This is the Accepted Manuscript of a conference paper presented at the 2020 31st Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), Saratoga Springs, NY, USA, published by IEEE under http://dx.doi.org/10.1109/ASMC49169.2020.9185201

The project iDev40 has received funding from the ECSEL Joint Under-taking under grant agreement No 783163. The JU receives support from the European Union's Horizon 2020 research and innovation programme. It is co-funded by the consortium members, grants from Austria, Germa-ny, Belgium, Italy, Spain and Romania. It is coordinated by Infineon Technologies Austria AG. The content of this article does not reflect the official opinion of the Joint Undertaking ECSEL. Responsibility for the information and views expressed in the article lies entirely with the authors.

Diese Maßnahme wird mitfinanziert durch Steuermittel auf Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Order Release Methods in Semiconductor Manufacturing: State-of-the-Art in Science and Lessons from Industry

Jacob Lohmer
Chair of Business Management, esp. Logistics
Technische Universität Dresden
Dresden, Germany
jacob.lohmer@tu-dresden.de

Christian Flechsig
Chair of Business Management, esp. Logistics
Technische Universität Dresden
Dresden, Germany
christian.flechsig@tu-dresden.de

Rainer Lasch
Chair of Business Management, esp. Logistics
Technische Universität Dresden
Dresden, Germany
rainer.lasch@tu-dresden.de

Konstantin Schmidt
Infineon Technologies Dresden GmbH & Co. KG
Dresden, Germany
konstantin.schmidt2@infineon.com

Benjamin Zettler
Infineon Technologies Dresden GmbH & Co. KG
Dresden, Germany
benjamin.zettler@infineon.com

Germar Schneider
Infineon Technologies Dresden GmbH & Co. KG
Dresden, Germany
germar.schneider@infineon.com

***Abstract*—This contribution presents an industry case study as well as an analysis of the state-of-the-art in science concerning order release methods in wafer manufacturing in the semiconductor industry. The release of orders into the fab significantly influences critical parameters such as WIP, cycle time and throughput. We examine the processes currently applied in industry, indicate the effects of this order release approach on the performance of high-mix, high-volume fabs and establish a link to the analyzed scientific literature to develop a concept for meaningful automation of the release decision.**

Keywords—order release, workload control, state-of-the-art review, case study, factory automation

I. INTRODUCTION

Since the 1980s, the semiconductor industry has seen an increase in academic attention. Four basic steps form the semiconductor manufacturing process: Wafer fabrication, sort (or probe), assembly and test. Two different facilities are involved in the typical manufacturing process: A frontend facility, where fabrication and sort take place and a backend facility for assembly and test. If the tasks in production planning and control are hierarchically structured, the presentation in Fig. 1 typically results.

In this paper, the focus is on the order release task in the frontend facility, which connects the planning and the control level by distributing the predetermined specifications in terms of time and quantity [1]. The result is a plan for the upcoming week(s) that defines which jobs should be started into production (specifically wafer fabrication) at what date and time. This decision strongly influences the manufacturing key performance indicators such as work-in-process (WIP), cycle time (CT), and throughput of the fab as the semiconductor manufacturing process is argued to be the most complex manufacturing process in existence [1]. This is due to re-entrant product flows, complicated and expensive equipment, hardly predictable yield rates, and equipment downtimes as well as fast-changing product mixes [1], [2].

This research was supported by the EU project iDev40. The project iDev40 has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 783163. The JU receives support from the European Union's Horizon 2020 research and innovation program. It is co-funded by the consortium members, as well as grants from Austria, Germany, Belgium, Italy, Spain, and Romania.

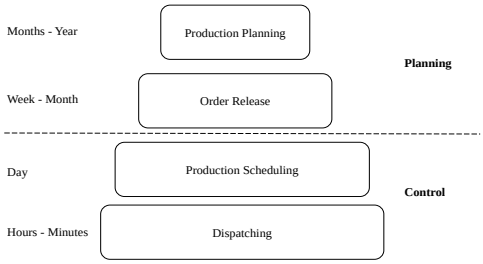

Fig. 1. Production planning and control hierarchy (based on [1])

Order release strategies in semiconductor manufacturing are also referred to as input regulation policies or lot release methods, since orders are dispatched and tracked as lots in fabs. The latest review of order release strategies specifically designed for semiconductor manufacturing dates back to 2002. Fowler et al. [2] indicated that there is still a need for subsequent research and especially for research-industry transfer, as few companies were entirely relying on automated scientific approaches. As the topic of optimization and automation of order release planning and the investigation of the influence of different order release strategies emerged in a recent research project, a state-of-the-art review of order release strategies is presented in this contribution. Additionally, we investigate the current practical implementations based on a case study within a major semiconductor manufacturer in Europe and establish a link to the analyzed literature to develop a concept for meaningful automation of the release decision.

