Simulation Analysis of Segmented CONWIP: Application to Reentrant Flow Lines

Yassin M. Shaalan, Ingy A. El-Khouly, and Khaled S. El-Kilany
Department of Industrial and Management Engineering
Arab Academy for Science, Technology, and Maritime Transport
Alexandria, Egypt.
yassin.shaalan@aast.edu, ingyelkhouly@aast.edu, kkilany@aast.edu

Abstract

Reentrant flow lines is a special type of production flow lines where a job may visit a machine or group of machines more than once. This reentrancy characteristic results in higher variability of cycle time and throughput rates when compared to traditional production flow lines. Complex reentrant flow of jobs are closely related to semiconductor wafer fabrication due to layering of wafers, which is one of the most complex processes in semiconductors manufacturing. This paper presents a simulation study that seeks achieving a target throughput rate while improving cycle time using the segmented CONWIP lot release policy, which basically entails division of the production line into segments of similar total processing times and controlling the WIP level in each segment using CONWIP. A highly re-entrant and representative wafer fabrication facility found in literature is used for this study. Different scenarios that varies the number of segments, WIP level of each segment, and their combinations are tested using simulation. Results analysis shows that line segmentation using different CONWIP levels affects the line performance. In fact, in some instances, compared to a single segment CONWIP and push system, using segmented CONWIP can achieve the same throughput rate in a shorter cycle time.

Keywords

Lot release policy; modelling and simulation; re-entrant flow; segmented CONWIP

1. Introduction

Reentrant flow shop is considered a complex manufacturing environment because of the repetitive use of the same facilities by the same job, which leads to conflicts among jobs at some facilities at different stages in the process (Graves *et al.*, 1983). A direct application of reentrant flow shops is the photolithography operation, which is the main operation of wafer fabrication of semiconductor manufacturing that is considered one of the most difficult manufacturing environments. Photolithography operation is the photo expose step that is performed during the creation of each wafer layer leading to reentrant flow (Wein, 1988). Reentrant flow means that the job will visit a machine or group of machines more than once during processing due to layering of wafers (Zoghby, 2002).

Scheduling of semiconductors manufacturing lines is of a great importance as it achieves many benefits such as increasing throughput and maintaining low work-in-process (WIP) levels (Chung and Jang, 2009). Improvements obtained from input control policies are larger than those obtained by sequencing rules as was proved by (Wein, 1988). CONWIP is one of the lot input control policies (described in the next section) and is tested in this study. Dividing the production line into segments based on the critical tool groups such as bottleneck and high-loading tool groups has been studied by (Chien and Hu, 2006), in which a proposed WIP control mechanism at different check points has reduced queue time and cycle time.

Simulation is a common approach that is used for cycle time estimation and performance analysis of semiconductors manufacturing systems, it can also be modeled and adjusted for different experimental purposes (Shanthikumar, Ding and Zhang, 2007). The simulation in this study is based on the data of one fab version of a semiconductor wafer fab model presented by (Wein, 1988) that he used to evaluate different input control policies and scheduling rules. Two performance measures are evaluated in this study; the average cycle time (CT), and the average throughput rate (TH). The simulation experiments are divided into three sets. First, different PUSH scenarios are tested by varying the time between arrivals (TBA) of lots to the model. Second, CONWIP scenarios are under taken with different number of WIP levels tested. Finally, the third set studies the effect of Segmented CONWIP on the performance of the line, by varying the number of segments and the WIP levels of these segments.

The aim of this study is to examine the application of Segmented CONWIP lot release policy to a highly reentrant flow line. Specifically, the study addresses the method of segmentation that should be used, analyzes the number of segmentations that should be applied to the line, and setting of the constant WIP levels of each segment.

The remainder of this paper is organized as follows: in the next section, a brief review of literature related to this work is given followed by Section 3 which provides a detailed description of the case study. Section 4 describes the simulation model development. Then, the different experiments, their results and analysis are provided in Section 5. Finally, the conclusions drawn from this work are pointed out in Section 6.

2. Literature Review

Production control systems with multiple stages are classified into push and pull systems (Kimura and Terada, 1981). (Spearman, Woodruff and Hopp, 1990) already established the fact that pull systems are superior systems to control production lines when compared to push. An important difference is that there is no limit for WIP level within the production line of a push system (Enns and Rogers, 2008); while, in a pull system the production of the upstream stage is triggered by consumption of downstream stages (Kang, 1996). Of course, WIP levels directly affects the throughput rate and cycle time of a line. Too much WIP would result in high throughput rates however it would probably result in longer cycle times as well; and vice versa.

WIP levels, for both push and pull systems, are generally controlled by the lot release policies applied to the production line, which are concerned with when to release a new lot into a production line. Such policies can be generally classified into two categories; static and dynamic. Static release strategies are those where the lots arrival rate is pre-determined, fixed, and follows a certain distribution e.g. deterministic or Poisson (Shi, Zhang and Li, 2008), this is basically a push system; while, CONWIP as a pull system beside other control policies are said to be dynamic ones (Chen *et al.*, 2014). One of the difficulties of applying pull system is to set the WIP level, which when set correctly, can achieve higher throughput rate with same WIP levels, lower variability of cycle times, and stable throughput behavior when compared to push systems (Sturm *et al.*, 1999).

2.1 CONWIP

CONWIP is a pull production system which was introduced and described by (Spearman, Woodruff and Hopp, 1990). It can be viewed as a lot release decision that rely on WIP as a performance indicator (Shanthikumar, Ding and Zhang, 2007). Generally speaking, CONWIP outperforms other push and pull systems in achieving higher throughput rate, lower cycle time and lower WIP. It is also easier to implement and use in complex environments (Jaegler *et al.*, 2017); however, that doesn’t ignore the fact that push models with low coefficient of variation of inter-arrivals can outperform CONWIP, as proven by simulation experiments by (Enns and Rogers, 2008). In addition, it is not easily applied to complex manufacturing environments like semiconductor manufacturing and should be modified to accommodate the complexities of such manufacturing environments (Kalisch, Ringel and Weigang, 2008).

