Article

Short-Term Scheduling Model of Cluster Tool in Wafer Fabrication

Ying-Mei Tu

Department of Industrial Management, Chung Hua University, Hsinchu 300, Taiwan; amytu@chu.edu.tw

Abstract: Since last decade, the cluster tool has been mainstream in modern semiconductor manufacturing factories. In general, the cluster tool occupies 60% to 70% of production machines for advanced technology factories. The most characteristic feature of this kind of equipment is to integrate the relevant processes into one single machine to reduce wafer transportation time and prevent wafer contaminations as well. Nevertheless, cluster tools also increase the difficulty of production planning significantly, particularly for shop floor control due to complicated machine configurations. The main objective of this study is to propose a short-term scheduling model. The noteworthy goal of scheduling is to maximize the throughput within time constraints. There are two modules included in this scheduling model—arrival time estimation and short-term scheduling. The concept of the dynamic cycle time of the product's step is applied to estimate the arrival time of the work in process (WIP) in front of machine. Furthermore, in order to avoid violating the time constraint of the WIP, an algorithm to calculate the latest time of the WIP to process on the machine is developed. Based on the latest process time of the WIP and the combination efficiency table, the production schedule of the cluster tools can be re-arranged to fulfill the production goal. The scheduling process will be renewed every three hours to make sure of the effectiveness and good performance of the schedule.

Keywords: wafer fabrication; cluster tool; short term scheduling; dynamic cycle time

1. Introduction

It is well known that semiconductor manufacturing is a cost intensive industry, particularly for modern 300 mm wafer manufacturing factories. Thus, productivity improvement is very important for factories. Cluster tools consist of several single-wafer processing chambers, for diverse semiconductor fabrication processes, shorter cycle time, faster process development, and better yield for less contamination [1]. Based on experiences over the past decade, this kind of equipment has been widely used for wafer fabrication processes and occupy over 60% capacity of modern wafer fabrication factories. Due to the complicated configuration of equipment, cluster tools also increase the difficulty of production planning and shop floor control significantly. Moreover, for some small and middle scale semiconductor factories, managers will consider combining un-relevant processes into one machine due to machine numbers and cost constraints. Thus, it will turn the difficulty of production planning and shop floor control into an arduous challenge.

In the small and middle scale semiconductor factories, without proper planning and control, the cluster tool will become the bottleneck machine and affect the overall output. However, because of the complicated configurations of the cluster tools, any slight relaxation will cause the loss of production capacity and fail to reach the expected output. This is particularly true for parallel machines, as processing one lot in multiple reaction chambers can shorten the product cycle time, but it will lose productivity simultaneously because processing one lot in only one reaction room results in the shortage of load port. In addition to the load port issue, there are certain layers on the dry etcher that need to be etched in two stages, with time constraints in between. Therefore, the question of how to make proper arrangements to obtain the maximum efficiency of the cluster tool is also a rather difficult problem for the shop floor control of the wafer fabrication factory.

As mentioned, the cluster tools consist of load ports for loading and unloading FOUPs, vacuum chambers (load locks), ATM and VTM robots to transfer wafer between load locks, process chambers and single-wafer process chambers (PM chamber) in a cluster tool as Figure 1, which operates as a small factory. Consequently, it reflects a general belief that the planning data such as capacity and cycle time are hard to estimate. Additionally, many special situations and difficulties also exist, such as multiple manufacturing processes, sharing components, unavailability of wafer temporary storage area between internal steps, moving on as soon as possible when the process is finished in the PM chamber and the association with different recipe combinations [2,3]. Moreover, regarding the productivity of the factory, many studies were focused on the scheduling of the cluster tool's internal process to improve the equipment productivity, which induced the scheduling algorithm to be imbedded into the equipment and hard to change. Based on the viewpoint of shop floor control, the major factor that affects the equipment productivity is the production combinations of the cluster. For this reason, the properly scheduling of the production sequence of the cluster tools is very essential and critical.
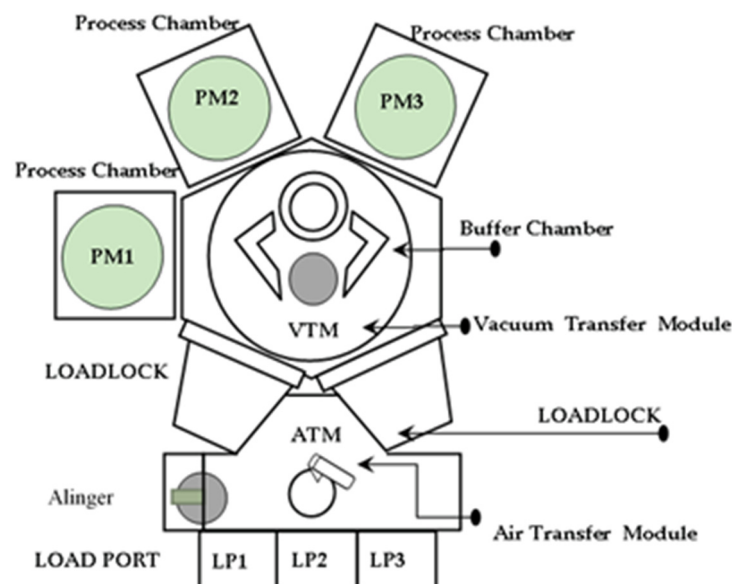

Figure 1. Cluster tool.