II. LITERATURE REVIEW – STATE-OF-THE-ART IN RESEARCH

To analyze the state-of-the-art in research on order release strategies in semiconductor manufacturing, we conducted a systematic search for peer-reviewed articles in journals and conference proceedings using the scientific databases EBSCO Business Source Complete, ScienceDirect and Google Scholar [3], [4]. A time criterion was not applied as we attempt to provide a comprehensive review of existing research efforts. Used search strings included "order" OR "lot" AND/OR "release" AND "semiconductor" OR "wafer", which returned 152, 46, and 168 articles, respectively. After removing duplicates and screening the abstracts for

indications of semiconductor manufacturing and specific order release keywords, 65 publications remained. These publications were thoroughly read and screened regarding specific research methodology, semiconductor reference, and whether new release strategies were applied. Articles examining existing semiconductor order release strategies were excluded for the analysis in Table I. 24 articles passed this stage and constituted our final literature set that was then analyzed. Table I shows the criteria used for the analysis and categorization of existing literature and the case study.

Criteria and categories for the analysis include order release method, objectives, focus, bottleneck consideration, outcome, limitations, and a case study indicator. The articles all used simulation to apply and test the proposed order release policies, except for [5] that combined simulation and optimization and [6] that focused on optimization solely. The column "order release method" indicates the specific method of releasing orders into the fab. All articles assessed utilized closed-loop methods. Closed-loop (CL) includes dynamic shop floor information and usually a "WIP cap" [7]. In contrast, open-loop (OL) refers to a release decision (based on queueing theory and reduced variability) that does not consider the current fab situation but only refers to exogenous information like demand. Frequently employed methods are a constant release (one lot every x minutes) and a random release (release all lots that are available at the beginning of a shift). Next, objectives that are controlled and measured in the contributions are indicated. Except for the case study and [1], which used a static release execution (e.g., once a week), the release methods are dynamic and allow lots to be released at any time. The focus of each method is elaborated to provide concise insights into the specific release process. If a bottleneck is considered for releasing orders, this is indicated in the next column. The outcome of conducted experiments regarding the performance of the proposed release policy and limitations of each contribution are provided next. The last column displays whether a case study was used to test and implement the release method or not. Several articles were not included in Table 1 as they focus on reviews of workload control [8] and CONWIP systems [9] or propose and test new dispatching rules only [10]–[12].

The analysis of identified literature shows that scholars and industry started with OL policies (random or constant release of lots into the fab). Since the 1990s, there has been a trend towards CL methods based on early works by Wein or Glassey and Resende [13], [14]. In general, CL methods outperform OL methods as they incorporate system information before deciding on lot releases and may dynamically change release plans when disturbances occur (see e.g. [15]). As Little's Law explains the interrelation of average CT and average WIP level, limiting the WIP by employing a "WIP cap" is a common approach to limit CT of lots in the fab. WR, CONWIP, CONLOAD, WIPLCtrl, CONSTWL, and more, all utilize the WIP level of bottleneck workstations or the overall shop floor to control the release of lots. These approaches have shown superior performance to a uniform or random release in simulation experiments. In general, 75 % of the proposed policies considered bottlenecks to set a workload limit. This implies that regular checks have to be carried out to ensure that the bottlenecks still persist and the influence of the alignment with the bottleneck on other work centers has to be determined.

Experiments to benchmark existing release policies were conducted by [16] and [15]. Performance measures CT and coefficient of variation (CoV) of the arrival process were considered in [16], which indicated that CONWIP outperforms a deterministic release policy in settings with low CoV values but not in case of high CoV values. In [17], the results from earlier studies on OL vs. CL methods are confirmed: CL methods outperform OL methods due to their ability to react to the real-time status of the fab. This includes WIP congestion and machine failures. At high throughput levels of 90% and more, the best policies are WR, CONWIP, and CONLOAD [17]. However, these methods are not easy to implement and maintain in real-life wafer fabs. WR limits the work amount in the fab based on the deterministic fab-wide bottleneck that is regularly shifting in regular operation. Besides, the distribution of WIP in the fab is not considered. CONLOAD extends the WR policy to limit the amount of bottleneck workload, also based on deterministic bottlenecks. CONWIP, on the other hand, ignores the distribution of workload completely and only checks whether lots leave the system to trigger a new release then. Identifying the WIP threshold is not an easy task either.