Although among the advantages of CONWIP, its simplicity and robustness in dealing with variable processing times and changing bottlenecks (Enns and Rogers, 2008); however, it is limited in dealing with the distribution of jobs along the line at the time of job release (Chao and Sivakumar, 2006). Also, the determination of the threshold WIP level is not simple; where extensive simulation experiments and analysis are usually required (El-Khouly, El-Kilany and El-Sayed, 2009; El-Khouly *et al.*, 2011).

CONWIP variants were reported in literature in an attempt to further improve the performance of this lot release policy. As an example, (Belisario, Azouz and Pierreval, 2015) studied the effect of varying WIP levels of CONWIP on the performance of the production system. Simulation optimization experiments by (Pierreval *et al.*, 2013) pointed out the significance of dynamic WIP levels as well. (El-Kilany, 2011) presented a continuous and a periodic review policy for review of WIP levels set using a CONWIP lot release policy. (Prakash and Chin, 2014) reviewed 15 variants of CONWIP that overcome limitations of CONWIP in dealing with complex and realistic systems, e.g. the determination of suitable WIP levels, the unbalance of WIP levels across multiple stages, and the response of CONWIP to sudden changes in production capacity and demand. (Prakash and Chin, 2014) also noted that these modified systems dominate simulation studies and are mostly applied to the semiconductor industry. Finally, segmented CONWIP (discussed next) was presented as one of the multi-control mechanisms that promises achievement of highest throughput levels.

2.2 Segmented CONWIP

An early study of segmentation of a serial Kanban production line was made by (Tayur, 1992). In his work the allocation of cards to cells was studied and how the partitioning of the line into cells can achieve a desired throughput with as few number of cards as possible.

A segmented WIP control of a wafer fab in Taiwan, presented by (Chien and Hu, 2006) to study the relation between WIP level and cycle time. However, in their study the reentrant flow was not considered and; hence, the segmentation of the production route was based on the critical tool groups such as bottleneck and high-loading tool groups. Results of their simulations showed that segmentation can reduce both cycle time and WIP level.

It should be noted that Segmented CONWIP lot release policy can be found in literature under other names such as multi-CONWIP as proposed by (Yang, Fu and Yang, 2007) and multi-loop CONWIP proposed by (Eng and Sin, 2013); where both proposed a strategy which divides the line into segments and proved it outperforms the single loop CONWIP as it decreased cycle time. Although both studies were applied to the semiconductor manufacturing; yet, one study was applied to a packaging line and the other was applied to an end of line assembly. Hence, reentrant flow which is a characteristic of wafer fabs was not considered.

3. TRC Fab Model Description

A representative fab presented in (Wein, 1988) is the case study of this work. Most of the data of this model is derived from real data gathered at an actual facility. This facility is the Hewlett-Packard Technology Research Center Silicon fab (referred to as the TRC fab), which is a relatively large development laboratory in Palo Alto, California. Wein fab (TRC fab) features 172 processing steps performed at 24 stations. The structural, operational, and numerical data are detailed in this section, along with a basic capacity analysis of the fab understudy.

3.1 Structural Data

Wafers are grouped in lots entering the fab; with each lot containing 24 wafers. The fab operates two 12-hours shift a day, 7 days a week. The job shop contains 24 single and multi-server stations with identical machines in multi-server stations. A total of 38 machines are located at 24 stations as follows: Stations 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 17, 18, and 20 are single server stations with one machine each. There are 2 identical machines at stations 1, 2, 3, 16, 19, 21, 22, 23, and 24. Station 14 has 3 identical machines, and finally station 14 has 4 identical machines.

3.2 Operational and Numerical Data

The operational data presented here is mainly related to the routing of lots. As mentioned earlier, the fab under study features 172 processing steps performed at 24 stations; hence, each lot visits the same station more than once (reentrant flow). The process flow across the 24 stations is as presented in Figure 1. Table 1 lists the steps served by each station, which shows that some stations are visited 10 times or more as stations 1, 13, 14, 15, 16, 17, 22, and 23. Hence, it is clear that the process under study exhibits a sufficiently complex re-entrant nature.

Enter → 1 → 2 → 13 → 14 → 23 → 15 → 20 → 22 → 23 → 22 → 17 → 13 → 14 → 15 → 23 → 16 → 24 → 23 → 22 → 17 → 1 → 8 → 4 → 22 → 22 → 1 → 2 → 8 → 13 → 14 → 18 → 23 → 15 → 16 → 23 → 18 → 22 → 1 → 1 → 13 → 14 → 23 → 15 → 16 → 24 → 23 → 22 → 17 → 1 → 2 → 8 → 9 → 21 → 22 → 1 → 4 → 22 → 22 → 1 → 2 → 13 → 14 → 23 → 15 → 16 → 24 → 24 → 23 → 22 → 17 → 24 → 1 → 2 → 7 → 1 → 3 → 22 → 13 → 15 → 23 → 22 → 22 → 22 → 17 → 13 → 14 → 18 → 23 → 15 → 16 → 20 → 23 → 1 → 17 → 1 → 1 → 3 → 13 → 14 → 16 → 24 → 23 → 22 → 17 → 9 → 21 → 1 → 3 → 13 → 14 → 15 → 23 → 15 → 16 → 24 → 23 → 22 → 17 → 1 → 3 → 13 → 14 → 23 → 15 → 16 → 23 → 15 → 16 → 24 → 23 → 22 → 17 → 1 → 3 → 10 → 22 → 12 → 6 → 22 → 6 → 1 → 1 → 4 → 10 → 19 → 23 → 1 → 10 → 13 → 14 → 16 → 21 → 12 → 13 → 14 → 18 → 23 → 15 → 15 → 15 → 16 → 19 → 23 → 22 → 17 → 11 → 13 → 14 → 15 → 21 → 23 → 5 → Exit.