The attention of this study on these problems is aimed at proposing a short-term scheduling model of the cluster tool within the established goals and constraints. Because it is a short-term scheduling model, the quality and speed of scheduling are the two most important factors. In order to maximize the output, this research is not only considering the current WIP in front of the machine, but also calculates and adds the WIP that will reach the machine in the upcoming three hours. In addition, the WIP with the time constraint is calculated and marked at the latest time that must be processed on the machine, and then the efficiency table of the machine processing combination developed in the previous studies [4,5] is used to arrange each machine within the next three hours processing schedule to guarantee that the resulting schedule is the best schedule in the current situation. Under this procedure, the best short-term schedule of the cluster tool can be arranged accordingly.

2. Literature Review

The short-term scheduling of the cluster tool focuses on the processing sequence of the WIP based on the operation correlation of the upstream and downstream machines and the status of the WIP in front of the machines. There is no doubt that the quality of production scheduling is always closely related to the performance of the factory and there were a

lot of studies in this field in the past [6–9]. As stated above, the definition of scheduling is the start and end times at each job that will be accomplished on the machine under a given scheduling objective [10]. Nonetheless, most of the so-called "scheduling" in the wafer fabrication field literature is only in the stage of "dispatching", rather than the actual scheduling. Due to the complexity of the wafer fabrication process and the uncertainties in the factory, the dispatching is mostly adopted as the method of site scheduling even in practice. Fortunately, due to the evolvement of IT technique in recent years, the real scheduling is applied in some equipment of semiconductor fabrication.

Almost all the current Manufacturing Execution System (MES) are Real Time Dispatching (RTD), which dispatching the work pieces or the next step of the machine in semiconductor fabrication factory. There are some new architectures and logic theories applied in MES. Toba et al. [11] took the minimization of the sum of the handling time and the estimated processing time as the principle of the workpiece assignment, and its effect reduced the waiting time of the workpiece at each processing stage. Lee and Hsiao [12] proposed that taking the buffering cycle time as the basis for the operation of the immediate dispatching system, the buffering unit could balance the quantity of rework and shorten the production cycle time. Chang et al. [13] focused on two-phase and twin-fab environment to improve their RTD system. Based on the concepts of balancing the factory of the machine load degree and avoiding the happening of inter-regional transportation to reduce the frequency has no effect of handling of waiting time and product performance improvement. All these explanatory studies revealed the considerations of all related parameters of the machine load were only focused on the situation of preventing scheduling and scheduling of experimental group. Lee et al. [14] used two modules, the offline multi-objective scheduling system and on-line real time dispatch system to improve the performance of semiconductor manufacturing system. Obviously, because of the complicated flow path required, it is worth worrying whether the tandem degree of effect between two modules decreases with the increase of interval time. According to the above literatures, most of RTD (Real Time Dispatching) systems were only judged by the current system state at the time of execution. Apparently, there are still few mechanisms that include the mechanism of pre-considering the status of the subsequent period, and the improvement of production performance is still limited.

As the result of the high degree of automation in advanced fabs, the APC (Advanced Process Control) system has become an indispensable part of plant operations. In the past, the simple dispatching principle was sometimes ignored which leads the scheduling became the focus of rethinking in the wafer fab again, and the RTR (Run to Run) Control section of APC was added with short-term scheduling method. There have been many studies on the application of scheduling methods in wafer manufacturing, which can be roughly divided into batch machine scheduling, such as furnace tube machine [15–20]. Machine scheduling with auxiliary units, such as Scanner [21,22], or region-specific scheduling [23–26]. In these past studies, most of the theoretical methods used are MIP (Mixed Integer Programing), GA (Genetic Algorithm), dynamic Programing (Dynamic Programing) plus GA or Branch-bound Algorithm, and so on. Besides, numbers of studies have developed heuristic solutions. Nevertheless, no matter which method is adopted, it can be found that many restrictions and assumptions exist or some real phenomena exist in the field are disregarded. It appears to some differences and troubles will be generated in practical application. Instead, some heuristic solutions take the actual operating conditions and limitations into account, allowing the practical community to adopt them even without best solution. However, it is doubtful whether the result is really close to the best solution. Contrary to the impression, it is also important to consider that this heuristic has the same effect when the field conditions are different.

3. Short-Term Scheduling Model

Scheduling belongs to short-range planning, which focuses on the quality and speed of scheduling, it needs to arrange a processing sequence that is closer to the target under

various constraints in a shorter time. In recent years, the cluster tool, such as dry etcher, CVD, became the bottleneck repeatedly in the small or medium-sized fabs. Generally, there are some processing limitations in the cluster tools and its internal process is complicated. The production capacity of the cluster tool, hence, is easily affected by the processing combination. If the field operator and even the manager cannot make a good decision to arrange the production sequence from the existing or upcoming WIP, the output of the plant will be affected deeply. This is the reason why a cluster tool needs an effective short-term scheduling method. A module to estimate the arrival time of WIP should be developed to take the coming WIP into account. Accordingly, there are two modules, arrival time estimation and shop floor scheduling, included in the short-term scheduling model.

