# Risk minimization through metrology in semiconductor manufacturing

Juan Alejandro Sendon Perez

N° d'ordre NNT : 2017LYSEM022

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON

opérée au sein de

l'École des Mines de Saint-Étienne

**École Doctorale** N° 488

**Sciences, Ingénierie, Santé**

**Spécialité de doctorat** : Génie Industriel

Soutenue publiquement le 15/09/2017, par :

**Alejandro Sendón Pérez**

---

# Risk minimization through metrology in semiconductor manufacturing

---

Devant le jury composé de :

| | |
|---|---|
| Claude Yugma, Maître de recherche, EMSE | Président |
| Nathalie Sauer, Professeur, Université de Lorraine | Rapporteure |
| Michel Tollenaere, Professeur, Institut Polytechnique de Grenoble | Rapporteur |
| Safia Kedad-Sidhoum, Maître de conférence, Université Pierre et Marie Curie | Examinatrice |
| Jacques Pinaton, Ingénieur, STMicroelectronics | Examinateur |
| Stéphane Dauzère-Pérès, Professeur, EMSE | Directeur de thèse |

| Spécialités doctorales | Responsables : | Spécialités doctorales | Responsables |
|---|---|---|---|
| SCIENCES ET GENIE DES MATERIAUX | K. Wolski Directeur de recherche | MATHEMATIQUES APPLIQUEES | O. Roustant, Maître-assistant |
| MÉCANIQUE ET INGENIERIE | S. Drapier, professeur | INFORMATIQUE | O. Boissier, Professeur |
| GENIE DES PROCEDES | F. Gruy, Maître de recherche | SCIENCES DES IMAGES ET DES FORMES | JC. Pinoli, Professeur |
| SCIENCES DE LA TERRE | B. Guy, Directeur de recherche | GENIE INDUSTRIEL | X. Delorme, Maître assistant |
| SCIENCES ET GENIE DE L'ENVIRONNEMENT | D. Graillot, Directeur de recherche | MICROELECTRONIQUE | Ph. Lalevée, Professeur |

## EMSE : Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'Etat ou d'une HDR)

| | | | | |
|---|---|---|---|---|
| ABSI | Nabil | CR | Génie industriel | CMP |
| AUGUSTO | Vincent | CR | Image, Vision, Signal | CIS |
| AVRIL | Stéphane | PR2 | Mécanique et ingénierie | CIS |
| BADEL | Pierre | MA(MDC) | Mécanique et ingénierie | CIS |
| BALBO | Flavien | PR2 | Informatique | FAYOL |
| BASSEREAU | Jean-François | PR | Sciences et génie des matériaux | SMS |
| BATTON-HUBERT | Mireille | PR2 | Sciences et génie de l'environnement | FAYOL |
| BEIGBEDER | Michel | MA(MDC) | Informatique | FAYOL |
| BLAYAC | Sylvain | MA(MDC) | Microélectronique | CMP |
| BOISSIER | Olivier | PR1 | Informatique | FAYOL |
| BONNEFOY | Olivier | MA(MDC) | Génie des Procédés | SPIN |
| BORBELY | Andras | MR(DR2) | Sciences et génie des matériaux | SMS |
| BOUCHER | Xavier | PR2 | Génie Industriel | FAYOL |
| BRODHAG | Christian | DR | Sciences et génie de l'environnement | FAYOL |
| BRUCHON | Julien | MA(MDC) | Mécanique et ingénierie | SMS |
| CAMEIRAO | Ana | MA(MDC) | Génie des Procédés | SPIN |
| CHRISTIEN | Frédéric | PR | Science et génie des matériaux | SMS |
| DAUZERE-PERES | Stéphane | PR1 | Génie Industriel | CMP |
| DEBAYLE | Johan | CR | Sciences des Images et des Formes | SPIN |
| DEGEORGE | Jean-Michel | MA(MDC) | Génie industriel | Fayol |
| DELAFOSSE | David | PR0 | Sciences et génie des matériaux | SMS |
| DELORME | Xavier | MA(MDC) | Génie industriel | FAYOL |
| DESRAYAUD | Christophe | PR1 | Mécanique et ingénierie | SMS |
| DJENIZIAN | Thierry | PR | Science et génie des matériaux | CMP |
| DOUCE | Sandrine | PR2 | Sciences de gestion | FAYOL |
| DRAPIER | Sylvain | PR1 | Mécanique et ingénierie | SMS |
| FAUCHEU | Jenny | MA(MDC) | Sciences et génie des matériaux | SMS |
| FAVERGEON | Loïc | CR | Génie des Procédés | SPIN |
| FEILLET | Dominique | PR1 | Génie Industriel | CMP |
| FOREST | Valérie | MA(MDC) | Génie des Procédés | CIS |
| FRACZKIEWICZ | Anna | DR | Sciences et génie des matériaux | SMS |
| GARCIA | Daniel | MR(DR2) | Sciences de la Terre | SPIN |
| GAVET | Yann | MA(MDC) | Sciences des Images et des Formes | SPIN |
| GERINGER | Jean | MA(MDC) | Sciences et génie des matériaux | CIS |
| GOEURIOT | Dominique | DR | Sciences et génie des matériaux | SMS |
| GONDRAN | Natacha | MA(MDC) | Sciences et génie de l'environnement | FAYOL |
| GONZALEZ FELIU | Jesus | MA(MDC) | Sciences économiques | FAYOL |
| GRAILLOT | Didier | DR | Sciences et génie de l'environnement | SPIN |
| GROSSEAU | Philippe | DR | Génie des Procédés | SPIN |
| GRUY | Frédéric | PR1 | Génie des Procédés | SPIN |
| GUY | Bernard | DR | Sciences de la Terre | SPIN |
| HAN | Woo-Suck | MR | Mécanique et ingénierie | SMS |
| HERRI | Jean Michel | PR1 | Génie des Procédés | SPIN |
| KERMOUCHE | Guillaume | PR2 | Mécanique et Ingénierie | SMS |
| KLOCKER | Helmut | DR | Sciences et génie des matériaux | SMS |
| LAFOREST | Valérie | MR(DR2) | Sciences et génie de l'environnement | FAYOL |
| LERICHE | Rodolphe | CR | Mécanique et ingénierie | FAYOL |
| MALLIARAS | Georges | PR1 | Microélectronique | CMP |
| MOLIMARD | Jérôme | PR2 | Mécanique et ingénierie | CIS |
| MOUTTE | Jacques | CR | Génie des Procédés | SPIN |
| NEUBERT | Gilles | | | FAYOL |
| NIKOLOVSKI | Jean-Pierre | Ingénieur de recherche | Mécanique et ingénierie | CMP |
| NORTIER | Patrice | PR1 | Génie des Procédés | SPIN |
| O CONNOR | Rodney Philip | MA(MDC) | Microélectronique | CMP |
| OWENS | Rosin | MA(MDC) | Microélectronique | CMP |
| PERES | Véronique | MR | Génie des Procédés | SPIN |
| PICARD | Gauthier | MA(MDC) | Informatique | FAYOL |
| PIJOLAT | Christophe | PR0 | Génie des Procédés | SPIN |
| PINOLI | Jean Charles | PR0 | Sciences des Images et des Formes | SPIN |
| POURCHEZ | Jérémy | MR | Génie des Procédés | CIS |
| ROUSSY | Agnès | MA(MDC) | Microélectronique | CMP |
| ROUSTANT | Olivier | MA(MDC) | Mathématiques appliquées | FAYOL |
| SANAUR | Sébastien | MA(MDC) | Microélectronique | CMP |
| STOLARZ | Jacques | CR | Sciences et génie des matériaux | SMS |
| TRIA | Assia | Ingénieur de recherche | Microélectronique | CMP |
| VALDIVIESO | François | PR2 | Sciences et génie des matériaux | SMS |
| VIRICELLE | Jean Paul | DR | Génie des Procédés | SPIN |
| WOLSKI | Krzystof | DR | Sciences et génie des matériaux | SMS |
| XIE | Xiaolan | PR1 | Génie industriel | CIS |
| YUGMA | Gallian | CR | Génie industriel | CMP |

Mise à jour : 03/02/2017

## Acknowledgements

After these last years, full of good memories, I would like to thank those people who have accompanied me and taught me throughout this wonderful experience. First of all, I would like to thank Stéphane Dauzère-Pérès and Jacques Pinaton for giving me the opportunity to work with them in this very interesting topic. Stéphane, thank you for sharing your wisdom with me and for teaching me how to find the right solutions during the thesis. Jacques, thank you for always helping me to grow as a professional through your experience and your advices.

I would like to thank all people of the SFL department, because I found a great atmosphere there for doing the thesis. It was very enriching share with you so many moments. I have enjoyed all our discussions about the different aspects of our countries during the *pause cafés* and the meals at the faboulous *cantine*. I do not want to forget either the SFL seminars where we had good times together in the best places of the region.

Of course, I would also like to thank all members of the Process Control department at STMicroelectronics for all your support and advices along this period, and of course for making my French better every day. From the PhDs/*stagiaires* to the permanent members of the department. In particular, I would like to thank Eric Tartiere for all the hours invested in discussions and for having always time for helping me.

During these years there was also time for fun and we have had a great time with a fantastic group in Aix. Thank you Laura, Carlos, Dario, Vinz, Babeth, Diego, Stefania, Jakey, Sebastian and Elisabeth for all the laughts together, the conversations, the beers, the trips and the parties at "Sunset".

Special mention to my family and my loved ones for all your support.

*Gracias papá y mamá por haberme apoyado desde pequeñito y estar siempre ahí. Si no fuera por vosotros, hoy no sería la persona que soy.*

*Y por supuesto, GRACIAS a ti María, porque sin ti nada de esto hubiera sido igual, gracias por ayudarme y aconsejarme cada día, por tu apoyo incondicional, por sacarme siempre una sonrisa, por hacerme mejor persona... pero sobre todo, gracias por salir de casa aquella tarde de Septiembre.*

Thank you very much. Merci beaucoup. Muchas gracias.

**Alejandro**
**San Sebastián, Noviembre 2017**

# Contents

# List of Tables

# List of Figures

# General Introduction

The complexity in semiconductor manufacturing processes is continuously growing. This is why, control steps are added in order to guarantee high product quality. To manufacture an integrated circuit (IC), many process steps are required and most of them are repeated, leading to cycle times of more than two months. Besides, the continuous reduction of the size of ICs and the variety of product mixes increases even more this complexity. Despite the necessity of control steps, most of them can be considered as unnecessary and therefore as operations with non-added value that increase final product costs. The objective of semiconductor companies is to find a balance between performing the right number of control operations and achieving a high yield.

This thesis consists in analyzing the different properties of metrology workshops, proposing novel approaches to optimize sampling rates and developing new dynamic strategies for risk reduction during semiconductor fabrication. In semiconductor manufacturing, many types of control exist. In this thesis the main metrology types are explained. A review of the different ways to manage risk is proposed and the types of sampling techniques found in the literature are detailed. A thorough analysis of metrology workshops in the site of Rousset of STMicroelectronics has been carried out. Their physical properties and also their characteristics in terms of measure qualification, sampling, lot dispatching strategy, risk levels and so on are considered.

Also, a new procedure was developed that helps to determine which sampling strategy fits better according to the metrology characteristics and risk values. Results showed that unnecessary measures can be avoided if a sampling strategy associated only to process machines with a constant sampling ratio is adopted for simple metrology systems (those that measure a wafer property which is not linked to the type of product). One part of the procedure for the sampling strategy change suggests to calculate new ratios for the new sampling method to be implemented. In this thesis, new approaches are proposed to optimize the sampling rates for different types of metrology tools respecting the metrology capacity and taking into account parameters such as throughput rates of process machines and metrology tools, and the probabilities of failure. The results showed the metrology capacity is better used and the method controls well process machines depending on their characteristics, paying more attention to those that are critical.

In the final part of the thesis, simulation models of several metrology workshops have been developed. These models reproduce the behavior of the workshops to better understand them and to evaluate the impact of new improvements.

The manuscript is organized as follows:

1. **Chapter 1** is dedicated to the industrial context. The terminology associated to semiconductor manufacturing is introduced. The history of the semiconductor industry is presented, the manufacturing process steps and the control techniques are introduced. The problem and the thesis objectives are also discussed.

2. **Chapter 2** defines the term risk in a semiconductor environment. The methods to manage the risks in semiconductor manufacturing are described and the risk indicator used in this thesis is explained. A review of the sampling techniques in semiconductor manufacturing is carried out, and our problems are positioned in the literature.

3. **Chapter 3** contains an analysis of all metrology workshops in the Rousset factory of STMicroelectronics which takes into account their main properties. An approach to analyze the queue times is proposed. A general procedure to perform a sampling strategy change is provided and numerical results are discussed.

4. **Chapter 4** details novel approaches to optimally assign a sampling rate, satisfying metrology capacity, to each process machine in a given workshop. The mathematical models that were developed are detailed and numerical results are discussed. Also, some industrial results and an industrial implementation are presented.

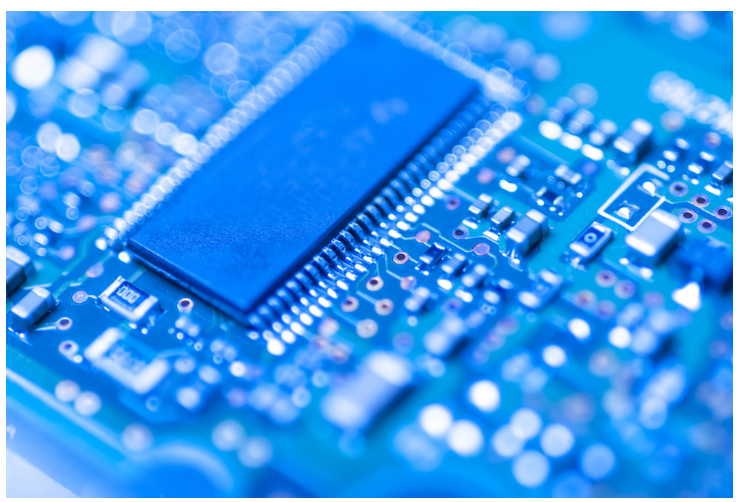
5. **Chapter 5** presents simulation models developed to study the behavior of some metrology workshops in semiconductor manufacturing fabs with the objective of controlling the risk on process machines. One model is dedicated to the Ion Implant workshop and the other model to compare dispatching and sampling policies in the thickness metrology workshop.

The manuscript ends with some general conclusions and perspectives for further work.

## Chapter 1

# Industrial Context

*This chapter provides the terminology associated to semiconductor manufacturing used in this manuscript. It includes history of the semiconductor industry, the manufacturing process steps and the control techniques used to ensure that process machines are properly working. Finally, the problem and the thesis objectives are described.*



## 1.1 Introduction

Nowadays, we are surrounded by electronic devices. From our daily life to our work environment the presence of elements using Integrated Circuits (ICs) is raising more and more. This is translated into an increase of the demand for the semiconductor industry that implies producing more products and reducing cycle times.

The continuing trend of IC complexity [47] [33] since the 1959, the year of the planar silicon transistor invention, has led to manufacture devices with a smaller size each year. Throughout the years, more process controls has been added to ensure the quality of the ICs, to achieve a good yield and to ensure that machines are working properly because of the reduction scale, reaching more than 700 process steps during a cycle time of around 2 months to get the final product.

Different kinds of controls are established along the route of process operations, and their characteristics depend on the nature of the process. Not efficiently using of these controls has a negative impact on product cycle times and creates large risk values to produce defect products. Novel techniques are required to reduce risk while keeping the standards of quality. The aim is to continue controlling production machines and at the same time to better use control operations.

This first chapter presents the field of semiconductor manufacturing, the process steps involved and the metrology tools used. Section 1.2 introduces the process methods in integrated circuit fabrication. In Section 1.3, the metrology systems used to control the production are introduced. Sections 1.4 and 1.5 focus on the description of the problem and the scientific objectives of the thesis respectively.

## 1.2 Semiconductor manufacturing

The semiconductor industry is based on the production of ICs which are devices made by connecting a great quantity of electronic components such as transistors, diodes, resistors and so on that are built over a silicon surface. Silicon is a semiconductor material with the property of conducting electricity or not depending on the treatments performed on it.

Since its birth, this industry has experimented a huge growth, becoming one of the largest industries in the world. The semiconductor market revenue for 2015 was estimated around U.S. $341 billion, and since 1990, maintains a 7.6% compounded annual growth rate [4].

The incredible evolution of the semiconductor industry was related to its capability of doubling the number of transistors on an IC per year and reducing the cost per transistor as well, following the well known **Moore's Law** [68] in 1965. Ten years later, Gordon E. Moore reformulated his law claiming that the number of transistors per IC will double every 18 months. Nowadays, due to physical limitations this law is becoming no longer valid and to continue this trend, scientists and engineers face new technological challenges [52].

### 1.2.1 Integrated circuits

An integrated circuit (IC) is a collection of electronic circuits, the principal component is the **transistor**. Nevertheless, an IC could not function without much simpler components as the diodes, resistances and capacitors. The essential attribute of the transistor comes from its capability to produce a stronger output signal compared to the input signal. Also it can be used as a controller switching on or off the current in a circuit. There are two main types of transistors, the bipolar which is a current-driven device, and the MOS which is voltage-driven. Most of the semiconductor companies manufacture integrated circuits composed by MOS transistors and in the Rousset site of STMicroelectronics, only MOS transistors are used. In Figure 1.1, packages containing a semiconductor die are shown with nanowires attached that allows the electrical connections.

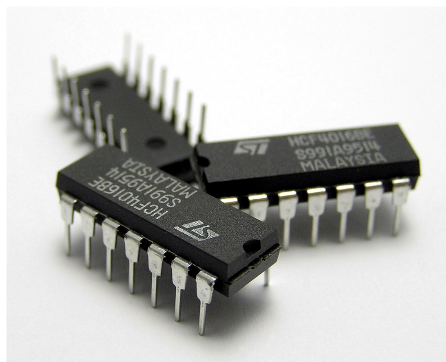



**Figure 1.1 –** *A group of packages containing IC chips inside.*

A MOS device comprises four terminals: A gate, a drain, a source and the body or substrate. In the last 40 years, the dimensions of the MOSFET (Metal Oxide Semiconductor Field Effect Transistor) have shrunk from a gate length of 5 $\mu$m (1970s) to less than 32 nm today in some industries, and are forecast to achieve values lower than 10 nm in about 2020. At the beginning of the semiconductor devide development (1950s), the bipolar junction transistor was dominant, but in the 1960s due to the development of CMOS (Complementary MOS) devices, with the advantage of a very low power dissipation , the MOSFET became the most selected device type in ICs. The size of the wafer was increased in the late 1960s because of the inclusion of more expensive and sophisticated process equipment. As explained before, in the 1970s, the MOSFET is introduced as a new device for the VLSI (Very Large Scale Integration). In the early 1980s, quality started to be a capital issue since the rise to leadership of the Japanese semiconductor industry, which increased the manufacture yield and factory efficiency. The size of the wafer was again augmented to reduce the cost of IC manufacturing and in parallel the not stoppable progression of the technology to a smaller feature size carried on. During the 1980s and 1990s, the yield was increased thanks to the automation of manufacturing equipment, the particle reduction and a more efficient management of the factories. Since the 1990s, the collaboration among semiconductor companies has increased in order to keep improving the problems related to the manufacturing devices that will continue reducing the feature size [61].

### 1.2.2 Manufacturing steps

The manufacturing of integrated circuits (ICs) starts with previous phase of the wafers preparation by slicing a silicon ingot. It is sliced into wafers with a 0.75 mm thickness approximately keeping a circular shape to minimize losses due to wafer handling in fabs. The sliced wafers are polished to have the surface as flat as possible. The manufacturing stages for the IC fabrication are commonly divided into two main stages:

1. **Front-End** processing: This stage corresponds to the wafer fabrication. It is divided into two major stages: Front-End Of Line (FEOL), where the active and passive parts of the components are fabricated (i.e. transistors, capacitors, etc.) and the Back-End Of Line (BEOL), where the metal components connect the devices and different layers [103]. This stage consists in multiple process steps that are continuously repeated in the production route.

2. **Back-End** processing: During this second stage, each die of the wafer is electronically tested, cut and separated. The bad dies will be scrapped in the next packaging step. The others will be packaged, their connections wire-bounded and a final test will be performed to validate the specifications.

The work of this thesis is exclusively focused on the Front-End processing. The surface of the wafer is covered layer by layer until the IC is completed. The number of layers depends on the type of product technology, the more complex technologies can reach up to 40 layers which is equivalent to more than 400 sequential process operations. The principal manufacturing steps are:

- *Oxidation*: The wafers are heated to high temperatures in presence of $O_2$, which produces the growth of a layer of Silicon Dioxide ($SiO_2$).

- *Deposition*: A multitude of thin film sheets of different materials are deposited on the wafer surface by means of several processes. The two more important are Chemical Vapor Deposition (CVD) and Physical Vapor Deposition (PVD). There are also the Plasma Enhanced Chemical Vapor Deposition (PECVD) which uses plasma techniques to produce the chemical reactions, and more recently, the atomic layer deposition (ALD).

- *Photolithography*: It is in charge of marking the patterns on the wafer. The wafer is completely coated with a film of light-sensitive photoresist polymer. Then a photo mask or reticle is placed just above the wafer and an ultraviolet light passes through the mask polymerizing the unhidden sections of the wafer surface. As a final stage, these sections are removed from the surface to continue developing the pattern.

- *Etch*: In this step the sections of the layer a the surface of the wafer not covered by the remaining pattern of the photolithography step are removed. There are two fundamental etching types: Wet etch which uses liquid chemicals and dry etch which uses plasma techniques.

- *Planarization*: To get a flat layer and eliminate undesirable substances, the wafer surface is polished. The most common planarization technique is Chemical Mechanical Polishing (CMP).

- *Ion implantation or doping*: In order to change the electrical properties of some regions, an ion bombardment is performed to alter the charge conditions of the material. Different types of dopants, energy and dose parameters are used.

Figure 1.2 shows the main process stages in the Front-end processing. The process flow is not linear but re-entrant. A lot comes back many times to the different manufacturing steps, repeating some process operations layer after layer on the wafer.

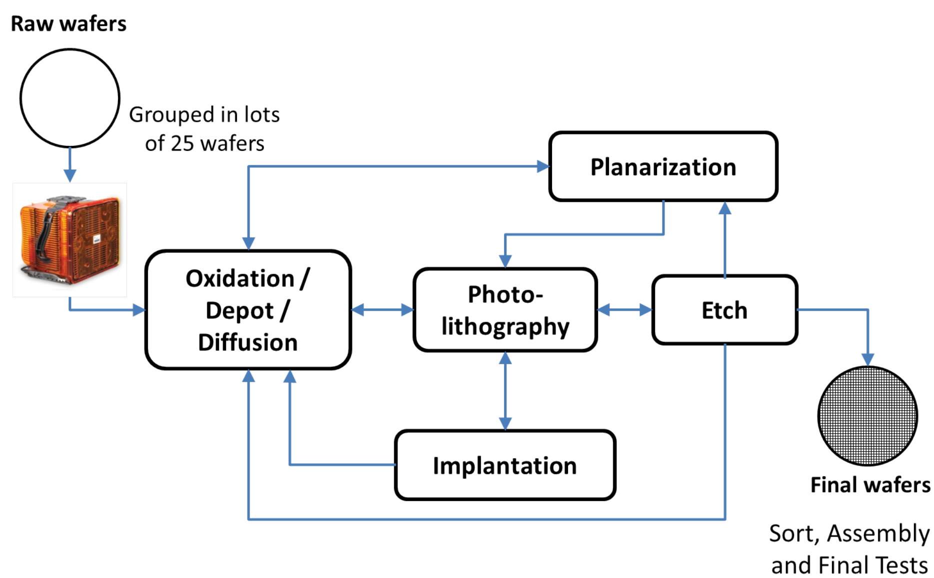



**Figure 1.2 –** *Front-end process steps [51].*

## 1.3 Control methods in semiconductor manufacturing

Due to the time invested to complete the whole process phase of chip manufacturing, the use of controls is required and every year the source of defects is gradually decreasing. To ensure its quality, it is necessary to control the state of the products (physical and electrical properties) once the process operation has finished, thereby validating that the result of the measure of the product is the expected one as well as, the good condition of the production machine. The scientific discipline responsible for measurement is called **Metrology**. Multiple data treatment techniques which, process the information obtained from wafer measurement and those coming from process equipment lead to raise yield improvement. These techniques are known as **Process Control** techniques and they lead to the decrease in the number of excursions, correct process drifts and reduce variability.

### 1.3.1 Process control techniques

There are several types of process controls and their aim is to reduce yield loss by controlling the machine and the process to guarantee that the expected results are achieved. Depending on the source of yield losses the control method can be divided in three categories [61]:

- Abnormality control methods: Based on finding unusual process behaviors and correcting them.
- Compensation (target tracking) control methods: Process behavior variation is expected, but not desirable.
- Advanced Process Control (APC): Combine both (compensation and abnormality detection).

Compensation control assumes a random noise in the expected non-random variation, while abnormality control detects abnormal variations (faults) in presence of normal variation that must be corrected. In APC there is an expected systematic variation and an unexpected variation. As for systematic variation the compensation is assumed and as soon as an unexpected variation is observed an alarm is generated to signal the situation. The major benefit of APC is the increment of the throughput which results in cost reductions, this increment is achieved thanks to the combination of on-line metrology speed-ups (for instance, the measurement of the wafer while processing) and the elimination of operator waiting times (e.g., performing and inspecting metrology results manually) [54].

The Table 1.1, shows the control methods considered in the previous categories explained above [61].

**Table 1.1 –** *Types of process control methods.*

| Abnormality control methods | Compensation control methods |
|---|---|
| Statistical process control (SPC) | Model-based process control |
| Statistical process monitoring (SPM) | Sensor-based process control |
| Multivariate SPC or SPM | Engineering process control |
| Real-time SPC | Algorithmic SPC |
| Equipment monitoring | Automatic process control |
| Excursion detection and control | Automated process control |
| Fault detection and classification (FDC) | Feedback/Feedforward control |
| Fault isolation | model predictive control |
| Fault prognosis | Run-to-run (R2R) control |
| Diagnosis | Real-time control (within a batch) |

The main process control techniques are summarized and briefly explained:

1. **Statistical Process Control (SPC)** is a method which ensures the stability of the process through statistical tools. According to the process state given by a SPC several actions can be taken from adjusting process parameters to stopping the production machine. SPC has continuously been evolving along the years with manufacturing technology changes [86].

2. **Fault Detection and Classification (FDC)** consists in monitoring the variations of the process by means of statistical models using parameters of the process machines (temperature, pressure, gas flow and so on) collected in real-time. As soon as an unexpected variation (fault) is detected the process machine is interrupted and corrective actions are taken.

3. **Run-to-run (R2R)** is a loop control technique that rectifies the process deviation from the defined target. Using equipment and lot information obtained from a run, the controller modifies the recipe parameters for the next run [54]. There are two modes for the control loops: Feed-Forward and Feed-Back. The first one, adjusts the parameters based on the results of a previous step, and the second one, adjusts the parameters using the results of the preceding run. A schematic model of a feed-back run-to-run controller used to CMP processes is presented in Figure 1.3.

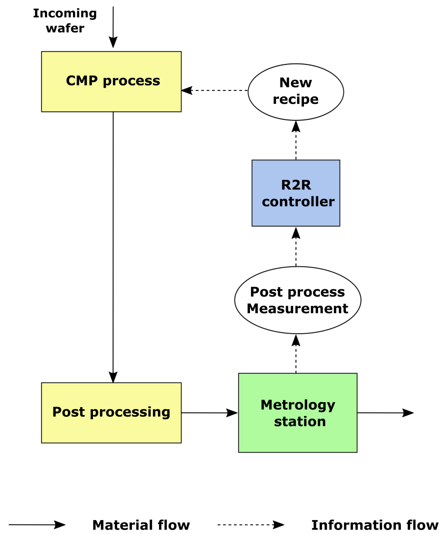



**Figure 1.3 –** *Model of a feed-dback R2R controller for CMP processes [32].*

4. **Virtual Metrology (VM)** is a technique based on predicting of the measurements of the wafer. It consists in predictive models generated using parameters collected by sensors such as temperature, pressure, and so on which forecast the electrical and physical parameters of the wafer [13]. The aim is to reduce the number of direct measurements and their substitution for virtual measures which can notify earlier when a process starts to be drifting.

### 1.3.2 Types of metrology systems

The unique way to ensure that a machine is processing fully is to check the physical properties of the product made. In semiconductor manufacturing the physical characteristics of the wafer are verified. According to these verifications, corrections are taken to straighten process machine behavior if necessary. The products built in semiconductor manufacturing go through sequential process operations such as oxidation, annealing, photolithography, etching, ion implantation, chemical mechanical polishing, and so on, where a large number of layers are deposited and at the end resulting in integrated circuits.

The physical properties reviewed are for instance film thickness, layer alignment (overlay measure), the level of particles or scratches on the surface, etc., but other parameters are also taken into account such as electrical parameters (resistivity of the sheet) or implant doses; an example of some process steps and their associated measure operations is shown in Figure 1.4. On one hand, some of the measurements can be done through human visual inspection, but on the other hand sophisticated tools provide the results of the measure.

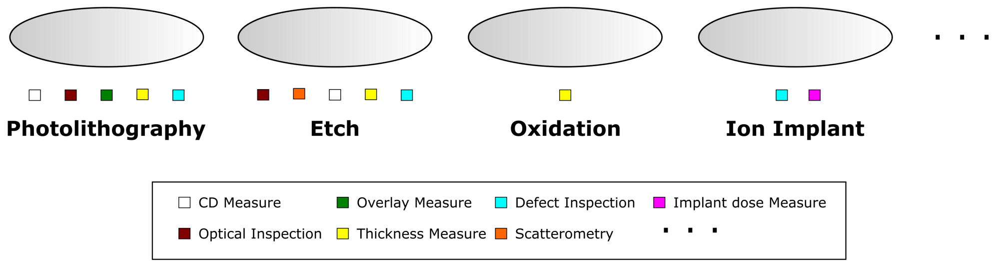



**Figure 1.4 –** *Example of some process steps and its respective metrology operations.*

In this thesis, only in-line metrology will be treated, which comprehends the physical measurements performed on the wafer that control IC manufacturing (transistor and interconnect fabrication). Metrology systems can be classified as either measurements performed in-situ, in-line or off-line [87] [35] [48] [79] [105]. In-situ metrology is the measure done by a sensor integrated inside the process chamber (real time measurement). In-line metrology is referred to the measures performed inside the fab during the production by a metrology tool independent of the process equipment. Off-line metrology is related to the measures done outside of the fab, for instance in an external laboratory.

The nature of metrology systems may impact on management constraints. For instance, the time to obtain measurement results and thus to take corrective actions in case of process drifts is considered short for in-situ metrology (measurement during processing), medium or quite large for in-line metrology (depending on the type of measure and the metrology workstation allocation) and considered the largest for off-line metrology (a long time between process and measurement). In another categorization for metrology systems based on in-situ metrology are referred to as “integrated tools” and in-line or off-line metrology are called “stand-alone tools”.

Metrology tool classification is inspired by the one proposed in [48], classified either by measurements performed on blanket thin films, measurements done on patterned thin films or measurement tools for particle and defect inspection. Blanket thin films measurement refers to wafers that, after being processed, are uniformly coated with a thin film, in a different way patterned thin films measurement are wafers which film has been previously patterned using photolithography; both are represented in Figure 1.5.

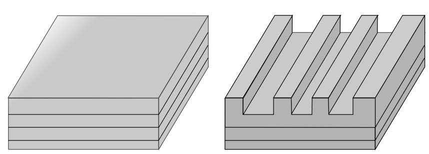


**Figure 1.5 –** *Blanket and patterned thin film representation.*

#### 1.3.2.1 Blanket thin film

This section comprises the metrology tools which perform the measure on wafers layered by a thin film.

1. **Interferometry** or also sometimes called reflectometry, is a widely employed optical technique used to measure film thickness of the top layer of a whole stack and to determine optical constants. It is used for CMP process operations, after taking off part of a blanket film. Normally interferometry is for optically transparent thin films and reflectometry for nontransparent films [61]. This technique consists on projected light from a source focused on the semiconductor wafer where the reflected light from the top and from the bottom of the film are measured by a detector. The reflected light intensity varies as a function of time depending on the thickness of the top layer due to constructive and destructive interferences caused by multiple reflections [48]. The technique is illustrated in Figure 1.6. The thickness of the film is measured as the periodic variation of reflectance according to wavelength due to the two paths followed by the light.

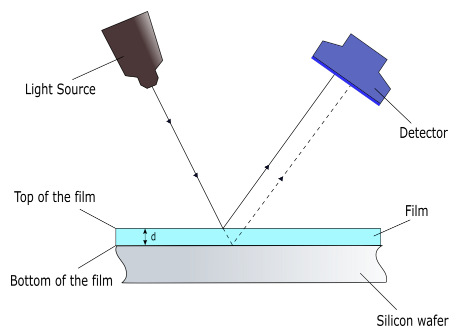



**Figure 1.6 –** *Interferometry used to measure film thickness of deposited layers.*

2. **Ellipsometry** is a measurement method of film thickness. It is based on the polarization change of the light when it reflects from the dielectric on the wafer surface. If the optical constants of the dielectric are known, the thickness can be determined. Changes in polarization are due to optical properties of the material (i.e., its complex refractive indices), its thickness, and the wavelength and angle of incidence of the light beam relative to the surface normal [48]. Spectroscopic ellipsometry refers to the use of wavelength-dependent data, it measures amplitude ratio and phase, and change versus wavelength at a single angle of incidence [81]. Typically, a model of the optical structure of the dielectric film or film stack on the silicon wafer is used [23]. It is a fundamentally more accurate technique than interferometry for obtaining film thickness and optical dielectric function information, as ellipsometry has the advantage of gathering information from both polarization states [48] [23], and is represented in a schematic way in Figure 1.7.

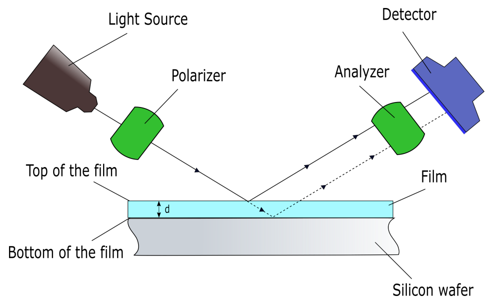



**Figure 1.7 –** *Schematic representation of the ellipsometry measure.*

3. **The four-point probe** is a device to measure sheet resistance and resistivity, the use of this instrument is considered as a destructive measurement and thus the measures are done over test wafers. Several process workshops such as ion implantation, diffusion or metal deposition, check sheet resistance to keep the yield of the devices and control the process [23]; for instance, the active dopant dose could be characterized by measuring resistivity and relating it to the active carrier concentration [61]. This method is based on placing in contact with the wafer surface four conductive probes in a linear array. Two probes carry a known current and the voltage across the other two is measured. By using Ohm's Law ($V = I * R$) sheet resistance is calculated. Two different configurations can be selected for the four-point probe measurement setup as is shown in Figure 1.8.

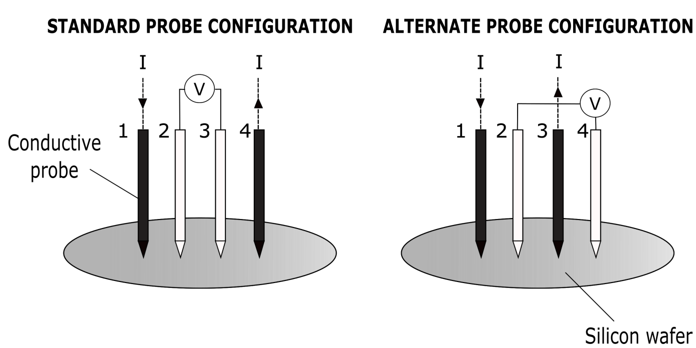



**Figure 1.8 –** *The different methods for four-point probe measurement.*

4. **Metrology systems for Doping Processes:** In addition to the four-point probe method that is also used by the diffusion and ion implant workshops for measuring the resistivity on doped layers, other tools or methods are used for the doping processes:

   (a) Implant Damage Measurement: One of the properties of the wafer that must be monitored after the implantation process is the level of damage produced on the wafer due to the implantation doses that alter its surface. There are two important measurement techniques to tackle this: Thermawave and optical dosimetry.

Both techniques have quite a good sensitivity to the lower and upper ranges of implant doses. **Thermawave** is based on reflectivity change of the silicon surface implanted. It is an optical reflection tool which is often named following the company's name. This metrology system emits thermal waves through an ion laser that heats the surface of the implanted wafer changing the volume of the silicon close to the surface. This volume change alters the optical properties of the surface that is probed by measuring the change of reflectivity using light from another laser [61]. The implant damage on the surface of the wafer is determined by this volume change. **The optical dosimetry** relies on getting dark polymers as a consequence of the implanted ions, and by using a test wafer it is quantified as the optical response of the polymers for the dose and energy used [23].

(b) Secondary Ion Mass Spectrometry (SIMS): The secondary ion mass spectrometry is an offline measurement which is employed to analyze the composition of the surface of an implanted wafer and its profile. The surface is sputtered away by an ion beam and the secondary ions rejected by the surface are registered and analyzed by a mass spectrometer. This type of measure enables measuring the dose and depth at the same time and in most cases it is considered as an offline measurement performed in an external laboratory [23].

#### 1.3.2.2 Patterned thin film

The following measure systems are used on wafers with patterned thin films, wafers that have been patterned after going though photolithography and etching process steps.

1. **Profilometry:** For measuring the topography of the surface of the wafer a metrology tool called "profilometer" is commonly used. This system is composed of a stylus sensor in contact with the surface measuring its profile. As the stylus encounters a step high on the surface, a signal identifies on a chart this profile variation. The advantage of this measure is its high sensitivity but as a drawback it is considered slow in measurement speed [81]. A schematic representation is provided in Figure 1.9.

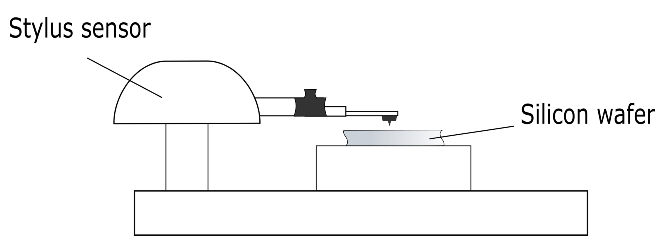



**Figure 1.9 –** *Profilometer system to measure the profile of the wafer surface.*

2. **Atomic Force Microscopy (AFM)** is a method that measures surface topography (surface properties and/or profiles) with a very high definition in the order of an atomic scale [48]. As represented in Figure 1.10, this device consists in scanning the wafer with a sharp tip fixed to the end of a cantilever arm. Once the tip is near to the surface the electrostatic forces between them will deflect the cantilever perpendicularly to the surface. These vertical displacements of the cantilever arm during scanning are turned into topographic information [23].

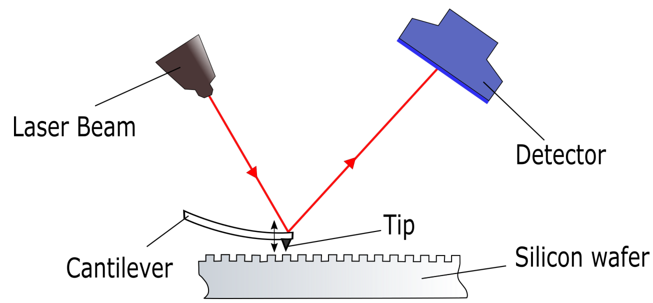



**Figure 1.10 –** *Elements that comprise the atomic force microscopy.*

3. **Critical Dimension Scanning Electron Microscopy (CD-SEM):** The critical dimension (CD), the linewidth of the trenches, is the smallest physical feature which can be built in semiconductor manufacturing, due to its continuous size reduction through the years it has rendered traditional metrology systems obsolete. However, linewidth measurements based on Scanning Electron Microscopy (SEM) can overcome the limitations of optical techniques for submicrometer geometry features [48]. CD-SEM relies on scanning the wafer surface by projecting a beam of electrons and signals produced by the interactions between the atoms of the surface and the electrons are collected in a secondary electron detector which provides the information related to its topography, as shown in Figure 1.11.

