacturing techniques. In this context, time-to-market for new technologies is critical. R&D fabs frequently run experimental lots to test process changes, identify and mitigate defect mechanisms, and enhance device performance. The success of an R&D fab is ultimately measured by how effectively it transfers mature technologies into industrial-scale manufacturing and meets performance targets.

More specifically, semiconductor R&D fabs differ significantly from industrial-scale ones in several key aspects, including production scale, level of production control, manufacturing *modus operandi*, and performance metrics (Kim et al. 2008; Chen et al. 2017; Borodin et al. 2024). While extensive research has focused on operations management for industrial-scale semiconductor manufacturing fabs (see e.g., Mönch et al. (2011)), there is a noticeable lack of studies aimed at improving the performance of R&D fabs (Ramamurthi et al. 2005).

This gap in the literature can be partly explained by two main factors: (i) The deployment of automation is not so straightforward as in industrial-scale fabs, where operating conditions are known beforehand and manufacturing schemes are predictable. Automation in R&D fabs is implemented cautiously, as these environments must remain highly versatile (Fischer et al. 2023). For instance, while Manufacturing Execution Systems (MES) in industrial-scale fabs enforce fixed recipes and standard operating procedures, MES in R&D fabs must allow for on-the-fly modifications. Since automation relies on standardized processes, it often conflicts with the inherent flexibility required in R&D. Engineers frequently need to manually create new recipes for experimental steps that are too novel to be predefined in the MES (Fischer

Table 1: Semiconductor manufacturing fabs: R&D *versus* industrial-scale

Aspect	R&D	Industrial-scale
Objectives	Innovation and acceleration of learning Low throughput	Market supply, cost efficiency, and yield High throughput
Key characteristics	Experimentation driven by engineers and statistical analysis: High tolerance to variability	Rigorous process control: Minimal tolerance to variability, see e.g., (Moyne and Iskandar 2017)
	Limited redundancy, extensive equipment diversity	Extensive redundancy in equipment
	Bypassable automation enabling customized executions (Fischer et al. 2023; Fischer et al. 2025)	(Quasi-) complete automation to standardize and stabilize operations
Philosophy of operations management and process control	Optimization for flexibility, exploration, and knowledge acquisition; time-to-market for new nodes is critical (Kuhl et al. 2004)	Optimization for stability, efficiency, and competitiveness, see e.g., (Moyne et al. 2016; Mönch et al. 2018)
Key performance indicators	Agility and learning: Number of experiments conducted, achieved technology milestones, speed of knowledge acquisition	Stability and efficiency: Throughput, cycle time, on-time delivery, productivity, and cost
Related literature in manufacturing sciences	Application still very limited, mainly focused on production control (dispatching), see e.g., (Ramamurthi et al. 2005; Kim et al. 2008; Kuhl et al. 2004; Borodin et al. 2024)	Extensive literature covering all decision levels of supply chain planning (strategic, tactical, operational) with advanced predictive and prescriptive tools, see e.g., (Mönch et al. 2011; Mönch et al. 2018)

et al. 2023). (ii) R&D fabs are subject to and tolerate significantly higher variability in both processes and outcomes (Loriferne et al. 2025).

One of the earliest contributions to R&D semiconductor manufacturing was proposed by Liao et al. (1996), who addressed the uncertainty caused by frequent equipment failures and yield fluctuations. They proposed a daily scheduling approach that enables quick and smooth sequencing adjustments, thereby supporting more agile manufacturing management.

Among the limited existing studies, Ramamurthi et al. (2005) examine production control strategies in semiconductor R&D fabs, aiming to reduce time-to-market through simulation-based methods. Their work evaluates how critical factors (such as bottleneck utilization, lot prioritization, dispatching policies, and due date constraints) impact key performance indicators like on-time delivery and cycle time. In another study, Borodin et al. (2024) assess and challenge a dispatching rule-based heuristic applied in real-life settings. To improve throughput and shorten turnaround time, Kim et al. (2008) propose a fab-wide scheduling approach that simultaneously schedules all lots across the full set of machines. Their approach combines mixed-integer programming with constraint programming to optimize overall fab efficiency.

In terms of capacity management and productivity enhancement in R&D settings, Kuhl et al. (2004) introduce a simulation model for a semiconductor and microsystems fabrication laboratory. The model assesses fab performance through key indicators such as cycle time, Work-In-Progress (WIP), and throughput of representative load lots.

Given this context and further in this direction, this paper focuses on modeling, simulating, and evaluating the performance of operations and activities management of an R&D semiconductor manufacturing system under conditions characterized by known unknowns. The proposed discrete-event simulation tool is designed as a foundation for future integration with prescriptive capabilities, enabling resilient operations management in environments where high uncertainty and (un)known unknowns are considered normal operating conditions.