3.1. Arrival Time Estimation

Generally, the product's turn ratio (TR) is applied to the look-ahead function of the arrival time of the WIP. Nonetheless, the conditions in the factory are changing quickly and the cycle time of product's step is varied, TR value of the product will be changed with the environment and different in every stage of product. Thus, the concept of dynamic cycle time of product step instead of TR is applied to estimate the arrival time of lot in front of cluster tool to increase the accuracy of scheduling result. Because the data is applied for the short-term scheduling only, to consider the upcoming WIP within 3 h will be satisfactory. Afterwards, the cycle time data needed to estimate the arrival time of lot will be about four steps prior to cluster tool. The cycle time of product steps include queue time and processing time eventually. The processing time is constant for each product step, but the queue time is varied by shop floor status. It exposes the importance of dynamic queue time for the accurate scheduling results. The following is the calculation and monitoring procedure of product step's queue time.

3.1.1. Step 1. Decide the Time Span (TS) of the Data Required to Calculate the Queue Time of Equipment

Generally, the long-range and steady-state historical data is used to calculate the product's cycle time data for long-term planning. Nevertheless, if there is a significant difference between the future product portfolio and the past's or the shop floor status fiercely changed, it is recommended to build a simulation system or use other mathematical models to calculate the queue time of each machine. As for short-term planning, the entire production system should not be far away from the situation compares to the past few days, so it is feasible to use the recently historical data to calculate the queue time data. As the situation of shop floor is not static, the queue time of WIP will also be changed since the condition of the machine, distribution of WIP, and combination of product released are changing every single day. For short-term planning or scheduling, the accuracy of any data will seriously affect results, and cycle time is of course no exception. Therefore, in the short-term planning, the time range of the information required for the cycle time of product should not be too long, otherwise the on-site situation will be covered by the long-term trend. On the other hand, if the time span is too short, just as short as half hour, the data will also generate some bias. Therefore, it is usually recommended to calculate the queue time data based on 3–7 days.

3.1.2. Step 2. Calculate Queue Time of Equipment

The time of a job is made to wait before processing is called queue time [27]. As mentioned in step 1, the queue time data within 3–7 days will be taken to calculate their average and standard deviation. The equations are as follows:

$$QT_i(TS) = \frac{\sum_{d=1}^{TS} \sum_{l_d=1}^{n_d} QT_{idl_d}}{\sum_{d=1}^{TS} n_d}$$

$$Std_{QT_i(TS)} = \sqrt{\frac{\sum_{d=1}^{TS} \left(QT_{idl_d} - QT_i(TS)\right)^2}{\sum_{d=1}^{TS} n_d}}$$

where,

$QT_i(TS)$: average queue time of workstation i within time span TS,

QT_{idl_d}: queue time of lot l in workstation i at day d,

$Std_{QT_i(TS)}$: standard deviation of queue time in workstation i within time span TS,

TS: the time span of queue time calculation (day).

3.1.3. Step 3. Monitor and Amend the Queue Time of Equipment

Generally, the short-term planning requires more accurate data to insure the validity of result. Although the queue time data is already a lagging indicator, it is still hoped that it will not be too far from the current and future factory conditions. Accordingly, the effectiveness of the queue time data will be monitored with the concept of a control chart. The module will set the queue time of the machine to be calculated every three hours to simplify the calculation.

As for the control chart, the average queue time in the time span TS is the center line, and the average queue time plus or minus n times of standard deviations is regarded as the upper and lower limits of the control chart. The value of n is determined by the user. In general, 3 standard deviations are used. However, if the user wants more strict control, 2 standard deviations can be used as the limit as well. When the control chart is built, the average queue time calculated every three hours can be put into the control chart to trace whether the on-site conditions are still within a tolerable range of difference. Once the average queue time within 3 h exceeds the limit range, it should go back to step 2 to recalculate the average queue time within time span TS for scheduling.

With the usable cycle time data, the WIP arrival time estimation can be estimated according to the following formula before scheduling, associates with exact date and time. The formula is as follows:

$$T_{wi} = TNow + \sum_{k=S_{\text{wn}}}^{S_i} QT_{ws_k} + PT_{ws_k}$$

$$w \in S_w$$

$$S_w = set\ of\ WIP$$

where,

S_i *will within the next* 4 *steps of this WIP,*

where,

T_{wi}: the time of WIP w will arrive to workstation i,

$TNow$: current time,

$QT_{ws_k}i$: average queue time of workstation i where WIP w at the step s_k,

PT_{ws_k}: processing time of WIP w at the step s_k,

S_i: the step processed on workstation i,

S_{wn}: current step of WIP w,

S_w: a set of WIP where S_i will within the next 4 steps of this WIP.