New approaches have been proposed to overcome the mentioned drawbacks, with a recent focus on minimizing mean CT and standard deviation of CT [18]–[21] as well as maximizing on-time delivery [1], [22], [23]. A reduction of average CT has various positive effects on the overall fab performance: minimized customer response time, lower WIP levels, and lower yield loss. Nevertheless, as indicated in [14], WIP at the bottleneck workstations is an important performance measure as well as it is assumed that the bottleneck limits the system capacity. When order release methods are discussed, the used dispatching rules are important as well, as both decisions are related. FIFO (here equal to FCFS) is most commonly used, followed by SRPT. In experiments with different release methods, SRPT achieved the best results at given throughput levels when compared with FIFO and LIFO [17]. The still existing focus on FIFO can be attributed to the easy implementation and the high variation of raw process times of semiconductor products.

When analyzing the case study category in Table I, it is apparent that an investigation of real wafer fabs and the effects of proposed release methods on their performance has not yet been carried out to a sufficient extent. Except for [19], which features a case study at Chartered Semiconductor in 2009 with 511 machines and 37 part types, the promising policies proposed in the literature have not been tested in a simulation model that is comparable to real-world high-volume, high-mix semiconductor manufacturing in a highly automated fab. For ease of simulation and comparison, the Intel Five Machine Six Steps (Mini Fab) is commonly used, e.g. [17], [22], [24]. It includes re-entrant flows, batch processors, machine failures, set-ups, and two main products, as well as one test wafer product. In particular, the small number of products has to be criticized, as the release method in high-volume manufacturing can quickly lead to the suppression of low volume products, which in turn leads to yield losses and ultimately lost sales.

III. State-of-the-Art in Industry – Case Study

In the case study, we conducted several interview rounds with different stakeholders to capture the actual release process in high-volume, high-mix wafer fabs. The details of the actual release policy are also indicated in Table I.