The remainder of this paper is structured as follows. Section 2 defines the problem under study. Section 3 presents the proposed discrete-event simulation framework. Section 4 discusses the results of the numerical experiments conducted on real-world industrial data. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2 PROBLEM DESCRIPTION

Unlike industrial-scale fabs, an R&D center in the semiconductor manufacturing domain does not produce physical products. Instead, it develops wafer manufacturing methods by inventing, improving, modifying, and integrating individual processes into cohesive flows. This inherently innovative nature of R&D demands significant human expertise and intervention. Operators must parameterize machines at each process step to fine-tune the processes. Moreover, redundancy in equipment is rare. Redundancy, when it exists, arises from the overlap of capabilities among machines. However, certain steps require specific machines and skilled operators, making the impact of machine failure uneven and highlighting the critical role of operator knowledge.

Non-redundancy also arises at the lot level. Once a lot becomes deployable for industrial-scale manufacturing, its development is usually complete, and it transitions out of the R&D center. Furthermore, low levels of production control mean that each lot may follow a unique route, which evolves throughout the process. This uniqueness makes traditional lot batching nearly impossible. Instead, *campaigns* are used, grouping similar process steps rather than lots, based on shared parameters.

In the context of complex and innovative processes, failures and scrap must be minimized. To mitigate these risks, wafers are often separated from their original lots, forming the so-called *son lots* to enable testing. This form of *lot filiation* is common for engineering steps. However, these son lots are not always tracked independently in the MES, as they may be embedded within the step duration. Some are treated implicitly within tasks, while others are considered separate lots.

Another form of lot filiation is the *definitive split*, where a single parent (or *father*) lot is deliberately divided into two derivative lots that will not recombine. This practice enables parallel experimentation with alternative process trajectories and outcomes within the scope of a single project.

The problem addressed in this paper involves the characterization, modeling, and simulation of a semiconductor manufacturing R&D system, with the goal of supporting decision-making in environments where high levels of uncertainty are inherent to day-to-day operations. The current focus is on the sufficiency of the representation and analysis of system dynamics to capture the effects of critical aspects, and the integration of optimization capabilities is left as a direction for future work.

3 DISCRETE-EVENT SIMULATION-BASED APPROACH

The discrete-event simulation, defined over a discrete time horizon T , is lot-oriented and is based on a Work-In-Process (WIP) object. The main building blocks are described in Section 3.1. The management of quantitative operations is provided in Section 3.2.

3.1 Description of Core Building Blocks

Entities and parameters. Consider a set of lots (i.e., jobs) $j \in \mathcal{J}$ in the WIP and their associated sets of operations (i.e., steps) $\mathcal{O}_j = \{o_{1,j}, o_{2,j}, \dots, o_{n_j,j}\}$ of cardinality n_j , and a set of machines \mathcal{M} located in a set of work areas \mathcal{Z} . Each work area $z \in \mathcal{Z}$ is associated with workers w grouped into teams W_z . Let d_m be the statistically estimated processing time of machine $m \in \mathcal{M}$, measured in simulation time units.

As explained in Section 2, lots in the R&D center under study are nearly unique, i.e., there are almost as many distinct lots as there are products. The processing of job j can be started if the following constraints are met:

- All needed resources (machine and qualified operator) must be free and available/ready for use,
- No ongoing son lots,
- No ongoing hold for the lot.

As proposed in (Loriferne et al. 2025), let us distinguish the following resource types in an R&D manufacturing system:

- *Machines*: Machines in an R&D fab are selected primarily for their ability to support experimental processes and accommodate a broad spectrum of process configurations, including novel and evolving steps. Since demand in terms of volume is typically very low in R&D settings, there may be only one or very few machines of each type. Consequently, machine downtime is generally higher in R&D, either due to lower utilization pressure or frequent reconfigurations by engineers. The machine downtime is characterized by Time To Repair (*TTR*) and Time Between Failures (*TBF*). These parameters are tracked, traced, and explicitly considered in the proposed simulation model.

In terms of architecture, the MES of the studied R&D cleanroom distinguishes two categories of machines: **(i) Mono-module machines**: Mono-module tools that process lots serially using one specific process. Each mono-module machine can have multiple capabilities, although these are not always cross-compatible or simultaneously available. **(ii) Multi-module machines**: Machines composed of multiple modules grouped under a so-called *mainframe*. These are logically subdivided into virtual machines (referred to as *virtuals* for short).

- *Workforce*: Workforce resources are modeled at the workshop level, where each team W_z is capable of operating most machines within its assigned work area $z \in \mathcal{Z}$. In the simulation, operators within a work area are considered interchangeable. To process a lot in a given work area $z \in \mathcal{Z}$, an operator must be available and not already engaged in another task. Nominal call behavior (i.e., the preemptive assignment of operators to lots) is not yet implemented, as it requires a higher level of data granularity that is currently unavailable in the simulation framework.