In addition, some WIPs will have the time constraint concern, including single time constraint (single TC) and time constraint loop (TC loop). The single TC means that the time constraint starts at the lot arriving the cluster tool and ends at the lot processed. TC loop stands for the time constraint covers a process segment. Due to the time constraint being covered between a process segment, the quota of time constraint can be allocated to use in the cluster tool portion which should be calculated in advance. Generally, the time constraint quota of each workstation in the loop is allocated by the ratio of its standard

cycle time. A lot may be with single TC and TC loop simultaneously, these two kinds of TC will be combined and taken the small one as the only TC of this lot to increase the scheduling efficiency. The formulas of TC calculation are as follows:

$$RTC_{wi} = Min(RSTC_{wi},\ RTCL_{wi})$$

$$RSTC_{wi} = \begin{array}{cc} STC_{wi} - QT_{wi} & \textit{lot with single TC} \\ 99999 & \textit{lot without single TC} \end{array}$$

$$RTCL_{wi} = \begin{array}{cc} TCL_{wi} - QT_{wi} & \textit{lot with TC loop} \\ 99999 & \textit{lot without single TC} \end{array}$$

$$TCL_{wi} = RTCL_w \times \frac{CT_{wi}}{\sum_{j=i}^{le} CT_{wj}}$$

where,
RTC_{wi}: remaining queue time limit of WIP w at workstation i,
$RSTC_{wi}$: remaining queue time limit of WIP w which belongs to single TC at workstation i,
$RTCL_{wi}$: remaining queue time limit of WIP w which belongs to TC loop at workstation i,
STC_{wi}: single TC of WIP w at workstation i,
TCL_{wi}: TC loop of WIP w at workstation i,
QT_{wi}: current queue time of WIP w at workstation i,
$RTCL_w$: remaining queue time limit of WIP w which belongs to TC loop,
CT_{wi}: standard cycle time of WIP w at workstation i,
le: the end workstation of TC loop.

Based on the algorithm, the flowchart of WIP arrival time estimation is showed as Figure 2.
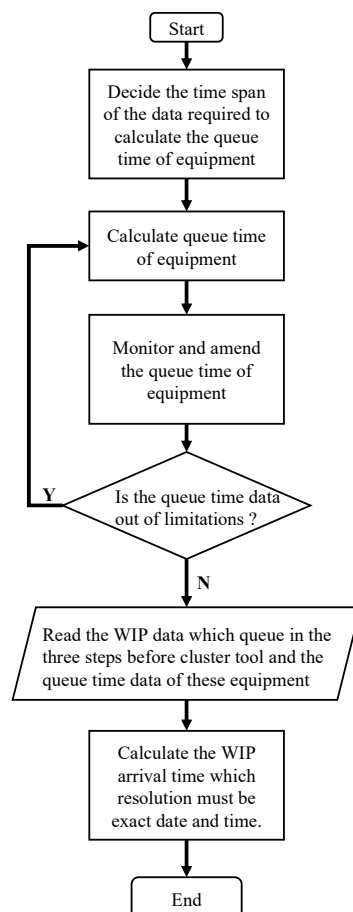

Figure 2. Flowchart of WIP arrival time estimation.

3.2. Short-Term Scheduling

Scheduling is a complex and tedious task. Although each machine in the field belongs to the same workstation, it has its own restrictions and configurations. In addition, each lot also has its own characteristics, such as: recipe to be processed, remaining queue time limit, delivery date, etc. In such a many-to-many combination, it is very difficult to arrange a better schedule under the target consequently. Short-term scheduling needs to be completed in a very short time. Many systems use rule base to filter out some unqualified combinations, and then exhaustively find the best schedule from a relatively small amount of data to complete it in a short period. Nonetheless, it is often found that sometimes the best solution is filtered out when the rule base is designed improperly in practice, so the best solution after exhaustion of the remaining combinations is not the real best solution. In order to avoid such a situation, the rule base plus exhaustive method is eliminated in this scheduling model. Basically, the main goal in this scheduling model is to maximize the output without exceeding the time constraint. It leads to the latest time that WIP must be processed on the machine should be first calculated and indicated, and then the efficiency table of the machine processing combination [4] (it is called as combination efficiency table) is applied to arrange the future 3-h schedule for each machine. It is well known that the status of the shop floor is varied quickly, to re-schedule every half hour to respond to the on-site conditions is necessary. Because this model arranges the processing of WIP according to the efficiency table of machine, it must be able to ensure that the resulting schedule is the best schedule under the current situation, and it does not need to be exhaustive. The following is a detailed description of the scheduling procedure.

3.2.1. Step 1. Set the Latest Time Which WIP Should Be Processed

In the previous stage, the remaining time constraint of lot has been calculated. Regardless of whether the lot has a time constraint or not, a remaining time constraint is given. If the lot without time constraint, its remaining time constraint is set to one larger value, for example: 99,999. The latest time which lot should be processed is calculated as follows:

$$LST_{wi} = T_{wi} + RTC_{wi}$$

where,
LST_{wi}: latest time which WIP w should be processed in workstation i.

Each lot has its own time limit for processing. Here, all WIPs will be divided into four zones based on its time limit. Lots with the machining time limit less than one hour belong zone one, which between one and two hours belong zone two, and so on for zone three. All lots with more than three hours belong to zone four. The lot in each zone is sorted according to its machining time limit from small to large, the rule as above to form a WIP urgency table.

3.2.2. Step 2. Estimate the Complete Time of Processing Lots

When the scheduling is executed, each machine of this workstation may be processing some lots. The time when the works on hand will be completed should be estimated in advance. The estimation method is as follows:

$$FT_{li} = PT_{li} + ST_{li} PT_{li} = TPT_{li} \times E_{li}$$

where,
FT_{li}: completed time of the lot l at machine i,
PT_{li}: processing time of the lot l at machine i,
ST_{li}: start time of the lot l to be processed at machine i. If current time is set as 0, ST_{li} will be negative value,
TPT_{li}: theoretical processing time of the lot l at machine i,
E_{li}: efficiency of the lot l in machine i under current processing combination.

