Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

SKIPPING ALGORITHMS FOR DEFECT INSPECTION USING A DYNAMIC CONTROL STRATEGY IN SEMICONDUCTOR MANUFACTURING

Gloria Luz Rodriguez-Verjan
Stéphane Dauzère-Pérès
Sylvain Housseman

Department of Manufacturing Sciences and Logistics
Ecole Nationale des Mines de Saint-Etienne - CMP
F-13541 Gardanne, FRANCE

Jacques Pinaton

STMicroelectronics Rousset
F-13106 Rousset, FRANCE

ABSTRACT

In this paper, we propose new ways for efficiently managing defect inspection queues in semiconductor manufacturing when a dynamic sampling strategy is used. The objective is to identify lots that can skip the inspection operation, i.e. lots that have limited impact on the risk level of process tools. The risk considered in this paper, called *Wafer at Risk* ($W@R$), is the number of wafers processed on a process tool between two defect inspection operations. An indicator *(GSI, Global Sampling Indicator)* is used to evaluate the overall $W@R$ and another associated indicator *(LSI, Lot Scheduling Indicator)* is used to identify the impact on the overall risk if a lot is not measured. Based on these indicators, five new algorithms are proposed and tested with industrial instances. Results show the relevance of our approach and that evaluating sets of lots for skipping performs better than evaluating lots individually.

1 INTRODUCTION

In semiconductor manufacturing, *yield* is an important indicator that reflects the ability to produce high quality products. In order to ensure high *yield*, regular inspections are introduced in the manufacturing process. However, inspection capacity is costly and limited. Besides, more inspections do not necessarily result in more quality (Leachman and Ding 2011). With the increase of the sampling rate, the *yield* also increases. However, after a certain limit, the queue of lots waiting for inspection grows, leading to longer delays for corrective actions with negative impacts on the *yield* (Tirkel and Rabinowitz 2012). To cope with this situation, several sampling techniques have been developed. These sampling techniques are classified according to their capacity to react regarding the production state. The three categories are: *Static*, *Adaptive* and *Dynamic sampling* (Nduhura Munga et al. 2013). In *Static Sampling*, the selection of lots is done at the beginning of the manufacturing process and does not change throughout production (Lee et al. 2001). *Adaptive sampling* is based on adapting the sampling rate according to the state of production (Sullivan et al. 2004, Mouli 2005 and Song-bor et al. 2003). In *Dynamic Sampling*, no rules are defined in advance, the selection of lots is done in real time and according to the information carried by the lot (Good and Purdy 2007 and Dauzère-Pérès et al. 2010). While a selected lot is waiting for inspection, the production state changes and the lot may lose its interest to be inspected. Hence, it can be removed from the inspection queue and moved to the next process operation. Some studies have already focused on methods to identify lots that can skip measurement. Purdy et al. (2005) propose a method to release lots in metrology queue, each lot is evaluated individually and the objective is to guarantee the measurement of lots with more recent information. The developed application is part of a sampling system which combines a number of separate sampling rules into a single sampling decision. Sahnoun et al. (2011) propose an algorithm to identify the lots to be skipped according to a risk indicator, called *Wafer at Risk* ($W@R$). A system composed of a buffer and an inspection tool is simulated and only the risk of one process tool is considered. Their results show that 33% of measures could be skipped without losing any information. In this paper we are considering the complexity of a real semiconductor manufacturing plant (fab) and experiments are conducted on industrial data.

The risk considered in this study is evaluated using the *Wafer at Risk (W@R)* indicator, which is the number of wafers processed on a process tool between two defect inspections operations. In general, the *W@R* of a process tool is incremented each time a lot is processed and is decremented when the results of a measure are obtained. In this study, we focus on the micro defect inspections where the flaws produced by particles are detected. The selection of lots is done dynamically and according to the production risk level. The indicator used to evaluate the global *W@R* is called the *Global Sampling Indicator (GSI)* proposed in Dauzère-Pérès et al. (2010).

This paper presents a new methodology to manage the defect inspection queues based on the GSI and LSI indicators. The objective is to identify lots that can skip the inspection operation with limited impact on the *W@R*. Measuring a lot in the inspection queue usually reduces the *W@R* of several process tools. The decision of skipping is evaluated according to the impact on the GSI if a lot is not inspected. Depending on the algorithm used for skipping, different sets of lots can be obtained, hence different impacts on the overall *W@R*. The proposed algorithms have been tested with industrial instances and are currently used at the defect inspection area of a 200mm semiconductor manufacturer.

The paper is organized as follows. Section 2 defines the problem and notations. Section 3 presents the proposed algorithms. Section 4 is devoted to the analysis of the numerical results. Finally, concluding remarks are given in Section 5.

2 PROBLEM DESCRIPTION

Our problem focuses on the defect inspection area. The main objective of defect inspection is the early detection of flaws produced by particles (May and Spanos 2006). Hence, a *defect inspection control plan* is defined by product and contains the list and position of inspection operations that have to be performed within the manufacturing route. It also includes the *coverage block* of each inspection operation, i.e. the list of process tools that can be controlled. Figure 1 shows a small portion of the production route for a product of technology “A”. Between some process operations, there are two defect inspection operations that can be performed (“1202” and “1330”). When a lot l is inspected in inspection operation “1202”, the *W@R* of the tool that processed l on the process operation “1200” is reduced. The same for the inspection operation “1330”, when lot l is controlled, the *W@R* on the tools that processed l from process operation “1210” to process operation “1290” can be reduced. In other words, a lot carries the *W@R* information of the process tools where it was processed and if the lot is inspected, only the *W@R* of the process tools that are covered can be reduced. Even if a lot has been waiting for a long time in the inspection queue, it does not mean that it can be skipped. There are some cases where this lot is the only one that can reduce the *W@R* of some process tools, notably because of the large number of processing tools and the design of defect inspection control plans, which differ between the products.