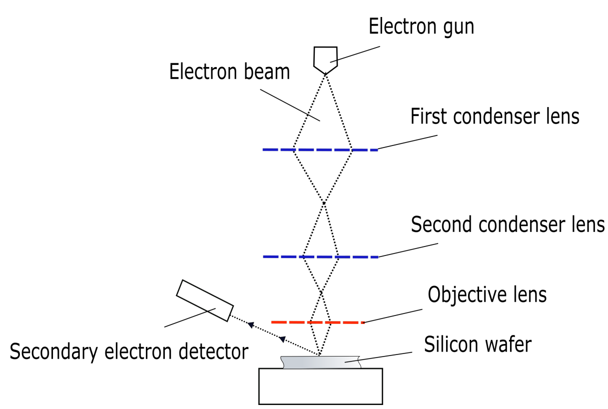



**Figure 1.11 –** *Representation of the Critical Dimension Scanning Electron Microscopy (CD-SEM).*

The main difference between in-line CD-SEMs and those used in laboratories is that the first are configured in very short working distances (distance between wafer and SEM electron beam lens) [61]. In Figure 1.12, some results of the CD-SEM measurements are displayed.

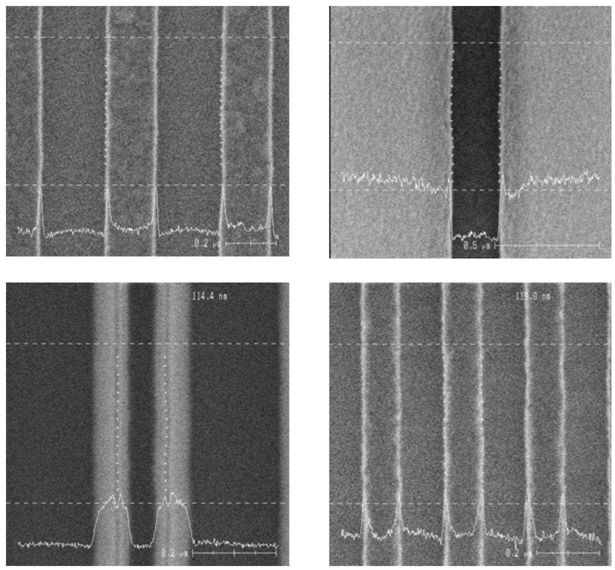


**Figure 1.12 –** *Examples of CD-SEM measurements (STMicroelectronics).*

4. **Scatterometry** is an optical metrology technique which is used to characterize surface roughness, defects, particle density on the surface, film thickness and the CD of patterned wafers with periodic structures [48]. Its operation relies on the measurement of the intensity of the scattered or diffracted light coming from the wafer, as the collected intensity variations enable to form linewidths and topographic shapes after carrying out its corresponding analysis through computerized models. The final results are shown in Figure 1.13.

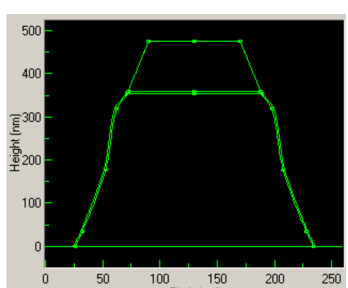



**Figure 1.13 –** *A scatterometer model after collecting diffracted lights (STMicroelectronics).*

Depending on the incident and measurement angle, some configurations can be adopted. When both are fixed it is called Fixed-angle scatterometer, whereas when both angles are variables it is called Variable-angle scatterometer or $2\Theta$ scatterometer [23]. In Figure 1.14 a schematic of the $2\Theta$ scatterometer is presented.

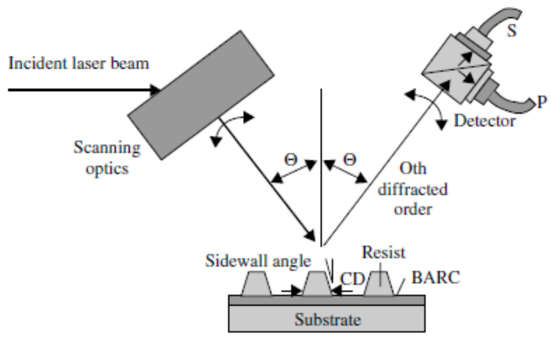



**Figure 1.14 –** *Schematic of a* $2\Theta$ *scatterometer [48].*

5. **Overlay** metrology controls the effects of the photolithography step due to the alignment of every patterned layer which must be verified. This measurement contributes in comparing the patterned design of the current layer with the patterned structure of the previous layer. Normally this metrology system uses an optical system that automatically evaluates how far the center of the target pattern in the top layer is from the center of the target pattern in the layer below [61]. Different types of misalignments are found, such as: by magnification, with trapezoidal form, with a rotation, by expansion, by rotation or translation as it is summarized in Figure 1.15. A shift between the reference level (in red) and the measured level (in blue) can be seen. If the result of the application of the resin on the surface of the wafer is unexpected, the wafer would be reworked.

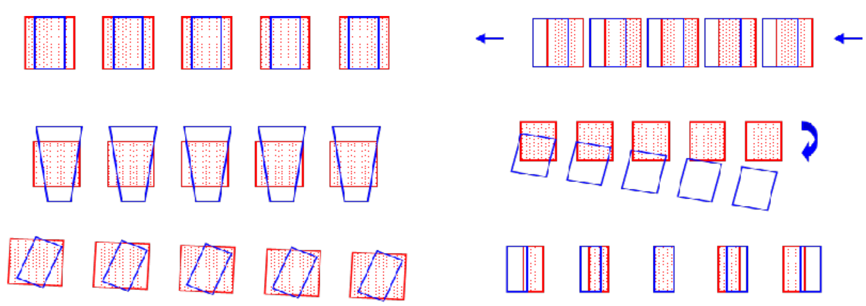


**Figure 1.15 –** *Possible misalignments detected by using overlay measurement (STMicroelectronics).*

6. **Film Thickness Measurement Using Acoustic Methods:** One way to measure metal film thickness is by using non-destructive techniques based on acoustic methods. The most important systems are the impulsive simulated thermal scattering (ISTS) and the picosecond ultrasonic laser sonar technology (PULSE). The acoustic wave travels parallel to the surface of the sample in ISTS, in contrast to picosecond laser ultrasonic sonar where the acoustic wave travels downward [61].

#### 1.3.2.3 Particle and defect inspection

In order to verify the surface of the wafer and identify deposited particles or physical defects, inspection techniques are required. There are several kinds of defects that may occur such as: particles, voids, patterns, scratches, residues, corrosion or extra-patterns. The objective is to decrease the ratio $defects/cm^2$, and thus improve yield by monitoring the possible effects arising from process machines or metrology tools along the manufacturing steps. The inspection tool must identify the presence of defects and their location on the wafer by a contrast mechanism which is to distinguish the defect from its surroundings [81]. The inspection systems most commonly used are **Dark-field**, **Bright-field** and **Electron beam**. The Dark-field system is equipped with a laser beam and collects the scattered light to generate an image, the Bright-field one uses a lamp and recovers the scattered light and also the reflected light, the first system is capable of finding smaller defects than the second one can. A third system, the electron beam, is in charge of finding conductivity and insulation defects. Finally, another technique used to analyze metal contamination is known as **TXRF** (total reflection x-ray fluorescence spectroscopy), where x-rays at a very low angle on the sample surface are reflected, exciting only the top few monolayers and, florescence x-rays are then emitted in various directions and collected by a detector [61].

## 1.4 Problem description

Along the years, semiconductor factories have been applying the best practices to improve their manufacturing yield [20] and particularly by using process control methods. It is important to preserve the process machines in a good condition, because if one starts producing in a unexpected way all wafers produced will be affected thereby generating possible excursions and thus increasing the amount of scrapped wafers.

There are two important aspects to take into account to ensure the complete coverage of process machines:

- The measurability of lots processed, which means the percentage of lots processed which are authorized to be measured.
- The sampling strategy employed. Sampling is the selection of a part of the population to determine some characteristics. In this case of importance is the frequency of measuring lots (sampling rate) and the strategy, being, which lots are selected over others?

For certain metrology systems is hard to qualify new process operations often because the recipes are composed by a high number of parameters [81] or software used are quite old and it is time consuming. Therefore, only a percentage of lots processed belong to the group of **measurable** lots and those which are not qualified to be measured are called **not measurable** lots. This particularity should be considered for every metrology area since depending on the nature of the workshop perhaps not all lots are candidates to be measured and thus reduce the risk.

Concerning sampling techniques, they can be classified into three main classes: Static, adaptive and dynamic [60].

- The start or **static** sampling technique defines the lots to be measured at the beginning of the production without any possible change.
- The **adaptive**, as the static techniques, also flags the lots to be measured at the start by its corresponding sampling rules but with the possibility to finally perform the measure or not depending on the information gathered during production.
- The third type, **dynamic** sampling measures in real time the best lots or wafers depending on metrology capacity and the current situation.

Drawbacks been shown for static sampling [59], where some process machines were never covered requiring at the end the use of dynamic policies.

To control process machines or products static sampling could be assigned with a given sampling ratio as "1/N", N being the number of lots processed with wafers inside at risk, which means that for a ratio of 1/3, one lot will be measured after 3 lots processed and if each lot contains 25 wafers, there will be 75 wafers at risk. W@R (wafers at risk) is an indicator to manage the risk level of a process machine, a product and so on. It indicates the number of wafers processed that could be at risk between two measures. There are metrology workshops, which cover a group of process machines, that follow a start sampling strategy by product and operation and has as the drawback of never having a constant ratio to cover the process machines, so their sampling rates vary following the distribution of lots to be processed in front of them, and therefore the wafers at risk are not controlled.

Besides, it has been found that some metrology workshops are not product dependent, but process machine dependent (e.g. on some optical measures) and sampling by equipment is required. Figure 1.16 shows an example of the differences between sampling by product and operation and sampling by equipment. This figure shows for three different products and operations (A, B and C) with different sampling rates (1/3, 1/4 and 1/5) how from this distribution of lots any lot will be measured until the ninth (1/9). The values of risk could vary without control, producing periods of lack of control and even over-control (measuring lots in a consecutive way). On the opposite, no matter what type of product or process operation the lot stems from, sampling by equipment will send a lot to be measured following the established sampling rate, for our example it will be always 1/4, and then the risk will be under control with a constant ratio.

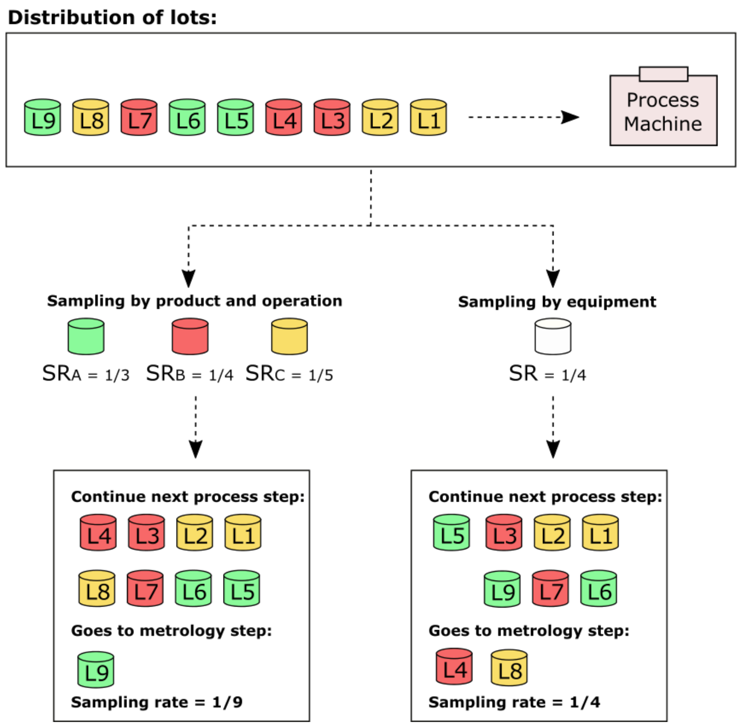



**Figure 1.16 –** *Sampling by product and operation and sampling by equipment.*

For the metrology workshops that could not switch completely to sampling by equipment because they are product dependent due to their nature (e.g. the thickness measure) a new sampling strategy is necessary, one that combines the coverage of process machines as well.

The first step is to categorize the general types of metrology workshops. In the first place the sampling policies used to adapt the workshops to sampling by equipment, by product and operation or in a combined form, additional considerations should be taken to create a general procedure for sampling strategy changes. Moreover, it is important to meticulously choose which characteristics are important for metrology workshop classification in terms of dispatching rules for the lots, their location in the fab with respect to the controlled process machines, the difficulty to create new qualifications, the queue times to perform the measure after the process operations and the current wafer at risk values. These properties will provide us information to design new dispatching or sampling policies [21] in case of necessity thereby decreasing queue times that will reduce the manufacturing cycle time, reducing exposition of the wafers in production to bad processing and with this, improving the yield [43].

As mentioned before, when a sampling strategy change requires a switch from sampling by product and operation to sampling by equipment, the accurate calculation of the sampling rates to put in the process machines is demanded, so one of the keys is to consider the optimization of sampling rates once the sampling change is made. Another axis of research, for any given metrology workshop with serious problems in qualifying products to measure, with only a small percentage of measurable lots is, how to decide which are the best products to qualify.

## 1.5 Thesis objectives

In the context of this thesis, we put emphasis on bringing a deep study of all metrology workshops and getting a global picture in terms of risk. One of the pillars of this work is to understand the current system of the sampling policies in the site of Rousset of STMicroelectronic, detect the possible improvements that could be done as it has been explained before to reduce the risk levels, doing a better use of metrology capacity and thus improving the cycle times.

The objectives that were designed to achieve this are the following:

- Analysis of the metrology workshop's behavior: study its running, find their properties, extract the levels of risk to evaluate the current status of each one.
- Development of a procedure to implement new sampling strategies: once the decision of making a sampling strategy change is taken, which steps must be followed? A general procedure should be created.
- Propose a new method to calculate sampling rates: optimize the sampling rates for the metrology workshops considering parameters coming from the production machines and metrology tools.
- Design simulation models to control dynamically the risk and check the impact of modifications on the workshops: study the behavior of several workshops through simulation models and verify possible improvements before implementing them.

## 1.6 Conclusion

This first chapter presents briefly the most important parts of the domain where it has been carried out, the semiconductor industry. The main process steps to build ICs have been introduced as well as the metrology systems that are responsible of keeping the quality of the products by controlling the well shape of the production machines. The huge number of process operations needed to get a final product added to the optional quantity of metrology steps show the degree of difficulty linked to the semiconductor environment.

In this thesis, we have focussed on discover how it is used each metrology station by the operators to find out the main characteristics of each one in order to establish the key points to improve the sampling parameters. Understand more deeply these characteristics will enable us to propose new actions in order to better manage the wafer at risk values of the metrology workshops.

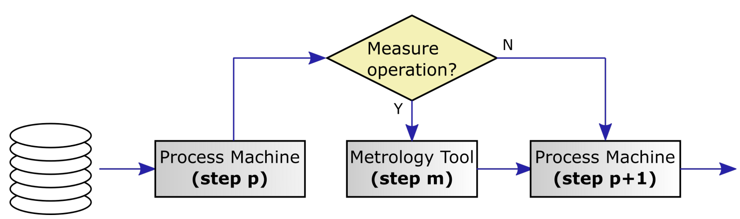
At the beginning of the project, an only sampling strategy was used, that it was product based, the main drawback is that during the day the production machines were not controlled regularly with a fix ratio.

The possibility of switching to another sampling policies has open different research doors: New strategies to cover just process machines or a combination of products and process machines, the optimization of the sampling rates once the sampling change is done, the development of a general procedure to know the convenience of performing a sampling strategy change and the previous analysis through simulations. In the next chapter, a literature review focused on risk control is provided.

# Chapter 2

# Literature review on risk controls and sampling techniques

*This chapter defines the term risk in semiconductor environment. It describes the methods to manage the risks in semiconductor manufacturing and the risk indicator used this thesis is explained. A state of the art on sampling techniques in semiconductor manufacturing is provided, and also our problems are positioned in the literature.*



## 2.1 Introduction

The semiconductor industries develop methods in order to detect process drifts as soon as possible for keeping high yields. Finding new ways to improve process and equipment reliability without overloading the production system has become a priority [19].

This chapter studies how the management of risk linked to production losses is handled by several research works in the literature and a classification of the main risk types is proposed. The goal is to show the variety of studies that consider term risk in different ways, but keeping the same objective of process and equipment improvement.

In order to achieve the quality of process and production machines, control operations are performed, but due to the high costs associated to measuring all lots and the risk of saturating metrology capacity, sampling strategies are implemented. In this chapter, the principal sampling strategies used in semiconductor manufacturing found in the literature are defined.

Section 2.2 presents several definitions found in the literature related to the term risk as used in a semiconductor environment and a classification containing all the works and the types of risk identified since 1990. This section also introduces the risk indicator chosen in this research and called W@R. In section 2.3, a classification of the different sampling strategies adopted in semiconductor manufacturing is presented, reviewing all studies of the last 10 years which have developed a sampling policy and their type (static, adaptive or dynamic sampling).

## 2.2 Risk definition

In this thesis we will not talk about risk in terms of the level of danger that could cause health problems to human beings or to the environment, but risk related to production or process equipment. Thus the risk indicators that compute the hazards for people due to illnesses, work accidents or occupational hygiene will be excluded from risk calculation research in semiconductor manufacturing.

The term risk could be defined as a situation of exposure to danger. Concretely, in an industrial environment it is the possibility of the occurrence of a failure, and represents the potential failures that may have an impact on products, equipment or production processes [9]. These failures generate an increase of cost and cycle time and yield loss.

### 2.2.1 Risk in Semiconductor Manufacturing

In semiconductor manufacturing, risk assessment has been studied since the early 1990s. As a result of a full review of literature, we have established eight main methods to define and parametrize the risk in a semiconductor environment:

1. **Failure Mode and Effects Analysis (FMEA)**: FMEA was originally developed in the 1950s by US Armed Forces engineers to study problems that arose from malfunctions of military devices. It is a technique not only to minimize risks, but also to define them whenever possible [88]. It is widely used in the automotive, aerospace and electronics industries. The risks are prioritized by a Risk Priority Number (RPN) which is the product of three indices: Occurrence (O) Severity (S) and Detection (D) of each individual risk and the goal is to analyze the 'root cause' and 'end effects' of potential failures in the system [73]. In 1994, one of the first studies which used the FMEA approach in semiconductor fabrication [107] showed how this approach helps in aligning work efforts to achieve yield improvement and risk reduction.

2. **Supply chain risk management (SCRM)**: SCRM is a method that implements strategies to manage both known and exceptional risks along the supply chain based on continuous risk evaluation with the objective of reducing exposure and ensuring continuity [108]. It is normally divided in four process steps: Identification, assessment, controlling, and monitoring of supply chain risks.

3. **Material at risk (MAR)**: In this case the risk is quantified, it is not considered as an event, a deviation from an objective or a probability. The MAR is an amount of material or product that is going to receive a physical treatment. A MAR value represents the quantity of material that corresponds to an entity, a process or a facility which is studied. In the literature, MAR has been interpreted by controlling the number of lots at risk [39] and in a more precise way through the number of wafers at risk (W@R) [21].

4. **Project risk management (PRM)**: The consideration and assessment of risks in projects is extended and has become an important area of project management [110]. There has been some propositions for the PRM procedure, among them, the Project Management Institute proposes four phases: Identification, quantification, response development and control [72]. Some studies have been carried out in the semiconductor manufacturing field [114].

5. **Bayesian networks (BN)**: Also known as belief networks, they are part of the family of probabilistic graphical models. A Bayesian network is a model where a set of random variables are represented (nodes) with their conditional dependencies (edges). They are applied in several fields such as medicine, image processing, engineering and also for risk analysis [102].

6. **Risk assessment survey**: This category encompasses all methods which are composed by check-lists or questionnaires with key questions to follow. The goal of this surveys is to find the places where a factory is most at risk [22], and this kind of analysis is carried out by experts who have knowledge in all disciplines.

7. **Mathematical model**: It consists in defining the risk by means of a mathematical function. It could be, for instance, a new parameter or indicator devised through a formula taking into account several variables.

8. **SPC risk alarms**: The aim of a SPC method is to decide whether a process is under statistical control or not. Two hypotheses are considered, whereas $H_0$ is a process under statistical process control and $H_a$ is a process out of statistical process control. There are two types of errors associated to these hypotheses. Type I error due to mistakenly rejecting $H_0$ (to emit a false alarm), and type II error owing to mistakenly accepting $H_a$ (miss issuing an alarm). The *manufacturer's risk* is the probability of committing a type I error causing unnecessary disruptions during production, while the *consumer's risk* is the probability of committing a type II error leading up to producing defective products [86].

The main approaches to manage risk found in the literature used in a large amount of works and the most noteworthy among them are discussed in Section 2.2.3.

### 2.2.2 Wafer at risk (W@R) indicator

In this thesis, the measure of risk used is the **Wafer at Risk** (**W@R**) indicator. W@R is the exposure level in number of wafers processed since the last measured lot, and it can be applied to different ambit such as recipe, process machine, product or technology. This indicator represents the number of wafers that are potentially in danger if a problem occurs.

Figure 2.1 illustrates the W@R evolution of a process machine over time. W@R increases when a lot is processed on a production machine and it decreases as soon as the results of a measure are obtained. Let us suppose that $NW(l)$ is the number of wafers in lot $l$, $W@R_m$ denotes the current wafers at risk of process machine $m$ (W@R evolves dynamically) and $W@R_m(l)$ denotes the wafers at risk when lot $l$ is completed on $m$. Then, $W@R_m$ and $W@R_m(l)$ are updated as follows when lot $l$ is completed on $m$: $W@R_m := W@R_m + NW(l)$ and $W@R_m(l) = W@R_m$. Until the measure of lot $l$, that was processed on process machine $m$, is performed, W@R$_m$ keeps increasing due the fact that other lots continue to be processed on $m$. When lot $l$ is measured, then $W@R_m := W@R_m - W@R_m(l)$.

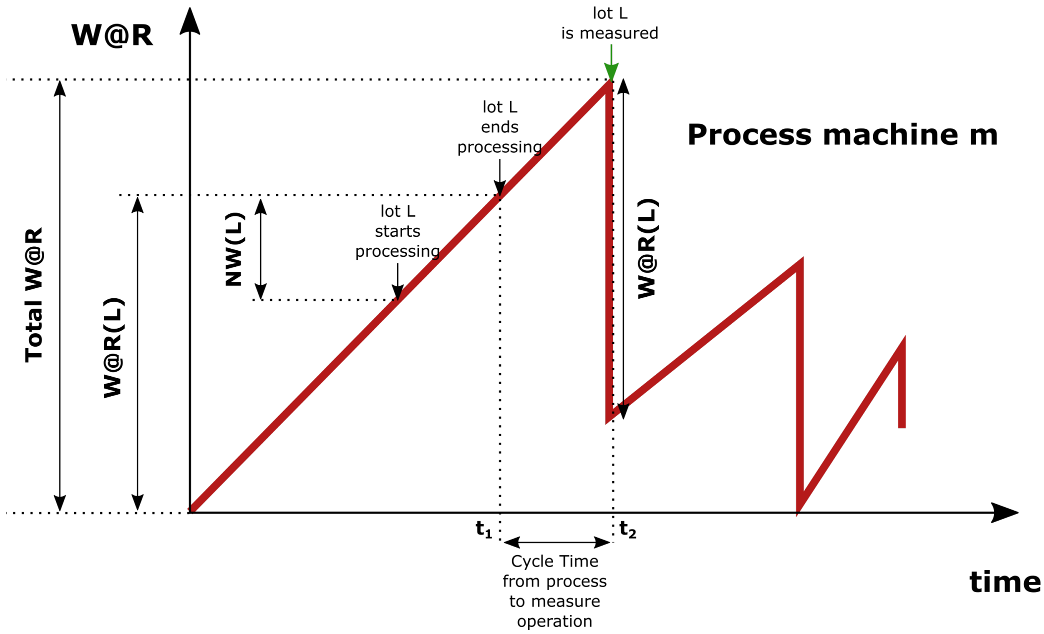



**Figure 2.1 –** *W@R profile for a process machine m.*

### 2.2.3 Literature review

In the literature, research works are focused on risk management using the methods explained previously. Concerning the use of a *FMEA* procedure, in 1994 Whitcomb and Rioux [107] implemented the technique from the automotive to the semiconductor field to improve the CVD process step. Five years later, Trahan and Pollock [100] developed an inverted FMEA that starts by listing the proposed changes to the existing system in order to minimize the risk contrary to the classic FMEA that seeks to identify what could fail. In 2001, Trammell et al. [101] proposed to apply a system of risk assessment methods early in the design phase that is a combination between FMEA and the Hazop technique. Other approaches focused on quality and costumer's satisfaction propose using in addition a QFD (Quality function deployment) matrix to consider the reliability demands from the customer. Some works help in the definition of an occurrence index [95] by means of a fuzzy inference system. More recently, Mili et al. [49] have developed a method using FMECA (FMECA extends FMEA by including a criticality analysis) and Ahire and Relkar [1] have established a relation between OEE (Overall Equipment Effectiveness) and FMEA results.

Nowadays, *supply chain risk management (SCRM)* has become a priority, and enterprises must concentrate not just on the performance of supply chain, but also on its risks. In 2000, Ericsson lost 400 million Euros after their supplier's semiconductor plant caught fire, Philip's semiconductor plant was damaged in a fire and Apple lost many customer orders during a supply shortage of DRAM chips after an earthquake hit Taiwan in 1999 [94]. Terwiesch et al. [96] using data gathered from a semiconductor company showed that both manufacturer and retailer would lose when sharing forecast information. On one hand, the manufacturer would penalize the retailer for unreliable forecasts by delaying the fulfilment of forecasted orders, on the other hand, the retailer would inflate their orders only to make use of excessive order cancellations. Houshyar et al. [30] through simulation models identified the probable reasons of disruptions in a semiconductor supply chain and classify the risks that can cause them. Chen and Wu [15] within the frame of supply chain risk management (SCRM) proposed a modified FMEA method to select new suppliers from a supply chain risk's perspective.

It is possible to identify the places where the factory is most at risk for scrapping multiple lots, as De Pinto [22] showed. Firstly, he applied a *risk assessment survey* to find the major risk zones of the fab and secondly, implemented a FMEA procedure to improve and manage the risk. By implementing this approach, in-line scrap was decreased by 27% over a one year period of time. Here, it is one example of how the risk can be quantified (percentage of scrapped wafers) which could be associated to a *material at risk* treatment of the risk.

In 1999, the concept of material at risk is used [109] to optimize a sample planning for wafer defect inspection. This measure of risk has been defined as lots at risk in literature [39] [10] [83] mainly to compare the effectiveness of sampling methods. Another way to better quantify the material at risk is through the number of wafers at risk, using the indicator called W@R, see Section 2.2.2. In the literature several studies have used W@R values to calculate risk levels. Dauzere-Péres et al. [21] developed through a *mathematical model* an indicator, called GSI, valid for multiple risk types on multiple tools, and implemented it using W@R values. Nduhura Munga et al. [55] propose an algorithm that helps operators to correctly dispatch lots on production tools based on their states for the CMP workshop, and a second study was carried out in Nduhura Munga et al. [58] for the defectivity workshop using the same mathematical index providing the set of process machines most likely to be the source of the excursion computing global risk indicators on production. For control plan design some works focused on the W@R reduction of process machines, as in Rodriguez-Verjan et al. [106], with the aim of remaining below a threshold limit of risk exposure, and as Bettayeb et al. [7] as well. Rodriguez-Verjan et al. [75] [76] compared the W@R levels on process machines to validate novel dispatching strategies and skipping algorithms. Sahnoun et al. [77] used W@R as a performance indicator of its sampling strategy and Sendón et al. [80] have developed simulation models for Ion Implant workshop for controlling risk on process equipment comparing W@R levels.

At the beginning of the 90s, $\alpha$ (manufacturer) and $\beta$ (consumer) risks started to be used in process control applications [42] [86]. To evaluate sampling plans, some studies are based on balancing between the material at risk and the false alarms generated (alpha risk, for SPC methods) as in Elliott et al. [25]. In order to determine the optimal in-line inspection sampling using traditional SPC policies, Tomlinson et al. [99] verified the percentage variation of lots at risk. Kuo et al. [39] evaluated the effectiveness of his SPC chart based adaptive sampling making a lot at risk comparison. Nurani et al. [62] proposed a method that reduces the probability of declaring an in-control process to be out-of-control ($\alpha$ risks) and they took into account all the relevant cost components such as wafer inspection costs and costs of running out of control. In 1999, Shindo et al. [84] developed analytical models and compared $\beta$ risks (probability that the excursion of the killer type defect is not detected) for effective process excursion monitoring.

In the literature, several works related to *project risk management (PRM)* can be found. Benavides et al [6] developed an approach to determine the optimal fab investment policy, to balance the risk of costly periods of low capacity utilization with the risk of costly periods of capacity shortages. Zafra-Cabeza et al. [114] [113] applied to the construction of semiconductor manufacturing facility a method to optimize the cost and time of the project considering principles from risk management and through model predictive control (MPC). Su and Chou [91] proposed a systematic methodology based on the company's strategic policies and costumer demand, to determine the benefits and risk priorities of each project.

*Bayesian networks* have been applied in some approaches for managing the risk, for instance, Chien et al. [17] used a Bayesian decision analysis to systematically structure the problem and effectively analyze alternative strategies for IC final testing. Bouaziz et al. [9] designed a general procedure to develop predictive risk models on the product and equipment in an uncertain context. A predictive behavior model based on Bayesian learning methods was proposed.

All papers surveyed related to risk assessment in semiconductor manufacturing are summarized in Table 2.1.

**Table 2.1 –** *Survey on approaches for risk management in semiconductor manufacturing.*

| | Year | FMEA | SCRM | Bayesian networks | Risk assessment survey | Material at risk | Project risk management | Mathematical model | SPC risk alarms |
|---|---|---|---|---|---|---|---|---|---|
| Lazaroff et al. [42] | 1991 | | | | | | | | * |
| Spanos [86] | 1992 | | | | | | | | * |
| Whitcomb and Rioux [107] | 1994 | * | | | | | | | |
| De Pinto [22] | 1996 | * | | | * | * | | | |
| Nurani et al. [62] | 1997 | | | | | * | | * | * |
| Kuo et al. [39] | 1997 | | | | | * | | | * |
| Tomlinson et al. [99] | 1997 | | | | | * | | | * |
| Akella et al. [2] | 1998 | | | | | | | | * |
| Babikian and Engelhard [3] | 1998 | | | | | | | | * |
| Benavides et al. [6] | 1999 | | | | | | * | * | |
| Elliott et al. [25] | 1999 | | | | | * | | | * |
| Shindo et al. [84] | 1999 | | | | | | | * | * |
| Trahan and Pollock [100] | 1999 | * | | | * | | | | |
| Williams et al. [109] | 1999 | | | | | * | | | |
| Trammel1 et al. [101] | 2001 | * | | | | | | | |
| Tan and Neo [93] | 2002 | * | | | | | | | |
| Terwiesch et al. [96] | 2002 | | * | | | | | | |
| Bousetta [10] | 2005 | | | | | * | | | * |
| Chien et al. [17] | 2007 | | | * | | | | * | |
| Shanthikumar [83] | 2007 | | | | | * | | * | * |
| Zafra-Cabeza et al. [114] | 2007 | | | | | | * | * | |
| Su and Chou [91] | 2008 | * | | | | | * | | |
| Tay et al. [95] | 2008 | * | | | | | | | |
| Zafra-Cabeza et al. [113] | 2008 | | | | | | * | * | |
| Mili et al. [49] | 2009 | * | | | | | | | |
| Dauzere-Péres et al. [21] | 2010 | | | | | * | | * | |
| Houshyar et al. [30] | 2010 | | * | | | | | | |
| Bouaziz et al. [9] | 2011 | * | | * | | | | | |
| Nduhura Munga et al. [55] | 2011 | | | | | * | | * | |
| Nduhura Munga et al. [58] | 2011 | | | | | * | | * | |
| Rastogi et al. [71] | 2011 | | * | | | | | * | |
| Verjan et al. [106] | 2011 | | | | | * | | * | |
| Bettayeb et al. [7] | 2012 | | | | | * | | * | |
| Rodriguez-Verjan et al. [75] | 2012 | | | | | * | | * | |
| Sahnoun et al. [77] | 2012 | | | | | * | | * | |
| Ahire and Relkar [1] | 2012 | * | | | | | | | |
| Rodriguez-Verjan et al. [76] | 2013 | | | | | * | | * | |
| Chen and Wu [15] | 2013 | * | * | | | | | | |
| Sendón et al. [80] | 2015 | | | | | * | | * | |

## 2.3 Sampling strategies in semiconductor manufacturing

The way to reduce risk values associated to process machines, products, and so on, is achieved by means of selecting certain processed lots to perform a measure operation. This is called sampling. Once the measure is done, depending on its result, the process could be considered as working as expected or not.

Measurement steps are mandatory to guarantee the requirements on quality of products and processes, and because of that measure operations are distributed throughout the production route, allowing to detect manufacturing problems. However, the high costs associated with these measurement devices and the enormous amount of time invested on controlling 100% of lots has led to devising **sampling strategies** as a necessity [90]. Besides, 100% measures still does not ensure 100% quality since measures are never totally reliable, and can easily introduce an error of the same order as the fraction of defectives [65].

In addition, measuring all lots does not ensure a high yield. Unnecessary measurements may occur leading to an increase of CT (cycle time) and yield decline due to long queue times of lots in metrology workshops that delay corrective actions [97]. Hence, long manufacturing CT may have a negative influence on yield [43].

A sampling strategy corresponds to the way of selecting lots to perform a measure. The decision of measuring certain lots depends on the availability of metrology capacity and the reduction of risk obtained. Nduhura-Munga et al. [60] propose a complete literature survey, where sampling techniques in semiconductor manufacturing are reviewed and their benefits and drawbacks are discussed. They introduce a sampling technique classification divided into three main categories: **Static**, **Adaptive** and **Dynamic** sampling.

### 2.3.1 Static sampling

This type of sampling is designed by fixed rules that will not change along the production route, for instance, by defining a certain amount of lots to sample (by flagging them) at the beginning of the route or assigning a fixed frequency or sampling rate to the process machines. Even if this technique is easy to implement, and thus is widely used because it helps to rather easily manage metrology capacity, it does not allow any quick reaction to continuous changes in the fab.

*Static sampling* has been used in the majority of semiconductor companies. Even if, for high-mix semiconductor plants, it is more recommended to use dynamic sampling strategies, some companies keep using static policies for several products or process machines, especially during the integration phase of new products or equipment. Wu and Pearn [111] proposed a new distribution for a sampling plan based on the process capability index $C_{pmk}$ to deal with product sentencing (acceptance determination). Chow and Ong [18] presented a novel push-pull sampling methodology which consists in providing the correct sampling data to the operators for them to run test lots with the correct sampling values with the aim of reducing human handling or dependency. Sahnoun et al. [77] studied the impact of the reduction of the inspection sampling rate on performances of the production system, and they used W@R values and the metrology time delay as performance indicators of the sampling strategies.

### 2.3.2 Adaptive sampling

Adaptive sampling has been previously defined by the rules of static sampling, but it adapts sampling rates depending on the state of production. For example, when a metrology tool breaks down, the sampling rates are reduced to not exceed the capacity of the metrology workshop by saturating the remaining metrology tools. Or, when a problem occurs in a process machine, the sampling rates increase to measure more lots in order to verify that process parameters are stable.

In the middle of the 90s, the change from static to adaptive sampling started [67]. Mouli and Scott [53] presented an Adaptive Metrology Sampling based on the evaluation of the risk score, by weighting each lot to make sampling decisions and prioritize a list of lots on metrology tools. Shanthikumar [83] developed models based on a capture rate of the inspection tools, which determines the modification of the sampling rates. Chen et al. [14] developed an integrated metrology sampling strategy to maximize the wafer-level control effectiveness subjected to throughput, APC and SPC constraints. Veetil et al. [104] proposed an intelligent selection of samples according to the probability distribution function of total process variation which is performed using a Quasi Monte Carlo technique. In 2010, Good et al. [27] designed a sampling compensation algorithm based on the minimum-norm integrated moving average IMA forecast.

### 2.3.3 Dynamic sampling

Dynamic sampling bases its criteria to select the best lots to measure in real time and depending on the current situation of production, without considering predefined rules before the measurement step. This sampling technique enables to better manage metrology capacity than static and adaptive sampling techniques.

The first works focusing on dynamic sampling started in 2005 with Purdy et al. [69] [70], whom proposed a method to efficiently manage metrology queues combining separated sampling rules into a unique sampling decision, and two years later, an algorithm capable to respond to changing factory conditions that allows for improved flexibility in defining sampling rules. Good and Purdy [28] introduced an algorithm for selecting the optimal wafers, given a set of selection rules, based on assigning a penalty to each of the sampling rules and then using a mixed-integer linear program to select the wafers which minimize the sum of the penalties. Holfeld et al. [29] developed an integrated lot-level and wafer-level sampling application deployed in an AMD's fab that better controls metrology cycle time while maintaining excellent product yields. To avoid over-sampling, Kaga et al. [36] proposed a sampling methodology that can separate systematic defects and categorize random defects by using design data.

Jansen et al. [34] developed a sampling strategy for improving the utilization of SEM resources, that Xu et al. [112] and Tolle et al. [98] continued by comparing different sampling approaches. Sun and Johnson [92] proposed a method and system based on weighted objectives for determining the optimal wafer sampling for maximum coverage. Lin et al. [46] introduced a new sampling method for in-line defect inspections, that provides sampling stability, satisfactory coverage of products in line and comprehensive inclusion of the process tools.

Dauzere-Péres et al. [21] presented a sampling, scheduling, and skipping algorithm to minimize risk dynamically, based on a Global Sampling Indicator (GSI) that gives a weight to each lot arriving at the measurement step. Housseman et al. [31] presented an industrial application of dynamic sampling based on the GSI indicator that identifies the lots that should be measured to minimize the overall risk level of a fab. Nduhura Munga et al. [55] developed an index called IPC to evaluate in real time the risk on production tools in order to optimize the number of controls.

Kurz et al. [41] proposed a sampling design in form of a sampling decision system relying on Virtual Metrology that provides a decision regarding the state of every wafer. In a later work following the same line, Kurz et al. [40] presented several sampling strategies based on the information coming from Virtual Metrology (expected value of measurement information, two-stage sampling model, wafer quality risk values).

Rodriguez-Verjan et al. [75] developed a new dispatching system which helps operators to identify in real time when the risk level of process tools is critical and actions have to be taken. In a next stage of their study, Rodriguez-Verjan et al. [76] presented an industrial application of skipping algorithms to effectively manage inspection queues.

Table 2.2 presents the works found in the literature in the last 10 years related to sampling techniques.