3.2.3. Step 3. Arrange the Next 3 h of Production Activities for Each Machine in This Workstation

The available machines can be sorted out in the order of time when all completion times of processing lots are estimated in Step 2. Because the cluster tool is a single-wafer multi-chamber machine, there should be more than one lot processing on the machine at the same time. Therefore, the same machine may appear several times in the completion time sequence. When scheduling, three sets of data will be referred to at the same time: machine completion schedule, WIP urgency table, and the combination efficiency table. Since the major target of scheduling is to keep all lots within the time constraint, the lots in the zone one of the WIP urgency table first will be arranged accordingly. Under lots' time constraint and refer to the combination efficiency table to find out the lot which is the best efficiency combination with the processing lots of the scheduled cluster tool. The start time and completion time of lot processing are arranged, and the completion time of the machine is added to the machine completion schedule. The detailed scheduling procedures are as follows.

Step 3.1. Read Data

The data required for scheduling includes: machine processing status, machine completion schedule, WIP urgency table, and combination efficiency table.

Step 3.2. Capacity check and allocate for each zone

Generally, the lots in zone one, zone two and zone three are with time constraint and should be started to process by the identical time. On these grounds, it has to make sure there are enough capacity to process these lots. In this step, the machine completion schedule should be matched to the lot quantity in each zone of WIP urgency table. If the numbers of machine within the first one hour is more than lot quantity, it means the machine capacity is surplus, otherwise, the capacity is insufficient and some lots will be over time constraint. This is obvious that if there is surplus capacity for zone one or zone two, it can support the WIPs in its previous zone to avoid over time constraint of lot. Otherwise, the capacity should be checked and lots need to be re-allocated. The reallocation procedures are as follows.

Step 3.2.1. Reallocate lots of zone one.

If the capacity for zone one is insufficient, the lots with longer queue time limit will be moved to zone two no matter the capacity of zone two is enough or not.

Step 3.2.2. Reallocate lots of zone two and zone three.

If the capacity for zone two or zone three is insufficient, there will have two situations to consider. The first one is the capacity of its previous zone is surplus, then if the surplus capacity of previous is sufficient to support the lots which are over capacity, moving the lots with shorter queue time limit to its previous zone, otherwise, moving the lots with longer queue time limit to its next zone.

Step 3.3. Schedule Lots

All lots in all zones will be scheduled in this step from zone one to zone four. The scheduling procedures are as follows:

Step 3.3.1. Arrange the lots in the selected zone of WIP urgency table.

The lot in zone one of the urgency table is arranged first. Moreover, the lots in zone one mean that these lots must be processed within one hour, and lots in zone two must be processed within two hours, and so on. Based on the latest start time of the first lot in the selected zone of the urgency table, the machines available before this time are listed as candidates and the best performance is selected from the candidates according to the combination efficiency table. Based on the matching machine, arrange the production schedule of this lot, including the start time and completion time, and add the machine completion time to the machine completion schedule and mark the time when the lot starts processing on this machine. All the lots in the selected zone are arranged in this order.

Step 3.3.2. Arrange the remaining capacity of machines in the selected zone's hour.

If there are machines still list in the machine completion schedule after the urgent lots in the selected zone scheduled, they should be arranged. The best matching lot will be selected according to the combination efficiency table, and the urgency of WIP is broken. The urgency of lot in the zone two to zone fours is cancelled. Arrange the best lot by the order of machine completion time until all machines list in the machine completion schedule of the selected zone hour are scheduled, and add the machine completion time to the machine completion schedule and mark that the lot starts on this machine processing time.