TABLE I. STATE-OF-THE-ART LITERATURE CLASSIFICATION

Reference, Year	Order release method	Objective	Focus	Bottle-neck	Outcome	Limitations	Case study
[14], 1988	*Starvation Avoidance (SA)* input method and a new dispatching method	Max. bottle-neck utiliza-tion, min. WIP	Bottleneck with virtual inventory (working hours for all jobs in the queue), release controlled to avoid starvation at the bottleneck	X	*SA* outperformed uniform and fixed-WIP release when compared on KPIs like delay or throughput. Release control has more impact on CT than dispatching	Single bottleneck, single product, global inventory control necessary	-
[13], 1988	*Workload Regulation* (*WR*) input and *workload balancing* dispatching method	Min. CT	The remaining process time at the bottleneck is monitored with a critical value. If the value is undercut, new jobs are released	X	Mean and variability of CT reduced by *WR* approach, release control has more impact on CT than dispatching	Single bottleneck, workload target needs to be estimated and set	X
[25], 1989	*Flow Rate Control*	Min. WIP, improve predictability	Threshold values for stock and output control production	-	WIP reduced, predictability increased when compared to uniform loading	Single product type, a simple simulation model	-
[26], 1990	*CONstant Work In Process (CONWIP)*	Min. inventory & flow times	Kanban-like cards that control the release of lots based on a constant WIP level	-	WIP, average, and variance of flow time reduced	System WIP level is limited without further information on the fab status	-
[27], 1990	*Flow Rate Hub Control (Two-boundary control)*	Min. total inventory & backlog cost	Policy controls the photolithography cell as the hub of production, rules similar to [25]	X	The proposed policy outperformed uniform loading in mean throughput and total cost	Tested against uniform release only, a small simulation study	-
[5], 1996	*Descending control rule (DEC)*	Min. CT	Simulation-optimization approach to use linear control rules for order release, approximate control set depends on the fab situation	X	Min. average inventory at the same throughput compared to deterministic input, *WR* and *CONWIP*	High-volume, but single product fab	X
[28], 1996	*Dynamic release control (DRCP)*	Min. mean CT and tardiness, max. throughput	Projected queue size in the system as a threshold to release jobs	-	Mean tardiness and waiting time decreased, esp. when FIFO is used as dispatching rule	3 products and 7 workstations only	-
[29], 1998	*Parametric workload regulating (PWR)*	Max. throughput, min. WIP and CT	A new job is released into the system when the future bottleneck workload falls below a critical value	X	CL rules outperformed OL, *WR* was slightly better than *PWR* on CT and throughput	Single product simulation experiment	-
[30], 1999	*CONstant LOAD (CONLOAD)*	Keeping bottleneck utilization at the target level	*CONLOAD* considers the amount of load for the bottleneck (processing times / avg. CT)	X	*CONLOAD* outperformed *CONWIP*, *WR* and *PUSH* in bottleneck utilization, but worse performances in changing product mixes	Simple simulation model, avg. CT of each product required	-
[31], 2001	*Layerwise CONWIP, Total CT*	Min. variability (WIP & CT)	*CONWIP* applied to each fabrication process layer, *Total CT* uses the total CT in the system to release lots	X	Variability of WIP and CT of the fab is reduced but mean values of WIP and CT increase	Focus on the bottleneck, simplification of other work centers	-
[6], 2005	*Multi-constraint based finite capacity mechanism*	Min. CT & machine changeover	A holistic approach combining lot prioritization and capacity allocation for a high product mix	X	CT decreased while minimizing machine conversion	Partial release of lots is not preferred	X
[32], 2006	*Revised constant WIP (RCONWIP)*	Achieving target WIP	System WIP and first workstation WIP threshold to release lots into the fab	X	Desired throughput of each product achieved while WIP smoothened and bottleneck utilization controlled	No comparison with other release methods	-
[22], 2008	*Dynamic classified WIP (DC-WIP)*	Min. CT and WIP, max. throughput and OTD	*DC-WIP* controls the WIP allocation according to the max. production rate of the bottleneck machine	X	*DC-WIP* outperformed an OL method and an avg. WIP method on CT, WIP, and ODR	Mini-fab model, deterministic bottleneck	-
[18], 2008 [19], 2009	*WIPLOAD Control (WIPLCtrl)*	Min. avg. CT & standard dev. of CT	Sum of rem. processing time of all jobs is the reference *WIPLOAD* level according to an expected throughput rate	X	*WIPLCtrl* outperformed *UNIF* and *CONWIP* for a given output: The higher the output, the higher the improvement.	Reference WIP depends on desired throughput, no job rework	X
[24], 2011	*Continuous and periodic WIP review*	Min. avg. CT, max. throughput	*CONWIP*-like system that allows the release level to decrease and increase by more than one lot at a time	X	Periodic review of WIP reduces arrival variability	Mini-fab model	-
[33], 2012	*WIP Balance* approach and due date control	Min. CT, tardy lots, avg. tardiness	Min. and max. workload for bottlenecks and non-bottlenecks to avoid starvation. Priority for tardy lots at upstream centers.	X	Low volume products were sped up, tardiness and CT reduced without affecting high volume products	Single dynamic bottleneck	-
[1], 2013	*Three shift release, job train release*	Min. CT, max. OTD	Case Study to start jobs periodically in three shifts each day, with addition of batch-optimized releases	-	Improvement of the number of jobs completed by day and OTD for three-shift release. Batching has a negative influence in most cases.	Only uniform methods tested	X
[20], 2014	*Effective-workload control (EWL-n-Ctrl)*	Min. WIP, avg. and stand. dev. of CT, max. throughput	Similar to *WIPLCtrl*, the weight of a lot is based on its remaining effective-workload on the shop floor. QTR as new dispatching rule	-	Performance measures improved when applying *EWL-n-Ctrl* and QTR dispatching	Low-volume fab model (4 products)	-
[21], 2014	*WR with extreme learning machine (WRELM)*	Max. util. & throughput, min. CT	Dynamic setting of critical value for *WR* based on real-time bottleneck status information	X	Performance of *WRELM* equal to best *WR*, improved throughput and CT when compared to average *WR* performance	Only 5 products and a single bottleneck workcenter	-
[7], 2016	*Release policy based on extreme learning machine (RPELM)*	Max. ODR, throughput, min. CT	Release based on a priority index consisting of processing and CT time, steps and lot priority (hot)	X	*RPELM* can improve OTD when compared with FIFO and EDD as release rules	The method was not tested against any pull-based release policy	-
[23], 2016	*Rel_M2*	Max. ODR and throughput, min. CT	Jobs are released depending on the primary bottleneck workload of each parallel manufacturing line	X	Compared with *CONWIP*, the method improves the ODR, CT, and throughput	Only compared with CONWIP on a small model	-
[17], 2018	*Constant workload (CONSTWL)*	Max. bottleneck utilization, min. CT	Modified *CONWIP* based on the overall shop floor workload	-	*CONSTWL* reduces WIP inventory at the bottleneck and outperforms *WR*, *CONWIP* and *CONLOAD* on delay/throughput	Mini-fab model, policy was not compared with recent policies	-
[34], 2019	*Constant batch machine workload (CONSTBWL)*	Max. bottleneck utilization, min. CT	Release based on batch machine workload	X	*CONSTWBL* with FIFO, LIFO, or SRPT reduced avg. CT and WIP inventory but increased the CT stand. dev.	The method was not tested against recent policies	-
Case company	Uniform release	CT, throughput, WIP balancing	Achieve optimal delivery reliability	-	WIP balancing for to be released product types only, no bottleneck consideration	Balanced product mix with prioritized releases based on due dates committed to customer while considering the capacity cap	X

CL = closed-loop, CT = cycle time, FIFO = First-In-First-Out, LIFO = Last-In-First-Out, max. = maximized, min. = minimized, OL = open-loop, ODR = on-time delivery ratio, OTD = on-time delivery, rem. = remaining, SRPT = Shortest-Remaining-Processing-Time, WIP = Work-in-process