**Table 2.2 –** *Survey on sampling strategies in semiconductor manufacturing.*

| | Year | Static sampling | Adaptive sampling | Dynamic sampling |
|---|---|---|---|---|
| Wu and Pearn [111] | 2006 | * | | |
| Holfeld et al. [29] | 2007 | | | * |
| Good and Purdy [28] | 2007 | | | * |
| Lensing and Stirton [45] | 2007 | | | * |
| Mouli and Scott [53] | 2007 | | * | |
| Purdy [69] | 2007 | | | * |
| Shanthikumar [83] | 2007 | | * | |
| Bunday et al. [11] | 2008 | | * | |
| Chow and Ong [18] | 2008 | * | | |
| Jansen et al. [34] | 2008 | | | * |
| Kaga et al. [36] | 2008 | | | * |
| Lee [44] | 2008 | | | * |
| Sun and Johnson [92] | 2008 | | | * |
| Chen et al. [14] | 2009 | | * | |
| Veetil et al. [104] | 2009 | | * | |
| Dauzere-Péres et al. [21] | 2010 | | | * |
| Good et al. [27] | 2010 | | * | |
| Lin et al. [46] | 2010 | | | * |
| Shanoun et al. [82] | 2010 | | * | |
| Shanoun et al. [78] | 2010 | | * | |
| Nduhura Munga et al. [55] | 2011 | | | * |
| Kurz et al. [41] | 2012 | | | * |
| Rodriguez-Verjan et al. [75] | 2012 | | | * |
| Sahnoun et al. [77] | 2012 | * | | |
| Rodriguez-Verjan et al. [76] | 2013 | | | * |
| Xu et al. [112] | 2013 | | | * |
| Housseman et al. [31] | 2014 | | | * |
| Kurz et al. [40] | 2015 | | | * |
| Tolle et al. [98] | 2016 | | | * |

To sum up, developing dynamic sampling strategies requires more investment in terms of time to develop its methodologies and implementation compared with static or adaptive sampling techniques. As to advantages, sampling strategies offer a precise control of the measures to be performed allowing to select the most suitable lot candidate to reduce risk, gain metrology capacity, reduce cycle time and increase yield. Hence, one part of our research work aims at proposing on new strategies by using dynamic sampling policies.

## 2.4 Conclusion

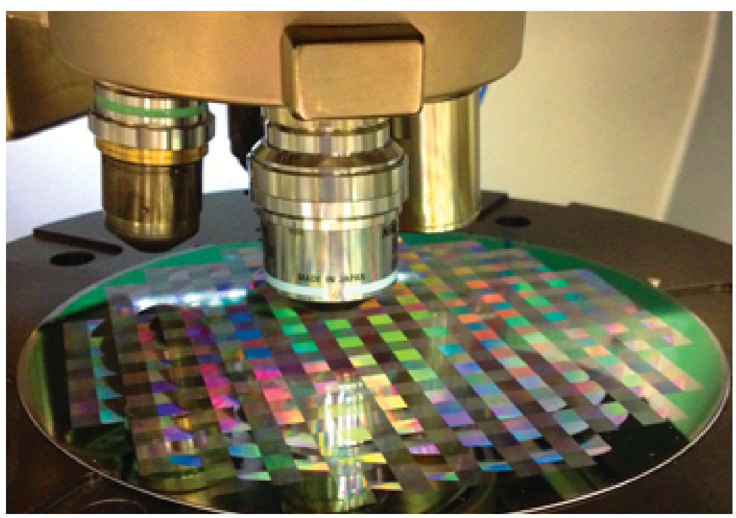
In this chapter, we surveyed the literature on risk in semiconductor manufacturing. We established eight main methods to define and parametrize the risk in a semiconductor environment: Failure Mode and Effects Analysis, Supply chain risk management, Material at risk, Project risk management, Bayesian networks, Risk assessment survey, Mathematical models and SPC risk alarms. Our research is focused on risk management by using the W@R as risk indicator which is the exposure level in a number of wafers processed since the last measured lot.

The sampling strategies used in semiconductor manufacturing are described: Static, adaptive and dynamic. Dynamic sampling is most suitable for high-mix semiconductor fabs because of it is more efficient in the selection of lots to measure, in its adaptation to the changes of the fab and in its management of metrology capacity. Thus part of our work proposes improvements by using dynamic sampling strategies. A new model to prioritize lots is introduced in Chapter 5.

Chapter 3

# Metrology workshop analysis and Sampling method improvement

*A complete analysis of all metrology workshops has been carried out, providing a guide identifying their characteristics and they are running efficiently. An analysis of queue times and approaches to reduce these queue times for the most risky metrology workshops are proposed. Finally a general procedure to define a sampling strategy change is provided and first numerical results are presented and discussed.*



## 3.1 Introduction

As explained in Chapter 1, to get the final product, wafers have to go through more than 400 process operations, most of these during a CT (Cycle Time) of approximately two months. Metrology aims of ensuring that machines are correctly processing wafers. This is why almost every process operation is followed by a metrology operation. However, an excessive use of metrology leads to longer CTs causing yield decrease [97]. A trade-off between the number of measures and the yield must be achieved.

In this chapter, the main metrology workshops in semiconductor manufacturing are analyzed, by pointing out their principal properties, and their behavior depending on the nature of the measure. A method for selecting dispatching or sampling rules according to the risk values and queue times of the metrology workshop is proposed. Finally, a sampling strategy change method has been developed whose goal is to define the right steps to identify the characteristics of the current sampling strategy for a certain metrology workshop, verify if there is an improvement margin in terms of risk, number of useless measures, and the corresponding follow-up stage.

This chapter is organized as follows. A brief introduction of the role of metrology in semiconductor manufacturing is described in 3.2. Section 3.3 includes the analysis of metrology workshops. In Section 3.4, a method for selecting dispatching or sampling rules is introduced. Section 3.5 introduces the sampling strategy change approach and first results with real data are shown.

## 3.2 Metrology in Semiconductor Manufacturing

Metrology, the science of measurement, defines measurement as the process of experimentally obtaining one or more values that can reasonably be attributed to a quantity [8]. Metrology is a key point in semiconductor manufacturing, since it helps to speed up the yield improvement and to keep the yield performance at every process step during production for new and mature products [87].

In the Rousset site of STMicroelectronics, 125.000 wafers are measured per week, which means that one wafer is measured every 4 seconds. The metrology area is composed of 46 people, 145 tools are currently used, as shown in Figure 3.1. There are 76 stand-alone metrology tools (for measures like thickness, overlay, critical dimension and so on), 27 defectivity tools (detection of contamination, particles, scratches, residues, voids, etc.), 27 handling tools (controlled storage tools and process enclosures to protect wafers from contaminants) and 15 embedded metrology tools (for electronic test applications).

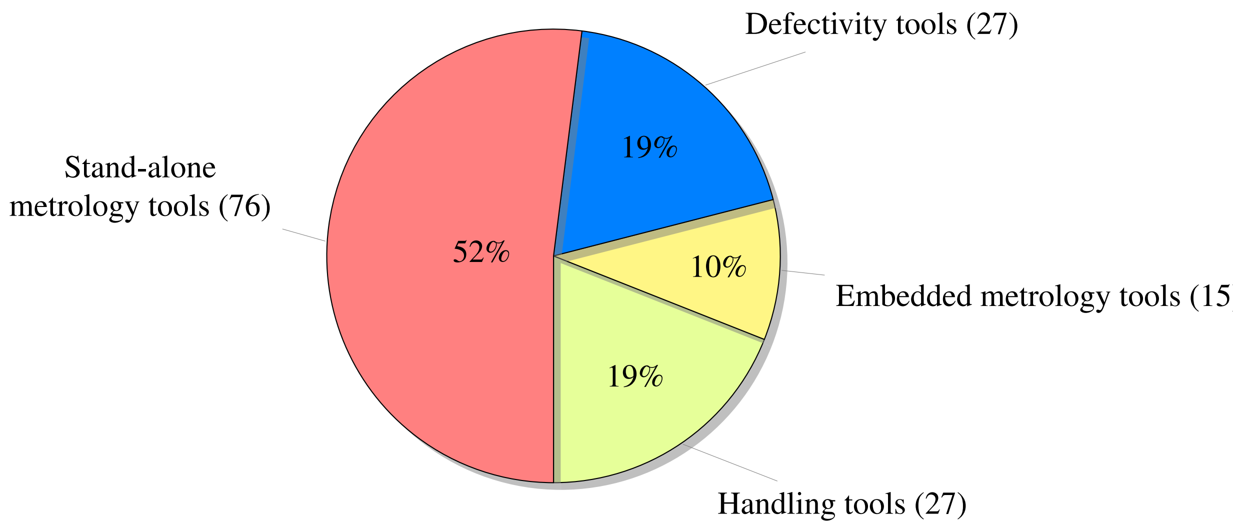



**Figure 3.1 –** *Distribution of metrology tools in the Rousset site of STMicroelectronics.*

Metrology is a key element and was the first semiconductor technology area to routinely work in the field of nanoelectronics [74]. Nowadays more than 40% of the production volumes are processed through metrology areas. Several functions in semiconductor manufacturing depend on metrology, as shown in Figure 3.2:

- **Quality**: In charge of verifying that the products are according to the design specifications, analyzing whether the characteristics of the product are in accordance with ISO standards and verifying customer's satisfaction.
- **Process Control**: Monitors the information obtained from process equipment and metrology tools by treating data to avoid process drifts to reduce process variability and to improve performance parameters.
- **Reticle assembling**: The semiconductor area where new masks for the photolithography step are designed following the layout designs.
- **Facilities**: Ensure that the cleanroom run effectively by checking the contamination levels, the gas flow management, the water entries, the temperature conditions, the handling of the raw materials for processing and so on.
- **Research and development**: Identifies new possibilities of improvement, proposing the use of new materials, developing new process technologies, providing new recipes or recipe changes to process in the most efficient routes.
- **Device**: The device department, also known as process integration is in charge of understanding the characterization techniques, developing and implementing fabrication processes and generating standard process specifications.

- **Manufacturing**: Includes all process areas such as etch, photolithography, ion implantation, diffusion, etc.

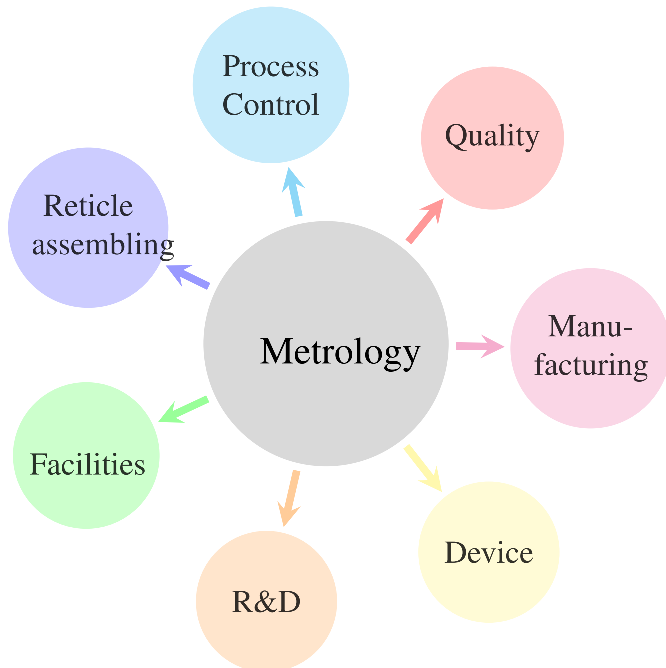



**Figure 3.2 –** *Influence of metrology in semiconductor functions.*

Semiconductor manufacturing is composed of a series of successive process stages in which wafers are heated, patterned and etched. The wafer surface is doped and films of various material are deposited to produce the integrated circuits. Along the production route, to guarantee the quality of final products, process monitoring is needed through required measurements that characterize physical parameters such as film thickness, uniformity and feature dimensions or electrical parameters such as resistance and capacitance [48]. The typical route of a product includes around 180 measures from start to shipping.

Metrology provides many benefits. The information obtained from measurements brings knowledge that will improve production (process improvements or effective actions). This is called “Productive Metrology” [35][38]. Metrology is the only way to achieve reliability. In order to ensure process repeatability to produce high-quality devices and circuits, each process stage must be strictly controlled [48] and the economic benefits of effective monitor policies increase with the complexity of manufacturing processes. The learning cycles, when electrical and physical data coming from a full-flow process are analyzed, are reduced (number and time) [23]. Metrology contributes with knowledge that will help to solve future problems or will improve the conditions for the next learning cycle. Thus, appropriate metrology practices can reduce manufacturing costs and cycle times of manufacturing through better characterization of tools and processes [87].

Because measurement operations are in series with respect to process operations, the main drawbacks of metrology are [35]:

- The work in process (WIP) is increased. By adding more metrology operations in the production route, the number of lots in the fab increases as well.
- The cycle time of the products is increased. It takes more time to complete a product, because extra time is necessary to perform metrology operations.
- The throughput is reduced, because the longer the lots remain in the production system, the more likely the process drifts to an undesirable state and starts manufacturing bad products.
- Extra costs and human resources are required.

Despite these drawbacks, metrology operations are required in wafer fabrication to deal with smaller and smaller transistor sizes. However, measurements could be minimized by using in an effective way. Each metrology workshop plays an important role. Its knowledge and capacity to adjust the settings of metrology tools, the use of appropriate metrics and procedures, the selection of the more suitable tools depending on the process and the regular verification of tool calibration will reduce the sources of error.

Also, the use of metrology is directly linked to the process or product maturity, as shown in Figure 3.3. The discovery of new materials used in processes entails the study of their impact in the process chain, how their chemical and electrical properties will perform, if they will react as expected with the other substances used on some process steps, if they will generate more particles or will contaminate more, and so on.

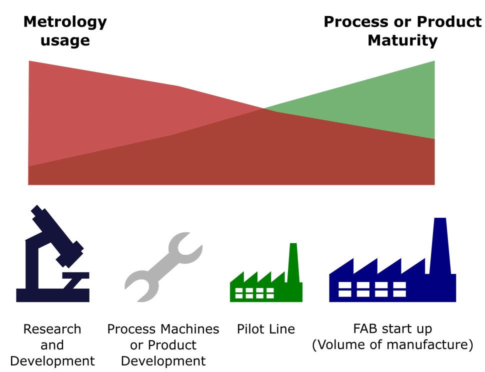



**Figure 3.3 –** *Metrology usage correlation with the process or product maturity.*

The same happens when new processes or production machines are introduced. It is mandatory to intensively control that the machines and check whether undesired defects are generated. In these initial stages, the use of metrology is required even more.

Once the new materials are well characterized and new processes or production machines are better known, the use of metrology and number of measures progressively decreases. Metrology tools must be also evaluated. There are two parameters that represent the precision of a measurement: Repeatability and reproducibility. Repeatability, see Figure 3.4, is the variation of repeated measurements made under identical conditions and reproducibility is the variation of the results when measurements are made under different conditions, such as reloading the wafer on different days. An example is presented in Figure 3.5. The precision of the measurement is calculated using (3.1), the square root of the sum of the squares of repeatability and reproducibility [23].

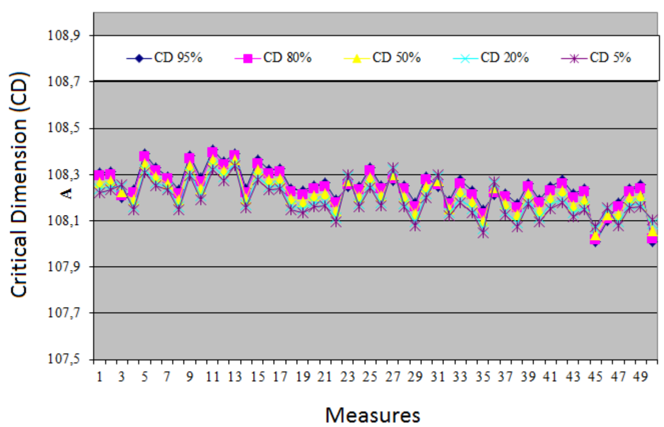



**Figure 3.4 –** *Repeatability chart for a CD measure, 50 measures were performed at the same point (STMicroelectronics).*

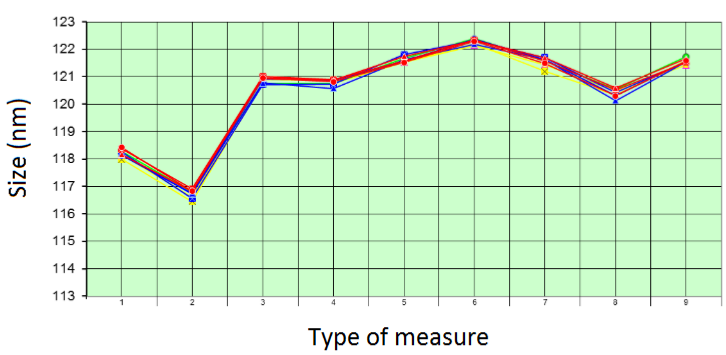



**Figure 3.5 –** *Reproducibility chart for 9 different measures. One measurement done per day on the same wafer (STMicroelectronics).*

$$\sigma_{measurement} = \sqrt{\sigma_{repeatability}{}^2 + \sigma_{reproducibility}{}^2} \tag{3.1}$$

In semiconductor manufacturing an additional complexity is that not all products are "measurable". This depends on the characteristics of the product and the type of measure. A recipe must be created for a metrology tool to perform a measure for a certain product in a given operation. A metrology recipe is composed of the configuration of the measurement parameters adapted to each measure type. Nowadays, around 430 metrology recipes are modified or created from scratch per week in the Rousset site of STMicroelectronics.

## 3.3 Analysis of metrology workshops

In order to improve the status of metrology workshops that exist in the Rousset site of STMicroelectronics, a study has been carried out taking into account several characteristics. The objectives are to detect the workshops that are most at risk, categorize each one, verify the impact of the metrology for the fab and find out possible improvement points. The workshop's characteristics are separated in four areas. A first area is related to the nature of the measure, its location in the fab, its difficulty to be qualified, the dispatching rules used that bring qualitative information to establish the link with high risk values. A second area is focused on the measurement and queue times and the sampling policies, which enable us to associate the risk levels with the time spent to perform a measure or the waiting times of the lots to propose new sampling or dispatching approaches. A third area concerns the sampling ratios and the global percentage of lots measured by workshop, that on one hand, let us verify whether the sampling rates are correct with respect to the risk levels, and on the other hand, provides information on which metrology workshops are more loaded. Finally, one last area contains risk data that allows us to identify metrology workshops that are not controlling consistently their assigned process machines. The different areas to analyze, as Figure 3.6 shows, are the following:

1. **Property table**: It comprises the wafer's verified properties, the associated metrology system (all types are explained in 1.3), the process steps controlled, the number of controlled process machines, the difficulty to qualify the metrology tools to perform the measure, their location in the fab and so on.
2. **Times and Sampling policies**: This table is composed with the measurement and queue average times for each metrology workshop, the number of previous metrology operations run before those and the current sampling policies. A complete review related to the types of sampling policies is presented in Chapter 2.
3. **Measure data**: The sampling rate values, the average of lots measured per week, and the corresponding percentage of lots measured for each metrology workshop are contained in this area.

4. **W@R values**: The values in terms of Wafer at risk (W@R) are represented here, their average and the average of maximum W@R values. The W@R indicator shows the number of wafers processed on a production machine since the last performed control, as it has been explained more in detail in Chapter 2.

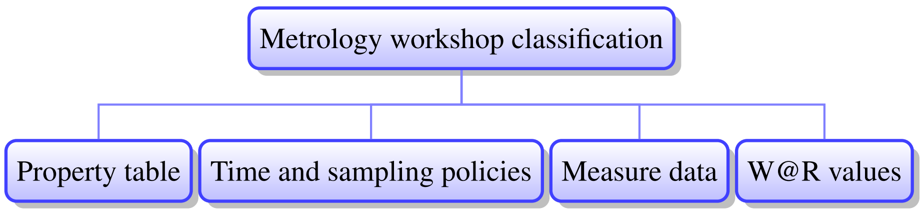



**Figure 3.6 –** *Metrology workshop classification organization chart.*

### 3.3.1 Property table

This section of the classification determines by metrology workshop (associated to a measure type) the physical properties of the wafer that are verified, the information obtained by the operators by using the metrology systems, the process steps controlled, the qualification difficulties, the situation of the metrology workshop with respect to the process machines in relation to the others. To better analyze the table, due to its size, it has been divided into four parts:

- **Table 3.1**: This part shows the metrology tools used, the wafer property verified and the information gathered, the process steps controlled, the number of metrology tools by metrology workshop and the number of process machines controlled.
- **Table 3.2**: Here is evaluated the difficulty to qualify the metrology tools to perform the measure, the differences between metrology tools that belong to a same metrology workshop (similar or different), whether the measure operation is automatic or performed by an operator, and the measurement time variability.
- **Table 3.3**: In this part is analyzed the lot dispatching strategy and the location of the metrology tools with respect to the production workshop (group of process machines) controlled.
- **Table 3.4**: The representability of the measure is shown and the frequency of defect detection as well.

#### 3.3.1.1 Property table (1/4)

Table 3.1 presents the first of the tables and as it can be observed, some of the most common metrology systems introduced in Section 1.3 are used like the Atomic Force Microscopy, the Ellipsometer and the CD-SEM for instance. The inspections for macro level scale are spread over the process steps by the measure types BIN, BLI, OIN and MDD. Apart from the measure systems for defectivity (DDM, DDP, DIP and DRS) that cover all process machines of the FAB, THM, which measures film thickness, covers the most with 6 control steps controlled (CVD, PVD, CMP, Diffusion, Photo and Etch) and 174 process machines.

**Table 3.1 –** *Property table (1/4): Types of process control methods.*

| General specifications | | | | | | ST specifications | |
|---|---|---|---|---|---|---|---|
| **Measure type** | **Metrology system** | **Wafer property verified** | **Metrology tool output** | **Process step controlled** | **Destructive or non-destructive** | **Number of metrology tools** | **Number of process machines controlled** |
| AFM | Atomic Force Microscopy | Mechanical properties of wafer surface, surface profile | Topographic images | Etch | Non destructive | 1 | 10 |
| BIN | Manual Optical Detection | Macro level: scratches on surface, particles | Visual inspections of wafer surface | CMP | Non destructive | 2 | 16 |
| BLI | Manual Optical Detection | Macro level: scratches on surface, particles | Visual inspections of wafer surface | CVD, PVD | Non destructive | 3 | 77 |
| CDS | Critical Dimension Scanning Electron Microscopy (CDSEM) | Critical Dimension(CD) | Top view of wafer structure, Critical Dimension(CD) | Photo, Etch | Non destructive | 9 | 67 |
| DDM | Scanning Electron Microscopy(SEM) | Micro level: Particles, scratches, crystalline defects, corrosions, voids on wafer surface | Very high-resolution images of wafer surface | All FAB | Non destructive | 6 | All process machines of the FAB |
| DDP | Scanning Electron Microscopy(SEM) | Micro level: Particles, scratches, crystalline defects, corrosions, voids on wafer surface | Very high-resolution images of wafer surface | All FAB | Non destructive | 4 | All process machines of the FAB |
| DIP | Scanning Electron Microscopy(SEM) | Micro level: Particles, scratches, crystalline defects, corrosions, voids on wafer surface | Very high-resolution images of wafer surface | All FAB | Non destructive | 7 | All process machines of the FAB |
| DRS | Scanning Electron Microscopy(SEM) | Micro level: Particles, scratches, crystalline defects, corrosions, voids on wafer surface | Very high-resolution images of wafer surface | All FAB | Non destructive | 2 | All process machines of the FAB |
| IDM | Therma Wave | Crystallographic defects on wafer surface and Dose uniformity | Wafer mapping | Ion Implantation | Non destructive | 2 | 18 |
| MDD | Manual Optical Detection | Macro level: scratches on surface, particles | Visual inspections of wafer surface | CMP, Photo | Non destructive | 2 | 27 |
| OIN | Manual Optical Detection | Macro level: scratches on surface, particles | Visual inspections of wafer surface | CMP, Photo, Etch | Non destructive | 5 | 101 |
| OVL | Overlay | layer-to-layer alignment | Alignment offset of the wafer respect to the reticle | Photo | Non destructive | 5 | 29 |
| TCU | Reflectometer | Thickness of the deposited layer | Thickness - Distance charts | CMP-CU | Non destructive | 7 | 8 |
| THA | Optical and acoustic system | Thickness of the deposited layer | Thickness - Distance charts | PVD, CMP-CU, Etch | Non destructive | 2 | 38 |
| THM | Ellipsometer and Scattero measures | Film thickness, optical constants, etch selectivity, removal rate | Thickness - Distance charts, Intensity charts, Reflectance - Wavelength charts | CVD, PVD, CMP, Diffusion, Photo, Etch | Non destructive | 11 | 174 |
| THP | Ellipsometer | Film thickness, optical dielectric function information, physical composition, roughness | Thickness - Distance charts, Intensity charts, Reflectance - Wavelength charts | CMP | Non destructive | 10 | 10 |

#### 3.3.1.2 Property table (2/4)

The next part of the property table is focused on the qualification part of the measure types, the differences between metrology tools, the operating of the measurements and the measurement time's variability. Table 3.2 summarizes this section.

**Table 3.2 –** *Property table (2/4).*

| | Difficulty of measurement qualification | | | Metrology tool type | | Automatic/Manual operation | | Measurement time variability | |
|---|---|---|---|---|---|---|---|---|---|
| **Measure type** | NO QUALIFICATION NEEDED | EASY | HARD | SIMILAR | DIFFERENT | AUTO | MANUAL | NO | YES |
| AFM | | X | | X | | X | | | X |
| BIN | X | | | | X | | X | | X |
| BLI | X | | | X | | | X | | X |
| CDS | | X | | | X | X | | | X |
| DDM | | | X | X | | X | | | X |
| DDP | | | X | X | | X | | | X |
| DIP | | | X | | X | X | | | X |
| DRS | | | X | X | | X | | | X |
| IDM | | | X | | X | X | | | X |
| MDD | | | X | X | | X | | | X |
| OIN | X | | | | X | | X | | X |
| OVL | | X | | | X | X | | | X |
| TCU | | X | | | X | X | | | X |
| THA | | X | | | X | X | | | X |
| THM | | X | | | X | X | | | X |
| THP | | X | | | X | X | | | X |

One of the most important things to consider for achieving a risk improvement for a metrology workshop is the degree of difficulty for the engineers to qualify a new product to measure.

In Table 3.2, it is assumed that qualifying a new product to measure means qualifying a product that is already mature, which was not measurable before and not qualifying a new product as such. Qualifying a product in a metrology tool means creating the recipe on the metrology tool to allow the machine to measure the lots belonging to a certain product and process by a given process machine.

To qualify a metrology tool, in the case of TCU/THM/THA/THP, is not time consuming and the time invested to create the recipe is around 4 hours which covers all routes and process operations of the product. Normally if it is about measuring a new non-mature product that is going to be introduced, the time required to create the recipe is longer than the products already known. It takes some weeks to discuss with the research and development department, and understand the features of the new product to control in order to carefully create the parameters for the machine and so on.

In the case of defectivity of DDM/DDP/DIP/DRS metrology tools, to qualify a metrology tool, it is hard to create the recipe and it is time consuming because the engineers must create a recipe for each product and process operation leading to time spent and wasted and huge amounts of recipes. In case of IDM, it is also hard to create metrology recipes an old software is used. However, there are some metrology workshops that do not need a qualification step to use the measurement tool, which is the case of BIN, BLI and OIN where the measure consists in just looking at the surface of the wafer and not doing any preliminary study about the possible interactions of new products with the metrology tools to set parameters that must be introduced in the tool.

The way to classify if tools are equivalent or different in a metrology workshop will be explained through our approach in Section 3.3.2.3. It is also important to take into account whether the metrology step is manual (made by an operator) or automatic (made directly by the metrology tool) because human intervention will create some variability and increase the queue time.

#### 3.3.1.3 Property table (3/4)

This part of the property table provides the different dispatching strategies adopted by each metrology workshop and their location inside the fab, and all of these considerations are shown in Table 3.3.

For the workshops BIN, BLI, TCU and THP the strategy used to measure the lots is a FIFO (First In, First Out) policy. For the defectivity workshop (DDM, DDP, DIP, DRS), the operators send lots to measure using a system called WIPPER D0 that creates the lot order according to time constraints, W@R levels, and service priorities, and priorities established daily are also considered. For the other workshops, the order is established by the team manager of the operators for each metrology tool that indicates which lots from a given process machine must be measured first.

Concerning the location of the metrology workshops with respect to the process workshops controlled, there is no metrology workshop considered that performs the measurement operation inside the process machine (in-situ measure) and neither outside the fab. In the Rousset site of STMicroelectronics, the farthest metrology workshops are those related to defectivity which are located in a lower floor of the fab, with the exception of the DDM that is in between the controlled process workshop, since it is in charge of doing only quality tasks.

**Table 3.3 –** *Property table (3/4).*

| | Lot dispatching strategy | | | | | Measurement location respect to the workshop controlled | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Measure type** | FIFO | LIFO | RANDOM | PRIORITY ORDER | OTHER | IN-SITU | INSIDE WORKSHOP | BETWEEN WORKSHOPS | SEPARATED AREA | OUTSIDE FAB |
| AFM | | | | X | | | | X | | |
| BIN | X | | | | | | | | X | |
| BLI | X | | | | | | | X | | |
| CDS | | | | X | | | | X | | |
| DDM | | | | X | | | | X | | |
| DDP | | | | X | | | | | X | |
| DIP | | | | X | | | | | X | |
| DRS | | | | X | | | X | | X | |
| IDM | | | | X | | | X | | | |
| MDD | | | | X | | | X | | | |
| OIN | | | | X | | | X | | | |
| OVL | | | | X | | | X | | | |
| TCU | X | | | | | | X | | | |
| THA | | | | X | | | | X | | |
| THM | | | | X | | | | X | | |
| THP | X | | | | | | X | | | |

The AFM, BLI, CDS, THA, THM workshops are in a shared space between the process machines covered. And the remaining metrology systems are near the process machines to be controlled.

#### 3.3.1.4 Property table (4/4)

The last part of the property table is presented in Table 3.4. It shows the representability of the measure when a lot is sent to the measure, as well as the frequency of defect detection.

The majority of the metrology workshops measure some wafers of a lot, which are sometimes always the same (for instance for defectivity: the first wafer, some wafer in the middle, and the last wafer), sometimes are random, and sometimes depend of the previous process operation. Rarely are all wafers of a lot measured, which is the case only for TCU and THP metrology workshops, because they cover an important step that needs to be controlled. For the THM and CDS workshops, at least one wafer coming from every chamber of the process machine will be measured.

**Table 3.4 –** *Property table (4/4).*

| | **Representability of the measure by lot** | | | **Frequency of defect detection** | | | |
|---|---|---|---|---|---|---|---|
| **Measure type** | ALL WAFERS | A WAFER BY CHAMBER | SOME WAFERS | never | low | medium | high |
| AFM | | | X | | X | | |
| BIN | | | X | X | | | |
| BLI | | | X | X | | | |
| CDS | | X | X | | | X | |
| DDM | | | X | | | X | |
| DDP | | | X | | | X | |
| DIP | | | X | | | X | |
| DRS | | | X | | | X | |
| IDM | | | X | | X | | |
| MDD | | | X | | X | | |
| OIN | | | X | | X | | |
| OVL | | | X | | X | | |
| TCU | X | | | | | X | |
| THA | | | X | | X | | |
| THM | | X | | | | X | |
| THP | X | | | | X | | |

## 3.3.2 Times and Sampling policies

Table 3.5 is called "Times and Sampling policies", and includes the measurement times and queue times calculated over 6 months by metrology workshop. To know the gap between the process step controlled and the metrology step, the average number of previous metrology operations has been calculated before performing those that belong to the metrology workshop. The BLI metrology workshop always performs its measure first right after the process step, and at the other extreme there is the OIN measure which has an average of around 2 metrology steps before its measure.

### 3.3.2.1 Sampling policies

Concerning the Sampling policies, in case of THP and TCU, all lots are measured, and for THA and THM all process families and process operations have their own sampling ratio. For AFM, IDM and MDD, in the static way, only some products (one product contains many process families) are qualified and all these lots are 100% measured and some process families and process operations have their own sampling ratio.

The defectivity workshop (DDP, DIP, DRS) follows a mix between sampling by product and by equipment. Some products are qualified and are measured using a given sampling ratio, and for the coverage of the process machines, dynamic policies are used, following a sophisticated method [76] to calculate the current risk levels, and depending on them, some

**Table 3.5 –** *Times and Sampling policies.*

| | Time | | | Sampling by product or process family | | | | Sampling by machine | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Measure type** | Measurement average (seconds) | Queue average (seconds) | **Average of Metrology Operations before this measure** | STATIC | ADAPTIVE | DYNAMIC | OTHER | STATIC | ADAPTIVE | DYNAMIC | OTHER |
| AFM | 2120 | NOT AVAILABLE | 0.16 | Some Prods, 100% | | Some Process families and Oper, 1/X | | | | | |
| BIN | NOT AVAILABLE | NOT AVAILABLE | 0.63 | | | | | W@R, 1/X | | | |
| BLI | 231 | 813 | 0 | | | | | W@R, 1/X | | | |
| CDS | 426 | 656 | 0.09 | | | All process families and Oper,1/X | | | | | |
| DDM | 217 | 4140 | 0.25 | | | | | X | | X | |
| DDP | 624 | NOT AVAILABLE | 0.28 | Some Prods, 1/X | | | | | | W@R | |
| DIP | 2576 | NOT AVAILABLE | 0.29 | Some Prods, 1/X | | | | | | W@R | |
| DRS | NOT AVAILABLE | NOT AVAILABLE | 1.43 | | | | | | | | |
| IDM | 281 | 947 | 0 | Some Prods, 100% | | Some Process families and Oper, 1/X | | | | W@R | |
| MDD | 293 | 1742 | 1.99 | Some Prods, 100% | | Some Process families and Oper, 1/X | | | | | |
| OIN | NOT AVAILABLE | NOT AVAILABLE | 2.27 | | | | | W@R, 1/X | | | |
| OVL | 100 | 1320 | 1.28 | | | | No sampling | | | | No sampling |
| TCU | NOT AVAILABLE | NOT AVAILABLE | 1 | | | | All lots measured | | | | All lots measured |
| THA | 369 | 2011 | 1.12 | | | All process families and Oper,1/X | | | | | |
| THM | 252 | 1491 | 0.50 | | | All process families and Oper,1/X | | | | | |
| THP | NOT AVAILABLE | NOT AVAILABLE | 0.70 | | | | All lots measured | | | | All lots measured |

lots are skipped and the more recent lots are measured. A special procedure is followed for the DDM that only perform quality tasks that cover the process machines in a static way (1 lot measured for each day) and in a dynamic way (1 lot measured each 1000 wafers processed).

The inspection measures (BIN, BLI, OIN) only use a sampling by equipment, with a fixed sampling ratio: 1/X, and W@R levels are controlled, and for OVL almost all lots are measured due to its huge metrology capacity.

#### 3.3.2.2 Measurement times

The study concerning measurement times evidenced a variety of values which depended on the measure type. Two measure types take longer to measure with an average above 1000 seconds, the remaining nine are below. OVL is the fastest measure and DIP is the slowest one, as it can be seen in Figure 3.7 and Figure 3.8.

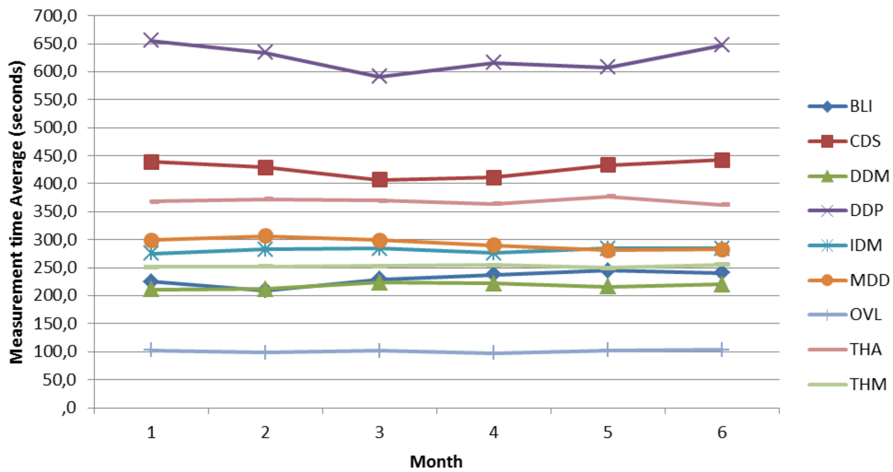



**Figure 3.7 –** *Average of measurement times for metrology workshops BLI, CDS, DDM, DDP, IDM, MDD, OVL, THA and THM.*

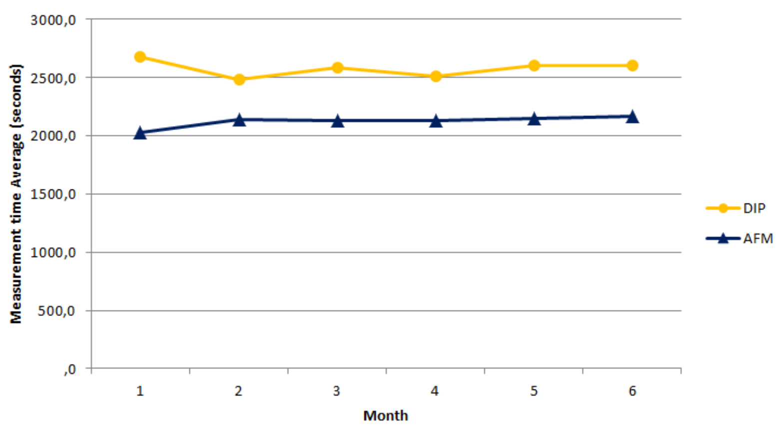
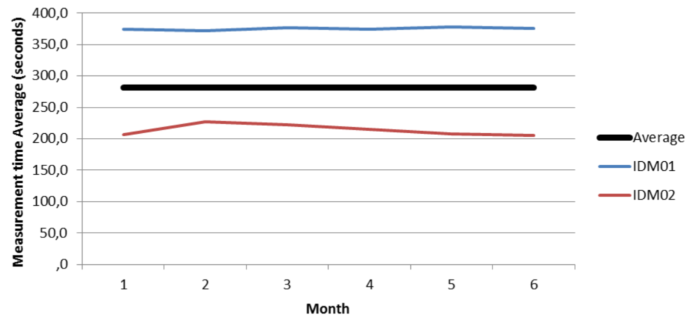
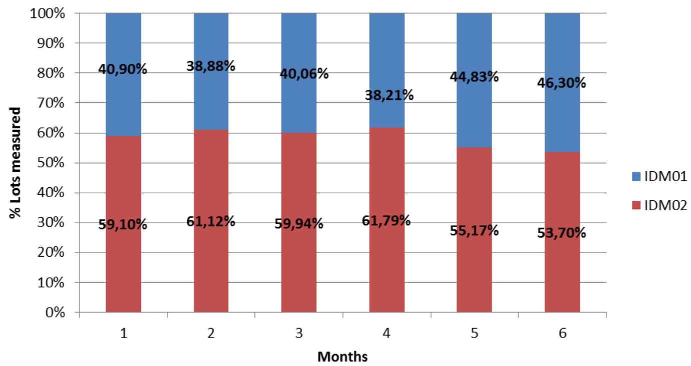



**Figure 3.8 –** *Average of measurement times for metrology workshops AFM and DIP.*

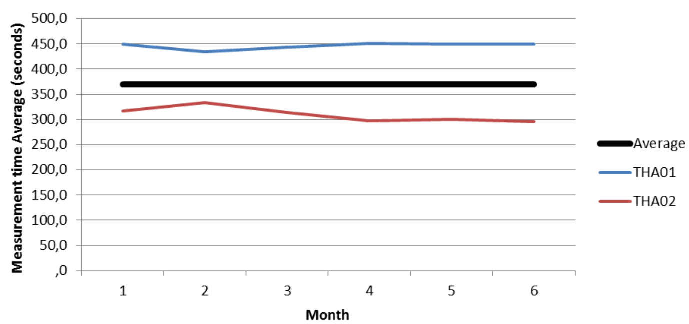
It can be observed that, for some of the metrology workshops, the fastest metrology tool measures a larger percentage of lots, such as the IDM (IDM02 faster than IDM01) and THA (THA02 faster than THA01) workshops as shown in Figure 3.9.