States and transitions. Machine behavior is modeled in terms of $\{\text{Up}, \text{Down}\}$ states and capability configuration. Following the approach proposed by Loriferne et al. (2025), machine failures are represented using two key stochastic parameters: TBF and MTTR. The values of these parameters are dynamically generated whenever a machine transitions between states, introducing time-varying randomness into the simulation. Both TBF and TTR are defined for each mainframe and its associated modules.

It is worthwhile mentioning that lots are not processed directly on mainframes or individual modules. Processing occurs on virtual machines. A virtual becomes eligible to process a lot only if the following conditions are met:

- The mainframe is Up, i.e., operational.
- All modules in the virtual machine path are also Up.

The processing time of a given lot is computed based on the characteristics of the assigned virtual machine. During processing, both the lot and the machine are considered occupied until one of two outcomes occurs: **(i)** The processing time elapses and the operation completes successfully, or **(ii)** A failure interrupts the operation. In the event of a failure, the lot is released without completing its step, while the machine remains unavailable for the duration of the repair time.

The simulation is designed to reflect real-world constraints on resource availability in the cleanroom and system dynamics. Algorithm 1 simulates the behavior of the system over one period of simulation $t \in T$, while accounting for jobs (i.e., lots) and their holds, machine activity and failures, and workforce availability. The manufacturing operations are managed as described in Section 3.2. It explicitly incorporates all key factors influencing cycle time per process step and processing speed. The model accounts for lot filiation and holds. Holds are implemented using fixed durations and a probabilistic triggering mechanism at each process step, calibrated using empirical historical data as described in (Loriferne et al. 2025).

Algorithm 1 Discrete-event simulation model associated with a given period $t \in T$.

```

for  $j \in \mathcal{J}$ , where  $\mathcal{J}$  is ordered by a dispatching policy do ▷ Lot processing and holds
  if  $\text{remainingTimeInProcess}(j) = 0$  and  $\text{isTreatable}(j)$  then
     $o_j^{\text{current}} \leftarrow$  current step  $o_{i,j}$  of job  $j$ 
     $m \leftarrow \text{machine}(o_j^{\text{current}})$ 
    if ( $\text{leftoverTime}(m) \leq 0$  and  $\text{isUp}(m)$ ) and  $\exists w \in W_z$  s.t.  $\text{occupancyTime}(w) \leq 0$  then
       $\text{leftoverTime}(m) \leftarrow d_m$ 
       $\text{remainingTimeInProcess}(j) \leftarrow d_m$ 
    end if
  else if  $\text{remainingTimeInProcess}(j) > 0$  then
     $\text{remainingTimeInProcess}(j) \leftarrow \text{remainingTimeInProcess}(j) - 1$ 
    if  $\text{remainingTimeInProcess}(j) \leq 0$  then
      if  $\text{rand}(0, 1) \leq \alpha$ , where  $\alpha \in (0, 1)$  is a hyper-parameter then
         $\text{remainingTimeInProcess}(j) \leftarrow \text{holdDuration}$ , as defined in (Loriferne et al. 2025)
      else
        Go to next step  $o_{i+1,j}$  of job  $j$ 
      end if
    end if
  end if
end for
for  $m \in \mathcal{M}$  do ▷ Machine activity and failures
   $\text{failure}(m) \leftarrow$  elapsed time since the last state change of  $m$  as defined in (Loriferne et al. 2025)
  if  $\text{leftoverTime}(m) > 0$  then
    if  $\text{failure}(m) \geq \text{TBF}(m)$  then
       $j' \leftarrow j \in \mathcal{J}$  s.t.  $o_j^{\text{current}}$  is processed by machine  $m$ 
      if  $j' \neq \text{Null}$  then
         $\text{remainingTimeInProcess}(j') \leftarrow 0$ 
      end if
       $\text{leftoverTime}(m) \leftarrow \text{TTR}(m)$ 
      Change the state of machine  $m$  to Down
    end if
     $\text{leftoverTime}(m) \leftarrow \text{leftoverTime}(m) - 1$ 
  else if State of machine  $m$  is Down then
    Set state of machine  $m$  to Up
    Generate randomly the next  $\text{TBF}$  as defined in (Loriferne et al. 2025)
  end if
end for
for  $z \in \mathcal{Z}$  do ▷ Workforce aspect
  for  $w \in W_z$  do
     $\text{occupancyTime}(w) \leftarrow \text{occupancyTime}(w) - 1$ 
  end for
end for

```

3.2 Production Planning and Control

Focus on campaign. As previously mentioned, the equipment park and its configuration in an R&D fab are dimensioned for flexibility and experimental capability scope. Machines in such environments

must often accommodate a wide variety of process settings, including entirely new or modified process steps. For this reason, machines are frequently unique or nearly so in their configuration and use. The different capabilities of a machine can be mutually exclusive, requiring time-consuming setup or teardown operations to switch from one capability to another.