Two control limits are used to manage the *W@R* on tools: The *Warning Limit* and the *Inhibit Limit*. The *Inhibit Limit (IL)* refers to the maximum value of acceptable *W@R* on the process tool. If the *W@R* reaches the IL, the tool must be stopped and a special control is performed on the tool. The *Warning Limit (WL)* is a limit after which actions have to be taken before the *W@R* reaches the IL. Figure 2 illustrates how the *W@R* evolves on a process tool. The *W@R* is incremented each time a lot is processed. Controlling a lot does not necessarily reduce the *W@R* of any machine. Let us consider that Lot 1 is processed before Lot 2, suppose that after a while, both lots are waiting in the inspection queue. Lot 2 is measured first and results are within the control limits, in consequence the *W@R* of the process tool is reduced by the amount of wafers that were processed on the tool before Lot 2 was processed. When Lot 1 is measured, the *W@R* is not decreased because the information brought by Lot 1 is redundant with the measure of Lot 2. In consequence, Lot 1 could be removed from the inspection queue without impacting the *W@R*

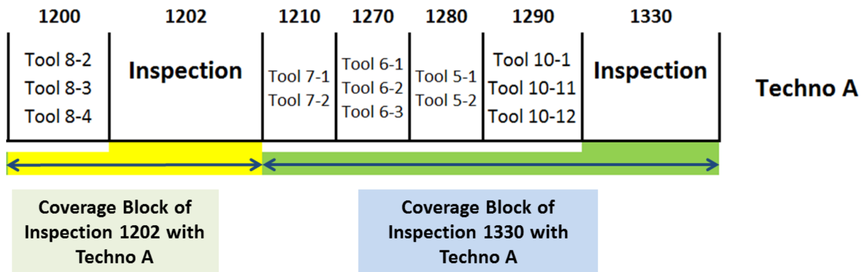

Figure 1: Extract of a small portion of a defect inspection control plan

on the related process tool. This situation can occur because lots can follow different paths after having been processed on the tool. Moreover, the defect inspection area does not select the lots using the FIFO (first-in-first-out) rule.

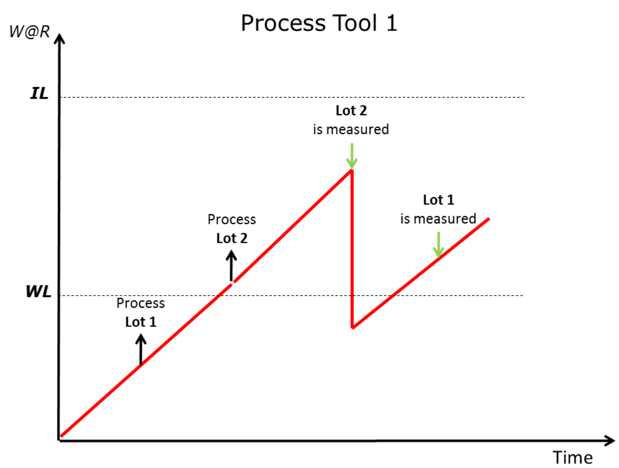

Figure 2: Wafer at risk evolution on process tool 1

Our problem is to identify which lots can be removed from the inspection queue due to redundant information in terms of $W@R$. We use the Global Sampling Indicator (GSI) introduced in Dauzère-Pérès et al. (2010) to evaluate the global risk in the fab. The proposed definition of the GSI can be applied to evaluate different risk contexts (e.g. recipes, products and tools). In this paper, we focus on process tools. Let us recall the following notations:

- R: The number of risk types. In our case it will be the number of considered process tools.
- IL_r: Inhibit Limit for tool r.
- RV_r: Current risk value on tool r. In our case it will be the $W@R$ on process tool r .
- $G_{r,l}$: Gain on risk of tool r if lot l is inspected.
- $NRV_{r,l}$: New risk value of tool r if lot l is inspected, i.e. $NRV_{r,l} = RV_r - G_{r,l}$. In our case it will be the *New Wafer at Risk* ($NW@R$).
- $NRV_r(S)$: New risk value of tool r if lots in set S are inspected. It is calculated as follows:

$$NRV_r(S) = \min_{l \in S} NRV_{r,l}$$

- α: Is a parameter used to give more or less emphasis on getting as far as possible from Inhibit Limits.

The Global Sampling Indicator (GSI) is used to evaluate the overall considered risk (in our case the $W@R$) when a set of lots S are inspected. It is calculated as follows:

$$GSI(S) = \sum_{r=1}^{R} \left(\frac{NRV_r(S)}{IL_r} \right)^{\alpha} \tag{1}$$

Through experimentation, Nduhura Munga (2012) studied the impact of the parameter α on the performance of the GSI. They observed that satisfactory results are achieved by setting $\alpha = 6$. This value will be used in all of our experiments. To evaluate the impact of lot l in the set of lots $S_{initial}$, we use the *LSI* indicator introduced in Nduhura Munga (2012). It is determined by calculating how much would be lost in terms of GSI if lot l is not measured. $LSI(L)$ is defined as the difference between $GSI(S_{initial} \setminus L)$ and $GSI(S_{initial})$. The smaller the value of $LSI(L)$, the less important is the lot. Let us suppose that there are 3 lots L1, L2 and L3 in the waiting queue of the defect inspection area. To define the impact of skipping the measurement of a lot, four combinations are evaluated. These combinations are obtained by removing each lot from the initial set of lots ($S_{initial}$):