Figure 1: Flow across the 24 stations to complete a single lot.

Table 1: Details on routing of lots across the fab.

Station	*Steps Served per Station*
1	S1, S21, S26, S38, S39, S49, S55, S59, S72, S75, S93, S95, S96, S107, S119, S133, S141, S142, S147
2	S2, S27, S50, S60, S73
3	S76, S97, S108, S120, S134
4	S23, S56, S143
5	S172
6	S138, S140
7	S74
8	S22, S28, S51
9	S52, S105
10	S135, S144, S148
11	S166
12	S137, S153
13	S3, S12, S29, S40, S61, S78, S85, S98, S109, S121, S149, S154, S167
14	S4, S13, S30, S41, S62, S86, S99, S110, S122, S150, S155, S168
15	S6, S14, S33, S43, S64, S79, S89, S111, S113, S124, S127, S158, S159, S160, S169
16	S16, S34, S44, S65, S90, S100, S114, S125, S128, S151, S161
17	S11, S20, S48, S70, S84, S94, S104, S118, S132, S165
18	S31, S36, S87, S156
19	S145, S162
20	S7, S91
21	S53, S106, S152, S170
22	S8, S10, S19, S24, S25, S37, S47, S54, S57, S58, S69, S77, S81, S82, S83, S103, S117, S131, S136, S139, S164
23	S5, S9, S15, S18, S32, S35, S42, S46, S63, S68, S80, S88, S92, S102, S112, S116, S123, S126, S130, S146, S157, S163, S171
24	S17, S45, S66, S67, S71, S101, S115, S129

Numerical data for processing times, machine failures, and lots inter-arrival time presented by (Wein, 1988) are all stochastic. The processing times for a lot of wafers are stochastic as it includes setups, operator unavailability, and rework. (Wein, 1988) used gamma distribution with shape parameter equal to two and Coefficient of Variation (CV) equal to 0.707; however, when these parameters were tested in this work, different results for the mean processing times were reported. For example, when using the gamma distribution with shape parameter of two, the mean processing time generated for station 14 (the bottleneck station) was equal to 15.75 hours, which is quite different from the mean processing time reported in (Wein, 1988), which was 7.82 hours. Hence, all machines mean processing times are modelled using a lognormal distribution, which when used in the simulation model, resulted in more accurate mean processing time and with almost the same CV.

Although, machine failures are mostly non-preemptive such as routine maintenance and machines adjustment; however, some unscheduled breakdowns are included as well; hence, machine failures are represented as another stochastic variable. (Wein, 1988) used the mean time between failures (MTBF) and mean time to repair (MTTR) to model such failures and fitted the data extracted from the TRC fab to a gamma distribution with shape factor of one half. Again, using the gamma distribution resulted in different means for both MTBF and MTTR. For this reason, exponential distribution with mean values reported for the MTBF and MTTR were used instead as it results in closer mean values when compared to reported values from the gamma distribution; however, with a lower CV.

Finally, the arrival rate of lots used in (Wein, 1988) had a mean of 0.0236 lots/hour following a Poisson distribution, which means a time between arrivals of 42.37 hours, in this work an exponential distribution is used instead.

3.3 Basic Capacity Analysis

A basic capacity analysis is made for the production line according to the operation parameters listed in Table 2. The station mean production rate (r^*) is calculated such that, $r^* = m/MPT$, where m is the number of machines per station and *MPT* is the mean processing time per step. Availability (*A*) is calculated based on *MTBF* and *MTTR*, where *A* = *MTBF*/(*MTBF* + *MTTR*). Effective production rate (r_e) is obtained by multiplying the mean production rate by the availability and rate of arrivals of lot per each station (r_a) is determined by multiplying the number of visits per lot (*NV*/*L*) by the input arrival rate which is 0.0236 lots/hour as mentioned earlier. Finally, the percentage utilization is calculated from the relation $U = r_a / r_e$. It is clear from the reported utilizations in the table that station 14 has the highest utilization of 90.14% and; hence, is the bottleneck of the fab.

Table 2. Basic Capacity Analysis

Station	*Mean Processing Time (hr/lot)*	*No. of Machines per Station*	*Number of Visits per Lot*	*Mean Time Between Failures*	*Mean Time To Repair*	*Availability (%)*	*Mean Production Rate (lot/hr)*	*Effective Production Rate (lot/hr)*	*Rate of Arrivals (lot/hr)*	*Utilization (%)*
	MPT	*m*	*NV/L*	*MTBF*	*MTTR*	*A*	r^*	r_e	r_a	*U*
1	1.55	2	19	42.18	2.22	95.00	1.29	1.23	0.45	36.58
2	4.98	2	5	101.11	10.00	91.00	0.40	0.37	0.12	32.29
3	5.45	2	5	113.25	5.21	95.60	0.37	0.35	0.12	33.63
4	4.68	1	3	103.74	12.56	89.20	0.21	0.19	0.07	37.15
5	6.14	1	1	100.55	6.99	93.50	0.16	0.15	0.02	15.50
6	7.76	1	2	113.25	5.21	95.60	0.13	0.12	0.05	38.31
7	6.23	1	1	16.78	4.38	79.30	0.16	0.13	0.02	18.54
8	4.35	1	3	13.22	3.43	79.40	0.23	0.18	0.07	38.79
9	4.71	1	2	10.59	3.74	73.90	0.21	0.16	0.05	30.08
10	4.05	1	3	47.53	12.71	78.90	0.25	0.19	0.07	36.34
11	7.86	1	1	52.67	19.78	72.70	0.13	0.09	0.02	25.52
12	6.10	1	2	72.57	9.43	88.50	0.16	0.15	0.05	32.53
13	4.23	4	13	22.37	1.15	95.11	0.95	0.90	0.31	34.11
14	7.82	3	12	21.76	4.81	81.90	0.38	0.31	0.28	90.14
15	0.87	1	15	387.20	12.80	96.80	1.15	1.11	0.35	31.82
16	2.96	2	11	1.00	0.00	100.00	0.68	0.68	0.26	38.42
17	1.56	1	10	119.20	1.57	98.70	0.64	0.63	0.24	37.30
18	3.59	1	4	1.00	0.00	100.00	0.28	0.28	0.09	33.89
19	13.88	2	2	46.38	17.42	72.70	0.14	0.10	0.05	45.06
20	5.41	1	2	36.58	9.49	79.40	0.18	0.15	0.05	32.16
21	7.58	2	4	36.58	9.49	79.40	0.26	0.21	0.09	45.06
22	1.04	2	21	118.92	1.08	99.10	1.92	1.91	0.50	26.01
23	1.09	2	23	1.00	0.00	100.00	1.83	1.83	0.54	29.58
24	3.86	2	8	55.18	12.86	81.10	0.52	0.42	0.19	44.93