To manage this complexity, campaigns are used as a scheduling approach for machine capabilities. A *campaign* typically spans a predefined period of duration, during which specific capabilities, deactivated by default, are temporarily activated. Conversely, some capabilities that are normally active may be deactivated for the duration of the campaign. Campaigns generally fall into three categories:

- **Size:** The machine supports only one wafer size at a time. Changing the accepted size requires human intervention, including maintenance and qualification.
- **Contamination:** The campaign temporarily alters the contamination level of the machine, either increasing or decreasing it.
- **Materials:** The chemical composition or target material used by the machine needs to be changed.

Campaigns are also characterized by different types of periodicity. Some campaigns follow a fixed periodic schedule, typically on a weekly or monthly basis. Other ones occur less frequently, such as every three months, and are planned accordingly. The remaining campaigns are project-driven, initiated due to specific constraints, and often have a short window of anticipation.

These campaigns are managed through the control of capabilities, specifically their activation or deactivation relative to a predefined standard state. Based on the level of control knowledge available, campaigns can be classified into four distinct levels:

- **Level 0:** This level reflects the lowest degree of knowledge, typically in engineering contexts. At this level, no specific capabilities are identified or provided.
- **Level 1:** At this level, there is partial knowledge regarding which capabilities should be activated or deactivated. However, the standard state of a machine is clearly defined, which distinguishes it from the next level.
- **Level 2:** In this case, there is no defined standard state. The machine operates in a permanent campaign mode. Each campaign specifies the capabilities to be activated, as no deactivation is required since the concept of a standard state (with default capabilities) is absent.
- **Level 3:** This level represents the highest degree of control. Both the standard state of a machine and the sets of capabilities to be activated and deactivated during a campaign are fully defined.

These campaigns are characterized by the following elements: a start date, a setup time, and lists of capabilities to be activated and deactivated. The teardown time is considered part of the setup time of the subsequent campaign. The standard state is also modeled as a campaign with its own set of capabilities. Since the setup time can vary depending on the preceding campaign, it is represented using a cost matrix.

Focus on scheduling and dispatching. No formal analytical specifications are available to determine which machines are qualified or which recipes are suitable for a given job and operation (Fischer et al. 2023). Instead, engineers typically rely on their experience and understanding of product requirements to select appropriate machines and recipes. The current dispatching approach consists of two main phases (Borodin et al. 2024):

- **Bin assignment:** Rather than explicitly modeling individual resources, a set of resource bins is defined. Each resource bin corresponds to a specific combination of priority level and operation type. These bins have fixed and limited capacities. Prior to sequencing, operations are assigned to resource bins based on job priority, which is defined according to R&D specifications.

- **Sequencing:** After bin assignment, jobs within each bin are ordered based on a compound priority score. This score accounts for several factors, including the adjusted processing speed required to meet the job deadline, the number of remaining operations in the job route, and the slack time.

This dispatching strategy is designed to support On-Time Delivery (OTD), as the R&D center operates under a make-to-order production model. Time-to-market for new technology is critical.

4 COMPUTATIONAL EXPERIMENTS

This section discusses the accuracy of the proposed simulation model based on the real-world instances described in Section 4.1. Numerical experiments are designed as detailed in Section 4.2. Section 4.3 provides an analysis of R&D fab performance and examines the impact of dispatching policies on key performance indicators.

4.1 Description of Instances

Numerical experiments have been performed on real-life instances provided by CEA-Leti, one of the three European research and technology organizations in semiconductor manufacturing. The dataset comprises hundreds of unique products and includes approximately 2,500 lots present in the Work-In-Progress (WIP) at any given time. During the simulation, lots are added daily to the WIP based on historical inflow patterns, replacing those that have completed processing.

Machine parameterization is also based on historical data. Specifically, four years of data, 2021–2025, have been used to model machines, with the chosen time span serving to mitigate the impact of the COVID-19 pandemic, during which the cleanroom was partially closed. Following the multi-resource modeling approach from (Loriferne et al. 2025), machines are represented using a modular approach, where each machine is decomposed into *modules* (corresponding to specific capabilities) and *mainframes* (physical assemblies).

4.2 Design of Experiments

Numerical experiments were conducted following an experimental plan structured around several aspects:

- **Accuracy of the simulation model:** The discrete-event simulation model described in Algorithm 1 has been validated against the ground truth to examine and shed light on the following key features:
 - *Multi-resource constraints:* Two modeling approaches are compared: **(i) As-is model:** This model corresponds to the machine model currently used to manage cleanroom operations; and **(ii) Proposed multi-resource model**, detailed in Section 3 and in (Loriferne et al. 2025).
 - *Machine failures and lot holds* modeled as proposed in (Loriferne et al. 2025).
- **Impact of critical features:** The combined effect of explicitly considering machine failures and lot holds has been estimated and reported.
- **Dispatching policies and cleanroom activity:** Section 4.3 analyzes three policies within the proposed simulation framework: **(i)** random lot shuffling, **(ii)** bin-level shuffling, and **(iii)** the current dispatching heuristic.