1. $GSI(S_{initial}) = GSI\{L_1, L_2, L_3\}$
2. $LSI\{L_1\} = GSI(S_{initial} \setminus L_1) - GSI\{S_{initial}\} = GSI\{L_2, L_3\} - GSI\{L_1, L_2, L_3\}$
3. $LSI\{L_2\} = GSI(S_{initial} \setminus L_2) - GSI\{S_{initial}\} = GSI\{L_1, L_3\} - GSI\{L_1, L_2, L_3\}$
4. $LSI\{L_3\} = GSI(S_{initial} \setminus L_3) - GSI\{S_{initial}\} = GSI\{L_1, L_2\} - GSI\{L_1, L_2, L_3\}$

When a set of lots is removed simultaneously, the LSI will be associated to that set of lots and not only to each lot removed independently. For instance, removing lots L1 and L2 will be evaluated by calculating $LSI\{L_1, L_2\} = GSI(S_{initial} \setminus S\{L_1, L_2\}) - GSI\{S_{initial}\} = GSI\{L_3\} - GSI\{L_1, L_2, L_3\}$. A threshold named T_{Metro} is used to decide whether or not a lot or a set of lots can be skipped. It can be interpreted as the minimum gain in terms of risk reduction that a lot or a set of lots should bring to stay in the waiting queue or as the maximum risk value that can be tolerated for degrading the initial *GSI*. The higher the value of T_{Metro}, the higher the risk that will be tolerated. All combinations associated with removing each $L \in S_{initial}$ from $S_{initial}$ are evaluated. However, different algorithms used to identify the lots to be skipped can lead to different decisions. Aside from the *LSI* criteria to decide if a lot can be skipped, there are some rules defined in advance to guarantee the measure of certain lots. These rules can be defined by the defect inspection team when there is a focus on measuring a particular group of lots. This information is used to create a set of lots that might skip the measure. This set of “skippable” lots is a subset of the lots in the inspection queue ($S_{Skippable} \subset S_{initial}$).

The next section describes briefly the five algorithms that have been developed and tested using industrial instances. In Section 4, some computational results are presented.

3 SKIPPING MECHANISM

This section summarizes the different algorithms that have been implemented for determining the lots that can skip the inspection operation. The objective is to maximize the number of skipped lots while satisfying the threshold. When equivalent solutions are found in terms of the number of lots to skip, the objective is to minimize the global risk. The algorithms are listed according to their complexity and capability to obtain solutions of better quality.

- **Algorithm 1**. The *LSI* for each lot L in set $S_{Skippable}$ is calculated only once. Each lot L for which $LSI(L) < T_{Metro}$ is skipped.
- **Algorithm 2**. The *LSI* for each lot L in set $S_{Skippable}$ is calculated. Each time a lot is identified for skipping, the *LSI* of the remaining lots is recalculated. This procedure is performed only once.

- **Algorithm 3 (greedy heuristic)**: The *LSI* for each lot *L* in set $S_{Skippable}$ is calculated. The lot with the smallest *LSI* is identified and if its *LSI* is smaller than T_{Metro}, the lot is skipped and the *LSI* of the remaining lots are recalculated. This procedure is performed until the lot with the smallest *LSI* cannot be skipped.
- **Algorithm 4 (add-remove local search heuristic)**: It is the same method as in Algorithm 3 except that each time a new lot is selected for skipping, the previous decisions are reviewed.
- **Algorithm 5 (branch and bound)**: The *LSI* is calculated for each lot returned by Algorithm 1. Lots are sorted by increasing *LSI*. A branch and bound method is applied, where bounds consider both the number of lots that can be skipped and the sum of *LSI*.

The first algorithm has two important weaknesses: it considers that lots have independent *LSI*, and that T_{Metro} is an individual threshold while $LSI(Lx,Ly) \geq LSI(Lx) + LSI(Ly)$, and T_{Metro} is global. However, this algorithm is used to reduce the set of lots that can skip the inspection operation (i.e. $S_{Skippable}$). If a lot in the first iteration has its *LSI* strictly larger than T_{Metro} ($LSI > T_{Metro}$), then it cannot not be skipped. Solutions obtained with Algorithm 2 satisfy the restriction of T_{Metro}, nevertheless the solution highly depends on the initial order of the set of lots. Algorithm 3 guarantees that the order of the set does not influence the lots selected to be skipped. Algorithm 4 is an improvement of Algorithm 3 because lots are evaluated by sets instead of evaluating each lot individually. Finally, Algorithm 5 is an exact method that gives an optimal solution. In the following example, Algorithms 1, 3 and 4 are used to illustrate the importance of evaluating sets of lots to skip instead of evaluating lots individually.

Let us suppose there are 5 lots in the defect inspection queue. The $W@R$ reductions that can be obtained by measuring each lot are given in Table 1. The column $W@R$ represents the current risk level of the process tool. The column $NW@R$ shows the risk level after measuring the lot and the column *IL* gives the value of the Inhibit Limit on the process tool. As illustrated in Figure 1, a lot can helps reducing the $W@R$ of several process tools. It depends on the product and the coverage block of the inspection operation. In this example, if lot L1 is inspected, the $W@R$ of process tools 12, 08 and 07 will be reduced.