4. Simulation Models Development

The simulation models of the TRC fab have been developed using the ExtendSim™ Suite v9.0.2 simulation environment from ImagineThat, Inc. These models are capable of running with different input lot release policies, and have the flexibility of varying the controlling parameter for each policy (such as inter-arrival time for PUSH and the constant WIP level for CONWIP) without the need to change the model itself.

4.1 Simulation Parameters

The different simulation parameters that are set for the experiments presented in the next section are the simulation runtime, number of replications, and the warm-up period. A single simulation run covers a time period of 4.5 years (39,420 hours) and the number of replications for each experiment is 20 replications, with the first year removed for simulation model warm-up period. This means that a total of 613,200 hours of data is used for each experiment to estimate the performance of the fab.

4.2 Experimental Design

Three sets of simulation experiments are carried out:

SET I. PUSH experiments: A number of PUSH scenarios are tested, each scenario has a different input value for time between arrivals (TBA) of lots, which follows an exponential distribution.

SET II. CONWIP experiments: A number of CONWIP scenarios are examined, each scenario has a different WIP level.

SET III. Segmented CONWIP experiments: This set of experiments is divided in to five groups, for each group the fab is divided into segments, with a minimum of 2 segments and a maximum of 6 segments. Every group has a number of scenarios tested, which is essentially varying the WIP levels of each segment.

The objective of the first set is to determine the target TH the fab should achieve. The second set of experiments are then used to determine the CT that be achieved while meeting the target TH; however, using the CONWIP release policy. Finally, the third set of experiments aims at achieving the target TH of PUSH while achieving a CT that is lower than that of the CONWIP; through the application of the segmented CONWIP lot release policy.

4.3 Performance Measures

The performance measures that are evaluated are the average cycle time (CT) in hours, and average throughput rate (TH) in lots/hour. Apparently, varying the WIP levels of the entire fab in the second of experiments or the WIP levels of each segment in the third set will result in a conflicting effect on CT and TH; where, lowering the WIP level reduces CT; yet, it decreases TH as well. On the other hand, increasing the WIP level increases TH; but, it increases CT as well. Thus, the objective is to improve the performance of the fab in terms of CT while achieving a target TH level.

5. Experimentations, Results and Analysis

As mentioned earlier, the ExtendSim^{TM} simulation environment is used to develop all models in this work. This section reports the results obtained from these models and their analysis. First, the results of the PUSH and CONWIP are presented (experiment sets I and II); target TH and CT are determined. Afterwards, results of the segmented CONWIP experiments are analyzed and the fab performance under this strategy is compared against the target TH and CT.

5.1 PUSH and CONWIP Experiments

To determine the target TH using the PUSH simulation model (Experiment Set I), TBA of lots is varied from pushing a lot every 37 hours to every 70 hours in increments of 1. TBA of 42 is selected as it results in the nearest value of TH and CT to that found in (Wein, 1988), as shown in Table 3, and a target TH of 0.023 lot/hr is set.

Table 3. Results of PUSH and Wein model.

	Input values		Reported values	
Model	TBA (hrs)	Distribution	TH (lot/hr)	CT (hrs)
PUSH	42	Exponential	0.02300	1,020.45
Wein	42.3728	Poisson	0.02279	999.10

Then, CONWIP as a lot release policy (Experiment Set II) is applied to the fab. CONWIP level across the fab is varied from 11 to 35 lots. At CONWIP level 21, the TH reported is 0.02309 lot/hr, which means that the target is achieved and at a lower CT (908.7 hours) when compared to that of PUSH (1,020.45 hours). Reported performance measures of all PUSH and CONWIP experiments are presented in Figure 2, which clearly shows that CONWIP consistently outperforms PUSH with respect to both TH and CT.

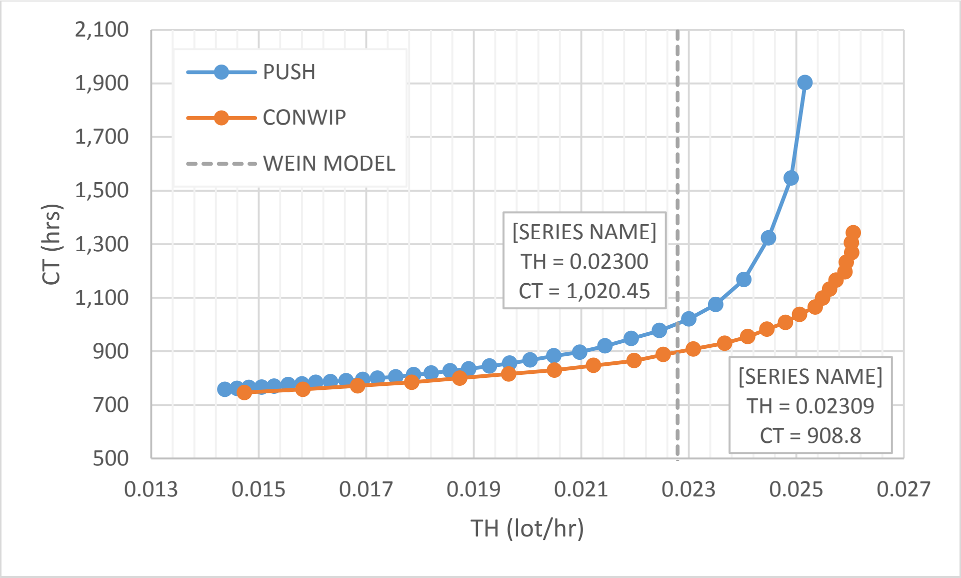

Figure 2: Operating curves of PUSH and CONWIP experiments.