The release type is periodical; each week, the lot release plan for the upcoming week is created once. The committed sales orders, which have to be released in the coming week to meet their promised delivery date, are passed from production planning to production control. Lots are sequenced in a pre-shop pool taking into account urgent orders as well as the WIP level and the die-bank stock of the corresponding product type. The die-bank inventory relates to the intermediate storage in the semiconductor production process where the customer order decoupling takes place (in the sense of postponement, refer to [35] for a detailed description of typical manufacturing steps and inventories of semiconductor manufacturing). The WIP level for the product types to be released is checked and used in the manual allocation of start dates and times. In this way, the loop is both open and closed, as some real-time information is used but does not directly impact the release decision (like in a pull approach). The lots are distributed uniformly over the following week, using a weekly capacity cap that is split up into days. Bottlenecks are not considered for the release decision, as they frequently change over time. The resulting plan is only changed in case of emergencies. Batching is carried out for lots with small quantities to ensure that full boxes (25 wafers) of raw material can be used on the shop floor. The overall goal is achieving optimum delivery reliability.

The dispatching rules in the fab are specific to each workstation, with global and local rule sets. The basic global approach is to use ODD (operational due date) for lots that have a tight due date and FIFO for all lots with a relaxed due date. During the observation period, the priorities of the different product types changed on a weekly basis. The process of assigning the priorities is neither comprehensible nor transparent for the production control team. Unlike in literature (e.g. [6]), lots in the pre-shop pool are not ranked according to a fixed order criticality index that considers due dates. Instead, the priority is determined and recorded by the responsible production planning employees themselves.

IV. Analysis and Discussion

Different process steps are strongly affected by subjective influences in the current process. The uniform distribution of lots to be released over the weekdays, taking into account the prioritized lots on the first days of the week, leads to variability on the shop floor and has an influence on the WIP distribution. When analyzing the actual production data in the case study, the impact of the release decision became apparent: The batching processes performed before releasing lots into production influence the performance of all subsequent operations. The results are an over- or underload of manufacturing equipment and spikes of WIP levels at workstations throughout the fab, affecting the queuing times at tools.

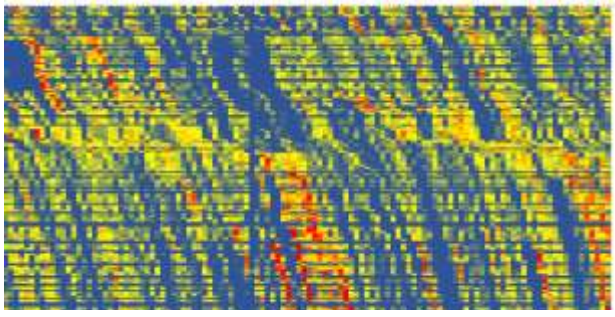

Fig. 2. WIP analysis of a representative product in the fab

Figure 2 shows the WIP levels at different operations over the production time of the lot. The red areas indicate a significantly increased WIP level. The very first operations influence the performance of the entire frontend production for this product type, based on the release decision that also provides batching suggestions. A detailed quantitative evaluation is not presented here for reasons of confidentiality.

When comparing the practical approach with the literature, it is apparent that no scientific approach is explicitly used, and the release method can be best described as a uniform release. Some of the basic ideas of the scientific approaches have nevertheless made their way into practice. The WIP level is considered in the release decision, as done by approaches like WR or CONWIP. However, it is only used at the particular moment of deciding statically on releases to further prioritize or deprioritize lots that must be released in any way. The overall WIP in the fab is not considered, although the introduction of new products inevitably influences the cycle times of all products through the re-entrant process flows. It is also conducive to consider bottlenecks. Unlike in the literature, in our case study, a bottleneck is not regarded as a critical resource that should not fall below a certain WIP level in order to guarantee a high utilization. Instead, a bottleneck is defined as a resource that has a high WIP level that should decrease in terms of the performance of all products and the associated cycle times. New lots should be released to the bottleneck after a decrease in WIP only. Based on a snapshot of the current fab status at the time of deciding on the static release decision, lots are released earlier (low or acceptable WIP at current bottlenecks) or later (high WIP).

The different release policies from the literature were discussed in internal meetings, with a general reluctance to rely on an automated pull-based approach entirely. CONWIP, in its original form, has been tested in the case company some years before and did not show encouraging results. However, the stakeholders expressed their assent to the need to improve the process. One of the participants suggested that the epitaxy processes, which are within the first steps of the fabrication process, should be considered in the batch generation process due to their long set-up times. The epitaxy machines are often set-up in advance based on the routings of the expected products and may then have to be converted at great expense if a different release is carried out. On the other hand, harmonizing the release decision with this tool group only makes sense if the scheduling (or dispatching) on the preceding workstations then process and pass on the batched lots accordingly. It would also be helpful to specifically consider bottlenecks, in the sense of machines that should not be left without processing stock. The production planning department in our case study plans the orders in such a way that the utilization of the machines is as close to the maximum as possible. However, this again ignores the current WIP situation. We see some potential for improvement here.