Table 1: Example of lots waiting to be measured in the defect inspection area

Lot	Process Tool	W@R	NW@R	IL
L1	Tool 07	960	481	1100
L1	Tool 08	948	486	1100
L1	Tool 12	625	425	2500
L2	Tool 05	179	104	500
L2	Tool 06	622	349	1200
L3	Tool 03	82	56	500
L3	Tool 04	79	52	500
L3	Tool 06	622	274	1200
L3	Tool 07	960	456	1100
L4	Tool 08	948	462	1100
L4	Tool 11	737	274	2500
L5	Tool 01	226	104	500
L5	Tool 02	31	1	500
L5	Tool 06	622	299	1200
L5	Tool 09	306	293	1100
L5	Tool 10	302	290	1100
L5	Tool 12	625	425	2500

Table 2: Example of GSI and LSI calculations (Iteration 1)

	$(NW@R/IL)^{\alpha}$					
Tools	$S_{initial}$	$S_{initial} \setminus L1$	$S_{initial} \setminus L2$	$S_{initial} \setminus L3$	$S_{initial} \setminus L4$	$S_{initial} \setminus L5$
Tool 1	0.00008	0.00008	0.00008	0.00008	0.00008	0.00853
Tool 2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Tool 3	0.00000	0.00000	0.00000	0.00002	0.00000	0.00000
Tool 4	0.00000	0.00000	0.00000	0.00002	0.00000	0.00000
Tool 5	0.00008	0.00008	0.00211	0.00008	0.00008	0.00008
Tool 6	0.00014	0.00014	0.00014	0.00024	0.00014	0.00014
Tool 7	0.00507	0.00507	0.00507	0.00699	0.00507	0.00507
Tool 8	0.00549	0.00549	0.00549	0.00549	0.00744	0.00549
Tool 9	0.00036	0.00036	0.00036	0.00036	0.00036	0.00046
Tool 10	0.00034	0.00034	0.00034	0.00034	0.00034	0.00043
Tool 11	0.00000	0.00000	0.00000	0.00000	0.00066	0.00000
Tool 12	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
GSI	0.01159	0.01159	0.01361	0.01363	0.01419	0.02024
LSI	0.00000	0.00000	0.00202	0.00204	0.00260	0.00865

Table 3: Example of LSI calculations

(a) Iteration 2

Lot	*LSI*	Decision
L2	0.002024	Skip
L3	0.436901	Not Skip
L4	0.404892	Not Skip
L5	0.008865	Not Skip

(b) Iteration 3

Lot	*LSI*	Decision
L2	–	–
L3	0.438925	Not Skip
L4	0.406916	Not Skip
L5	0.010890	Not Skip

Table 2 presents the information of the *GSI* for the initial set of lots and the *LSI* for each lot. Suppose that the predefined T_{Metro} is 0.007. If all lots are measured, $GSI(S_{initial}) = 0.01159$. If Lot L1 is removed from the queue, $GSI(S_{initial} \setminus L1) = 0.01159$ and $LSI(L1) = 0.000$, thus Lot L1 can be skipped. This is because, although the $W@R$ can be reduced by inspecting L1, the $W@R$ is also reduced by inspecting other lots that are in the queue (i.e. L3, L4 and L5). If lot L5 is removed from the queue, $GSI(S_{initial} \setminus L5)$ would be 0.002024 and $LSI(L5) = 0.00865$. Since $LSI(L5)$ is larger than T_{Metro}, L5 cannot be skipped because it is the only lot that reduces the risk on tools 1, 2, 9 and 10. It is important to note that the *LSI* has been calculated only with one iteration. If Algorithm 1 is used to skip lots with a *LSI* smaller than T_{Metro}, then the final decision is to skip lots L1, L2, L3 and L4. However, when the *LSI* of lot L2 is calculated, lots L3 and L5 are in the queue. When the *LSI* of L3 is calculated, lots L1, L2 and L5 are in the queue. Thus, skipping simultaneously all the lots with $LSI < T_{Metro}$ can lead to uncontrolled (and thus undesirable) situations. Even if Algorithm 1 cannot be used to take the final decision of skipping, it is used to reduce the number of lots in set $S_{Skippable}$. In this example it can be observed that L5 cannot be skipped. In the following, the mechanism of Algorithm 3 is explained. Let us consider that, in the first iteration, L1 is skipped and the LSI of the remaining lots is recalculated (See Table 3). In the second iteration, L2 has the smallest *LSI* which is smaller than T_{Metro}, and thus will be skipped. In the third iteration, the smallest *LSI* is obtained with L5 but it cannot be skipped because it is larger than T_{Metro}. The final decision would be to skip lots L1 and L2.

Different lots may be selected when computing *LSI* for sets of lots rather than computing *LSI* for lots individually. Table 4 gives the *LSI* when sets of lots are considered. It can be observed that the set of lots $\{$L2, L3, L4$\}$ has the smallest *LSI* (0.006673), and therefore can be skipped. Let us note that, compared to the previous solution, L1 is not skipped. This is due to the fact that, when $LSI(L1)$ is calculated, lots L3 and L4 are in the queue and, by skipping L1, their *LSI* will increase. When the *LSI* is calculated for sets of lots, it will be preferable to leave L1 in the queue and to skip L3 and L4. Then, evaluating a set of lots for skipping performs better than evaluating each lot individually. This is why Algorithms 4 and 5 have been developed.