Numerical experiments were performed on a desktop computer equipped with an Intel i5 processor ($2 \times 2.80\text{Hz}$) and 8 GB of DDR4 RAM. With a time granularity of 5 minutes, simulating a two-month period required approximately 13 minutes. To simulate multiple instances, a computing farm consisting of heterogeneous machines was utilized, resulting in variable execution times. The findings reported in this paper are derived from 20 replications of the simulation, which were empirically determined to be sufficient for validating the current version of the model and for generating managerial insights.

4.3 Analysis of the Obtained Results

Validation of the sufficiency of the simulation accuracy. The accuracy of the simulation model has been validated in terms of *cleanroom activity*. The activity serves as a key performance indicator, corresponding to the number of processed steps per day.

Figure 1 illustrates the capability of the proposed simulation model to replicate real-world activity patterns within the cleanroom. Firstly, it is essential to note that, statistically speaking, the proposed simulation approach, which integrates multi-resource modeling, lot holds, and machine failures, closely replicates the actual cleanroom activity. From Figure 1, it is also evident that the as-is machine model currently used to manage cleanroom operations consistently results in a lower number of processed steps per day, with associated mean processing times appearing to be overestimated.

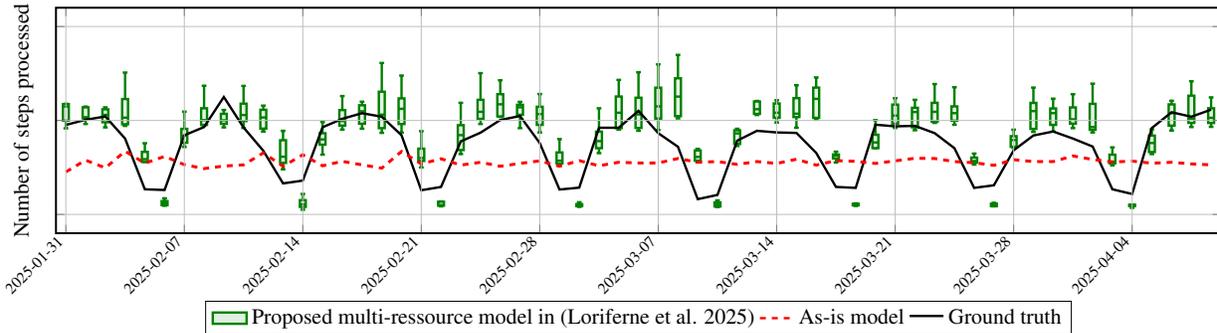


Figure 1: Comparison between ground truth and simulation results, accounting for multi-resource constraints, machine failures, and lot holds: *Number of steps processed per day*. The black line represents the historical activity, while the boxplots represent the simulated daily activity over 20 replications. The y-axis values have been intentionally omitted to preserve confidentiality.

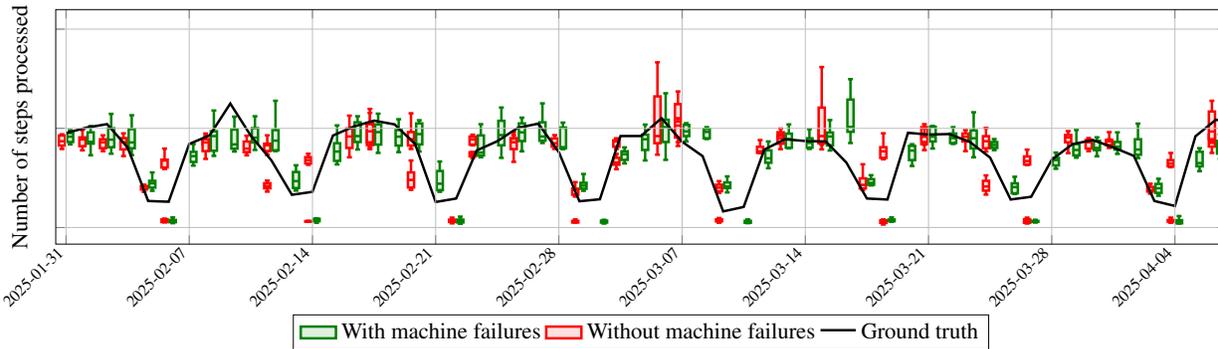


Figure 2: Simulation results including the proposed multi-resource model, lot holds, and with/without machine failures: *Number of steps processed per day*. The black line represents the historical activity, while the boxplots represent the simulated daily activity over 20 replications.