Comparing the results reported from the selected PUSH and CONWIP scenarios, it is shown in Table 4 that CONWIP improved the performance of the fab, it achieved the target TH with 10.94% CT improvement. The objective now is achieving the target throughput rate reported by the PUSH at a shorter CT than that of the CONWIP.

Table 4: PUSH and CONWIP results.

Experiment	TH (lot/hr)	CT (hrs)
PUSH	0.02300	1020.4
CONWIP	0.02309	908.8
Percentage improvement in CT (%)		10.94

Further investigation is undertaken to analyze the improvement achieved by CONWIP. This analysis looks at the performance of the queues at all stations under PUSH and CONWIP lot release policies. Figure 3 compares the PUSH and CONWIP based on the average and maximum number of lots waiting in queues at all stations.

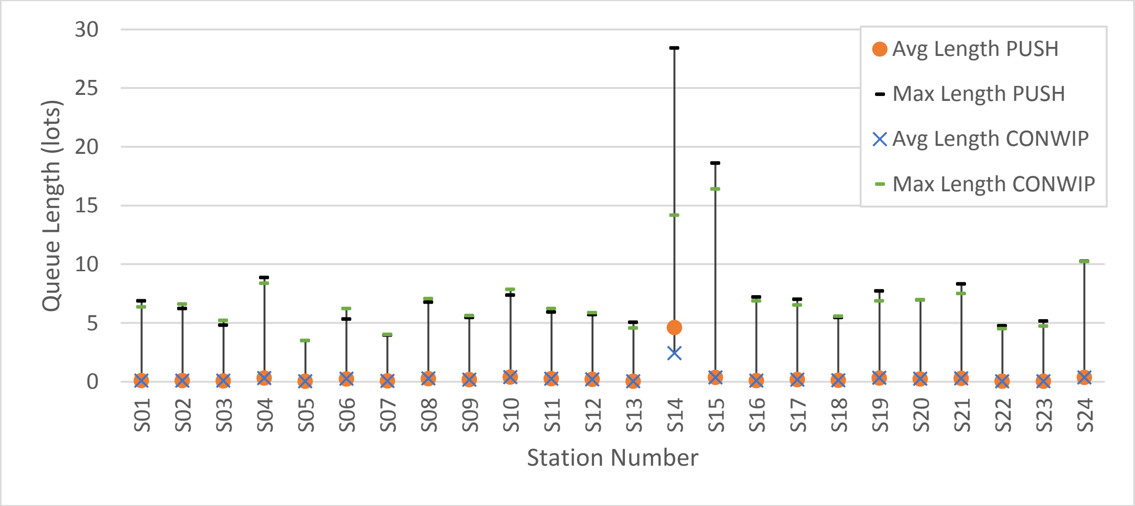

Figure 3: Queue length performance at all stations.

The figure shows that all stations have the same average values of queue length with slight differences in the maximum values of both measures. The only significant difference is found at Station 14 (the bottleneck station);

where, CONWIP has noticeable lower maximum and average values of queue length when compared to PUSH. The same performance applies to the queue waiting times at all stations. Reduction of queue length and queue waiting times directly affects the fab CT and means achieving the target TH with a lower WIP level. Another observation drawn from the chart in Figure 3 is that when using CONWIP, the WIP level is not distributed equally over the different stations, which means that except for the bottleneck station, other stations may be starved occasionally, which can degrade the line performance in terms of TH.

Therefore, due to the aforementioned observations, it is suggested to divide the fab to a number of segments, each segment having its own CONWIP level and its own bottleneck, which would possibly lead to improvements in queue performance and consequently improvements in CT. Furthermore, dividing the line into segments with fixed WIP levels can also help in better distribution of the WIP along the line and; hence, would lead to better utilization of all stations leading to overall improvements in TH as well.

5.2 Segmented CONWIP Experiments

The third and final set of experiments evaluates the fab TH and CT using the segmented CONWIP lot release control policy. The production line is divided into segments and each segment has its own constant WIP level. To complete one lot 172 steps are processed consuming a total MPT of 549.3 hrs. Due to the re-entrant flow nature of the fab understudy, segmentation of the fab is based on the total mean processing time; where, the total MPT is divided by the number of segments tested so that each segment should have similar total MPT. As mentioned earlier, this set of experiments is divided into 5 groups of experiments. Each group simulates the fab being divided into 2, 3, 4, 5, and 6 Segments. The mean processing time of each segment is given in Table 5 and steps performed in each segment are as presented in Figure 4.

Table 5: Total mean processing time of each segment in hours.