Table 4: Example of LSI calculations for sets of lots

Set of Lots	*LSI*
L1, L2, L4	0.406916
L1, L2, L5	0.010890
L1, L3, L4	0.841793
L1, L3, L5	0.446132
L1, L4, L5	0.413757
L2, L3, L4	0.006673
L2, L3, L5	0.031869
L2, L4, L5	0.013273
L3, L4, L5	0.013660

4 NUMERICAL RESULTS ON INDUSTRIAL DATA

Our algorithms were developed with the R software (R Development Core Team 2011). The computational experiments in this section compare the efficiency of the five algorithms on a set of 12 industrial instances that we randomly selected. Moreover, different values of T_{Metro} are analyzed, and Figure 3 shows the impact of T_{Metro} on the number of lots that are skipped. The curves correspond to the average number of lots skipped for Algorithms 2 and 5, depending on the value of T_{Metro}. A significant improvement is observed between Algorithm 2 and 5. This is mainly due to the fact that Algorithm 5 evaluates sets of lots for skipping contrary to Algorithm 2 which evaluates lots individually. In particular, when $T_{Metro} \in (0.049, 0.061)$, the average number of skipped lots with Algorithm 5 increases from 7.5 to 8.7, contrary to Algorithm 2 where the average number of skipped lots increases from 7.2 to 7.4. This result shows the importance of the setting of T_{Metro}. Moreover, increasing the number of skipped lots should be carefully considered because it is a consequence of the sampling algorithm, and it impacts the global risk of the factory and the available capacity at the inspection area. A first improvement of the analysis performed in this paper would consist in varying T_{Metro} in more details, which would probably result in a set of step curves.

Table 5 details the results for different values of T_{Metro} (0.001, 0.005, 0.01, 0.05 and 0.1). The lots selected by Algorithm 1 corresponds to the lots for which $LSI \leq T_{Metro}$ in the first iteration. Only these lots will be considered for skipping by the other algorithms. Hence, the set $S_{Skippable}$ is reduced using Algorithm 1. Let us focus on the results obtained with $T_{Metro} = 0.005$. It can be observed that, when the number of lots in $S_{Skippable}$ increases (more than 10 lots in this particular case), the heuristics (Algorithms 2, 3 and 4) are not efficient enough to find the best solution. Thus, the Branch and Bound algorithm (Algorithm 5) finds better solutions. This is the case for Instances 1 and 10 where a smaller value of *LSI* is determined with the same number of skipped lots. Besides, Algorithm 5 determines a solution for Instance 7 with more lots while respecting the value of T_{Metro}. In most of the cases, Algorithms 3, 4 and 5 give better results in terms of the final *LSI* than Algorithm 2. This is because, in Algorithm 2, lots are evaluated individually and the order of the set of lots highly influences the final decision. For example, let us consider two lots (L_x and L_y) that reduce the risk on the same tools, but L_x reduces the risk more than L_y. If L_x is evaluated

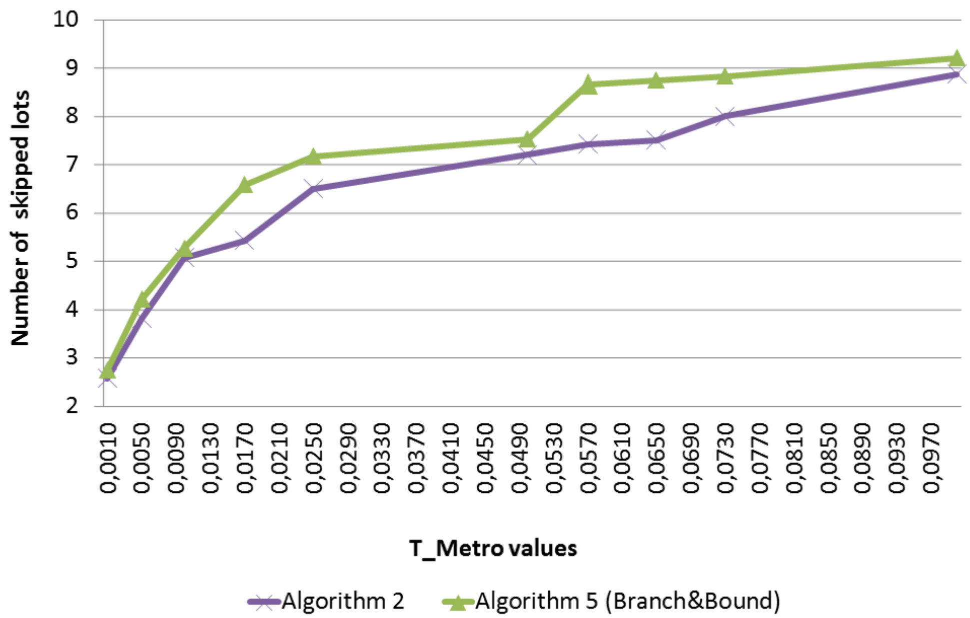

Figure 3: Average number of skipped lots depending on T_{Metro} for Algorithms 2 and 5.

first, it will be skipped because L_y is in the waiting queue. Then, when $LSI(L_y)$ is calculated, it will not be skipped because L_x is no longer in the waiting queue. When using a smaller value of T_{Metro}, there are two cases without lots in $S_{Skippable}$ (i.e. Instances 4 and 12), and one case with only one lot in $S_{Skippable}$ (i.e. Instance 8). In Instance 8, the final LSI is equal to zero, this means that if the identified lot is skipped, the risk will not be impacted. However, as in the example of Section 3, this does not mean that the selected lot for skipping will be selected when the value of T_{Metro} increases. The overall number of lots that are skipped significantly increases with the value of T_{Metro}.

Finally, let us summarize the highlights and drawbacks of the proposed algorithms. Algorithm 1 is the simplest but it cannot be used to take the final decision of skipping, it is used to reduce the set of lots which can be skipped, which reduces the calculation time for the other algorithms. With Algorithm 2, we can obtain solutions that satisfy the condition of T_{Metro}, but the order in which lots are considered highly influences the solution. The quality of solutions obtained with Algorithm 3 does not depend on how lots are ordered, but lots are still evaluated individually. With Algorithm 4, solutions are improved because lots are evaluated by sets and not individually. Finally, Algorithm 5 finds the optimal solution but the calculation time quickly increases if the number of lots in $S_{Skippable}$ increases.