We examined the production conditions in our case study in more detail to prepare a future implementation of order release mechanisms with control of workload, referred to as workload control (WLC) mechanisms in literature [36]. As workload control (input/output control) is a promising tool to control the production in different industries, several criteria for evaluation of the “best fit” of workload control to a specific case study have been proposed [37].

The factors and characteristics were analyzed for the production method in our case study (high volume, high mix) in order to be able to make an initial estimate of the fit of the workload control (see Table II). The evaluation ranges from "+" indicating promising conditions for the application of the WLC ("best fit") to "-" for a poor fit for a reasonable implementation of the WLC.

Orders arrive regularly during the week, but are only processed in bundles by the production control staff. This could be addressed in the new concept and modified to a regular arrival rate easily. A pre-shop pool is available and used to absorb any inter-arrival time variability. In addition, there is usually more than one raw material that can be used for a product. As semiconductor manufacturing has cycle times of several weeks, the due dates are typically quite tight. However, delivery commitments are often made weekly, i.e., a product has seven different acceptable delivery dates within a week, which allows some flexibility for the release decision. Again, the pre-shop pool of raw materials allows coping with variability in due dates. Processing times, on the other hand, are very long and do not vary much between products, so there is little flexibility to balance workloads in this context. The ratio of set-up time to processing time is advantageous in most areas, with a lot of processing time in relation to set-up time. In some factory areas, however, set-up time can be almost twice as long as processing time. This characteristic was therefore rated as indifferent (o) in Table II. Routing sequence variability was rated as a good fit as several different products are manufactured. Routing lengths and length variability must rather be evaluated as a bad fit since semiconductor manufacturing with re-entrant flows is complex and lengthy routing plans, with significant variabilities. Turning to route flexibility: In general, there is a certain flexibility of processing, but in many production areas, investments in machines are associated with such high costs that only single machines are qualified for a product type. In these areas, it is challenging to balance the workload, so the general fit was rated as rather poor. Pre-assembly is rather simple as the chips are separated and no further semi-finished products for joint assembly have to be brought together here.

TABLE II. WORKLOAD CONTROL "BEST FIT"

Contextual factors	**Characteristics**	**Rating for case study**
Order arrival intensity	High arrival rate of small jobs allows greater release flexibility	+
Inter-arrival time	Pre-shop pool absorbs high inter-arrival time variability	+
Due date tightness	Loose due dates lead to flexibility	o
Due date variability	Buffering in pre-shop pool suits a high variability in due dates	+
Processing time lumpiness and variability	Short processing times and a high variability provide flexibility to balance workloads	-
Set-up/processing time ratio	A low ratio is preferred for workload control	o
Routing sequence variability	High variability provides flexibility due to more options	+
Routing length and length variability	Short lengths with a high variability preferred	-
Routing flexibility	High flexibility allows balancing workloads across work centers	-
Level of convergence	Few assembly processes preferred to focus on release step	+

The analysis shows that there is at least a decent fit of the WLC concept. Some areas need to be carefully assessed and addressed in the implementation of a future concept. Besides these issues, we are preparing to more general implementation issues like the need for "quick wins", training of users and involving all stakeholders in the process of developing and deploying a new order release concept [36].

V. CONCLUSION AND OUTLOOK

Order release decisions that affect key performance indicators such as WIP, cycle time, and fab throughput vary between science and practice in semiconductor manufacturing. In this paper, the scientific methods are analyzed and compared with the practical approach in a high-mix, high-volume fab. To automate the release decision, suitable features from the literature are indicated and the fit to the WLC concept is analyzed for the case study to develop a suitable future customized release approach that can be evaluated using simulation.

For further research, we plan to implement several release methods from the literature in the approved and confirmed simulation model of the real factory. This should provide us with insights into the actual influence of the release decision. Besides, we also plan to investigate the influence of priority distributions in the current release method. A uniformly distributed release, without specific prioritization of products at the release stage, could provide some insights into the general behavior of the fab and reveal the potential for improvement of the release policy.

REFERENCES

[1] L. Mönch, J. W. Fowler, and S. J. Mason, *Production Planning and Control for Semiconductor Wafer Fabrication Facilities*, vol. 52. New York, NY: Springer New York, 2013.

[2] J. W. Fowler, G. L. Hogg, and S. J. Mason, "Workload control in the semiconductor industry," *Prod. Plan. Control*, vol. 13, no. 7, pp. 568–578, 2002, DOI.10.1080/0953728021000026294.