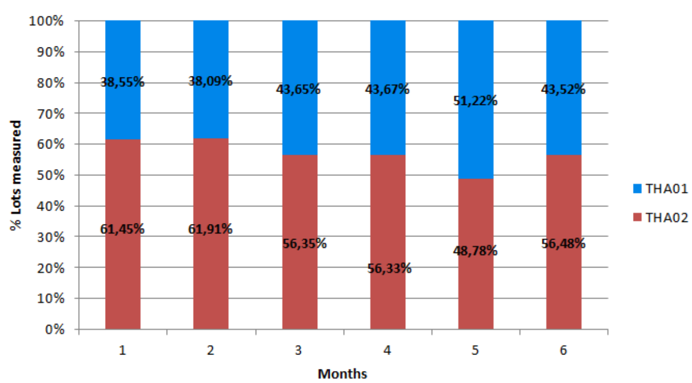

**(a)** *Measurement times - IDM*




**(b)** *Lots measured by metrology tool - IDM*



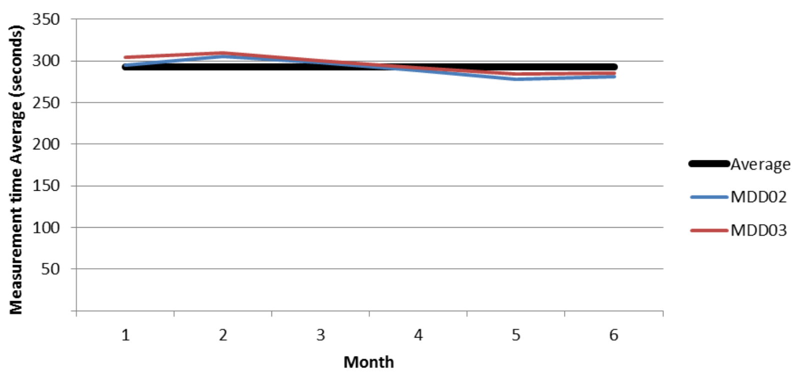

**(c)** *Measurement times - THA*



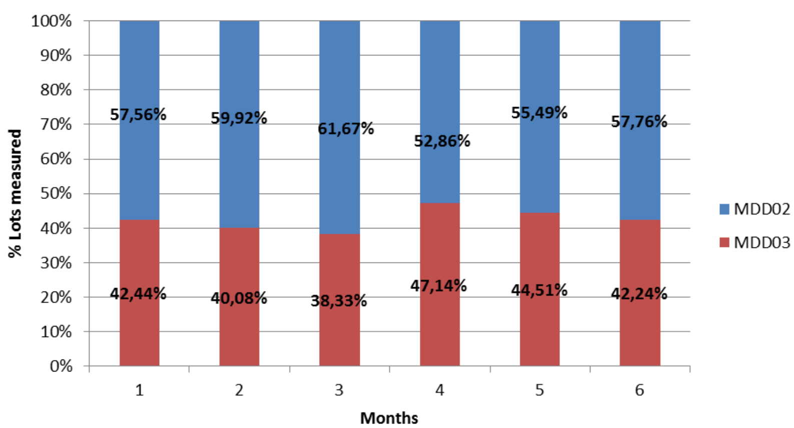

**(d)** *Lots measured by metrology tool - THA*

**Figure 3.9 –** *Average of measurement times and percentages of lots measured for IDM and THA.*

There are other cases, like for MDD (see Figure 3.10) where the measurement times are quite similar. The measurement time differences are almost zero. However, MDD02 always measures more lots than MDD03, maybe because of qualification reasons.



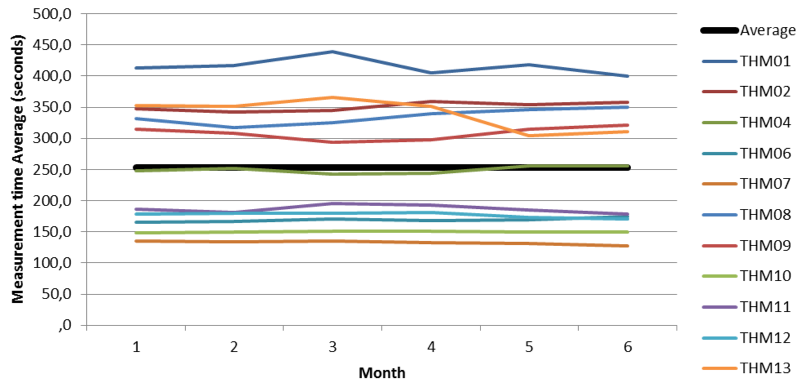

**(a)** *Measurement times - MDD*



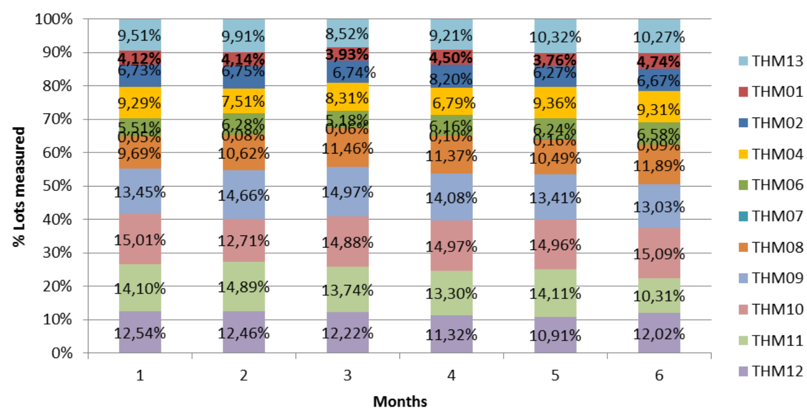

**(b)** *Lots measured by metrology tool - MDD*

**Figure 3.10 –** *Average of measurement times and percentages of lots measured for MDD.*

In the case of THM measurement, it is shown in Figure 3.11 that there are five metrology tools out of the eleven which, more or less, are in the same range (between 150 and 200 seconds to measure). Nevertheless, the other metrology tools present larger differences, and THM01 is the slowest tool. Concerning the number of lots measured by metrology tool, the fact that each month the percentage of lots measured does not vary excessively is significant, which means that this metrology workshop is constantly under control.




**(a)** *Measurement times - THM*




**(b)** *Lots measured by metrology tool - THM*

**Figure 3.11 –** *Average of measurement times and percentages of lots measured for THM.*

#### 3.3.2.3 Metrology tool differentiation approach

How can it be determined whether the metrology tools for the same measure are similar or not? In order to distinguish between metrology tools and to classify them into similar or different tools, the following approach is proposed:

1. Calculate the average of measurement times for each metrology tool.
2. Find the fastest metrology tool (smallest measurement time).
3. Divide all measurement times by the fastest one.
4. Establish a threshold, $T$, to differentiate one metrology tool from another, in our case, $T = 10\%$ has been chosen.

This approach has been applied, and some cases are presented in Table 3.6.

**Table 3.6 –** *Metrology tool differentiation approach.*

| Measure type | Metrology tool | Measurement time average (s) | Fastest measurement time (s) | Difference between fastest tool | Gap with the fastest tool | Similar or Different tool |
|---|---|---|---|---|---|---|
| MDD | MDD02 | 291 | 291 | 100% | - | SIMILAR |
| MDD | MDD03 | 296 | 291 | 98% | 2% | SIMILAR |
| BLI | BLI01 | 242 | 227 | 94% | 6% | SIMILAR |
| BLI | BLI02 | 227 | 227 | 100% | - | SIMILAR |
| BLI | BLI03 | 241 | 227 | 98% | 2% | SIMILAR |
| IDM | IDM01 | 375 | 214 | 57% | 43% | DIFFERENT |
| IDM | IDM02 | 214 | 214 | 100% | - | DIFFERENT |
| THA | THA01 | 446 | 309 | 69% | 31% | DIFFERENT |
| THA | THA02 | 310 | 309 | 100% | - | DIFFERENT |
| OVL | OVL03 | 143 | 80 | 56% | 44% | DIFFERENT |
| OVL | OVL04 | 131 | 80 | 61% | 39% | DIFFERENT |
| OVL | OVL05 | 80 | 80 | 100% | - | DIFFERENT |
| OVL | OVL06 | 80 | 80 | 100% | - | DIFFERENT |
| OVL | OVL07 | 96 | 80 | 83% | 17% | DIFFERENT |

As shown in Table 3.6, according to our criteria, some metrology tools are similar and others are different. For instance, for the MDD measure, the fastest metrology tool is MDD03 with an average measurement time of 291 seconds, and the difference with the other tool is 2%; thus both tools are considered as equal. Another case of similar tools concerns the BLI measure, of which the fastest is BLI02 and the gap with the other two tools is lower than 10%, and therefore they are considered to be similar tools.

There are metrology workshops composed of different types of tools, as presented in Table 3.6 with the examples of the IDM, THA and OVL measures. For IDM it is quite easy to see that the two metrology tools are different with a gap of 43%, as well as for the THA with a gap of 31%. However, for the OVL measure, three metrology tools (OVL03, OVL04 and OVL07) are clearly different compared to the fastest tool OVL06, but the metrology tool OVL05 has an average measurement time equal to OVL06. However, since there is at least one metrology tool which is different in the metrology tool group, all metrology tools are considered to be different.

#### 3.3.2.4 Queue times

In order to understand high risk levels, to decide whether to change the dispatching rules or implement a new sampling strategy, it is important to know how long the lots have been waiting to be measured. In this study, as it is represented in Figure 3.12, the queue time for a measured lot is considered from the completion of the process operation until the start of the measure operation.

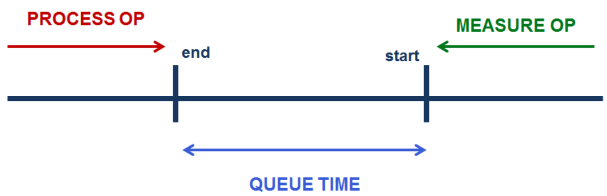



**Figure 3.12 –** *Queue time consideration.*

Due to the complexity of extracting data, the information related to the queue times has been obtained for only for 9 metrology workshops, as shown in Figure 3.13, for each measure type and for six months. Lots measured by DDM are those that wait longer after the process step. These metrology tools are special because they only measure lots that perform quality tasks when a limit of any given number of lots are processed or limited in time (for instance, a quality task is performed if during 24 hours no measure was done). The measure with a lower value of queue time is CDS, which is natural, because it has 9 metrology tools that are spread out between workshops.

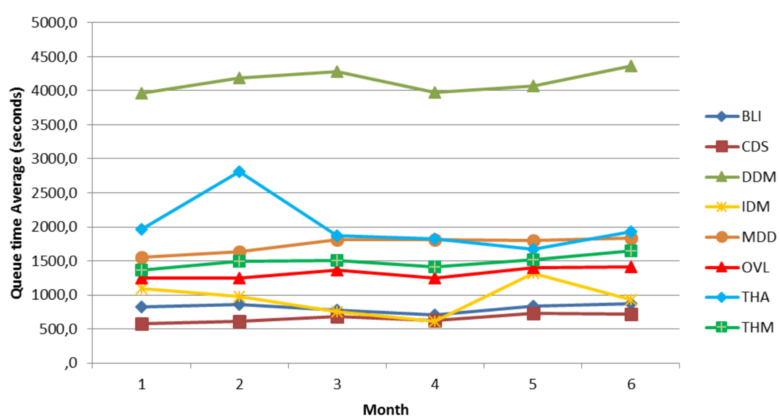



**Figure 3.13 –** *Average of queue times by metrology workshop during six months.*

For the average of queue times for all periods shown in Figure 3.14, again, DDM has the largest value, and CDS the lowest one. But these values bring two questions: How do we know if these values can be considered as reasonable? Is it possible to reduce them by applying other dispatching rules or implementing new sampling strategies? These questions will be tackled in Section 3.4.

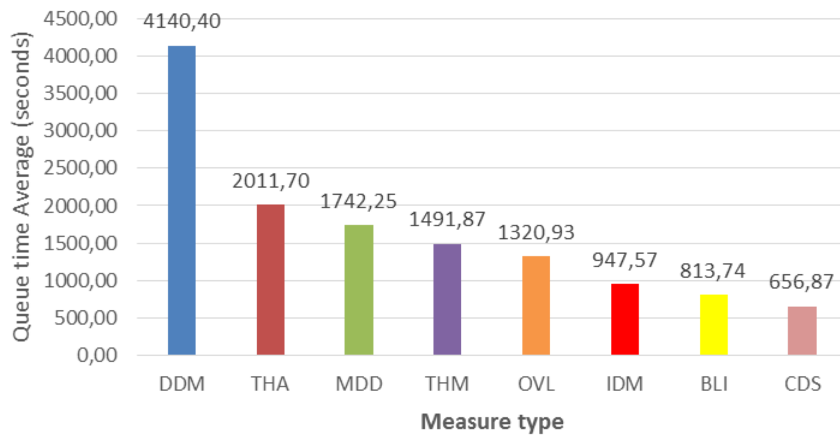



**Figure 3.14 –** *Average of queue times for the metrology workshops.*

### 3.3.3 Measure data

In Table 3.7, the sampling rates and the number of lots measured per week have been collected over 6 months. Some metrology workshops were removed from the table because it was not possible to get data.

**Table 3.7 –** *Measure data.*

| | Sampling rate | | | | |
|---|---|---|---|---|---|
| **Measure type** | Average total | Average minimum | Average maximum | **Lots measured per week** | **%Lots measured** |
| AFM | 56.0 | 2.3 | 278.2 | 338 | 0.9% |
| BIN | 3.1 | 1.9 | 3.8 | 743 | 1.9% |
| BLI | 5.4 | 1 | 25.7 | 2340 | 6.0% |
| CDS | 2.2 | 1 | 10.1 | 12407 | 31.9% |
| DDM | 55.2 | 2.8 | 202.8 | 239 | 0.6% |
| IDM | 14.0 | 5.6 | 41.2 | 1041 | 2.7% |
| MDD | 6.9 | 2.9 | 14.1 | 1918 | 4.9% |
| OIN | 190.7 | 1 | 662.9 | 1030 | 2.6% |
| OVL | 1.6 | 1 | 6.4 | 9475 | 24.4% |
| THA | 10.1 | 1 | 37.4 | 901 | 2.3% |
| THM | 34.8 | 1 | 668.9 | 8451 | 21.8% |

#### 3.3.3.1 Sampling rates

The sampling rate the number of lots processed before performing a measure, and the **average** (over time and over all machines in a workshop) sampling rate has been calculated by metrology workshop using (3.2).

$$\frac{\sum_{m=1}^{M} \frac{\sum_{w=1}^{W} \frac{LP_{m,w}}{LM_{m,w}}}{W}}{M} \tag{3.2}$$

Where $M$ is the number of machines, $W$ is the number of weeks, $LP_{m,w}$ is the number of lots processed by machine $m$, in week $w$, $LM_{m,w}$ is the number of lots measured that has been processed by machine $m$ in week $w$.

The minimum and maximum average (over time) sampling rates among all machines are calculated using (3.3) and (3.4).

$$\text{Min}_{m=1,\dots,M} \frac{\sum_{w=1}^{W} \frac{LP_{m,w}}{LM_{m,w}}}{W} \tag{3.3}$$

$$\text{Max}_{m=1,\dots,M} \frac{\sum_{w=1}^{W} \frac{LP_{m,w}}{LM_{m,w}}}{W} \tag{3.4}$$

On the one hand, among all average sampling rates for each metrology workshop, we found reasonable values such as 3.1, 5.4, 2.2, 14, 6.9, 1.6, and 10.1, for metrology workshops BIN, BLI, CDS, IDM, MDD, OVL, and THA respectively. Even if the difference between the minimum and maximum average sampling rates is quite large as for BLI (its minimum value is 1 and its maximum value is 25.7) or IDM (its minimum value is 5.6 and its maximum value is 41.2). The metrology workshop with the least difference is BIN, with all its process machines well distributed, as observed in Figure 3.15, with an average sampling rate of 3.1, a minimum average sampling rate of 1.9 and a maximum average sampling rate of 3.8.

On the other hand, we found a workshop with a pretty high average sampling rate. For example, the average sampling rate for the ten process machines of the AFM measure type was 56, as can be seen in Figure 3.16. Except for machines 2 and 5, with average sampling rates of 246.7 and 278.2 respectively because they only have 7 and 4 lots measured, each one during 6 months, the other process machines have an average sampling rate of 4.4 and an average number of measured lots of 42.1. Thus, for some metrology workshops, there are

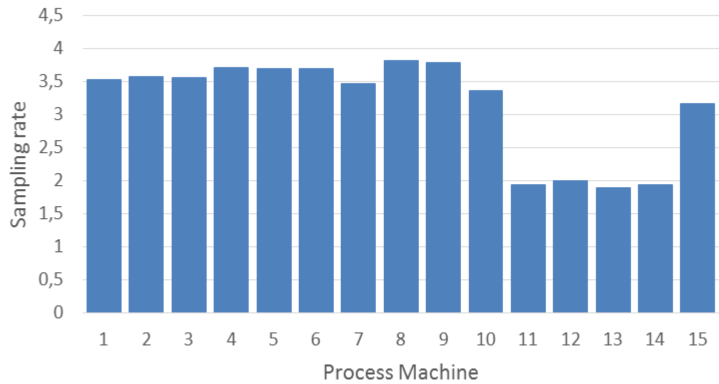



**Figure 3.15 –** *Sampling rates of process machines of BIN metrology workshop.*

machines that are less controlled (in this example machines 2 and 5, with an average number of processed lots per week of 263.9 and 281.6, and an average number of measured lots of 0.3 and 0.2). Maybe a limit should be used so that these cases will not count, to not interfere with and impact the global value of the average sampling rate.

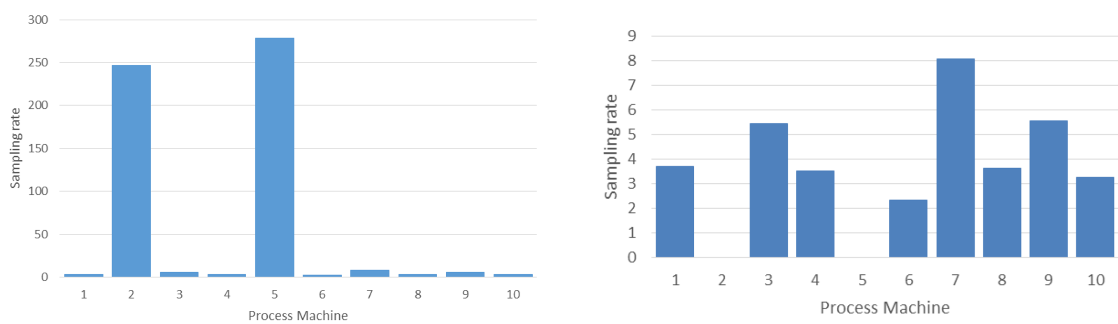



**Figure 3.16 –** *Sampling rates of Process machines of AFM metrology workshop.*

#### 3.3.3.2 Lots measured per week

There are metrology workshops more important than others, which depend on the property of the measured wafer, the complexity of the process operation performed, the difficulty of getting a good final result or the high risk of the production machines. As it can be seen in Figure 3.17, the three metrology workshops with the larger volume of measured lots are CDS, OVL and THM, which measure the Critical Dimension, the alignment of the different layers and the thickness of the wafer. The average of sampling rates for CDS and OVL are 2.2 and 1.6 respectively, which means that these two metrology workshops are important and have more capacity.

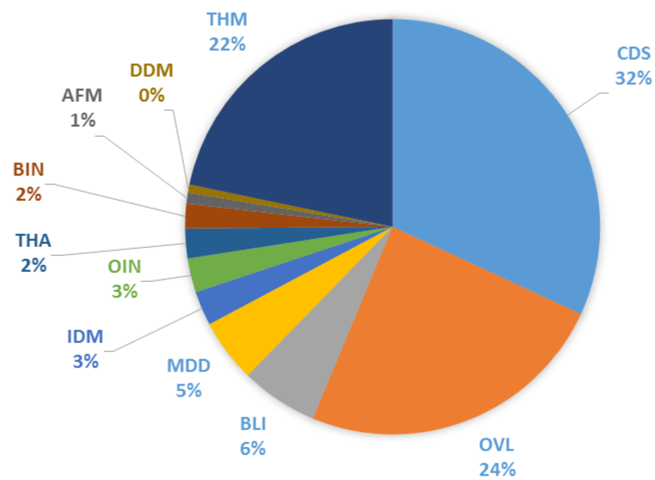



**Figure 3.17 –** *Weight of every metrology workshop in terms of lots measured.*

### 3.3.4 W@R values

The last table is focused on the values of the indicator called wafer at risk (W@R, or wafer at risk is the number of wafers processed since the last measure) of process machines covered by each metrology workshop.

The calculation of the W@R has been divided into two points of view. The first one, is the average of W@R values per week and for each machine of every metrology workshop during 6 months (25 weeks). For a given process machine $i$ its **Average W@R** has been calculated as shown in (3.5), being $W = 25$.

$$\frac{\sum_{w=1}^{W} \frac{\sum \text{W@R} \times \text{time (seconds)}}{\text{Total time in a week (seconds)}}}{W} \tag{3.5}$$

A typical profile of the average W@R of machine $i$ during a week is shown in Figure 3.18. Note that, in some periods of the week, the lots with 25 wafers, are processed with positive risk levels for some time, until the lot is measured and the risk is reduced to zero.

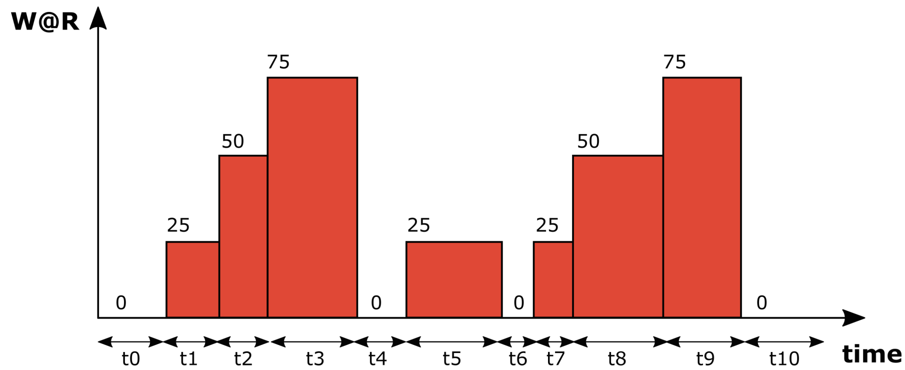



**Figure 3.18 –** *Example of an Average W@R chart for a process machine during a week.*

The maximum values of W@R reached before performing the measure (i.e. when the W@R is reduced) have been recorded in the same period. For a given process machine *i*, the average of maximum W@R values has been calculated as shown in (3.6):

$$\frac{\sum_{w=1}^{W} \frac{\sum \text{W@R value achieved by the process machine i before the measure}}{\sum \text{measures of lots processed by machine i}}}{W} \tag{3.6}$$

The maximum values of W@R for a process machine are the peaks in Figure 3.19.

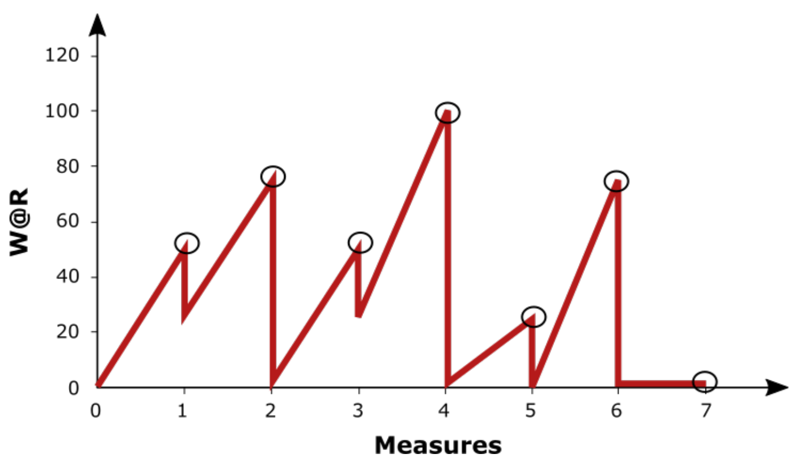



**Figure 3.19 –** *Example of a chart of the maximum W@R values for a process machine during a week.*

The results are shown in Table 3.8.

**Table 3.8 –** *W@R values by metrology area.*

| | W@R | | | | | | |
|---|---|---|---|---|---|---|---|
| | Values by machine | | | Maximum values by machine | | | |
| **Measure type** | Average total | Average minimum | Average maximum | Average total | Average minimum | Average maximum | Maximum value achieved |
| AFM | 4911 | 62 | 26520 | 4080 | 55 | 23694 | 51247 |
| BIN | 35 | 19 | 48 | 74 | 46 | 92 | 351 |
| BLI | 101 | 0 | 581 | 134 | 23 | 640 | 2953 |
| CDS | 60 | 11 | 515 | 54 | 19 | 245 | 6159 |
| DDM | 10174 | 33 | 119813 | 4620 | 70 | 34266 | 79010 |
| IDM | 423 | 167 | 1114 | 347 | 152 | 1014 | 15551 |
| MDD | 160 | 60 | 372 | 169 | 71 | 344 | 3745 |
| OIN | 10572 | 0 | 78024 | 8099 | 24 | 38867 | 84186 |
| OVL | 33 | 15 | 170 | 40 | 23 | 154 | 1195 |
| THA | 229 | 8 | 761 | 247 | 29 | 877 | 4902 |
| THM | 5857 | 0 | 147445 | 2312 | 21 | 63314 | 102134 |

The metrology workshop with the most stable W@R values is BIN, with a total average of 35, where the minimum W@R for a production machines is 19 and the maximum W@R is 48. Figure 3.20 gathers the cumulative average W@R of all the production machines per week. The average maximum W@R is 74, close to the value of the Average W@R, and a maximum value of 351 during 6 months.

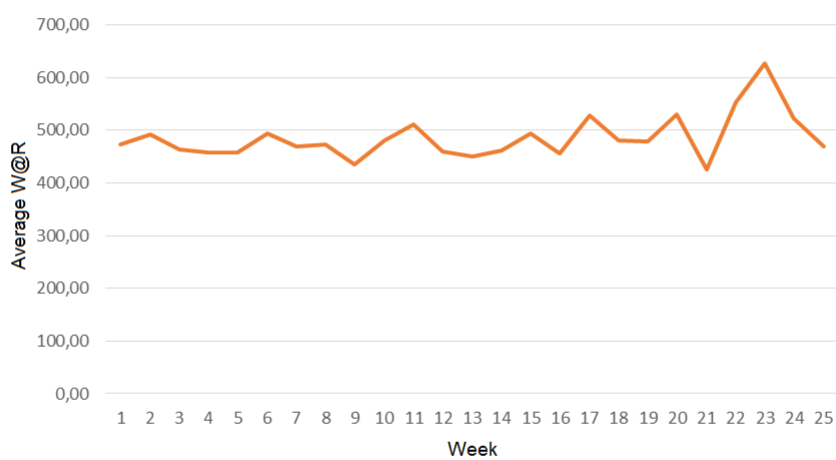



**Figure 3.20 –** *Sum of the average W@R of all process machines covered by the BIN measure*

In the case of the BLI measure, the average W@R is 101, with a minimum value of 0 and a maximum 581. The value of 0 is due to the fact that one machine processes only one or two each week, and they are always measured. The majority of process machines have an average

W@R of around 50 except for some with a larger W@R, which increases the final average, and the average W@R profile of all the process machines is represented in Figure 3.21. The maximum W@R reached is 2953 which, compared with an average maximum W@R of 134, means that some process machines could be uncontrolled compared to others. The minimum W@R of 23 means that at least one process machine has all of its lots measured. Similar cases are the CDS, MDD and OVL measures.

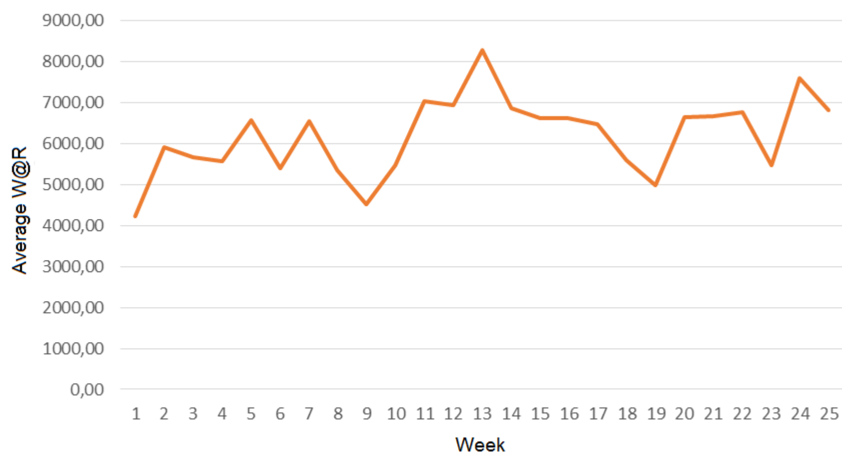



**Figure 3.21 –** *Sum of the average W@R of all process machines covered by the BLI measure*

The impact of some process machines could provide large W@R values. For instance, the AFM measure has an average W@R of 4911, with a minimum average of 62 and a maximum of 26520. When diving into the data, 2 of the 10 process machines covered by this measure have 4 and 7 measures for the 6 months considered, when the other machines have an average of 42 measures per week. The average W@R would be 176 if these two uncontrolled machines (with W@R averages of 21187 and 26520) would not have been taken into account. In Figure 3.22, it can be observed that, from week number 3, the average W@R starts to increase due to these machines, its lots are processed but not measured and it is not until week 13 that some measures are performed. There are two possible explanations: (1) two pretty risky machines are left totally uncontrolled, or (2) the engineers are controlling these machines through another metrology system. In this latter case, the small volume of lots measured is only tested or taken as alternatives of the measures and thereby the two process machines should not be covered by this metrology workshop anymore or at least for the W@R calculation. The same happens for other metrology workshops such as THM and DDM.

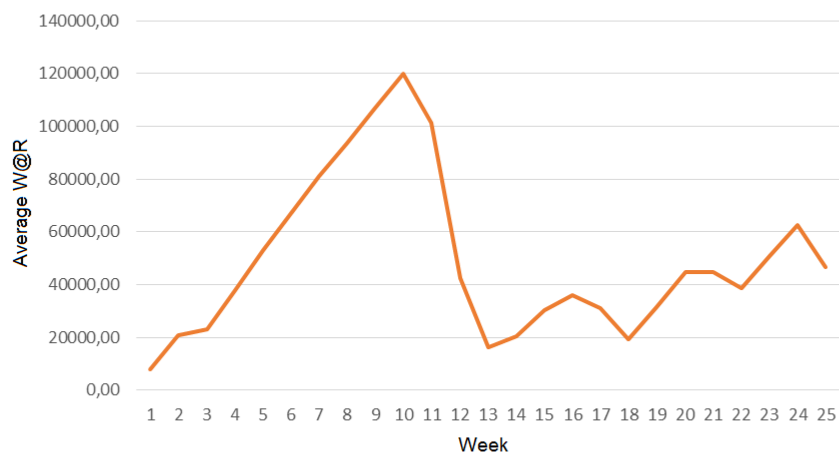



**Figure 3.22 –** *Sum of the average W@R of all process machines covered by the AFM measure*

The IDM measure comprises machines with pretty large W@R values but homogeneous. The minimum value of the average W@R of the machines controlled is 167 and the remainder of the 18 machines have values lower than 500, except 4 of them, which have a maximum value of 1114. IDM has the largest maximum W@R value with 15551 among the well covered metrology workshops (BIN, BLI, CDS, MDD, OVL, THA).

## 3.4 Assigning dispatching and sampling policies

As already discussed, how can we know if the queue time values for a given group of metrology tools are reasonable or if some actions should be taken? In this thesis, an approach is proposed that allows evaluating if the sampling and dispatching system of lots in a metrology area should be changed.

### 3.4.1 Algorithm

The approach focuses on helping industries with large risk values or, equivalently, large periods during which machines are processing products without measuring and checking that the production machines are in good condition. Also, the method proposed is solely centered on the queue times (explained in section 3.3.2.4) and the measurement times (especially lower than 5 hours) of metrology tools. The semiconductor manufacturing industry is thus a suitable candidate that corresponds extremely well to these characteristics.

The types of dispatching policies considered here are:

1. **FIFO**: First in, first out. The first lot that arrives in the metrology area and flagged for measurement is the first lot to be measured.

2. **LIFO**: Last in, first out. The last lot that arrives in the metrology area and flagged for measurement is the first lot to be measured.

3. **Priority order**: The lots reaching the metrology area follow an order previously established. This order for instance can be determined by increasing delivery dates or by giving more importance to the lots coming from process machines with the largest levels of risk.

4. **Other**: Other dispatching methods not considered above, for example a random choice of the next lot to measure.

Besides applying a dispatching rule, we consider the possibility of going one step further in terms of better using the metrology capacity and getting lower values of risk by means of sampling techniques [60]. In particular this can be done by using a dynamic approach [76] where skipping policies are implemented in order to manage efficiently metrology queues and to first select lots coming from process machines with larger risks.

In order to establish which of the methods is the most appropriate, as shown in Table 3.9, three criteria are considered:

1. **Physical space required**: Evaluates the method based on the required physical place for the lots waiting to be measured.

2. **Resources required to implement**: Evaluates the method based on the required resources to implement in terms of time, money and complexity.

3. **Efficiency**: Evaluates the method based on its expected results in terms of risk and on how smart are the measures performed (measure the lots that reduce the risk the most).

The methods are ordered by criterion from best (value of 1) to worst (value of 4). At the end, the method with the lowest value is the best one to select.

**Table 3.9 –** *Dispatching methods and sampling strategies evaluated by criterion.*

| **Method** | **Physical place required** | **Resources required to implement** | **Efficiency** | **TOTAL** |
|---|---|---|---|---|
| FIFO | 3 | 1 | 3 | 2.3 |
| LIFO | 3 | 1 | 4 | 2.7 |
| Priority order | 2 | 2 | 2 | 2 |
| Dynamic Sampling | 1 | 3 | 1 | 1.3 |

The algorithm of the method is shown in Figure 3.23.

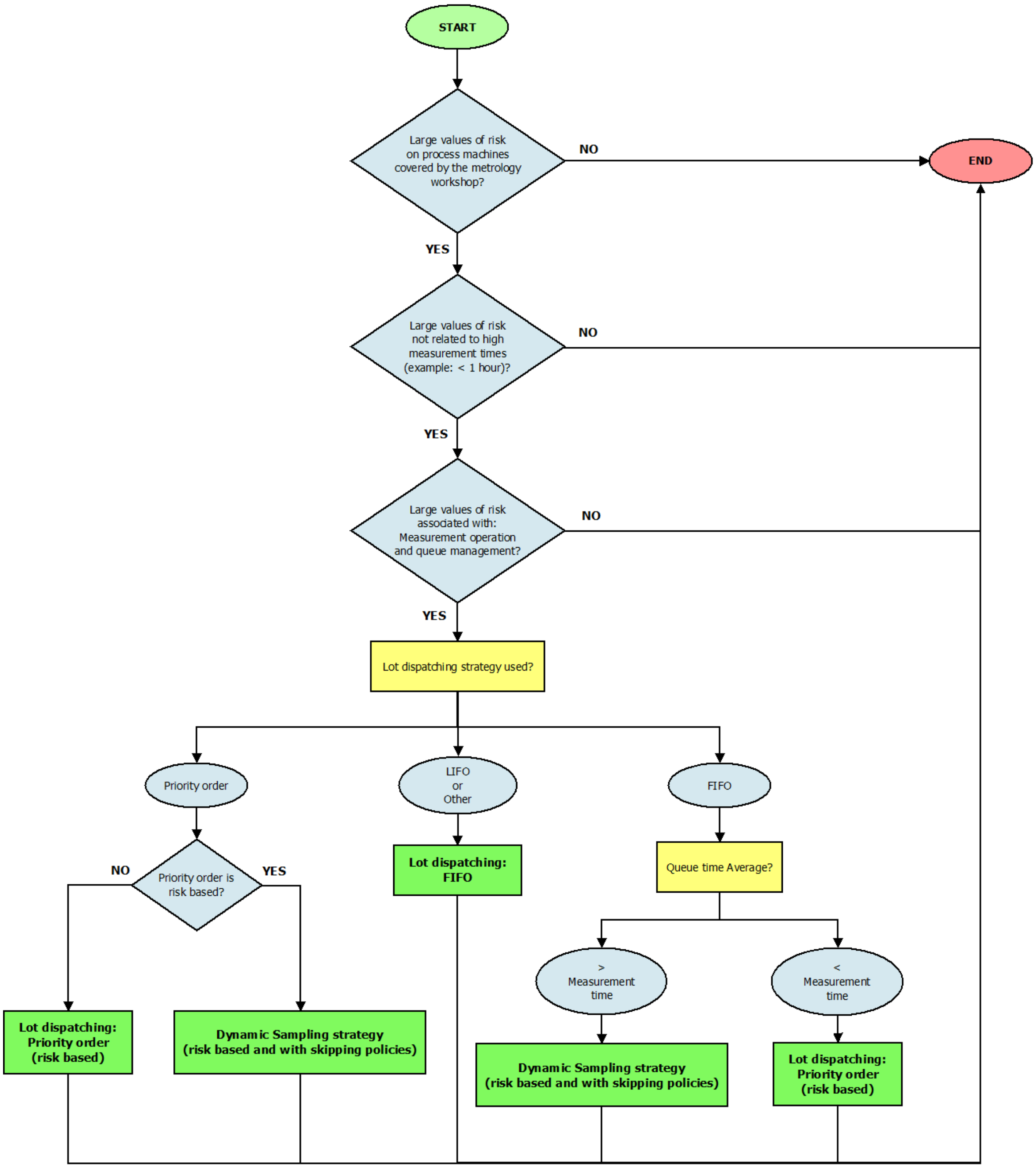



**Figure 3.23 –** *Assignment method of dispatching and sampling policies for Metrology workshops.*

This approach could lead to implementing a new method, different to the one currently used, in order to improve the W@R values obtained in a metrology workshop by minimizing the queue time.

As explained in Section 3.3.2, some metrology workshops will not carry previous useless metrology operations (i.e. that do not reduce the W@R) for the machine controlled after the process operation (as BLI with 0, in Table 3.5 "Times and Sampling policies", field: “Average of Metrology Operations before this measure”), but it could happen for others.

In the approach proposed here, the queue times and consequently the W@R levels will improve entirely for the metrology workshop such that the measurement operations follow the process operations of the controlled machines. For metrology workshops that have operations coming from other metrology systems (as for example BIN with 0.63, in Table 3.5, which indicates that in 63% of the time lots have a different measurement operation before), only the time from the last measurement operation and the measurement operation which reduces the W@R will be improved.

The first case is shown in Figure 3.24 with a W@R chart for a given process machine. When Lot $L_1$ is processed, the W@R increases as much as the number of wafers in $L_1$ during the “Process Time”. $L_1^{Ps}$ represents when the process operation starts for $L_1$ and its ending is represented as $L_1^{Pe}$. Later on, $L_1$ starts the measurement operation ($L_1^{Ms}$) which takes the measurement time ($L_1^{Me}$). At that moment the W@R for the machine will be reduced as much as the number of wafers after $L_1$ has been processed, the final W@R value will be the difference between the W@R reduced after the measurement and the current W@R increased due to the fact that the process machine is processing a new lot ($L_2$).