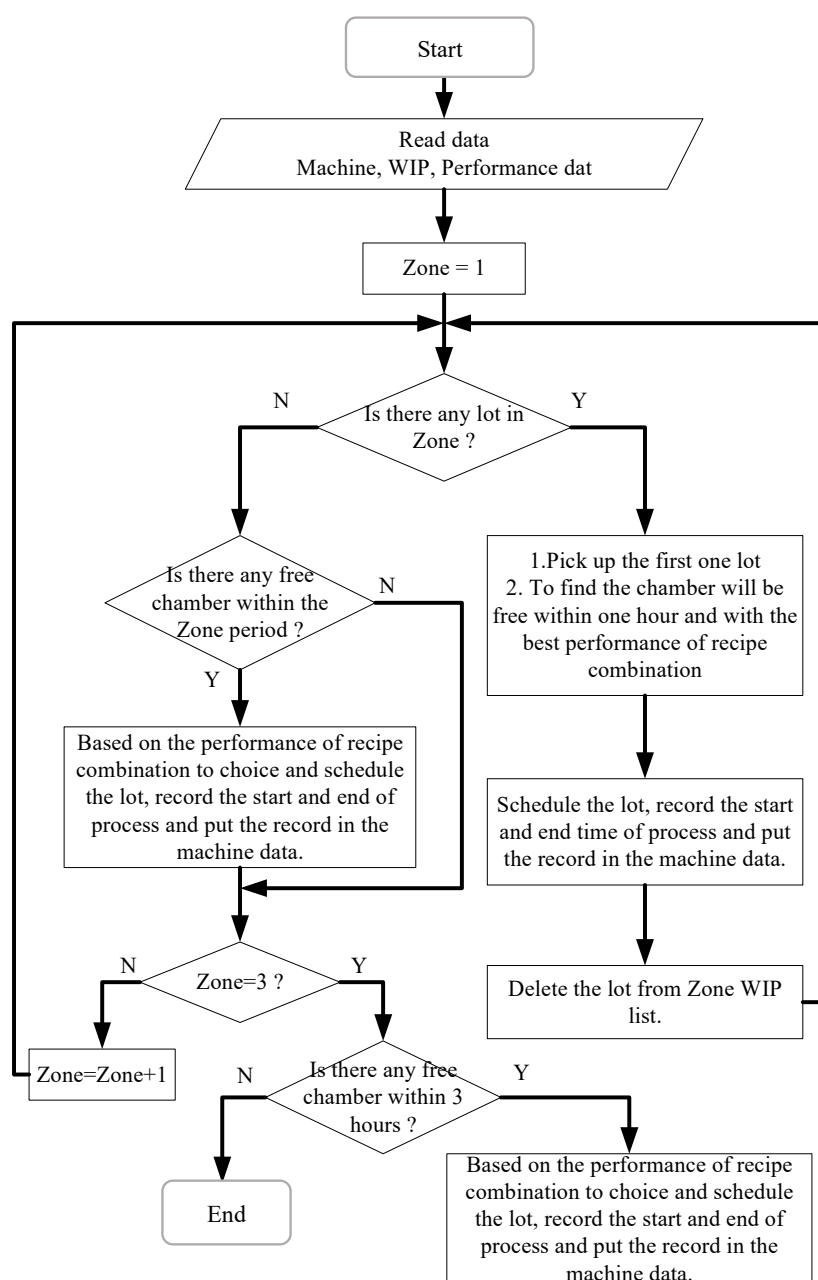Based on the proposed algorithm, the flowchart of short-term scheduling is showed as Figure 3.

Figure 3. Flowchart of short-term scheduling algorithm for cluster tool.

4. Performance Comparison with Dissimilar Model

For validating the performance of proposed model, a simplified simulation model with a parallel cluster tool is established and to compare the performance results with FIFO rule. Although there is only one cluster tool in the system, the behaviors of cluster tool are established as the production tool and has been validated [4]. Besides, an arrival list of wafer lots for 33 days is created in advance for expediting the validation process and the lookahead function is ignored in such a situation. The configuration of parallel cluster tool, such as dry etcher or CVD equipment, is as Figure 1 with three chambers and three load ports. The product mix and combination table are as Tables 1 and 2. Because the arrival list is lengthy, Table 3 only shows portion of list. The time constraints lots are set only on the recipe TR3 and the remaining time constraints of lot is set between 2 h and 24 h randomly. Furthermore, the heavy loading situation is applied for demonstrating the effects of the proposed model. In the heavy loading environment, the better combination provided by proposed model can be selected and the difference from FIFO model can be appeared. Regarding to the performance indexes, the proposed model took product throughput, cycle time and the ratio of over time constraints into account. Tables 4–6 show the performances of proposed model and FIFO model with taking TC (Remining Time constraints) and without TC.

Table 1. Product mix and information of recipe.

Recipe	PT (Sec)	WAC1 (Sec/Lot)	WAC2 (Sec/PCS)	WAC3 (Sec/Lot)	Stable + Others (Sec/PCS)	WPH	Rate
TR1	150	200	80	80	30	13.27	60%
TR2	310	200	80	80	30	8.35	30%
TR3	250	200	80	80	30	9.70	10%

Table 2. Combination table.

Priority	Combination
1	TR1 + TR2 + TR3
2	TR2 + TR3 + TR3
3	TR3 + TR2 + TR2
4	TR2 + TR2 + TR2
5	TR2 + TR1 + TR1
6	TR1 + TR3 + TR3
7	TR1 + TR2 + TR2
8	TR3 + TR3 + TR3
9	TR3 + TR1 + TR1
10	TR1 + TR1 + TR1

Table 3. Arrival list.

Arrival Time	Recipe	TC Flag	Remaining TC (min)
0.0000	R3	1	119
0.0000	R1	0	0
0.0000	R3	1	393
0.0000	R1	0	0
0.0000	R2	0	0
0.0000	R2	0	0
35:34.5800	R1	0	0
1:16:00.8173	R1	0	0
2:02:07.1239	R3	1	68
2:43:59.7596	R1	0	0
3:22:57.2514	R1	0	0
3:53:14.4643	R2	0	0
4:24:22.9500	R2	0	0

Table 4. Performance of proposed model.

	CT_Mean (Min/Lot)	CT_Dev (Min/Lot)	PT_Mean (Min/Lot)	PT_Dev (Min/Lot)	Throughput (Lot)	Over_TC (Lot)
R1	3144	1010	114	1	591	
R2	7541	3138	181	1	220	
R3	233	64	159	1	110	
Total	3846	2853	136	29	921	0

Table 5. Performance of FIFO model with taking TC into account.