[3] C. F. Durach, J. Kembro, and A. Wieland, "A New Paradigm for Systematic Literature Reviews in Supply Chain Management," *J. Supply Chain Manag.*, vol. 53, no. 4, pp. 67–85, 2017, DOI.10.1111/jscm.12145.

[4] A. M. T. Thomé, L. F. Scavarda, and A. J. Scavarda, "Conducting systematic literature review in operations management," *Prod. Plan. Control*, vol. 27, no. 5, pp. 408–420, 2016, DOI.10.1080/09537287.2015.1129464.

[5] C. R. Glassey, J. G. Shanthikumar, and S. Seshadri, "Linear control rules for production control of semiconductor fabs," *IEEE Trans. Semicond. Manuf.*, vol. 9, no. 4, pp. 536–549, 1996, DOI.10.1109/66.542169.

[6] W. Liu, T. J. Chua, T. X. Cai, F. Y. Wang, and W. J. Yan, "Practical lot release methodology for semiconductor back-end manufacturing," *Prod. Plan. Control*, vol. 16, no. 3, pp. 297–308, 2005, DOI.10.1080/09532780500088043.

[7] L. Li, Z. Chen, Q. Yu, and N. Xiang, "Learning-based release control of semiconductor wafer fabrication facilities," *Proc. - Winter Simul. Conf.*, vol. 2016-Febru, pp. 2965–2973, 2016, DOI.10.1109/WSC.2015.7408400.

[8] M. Thürer, M. Stevenson, and C. Silvaa, "Three decades of workload control research: A systematic review of the literature," *Int. J. Prod. Res.*, vol. 49, no. 23, pp. 6905–6935, 2011, DOI.10.1080/00207543.2010.519000.

[9] J. Prakash and J. F. Chin, "Modified CONWIP systems: A review and classification," *Prod. Plan. Control*, vol. 26, no. 4, pp. 296–307, 2014, DOI.10.1080/09537287.2014.898345.

[10] Y. Kim, J. Kim, S. Lim, and H. Jun, "Due-Date Based Scheduling and Control Policies in a Multiproduct Semiconductor Wafer Fabrication Facility," *IEEE Trans. Semicond. Manuf.*, vol. 11, no. 1, pp. 155–164, 1998.

[11] Y. Kim, J. Kim, B. Choi, and H. Kim, “Production Scheduling in a Semiconductor Wafer Fabrication Facility Producing Multiple Product Types With Distinct Due Dates,” *IEEE Trans. Robot. Autom.*, vol. 17, no. 5, pp. 589–598, 2001.

[12] N. Bahaji and M. E. Kuhl, “A simulation study of new multi-objective composite dispatching rules, CONWIP, and push lot release in semiconductor fabrication,” *Int. J. Prod. Res.*, vol. 46, no. 14, pp. 3801–3824, 2008, DOI.10.1080/00207540600711879.

[13] L. M. Wein, “Scheduling semiconductor wafer fabrication,” *IEEE Trans. Semicond. Manuf.*, vol. 1, no. 3, pp. 115–130, 1988, DOI.10.1109/66.4384.

[14] C. R. Glassey and M. G. C. Resende, “Closed-loop job release control for VLSI circuit manufacturing,” *IEEE Trans. Semicond. Manuf.*, vol. 1, no. 1, pp. 36–46, 1988, DOI.10.1109/66.4371.

[15] R. Singh, “Experimental investigation for performance assessment of scheduling policies in semiconductor wafer fabrication — a simulation approach,” *Int. J. Adv. Manuf. Technol.*, pp. 1503–1520, 2018, DOI.10.1007/s00170-018-2414-y.

[16] R. Sandell and K. Srinivasan, “Evaluation of lot release policies for semiconductor manufacturing systems,” in *Winter Simulation Conference Proceedings*, 1996, pp. 1014–1022.

[17] R. Singh and M. Mathirajan, “Experimental investigation for performance assessment of scheduling policies in semiconductor wafer fabrication — a simulation approach,” *Int. J. Adv. Manuf. Technol.*, pp. 1503–1520, 2018, DOI.10.1007/s00170-018-2414-y.

[18] C. Qi, A. I. Sivakumar, and S. B. Gershwin, “Impact of production control and system factors in semiconductor wafer fabrication,” *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 3, pp. 376–389, 2008, DOI.10.1109/TSM.2008.2001214.

[19] C. Qi, A. I. Sivakumar, and S. B. Gershwin, “An efficient new job release control methodology,” *Int. J. Prod. Res.*, vol. 47, no. 3, pp. 703–731, 2009, DOI.10.1080/00207540701455335.

[20] Y. Li, Z. Jiang, and W. Jia, “An integrated release and dispatch policy for semiconductor wafer fabrication,” *Int. J. Prod. Res.*, vol. 52, no. 8, pp. 2275–2292, 2014, DOI.10.1080/00207543.2013.854938.