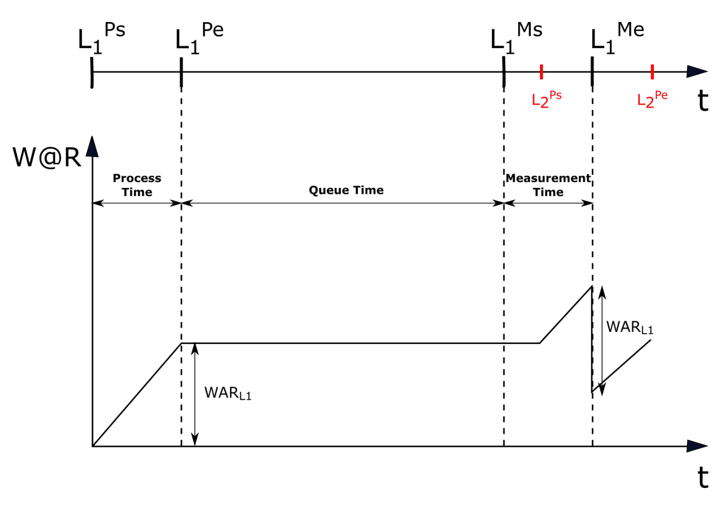



**Figure 3.24 –** *Times for a process machine directly covered by the metrology tool that reduces the W@R.*

The second case, where other measurement operations precede the measure which reduces the W@R levels is illustrated in Figure 3.25. The approach is only able to reduce the value of *Real Queue Time* since the new dispatching rules are recommended for the metrology area which covers the controlled process machine.

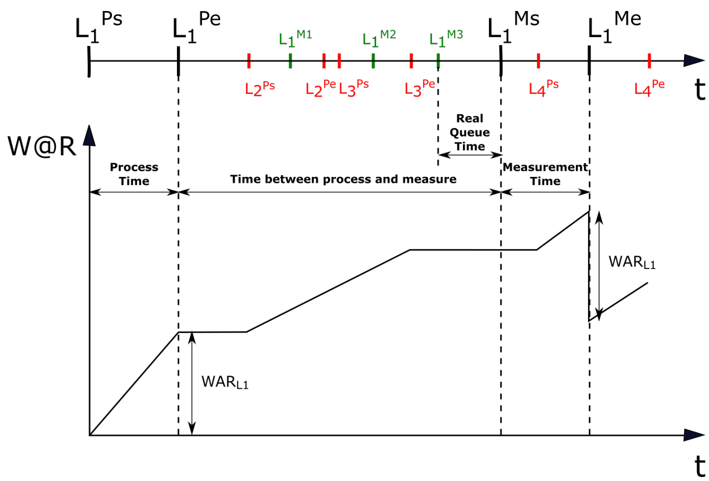



**Figure 3.25 –** *Times for a process machine with some process and metrology operations before arriving to the metrology tool that reduces the W@R.*

After $L_1$ is processed, some other lots ($L_2$, $L_3$ and $L_4$) are also processed by the same process machine increasing thereby the W@R value. $L_1$, before being measured by the metrology tool that directly covers the process machine and which will be able to reduce its W@R, will pass by other metrology tools ($M_1$, $M_2$ and $M_3$) increasing the queue time. In this case, only the *Real Queue Time* will be improved (the portion of the queue time between $L_1^{M3}$ and $L_1^{Ms}$).

### 3.4.2 Application

The "*Queue times*" part, in Section 3.3.2.4, presents the values by metrology workshop showing which of them are the fastest and the slowest. Normally the average measurement times are more stable (around 370 seconds) than the queue times (values between 1673 and 1960 and for one month stretching to 2811 seconds) as it can be seen for the THA metrology workshop in Table 3.10.

**Table 3.10 –** *Measurement times and queue times by month for the THA metrology workshop.*

| Measure type | Measurement times | Queue times | Month |
|---|---|---|---|
| THA | 368.3 | 1960.7 | 1 |
| THA | 372.3 | 2811.8 | 2 |
| THA | 370.1 | 1870.5 | 3 |
| THA | 364 | 1828.2 | 4 |
| THA | 377.3 | 1673.5 | 5 |
| THA | 362.8 | 1925.5 | 6 |

According to the comparison between measurement time and queue time as shown in Figure 3.26, it is quite clear for this metrology workshop that, on average, the time invested for a lot to be measured is normally larger than the measurement time, but other considerations have to be taken into account. In Chapter 5, some simulations are done to compare dispatching policies (FIFO, LIFO, and so on).

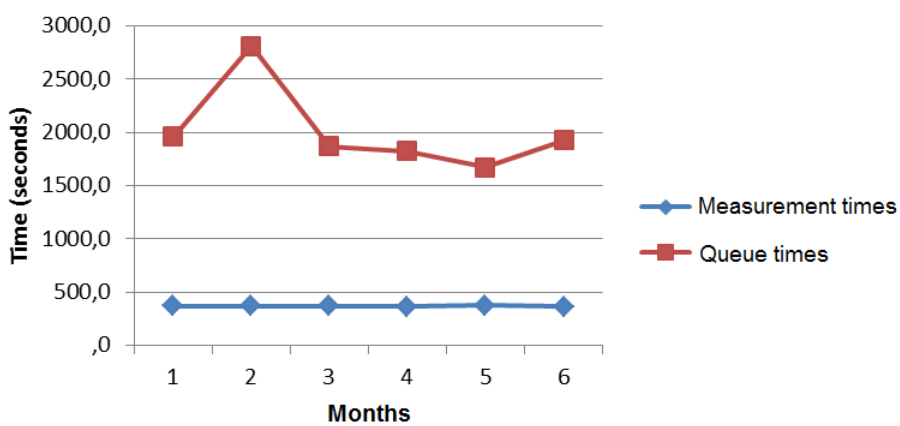



**Figure 3.26 –** *Measurement time vs Queue time.*

## 3.5 Approach for changing Sampling strategy

In order to detect possible points of improvement in a metrology workshop in charge of controlling several process machines, a procedure to verify the state of the current sampling strategies has been developed. The aim is to evaluate how a metrology workshop is currently running, to verify if it is not running as expected and to guide towards achieving the desired behavior. The use of this approach must be frequent to check for the utility of using a given metrology system and to correct the current sampling policy.

The procedure has been divided into four main stages, as shown in Figure 3.27:

1. Identification of the properties of the metrology system.
2. Analysis of the link between the group of process machines and the metrology tools.
3. Verification of the possible gap to improve and selection of the key performance indicators.
4. Identification of the current sampling strategy, analysis of the implementation of a new sampling strategy, and calculation of new sampling ratios.

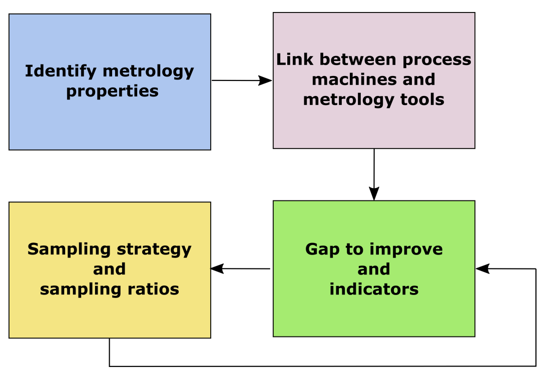



**Figure 3.27 –** *Main stages of the approach for changing Sampling strategy.*

The approach starts by identifying the metrology tool properties, followed by the link between process machines and metrology tools. The third stage is responsible for defining the right indicators to analyze the possible margin of improvement. If the indicators are good, the current sampling is kept. Otherwise, the approach moves to the fourth stage and a new sampling strategy must be implemented. After adopting a new sampling strategy, the indicators are again calculated to ensure a suitable trend of change.

### 3.5.1 Algorithm

1. **Identify metrology properties**

   At the beginning, the set of metrology tools that comprises the measure type should be identified. As a first step, the potential risk covered by the measure should be checked, in other words, if the measure serves to prevent a risk or otherwise if it has become useless. In this case, the measure should be removed. The control limits also must be reviewed or new ones must be created if there are no limits to validate the measure. In Figure 3.28 this stage of the sampling strategy change approach is systematized.

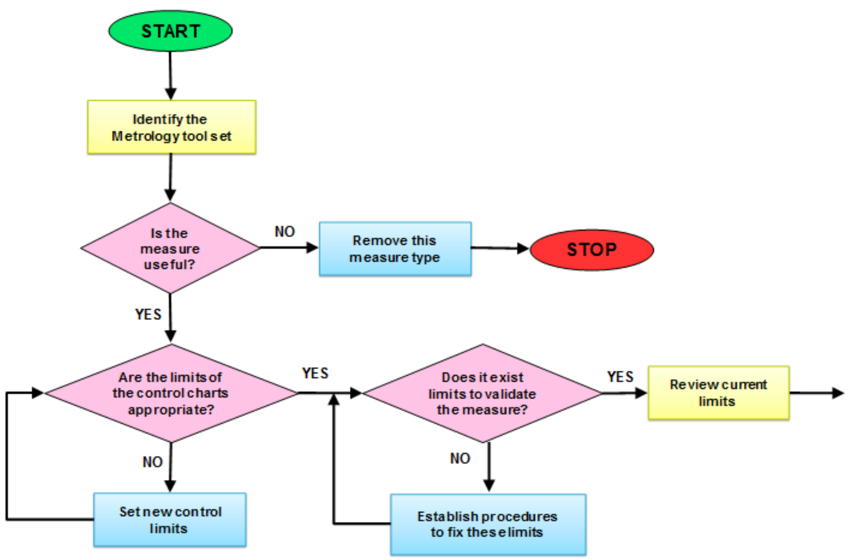



**Figure 3.28 –** *Approach for changing Sampling strategy - Identify metrology properties.*

2. **Link between process machines and metrology tools**

   Once the metrology system is validated, the process machines that will be controlled by the measure must be identified. It is important to know exactly the number of machines to avoid skipping any machine and letting them become uncontrolled. An analysis of the current sampling strategy is necessary, to know exactly how the metrology workshop is running. Then, as a final step, a coverage table and a control plan model need to be created and updated, otherwise they should be modified. The coverage table is only used in case some process machines are covered (their risk decrease) by the same measure. Figure 3.29 shows a diagram of this stage.

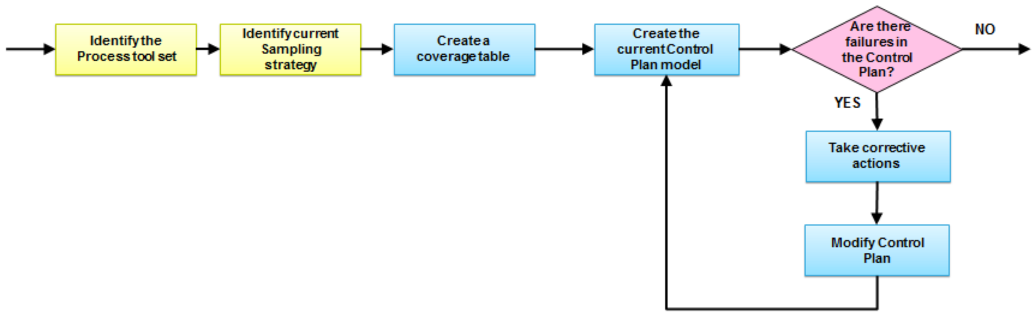



**Figure 3.29 –** *Approach for changing Sampling strategy - Link between process machines and metrology tools.*

3. **Gap to improve and indicators**

   Before performing some changes to the sampling strategy, indicators need to be calculated to verify afterwards whether the process has improved or has worsened. The indicators could be the W@R values by process machines, the OEE (Overall Equipment Efficiency), CT (cycle time), WFY (wafer fab yield) and so on.

   A preliminary study of the indicators should be done, and if the indicators present satisfactory values, the current sampling strategy will be kept, otherwise the gap for improvement should be analyzed.

   If there is no margin to improve in terms of implementing new sampling policies such as modifying the current sampling rate values or changing the sampling strategy, then the current dispatching system of lots is studied in order to propose new dispatching rules; see Section 3.4 where a method for assigning a dispatching and sampling policy is proposed.

   Finally, if a sampling strategy is kept, a sanity check model is created to regularly control that the results are the expected ones, and a W@R by process machine follows up as well. Figure 3.30 presents this stage of the approach for changing the sampling strategy.

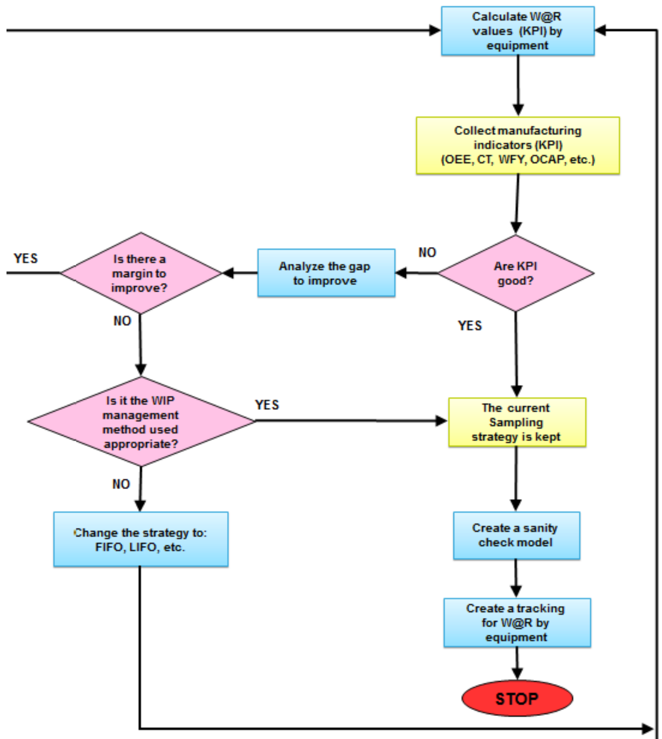



**Figure 3.30 –** *Approach for changing Sampling strategy - Gap to improve and indicators.*

4. **Sampling strategy and sampling ratios**

   In this stage of the sampling strategy change (see Figure 3.31), new sampling ratios must be calculated to assign them to the respective process machines. In Chapter 4, new methods are proposed to optimize the sampling rates of process machines taking into account parameters such as process throughputs, measurement times, probability of failure for process machines and metrology tools.

   As a final step, the new sampling strategy must be activated, and then return to the third stage: "Gap to improve and indicators", to verify if the W@R and key performance indicators are the expected ones.

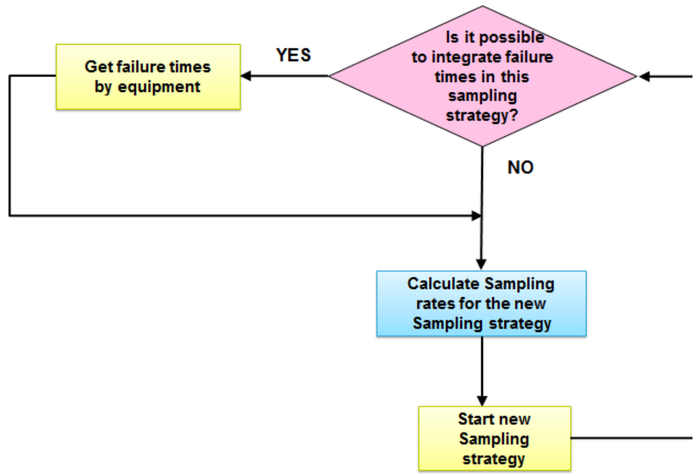



**Figure 3.31 –** *Sampling strategy change approach - Sampling strategy and sampling ratios.*

### 3.5.2 Industrial results

The procedure proposed here has been applied to perform a sampling strategy change for the BLI measure at the Rouset site STMicroelectronics. This measure verifies the surface of the wafer looking for defects at a macro level scale. The sampling policy used for measuring lots was assigning sampling ratios by product and operation, and thus the sampling rates by process machine were not constant as exposed in Section 1.4.

Stages 1 and 2 of the procedure were validated and the indicators were gathered in stage 3 ("*Analyze the gap to improve*"). It was considered that the W@R (that are directly linked to the sampling rates) of process machines were smaller than desired and inconstant with respect to the defects found in the wafers by using this measure. The decision to move to sampling by equipment was taken as shown in Figure 3.32.

With the method used in Chapter 4, the sampling rates have been optimized and assigned for all process machines covered by the BLI metrology workshop. After discussion with the engineers responsible for the set of process machines, it has been decided to apply a sampling ratio of *1/5* for all process machines (W@R value of 125 by machine).

During a period of three months, the change to a sampling by equipment strategy for the process machines associated to the BLI measure has been carried out for some process machines.

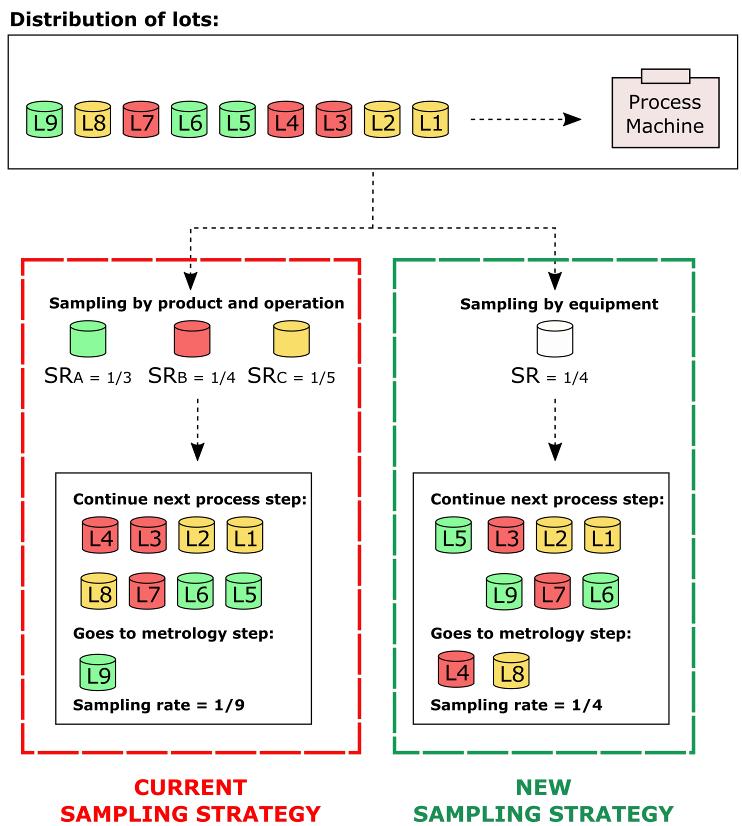
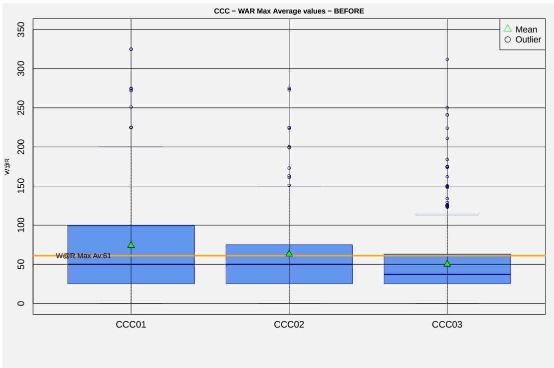


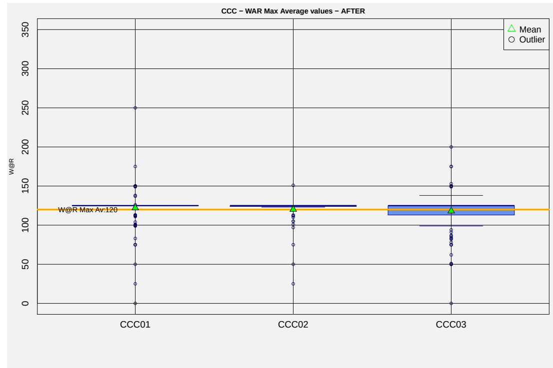

**Figure 3.32 –** *Sampling change from sampling by product and operation to sampling by equipment.*

Figure 3.33 provides the evolution of the W@R values with boxplot graphics for three process machines. The data for the *Before* and *After* periods consist in a production month, with the same number of days. The change to a sampling by equipment strategy provided a constant sampling rate that ensures the expected control of the machine. The process group *CCC* had an average maximum W@R of 61 and every machine had different means with a variability of W@R values. However, after the sampling change, the average for all machines was 120 (almost 125), and the means are in line with the average as well and the variability has significantly been reduced.

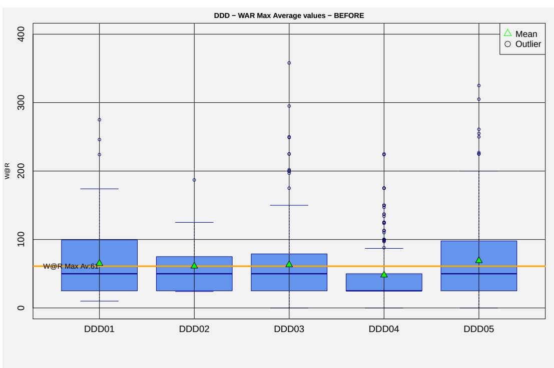
The process groups *DDD* and *EEE* changed their average maximum W@R from 61 and 67 respectively to 117, both of them keeping the mean of all their machines in the average, which means that all process machines are well balanced. There has been a reduction of the maximum W@R values, especially for the *EEE* machines. The values achieved before the sampling strategy change were 325, 350 and 825 for *CCC*, *DDD*, and *EEE* respectively and after the change a value of 250 is achieved for the three process groups. In terms of outliers, the boxplot chart shows how the number of outliers has been reduced probing that the sampling is now better controlled.



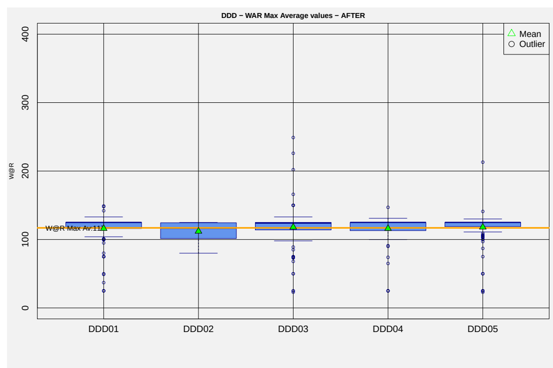

**(a)** *CCC process machines - Before*



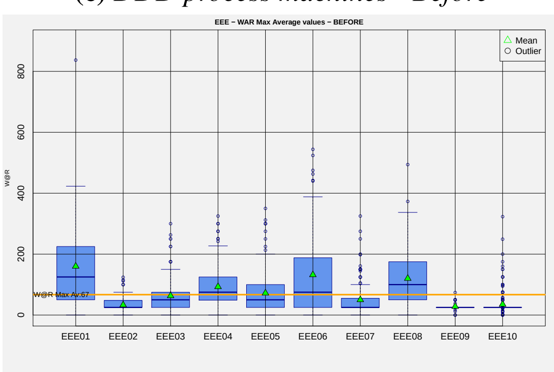

**(b)** *CCC process machines - After*



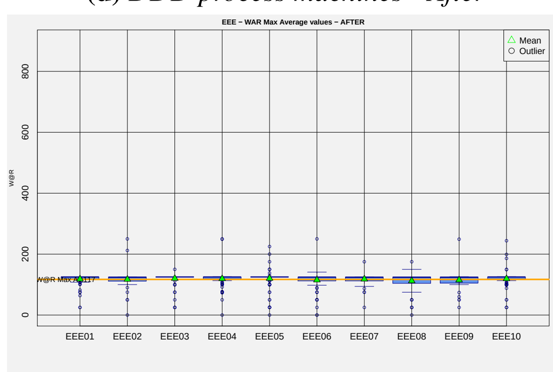

**(c)** *DDD process machines - Before*




**(d)** *DDD process machines - After*




**(e)** *EEE process machines - Before*




**(f)** *EEE process machines - After*

**Figure 3.33 –** *Average maximum W@R boxplots with results for BLI measure.*

Apart from the **W@R** values, two other key performance indicators helped to check if there was an improvement with the sampling change: **Risk variability** and **ΔCT** (cycle time difference).

The global **risk variability** for the BLI metrology type has been reduced. Even if, not all process machines moved to a sampling by equipment strategy, this first change helped to reduce the standard deviation, $\sigma$, of the W@R values from 91 to 59. The impact of the sampling strategy change on the global variability is shown in Figure 3.34.

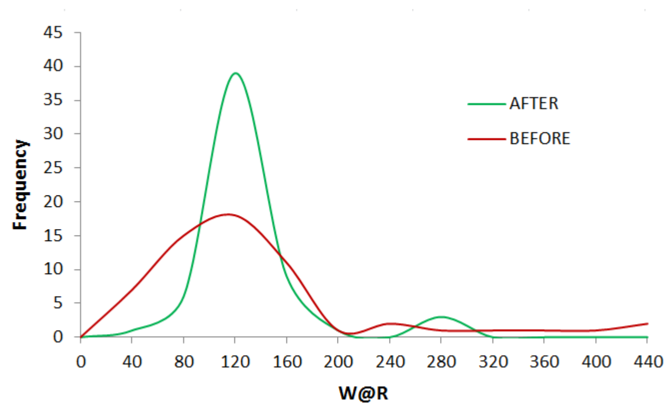



**Figure 3.34 –** *Risk variability reduction after implementing the sampling by equipment.*

The equation (3.7) is used to measure the cycle time gain, **ΔCT**. $L_m$ is the number of measured lots, $\overline{MT}$ is the average measurement time in seconds.

$$\Delta\mathrm{CT} = L_m^{\mathrm{before}} \times \overline{MT^{\mathrm{before}}} - L_m^{\mathrm{after}} \times \overline{MT^{\mathrm{after}}} \tag{3.7}$$

The number of measures has been reduced from $L_m^{\mathrm{before}} = 9203$ to $L_m^{\mathrm{after}} = 7666$ with an average measurement time of $\overline{MT^{\mathrm{before}}} = 256$ and $\overline{MT^{\mathrm{after}}} = 250$. The cycle time gain is $\Delta\mathrm{CT} = 437537$ seconds.

## 3.6 Conclusion

In this chapter the importance of metrology in semiconductor manufacturing and its impact on the final product has been discussed. Also the most important types of metrology systems used nowadays in semiconductor manufacturing have been reviewed.

A deep analysis of the metrology workshops of the Rousset site of STMicroelectronics has been carried out, taking into account several parts such as: Physical properties verified, number of process machines controlled, difficulty of performing measurement qualification, differences between metrology tools, lot dispatching strategy, location of workshop, number of measured wafers, sampling policies, measurement and queue times, sampling rates, number of measurements done and W@R values.

The proposed approach to build an entire classification of metrology workshops is composed of four main tables of characteristics. This classification provides a way to understand how a metrology workshop is operated through key parameters which will help to decide how to improve the performance of the workshop.

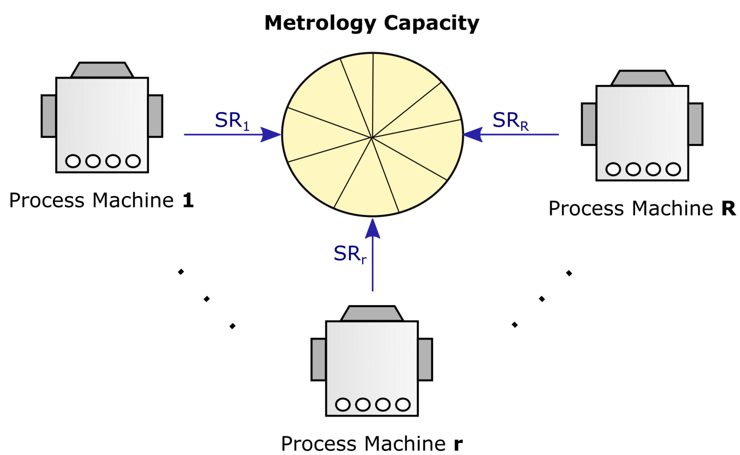
A new methodology has been proposed that enables to reduce high risk values by reducing the queue times depending on the current dispatching strategies in the metrology workshop.

A novel procedure to perform sampling strategy changes has been developed to regularly verify the correct use of the sampling policies for metrology workshops. In a first stage, the utility of the measure and the current measure limits are verified, and in a second stage, the number of associated process machines and their links are checked. The possibilities of designing a coverage table or not and an updated control plan are considered. In a third stage, the selection of the right indicators is done and then the easiness to move to a new sampling strategy is evaluated. In the final stage, the calculation of the sampling rates and the implementation of the new sampling strategy are performed. The results of the sampling strategy change implemented following the procedure at the Rousset site of STMicroelectronics has been explained, showing that the expected W@R values were achieved (decreasing the maximum values and getting the expected average). This led to a better control of the process machines and a reduction of the risk variability. Also, by reducing the number of measures of the measure type which needed less measures, a cycle time gain was also obtained improving the yield of lots.

# Chapter 4

# Static Sampling rate optimization

*In this chapter, novel approaches to optimally assign a sampling rate, respecting the metrology capacity, to each process machine in a given workshop is presented. Several mathematical models have been developed depending on the nature of the metrology tools. Numerical experiments have been carried out to validate the approaches. Some industrial results and the industrial implementation are presented at the end of the chapter.*



## 4.1 Introduction

Nowadays, semiconductor manufacturing has become a complex industry with hundreds of process operations are involved. To ensure the quality of products, numerous in-line metrology operations to control the previous production step are needed. Metrology helps to manage possible risks, in particular failures of process machines.

As explained in Chapter 3, depending on the metrology type, different properties of the wafer are controlled, such as possible defects on its surface, its mechanical properties, the thickness of the deposited layers, the implant dose, and so on. Apart from their benefits, metrology operations also affect other aspects of the production flow. The WIP is increased, extra costs are required, cycle time is increased and scheduling decisions are disrupted.

Significant research related to the use of metrology in semiconductor manufacturing has been carried out. Chien et al. [16] propose a cost-based heuristic that considers the trade-off between the related costs (inspection costs, false-alarm costs, out-of-control costs) and the corresponding probabilities by using different sampling rates. Colledani [19] shows that allocating a large number of metrology stations does not always provide higher conforming part production rate. Gilenson et al. [26] present a trade-off between cycle time and yield in a semiconductor production line, based on the impact of a constant inspection operation rate on both performance indicators.

In this chapter, a novel approach to optimize the sampling rates for a group of process machines covered by a metrology workshop is proposed. To determine the sampling rates, a limited metrology capacity is considered and failure rates for metrology tools and process machines are taken into account. This approach is focused on the types of metrology workshops in which sampled lots are measured right after the process operation in order to verify the good conditions of process machines.

This chapter is structured as follow: Section 4.3 presents mathematical models for these cases: with a unique metrology tool, identical metrology tools (same measurement throughput) and different metrology tools (different measurement throughputs) to perform the measures. Numerical experiments to validate the mathematical models are presented and discussed in Section 4.4. Section 4.5 shows some results on industrial data and Section 4.6 presents the industrial tool developed for STMicroelectronics. Finally, concluding remarks and perspectives are presented in section 4.7.

## 4.2 Mathematical models

Several metrology tools $t = \{1, \ldots, T\}$ measure the products of several process machines $r = \{1, \ldots, R\}$. The process machines are modeled as Bernoulli experiments and are differentiated by their probability of failure $\boldsymbol{p_r}$. The throughput rates are denoted by $\boldsymbol{TP_r}$ for the process machines and by $\boldsymbol{TM_r^t}$ for the metrology tool $t$ that measures a wafer processed on machine $r$. The sampling period, number of production cycles on a process machine $r$ between two consecutive measures, is referred as $\boldsymbol{SP_r}$. We assume that the production of a defective machine is fully scrapped and there is no difference between the value of wafers on the different process machines.

### 4.2.1 Unique metrology tool (Model (PI1))

This first model is developed considering a unique metrology tool $t$, with a measurement rate $\boldsymbol{TM_r}$, in charge of covering a group of process machines by measuring the wafers processed on them. The decision variables are the sampling periods $SP_r$, $\forall r = \{1, \ldots, R\}$. Setting $SP_r$, the throughput of defective wafers and the consumption of the metrology tool capacity, $g_r(SP_r)$, are determined.

$\boldsymbol{SP^{max}}$ is the maximum sampling rate value over which the quality control is unacceptable. If a failure occurs in a first production cycle, all the following $SP_r$ wafers will be scrapped, if it occurs in a second production cycle $SP_r - 1$ wafers will be scrapped, and so on. We assume that the production of a machine in good condition is perfect. By adding $\frac{TP_r}{SP_r}$, the rate at which the measurement is performed at the station, $\boldsymbol{WL_r(SP_r)}$ is the overall expected rate of defective wafers produced by a machine $r$ covered by a sampling period $SP_r$, called **Wafer Loss**, and is determined by (4.1):

$$WL_r(SP_r) = \frac{p_r TP_r}{SP_r} \sum_{i=0}^{SP_r - 1} (SP_r - i)(1 - p_r)^i \tag{4.1}$$

The portion of metrology capacity consumed by the process machine $r$ for a given sampling period $SP_r$, $g_r$, can be written:

$$g_r(SP_r) = \frac{TP_r}{SP_r TM_r} \tag{4.2}$$

The optimization problem (P) can be formulated as:

$$\min \sum_{r=1}^{R} WL_r(SP_r) \tag{4.3}$$

$$s.t.$$

$$\sum_{r=1}^{R} g_r(SP_r) \leq 1, \tag{4.4}$$

$$SP_r = \{1, \ldots, SP^{max}\}, \quad r = \{1, \ldots, R\}. \tag{4.5}$$

By linearising Constraints (4.3) and (4.4) in which the decision variables $SP_r$ are in the denominator, the problem (P) can be rewritten as an integer linear program (ILP). Let us define $s$ as the index for the sampling period, i.e, $s = \{1, \ldots, SP^{max}\}$ and the binary variable $u_r^s$ which is equal to 1 if the sampling period of process machine $r$ is $s$, i.e. $SP_r = s$ and 0 otherwise. Hence, (P) can be reformulated as the ILP below, denoted by (PI1):

$$\min \sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} WL_r(s) u_r^s \tag{4.6}$$

$$s.t.$$

$$\sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} g_r(s) u_r^s \leq 1, \tag{4.7}$$

$$\sum_{s=1}^{SP^{max}} u_r^s = 1, \quad r = \{1, \ldots, R\}, \tag{4.8}$$

$$u_r^s \in \{0, 1\}, \quad r = \{1, \ldots, R\}; s = \{1, \ldots, SP^{max}\}. \tag{4.9}$$

This problem is a Multiple-Choice Knapsack Problem (MCKP) as presented by Sinha and Zoltners [85], this problem has been studied along the years [115] [66], and it is basically about objects that are arranged in classes, and one an only object must be selected per class. In our problem, sampling periods are objects and process machines are classes. Our goal is to develop fast and simple heuristics, easy to implement, that are competitive in terms of solution quality with an exact approach, whether it is the one proposed by Pisinger [66] or the resolution of (PI1) with a standard solver.

The continuous version of the MCKP, usually called linear multiple-choice knapsack problem (LMCKP) in the literature, corresponds to (PI) in which the integrity Constraints (4.9) are replaced by:

$$u_r^s \in [0, 1], \quad r = \{1, \ldots, R\}, \quad s = \{1, \ldots, SP^{max}\}. \tag{4.10}$$

### 4.2.2 Identical metrology tools (Model (PI2))

The case where multiple identical metrology tools $t = \{1, \dots, T\}$ control a group of process machines $r = \{1, \dots, R\}$ is considered. The throughput rate of the metrology tools when measuring lots processed on machine $r$, $\boldsymbol{TM_r}$, will be the same for all metrology tools. One and only one metrology tool is assigned to control the totality of the production of each machine, thus $T \leq R$. We assume that the production of a machine in good condition is perfect. The assignment of process machine $r$ to metrology tool $t$ is modeled using a binary variable $v_r^t, \forall r = \{1, \dots, R\}, \forall t = \{1, \dots, T\}$.

Following the same steps than Section 4.2.1, the problem (P) related to these conditions can be rewritten as an Integer Linear Program (ILP) since, for each process machine $r$, $SR_r$ must be chosen in the set of all possible sampling periods $\{1, \dots, SP^{max}\}$. However, for this problem, it is also necessary to assign each process machine to one only metrology tool. Let us define the binary variable $w_r^{s,t} = 1$ if metrology tool $t$ is assigned to control the lots processed on machine $r$ with a sampling rate $s$, and 0 otherwise. The ILP denoted by (PI2) can be stated as:

$$\min \sum_{t=1}^{T} \sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} WL_r(s) w_r^{s,t} \tag{4.11}$$

$$s.t.$$

$$\sum_{t=1}^{T} \sum_{s=1}^{SP^{max}} w_r^{s,t} = 1, \quad \forall r = \{1, \dots, R\}, \tag{4.12}$$

$$\sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} g_r(s) w_r^{s,t} \leq 1, \quad \forall t = \{1, \dots, T\}, \tag{4.13}$$

$$w_r^{s,t} \in \{0, 1\}, \quad \forall r = \{1, \dots, R\}; t = \{1, \dots, T\}; s = \{1, \dots, SP^{max}\}. \tag{4.14}$$

Constraint(4.12) ensures that one and only one sampling rate $s$ and one and only one metrology tool $t$ are selected for each process machine $r$. The capacity of each metrology tool is satisfied and it will not be exceeded by means of Constraint(4.13).

### 4.2.3 Different metrology tools (Model (PI3))

In this case, we assume that different metrology tools $t = \{1, \dots, T\}$ that control a group of process machines $r = \{1, \dots, R\}$ is considered. The throughput rate of a metrology tool $t$ when measuring lots processed on machine $r$ is denoted by $\boldsymbol{TM_r^t}$. The measure operation is considered as imperfect and returns with probability $\boldsymbol{\alpha_r^t}$ a false negative (the measurement result is good while $r$ does not work properly).

The assignment of process machine $r$ to metrology tool $t$ is modeled using a binary variable $v_r^t, \forall r = \{1, \dots, R\}, \forall t = \{1, \dots, T\}$. Wafers are reworked or scrapped at a certain rate $WL_r^t(SP^{max})$ for process machines $r$ with a sampling rate $SP_r$ covered by metrology tools $t$.