Let us now investigate the impact of machine failures and lot holds using the proposed simulation model. Figure 2 illustrates the effects of machine failures on cleanroom activity. While the average level of WIP activity remains relatively stable, its variability increases significantly when machine failures are explicitly considered. Due to the complexity of the system and the presence of many unique machines, activity is unevenly distributed across them. This phenomenon is closely related to the so-called *big runners*, referring to machines that handle the majority of activity within each area of the cleanroom. Their impact

on activity patterns is more pronounced in R&D contexts, where machine redundancy is very limited. These findings highlight that adaptability to machine failures and maintenance can serve as an important optimization lever.

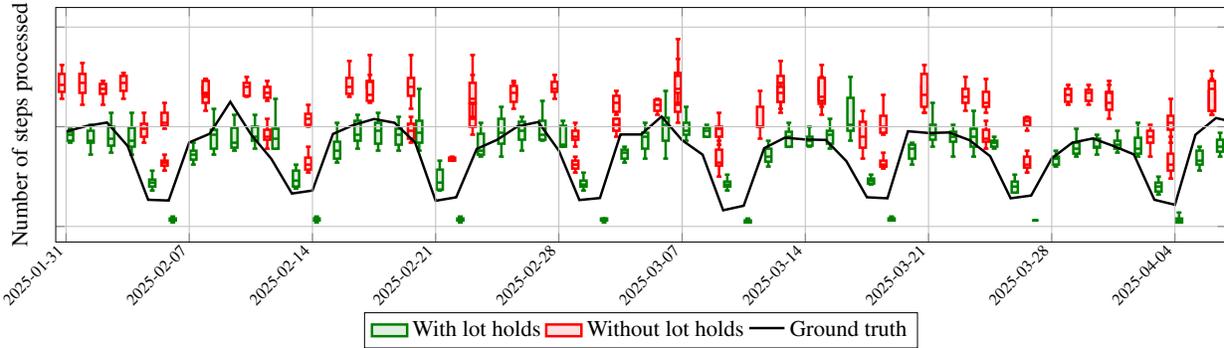


Figure 3: Simulation results including the proposed multi-resource model, machine failures, and with/without lot holds: *Number of steps processed per day*. The black line represents the historical activity, while the boxplots represent the simulated daily activity over 20 replications.

Figure 3 shows the effects of lot holds on cleanroom activity. The alignment between the ground truth and the simulation results validates the empirical approximation applied to model lot holds. This alignment allows decision-makers to quantify the potential gains achievable by reducing lot holds.

In the studied R&D center, engineering support is required for the majority of processing steps. For new processing steps, processing times are initially unknown, and subject matter experts must estimate the time needed for their completion. Conversely, while the processing times of mature steps are generally better controlled, they often still depend on feedback from engineering, inline, and offline teams before proceeding to subsequent steps. This tuning process can introduce significant delays.

The results provided in Figure 3 can be seen as the theoretical maximum activity achievable under ideal operating conditions, assuming lot holds are resolved immediately. The profile defined by red boxplots serves as a best-case reference, highlighting the potential margin for improvement in operational efficiency.

Performance assessment. Figure 4 illustrates the impact of WIP size on global activity in the cleanroom and mean lot speed. As expected, increasing the WIP size leads to higher global activity, since a greater number of machines can be utilized simultaneously. However, this comes at the cost of reduced lot speed. The slowdown is primarily due to longer runner queues and increased machine occupancy, which introduce additional delays in lot processing.

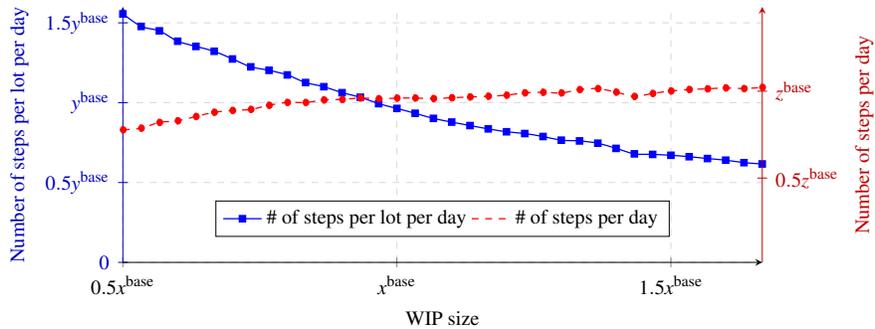


Figure 4: Average cleanroom activity per day and lot activity per day as a function of WIP size, where point $(x^{base}, y^{base}, z^{base})$ denotes the current baseline coordinates.

Additional experiments have been conducted to investigate the behavior of the currently deployed dispatching rule under conditions of high variability and uncertainties. To this end, three policies have been implemented so far and compared in terms of daily activity, as illustrated in Figure 5. The first policy corresponds to the as-is situation, serving as a baseline for comparison. The second policy corresponds to a completely shuffled WIP, ignoring all dispatching prescriptions and lot splitting into bins as detailed in Section 3.2. The third policy presents an intermediate case. It also shuffles the WIP but maintains the separation between bins.