Table 5: Number of skipped lots and final *LSI* for different values of T_{Metro}

Instance	TypeAlgo	**Number of skipped lots** (Related LSI)				
		T_{Metro}				
		0.001	0.005	0.01	0.05	0.1
1	Algorithm 1	**6** (0.001279)	**12** (0.284293)	**13** (0.291487)	**19** (0.435451)	**22** (0.722157)
	Algorithm 2	**5** (0.000653)	**7** (0.004930)	**7** (0.008011)	**10** (0.049931)	**14** (0.095832)
	Algorithm 3	**5** (0.000626)	**7** (0.004930)	**8** (0.008628)	**12** (0.045453)	**15** (0.098124)
	Algorithm 4	**5** (0.000626)	**7** (0.004930)	**8** (0.008628)	**12** (0.035398)	**15** (0.093280)
	Algorithm 5	**5** (0.000626)	**7** (0.004219)	**8** (0.007871)	**12** (0.035398)	**15** (0.093280)
2	Algorithm 1	**4** (0.070563)	**10** (0.131550)	**11** (0.167974)	**13** (0.194752)	**15** (0.312755)
	Algorithm 2	**3** (0.000882)	**5** (0.004762)	**7** (0.007275)	**10** (0.047746)	**11** (0.084170)
	Algorithm 3	**3** (0.000115)	**6** (0.003970)	**7** (0.006507)	**10** (0.046978)	**11** (0.083402)
	Algorithm 4	**3** (0.000115)	**6** (0.003970)	**7** (0.006507)	**10** (0.043171)	**11** (0.079595)
	Algorithm 5	**3** (0.000115)	**6** (0.003970)	**7** (0.006507)	**10** (0.043171)	**11** (0.079595)
3	Algorithm 1	**4** (0.002080)	**9** (0.099745)	**9** (0.099745)	**14** (0.209593)	**14** (0.209593)
	Algorithm 2	**1** (0.000989)	**3** (0.004692)	**6** (0.009018)	**9** (0.047786)	**12** (0.093494)
	Algorithm 3	**2** (0.000532)	**5** (0.003150)	**7** (0.007983)	**10** (0.045235)	**12** (0.091388)
	Algorithm 4	**2** (0.000532)	**5** (0.003150)	**7** (0.007983)	**10** (0.045235)	**12** (0.091388)
	Algorithm 5	**2** (0.000532)	**5** (0.003150)	**7** (0.007983)	**10** (0.045235)	**12** (0.091388)
4	Algorithm 1	**0** (0.000000)	**2** (0.005700)	**3** (0.013162)	**6** (0.084101)	**6** (0.084101)
	Algorithm 2	**0** (0.000000)	**1** (0.004455)	**2** (0.005700)	**4** (0.041514)	**6** (0.084101)
	Algorithm 3	**0** (0.000000)	**1** (0.001171)	**2** (0.005700)	**4** (0.032991)	**6** (0.084101)
	Algorithm 4	**0** (0.000000)	**1** (0.001171)	**2** (0.005700)	**4** (0.032991)	**6** (0.084101)
	Algorithm 5	**0** (0.000000)	**1** (0.001171)	**2** (0.005700)	**4** (0.032991)	**6** (0.084101)
5	Algorithm 1	**4** (0.001457)	**5** (0.005294)	**5** (0.005294)	**11** (0.166730)	**12** (0.227748)
	Algorithm 2	**3** (0.000493)	**4** (0.001457)	**5** (0.005294)	**7** (0.048881)	**7** (0.091766)
	Algorithm 3	**3** (0.000493)	**4** (0.001457)	**5** (0.005294)	**7** (0.038333)	**9** (0.081920)
	Algorithm 4	**3** (0.000493)	**4** (0.001457)	**5** (0.005294)	**7** (0.038333)	**9** (0.081920)
	Algorithm 5	**3** (0.000493)	**4** (0.001457)	**5** (0.005294)	**7** (0.038333)	**9** (0.081920)
6	Algorithm 1	**0** (0.000000)	**4** (0.008165)	**5** (0.017027)	**9** (0.125484)	**10** (0.211769)
	Algorithm 2	**0** (0.000000)	**2** (0.003708)	**4** (0.008165)	**6** (0.027115)	**8** (0.096540)
	Algorithm 3	**0** (0.000000)	**2** (0.003396)	**4** (0.008165)	**6** (0.027115)	**8** (0.083012)
	Algorithm 4	**0** (0.000000)	**2** (0.003396)	**4** (0.008165)	**6** (0.027115)	**8** (0.083012)
	Algorithm 5	**0** (0.000000)	**2** (0.003396)	**4** (0.008165)	**6** (0.027115)	**8** (0.083012)
7	Algorithm 1	**7** (0.001262)	**13** (0.396196)	**14** (0.401685)	**18** (0.522087)	**19** (0.595998)
	Algorithm 2	**6** (0.000512)	**7** (0.004225)	**8** (0.009715)	**12** (0.045763)	**14** (0.099489)
	Algorithm 3	**6** (0.000385)	**7** (0.001262)	**9** (0.009661)	**12** (0.038781)	**14** (0.081529)
	Algorithm 4	**6** (0.000385)	**7** (0.001262)	**9** (0.009661)	**12** (0.037227)	**14** (0.081529)
	Algorithm 5	**6** (0.000385)	**8** (0.004317)	**9** (0.008450)	**12** (0.037227)	**14** (0.081529)
8	Algorithm 1	**1** (0.000000)	**2** (0.003037)	**2** (0.003037)	**5** (0.091761)	**5** (0.091761)
	Algorithm 2	**1** (0.000000)	**2** (0.003037)	**2** (0.003037)	**4** (0.042092)	**5** (0.091761)
	Algorithm 3	**1** (0.000000)	**2** (0.003037)	**2** (0.003037)	**4** (0.042092)	**5** (0.091761)
	Algorithm 4	**1** (0.000000)	**2** (0.003037)	**2** (0.003037)	**4** (0.042092)	**5** (0.091761)
	Algorithm 5	**1** (0.000000)	**2** (0.003037)	**2** (0.003037)	**4** (0.042092)	**5** (0.091761)
9	Algorithm 1	**8** (0.001878)	**10** (0.007715)	**10** (0.007715)	**10** (0.007715)	**12** (0.087647)
	Algorithm 2	**4** (0.000447)	**6** (0.004665)	**10** (0.007715)	**10** (0.007715)	**12** (0.087647)
	Algorithm 3	**4** (0.000000)	**8** (0.004741)	**10** (0.007715)	**10** (0.007715)	**12** (0.087647)
	Algorithm 4	**4** (0.000000)	**8** (0.004741)	**10** (0.007715)	**10** (0.007715)	**12** (0.087647)
	Algorithm 5	**4** (0.000000)	**8** (0.004741)	**10** (0.007715)	**10** (0.007715)	**12** (0.087647)
10	Algorithm 1	**9** (0.111135)	**13** (0.120418)	**13** (0.12418)	**19** (0.488045)	**21** (0.682193)
	Algorithm 2	**6** (0.000985)	**9** (0.004572)	**11** (0.009628)	**13** (0.041486)	**13** (0.098187)
	Algorithm 3	**7** (0.000724)	**9** (0.003832)	**10** (0.006745)	**13** (0.029081)	**15** (0.077084)
	Algorithm 4	**7** (0.000724)	**9** (0.003832)	**11** (0.008887)	**13** (0.029081)	**15** (0.077084)
	Algorithm 5	**7** (0.000724)	**9** (0.003759)	**11** (0.008887)	**13** (0.029081)	**15** (0.077084)
11	Algorithm 1	**6** (0.001644)	**9** (0.29586)	**13** (0.128561)	**13** (0.128561)	**13** (0.128561)
	Algorithm 2	**4** (0.000670)	**7** (0.003612)	**8** (0.009209)	**11** (0.043792)	**12** (0.074949)
	Algorithm 3	**5** (0.000951)	**7** (0.003612)	**8** (0.006475)	**11** (0.043792)	**12** (0.074949)
	Algorithm 4	**5** (0.000951)	**7** (0.003612)	**8** (0.006475)	**11** (0.043792)	**12** (0.074949)
	Algorithm 5	**5** (0.000951)	**7** (0.003612)	**8** (0.006475)	**11** (0.043792)	**12** (0.074949)
12	Algorithm 1	**0** (0.000000)	**2** (0.003039)	**4** (0.024960)	**6** (0.074434)	**7** (0.124853)
	Algorithm 2	**0** (0.000000)	**2** (0.003039)	**2** (0.009072)	**4** (0.044788)	**6** (0.074434)
	Algorithm 3	**0** (0.000000)	**2** (0.003039)	**2** (0.009072)	**5** (0.047339)	**6** (0.074434)
	Algorithm 4	**0** (0.000000)	**2** (0.003039)	**2** (0.009072)	**5** (0.047339)	**6** (0.074434)
	Algorithm 5	**0** (0.000000)	**2** (0.003039)	**2** (0.009072)	**5** (0.047339)	**6** (0.074434)