[21] Z. B. Chen, X. W. Pan, L. Li, Q. J. Chen, and W. S. Xu, “A new release control policy (WRELM) for semiconductor wafer fabrication facilities,” *Proc. 11th IEEE Int. Conf. Networking, Sens. Control. ICNSC 2014*, pp. 64–68, 2014, DOI.10.1109/ICNSC.2014.6819601.

[22] R. Sun and Z. Wang, “DC-WIP - a new release rule of multi-orders for semiconductor manufacturing lines,” *7th Int. Conf. Syst. Simul. Sci. Comput. ICSC 2008*, pp. 1395–1399, 2008, DOI.10.1109/ASC-ICSC.2008.4675591.

[23] N. Xiang and L. Li, “A Release Control Policy Based on the Primary Bottleneck Workstation Utility of Multiple Semiconductor Manufacture Lines,” in *Advances in Computer Science Research, (ACSR)*, 2016, vol. 52, pp. 343–347.

[24] K. S. El-Kilany, “Wafer lot release policies based on the continuous and periodic review of WIP levels,” *IEEE Int. Conf. Ind. Eng. Eng. Manag.*, pp. 1700–1704, 2011, DOI.10.1109/IEEM.2011.6118206.

[25] S. Lou and P. W. Kager, “A Robust Production Control Policy for VLSI Wafer Fabrication,” *IEEE Trans. Semicond. Manuf.*, vol. 2, no. 4, pp. 159–164, 1989, DOI.10.1109/66.44620.

[26] M. L. Spearman, D. L. Woodruff, and W. J. Hopp, “CONWIP: A pull alternative to kanban,” *Int. J. Prod. Res.*, vol. 28, no. 5, pp. 879–894, 1990, DOI.10.1080/00207549008942761.

[27] S. Lou, H. Yan, S. Sethi, A. Gardel, and P. Deosthali, “Hub-centered production control of wafer fabrication,” in *IEEE/SEMI ASMC 1990*, 1990, pp. 27–32.

[28] J. Kim, R. C. Leachman, and B. Suh, “Dynamic release control policy for the semiconductor wafer fabrication lines,” *J. Oper. Res. Soc.*, vol. 47, no. 12, pp. 1516–1525, 1996, DOI.10.1057/jors.1996.195.

[29] Y. D. Kim, D. H. Lee, J. U. Kim, and H. K. Roh, “A Simulation Study on Lot Release Control, Mask Scheduling, and Batch Scheduling in Semiconductor Wafer Fabrication Facilities,” *J. Manuf. Syst.*, vol. 17, no. 2, pp. 107–117, 1998.

[30] O. Rose, “CONLOAD – A new lot release rule for semiconductor fabs,” *Proc. 1999 Winter Simul. Conf.*, vol. 31, no. 3, pp. 221–225, 1999.

[31] O. Rose, “CONWIP-like Lot Release for a Wafer Fabrication Facility with Dynamic Load Changes,” *Proc. SMOMS ’01*, no. April, pp. 41–46, 2001.

[32] S. H. Chung and C. M. Lai, “Job releasing and throughput planning for wafer fabrication under demand fluctuating make-to-stock environment,” *Int. J. Adv. Manuf. Technol.*, vol. 31, no. 3–4, pp. 316–327, 2006, DOI.10.1007/s00170-005-0185-8.

[33] Z. Zhou and O. Rose, “Wip balance and due date control in a wafer fab with low and high volume products,” in *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC)*, 2012, pp. 1–8.

[34] R. Singh and M. Mathirajan, “Investigation of different inputs and a new release policy in the proposed simulation model for wafer fabrication system,” *Sadhana - Acad. Proc. Eng. Sci.*, vol. 44, no. 2, pp. 1–19, 2019, DOI.10.1007/s12046-018-1006-8.

[35] L. Mönch, R. Uzsoy, and J. W. Fowler, “A survey of semiconductor supply chain models part I: semiconductor supply chains, strategic network design, and supply chain simulation,” *Int. J. Prod. Res.*, vol. 7543, pp. 1–20, 2017, DOI.10.1080/00207543.2017.1401234.

[36] L. Hendry, Y. Huang, and M. Stevenson, “Workload control: Successful implementation taking a contingency-based view of production planning and control,” *Int. J. Oper. Prod. Manag.*, vol. 33, no. 1, pp. 69–103, 2013, DOI.10.1108/01443571311288057.

[37] P. Henrich, M. Land, and G. Gaalman, “Exploring applicability of the workload control concept,” *Int. J. Prod. Econ.*, vol. 90, no. 2, pp. 187–198, 2004, DOI.10.1016/S0925-5273(03)00126-9.