The portion of capacity consumed on the metrology tool $t$ by covering a process machine $r$ with a sampling period $SP_r$, can be written:

$$g_r^t(SP_r) = \frac{TP_r^t}{SP_r TM_r^t} \tag{4.15}$$

The optimization problem (P) can be formulated as:

$$\min \sum_{t=1}^{T} \sum_{r=1}^{R} WL_r^t(SP_r) v_r^t \tag{4.16}$$
$$s.t.$$

$$\sum_{r=1}^{R} g_r^t(SP_r) v_r^t \leq 1, \quad \forall t = \{1, \ldots, T\}, \tag{4.17}$$

$$\sum_{t=1}^{T} v_r^t = 1, \quad \forall r = \{1, \ldots, R\}, \tag{4.18}$$

$$SP_r \in 1, \ldots, SP^{max}, \quad \forall r = \{1, \ldots, R\}, \tag{4.19}$$

$$v_r^t \in 0, 1, \quad \forall r = \{1, \ldots, R\}; t = \{1, \ldots, T\}. \tag{4.20}$$

The Wafer Loss for a process machine $r$ can be reformulated. Let $j$ denote the number of measurements on a metrology tool $t$ of lots processed on $r$ until a measure operation is performed. The value of $j$ can be $j = 1$ with a probability of $(1 - \alpha_r^t)$, $j = 2$ with a probability of $\alpha_r^t(1 - \alpha_r^t)$, and so on. A sampling cycle on a process machine $r$ is a series of $jSP_r^t$ Bernoulli experiments, each of which corresponds to the production of a lot. Therefore, if a failure occurs in the first production cycle, all the following $jSP_r^t$ lots (the number of lots produced until the next reliable measure takes place) are scrapped. Similarly, if a failure occurs in the second production cycle, $jSP_r^t - 1$ lots will be scrapped, and so on. In the case of a failure happens in the last production cycle before the next reliable measure, only one lot will be scrapped. The expected number of defective lots from $r$ between two reliable measures performed on $t$ is denoted by $WLC_r^t$ (Wafer Loss Count) and can be written:

$$\begin{aligned} WLC_r^t(SP_r^t, j) &= jSP_r^t p_r + (jSP_r^t - 1)(1 - p_r)p_r + \cdots + 1(1 - p_r)^{jSP_r^t - 1} p_r = \\ &= p_r \sum_{i=0}^{jSP_r^t - 1} (jSP_r^t - i)(1 - p_r)^i \end{aligned} \tag{4.21}$$

Wafer Loss Count, when considering the probability of an inspection to be reliable, allows to compute the total expected Wafer Loss for a process machine **r**:

$$
\begin{aligned}
WL_r^t(SP_r) &= (1-\alpha_r^t)WLC_r^t(SP_r,1) + \alpha_r^t(1-\alpha_r^t)WLC_r^t(SP_r,2) + (\alpha_r^t)^2(1-\alpha_r^t)WLC_r^t(SP_r,3)+ \\
&+ \cdots + (\alpha_r^t)^{j-1}(1-\alpha_r^t)WLC_r^t(SP_r,j) + \cdots = \\
&= (1-\alpha_r^t)\sum_{j=1}^{\infty}(\alpha_r^t)^{j-1}WLC_r^t(SP_r,j) = \\
&= (1-\alpha_r^t)p_r\sum_{j=1}^{\infty}\left[(\alpha_r^t)^{j-1}\sum_{i=0}^{jSP_r-1}(jSP_r-i)(1-p_r)^i\right]
\end{aligned}
\tag{4.22}
$$

As done in previous sections, (P) can be rewritten as an Integer Linear Program (ILP), since for each process machine $r$, $SP_r$ must be chosen in the set of all possible sampling periods $\{1,\ldots,SP^{max}\}$. Let define the binary variable $w_r^{s,t} \in \{0,1\}$. $w_r^{s,t} = 1$ if process machine $r$ is assigned to metrology tool $t$ with a sampling rate of $s$, and $w_r^{s,t} = 0$ otherwise. Hence, (P) can be reformulated as the ILP below, denoted by (PI3):

$$
\min \sum_{t=1}^{T}\sum_{r=1}^{R}\sum_{s=1}^{SP^{max}} WL_r^t(s)w_r^{s,t} \tag{4.23}
$$

$s.t.$

$$
\sum_{r=1}^{R}\sum_{s=1}^{SP^{max}} g_r^t(s)w_r^{s,t} \leq 1, \quad \forall t = \{1,\ldots,T\}, \tag{4.24}
$$

$$
\sum_{t=1}^{T}\sum_{s=1}^{SP^{max}} w_r^{s,t} = 1, \quad \forall r = \{1,\ldots,R\}, \tag{4.25}
$$

$$
w_r^{s,t} \in 0,1, \quad \forall r = \{1,\ldots,R\}; t = \{1,\ldots,T\}; s = \{1,\ldots,SP^{max}\}. \tag{4.26}
$$

Constraint (4.24) ensures that the metrology capacity is satisfied and Constraints (4.25) and (4.26) that one and only one sampling period is selected for each process machine.

Defined as such, the problem is a Multilevel Generalized Assignment Problem (see Ceselli and Righini [12] or Park et al. [63]). A special instance of this problem occurs when the metrology tools are all identical, i.e. when both $\alpha_r^t$ and $TM_r^t$ are independent of $t$. In this case, neither the objective function nor the constraint are dependent of the metrology tool, and the problem is in fact a Multiple Choice Multiple Knapsack Problem (MCMKP).

If, in addition, we replace Constraint (4.26) by $w_r^{s,t} \in [0,1]$, the problem reduces to a Multiple Choice Knapsack Problem with capacity $T$.

## 4.3 Problem resolution

Different heuristics are proposed to solve the problems.

### 4.3.1 Unique metrology tool

There is at least one optimal solution of the LMCKP such that all classes (in our case, all process machines) except one have exactly one selected object (in our case, one sampling period) [85], i.e. no fractional variable, and the remaining class has at most two partially selected objects, i.e. two fractional variables. Moreover, these two partially selected objects are cost-wise adjacent choices which correspond, in our case, to two adjacent variables $u_r^s$ and $u_r^{s+1}$. This property can be used to derive $O(n)$ (in our case, $n = R.SP^{max}$) [24] [116], Pisinger also proposes a partitioning algorithm [66] and presents a simple greedy algorithm that solves the problem in $O(nlogn)$.

In **Heuristic UMT-**$H_1$, $u_r^s$ is set to 0 and $u_r^{s+1}$ is set to 1 to ensure that the metrology capacity constraint (4.7) is satisfied. The time complexity of this heuristic is thus the same as the time complexity of the algorithm to solve the LMCKP, i.e. $O(R.SP^{max})$ in the best case.

**Algorithm 4.1** Heuristic UMT-$H_1$

1: Solve the LMCKP.
2: **if** $\exists r \in 1, \ldots, R$ and $s \in 1, \ldots, SP^{max} - 1$ such that $u_r^s > 0$ and $u_r^{s+1} > 0$ **then**
3: $u_r^s = 0$ and $u_r^{s+1} = 1$.
4: **end if**

**Heuristic UMT-**$H_2$ is a greedy heuristic that extends the previous heuristic by iteratively rounding the two fractional variables $u_r^s$ and $u_r^{s+1}$ to 0 and 1 respectively, and then fixing these variables until the solution of the LMCKP is no longer continuous. The time complexity of Heuristic 2 is thus R times the time complexity of the algorithm to solve the LMCKP, i.e. $O(R^2.SP^{max})$ in the best case.

**Algorithm 4.2** Heuristic UMT-$H_2$

1: Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
2: Solve the LMCKP.
3: **while** $\exists r \in 1, \ldots, R$ and $s \in 1, \ldots, SP^{max} - 1$ such that $u_r^s > 0$ and $u_r^{s+1} > 0$ **do**
4: $\mathcal{F} \leftarrow \mathcal{F} \cup \{(r, s+1)\}$.
5: Solve the LMCKP while initially fixing $u_{rr}^{ss} = 1, \forall (rr, ss) \in \mathcal{F}$.
6: **end while**

**Heuristic UMT-**$H_3$ differs from Heuristic UMT-$H_2$ in the rounding phase in which the largest of the two variables $u_r^s$ and $u_r^{s+1}$ is set to 1 and the other one to 0. However, to ensure that the metrology capacity constraint (4.7) is satisfied, the decision of fixing $u_r^s$ to

1 is reversed when it leads to an infeasible LMCKP, i.e. $u_r^s$ is replaced by $u_r^{s+1}$. The time complexity of this heuristic is similar than the previous one, $O(R^2.SP^{max})$ in the best case.

**Algorithm 4.3** Heuristic UMT-$H_3$

Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Solve the LMCKP.
**while** $\exists r \in 1, \dots, R$ and $s \in 1, \dots, SP^{max} - 1$ such that $u_r^s > 0$ and $u_r^{s+1} > 0$ **do**
    **if** $u_r^{s+1} > u_r^s$ or $|\mathcal{F}| = R - 1$ **then**
        $\mathcal{F} \leftarrow \mathcal{F} \cup \{(r, s+1)\}$.
    **else**
        $\mathcal{F} \leftarrow \mathcal{F} \cup \{(r, s)\}$.
    **end if**
    Solve the LMCKP while initially fixing $u_{rr}^{ss} = 1, \forall (rr, ss) \in \mathcal{F}$.
    **if** LMCKP is infeasible **then**
        $\mathcal{F} \leftarrow \mathcal{F} \cup \{(r, s+1)\} - \{(r, s)\}$.
        Solve the LMCKP while initially fixing $u_{rr}^{ss} = 1, \forall (rr, ss) \in \mathcal{F}$.
    **end if**
**end while**

### 4.3.2 Identical metrology tools

The first heuristic, called **Heuristic IMT-$H_1$**, combines all metrology tools into one, whose capacity is equal to $T$, i.e. Constraint (4.13) in the ILP is replaced by:

$$\sum_{t=1}^{T} \sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} g_r(s) w_r^{s,t} \leq T. \tag{4.27}$$

The resulting problem is a MCKP. The Heuristic IMT-$H_1$ obtains the sampling rates, $SP_r$ by solving the MCKP with the Heuristic UMT-$H_1$ (4.1) for a unique metrology tool, then a greedy heuristic is applied to build a feasible solution. At each iteration, the process machine $r$ not yet assigned to a metrology tool which consumes the largest metrology capacity is assigned to the metrology tool $t$ with the largest remaining capacity. If the capacity of $t$ is exceeded, then $SP_r$ is increased until either the capacity of $t$ is enough or $SP_r = SP^{max}$. In the latter case, it means that the solution is not feasible.

Its time complexity is at most $O(\max(R.SP^{max}log(R.SP^{max}), R.T.SP^{max}))$, which probably can be reduced to $O(R.SP^{max}log(R.SP^{max}))$.

**Algorithm 4.4** IMT-$H_1$

Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Determine $\mathcal{F}$ by solving the MCKP for the problem where (4.13) is replaced by (4.27) with the rounding Heuristic UMT-$H_1$ of Algorithm (4.1).
Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool). $\mathcal{G} \leftarrow \emptyset$.
Let $Capa_t$ be the capacity used on metrology tool $t$. $Capa_t = 0, t = \{1, \ldots, T\}$.
**while** $\mathcal{F} \neq \emptyset$ **do**
    Find the process machine $r$, such that $\exists (r, s) \in \mathcal{F}$, with the largest metrology capacity $g_r(s)$.
    Find the metrology tool $t$ with the largest remaining capacity, i.e. with the smallest $Capa_t$.
    **if** $Capa_t + g_r(s) \leq 1$ **then**
        Assign $r$ to $t$ with sampling rate $s$, i.e. $\mathcal{G} \leftarrow \mathcal{G} \cup \{(r, s, t)\}$ and $Capa_t = Capa_t + g_r(s)$.
    **else**
        Increase $s$ until either $Capa_t + g_r(s) \leq 1$ or $s = SP^{max}$.
        **if** $Capa_t + g_r(s) \leq 1$ **then**
            Assign $r$ to $t$ with sampling rate $s$, i.e. $\mathcal{G} \leftarrow \mathcal{G} \cup \{(r, s, t)\}$ and $Capa_t = Capa_t + g_r(s)$.
        **else**
            Assign $r$ to $t$ with sampling rate $SP^{max}$, i.e. $\mathcal{G} \leftarrow \mathcal{G} \cup \{(r, SP^{max}, t)\}$ and $Capa_t = 1$ and the problem is infeasible.
        **end if**
    **end if**$\mathcal{F} \leftarrow \mathcal{F} - \{(r, s)\}$
**end while**

Another heuristic is proposed, **Heuristic IMT-$H_1^+$** which is IMT-$H_1$ combined with an improving phase that is described in Algorithm 4.5. A Multi-Choice Knapsack Problem is solved for each metrology tool with the process machines assigned that were previously determined by using the Heuristic IMT-$H_1$. This is done with Heuristic UMT-$H_2/H_3$ by combining Heuristic UMT-$H_2$ (4.2) and Heuristic UMT-$H_3$ (4.3). The new solution for each metrology tool is only kept if it improves the current solution. Heuristic IMT-$H_1^+$ does not increase the complexity of Heuristic IMT-$H_1$, its time complexity is at most $O(R.SP^{max})$.

**Algorithm 4.5** Heuristic IMT-$H_1^+$ (improving phase)

1: Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool) determined by Heuristic IMT-$H_1$ of Algorithm (4.4), i.e. $\exists s \in 1, \ldots, SP^{max}$ and $\exists t \in 1, \ldots, T$ such that $(r, s, t) \in \mathcal{G}, \forall r \in 1, \ldots, R$.
2: **for** $t \in 1, \ldots, T$ **do**
3: $\quad \mathcal{G}' \leftarrow \mathcal{G}$.
4: $\quad$ Solve a Multi-Choice Knapsack Problem with Heuristic UMT-$H_2/H_3$ of Algorithms (4.2) and (4.3) for the process machines $r$ assigned to metrology tool $t$, i.e. $\exists s \in 1, \ldots, SP^{max}$ such that $(r, s, t) \in \mathcal{G}$.
5: $\quad$ Update $\mathcal{G}'$ with the new sampling rates for metrology tool $t$.
6: $\quad$ **if** $\displaystyle\sum_{r=1}^{R} \sum_{s=1;(r,s,t)\in\mathcal{G}'}^{SP^{max}} WL_r^t(s) < \sum_{r=1}^{R} \sum_{s=1;(r,s,t)\in\mathcal{G}}^{SP^{max}} WL_r^t(s)$ **then**
7: $\quad\quad \mathcal{G} \leftarrow \mathcal{G}'$.
8: $\quad$ **end if**
9: **end for**

### 4.3.3 Different metrology tools

In this section, we propose a Lagrangian relaxation heuristic which uses seven heuristics in the step where a feasible solution is constructed from the usually infeasible solution of the relaxed problem.

#### 4.3.3.1 General scheme

The Lagrangian dual problem (LDP) is defined as follows, where $\lambda_t$ are the Lagrangian multipliers associated to relaxing Constraints (4.24) (capacity constraints) in (ILP):

$$\max_{\lambda_t \geq 0;\ t=1,\ldots,T} \min \sum_{t=1}^{T} \sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} (WL_r^t(s) + g_r^t(s)\lambda_t) w_r^{s,t} - \sum_{t=1}^{T} \lambda_t \tag{4.28}$$

$s.t.$

$$\sum_{t=1}^{T} \sum_{s=1}^{SP^{max}} w_r^{s,t} = 1, \quad r = \{1, \ldots, R\}, \tag{4.29}$$

$$w_r^{s,t} \in \{0, 1\}, \quad r = \{1, \ldots, R\}; t = \{1, \ldots, T\}; s = \{1, \ldots, SP^{max}\}. \tag{4.30}$$

For given Lagrangian multipliers $\lambda_t$, the Lagrangian relaxed problem (LRP($\lambda_t$)) is:

$$\min \sum_{t=1}^{T} \sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} (WL_r^t(s) + g_r^t(s)\lambda_t) w_r^{s,t} - \sum_{t=1}^{T} \lambda_t \tag{4.31}$$

$s.t.$

$$\sum_{t=1}^{T} \sum_{s=1}^{SP^{max}} w_r^{s,t} = 1, \quad r = \{1, \ldots, R\}, \tag{4.32}$$

$$w_r^{s,t} \in \{0, 1\}, \quad r = \{1, \ldots, R\}; t = \{1, \ldots, T\}; s = \{1, \ldots, SP^{max}\}. \tag{4.33}$$

The relaxed problem is relatively easy to solve, since it is possible to independently solve a sub-problem for each process tool $r$ by selecting the metrology tool $t^*$, $t^* = 1, \ldots, T$, and the sampling rate $s^*$, $s^* = 1, \ldots, SP^{max}$, such that $(WL_r^{t^*}(s^*) + g_r^{t^*}(s^*)\lambda_{t^*})$ is the smallest, i.e. $(WL_r^{t^*}(s^*) + g_r^{t^*}(s^*)\lambda_{t^*}) = \min_{t=1,\ldots,T;\ s=1,\ldots,SP^{max}}(WL_r^t(s) + g_r^t(s)\lambda_t)$. Then, $w_r^{s^*,t^*} = 1$, and $w_r^{s,t} = 0$ for $t = 1, \ldots, T$ and $s = 1, \ldots, SP^{max}$ such that both $t \neq t^*$ and $s \neq s^*$.

The general scheme of the Lagrangian relaxation heuristic (see e.g. [64]) is presented in Algorithm 4.6.

**Algorithm 4.6** Lagrangian Relaxation Heuristic

1: *Step 1:* **Initialization.**
2: a. Initialize all multipliers to 0, i.e. $\lambda_t = 0$, $t = 1, \ldots, T$.
3: b. Set iteration number $k = 1$.
4: c. Initialize step length, $\alpha$.
5: d. Initialize lower bound $LB = -\infty$.
6: *Step 2:* **Solving the relaxed problem.** Solve Lagrangian relaxed problem (LRP($\lambda_t$)) for current values of multipliers $\lambda_t$ and calculate current lower bound $LB^k$.
7: *Step 3:* **Incumbent saving.** If $LB < LB^k$, then $LB := LB^k$
8: *Step 4:* **Feasibility heuristics.** Use the values of $w_r^{s,t}$ obtained in *Step 2* to find feasible solutions using the heuristics proposed in Section 4.3.3.2.
9: *Step 5:* **Updating multipliers.** Lagrangian multipliers are updated using the subgradient optimization method.
10: *Step 6:* **Stopping conditions.** If any stopping condition is met, save the best solution obtained so far and stop.
11: *Step 7:* **Update step length.** Update $\alpha$.
12: *Step 8:* Increment $k$ and go to Step 2.

#### 4.3.3.2 Feasibility of heuristics

Seven feasible heuristics are proposed, whose impact will be analyzed in Section 4.4.3 based on computational results. The first feasible heuristic, detailed in Algorithm 4.7, is the most straightforward. The metrology tool $t^*$ assigned to process machine $r$ is the one assigned when solving the Lagrangian relaxed problem, i.e. $\exists s = 1, \ldots, SP^{max}$ such that $w_r^{s,t^*} = 1$. Only the sampling rates are adjusted to satisfy the capacity constraints of metrology tools, that are considered one at a time. For a given metrology tool, the selection of the machine and the sampling rate to increase to reduce capacity consumption is based on the ratio between the risk increase and the capacity decrease. The process is repeated until the metrology capacity of $t$ is satisfied.

**Algorithm 4.7** Heuristic DMT-$H_1$

1: Let $\mathcal{F}(t)$ be the set of fixed pairs (machine, sampling period) for metrology tool $t$.
2: Initialize $\mathcal{F}(t)$ by using the optimal solution of the Lagrangian relaxed problem, i.e. $(r, s) \in \mathcal{F}(t)$ if $w_r^{s,t} = 1$.
3: Let $Capa_t$ be the capacity used on metrology tool $t$.
4: Initialize $Capa_t = \sum_{r=1}^{R} \sum_{s=1;(r,s)\in\mathcal{F}(t)}^{SP^{max}} g_r^t(s),\ t \in 1, \ldots, T$.
5: **for** $t = 1, \ldots, T$ **do**
6:     **while** $Capa_t > 1$ **do**
7:         Find the process machine $r^*$ and sampling period $s^*$ with the minimal ratio $\frac{WL_{r*}^t(s+1)-WL_{r*}^t(s^*)}{g_{r*}^t(s^*)-g_{r*}^t(s^*+1)}$ such that $(r^*, s^*) \in \mathcal{F}(t)$, i.e. $\frac{WL_{r*}^t(s^*+1)-WL_{r*}^t(s^*)}{g_{r*}^t(s^*)-g_{r*}^t(s^*+1)} = \min_{(r,s)\in\mathcal{F}(t)} \frac{WL_r^t(s+1)-WL_r^t(s)}{g_r^t(s)-g_r^t(s+1)}$.
8:         Change sampling rate of $r^*$ from $s^*$ to $s^*+1$, i.e. $\mathcal{F}(t) \leftarrow \mathcal{F}(t) - \{(r^*, s^*)\} \cup \{(r^*, s^*+1)\}$ and $Capa_t = Capa_t - g_{r^*}(s^*) + g_{r^*}(s^* + 1)$.
9:     **end while**
10: **end for**

In all other feasible heuristics, an assignment phase is used where the metrology tool selected for the process machine can change from the one assigned when solving the Lagrangian relaxed problem. All of them have two phases: An assignment and an improving phase.

In the **assignment phase**, presented in Algorithm (4.8) the process machine $r$ is associated to the metrology tool $t$ with a sampling rate $s$, and the metrology capacity is updated. If the capacity of $t$ is exceeded, then $SP_r$ is increased until either the capacity of $t$ is enough or $SP_r = SP^{max}$. In the latter case, it means that the solution is not feasible.

**Algorithm 4.8** Assignment phase $(r^*, t^*, s^*, Capa_{t^*}, \mathcal{G})$

**if** $Capa_{t^*} + g_{r^*}^{t^*}(s^*) \leq 1$ **then**
  Assign $r^*$ to $t^*$ with sampling rate $s^*$, i.e. $\mathcal{G} \leftarrow \mathcal{G} \cup \{(r^*, s^*, t^*)\}$.
  $Capa_{t^*} = Capa_{t^*} + g_{r^*}^{t^*}(s^*)$.
**else**
  **while** $Capa_{t^*} + g_{r^*}^{t^*}(s^*) > 1$ and $s^* \leq SP^{max}$ **do**
    $s^* = s^* + 1$.
  **end while**
  **if** $Capa_{t^*} + g_{r^*}^{t^*}(s^*) \leq 1$ **then**
    Assign $r^*$ to $t^*$ with sampling rate $s^*$, i.e. $\mathcal{G} \leftarrow \mathcal{G} \cup \{(r^*, s^*, t^*)\}$.
    $Capa_{t^*} = Capa_{t^*} + g_{r^*}^{t^*}(s^*)$.
  **else**
    Assign $r^*$ to $t^*$ with sampling rate $s^* = SP^{max}$, i.e. $\mathcal{G} \leftarrow \mathcal{G} \cup \{(r^*, SP^{max}, t^*)\}$.
    $Capa_{t^*} = Capa_{t^*} + g_{r^*}^{t^*}(SP^{max})$ and problem is infeasible.
  **end if**
**end if**

As a final stage of each heuristic, an **improving phase** is performed, presented in Algorithm (4.9). A Multi-Choice Knapsack Problem is solved for each metrology tool with the corresponding process machines assigned that were previously determined. This is done with Heuristic UMT-$H_2/H_3$ by combining Heuristic UMT-$H_2$ (Algorithm( 4.2)) and Heuristic UMT-$H_3$ (Algorithm( 4.3)). The new solution for each metrology tool is only kept if it improves the current solution.

**Algorithm 4.9** Improving phase($\mathcal{G}$)

Set $\mathcal{G}' \leftarrow \mathcal{G}$.
**for** $t = 1, \ldots, T$ **do**
  Solve a Multi-Choice Knapsack Problem with Heuristic UMT-$H_2/H_3$ (combining Algorithms (4.2) and (4.3) for the process machines $r$ assigned to metrology tool $t$, i.e. $\exists s = \{1, \ldots, SP^{max}\}$ such that $(r,s,t) \in \mathcal{G}$.
  Update $\mathcal{G}'$ with new sampling rates for metrology tool $t$.
**end for**
**if** $\sum_{t=1}^{T} \sum_{r=1}^{R} \sum_{s=1;(r,s,t)\in\mathcal{G}'}^{SP^{max}} WL_r^t(s) < \sum_{t=1}^{T} \sum_{r=1}^{R} \sum_{s=1;(r,s,t)\in\mathcal{G}}^{SP^{max}} WL_r^t(s)$ **then**
  $\mathcal{G} \leftarrow \mathcal{G}'$.
**end if**

For the first three heuristics, as a first stage the sampling rates, $SP_r$, are obtained by using the optimal solution of the Lagrangian relaxation problem. The **Heuristic DMT-$H_2$** shown in Algorithm (4.10) is metrology tool based. At each iteration, the metrology tool $t$ with the largest remaining capacity is selected. Then, the process machine $r$ which provides the larger portion of metrology capacity consumed for $t$ ($g_r^t(s)$) among the process machines that are not assigned yet is selected. The metrology capacity must be satisfied by assigning the process machine to $t$.

**Algorithm 4.10** Heuristic DMT-$H_2$

Let $\mathcal{R} = \{1, \ldots, R\}$ be the set of process machines.
Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Initialize $\mathcal{F}$ by using the optimal solution of the Lagrangian relaxation problem.
Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool). $\mathcal{G} \leftarrow \emptyset$.
Let $Capa_t$ be the capacity used on metrology tool $t$. $Capa_t = 0, t = 1, \ldots, T$.
**while** $\mathcal{R} \neq \emptyset$ **do**
    Select $t^*$, the metrology tool with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\ldots,T}(Capa_t)$.
    Determine $\Delta_r = \{g_r^{t^*}(s) - \min_{t'=1,\ldots,T; t' \neq t^* \text{ and } Capa_{t'} + g_r^{t'}(s) \leq 1}(g_r^{t'}(s))\} \quad \forall r \in \mathcal{R}$ and $s$ such that $(r, s) \in \mathcal{F}$.
    Let $\mathcal{V} = \{r \in \mathcal{R} | \Delta_r = \max_{r' \in \mathcal{R}}(\Delta_{r'})\}$.
    Select $r^* \in \mathcal{V}$ and $s^*$ such that $(r^*, s^*) \in \mathcal{F}$ and $t^*$ such that $g_{r^*}^{t^*}(s^*) = \max_{r \in \mathcal{V}, (r,s) \in \mathcal{F}}(g_r^{t^*}(s))$.
    **Assignment phase**$(r^*, t^*, s^*, Capa_{t^*}, \mathcal{G})$.
    $\mathcal{R} \leftarrow \mathcal{R} - \{r^*\}$.
**end while**
**Improving phase**$(\mathcal{G})$.

**Heuristic DMT-$H_3$** presented in Algorithm (4.11) is process machine based. The differences with the previous one are in blue. The first step is to search by process machine $r$ with which metrology tool $t$ the portion of metrology capacity consumed ($g^t_r(s)$) is the largest, then, select the process machine $r$ that presents the largest one. In next step, the metrology tool $t$ that presents the minimum portion of metrology capacity, $g^t_r(s)$, for the process machine $r$ and that respects the metrology capacity, is assigned. In case of not finding a metrology tool that assining a process machine $r$ with a sampling rate $s$ satisfies the metrology capacity, the metrology with the lowest utilization is selected.

**Algorithm 4.11** Heuristic DMT-$H_3$

Let $\mathcal{R} = \{1, \ldots, R\}$ be the set of process machines.
Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Initialize $\mathcal{F}$ by using the optimal solution of the Lagrangian relaxation problem.
Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool). $\mathcal{G} \leftarrow \emptyset$.
Let $Capa_t$ be the capacity used on metrology tool $t$. $Capa_t = 0, t = 1, \ldots, T$.
**while** $\mathcal{R} \neq \emptyset$ **do**
    Calculate $\Delta^t_r = \{g^t_r(s) - min_{t'=1,\ldots,T;t' \neq t \text{ and } Capa_{t'} + g^{t'}_r(s) \leq 1}(g^{t'}_r(s))\}$ $\forall r \in \mathcal{R}$ and $s$ such that $(r, s) \in \mathcal{F}, \forall t \in \mathcal{T}$.
    Let $\mathcal{V} = \left\{(r, t) \in \mathcal{R} \times \mathcal{T} | \Delta^t_r = \max_{r' \in \mathcal{R}, t' \in \mathcal{T}}(\Delta^{t'}_{r'})\right\}$.
    Select $(r^*, t) \in \mathcal{V}$and $s^*$ such that $(r^*, s^*) \in \mathcal{F}$ and $g^t_{r^*}(s^*) = \max_{(r,t) \in \mathcal{V},(r,s) \in \mathcal{F}}(g^t_r(s))$.
    Let $\mathcal{W} = \{t \in \mathcal{T} | g^t_{r^*}(s^*) = \min_{t'=1,\ldots,T;t' \neq t \text{ and } Capa_{t'} + g^{t'}_{r^*}(s^*) \leq 1}(g^{t'}_{r^*}(s^*))\}$.
    Select $t^* \in \mathcal{W}$ with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\ldots,T}(Capa_t)$.
    **if** $\nexists t \in V^t$ such that $Capa_t + g^t_{r^*}(s) \leq 1$ **then**
        Select $t^* \in \mathcal{T}$ with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\ldots,T}(Capa_t)$.
    **end if**
    **Assignment phase**$(r^*, t^*, s^*, Capa_{t^*}, \mathcal{G})$.
    $\mathcal{R} \leftarrow \mathcal{R} - \{r^*\}$.
**end while**
**Improving phase**$(\mathcal{G})$.

**Heuristic DMT-$H_4$** shown in Algorithm (4.12), is a modification of Heuristic DMT-$H_2$. The main difference is that after choosing the metrology tool $t$ in first place and the process machine $r$ right after, the capacity gap available that will remain by assigning $r$ for each metrology tool is calculated. The metrology tool $t$ that has the minimum capacity gap available, $\delta_t$, will be associated to the process machine $r$.

**Algorithm 4.12** Heuristic DMT-$H_4$

Let $\mathcal{R} = \{1, \ldots, R\}$ be the set of process machines.
Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Initialize $\mathcal{F}$ by using the optimal solution of the Lagrangian relaxation problem.
Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool). $\mathcal{G} \leftarrow \emptyset$.
Let $Capa_t$ be the capacity used on metrology tool $t$. $Capa_t = 0$, $t = 1, \ldots, T$.
**while** $\mathcal{R} \neq \emptyset$ **do**
    Select $t^*$, the metrology tool with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\ldots,T}(Capa_t)$.
    Determine $\Delta_r = \{g_r^{t^*}(s) - \min_{t'=1,\ldots,T; t' \neq t^* \text{ and } Capa_{t'} + g_r^{t'}(s) \leq 1}(g_r^{t'}(s))\}$ $\forall r \in \mathcal{R}$ and $s$ such that $(r, s) \in \mathcal{F}$.
    Let $\mathcal{V} = \{r \in \mathcal{R} | \Delta_r = \max_{r' \in \mathcal{R}}(\Delta_{r'})\}$.
    Select $r^* \in \mathcal{V}$ and $s^*$ such that $(r^*, s^*) \in \mathcal{F}$ and $t^*$ such that $g_{r^*}^{t^*}(s^*) = \max_{r \in \mathcal{V}, (r,s) \in \mathcal{F}}(g_r^{t^*}(s))$.
    Calculate $\delta_t = \{\min_{t=1,\ldots,T\,;\, Capa_t + g_{r^*}^t(s^*) \leq 1}(1 - (Capa_t + g_{r^*}^t(s^*)))\}$. $\delta_{t^*}$ is the minimum capacity gap available in metrology tool $t^*$ by adding $r^*$.
    Select $t^*$, the metrology tool with the lowest $\delta_t$, i.e. such that $\delta_{t^*} = \min_{t \in \{1,\ldots,T\}} \delta_t$ and assign $r^*$ to $t^*$.
    **Assignment phase**$(r^*, t^*, s^*, Capa_{t^*}, \mathcal{G})$.
    $\mathcal{R} \leftarrow \mathcal{R} - \{r^*\}$.
**end while**
**Improving phase**$(\mathcal{G})$.

Three other heuristics DMT-$H_5$, DMT-$H_6$ and DMT-$H_7$ are proposed, which are based on heuristics DMT-$H_2$, DMT-$H_3$ and DMT-$H_4$, respectively, and the main differences are marked in green. The initial assignment of process machines to metrology tools is obtained by using the optimal solution of the Lagrangian relaxation problem. The sampling rates, $SP_r$, are then determined by solving the MCKP with Heuristic UMT-$H_1$ (Algorithm (4.1)) for a unique metrology tool with a metrology capacity of $T$.

The **Heuristic DMT-$H_5$** shown in Algorithm (4.13) is metrology tool based and is a modification of DMT-$H_2$.

**Algorithm 4.13** Heuristic DMT-$H_5$

Let $\mathcal{R} = \{1, \ldots, R\}$ be the set of process machines.
Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Let $\mathcal{M}$ be the set of fixed pairs (process machine, metrology tool). $\mathcal{M} \leftarrow \emptyset$.
Initialize $\mathcal{M}$ by using the optimal solution of the Lagrangian relaxation problem.
Determine $\mathcal{F}$ by solving the Multi-Choice Knapsack Problem with Heuristic 1 of Algorithm (4.1) for the process machines $r = 1, \ldots, R$ considering a unique metrology tool with a $Capa = \sum_{t=1}^{T} Capa_t$. Use $g_r^t(s)$ and $WL_r^t(s)$ with the metrology allocation obtained in $\mathcal{M}$.
Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool). $\mathcal{G} \leftarrow \emptyset$.
Let $Capa_t$ be the capacity used on metrology tool $t$. $Capa_t = 0$, $t = 1, \ldots, T$.
**while** $\mathcal{R} \neq \emptyset$ **do**
    Select $t^*$, the metrology tool with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\ldots,T}(Capa_t)$.
    Determine $\Delta_r = \{g_r^{t^*}(s) - \min_{t'=1,\ldots,T; t' \neq t^* \text{ and } Capa_{t'} + g_r^{t'}(s) \leq 1}(g_r^{t'}(s))\}$ $\forall r \in \mathcal{R}$ and $s$ such that $(r, s) \in \mathcal{F}$.
    Let $\mathcal{V} = \{r \in \mathcal{R} | \Delta_r = \max_{r' \in \mathcal{R}}(\Delta_{r'})\}$.
    Select $r^* \in \mathcal{V}$and $s^*$ such that $(r^*, s^*) \in \mathcal{F}$ and $t^*$ such that $g_{r^*}^{t^*}(s^*) = \max_{r \in \mathcal{V}, (r,s) \in \mathcal{F}}(g_r^{t^*}(s))$.
    **Assignment phase**$(r^*, t^*, s^*, Capa_{t^*}, \mathcal{G})$.
    $\mathcal{R} \leftarrow \mathcal{R} - \{r^*\}$.
**end while**
**Improving phase**$(\mathcal{G})$.

**Heuristic DMT-$H_6$** presented in Algorithm (4.14) is process machine based and is a modification of DMT-$H_3$.

**Algorithm 4.14** Heuristic DMT-$H_6$

Let $\mathcal{R} = \{1, \ldots, R\}$ be the set of process machines.
Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Let $\mathcal{M}$ be the set of fixed pairs (process machine, metrology tool). $\mathcal{M} \leftarrow \emptyset$.
Initialize $\mathcal{M}$ by using the optimal solution of the Lagrangian relaxation problem.
Determine $\mathcal{F}$ by solving the Multi-Choice Knapsack Problem with Heuristic 1 of Algorithm (4.1) for the process machines $r = 1, \ldots, R$ considering a unique metrology tool with a $Capa = \sum_{t=1}^{T} Capa_t$. Use $g_r^t(s)$ and $WL_r^t(s)$ with the metrology allocation obtained in $\mathcal{M}$.
Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool). $\mathcal{G} \leftarrow \emptyset$.
Let $Capa_t$ be the capacity used on metrology tool $t$. $Capa_t = 0$, $t = 1, \ldots, T$.
**while** $\mathcal{R} \neq \emptyset$ **do**
    Calculate $\Delta_r^t = \{g_r^t(s) - min_{t'=1,\ldots,T; t' \neq t \text{ and } Capa_{t'} + g_r^{t'}(s) \leq 1}(g_r^{t'}(s))\}$ $\forall r \in \mathcal{R}$ and $s$ such that $(r, s) \in \mathcal{F}$, $\forall t \in \mathcal{T}$.
    Let $\mathcal{V} = \left\{(r, t) \in \mathcal{R} \times \mathcal{T} | \Delta_r^t = \max_{r' \in \mathcal{R}, t' \in \mathcal{T}}(\Delta_{r'}^{t'})\right\}$.
    Select $(r^*, t) \in \mathcal{V}$and $s^*$ such that $(r^*, s^*) \in \mathcal{F}$ and $g_{r^*}^t(s^*) = \max_{(r,t) \in \mathcal{V}, (r,s) \in \mathcal{F}}(g_r^t(s))$.
    Let $\mathcal{W} = \{t \in \mathcal{T} | g_{r^*}^t(s^*) = \min_{t'=1,\ldots,T; t' \neq t \text{ and } Capa_{t'} + g_{r^*}^{t'}(s^*) \leq 1}(g_{r^*}^{t'}(s^*))\}$.
    Select $t^* \in \mathcal{W}$ with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\ldots,T}(Capa_t)$.
    **if** $\nexists t \in V^t$ such that $Capa_t + g_{r^*}^t(s) \leq 1$ **then**
        Select $t^* \in \mathcal{T}$ with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\ldots,T}(Capa_t)$.
    **end if**
    **Assignment phase**$(r^*, t^*, s^*, Capa_{t^*}, \mathcal{G})$.
    $\mathcal{R} \leftarrow \mathcal{R} - \{r^*\}$.
**end while**
**Improving phase**$(\mathcal{G})$.

**Heuristic DMT-$H_7$** described in Algorithm (4.15) is metrology tool based and is a modification of DMT-$H_4$.

**Algorithm 4.15** Heuristic DMT-$H_7$

Let $\mathcal{R} = \{1, \dots, R\}$ be the set of process machines.
Let $\mathcal{F}$ be the set of fixed pairs (process machine, sampling period). $\mathcal{F} \leftarrow \emptyset$.
Let $\mathcal{M}$ be the set of fixed pairs (process machine, metrology tool). $\mathcal{M} \leftarrow \emptyset$.
Initialize $\mathcal{M}$ by using the optimal solution of the Lagrangian relaxation problem.
Determine $\mathcal{F}$ by solving the Multi-Choice Knapsack Problem with Heuristic 1 of Algorithm (4.1) for the process machines $r = 1, \dots, R$ considering a unique metrology tool with a $Capa = \sum_{t=1}^{T} Capa_t$. Use $g_r^t(s)$ and $WL_r^t(s)$ with the metrology allocation obtained in $\mathcal{M}$.
Let $\mathcal{G}$ be the set of fixed triplets (process machine, sampling period, metrology tool). $\mathcal{G} \leftarrow \emptyset$.
Let $Capa_t$ be the capacity used on metrology tool $t$. $Capa_t = 0, t = 1, \dots, T$.
**while** $\mathcal{R} \neq \emptyset$ **do**
  Select $t^*$, the metrology tool with the lowest current utilization, i.e. $Capa_{t^*} = \min_{t=1,\dots,T}(Capa_t)$.
  Determine $\Delta_r = \{g_r^{t^*}(s) - \min_{t'=1,\dots,T; t' \neq t^* \text{ and } Capa_{t'} + g_r^{t'}(s) \leq 1}(g_r^{t'}(s))\}$ $\forall r \in \mathcal{R}$ and $s$ such that $(r, s) \in \mathcal{F}$.
  Let $\mathcal{V} = \{r \in \mathcal{R} | \Delta_r = \max_{r' \in \mathcal{R}}(\Delta_{r'})\}$.
  Select $r^* \in \mathcal{V}$ and $s^*$ such that $(r^*, s^*) \in \mathcal{F}$ and $t^*$ such that $g_{r^*}^{t^*}(s^*) = \max_{r \in \mathcal{V}, (r,s) \in \mathcal{F}}(g_r^{t^*}(s))$.
  Calculate $\delta_t = \{\min_{t=1,\dots,T\,;\,Capa_t + g_{r^*}^t(s^*) \leq 1}(1 - (Capa_t + g_{r^*}^t(s^*)))\}$. $\delta_{t^*}$ is the minimum capacity gap available in metrology tool $t^*$ by adding $r^*$.
  Select $t^*$, the metrology tool with the lowest $\delta_t$, i.e. such that $\delta_{t^*} = \min_{t \in \{1,\dots,T\}} \delta_t$ and assign $r^*$ to $t^*$.
  **Assignment phase**$(r^*, t^*, s^*, Capa_{t^*}, \mathcal{G})$.
  $\mathcal{R} \leftarrow \mathcal{R} - \{r^*\}$.
**end while**
**Improving phase**$(\mathcal{G})$.

## 4.4 Numerical experiments

In this section, the performance of the previous heuristics is analyzed on numerous randomly generated instances. The results obtained are compared with the ILP and the standard solver IBM ILOG CPLEX 12.6.

### 4.4.1 Unique metrology tool

The number process of machines is chosen in the set $\{5, 10, 15, 20, 30\}$ and the characteristics of each process machine $r$ are defined as follows. The probability of failure $p_r$ is generated from a uniform distribution $U[p_{min}; p_{max}]$, where $p_{min}$ is kept constant ($p_{min} = 0.01$) and $p_{max}$ is chosen in the set $\{0.02, 0.05, 0.1, 0.2\}$. The throughput rate $TP_r$ is generated from a distribution $U[TP_{min}; TP_{max}]$, where $TP_{max} = 1000$ and $TP_{min}$ is chosen in the set $\{10, 100, 500, 900\}$. The measurement rate $TM_r$ is determined using the ratio $\frac{R \cdot \overline{TP_r}}{TM_r}$ chosen the set $\{5, 10, 20, 30\}$ where $\overline{TP_r}$ is the average throughput rate for the considered instance. A maximum sampling rate is set, $SP^{max} = 500$ for all machines.

Combining these parameters leads to 320 instances, with 10 instances generated for each instance with different randomly generated values of $p_r$ and $TP_r$. Thus, a total of 3200 different experiments were conducted.