5 CONCLUDING REMARKS AND PERSPECTIVES

In this paper, a new approach for managing the defect inspection queues is introduced. The objective is to identify lots that can be skipped with limited impacts on the overall risk in the fab. The risk considered is *the Wafer at Risk* ($W@R$) on process tools and a Dynamic Sampling system is used to select the lots for inspection. Various skipping algorithms were proposed and evaluated on industrial instances. The Add-Suppression and Branch-and-Bound (Algorithms 4 and 5) give the best solutions compared with Algorithms 1, 2 and 3. Nevertheless, the calculation time of Algorithm 5 quickly increases if the number of lots in the set $S_{Skippable}$ increases. The improvements brought by the best algorithms may look marginal, but the financial impact of an unobserved dysfunction dragged on for a while can be significant. Among the proposed algorithms, one has been implemented and is currently used in the defect inspection area of STMicroelectronics in Rousset.

Only the defect inspection area is considered in this study and future work will focus on the application of this mechanism to other metrology queues. An additional future research is the improvement of the skipping mechanism by considering the inspection time and the different inspection tool types. Defining an objective function using the inspection time allows for more adapted solutions to specific situations. We expect that this modification would imply skipping as many or less lots for similar $W@R$ reduction, while it will help to balance the workload of inspection tools. In addition, it could be interesting to dynamically choose the value of the threshold T_{Metro}. This could be done by fixing the number of lots to skip and defining other algorithms to find the least risky set of lots to skip considering the maximum number of skipped lots as a constant.

6 ACKNOWLEDGMENTS

This study has been done within the framework of a joint collaboration between STMicroelectronics in Rousset, France, and the Center of Microelectronics in Provence of the Ecole des Mines de Saint-Etienne in Gardanne, France. This research is part of the project MAGE, financed by the ”Programme d'Investissement d'Avenir” of the French Government. The authors would also like to thank the ANRT (Association Nationale de la Recherche et de la Technologie) which has partially financed this work. Finally, the authors are grateful to Eric Tartière, leader of the $W@R$ implementation group at ST Rousset, for his constructive comments.

REFERENCES

Dauzère-Pérès, S., J.-L. Rouveyrol, C. Yugma, and P. Vialletelle. 2010. "A Smart Sampling Algorithm to Minimize Risk Dynamically". In *Proceedings of the 2010 IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, 307–310.