	CT_Mean (Min/Lot)	CT_Dev (Min/Lot)	PT_Mean (Min/Lot)	PT_Dev (Min/Lot)	Throughput (Lot)	Over_TC (Lot)
R1	4950	1947	114	2	529	
R2	5172	1941	181	1	259	
R3	200	29	159	2	110	2
Total	4432	2414	139	30	898	2

Table 6. Performance of FIFO model without taking TC into account.

	CT_Mean (Min/Lot)	CT_Dev (Min/Lot)	PT_Mean (Min/Lot)	PT_Dev (Min/Lot)	Throughput (Lot)	Over_TC (Lot)
R1	4380	1775	115	2	541	
R2	4601	1755	181	1	266	
R3	4266	1900	159	2	94	94
Total	4433	1784	139	30	901	94

The results reveal the proposed model is better than FIFO model from all perspectives include throughput, recipe average cycle time of lot and the number lots of over time constraints. The only one point can be further addressed in this proposed model is the difference of the mean cycle time of each recipe. From the Table 6, it reveals the mean cycle time of pure FIFO model of each recipe is closer than other models. Nonetheless, Table 5 shows the mean cycle time of FIFO model with time constraint factor of each recipe also has significant difference. Therefore, it is reasonable that the mean cycle time of each recipe of the proposed model has more difference than FIFO model. The average cycle time of the proposed model is shorter than other models even though the mean cycle time of recipe is diversified.

5. Conclusions

Cluster tools are not only numerous in the fab, but also treated as an important process tools. They increase the difficulties in the shop floor management and control because of thier complex structure. A short-term scheduling model for cluster tools to arrange a better schedule under the goals and constraints is proposed at the heart of this study. Generally, quality and speed of the scheduling are the necessary characteristics for an on-site short-term scheduling model. In the scheduling model, the main goal is to maximize the output of all the WIP under the principle of not exceeding the queue time limit. For verifying the necessaries and goals of short-term scheduling, two modules are developed—arrival time estimation and short-term scheduling. The concept of the dynamic cycle time of the product's step is applied to estimate the arrival time of the WIP in front of machine. Moreover, an algorithm to calculate the latest time of the WIP to process on the machine is developed as well to avoid to violating the time constraint of the WIP. Based on the latest process time of the WIP and the combination efficiency table, the production schedule of the cluster tools can be arranged to fulfill the production goal properly. Finally, the short-term scheduling model can provide easily and well controlled cluster tools for the shop floor management and control. The results strongly support the viability of the proposed model from various perspectives.

Regarding further works, in order to reduce the impact by uncertainty and/or inaccuracy, the fuzzy preprocessing techniques of the data can be applied [28–31].

Conflicts of Interest: The author declare no conflict of interest.

References

1. Liao, D.Y. Automation and Integration in Semiconductor Manufacturing. In *Semiconductor Technologies*; Exclusive Books: Johannesburg, South Africa, 2010; pp. 39–56.
2. Kim, D.K.; Lee, T.E.; Lim, H.J. Optimal scheduling of transient cycles for single-armed cluster tools with parallel chambers. *IEEE Trans. Autom. Sci. Eng.* **2016**, *13*, 1165–1175. [CrossRef]
3. Lee, T.E. A review of scheduling theory and methods for semiconductor manufacturing cluster tools. In Proceedings of the 40th Conference on Winter Simulation, 2008 Winter Simulation Conference, Miami, FL, USA, 7–10 December 2008; pp. 2127–2135.
4. Tu, Y.M. Throughput Estimation Model of Cluster Tool in Semiconductor Manufacturing. *Key Eng. Mater.* **2019**, *814*, 196–202. [CrossRef]
5. Tu, Y.M. Capacity estimation model of cluster tool in wafer fabrication. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 1877–1882.
6. Lim, Y.; Yu, T.S.; Lee, T.E. Adaptive Scheduling of Cluster Tools with Wafer Delay Constraints and Process Time Variation. *IEEE Trans. Autom. Sci. Eng.* **2019**, *17*, 375–388. [CrossRef]
7. Yu, Q.; Yang, H.; Lin, K.Y.; Li, L. A Predictive Dispatching Rule Assisted by Multi-Layer Perceptron for Scheduling Wafer Fabrication Lines. *J. Comput. Inf. Sci. Eng.* **2020**, *20*, 031001. [CrossRef]
8. Wang, L.C.; Chu, P.C.; Lin, S.Y. Impact of capacity fluctuation on throughput performance for semiconductor wafer fabrication. *Robot. Comput. Integr. Manuf.* **2019**, *55*, 208–216. [CrossRef]
9. Weigang, J.; Ringel, R.; Kalisch, S.; Quarg, T. Method and system for scheduling a stream of products in a manufacturing environment by using process-specific WIP limits. U.S. Patent No. 8,185,226, 22 May 2012.
10. Uzsoy, R.; Lee, C.Y.; Martin-Vega, L.A. A Review of Production Planning and Scheduling Models in Semiconductor Industry Part I: System Characteristics, Performance Evaluation and Production Planning. *IIE Trans.* **1992**, *24*, 47–60. [CrossRef]
11. Toba, H.; Izumi, H.; Hatada, H.; Chikushima, T. Dynamic load balancing among multiple fabrication lines through estimation of minimum inter-operation time. *IEEE Trans. Semicond. Manuf.* **2005**, *18*, 202–213. [CrossRef]
12. Lee, Y.Y.; Hsiao, C.C. Cross area lot arrangement—Shortest inter cycle time. In Proceedings of the 2007 International Symposium on Semiconductor Manufacturing Conference, Santa Clara, CA, USA, 15–17 October 2007.
13. Chang, C.S.; Lo, H.L.; Pan, C.C.; Wu, D.L. Advanced effective dispatching method for 300mm Twin-Phase Twin-Fab. In Proceedings of the 2009 IEEE/SEMI Advanced Semiconductor Manufacturing Conference, Berlin, Germany, 10–12 May 2009; pp. 76–79.
14. Lee, Y.F.; Jiang, Z.B.; Liu, H.R. Multiple-objective scheduling and real-time dispatching for the semiconductor manufacturing system. *Comput. Oper. Res.* **2009**, *36*, 866–884. [CrossRef]
15. Mathirajan, M.; Sivakumar, A.I.; Kalaivani, P. An efficient simulated annealing algorithm for scheduling burn-in oven with non-identical job sizes. *Int. J. Appl. Manag. Technol.* **2004**, *2*, 117–138.
16. Kim, Y.D.; Joo, B.J.; Choi, S.Y. Scheduling wafer lots on diffusion machines in a semiconductor wafer fabrication facility. *IEEE Trans. Semicond. Manuf.* **2010**, *23*, 246–254.
17. Kurz, M.E.; Mason, S.J. Minimizing total weighted tardiness on a batch-processing machine with incompatible job families and job ready times. *Int. J. Prod. Res.* **2008**, *46*, 131–151. [CrossRef]
18. Klemmt, A.; Weigert, G.; Werner, S. Optimisation approaches for batch scheduling in semiconductor manufacturing. *Eur. J. Ind. Eng.* **2011**, *4*, 338–359. [CrossRef]
19. Yugma, C.; Dauzere-Peres, S.; Derreumaux, A.; Sibille, O. A Batch Optimization Sofware for diffusion area scheduling in semiconductor manufacturing. In Proceedings of the 2008 IEEE/SEMI Advanced Semiconductor Manufacturing Conference IEEE, Cambridge, MA, USA, 5–7 May 2008; pp. 327–332.

