Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

ABOUT AI-BASED REAL-TIME DISPATCHING AS COMPARED TO OPTIMIZED SCHEDULING IN SEMICONDUCTOR MANUFACTURING

Peter Lendermann[1]

[1]D-SIMLAB Technologies Pte Ltd, Singapore, SINGAPORE

ABSTRACT

With the rise of Artificial Intelligence (AI) in recent years, AI techniques have also become a hot topic in Semiconductor Manufacturing. This presentation will look at Reinforcement Learning enabled WIP flow management and discuss why some recent advancements in the field of AI-enabled dispatching have to be taken with caution and why such techniques bear much more potential for production scheduling instead.

1 REINFOREMENT LEARNING BASED REAL-TIME DISPATCHING

In the light of recent advancements in Artificial Intelligence (AI), there is a huge potential for application of AI techniques in the Semiconductor Manufacturing industry to further increase the efficiency and performance of factories. In this setting, Reinforcement Learning (RL) enabled real-time dispatching has been considered an important step towards realization of an "Autonomous Control Room" as described by Mitra (2023). For example, Stöckermann et al. (2023) demonstrated how a Dispatch Agent enabled by Deep RL with evolution strategies can be trained in an industrial-grade Discrete Event Simulation testbed. Sood et al. (2024) showed how wafer fab operations can be supported using multi-agent RL. In both cases, measurable performance enhancement potential with regard to rule-based dispatching techniques was demonstrated for the respective wafer fabrication facility. Such better KPI performance should actually be expected because a significantly larger number of factors is taken into consideration for deriving such dispatching decisions.

However, a wafer fab is a complex dynamic system. Batching or queue time constraints can only be considered indirectly though dispatching parameters that first need to be determined empirically or in a high-fidelity simulation testbed. The question then arises how likely is it that a sufficiently comparable situation would have already been observed (and analyzed) during a simulation-enabled training period and whether sufficient statistics would have been collected for each situation. Ultimately, even though some KPI performance enhancement with regard to rule-based dispatching can be demonstrated, such situation-based real-time dispatch decisions still constitute a sequence of independently determined local optima.

2 GENERATION OF SCHEDULES ACROSS EQUIPMENT GROUPS

To overcome some inherent shortcomings of dispatching, scheduling across multiple equipment spanning several subsequent production steps has evolved as a methodology to further enhance the performance of Semiconductor Manufacturing operations as to increase throughput and reduce cycle times. A number of powerful commercial scheduling solutions have been developed and successfully deployed in wafer fabs across the world in recent years. The underlying principle is to assign ('schedule') WIP lots to production equipment in sequence such that a set of KPIs is likely to be met in the best possible manner over a defined scheduling horizon. This needs to be done on a regular basis to account for all kinds of external or internal changes to the production environment and its drivers such as equipment downs (which to a significant extent are stochastic) or demand changes.

A major advantage of scheduling is that – as compared to dispatching – even more factors such as queue time constraints or batching can directly be considered, especially for the calculation and assessment of schedule candidates. An integrated optimum across several equipment groups, comprising several dispatching decisions to be executed within the immediate future, will result in better performance as compared to the aggregated effect of multiple, individually optimized dispatching decisions during the same time horizon. Also, generation of such schedules does not require past experience, rather scheduling can be considered "training and learning on the spot". In this setting, it should be iterated that the above-mentioned KPI enhancements associated with RL-enabled dispatching have been proven with regard to rule-based dispatching but not with regard to optimized scheduling. In fact, optimized scheduling would be inherently superior to dispatching for the above-mentioned reasons, no matter whether it is rule-based or RL enabled.

3 WHERE AI TECHNIQUES CAN ADD VALUE

The above-described limitations of RL enabled dispatching as compared to optimized scheduling do not imply that such RL techniques cannot add value to the generation of optimized dispatch sequences. As illustrated in Figure 1, scheduling is about searching for the best of many possible sequences and allocations of lots in the pipeline to equipment during a defined time horizon, under the assumption that for stochastic events the most likely path would be followed.

As compared to real-time dispatching, because of many iterations required, the finite duration of schedule generation is the only real drawback of scheduling as random events can have some effect on schedule validity and optimality. Along the process of determining an optimized schedule, the identification of a potentially good initial schedule and subsequently the search for better schedule candidates is a complex task where many factors and hence many alternative schedule candidates are to be considered. So an RL enabled trained Scheduling Agent would know which schedule change could possibly have how much (positive) effect on the KPI performance associated with the scheduling objective. Only few iterations would then be required to identify schedule candidates leading to a better KPI performance, so an optimal (or rather for all practical purposes sufficiently good) schedule can actually be found much faster which also means that the finite duration for the generation of such schedules would be further reduced.

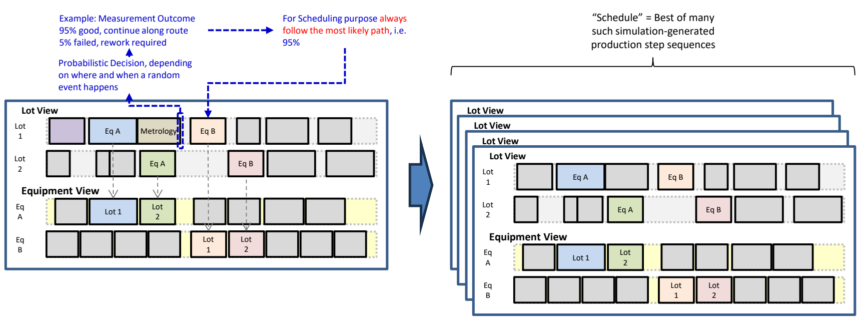

Figure 1: Simulation-based scheduling.

This approach would combine the advantages of optimized scheduling with the power of RL techniques. In the presentation, these considerations will be substantiated further through a sequence of graphical animations and numerical examples.

REFERENCES

Mitra, D.B. 2023. "Semiconductor Autonomous Smart Factory of the Future". In *SEMICON WEST Smart Manufacturing Pavilion*, San Francisco, 2023.

Sood, I., *et. al.* 2024. "Supporting Fab Operations Using Multi-Agent Reinforcement Learning". In *35th Annual SEMI Advanced Semiconductor Manufacturing Conference*, Albany NY, USA.

Stöckermann, P., *et. al.* 2023. "Dispatching in Real Frontend Fabs with Industrial Grade Discrete-Event Simulations by Deep Reinforcement Learning with Evolution Strategies". In *2023 Winter Simulation Conference (WSC)*, 3047-3058.