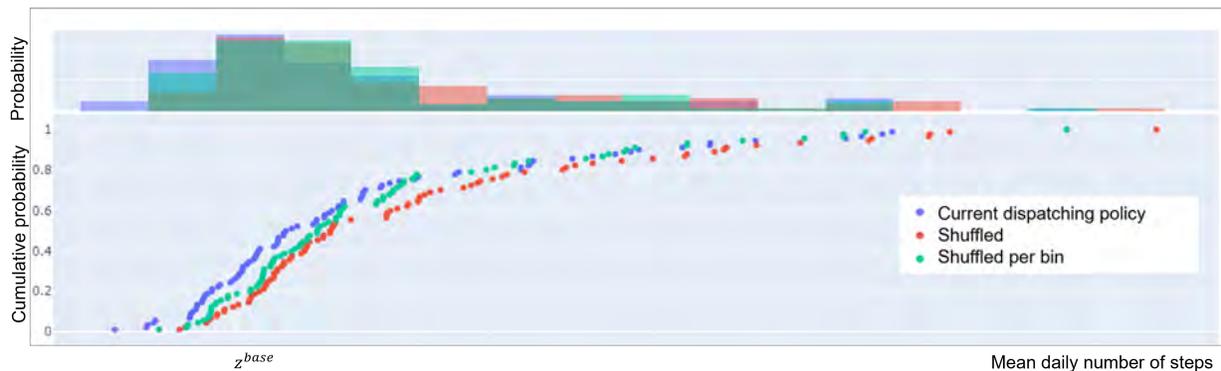


Figure 5: Empirical probability distributions of mean daily activity for three dispatching policies: **(i)** Random lot shuffling, **(ii)** bin-level shuffling, and **(iii)** current dispatching policy.

Together with the cleanroom activity, the distribution of On-Time Delivery (OTD) is analyzed to assess its alignment with real-world performance. However, it should be noted that OTD heavily depends on the assignment of lots to bins, which is influenced by external factors beyond the cleanroom environment and is therefore difficult to reproduce accurately.

It is observed that shuffling the sequence proposed by the current dispatching rule-based heuristic increases overall activity. This improvement is likely due to a more diverse workload distribution across machines, resulting in better average performance, regardless of whether the shuffle is applied to the entire WIP or separately within each bin. However, this gain comes at a cost. High-priority lots, which are the most critical, experience significant delays, while lower-priority lots tend to finish much earlier than necessary. Early completion for lower-priority lots is generally not beneficial. Approximately one-third of the lots were significantly accelerated, while the remaining two-thirds suffered delays. As expected, the current dispatching approach is more effective in terms of OTD. It better aligns completion times with delivery priorities, minimizing the gap between scheduled and actual delivery dates.

It is worth noting the long right tails of all distributions in Figure 5. The maximal achievable activity lies far above the mean, not only under the current dispatching policy but also across the other random-based scenarios. This observation highlights the significant combined impact of the frequency of lot holds and machine failures. Addressing these conditions is part of our ongoing research on supporting operations management of the R&D cleanroom.

5 CONCLUSION AND PERSPECTIVES

This paper presents the key entities, their interactions, and the contextual characteristics specific to research and development semiconductor manufacturing systems. A simulation-based approach is proposed and validated using real-world data and settings. The simulation tool is designed to be equipped with decision-making capabilities, enabling it to propose resilient solutions in response to the high levels of uncertainty that are intrinsic to normal R&D operating conditions.

A deeper understanding of the various resource aspects provides valuable insights into the criticality of specific machines and the distribution of workloads across workshops. In collaboration with the reporting team, additional what-if scenarios and performance indicators are currently being developed to further extend the decision analytics scope of the simulation-based tool.