		Line Segmentation				
		2 Segments	3 Segments	4 Segments	5 Segments	6 Segments
Segment	1	268.87	175.92	134.19	106.57	82.97
	2	280.43	186.65	134.68	103.89	92.95
	3		186.73	136.12	109.94	92.95
	4			144.31	104.18	93.70
	5				124.72	86.81
	6					99.92
Average		274.65	183.1	137.325	109.86	91.55

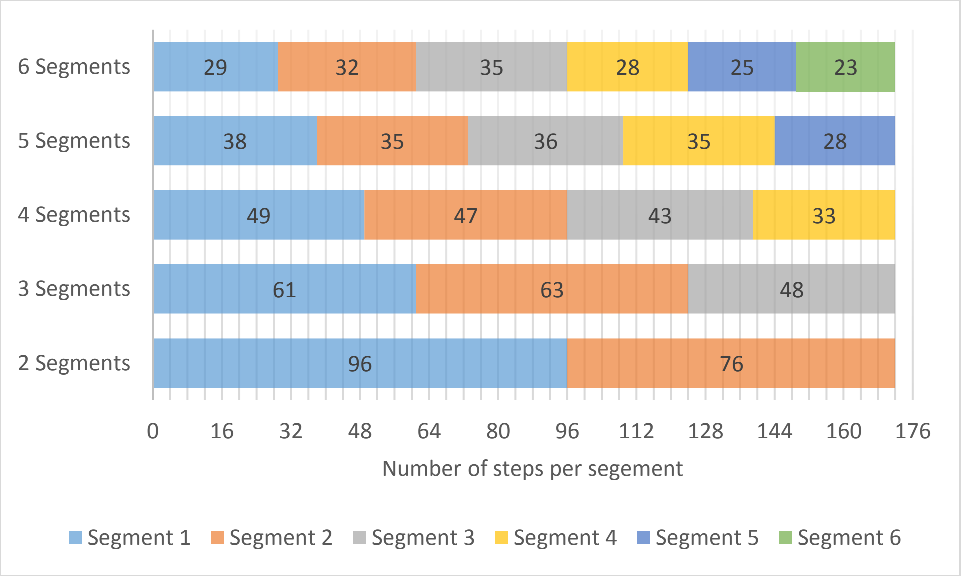

Figure 4: Number of steps per segment for all sub-groups of experiments.

For each experiments group (2-6 segments) different WIP level for each segment is tested. A total number of 1,924 scenarios are simulated to evaluate the performance of the fab when Segmented CONWIP lot release control strategy is applied. The detailed information about the minimum and a maximum WIP level applied for each segment is given in Table 6. The ExtendSim^{TM} Scenario Manager is used to generate the different scenarios so that all possible combinations of WIP levels at each segment is tested and it also manages the execution of the simulation and reporting the TH and CT for each experiment according to the simulation parameters mentioned in the previous section. Table 6 shows the number of scenarios generated and tested for each group of experiments.

Table 6: Limits of WIP level and number of scenarios carried out for all sub-groups.

		WIP level (Min-Max)				
		2 Segments	3 Segments	4 Segments	5 Segments	6 Segments
Segment	1	03-18	02-10	04-07	04-05	03-06
	2	03-18	02-10	04-07	04-07	03-07
	3		02-10	04-07	04-07	03-06
	4			04-07	04-07	03-07
	5				04-07	03-06
	6					03-07
No. of scenarios		244	368	256	393	663

After running all the above mentioned scenarios, a *best scenario* from every group is selected. A best scenario is the one that, based on the setting of the WIP level for each segment, is capable of achieving the target TH (0.023 lot/hr) at a minimum CT, which is then compared to the CT reported by CONWIP 21 (908.7 hrs) to determine whether or not CT has improved.

WIP levels for the best scenarios of each experiments group are presented in Table 7 along with the CT and the CT improvement percentage. The table shows that all segmentations tested are capable of achieving the target TH; however, tested levels of WIP per segment improved the CT in three out of five line segmentations only (2, 4 and 5 segments). Compared to CONWIP, the highest reported CT improvement (3.23%) is achieved when dividing the fab into 5 segments; followed by, CT reported for the fab divided into 2 and 4 segments, with an improvement of 1.28% and 2.98%; respectively.

Table 7: Improvement in CT of selected scenarios for all line segmentations.

		WIP levels of best scenario selected				
		2 Segments	3 Segments	4 Segments	5 Segments	6 Segments
Segment	1	10	7	5	4	4
	2	14	8	7	5	4
	3		8	7	7	6
	4			7	7	6
	5				7	6
	6					6
CT (hrs)		897.81	908.78	881.64	**879.33**	925.58
CT Improvement		1.21%	-0.01%	2.98%	**3.23%**	-1.86%

Further investigation is undertaken for all scenarios tested on the fab with 5 segments. VALID/INVALID TH and CT values reported from the fab running the 5 segments are defined in Table 8.

Table 8: Defining VALID/INVALID values of TH and CT.

		Throughput Rate	
		VALID	INVALID
Cycle Time	VALID	TH ≥ 0.023 lot/hr CT < 908.7 hrs	TH ≤ 0.023 lot/hr CT < 908.7 hrs
	INVALID	TH ≥ 0.023 lot/hr CT > 908.7 hrs	TH ≤ 0.023 lot/hr CT > 908.7 hrs

Out of the 393 scenarios tested; 259 scenarios are not able to achieve the target TH (INVALID TH); 208 scenarios of them are with longer CT (INVALID CT) and 51 scenarios of them with shorter CT (VALID CT). However, 134 scenarios are able to achieve the target TH (VALID TH); 68 scenarios of them with longer CT (INVALID CT) and 66 scenarios of them with shorter CT (VALID CT), and that is the improvement region of this sub-group.