Table 4.1 presents the results by heuristic ordered by the number of process machines used. Each of the three heuristics are presented in comparison with the optimal solution. A combination of Heuristic UMT-$H_2$ and Heuristic UMT-$H_3$, Heuristic UMT-$H_2/H_3$, taken the best value obtained from two heuristics is proposed because of they do not perform well in some cases. For each of the four heuristics, the proportion of cases for which the optimal solution is reached (*Hits*) is provided. Table 4.1 shows the average increase and the upper bound (*Worst*) in the objective function as compared with the optimal value. Since Heuristic UMT-$H_1$ is dominated by Heuristic UMT-$H_2$, we concentrate our analysis on Heuristic UMT-$H_2$, Heuristic UMT-$H_3$ and the combined Heuristic UMT-$H_2/H_3$. Not surprisingly, we observe that the higher the number of process machines used, the better all heuristics perform. The excellent results of Heuristic UMT-$H_2/H_3$ is when $R$=10 or more. We also remark that it does not mean that the heuristics tend to reach the optimum more easily, but that a larger number of machines means more flexibility in the metrology capacity assignment, and the closer the integer problem to its continuous approximation.

**Table 4.1 –** *Results on all randomly generated instances by number of machines R - Model (PI1).*

| | Heuristic UMT-$H_1$ | | | Heuristic UMT-$H_2$ | | | Heuristic UMT-$H_3$ | | | Heuristic UMT-$H_2/H_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $R$ | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) |
| 5 | 14.1 | 0.6 | 9.7 | 26.3 | 0.2 | 4 | 24.4 | 0.5 | 13.5 | 33.6 | 0.1 | 1.5 |
| 10 | 6.9 | 0.4 | 4.9 | 16.4 | 0.1 | 1 | 16.7 | 0.3 | 8.4 | 23 | 0.1 | 1 |
| 15 | 5.8 | 0.2 | 2.9 | 14.4 | 0.1 | 0.6 | 14.1 | 0.2 | 2.1 | 19.8 | 0 | 0.6 |
| 20 | 4.1 | 0.2 | 2.1 | 9.8 | 0 | 0.4 | 9.4 | 0.2 | 2.7 | 13.8 | 0 | 0.3 |
| 30 | 2.2 | 0.1 | 1.9 | 6.9 | 0 | 0.4 | 7.7 | 0.1 | 1.5 | 11.6 | 0 | 0.4 |

Pursuing the analysis in which $R$=5, the next parameter analyzed is the ratio $\frac{R \cdot \overline{TP_r}}{TM_r}$. The higher this ratio, the higher the throughput of the process machine compared to the metrology capacity. While this indicates a higher stress on the operations, Table 4.2 shows how it actually makes it easier for the heuristics to reach a good solution. Again, the reason lies in the fact that a higher $\frac{R \cdot \overline{TP_r}}{TM_r}$ ratio leads to higher values of $SP_r$ in the solution, and thus smaller capacity steps between adjacent sampling rates, bringing the integer problem closer to the continuous approximation. Note that although solutions are of better quality when the ratio increases, the reason is not that more optimal solutions are found.

**Table 4.2 –** *Results on randomly generated instances for R=5 by ratio* $\frac{R \cdot \overline{TP_r}}{TM_r}$ *- Model (PI1).*

| | Heuristic UMT-$H_1$ | | | Heuristic UMT-$H_2$ | | | Heuristic UMT-$H_3$ | | | Heuristic UMT-$H_2/H_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\frac{R \cdot \overline{TP_r}}{TM_r}$ | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) |
| 5 | 31.3 | 1.2 | 9.7 | 48.8 | 0.4 | 4 | 49.4 | 0.9 | 13.5 | 64.4 | 0.2 | 1.5 |
| 10 | 14.4 | 0.7 | 6.1 | 28.1 | 0.2 | 1.4 | 24.4 | 0.5 | 2.3 | 35.6 | 0.2 | 1 |
| 20 | 8.1 | 0.4 | 2.3 | 18.8 | 0.1 | 0.6 | 13.8 | 0.3 | 2.1 | 20.6 | 0.1 | 0.5 |
| 30 | 2.5 | 0.2 | 1.4 | 9.4 | 0.1 | 0.3 | 10 | 0.2 | 1.3 | 13.8 | 0.1 | 0.3 |

Tables 4.3 and 4.4 present the results for different values of $TP_{min}$ and $p_{max}$. Let us recall that in our experimentation, $TP_{max}$ and $p_{min}$ are kept constant, so that $TP_{min}$ and $p_{max}$ express, in fact, a measure of both the nominal values and the allowed diversity of process machines in terms of capacity and failure rate among the 10 randomly generated instances.

**Table 4.3 –** *Results on randomly generated instances for R=5 and by parameter* $TP_{min}$ *- Model (PI1).*

| | Heuristic UMT-$H_1$ | | | Heuristic UMT-$H_2$ | | | Heuristic UMT-$H_3$ | | | Heuristic UMT-$H_2/H_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $TP_{min}$ | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) |
| 10 | 10.6 | 0.9 | 9.7 | 27.5 | 0.2 | 4 | 28.8 | 0.4 | 4.9 | 38.8 | 0.1 | 1.2 |
| 100 | 13.1 | 0.6 | 5.8 | 26.3 | 0.2 | 2.4 | 26.9 | 0.5 | 13.5 | 35 | 0.1 | 1 |
| 500 | 16.3 | 0.5 | 5.3 | 23.1 | 0.2 | 1.3 | 22.5 | 0.4 | 5.1 | 30 | 0.2 | 1.3 |
| 900 | 16.3 | 0.5 | 3.1 | 28.1 | 0.2 | 1.7 | 19.4 | 0.5 | 4.2 | 30.6 | 0.2 | 1.5 |

**Table 4.4 –** *Results on randomly generated instances for R=5 and by parameter $p_{max}$ - Model (PI1).*

| | Heuristic UMT-$H_1$ | | | Heuristic UMT-$H_2$ | | | Heuristic UMT-$H_3$ | | | Heuristic UMT-$H_2/H_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p_{max}$ | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) | Hits (%) | Average (%) | Worst (%) |
| 0.02 | 16.9 | 0.7 | 9.4 | 28.8 | 0.3 | 4 | 25.6 | 0.5 | 4.8 | 33.1 | 0.2 | 1.3 |
| 0.05 | 15 | 0.6 | 5.6 | 29.4 | 0.2 | 1.6 | 27.5 | 0.4 | 3.7 | 40 | 0.1 | 1.5 |
| 0.1 | 11.3 | 0.6 | 9.7 | 25.6 | 0.2 | 3.2 | 23.8 | 0.6 | 13.5 | 31.9 | 0.1 | 0.8 |
| 0.2 | 13.1 | 0.5 | 5.3 | 21.3 | 0.2 | 1.3 | 20.6 | 0.4 | 3.7 | 29.4 | 0.1 | 1.3 |

The worst case measure is of special interest since it reflects the more extreme instance of the group. Indeed, no significant trend of the averages in any of cases presented in either Tables 4.3 or 4.4 is observed. Note that $TP_{min}$ has an impact, although limited, on the results. Higher values of this parameter means higher throughput rates on all machines, larger $SP_r$ values, and as explained before, better heuristic performance measures.

While one could expect the failure rate to have a similar effect on the heuristic behavior (the larger $p_r$, the smaller the sampling rate for an equivalent Wafer Loss), it does not seem to have any clear impact on the results. This is most probably due to the fact that $p_r$ does not affect the capacity consumption but the objective function only. In addition, the limited range these values were taken from, which is consistent with the machine behavior in practice, may reduce even further any possible effects on the results.

Finally, note that, in the vast majority of the cases across all four tables, the worst case performance of the combined heuristic, Heuristic UMT-$H_2/H_3$, is significantly better than either worst cases for Heuristic UMT-$H_2$ and Heuristic UMT-$H_3$ taken separately. This shows how both heuristics perform differently enough in difficult cases to provide a safety net, so to communicate, for one another, and to make their combination efficient.

### 4.4.2 Identical metrology tools

The number of process machines is chosen in the set $\{5, 10, 20, 40\}$ and the number of metrology tools in the set $\{3, 5\}$. The characteristics of each process machine $r$ are defined as follows. The probability of failure $p_r$ is generated from a uniform distribution $U[p_{min}; p_{max}]$, where $p_{min}$ is kept constant ($p_{min} = 0.01$) and $p_{max}$ is chosen in the set $\{0.05, 0.2\}$. The throughput rate $TP_r$ is generated from a distribution $U[TP_{min}; TP_{max}]$, where $TP_{max} = 1000$ and $TP_{min}$ is chosen in the set $\{100, 900\}$. The measurement rate $TM_r$ is determined using the ratio $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r}$ chosen the set $\{5, 10, 30\}$ where $\overline{TP_r}$ is the average throughput rate for the considered instance. A maximum sampling rate is set, $SP^{max} = 500$ for all machines.

Combining these parameters leads to 96 instances, with 30 instances generated for each instance with different randomly generated values of $p_r$ and $TP_r$. Thus, a total of 2880 different experiments were conducted.

Table 4.5 presents the number of instances, out of total of 360, for which Heuristic IMT-$H_1$ finds a feasible solution. As expected, the only case where a feasible solution is always found is when $R = T$. However, note that Heuristic IMT-$H_1$ finds a feasible solution in the vast majority of the instances

**Table 4.5 –** *Number of feasible solutions (out of 360) determined by IMT-$H_1$ - Model (PI2).*

| | T | |
|---|---|---|
| $R$ | 3 | 5 |
| 5 | 351 | 360 |
| 10 | 356 | 355 |
| 20 | 349 | 353 |
| 40 | 333 | 349 |

The performances of Heuristics IMT-$H_1$ and IMT-$H_1^+$ are compared. IMT-$H_1^+$ improves over IMT-$H_1$ in several ways. First, no matter the instance characteristics, IMT-$H_1^+$ always finds a feasible solution. Table 4.6 presents the improvement on the objective function brought by IMT-$H_1^+$ over IMT-$H_1$ (only for instances for which IMT-$H_1$ finds a feasible solution). Note that the improvement brought by IMT-$H_1^+$ is significant. The largest gains are obtained for low values of $R$ and $T$, for which the capacity $g_r(s)$ is usually the largest, thus leading to large increase of $s$, and the Wafer Loss, to make the solution feasible in IMT-$H_1$. In all instances, there are at least some for which IMT-$H_1^+$ does not improve over IMT-$H_1$.

**Table 4.6 –** *Comparison between IMT-$H_1$ and IMT-$H_1^+$ - Model (PI2).*

| | | T | |
|---|---|---|---|
| $R$ | | 3 | 5 |
| 5 | Avg | 2.6% | 0% |
| | Max | 7.3% | 0% |
| 10 | Avg | 2.3% | 1.1% |
| | Max | 8.5% | 6.1% |
| 20 | Avg | 0.9% | 0.4% |
| | Max | 2.6% | 1.8% |
| 40 | Avg | 0.7% | 0.2% |
| | Max | 2.3% | 1% |

Having assessed that IMT-$H_1^+$ strongly dominates IMT-$H_1$, only the performances of IMT-$H_1^+$ are now analyzed. Table 4.7 compares the results of IMT-$H_1^+$ to the upper bounds of the solver. The cases where $LB = UB$ ("Opt.", the solution obtained by the solver is guaranteed to be optimal) and where $LB \neq UB$ ("Non-opt.") are separated. Note that, when $T = R = 5$, both the solver and IMT-$H_1^+$ give optimal solutions. This is expected, since each metrology tool can be fully dedicated to one process machine. The optimal solution consists in filling the metrology capacity of each tool, and is therefore straightforward. Some instances, in particular with lower values of $R$ and $T$, are more challenging for the heuristic when $LB = UB$. For $T = 3$ and in the worst case, IMT-$H_1^+$ yields a solution with an expected Wafer Loss which is larger than the optimum by 6.4% for $R = 5$ and 3.4% for $R = 10$. However, on average, the performance of IMT-$H_1^+$ is excellent. Note also that, when $LB \neq UB$

and in some cases, IMT-$H_1^+$ finds a better solution than the solver. In all cases but one, the minimum difference is negligible.

**Table 4.7 –** *Comparison between IMT-$H_1^+$ and UB - Model (PI2).*

| | | T | | | |
|---|---|---|---|---|---|
| | | 3 | | 5 | |
| $R$ | | Non-opt. | Opt. | Non-opt. | Opt. |
| 5 | Avg | 0.2% | 0.7% | | 0% |
| | Min | 0% | 0% | | 0% |
| | Max | 0.4% | 6.4% | | 0% |
| 10 | Avg | 0.4% | 0.9% | 0.2% | 0.9% |
| | Min | 0% | 0% | -0.1% | 0% |
| | Max | 2% | 3.4% | 1.7% | 3.7% |
| 20 | Avg | 0.1% | 0.4% | 1% | % |
| | Min | 0% | 0% | 0.2% | % |
| | Max | 0.9% | 1.8% | 2% | % |
| 40 | Avg | 0.1% | 0.2% | 0.1% | 0.4% |
| | Min | 0% | 0% | 0% | 0.2% |
| | Max | 0.3% | 0.6% | 0.6% | 0.6% |

In the next three tables (4.8 to 4.10) the results are presented by the instance characteristics. The tables present the impact of the ratio $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r}$, of $TP_{min}$ and of $p_{max}$ on the average performance of IMT-$H_1^+$ compared to the upper bounds provided by the solver. The results seem excellent across the boards, which confirms the robustness of IMT-$H_1^+$ when faced to very different instances.

Table 4.8 prompts some remarks. First, the higher $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r}$, the better results of IMT-$H_1^+$. This is explained by the fact that higher $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r}$ values lead to a higher stress on the metrology capacity, thus larger values of the sampling rates, and therefore smaller differences for $WL_r(s)$ between values of $s$, which allows to get closer to the optimal solution.

**Table 4.8 –** *Impact of Ratio on the comparison between IMT-$H_1^+$ and UB - Model (PI2).*

| | | Ratio | | |
|---|---|---|---|---|
| $R$ | T | 5 | 10 | 30 |
| 5 | 3 | 1.3% | 0.6% | 0.1% |
| | 5 | 0% | 0% | 0% |
| 10 | 3 | 1.2% | 0.7% | 0.3% |
| | 5 | 1.3% | 0.7% | 0.2% |
| 20 | 3 | 0.6% | 0.3% | 0.1% |
| | 5 | 0.8% | 0.4% | 0.1% |
| 40 | 3 | 0.2% | 0.1% | 0% |
| | 5 | 0.3% | 0.1% | 0% |

In order to discuss the next two tables, let us recall that $TP_r$ values are randomly chosen between $TP_{min}$ and $TP_{max} = 1000$, which means that with $TP_{min} = 900$, the values of the production rates are not only larger, but also that the range of allowed production rates in the instance is greatly reduced compared to the case where $TP_{min} = 100$. The same holds with the failure probabilities. Since $p_{min} = 0.01$, instances characterized by $p_{max} = 0.05$ allow for a narrower range of values. In both tables, these instances (with similar machines) seem to be more challenging for our heuristic.

**Table 4.9 –** *Impact of $TP_{min}$ on the comparison between IMT-$H_1^+$ and UB - Model (PI2).*

| | | $TP_{min}$ | |
|---|---|---|---|
| $R$ | T | 100 | 900 |
| 5 | 3 | 0.4% | 0.9% |
| | 5 | 0% | 0% |
| 10 | 3 | 0.6% | 0.9% |
| | 5 | 0.7% | 0.8% |
| 20 | 3 | 0.2% | 0.4% |
| | 5 | 0.4% | 0.4% |
| 40 | 3 | 0.1% | 0.1% |
| | 5 | 0.1% | 0.1% |

**Table 4.10 –** *Impact of $p_{max}$ on the comparison between IMT-$H_1^+$ and UB - Model (PI2).*

| | | $p_{max}$ | |
|---|---|---|---|
| $R$ | T | 0.05 | 0.2 |
| 5 | 3 | 0.7% | 0.6% |
| | 5 | 0% | 0% |
| 10 | 3 | 0.9% | 0.5% |
| | 5 | 0.8% | 0.6% |
| 20 | 3 | 0.4% | 0.2% |
| | 5 | 0.5% | 0.4% |
| 40 | 3 | 0.1% | 0.1% |
| | 5 | 0.2% | 0.1% |

In order to confirm our previous analysis, Table 4.11 presents for each separate combination of $R$ and $T$, the results for each pair of values $p_{max}$ and $TP_{min}$. Indeed, the heuristic systematically performs at its best when $p_{max} = 0.2$ and $TP_{min} = 100$, i.e; when the dissimilarity between machines is the greatest. The worst performance of IMT-$H_1^+$ is observed in the opposite situation, i.e. when $p_{max} = 0.05$ and $TP_{min} = 900$.

**Table 4.11 –** *Impact of $p_{max}$ on the comparison between IMT-$H_1^+$ and UB - Model (PI2).*

| | | $TP_{min}$ | | | |
|---|---|---|---|---|---|
| | | 100 | | 900 | |
| | | $p_{max}$ | | $p_{max}$ | |
| $T$ | R | 0.05 | 0.2 | 0.05 | 0.2 |
| 3 | 5 | 0.41% | **0.37**% | 0.98% | 0.89% |
| | 10 | 0.67% | **0.48**% | 1.15% | 0.58% |
| | 20 | 0.26% | **0.17**% | 0.52% | 0.31% |
| | 40 | 0.10% | **0.06**% | 0.18% | 0.09% |
| 5 | 5 | 0% | 0% | 0% | 0% |
| | 10 | 0.74% | **0.64**% | 0.88% | 0.66% |
| | 20 | 0.44% | **0.36**% | 0.47% | 0.39% |
| | 40 | 0.14% | **0.08**% | 0.17% | 0.11% |

### 4.4.3 Different metrology tools

A comparison between heuristics for identical and different metrology tools has been carried out. We wanted to check that the Lagrangian heuristics introduced in Section 4.3.3 provide better results than the heuristics in Section 4.3.2 for identical tools. Using the 2880 instances of Section 4.4.2, the heuristics DMT-$H_1$, DMT-$H_2$, DMT-$H_3$, DMT-$H_4$, DMT-$H_5$, DMT-$H_6$ and DMT-$H_7$ provide strictly better results in 57.3% of the instances (1651 instances) and the same results in 42.7% (1229 instances). Table 4.12 presents the results. In the second column it is shown the times that each heuristic provides the best Wafer Loss value (it could be shared with another heuristic), the third column when the heuristic is the only one to get the best Wafer Loss value and the last column shows the difference with the Lower Bound.

**Table 4.12 –** *Heuristic comparison - Model (PI3).*

| | Best Wafer Loss Value | | Only best Wafer Loss Value | | Difference Heuristic-LB | |
|---|---|---|---|---|---|---|
| Heuristic | Times | % of instances | Times | % of instances | Times | % of instances |
| DMT-$H_1$ | 0 | 0% | 0 | 0% | 175.1% | 78.9% |
| DMT-$H_2$ | 1431 | 49.7% | 0 | 0% | 29.5% | 25.3% |
| DMT-$H_3$ | 1431 | 49.7% | 0 | 0% | 29.5% | 25.3% |
| DMT-$H_4$ | 1434 | 49.8% | 124 | 4.31% | 29.5% | 25.3% |
| DMT-$H_5$ | 1230 | 42.7% | 0 | 0% | 29.5% | 25.3% |
| DMT-$H_6$ | 1230 | 42.7% | 0 | 0% | 29.5% | 25.3% |
| DMT-$H_7$ | 1259 | 43.7% | 595 | 20.7% | 29.5% | 25.3% |

As expected, Heuristic DMT-$H_1$ does not provide good results when the metrology tools are equal (same measurement throughput). The average of the difference between the Wafer-Loss of DMT-$H_1$ and the Lower-Bound is of 175.1%. The rest of heuristics share the best WaferLoss values from 42.71% to 49.8% of the instances.

Seeing the results, it is possible that DMT-$H_4$ dominates DMT-$H_2$ and DMT-$H_3$, and that DMT-$H_7$ dominates DMT-$H_6$ and DMT-$H_5$. Table 4.13 and 4.14 compare the instances with the same values for the two groups of heuristics.

**Table 4.13 –** *Heuristic comparison: DMT-$H_2$, DMT-$H_3$, DMT-$H_4$.*

| | instances with same Wafer Loss | |
|---|---|---|
| Heuristic | DMT-$H_2$ | DMT-$H_3$ |
| DMT-$H_2$ | - | 1431 |
| DMT-$H_3$ | 1431 | - |
| DMT-$H_4$ | 1304 | 1304 |

**Table 4.14 –** *Heuristic comparison: DMT-$H_5$, DMT-$H_6$, DMT-$H_7$.*

| | instances with same Wafer Loss | |
|---|---|---|
| Heuristic | DMT-$H_2$ | DMT-$H_3$ |
| DMT-$H_5$ | - | 1230 |
| DMT-$H_6$ | 1230 | - |
| DMT-$H_7$ | 585 | 585 |

We found that when several metrology tools are equal DMT-$H_2$ and DMT-$H_3$ provide the same best Wafer Loss values in the same instances, thus, the use of just only one among them would be enough, and DMT-$H_4$ does not dominates DMT-$H_2$ and DMT-$H_3$ because the best Wafer Loss values are in different instances. The same happens with DMT-$H_5$, DMT-$H_6$ and DMT-$H_7$. DMT-$H_5$ and DMT-$H_6$ bring the best Wafer Loss values in the same instances and using just one among them is enough, and DMT-$H_7$ does not dominates the other heuristics.

For the new instances, the number of process machines is chosen in the set $\{5, 10, 20, 40\}$ and the characteristics of each process machine $r$ are defined as follows. The probability of failure $p_r$ is generated from a uniform distribution $U[p_{min}; p_{max}]$, where $p_{min}$ is kept constant ($p_{min} = 0.01$) and $p_{max}$ is chosen in the set $\{0.05, 0.2\}$. The throughput rate $TP_r$ is generated from a distribution $U[TP_{min}; TP_{max}]$, where $TP_{max} = 1000$ and $TP_{min}$ is chosen in the set $\{100, 900\}$.

The number of metrology tools is in the set $\{3, 5\}$, and their reliability $\alpha_r^t$ is randomly generated from three cases that are considered to generate the measurement rate $TM_r^t$ for metrology tools values, which is determined using the ratio $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r^t}$. In the first case, all metrology tools are equally fast, and their measurement rate is independent of the different products processed on process machines $r$. The ratio $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r^t}$ is chosen in the set $\{5, 10, 30\}$, where $\overline{TP_r}$ is the average throughput rate for the considered instance, thus leading to a unique measurement rate for all tools and machines. In the second case, the measurement rate depends on the metrology tool, but remains independent of the different products processed on process machines $r$. One value of $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r^t}$ is randomly chosen from a uniform distribution. The distribution range is first at $U[2.5, 7.5]$, then at $U[5, 15]$ and finally at $U[15, 45]$. These ranges are set around the values chosen for the fixed case, and allow for the fastest tool to run at most three times faster than the slowest one. The last instance, we allow any value of $TM_r^t$, regardless of others, based on the ratio $\frac{R \cdot \overline{TP_r}}{T \cdot TM_r^t}$ taken from the same uniform distributions previously mentioned.

A maximum sampling rate is set, $SP^{max} = 500$ for all machines. Combining these parameters leads to 864 instances, with 10 instances generated for each instance with different randomly generated values of $p_r$ and $TP_r$. Thus, a total of 8640 different experiments were conducted. The calculation of $WL_r^t(SP_r)$ includes an infinite sum that is calculated iteratively. In order to keep the accuracy level uniform between experiments, the calculations are stopped when the values of two consecutive $WL_r^t(SP_r)$ differ by less than 0.1%.

Table 4.15 shows the heuristics comparison for all the instances. There is no a dominant heuristics and all of them provide among all the instances an only best Wafer Loss value. In the case of uniform policy, DMT-$H_1$, DMT-$H_4$ and DMT-$H_7$ get the only best Wafer Loss value. For a random policy, DMT-$H_6$ is clearly the most efficient heuristic with 1538 instances with the only best Wafer Loss value. For the connected policy, the results are spread out among all heuristics.

**Table 4.15 –** *Heuristic comparison.*

| | Best Wafer Loss Value | | | | | Only best Wafer Loss Value | | | | | Difference Heuristic-LB | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Policy | | | | | Policy | | | | |
| Heuristic | Times | % of instances | Connect | Random | Uniform | Times | % of instances | Connect | Random | Uniform | % Average | % STDEV |
| DMT-$H_1$ | 1623 | 18.8% | 240 | 820 | 563 | 1495 | 17.3% | 223 | 714 | 558 | 22.8% | 19.3% |
| DMT-$H_2$ | 1732 | 20% | 471 | 0 | 1261 | 309 | 3.6% | 309 | 0 | 0 | 32.7% | 25.5% |
| DMT-$H_3$ | 1960 | 22.7% | 389 | 310 | 1261 | 491 | 5.7% | 319 | 172 | 0 | 23.5% | 19.9% |
| DMT-$H_4$ | 1739 | 20.1% | 436 | 95 | 1208 | 462 | 5.3% | 267 | 45 | 150 | 28.4% | 23% |
| DMT-$H_5$ | 1158 | 13.4% | 466 | 0 | 692 | 337 | 3.9% | 337 | 0 | 0 | 33.9% | 26.1% |
| DMT-$H_6$ | 3155 | 36.5% | 695 | 1768 | 692 | 2145 | 24.8% | 607 | 1538 | 0 | 21.2% | 19% |
| DMT-$H_7$ | 1556 | 18% | 651 | 269 | 636 | 936 | 10.8% | 479 | 102 | 355 | 26.9 | 20.9% |

## 4.5 Industrial results

In the frame of several sampling strategy changes performed in STMicroelectronics at Rousset site, the calculation of sampling rates for process machines associated to metrology tools was needed in order to control lots through a sampling by equipment strategy.

A previous analysis of the failure probability influence on the final results was done. For a similar metrology tool workshop, with 34 process machines covered with probability to failure $p_r = 0.1$, except for machine 1 that is decreased while the one of the machine 2 is increased. The results in Table 4.16 show that, as expected, the sampling period of machine 1 is larger than the rest because it is a safer machine while the one of machine 2 is smaller than the others since it is a more dangerous machine that needs to be controlled more regularly.

**Table 4.16 –** *Results on an industrial instance with different failure probabilities for two machines.*

| Machines | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Optimized $SP_r$ | **7** | **2** | 4 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 4 |
| Machines | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 |
| Optimized $SP_r$ | 3 | 4 | 4 | 4 | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |

The approach explained in this chapter was also used for two cases: a metrology workshop where the metrology tools are equal (same measurement throughputs) and a metrology workshop where the metrology tools are different (different measurement throughputs).

All parameters used in the approach for both cases are extracted from actual industrial data, the only data more difficult to obtain were the probabilities to failure, $p_r$, and the probabilities that the measures give bad results, $\alpha_r^t$. We consider for the calculation $p_r = 0.1$ for all process machines and $\alpha_r^t = 0.02$ for each metrology tool by process machine.

To obtain the optimized sampling rates, the last mathematical model is used, where seven heuristics are compared keeping the best value of Wafer Loss among them, because of it covers all types of metrology tools, as it is shown in section 4.4.3.

### 4.5.1 Identical metrology tools

The metrology workshop studied verifies the wafer properties in macro level (scratches on surface, particles, and so on), it consists on two metrology tools with a global average of measurement throughput for all process machines of $TM_r^1 = 298.89$ seconds and $TM_r^2 = 297.95$ seconds. Thus, they are considered as similar metrology tools. The metrology systems cover 26 process machines with different process throughputs, $TP_r$.

Table 4.17 shows the Wafer Loss values by heuristic. Heuristics DMT-$H_3$ and DMT-$H_6$ provide a value of 21.86, the best one. Both metrology tools are equally charged with a coverage of 13 process machines each one.

**Table 4.17 –** *Results on industrial data for similar metrology tools (Wafer Loss).*

| Heuristics | DMT-$H_1$ | DMT-$H_2$ | DMT-$H_3$ | DMT-$H_4$ | DMT-$H_5$ | DMT-$H_6$ | DMT-$H_7$ |
|---|---|---|---|---|---|---|---|
| Wafer Loss | 22.18 | 22.32 | **21.86** | 21.88 | 22.37 | **21.86** | 21.95 |

Table 4.18 presents the sampling rates by process machine. The optimized sampling rates are obtained through Heuristic DMT-$H_3$, and the actual sampling rates are the average of lots processed before sending a lot to the measure in a one month period. It should be noted that the sampling strategy used when the actual data was extracted is a sampling by product and process operation strategy, therefore, some process machines are uncontrolled due to the lots distribution like the case of machine 6 with an actual sampling rate of 31.23.

The optimized sampling rates proposed have an average of 6.11 and the average of the actual sampling rates a value of 7.04 (excluding machine 6). The optimized sampling rates are balanced with a rate of $SP_r = 6$, except for machines 1, 10, 17 and 25 with 7 because the ratio $\frac{TP_r}{TM_r^t}$ allows a more relaxed control of the lots processed on them. The opposite happens for machine 6 which needs a smaller sampling rate, $SP_r = 5$.

**Table 4.18 –** *Results on industrial data for similar metrology tools (Sampling rates).*

| Machines | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Actual $SP_r$ | 3.3 | 2.5 | 4 | 6 | 5.8 | 31.23 | 14.33 | 13.6 | 5.1 | 4.7 | 4.83 | 6 | 8 |
| Optimized $SP_r$ | **7** | **6** | **6** | **6** | **6** | **5** | **6** | **6** | **6** | **7** | **6** | **6** | **6** |
| Machines | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| Actual $SP_r$ | 4.9 | 4.9 | 6.8 | 6 | 6.7 | 6.4 | 10 | 9 | 17 | 6 | 8.5 | 5.4 | 6.2 |
| Optimized $SP_r$ | **6** | **6** | **6** | **7** | **6** | **6** | **6** | **6** | **6** | **6** | **6** | **7** | **6** |

## 4.5.2 Different metrology tools

The metrology workshop studied, that uses two different metrology tools, controls the crystallographic defects on wafer surface and dose uniformity after the ion implant operations. A metrology tool is faster than the other one, with a global average of measurement throughput for all process machines of $TM_r^1 = 419.16$ seconds and $TM_r^2 = 199.16$ seconds. Thus, they are considered as different metrology tools. The metrology tools cover 18 process machines with different process throughputs, $TP_r$.

Several Wafer Loss values are calculated by heuristic as it is shown int Table 4.19. Heuristic DMT-$H_7$, with 23.36, obtains the best Wafer Loss value. The metrology tool 1, which is the slowest one, covers 8 process machines and the tool 2, which is the fastest tool, covers 10 machines.

Table 4.20 presents the sampling rates by process machine for different metrology tools. The optimized sampling rates are obtained through Heuristic DMT-$H_7$. The actual sampling

**Table 4.19 –** *Results on industrial data for different metrology tools (Wafer Loss).*

| Heuristics | DMT-$H_1$ | DMT-$H_2$ | DMT-$H_3$ | DMT-$H_4$ | DMT-$H_5$ | DMT-$H_6$ | DMT-$H_7$ |
|---|---|---|---|---|---|---|---|
| Wafer Loss | 23.68 | 23.43 | 23.41 | 23.38 | 23.46 | 23.42 | **23.36** |

rates present two large values for machines 13 and 15, with an average of sampling rates of 52 and 146.2 respectively. These two machines are uncontrolled because of the sampling strategy by product and process operation that is currently being used.

The optimized sampling rates proposed have an average of 8.16 and, removing the values of machines 13 and 15, the average of the actual sampling rates is 9 . The optimized sampling rates make a better use of metrology capacity, a half of the machines present a sampling rate of $SP_r = 8$. Machines 10, 11 and 14 with a sampling rate of $SP_r = 7$ require a larger portion of metrology, $g_r^t(SP_r)$, thus must be more controlled. And machines 2, 3, 4, 6, 12 and 17 with $SP_r = 9$ are less critical.

**Table 4.20 –** *Results on industrial data for different metrology tools (Sampling rates).*

| Machines | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Actual $SP_r$ | 10 | 9.26 | 10 | 7.4 | 6.6 | 6.2 | 5.8 | 5.8 | 7.7 | 13.8 | 5.3 | 5.9 | 52 | 9 | 146.2 | 13.1 | 14.7 | 13.24 |
| Optimized $SP_r$ | **8** | **9** | **9** | **9** | **8** | **9** | **8** | **8** | **8** | **7** | **7** | **9** | **8** | **7** | **8** | **8** | **9** | **8** |

## 4.6 Industrial implementation

An industrial application, called Sampling Decision System, has been developed in order to calculate sampling rates to perform sampling strategy changes. This tool helped to get the current sampling rates used for the process machines covered by the existing metrology workshops in fab and to optimize them bringing the possible margin of improvement by using these new rates for the new sampling strategies. The tool has two main options, the optimization part and the historical data extraction.

In the optimization part, the user can select among all the metrology workshops to get the optimized sampling rates for all process machines associated to the metrology tools of the workshop. The application has integrated the algorithms explained in the mathematical models. The parameters needed such as $TP_r$ or $TM_r^t$ are regularly downloaded and collected into a folder, the application adapts the data into the correct format to perform the calculations.

Figure 4.1 shows a window of the Sampling Decision System where appears the optimized sampling rates for each process machine of a given metrology workshop.

In the historical data section, the user can extract a WAR profile by process machine associated to a metrology workshop with the current sampling rates used to verify if its

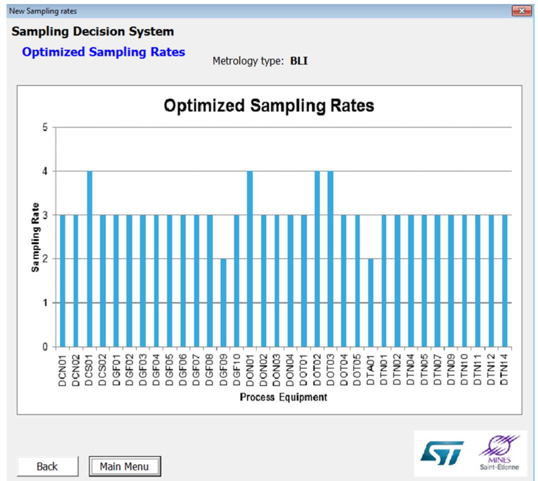



**Figure 4.1 –** *Screenshot of the optimized sampling rates of the Sampling Decision System.*

behavior is the expected one. Another option allows to see the global vision of the metrology workshop by plotting boxplot graphs by process machine type with the WAR values.

Figure 4.2 shows a WAR profile with the sampling values for a process machine of the historical data section and figure 4.3 the second option with the global vision of a metrology workshop.

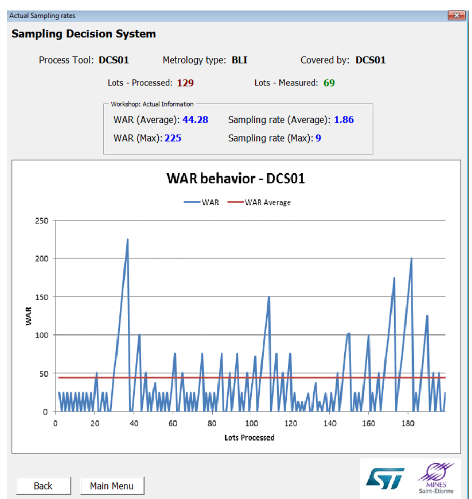



**Figure 4.2 –** *Screenshot of WAR profile for a process machine of the Sampling Decision System.*

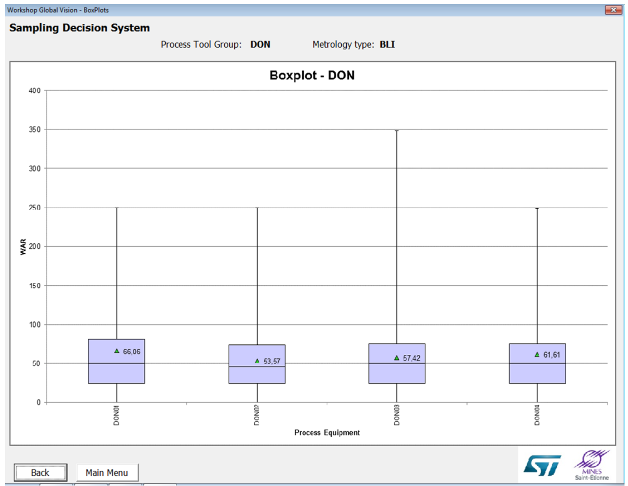



**Figure 4.3 –** *Screenshot showing the WAR global vision of the Sampling Decision System.*

The application has been developed with VBA. Figure 4.4 show the work flow of the application, the raw data with the process equipment and metrology tools information is regularly downloaded and collected into a folder where the application has access. The data is downloaded from the industrial database with the correct format and specification made by a previous database treatment. When the user selects an option of the application, the data is directly taken from the file and treated.

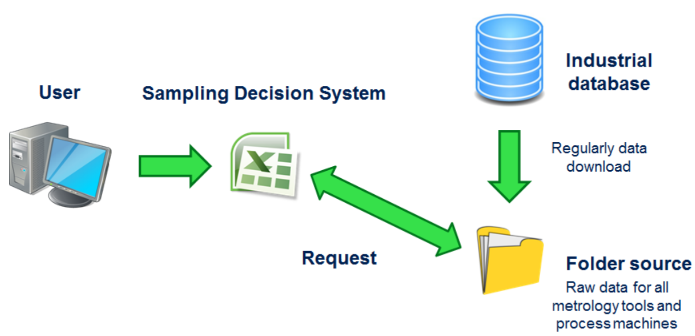



**Figure 4.4 –** *Work flow of the Sampling Decision System.*

## 4.7 Conclusion

This chapter presents a novel approach to optimize the sampling rates of heterogeneous process machines covered by several metrology tools with different specifications. Several mathematical models are proposed, considering different metrology coverage: a unique metrology tool, a group of identical metrology tools and different metrology tools. The logical path followed developing various models until build a robust method that considers all types of metrology tools is explained in this chapter. The method presented considers different heuristics which bring their own results and communicate between them to make their combination efficient.

The mathematical model 1 is modelled as a MCKP and various simple heuristics to implement without the use of optimization tools are proposed based on a property of the linear MCKP. This model is considered for using a unique metrology tool. Heuristic UMT-$H_1$ is based on rounding two fractional variables of this property, Heuristic UMT-$H_2$ is a greedy heuristic that extends the previous heuristic by iteratively rounding and Heuristic UMT-$H_3$ presents a difference in the rounding phase. The mathematical model 2 is also modelled as a MCKP problem and two heuristics are proposed. This model is focused on using a group of metrology tools with identical characteristics (same measurement throughput). Heuristic IMT-$H_1$ determines the sampling rates by using the previous Heuristic UMT-$H_1$ and the the process machines are assigned to the metrology tool with the largest capacity available. Heuristic IMT-$H_1^+$ is an improving phase of IMT-$H_1$ where a combination of Heuristic UMT-$H_2$ and Heuristic UMT-$H_3$ (Heuristic UMT-$H_2/H_3$) is used. The mathematical model 3 is modelled as a MCKP problem and as a DLP problem (Lagrangian Dual Problem). Seven heuristics are proposed. Heuristic DMT-$H_1$ solves the Lagrangian relaxation problem. Heuristics DMT-$H_2$, DMT-$H_3$ and DMT-$H_4$ optimize the sampling rates previously obtained from DMT-$H_1$ with different metrology tool assignment approaches and the Heuristic UMT-$H_2/H_3$. And heuristics DMT-$H_5$, DMT-$H_6$ and DMT-$H_7$ use in first place the metrology assignment done by DMT-$H_1$, then Heuristic 1 and then the same procedure than the previous three heuristics.

Numerical experiments were executed on randomly generated instances showing the efficiency of the heuristics of each mathematical model and the impact of the critical parameters. For the same instances, the mathematical model 3 provides better results in a 57.3% than the mathematical model 2 and the rest of the instances with the same value. Thus, the mathematical model 3 ensures to get the best results.