20. Yurtsever, T.; Kutanoglu, E.; Johns, J. Heuristic based scheduling system for diffusion in semiconductor manufacturing. In Proceedings of the Winter Simulation Conference, Winter Simulation Conference, Austin, TX, USA, 13–16 December 2009; pp. 1677–1685.

21. Cakici, E.; Scott, J.M. Parallel machine scheduling subject to auxiliary resource constraints. *Prod. Plan. Control* **2007**, *18*, 217–225. [CrossRef]

22. Mazumdar, C.S.; Mathirajan, M.; Gopinath, R.; Sivakumar, A.I. Tabu search methods for scheduling a burn-in oven with non-identical job sizes and secondary resource constraints. *Int. J. Oper. Res.* **2008**, *3*, 119–139. [CrossRef]

23. Chien, C.F.; Chen, C.H. A novel timetabling algorithm for a furnace process for semiconductor fabrication with constrained waiting and frequency-based setups. *Or Spectr.* **2007**, *29*, 391–419. [CrossRef]

24. Bixby, R.; Burda, R.; Miller, D. Short-interval detailed production scheduling in 300mm semiconductor manufacturing using mixed integer and constraint programming. In Proceedings of the 17th Annual SEMI/IEEE ASMC 2006 Conference, IEEE, Boston, MA, USA, 22–24 May 2006; pp. 148–154.

25. Fordyce, K.; Bixby, R.; Burda, R. Technology that upsets the social order: A paradigm shift in assigning lots to tools in a wafer fabricator—The transition from rules to optimization. In Proceedings of the 40th Conference on Winter Simulation, Winter Simulation Conference, Miami, FL, USA, 7–10 December 2008; pp. 2277–2285.

26. Cochran, J.K.; Horng, S.M.; Fowler, J.W. A multi-population genetic algorithm to solve multi-objective scheduling problems for parallel machines. *Comput. Oper. Res.* **2003**, *30*, 1087–1102. [CrossRef]

27. Swamidass, P.M. (Ed.) *Encyclopedia of Production and Manufacturing Management*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2000.

28. Postorino, M.N.; Versaci, M. A geometric fuzzy-based approach for airport clustering. *Adv. Fuzzy Syst.* **2014**, *2014*, 201243. [CrossRef]

29. Mahela, O.P.; Khan, B.; Alhelou, H.H.; Siano, P. Power quality assessment and event detection in distribution network with wind energy penetration using stockwell transform and fuzzy clustering. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6922–6932. [CrossRef]

30. Bagherinia, A.; Minaei-Bidgoli, B.; Hosseinzadeh, M.; Parvin, H. Reliability-based fuzzy clustering ensemble. *Fuzzy Sets Syst.* **2021**, *413*, 1–28. [CrossRef]

31. Wang, Q.; Yang, X. Investigating the sustainability of renewable energy–An empirical analysis of European Union countries using a hybrid of projection pursuit fuzzy clustering model and accelerated genetic algorithm based on real coding. *J. Clean. Prod.* **2020**, *268*, 121940. [CrossRef]