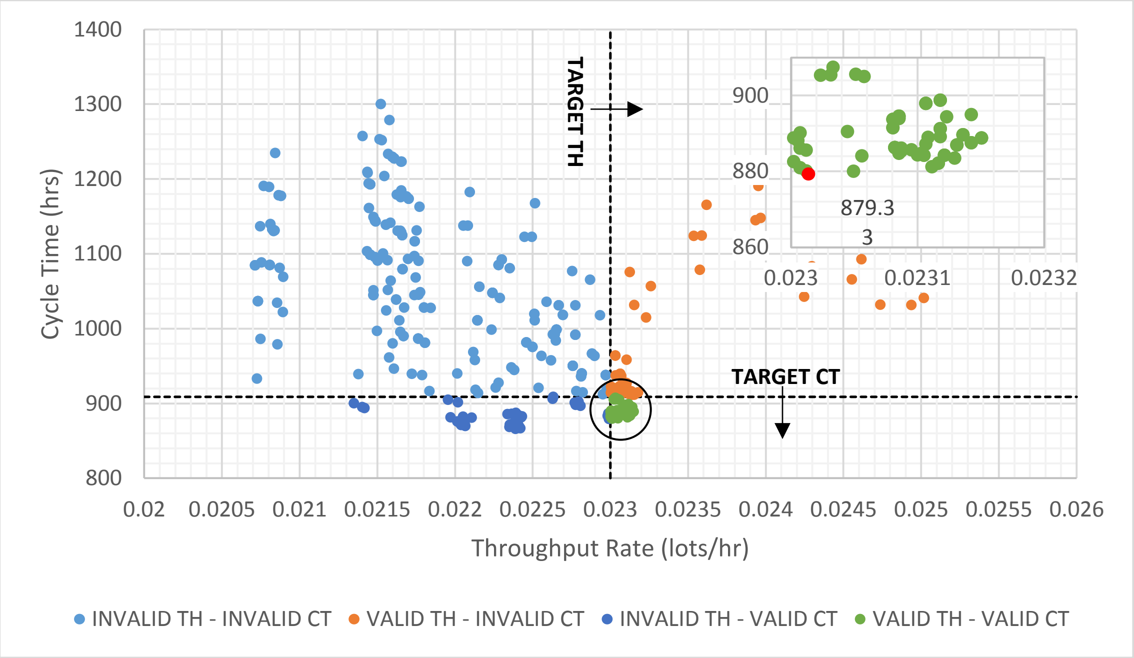

Figure 5: TH and CT reported of all 5 segments scenarios.

These are the four regions easily identified in Figure 5; where, 66 scenarios out of the 393 scenarios tested (16.79 %) of this segmented CONWIP group (green data series shown and magnified in the top-right corner of the chart) can outperform CONWIP and the minimum CT that can be achieved is 879.33 hours.

6. Conclusions

Three sets of simulation experiments have been conducted to test the performance of a highly reentrant wafer fab previously studied by Wein. The focus of this study has been directed towards applying segmented CONWIP to a reentrant flow shop and investigating its ability to further improve the CONWIP lot release policy. Segmented CONWIP is an approach that depends on dividing a production line into segments. When it comes to reentrant flow shop like the fab under study, segmentation of the fab is by processing step and not by stations; where, all segments should have similar total processing time.

Simulation results have shown that CONWIP experiments improved CT compared to PUSH. Segmented CONWIP with two, four and five segments resulted in improvements of CT compared to CONWIP while achieving the target TH set using the PUSH experiments. The three and six segments models achieved the target TH; however, it did not result in any improvements in terms of CT.

From this study, it was evident that the setting of the WIP level of each segment beside the number of segments used to divide the line has direct impact on both TH and CT. Yet, finding the right combination of values for these parameters are not easily achieved. Further analysis of this lot release policy is still ongoing; where more values for the WIP level of each segment and dividing the line into larger number of segment are under study.

Acknowledgements

The models developed for this work are built using the ExtendSim^{TM} Suite v9.0.2 simulation environment from ImagineThat, Inc. The tool has been offered to the Department of Industrial and Management Engineering, AASTMT, as a grant for teaching and research purposes as part of the ExtendSim Adopter Program.

References

Belisario, L. S., Azouz, N. and Pierreval, H. (2015) 'Adaptive ConWIP: Analyzing the impact of changing the number of cards', in *2015 International Conference on Industrial Engineering and Systems Management (IESM)*. IEEE, pp. 930–937. doi: 10.1109/IESM.2015.7380266.

Chao, Q. and Sivakumar, A. I. (2006) 'Job release based on WIPLOAD control in semiconductor wafer fabrication', *Proceedings of the Electronic Packaging Technology Conference, EPTC*, (2), pp. 665–670. doi: 10.1109/EPTC.2006.342793.

Chen, Z. B. *et al.* (2014) 'A new release control policy (WRELM) for semiconductor wafer fabrication facilities', *Proceedings of the 11th IEEE International Conference on Networking, Sensing and Control, ICNSC 2014*, pp. 64–68. doi: 10.1109/ICNSC.2014.6819601.

Chien, C.-F. and Hu, C.-H. (2006) 'Segmented WIP Control for Cycle Time Reduction', in *2006 IEEE International Symposium on Semiconductor Manufacturing*. IEEE, pp. 265–268. doi: 10.1109/ISSM.2006.4493079.

Chung, J. and Jang, J. (2009) 'A WIP Balancing Procedure for Throughput Maximization in Semiconductor Fabrication', *IEEE Transactions on Semiconductor Manufacturing*, 22(3), pp. 381–390. doi: 10.1109/TSM.2009.2017666.

El-Khouly, I. A. *et al.* (2011) 'A comparison of two different approaches to multi-criteria optimisation of semiconductor fabrication', in *9th International Industrial Simulation Conference 2011, ISC 2011*, pp. 144–149. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84898401907&partnerID=tZOtx3y1.

El-Khouly, I. A., El-Kilany, K. S. and El-Sayed, A. E. (2009) 'Modelling and simulation of re-entrant flow shop scheduling: An application in semiconductor manufacturing', in *2009 International Conference on Computers & Industrial Engineering*. IEEE, pp. 211–216. doi: 10.1109/ICCIE.2009.5223754.

El-Kilany, K. S. (2011) 'Wafer lot release policies based on the continuous and periodic review of WIP levels', in *2011 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, pp. 1700–1704. doi: 10.1109/IEEM.2011.6118206.

Eng, C. K. and Sin, L. K. (2013) 'CONWIP based control of a semiconductor end of line assembly', *Procedia Engineering*. Elsevier B.V., 53, pp. 607–615. doi: 10.1016/j.proeng.2013.02.078.

Enns, S. T. and Rogers, P. (2008) 'Clarifying CONWIP versus push system behavior using simulation', in *2008 Winter Simulation Conference*. IEEE, pp. 1867–1872. doi: 10.1109/WSC.2008.4736277.