Good, R. P., and M. A. Purdy. 2007. "An MILP Approach to Wafer Sampling and Selection". *IEEE Transactions on Semiconductor Manufacturing* 20 (4): 400–407.

Leachman, R., and S. D. Ding. 2011. "Excursion Yield Loss and Cycle Time Reduction in Semiconductor Manufacturing". *Transactions on Automation Science and Engineering* 8 (1): 112–117.

Lee, J. H., S. J. You, and S. C. Park. 2001. "A New Intelligent SOFM-Based Sampling Plan for Advanced Process Control". *Expert Systems with Applications* 20 (2): 133–151.

May, G. S., and C. J. Spanos. 2006. "Fundamentals of Semiconductor Manufacturing and Process Control". John Wiley & Sons.

Mouli, C. 2005. "Adaptive Sampling Technology - the next step to factory efficiency". *EuroAsia Semiconductor Magazine*:1–3.

Nduhura Munga, J. 2012. *Implementing and Optimizing Dynamic Control Plans in Semiconductor Manufacturing*. Ph. D. thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, Gardanne, France.

Nduhura Munga, J., G. Rodriguez-Verjan, S. Dauzère-Pérès, C. Yugma, P. Vialletelle, and J. Pinaton. 2013. "Literature Review on Sampling Techniques in Semiconductor Manufacturing". In *IEEE Transactions on Semiconductor Manufacturing*.

Purdy, M., C. Nicksic, and K. Lensing. 2005. "Method for Efficiently Managing Metrology Queues". In *Proceedings of the 2005 IEEE International Symposium on Semiconductor Manufacturing*, 71–74.

R Development Core Team 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Sahnoun, M., S. Bassetto, S. Bastoini, and P. Vialletelle. 2011. "Optimisation of the Process Control in a Semiconductor Company: Model and Case Study of Defectivity Sampling". *International Journal of Production Economics* 49 (13): 3873–3890.

Song-bor, L., R. L. Ta-Yung, L. Janson, and C. Yu-Ching. 2003. "A Capacity-Dependence Dynamic Sampling Strategy". In *Proceedings of the 2003 IEEE International Symposium on Semiconductor Manufacturing*, 312–314.

Sullivan, D. B., E. W. Conrad, and J. S. Smyth. 2004. "Overlay Metrology Sampling Capability Analysis and Implementation in Manufacturing". In *IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, 208–212.

Tirkel, I., and G. Rabinowitz. 2012. "The relationship between yield and flow time in a production system under inspection". *International Journal of Operation Research* 50 (14): 3686–3697.

AUTHOR BIOGRAPHIES

GLORIA LUZ RODRIGUEZ VERJAN Obtained her MSc degree in industrial engineering from the Ecole des Mines de Saint-Etienne, France. She is currently working as Research Engineer at a Franco-Italian semiconductor manufacturer, and working towards a doctoral degree within the Center of Microelectronics in Provence (CMP) of the Ecole des Mines de Saint-Etienne. From 2006 to 2009 she worked as logistics analyst at P&A Renault located in Bogotá, Colombia. She has also been part-time lecturer in simulation at the Ecole des Mines de Saint-Etienne in France, Sabana University and Javeriana University in Colombia. Her research work has been focused in simulation and logistics operations, and most recently in optimization of semiconductor manufacturing operations. Her email address is rodriguez@emse.fr or gloria.rodriguez@st.com.

STEPHANE DAUZERE-PERES is Professor and Director of the Center of Microelectronics in Provence of the EMSE. He received the Ph.D. degree from the Paul Sabatier University in Toulouse, France, in 1992; and the H.D.R. from the Pierre and Marie Curie University, Paris, France, in 1998. He was a Postdoctoral Fellow at the Massachusetts Institute of Technology, U.S.A., in 1992 and 1993, and Research Scientist at Erasmus University Rotterdam, The Netherlands, in 1994. He has been Associate Professor and Professor from 1994 to 2004 at the Ecole des Mines de Nantes in France where he headed the team Production and Logistic Systems (about 20 members) between 1999 and 2004. He was invited Professor at the Norwegian School of Economics and Business Administration, Bergen, Norway, in 1999. Since March 2004, he is Professor at the Ecole des Mines de Saint-Etienne, where he headed the research department Manufacturing Sciences and Logistics (SFL, about 20 members) from 2004 to 2013. His research interests broadly include modeling and optimization of operations at various decision levels (from real-time to strategic) in manufacturing and logistics, with a special emphasis on semiconductor manufacturing. He has published 45 papers in international journals and contributed to more than 100 communications in conferences. Stéphane Dauzère-Pérès has coordinated multiple academic and industrial research projects, and also five conferences. His email address is Stephane.Dauzere-Peres@emse.fr.

SYLVAIN HOUSSEMAN is a PostDoc fellow at the Center of Microelectronics in Provence of the Ecole des Mines de Saint-Etienne, where he received his Ph.D. degree in 2011. His research work focuses on applied optimization and simulation techniques for logistics and scheduling. His email address is

sylvain.housseman@gmail.com or sylvain.housseman@mines-stetienne.fr.

JACQUES PINATON is manager of Process Control System group at ST Microelectronics Rousset, France. He is an engineer in metallurgy from the Conservatoire National des Arts et metiers dAix en Provence. He joined ST in 1984. After 5 years in the process engineering group, he joined the device department to implement SPC and Process Control methodology and tools. He participated in the start up of 3 generations of fabs. He is leading various Rousset R&D programs on manufacturing science including programs on automation, APC, and diagnostics. His E-mail address is jacques.pinaton@st.com.