REFERENCES

- Borodin, V., V. Fischer, A. Roussy, and C. Yugma. 2024. "Scheduling Semiconductor Manufacturing Operations in Research and Development Environments". In *2024 35th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–6. Albany, NY, USA, May 13th–16th, 2024.
- Chen, Y., D. Kayarat, and J. Binford. 2017. "Multi Objective Optimization of Process R&D Goals in Automated Semiconductor Manufacturing Systems". In *Proceedings of the IIE Annual Conference*, 524–529. Pittsburgh, Pennsylvania, USA, May 20th–23rd, 2017.
- Fischer, V., O. Landré, and M. Duranton. 2023. "Automation in R&D: Complying with Contradictory Constraints of Seemingly Incompatible World". In *2023 34th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–4. Saratoga Springs, NY, USA, May 1st–4th, 2023.
- Fischer, V., O. Landré, and B. Vuailat. 2025. "A Novel Statechart Modeling Language Adapted to R&D Automation Development: Topic: FA: Factory Automation". In *2025 36th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–5. Albany, NY, USA, May 5th–8th, 2025.
- Kim, Y.-H., J.-H. Lee, and D.-S. Sun. 2008. "The Operational Optimization of Semiconductor Research and Development Fabs by FAB-Wide Scheduling". *The Transactions of the Korean Institute of Electrical Engineers* 57(4):692–699.
- Kuhl, M. E., K. D. Hirschman, V. Ramamurthi, and G. R. Laubisch. 2004. "Capacity and Productivity Modeling for Research Laboratories Using a Representative Product Load". In *2004 IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop (IEEE Cat. No. 04CH37530)*, 157–161. Boston, MA, USA, May 4th–6th, 2004.
- Liao, D.-Y., S.-C. Chang, K.-W. Pei, and C.-M. Chang. 1996. "Daily Scheduling for R&D Semiconductor Fabrication". *IEEE Transactions on Semiconductor Manufacturing* 9(4):550–561.
- Loriferne, B., G. Berthoux, V. Borodin, V. Fischer, and A. Roussy. 2025. "Modeling Resources in Research and Development Semiconductor Manufacturing Shop Floors". In *2025 36th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, 1–6. Albany, NY, USA, May 5th–8th, 2025.
- Mönch, L., J. W. Fowler, S. Dauzère-Pérès, S. J. Mason, and O. Rose. 2011. "A Survey of Problems, Solution Techniques, and Future Challenges in Scheduling Semiconductor Manufacturing Operations". *Journal of Scheduling* 14:583–599.
- Mönch, L., R. Uzsoy, and J. W. Fowler. 2018. "A Survey of Semiconductor Supply Chain Models Part I: Semiconductor Supply Chains, Strategic Network Design, and Supply Chain Simulation". *International Journal of Production Research* 56(13):4524–4545.
- Moyne, J., and J. Iskandar. 2017. "Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing". *Processes* 5(3):39.
- Moyne, J., J. Samantaray, and M. Armacost. 2016. "Big Data Capabilities Applied to Semiconductor Manufacturing Advanced Process Control". *IEEE Transactions on Semiconductor Manufacturing* 29(4):283–291.
- Ramamurthi, V., M. Kuhl, and K. Hirschman. 2005. "Analysis of Production Control Methods for Semiconductor Research and Development Fabs Using Simulation". In *Proceedings of the Winter Simulation Conference (WSC)*, 9–. IEEE <https://doi.org/10.1109/WSC.2005.1574504>.

AUTHOR BIOGRAPHIES

BAPTISTE LORIFERNE is a PhD student at Mines Saint Etienne, France, completing his thesis in industrial engineering, automation, and computer science in the premises of CEA-Leti. He holds a B.S. in Computer Science and Applied Mathematics, as well as a Master of Science in Operations Research. His research interests include operations research and management, combinatorial optimization, scheduling, and semiconductor manufacturing. His e-mail address is baptiste.loriferne@cea.fr.

GAËLLE BERTHOUX is an IT Engineer at CEA-Leti. She works on specialized KPIs for cleanroom applications, based on Manufacturing Execution Systems (MES) activity data treatment. She previously worked on information systems and business requirements in diverse technological environments, initially in the public sector and then in MES within the semiconductor industry. Her email address is gaelle.berthoux@cea.fr.

VALERIA BORODIN is an Associate Professor at IMT Atlantique, France, where she is actively involved in academic–industrial collaborations and knowledge transfer initiatives. She received her Ph.D. in Optimization and Systems Safety from the University of Technology of Troyes, France. Her research focuses on quantitative operations management in manufacturing

and logistics systems, including mathematical modeling, optimization, simulation, and data analytics. Her email address is valeria.borodin@imt-atlantique.fr.

VINCENT FISCHER has been working in the semiconductor industry for the past 20 years, in the areas of factory automation and electronic design automation. He is currently the section manager for operations planning, cleanroom logistics, supplies, and cleanroom IT. He has been leading the cleanroom automation team at CEA-Leti for 5 years. He has also developed customization modules on the Leti MES solution. Prior to joining Leti, he worked at different EDA companies, developing software for yield optimization and high-sigma yield analysis. He has an MSc and PhD in Electrical Engineering, both from the Grenoble Alpes University, France. His email address is vincent.fischer@cea.fr

AGNÈS ROUSSY is Professor at the CMP of Mines de Saint Etienne. She received the Ph.D degree from INP-Grenoble, France. She was a Postdoctoral Fellow at the University of Twente, NL. She has been Associate Professor and since 2020, she is Professor at Mines de Saint Etienne in France. She works in the field of Advanced Process Control applied to semiconductor manufacturing. Her e-mail address is roussy@emse.fr.