Graves, S. C. *et al.* (1983) 'Scheduling of re-entrant flow shops', *Journal of Operations Management*, 3(4), pp. 197–207. doi: 10.1016/0272-6963(83)90004-9.

Jaegler, Y. *et al.* (2017) 'The ConWip production control system: a systematic review and classification', *International Journal of Production Research*. Taylor & Francis, 7543(October), pp. 1–22. doi: 10.1080/00207543.2017.1380325.

Kalisch, S., Ringel, R. and Weigang, J. (2008) 'Managing WIP and cycle time with the help of Loop Control', in *2008 Winter Simulation Conference*. IEEE, pp. 2298–2304. doi: 10.1109/WSC.2008.4736334.

Kang, J. (1996) *A Method for Target Scheduling of Semiconductor Wafer Fabrication Based on Event-Based Optimization Modeling and Discrete Event Simulation*. University of California, Berkeley. doi: 10.16953/deusbed.74839.

Kimura, O. and Terada, H. (1981) 'Design and analysis of Pull System, a method of multi-staǵe production control', *International Journal of Production Research*, 19(3), pp. 241–253. doi: 10.1080/00207548108956651.

Pierreval, E. *et al.* (2013) 'A Simulation Optimization Approach for Reactive ConWIP Systems', in *2013 8th EUROSIM Congress on Modelling and Simulation*. IEEE, pp. 415–420. doi: 10.1109/EUROSIM.2013.78.

Prakash, J. and Chin, J. F. (2014) 'Modified CONWIP systems: A review and classification', *Production Planning and Control*. Taylor & Francis, 26(4), pp. 296–307. doi: 10.1080/09537287.2014.898345.

Shanthikumar, J. G., Ding, S. and Zhang, M. T. (2007) 'Queueing Theory for Semiconductor Manufacturing Systems: A Survey and Open Problems', *IEEE Transactions on Automation Science and Engineering*, 4(4), pp. 513–522. doi: 10.1109/TASE.2007.906348.

Shi, L., Zhang, X. and Li, L. (2008) 'Simulation and analysis of scheduling rules for semiconductor manufacturing line', in *2008 IEEE International Conference on Industrial Technology*. IEEE, pp. 1–5. doi: 10.1109/ICIT.2008.4608675.

Spearman, M., Woodruff, D. and Hopp, W. (1990) 'CONWIP: a pull alternative to kanban', *International Journal of Production Research*, 28(5), pp. 879–894. doi: 10.1080/00207549008942761.

Sturm, R. *et al.* (1999) 'Advanced WIP control for make-to-order wafer fabrication', in *10th Annual IEEE/SEMI. Advanced Semiconductor Manufacturing Conference and Workshop. ASMC 99 Proceedings (Cat. No.99CH36295)*. IEEE, pp. 31–36. doi: 10.1109/ASMC.1999.798176.

Tayur, S. R. (1992) 'Properties of serial kanban systems', *Queueing Systems*, 12(3–4), pp. 297–318. doi:

10.1007/BF01158805.

Wein, L. M. (1988) 'Scheduling semiconductor wafer fabrication', *IEEE Transactions on Semiconductor Manufacturing*, 1(3), pp. 115–130. doi: 10.1109/66.4384.

Yang, T., Fu, H. P. and Yang, K. Y. (2007) 'An evolutionary-simulation approach for the optimization of multi-constant work-in-process strategy-A case study', *International Journal of Production Economics*, 107(1), pp. 104–114. doi: 10.1016/j.ijpe.2006.02.014.

Zoghby, J. M. (2002) *Critical Arc Strategies for the Reentrant Job Shop Scheduling Problem with Setups*. The University of Texas at Austin.

Biographies

Yassin M. Shaalan is a M.Sc. student at the Department of Industrial and Management Engineering, Arab Academy for Science, Technology, and Maritime Transport (AASTMT). Shaalan received his BSc. In Industrial and Management Engineering from AASTMT in 2005, he used simulation using Extend in his graduation project to improve the performance of a food production plant by modifying the facility layout. He joined the Industry Service Complex (ISC) inside the AASTMT campus in 2007. He worked as supervisor of the AASTMT industrial educational workshops from 2008 to 2010. He is now the coordinator of R&D projects inside the ISC, where he supervises the implementation of engineering undergraduate projects and other R&D projects inside the workshops. In addition he is a vocational trainer and a Pearson HND level 5 (Mechanical Program) lecturer, teaching Materials Engineering and Mechanical Workshop Practices since 2013 and up to date.

Ingy A. El-Khouly is an Assistant Professor at the Department of Industrial and Management Engineering at the Arab Academy for Science, Technology & Maritime Transport (AASTMT), she joined the AASTMT in 2006. El-Khouly has received her Ph.D. in Mechanical and Manufacturing Engineering from Dublin City University (2015), Ireland; where her research work included the investigation of WIP management for lot flow control of a representative wafer fabrication facility, which is arguably the most technologically complex stage in semiconductor manufacturing, using modelling and simulation. Currently, her research interest's lies in production planning and scheduling, optimization using simulation, and development of component based simulation models.

Khaled S. El-Kilany is a Professor of Industrial Engineering at the Department of Industrial and Management Engineering at the AASTMT, which is accredited by the Engineering Accreditation Commission of ABET since 2010. The department offers both B.Sc. and M.Sc. degrees in Industrial and Management Engineering. Prof. El-Kilany is currently the head of department since February 2009. He is a senior member of the IISE and is a reviewer of several journal, conferences, and textbooks. He has received his Ph.D. in Mechanical and Manufacturing Engineering from Dublin City University, Ireland; where his research work included modeling and simulation of automated material handling system of Intel's wafer fabrication facility Fab24, which was the second wafer fabrication facility in the world that produces 300mm wafers. His research interests lies in the analysis and improvement of manufacturing systems performance; specifically, material flow, production planning and scheduling, and WIP management using discrete-event systems simulation and optimization techniques.