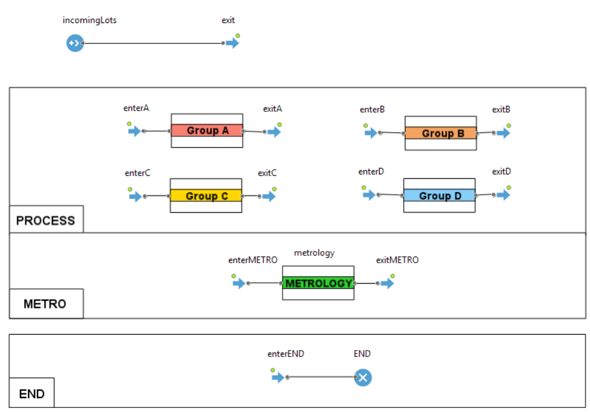
The approach was also validated on industrial data and the results are discussed. The balanced results and the better use of metrology capacity for the industrial instances led to develop an industrial application integrating the approach. The application allows to optimize the current sampling rates of the process machines covered by metrology workshop in fab and to extract current data in order to optimize the metrology usage and to get the possible improvement margin.

The optimized sampling rates got from the application are used for sampling strategy changes that are currently being performed for some metrology workshops in the company.

Chapter 5

# Dynamic risk management in Semiconductor Manufacturing

*This chapter presents and analyzes the results of simulation models implemented to study the behavior of some workshops in semiconductor manufacturing facilities (fabs) with the objective of controlling the risk on process equipment. Two case studies are considered: (1) Risk management in the Ion Implant workshop and (2) Sampling and dispatching policies in the thickness metrology workshop. The different components of the models are explained, and numerical results on industrial data are discussed.*



## 5.1 Introduction

The strategies of semiconductor manufacturing companies to improve their performance are based on performance measures such as yield, throughput and cycle time. To reach these objectives, control plans are used to supervise the behavior of process tools. Due to the complexity of this industry, a permanent review to check the effectiveness of control steps is necessary [5].

To manufacture an integrated circuit, wafers pass through numerous process steps such as Photolitography, Etching, Chemical deposition and so on. The first part of the chapter is focused on a particular step of semiconductor manufacturing: The Ion Implant workshop. The aim is to analyze the behavior of its process tools and the risk levels through a simulation model, and to propose and validate approaches to reduce the risk. In the second part of the chapter, for a given metrology workshop in charge of cover the thickness measurement, several dispatching policies (FIFO, LIFO, etc.) and sampling policies are compared in order to get the best strategy to achieve lower risk values. A sampling method to select the best lots to reduce risk and that provides gains in terms of cycle time and queue times is proposed.

Control methods are used to reduce risk. Different approaches can be found in the literature to manage risk. Risk analysis preventive tools such as Failure Mode and Effects analysis (FMEA) are used to identify potential failures of a process before using it and to evaluate its subsequent effects [50]. A prioritization of risk management measures using a risk-based approach is proposed in [89]. Khan and Haddara [37] introduce a method following the same approach but also taking into account a preventive maintenance plan for process equipment. The importance of optimizing control plans in semiconductor manufacturing is studied by Nduhura Munga et al. [57], they show that risk can be reduced without adding control measurement capacity.

This chapter is divided in two parts. The part related to the Ion Implant workshop is in Section 5.2 with the following structure: Section 5.2.1 defines the characteristics of the problem and Section 5.2.2 presents the simulation model. In Section 5.2.3, numerical results using industrial data help to analyze the current risk on process tools, and to compare with the case where additional products are qualified. The comparison of dispatching and sampling strategies is presented in Section 5.3. The different dispatching rules and sampling strategies used in this study are presented, and the key indicators to improve such as risk, cycle time and queue times are explained. The simulation model is detailed in Section 5.3.2. The numerical results are discussed in Section 5.3.3. Finally, conclusions and future perspectives are presented in Section 5.4.

## 5.2 Ion Implant workshop

### 5.2.1 Problem description

Ion Implantation in semiconductor manufacturing is a doping technique. Specific regions can be implanted with a precise control of doping levels modifying the conductivity of the semiconductor. The ion impacts alter the elemental composition of the wafer, otherwise each individual ion can cause point defects over the wafer surface. When a lot is measured to check that the process machine is not defective, the crystallographic damage is verified.

A route is the sequence of process operations required to obtain the final product. Along its route, each lot that usually contains 25 wafers enters several times into the Ion Implant workshop to receive different processes depending on the dose levels (Arsenic, Boron, Phosphorus, etc.), energy (from 2 to 3000 KeV) and implant angles. The workshop is divided in various process machine groups with different properties adapted to each possible treatment. There is a metrology area nearby where some lots are measured after being processed. Usually, a lot is only processed on one machine in the workshop before being measured or directly continuing to other workshops. In some cases, a lot may have to go through two or three consecutive Ion Implant operations. In these cases, the lot can always be measured after the first operation and sometimes after the following operations depending on the characteristics of the route of the lot.

Over the years numerous sampling techniques in semiconductor manufacturing have been developed [60]. An important characteristic to take into account for the workshop we are considering is that there is no Sampling Rate (defined as "$1/N$", i.e. one lot is measured after "$N$" lots are processed) to perform a measure. It will also be interesting to tackle this issue in the future because it has been shown that, if the selection of lots to measure is done dynamically according to risk levels, useless measures can be avoided as proposed by Dauzère-Pérès et al. [21]. In our study, the lots belonging to a measurable product are flagged before entering the workshop for the first time, and they are measured every time they visit the workshop just after they have been processed.

In this study, the risk is evaluated as the number of wafers processed on a process machine since the latest control performed for this machine. The indicator is called $W@R$ (Wafers at Risk) and corresponds to the possible loss in number of wafers in case of malfunction of the machine during the process. Let $NW(l)$ denotes the number of wafers in lot $l$, $W@R_m$ denotes the current wafers at risk of process machine $m$ ($W@R_m$ evolves dynamically) and $W@R_m(l)$ denotes the wafers at risk when lot $l$ is completed on $m$. Then, $W@R_m$ and $W@R_m(l)$ are updated as follows when lot $l$ is completed on $m$: $W@R_m = W@R_m + NW(l)$ and $W@R_m(l) = W@R_m$. If lot $l$ that was processed on process machine $m$ is measured, then $W@R_m$ is updated (i.e. decreased) by reducing its value by the number of wafer processed since the last measure (see Figure 5.1 as an example). For the Ion Implant workshop, a lot is measured just after having been processed (contrary to defectivity control considered for instance in Rodriguez-Verjan et al. [76]), hence $W@R_m := 0$ when a lot processed on $m$ is measured (see Figure 5.2 as an example).

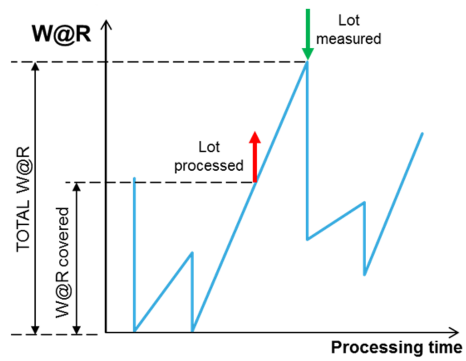



**Figure 5.1 –** *Wafer at risk (W@R) behavior for a process machine in the general case.*

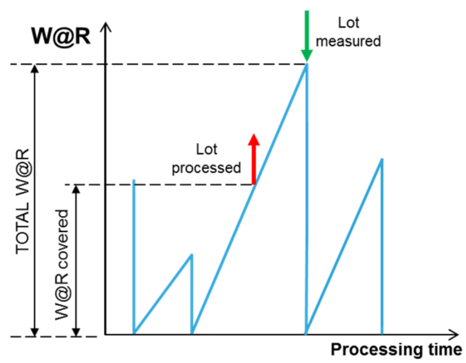



**Figure 5.2 –** *Wafer at risk (W@R) behavior for a process machine of Ion Implant workshop.*

This study is focused on building an Ion Implant simulation model with all the process and metrology tools based on the real behavior and data in a fab. The goal is to supervise the risk of every single process machine in terms of $W@R$ values. Figure 5.1 illustrates the main information obtained by representing a $W@R$ chart.

The $W@R$ corresponds to the number of wafers that could be potentially impacted if a problem occurs. As explained, the $W@R$ of a process machine increases each time a lot is processed on the machine, and it decreases when a measure of a lot processed on the machine is completed. Measuring a lot, which confirms that it was processed correctly on a machine, validates the quantity of wafers in the previous lots processed on the machine. In this case, the $W@R$ is reduced which means that the risk is decreased. The idea is to get dynamic $W@R$ values for lots processed and measured to better analyze how the workshop is operating. With a clearer view on how the real Ion Implant workshop is running, the following steps are: Propose new approaches to reduce the risk, introduce them into the simulation model and apply the effective approaches in the real workshop to change the current policies. Our goal is to later extend this work to other workshops.

### 5.2.2 Simulation model

The simulation model is divided into four parts: (1) Loading real data, (2) Model description explaining the flow of lots in the process machines and metrology tools, (3) Product qualification modeling, which is used to increase the number of lots to measure and (4) Model outputs, showing $W@R$ values and metrology tool indicators.

#### 5.2.2.1 Loading real data

It is essential for us to accurately simulate what happens in reality to validate potential improvements. To achieve this, industrial data are used as inputs to inject lots in the model. The data are taken from reports that contain what actually happened in the fab. A state chart manages this flow of lots. Lots processed and measured are registered in a table ordered by date. There, each row contains: The date when a lot was processed or measured, the machine which performed the operation, the lot ID, the operation, the number of wafers in the lot, the type of operation (process or measurement), the route name and the product name. When a simulation starts, the data are read and the lots enter in the model following the ordered sequence.

#### 5.2.2.2 Model description

The workshop has 17 process machines divided in 4 groups with different properties (Group A, Group B, Group C, Group D). Each group has its own behavior. Hence, if the $W@R$ values of machines of the same group are compared, they should be of the same order of magnitude, but process machines of different groups could be different. There are two metrology tools, Metro01 and Metro02, both with the same specifications except that one is faster.

The simulation model reproduces the arrival of the lots in the workshop, their paths across various process machines or metrology tools and the exit towards the next process steps. The lot information gathered in Section 5.2.2.1 is associated to the current lot injected with all its parameters by the source called "incomingLots”. As soon as a lot enters in the system, it is dispatched to the process area or the metrology area. In the case of the process area, there are four options according to the process machine. Figure 5.3 shows the structure of the simulation model.

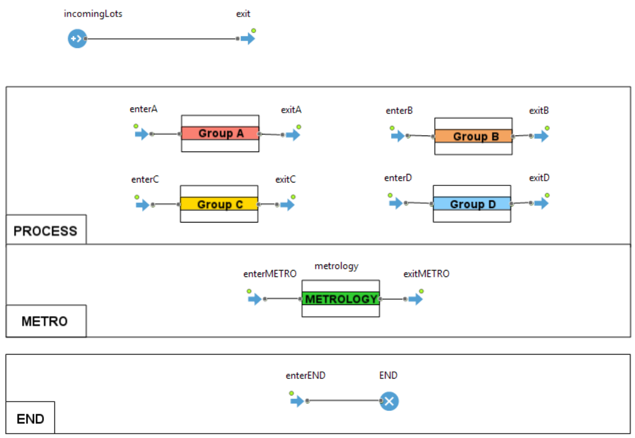



**Figure 5.3 –** *Ion Implant workshop structure.*

Once in a process machine group, a lot enters the machine fixed by its input parameter, and then two of its parameters are updated: Process machine and operation type. The purpose of refreshing these parameters is to confirm that the lot has been processed by the right machine. This is useful to evaluate future scheduling approaches, which could select another process machine instead of the one selected in the real data. After leaving the process machine, for instance machine *A01*, two counters for this particular machine increase: $W@R$ and $W@R_{max}$. The first one increases by the number of wafers in the lot, and the second counter only increases if the new $W@R$ value is larger than the current $W@R_{max}$. The metrology area operates equivalently. The lot is assigned to the fixed metrology tool and two parameters are overwritten with the selected metrology tool and to confirm that the operation type consists of a measure. Again, the idea of updating these two parameters is to propose and evaluate future approaches based on switching metrology tools. The configurations of process machine groups and metrology tools are shown in Figure 5.4.

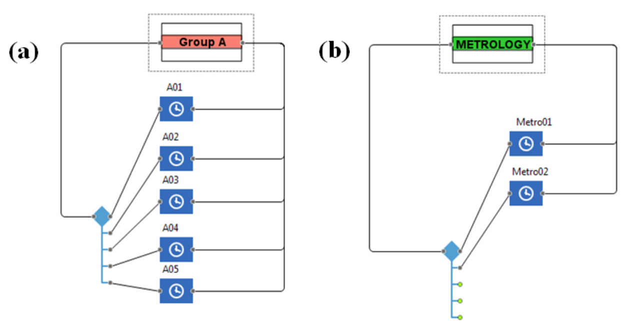



**Figure 5.4 –** *Process machine groups (a) and metrology tools (b).*

When a lot exits from a process machine group or from the metrology area,the counters of processed lots and measured lots increase. The $W@R$ value of a process machine decreases right after the measure of a lot that has been processed on this process machine. Hence, when a lot leaves the metrology area, the $W@R$ value from the machine that has performed the process operation is set to zero, otherwise the value of $W@R_{max}$ remains unchanged. During the simulation, there is a table that dynamically shows both values for every process machine.

#### 5.2.2.3 Product qualification modelling

Qualifying a product to be measured is highly time consuming for the engineers, since they must elaborate the recipe for the metrology equipment. Hence, a limited percentage of lots is available for measurement among all processed lots. Qualifying more products has been tested using the simulation model to evaluate the benefits of this strategy. Let us assume that it is possible to qualify any product. The lot, just after being injected in the workshop and before going to process or metrology, is added to a queue of a fictitious "qualification area". The qualification area must check if the incoming lot belongs to a qualified product and, in this case, the lot is flagged to be measured. The model of this area is shown in Figure 5.5.

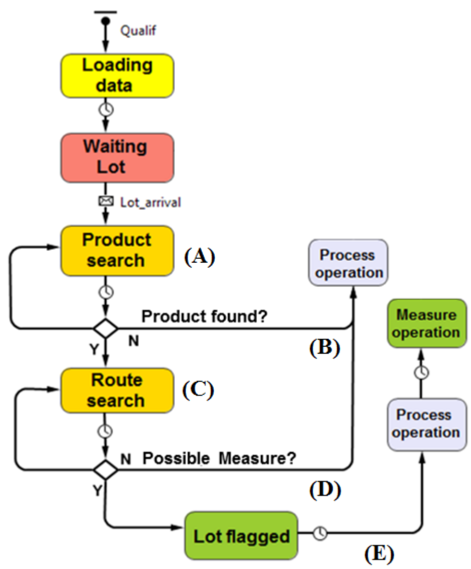



**Figure 5.5 –** *Product qualification diagram.*

There are three main characteristics in the simulation model associated with qualifying a lot: The product, the route and the process operation. As inputs, two tables are introduced. The first table is a list with the new products to qualify, and the second table includes all current routes in the fab with their respective Ion Implant process operations. The routes also indicate, for every operation, whether it is possible to perform a measure or not (measuring after some process operations is not allowed, in particular for data collection reasons that could create noise in the Statistical Process Control charts monitoring the workshop).

Every time a lot enters the product qualification area (see Figure 5.5), the first check (A) consists in verifying whether the lot belongs to a product that is qualified to be measured, otherwise, the lot is skipped (B) and is only processed. Then, the second check (C) consists in verifying whether the lot is at an operation in the route that allows a measure to be performed. If this is not the case, the lot is skipped (D) and is only processed.

As a final stage, if the operation can be measured (E), the lot is flagged and, after being processed, is sent to metrology. There, metrology tools now have a queue, and are configured to perform the measure in 7 and 5 minutes for Metro01 and Metro02, respectively. The queue management aims at measuring lots as quickly as possible. So, if both metrology tools are available, the first option is Metro02 because it is the fastest tool. But, when both tools are busy and there are lots waiting to be measured, the queue management software calculates to which metrology tool the lot will be first assigned.

#### 5.2.2.4 Model outputs

Our objective is manage the risk in terms of $W@R$ for every process machine. Having relevant data and output charts to perform the analysis are key points of the simulation model. Thanks to the charts, we can detect whether a process machine is over or under controlled. Each process machine $m$ has two parameters that are updated dynamically: $W@R_m$ (as already discussed in Section 5.2.1) and $W@R^m_{max}$. The updating of $W@R_m$ is discussed in Section 5.2.1 and, once $W@R_m$ is updated, $W@R^m_{max}$ is updated as follows: $W@R^m_{max} := \max(W@R^m_{max}, W@R_m)$, i.e. $W@R^m_{max}$ only increases over time.

Each time a lot finishes at a process machine group or at the metrology area, a new row is registered in a table that is generated as an output at the end of simulation. This new row comprises: The time when the lot was processed or measured (in seconds), the lot ID, the product name, the number of wafers, the operation type (process or measure) and the process machine or metrology tool. The $W@R$ parameters are also updated.

To sum up, while the simulation model is running, we can see how the $W@R$ values of process machines vary. When the simulation ends, two $W@R$ tables are created: one with the $W@R$ values of lots when they are completed on their process machines, i.e. $W@R_m(l)$ for each lot $l$ processed on machine $m$, and another table with the value of $W@R^m_{max}$ for each process machine $m$. Also, the set $ML(m)$ of measured lots for process machine $m$ is stored.

### 5.2.3 Numerical experiments

The numerical results in this section are obtained through the proposed simulation model that was implemented using the Anylogic software and with two weeks of real data from STMicroelectronics in Rousset, France. The data include around 21000 process operations and 1900 measures of lots. Two indicators for each process machine $m$ are used: $W@R^m_{max}$ (as discussed in Section 5.2.2.4) and $W@R^m_{Average}$, which is the average of $W@R$ values attained just before a lot processed on $m$ was measured, i.e. the average of $W@R$ values of lots in the set $ML(m) : W@R^m_{Average} = \sum_{l \in ML(m)} \frac{W@R_m(l)}{|ML(m)|}$.

First, the current risk on process machines is presented in Sections 5.2.3.1 and 5.2.3.2. Then, Sections 5.2.3.3 and 5.2.3.4 show the possible improvements when a qualification strategy is followed.

#### 5.2.3.1 System description

One of the constraints of the Ion Implant workshop is that not all lots can be measured to reduce the risk. As already mentioned in Section 5.2.2.3, this is due to the difficulty for engineers to qualify a new product on metrology tools. This is why there are often no lots qualified to be measured that are available, and process machines reach large $W@R$ values until a measure is performed to reduce the risk.

Table 5.1 shows the percentage of lots processed during two weeks in Ion Implant. It is divided into 6 products that are qualified to perform a measure and the other products. The impact of lots that are not covered for the metrology tools is 75.5%, and only 24.5% of lots are candidates to be measured. This leads to long periods without controlling process machines.

**Table 5.1 –** *Percentage of processed lots per product (only products 1 to 6 can be measured).*

| Products | Processed lots (%) |
|---|---|
| Product 1 | 7.8% |
| Product 2 | 3.1% |
| Product 3 | 0.8% |
| Product 4 | 5.9% |
| Product 5 | 5.8% |
| Product 6 | 1.1% |
| Other products | 75.5% |
| Total | 100% |

#### 5.2.3.2 Current situation

The $W@R$ charts obtained after running the simulation model help to quickly find for which process machines we need to take action in order to improve their risk level, see for example Figure 5.6 where the behavior of two machines belonging to group A is shown. In the graphs, every time a lot is processed, the $W@R$ of the corresponding process machine is increasing, and the peaks show the maximum $W@R$ values that are attained before a measurement is performed. For confidentiality reasons, the axes have been normalized.

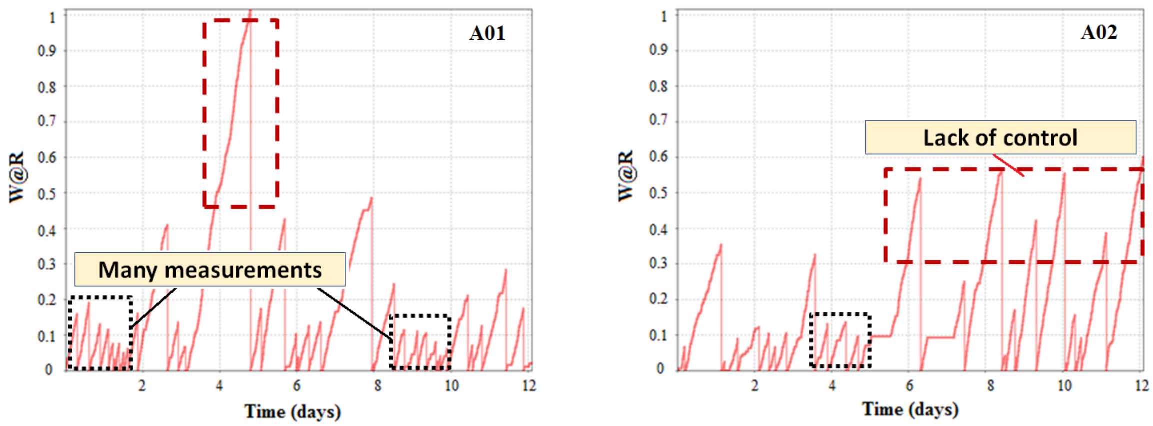



**Figure 5.6 –** *W@R evolution for process machines A01 and A02 for real data.*

Normally, there are some areas with lack of control. The goal is to reduce the maximum *W@R* values to remain below the *W@R* limits and even to establish lower *W@R* limits. Looking at Figure 5.6 and a limit of 0.5, the *W@R* value exceeds the limit and a control operation has to be performed as soon as possible. When analyzing the measurement effectiveness for process machine *A01*, too many measures are sometimes performed, for instance before the beginning of the second day and between days 8 and 10. It is a waste of measurement capacity since control operations do not bring enough added value.

At this stage, we want to provide a general view of the risk in the workshop and show that it is possible to avoid reaching large *W@R* values. As future perspectives, we want to develop a dynamic system to only select the right lots to measure and to avoid useless measurements.

### 5.2.3.3 Qualifying a single product

First, finding the best products to qualify is needed. With a treatment of the data obtained by running the simulation model on real data, a list is elaborated that takes all lots processed of products not covered into account and the routes with Ion Implant operations that can be measured. In this first study, the products that bring the largest number of additional measures are considered. Taking more aspects into account is considered as a future perspective.

The first six products are selected and each qualification has been simulated separately. Figure 5.7 shows the *W@R* evolution for process machines *A01* and *C04* when products 7 or 12 are qualified respectively. These two process machines have experimented the most remarkable *W@R* reductions compared to the other machines. Also, qualifying one of these two products provide the largest *W@R* reductions. The figure shows that, after qualification, most of the high peaks are reduced. For example, for machine *A01*, the *W@R* between day 2 and day 5 is better distributed than before and, for machine *C04*, it has considerably decreased between day 10 and day 12.

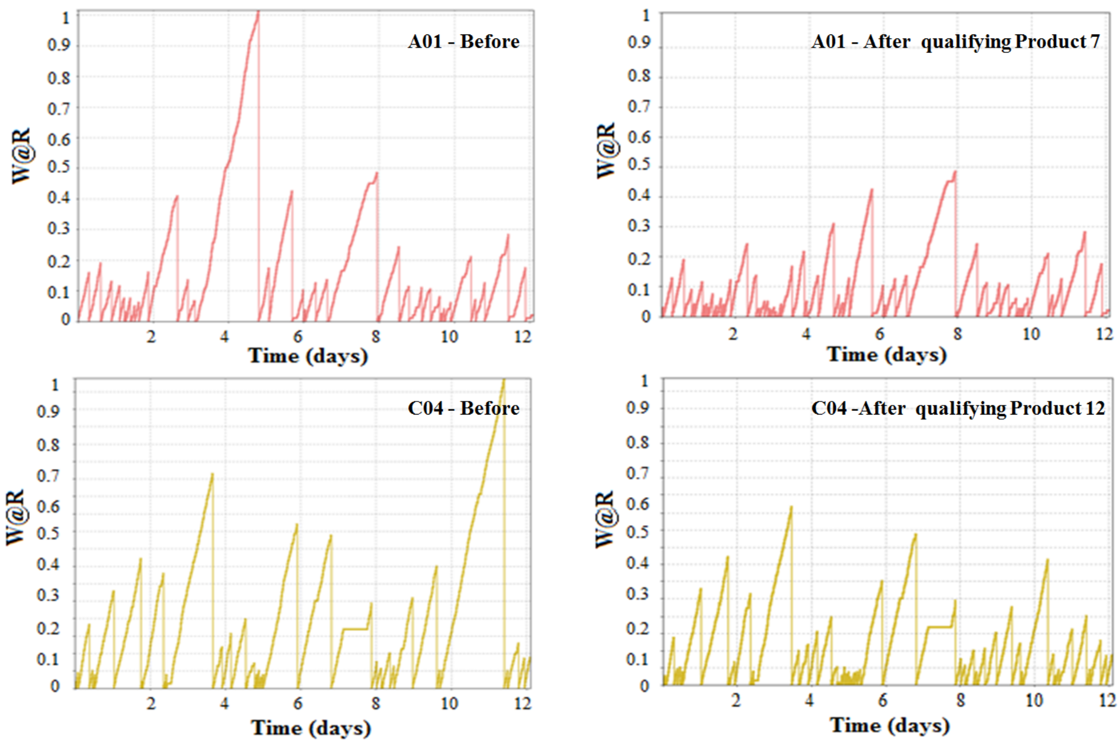



**Figure 5.7 –** *W@R for machines A01 and C04 before and after qualifying Products 7 and 12, respectively.*

Qualifying new products offers opportunities to reduce the risk, but selecting the product that induces the most additional measures may not be the best alternative. Qualifying other products could be more interesting if their lots are processed on machines that are at risk. Looking at Table 5.2, it is interesting to see that, when qualifying Product 12 and although it has less measured lots than the other products, larger reductions for $W@R_{max}$ of the process machines are obtained compared to the other products (except Product 7). For $W@R_{average}$, Product 12 has only 1% of difference when compared to Product 8, which is initially the second best option.

**Table 5.2 –** *Difference of global W@R reduction depending on the qualified product.*

| | Products | | | | | |
|---|---|---|---|---|---|---|
| Indicator | Product 7 | Product 8 | Product 9 | Product 10 | Product 11 | Product 12 |
| **Gain - $W@R_{max}$** | 20% | 18% | 16% | 14% | 13% | 19% |
| **Gain - $W@R_{average}$** | 29% | 22% | 22% | 21% | 20% | 21% |

When comparing the $W@R$ evolution of all process machines between the real case and qualifying Product 7 in Figure 5.8, it can be seen that many machines reach the same $W@R_{max}$ values without any changes; e.g. process machines *A02*, *B04* and *C03*. However,

$W@R_{average}$ is generally considerably reduced. It is interesting to see that, for process machines *A02* and *C03*, the same $W@R_{max}$ value is attained before and after qualifying products: $W@R_{max}(A02) = 0.55$ and $W@R_{max}(C03) = 0.24$. The behavior of their $W@R_{average}$ is different, since $W@R_{average}$ of process machine *A02* is almost divided by 2 (from 1 to 0.55), while $W@R_{average}$ of process machine *C03* practically remains the same (from 0.19 to 0.16). In Figure 8, because the values are normalized, it is not possible to directly compare chart (a) and chart (b). Also, by definition, $W@R_{average}$ is always smaller than or equal to $W@R_{max}$.

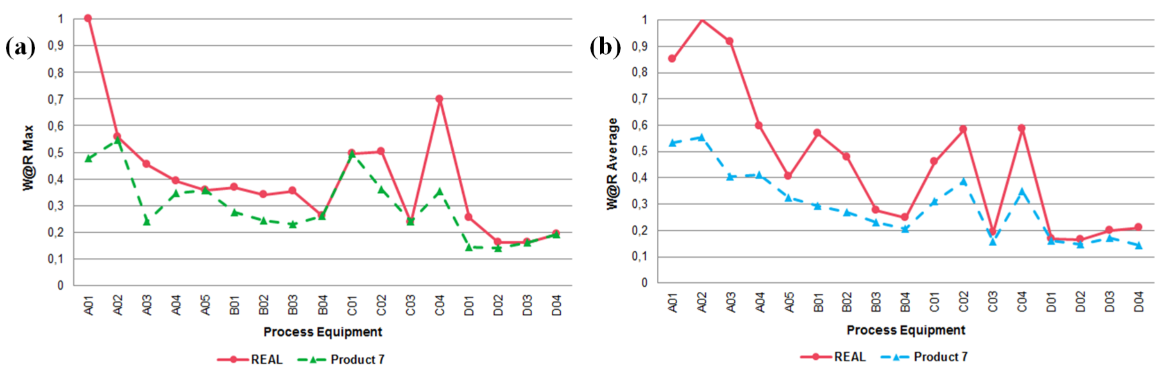



**Figure 5.8 –** *Comparison of $W@R_{max}$ (**a**) and $W@R_{average}$ (**b**) between real case and if Product 7 is qualified.*

Hence, as it is shown through our simulations, by qualifying products, the risk is decreased but not all process machines are covered for all the cases, since even if the values of $W@R_{average}$ are reduced, large maximum values are still reached. The next step is to qualify two products to reduce the risk even more.

### 5.2.3.4 Qualifying two products

As a starting point, only the combination of two products is considered due to the complexity for the workshop to qualify products. To select products, the goal is to cover the largest number of process machines and to decrease the $W@R$ values as much as possible. The best option is to select products 8 and 12 to qualify since it is the combination that leads to the largest $W@R$ reduction for the 17 process machines. A comparison between the real case, qualifying product 8 and qualifying both products 8 and 12 is shown in Figure 5.9. The reduction is sometimes quite large, for example for process machine *A03* (from 0.78 to 0.53) and, in other cases, less significant, for example for process machine *D02* (from 0.14 to 0.12). This combination also covers process machines that were not improved with only one qualified product and leads to improvement on process machine *A05* (from 0.36 to 0.28) and process machine *B01* (from 0.37 to 0.18).

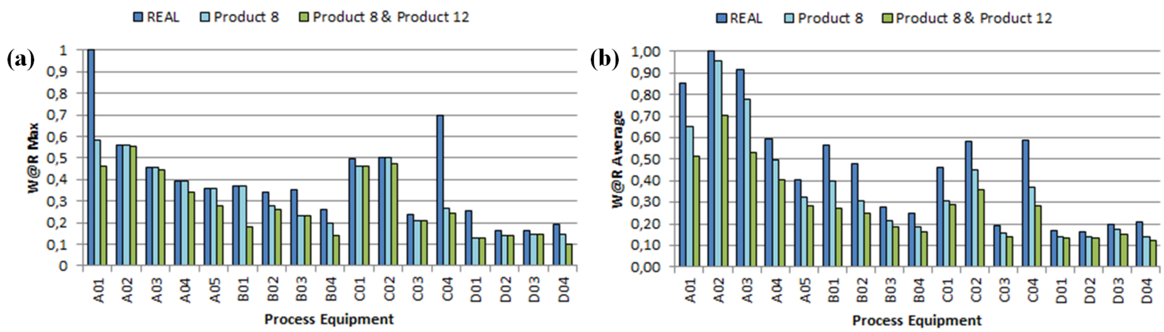



**Figure 5.9 –** *Comparison with qualifying products 8 and 12 for $W@R_{max}$ (**a**) and $W@R_{average}$ (**b**).*

Due to the complexity for engineers to perform new qualifications, the goal is to qualify a little products as possible to cover the largest number of process machines and to reduce the risk much as possible. Qualifying more than two products could be considered if it brings significant gains in terms of $W@R$ reduction.

## 5.3 Comparing sampling and dispatching policies

A group of process machines covered by a metrology workshop can reach large risk values depending on how lots are sampled and dispatched on metrology tools. In this section, several methods are studied to sample lots after they are completed on a group of process machines, and to dispatch lots on metrology tools that measure the film thickness of the wafers.

### 5.3.1 Problem description

The metrology workshop under study measures an average of 8451 lots per week, approximately 22% of total measures in the fab. It is composed of 20 metrology tools that measure lots processed by 23 different process machine groups.

There are particular machines that do not work as a mainframe (a mainframe includes multiple chambers), but as an independent chamber, which means that every chamber performs different process operations. Sometimes, lots are processed in certain chambers and not necessarily in all of them. This is why every chamber is considered as an independent machine. In total, there are 259 process machines to control. Machines that belong to the studied process groups but are only controlled few times per week or months have been discarded of this study.

As in Section 5.2, the risk is managed through the $W@R$ indicator, considering the current wafer at risk value for a given process machine $m$ when a new lot $l$ with $NW(l)$ wafers is processed as: $W@R_m = W@R_m + NW(l)$.

To update a $W@R$ value for a machine when a lot is measured, a permanent indicator has been used as in previous research [56] [58].

A permanent indicator for a process machine $m$, $PI_m$, increases by the number of wafers in lot $l$ processed by $m$:

$$PI_m = PI_m + NW(l) \tag{5.1}$$

Once a lot $l$ is processed on a process machine $m$, the current permanent index $PI_m$ of the lot $l$ for $m$ is associated:

$$PI_m(l) = PI_m \tag{5.2}$$

When a lot $l$ processed on $m$ is measured, the risk associated to $m$ will be updated by reducing the risk that corresponds to the lot by means of the permanent indicator:

$$W@R_m = \min(W@R_m, PI_m - PI_m(l)) \tag{5.3}$$

The main goal of this study is to gather several parameters for the normal behavior of the metrology workshop in terms of queue times of lots and risk values of process machines and then compare them with new simple strategies for dispatching lots on metrology tools and also new sampling techniques prioritizing the machines with highest risk.

New dispatching rules and sampling policies are proposed:

- Normal behavior with skipping policy: The current situation is respected and only lots that do not provide a risk reduction are skipped.
- FIFO (First In, First Out): The oldest lots in the metrology queue are taken first to be measured.
- LIFO (Last In, First Out): The most recently processed lots in the metrology queue are taken first to be measured.
- Skipping sampling policy (SSP): The oldest lots in the metrology queue that do not provide a risk reduction are skipped.
- Skipping and priority sampling policy (SPSP): The lots are skipped as in SSP, and the first lot to measure is the one that has been processed on the riskiest process machine.

### 5.3.2 Simulation model

The simulation model is composed of these parts:

1. **Loading data**: This part manages the generation of lots in the model.
2. **Process equipment area**: Lots are sent to this area to be processed by the same machine than the reality. $W@R$ counters by machine are updated after processing lots.
3. **Lot transport and waiting times**: The time spent by a lot from the process step to the measure step is simulated through this part of the simulation model.
4. **Metrology area**: This area contains the metrology queues, the metrology tools and the different approaches proposed to simulate the new sampling strategies.
5. **Parameters update and model outputs**: Before lots leave the metrology workshop, some parameters are updated and gathered for analysis.

One parameter of the simulation model allows the running mode to be selected: Normal, Normal with skipping option, FIFO, LIFO, SSP or SPSP.

#### 5.3.2.1 Loading data

The simulation model uses industrial data as inputs and creates lots following the actual historical order. The generated lots enter into the different parts of the model depending on whether it is a process operation or measure operation. About 5 days of data are used.

A data table, ordered by date, contains in every row the information of each lot: Number of wafers, the process machine that performed the process operation, the lot ID, type of operation (process or measurement) and the actual processing order of the lot. The processing order indicates the exact order in which the lot entered in the processing part of the model.

This table is pre-processed, and the processed lots that in the reality were later on measured are marked as flagged. This is for taking as candidates to measure using the other modes (FIFO, LIFO and Sampling) just the lots that were really measured.

Each row of the table includes the time that the next lot will be introduced into the model (the time difference of the dates between the actual row and the next one). When a simulation run starts, the data are read through a flowchart starting with the first row.

Figure 5.10 shows the flowchart that manages lots entering in the simulation model.

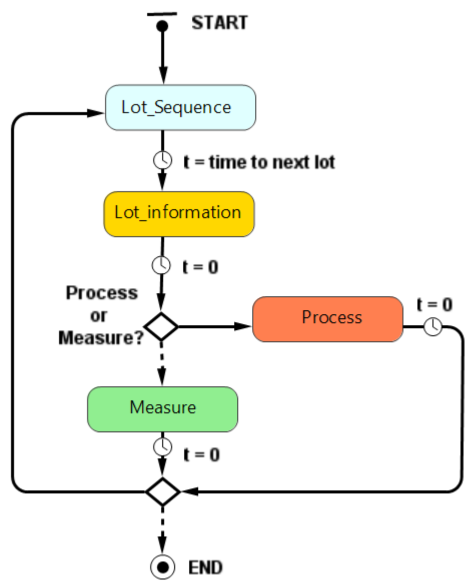



**Figure 5.10 –** *Loading data in the simulation model.*

Measured lots (the information rows of measurement operations contain the time when the lots were measured) in the data input are only considered in the normal mode to send lots to the metrology area in order to calculate the indicators associated to what happened in reality.

For the other modes, only the lots processed (the information rows of process operations) are considered, and only the flagged lots will continue their path after the process operation to enter the metrology area.

Figure 5.11 shows how the simulation model distributes lots depending on the operation type and the selected mode.

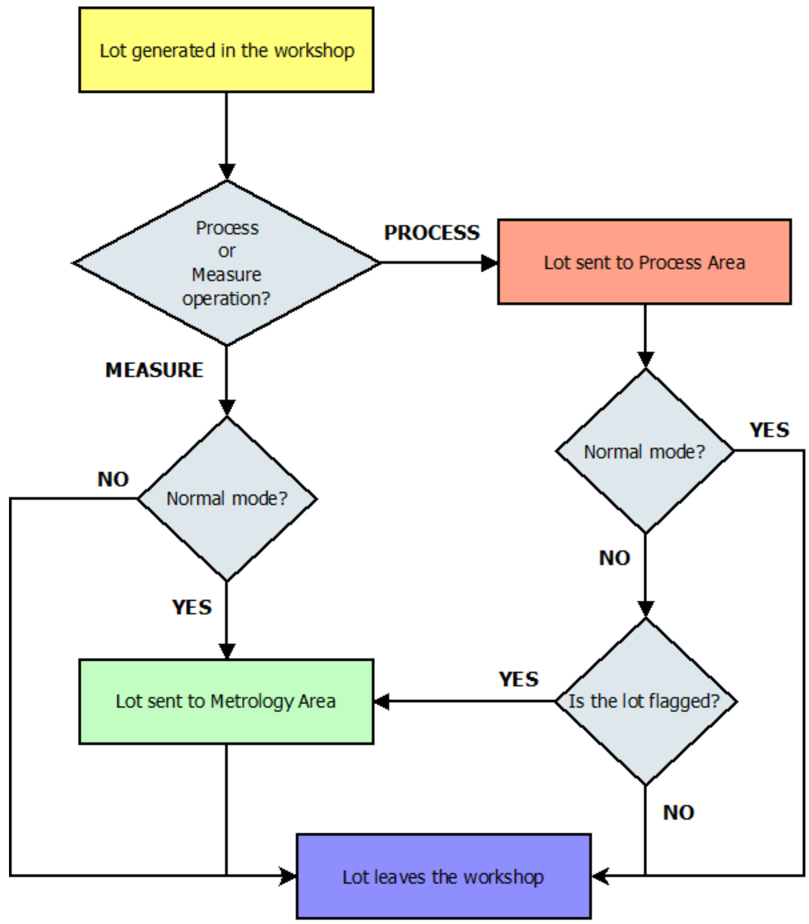



**Figure 5.11 –** *Lot distribution in the simulation model.*

#### 5.3.2.2 Process equipment area

The metrology workshop covers the risk of 259 process machines distributed in 23 process equipment groups. Once a lot is created in the simulation, it is sent to the corresponding group that performed the process operation just when the operation is finished.

Figure 5.12 shows the process equipment groups in the main section of the simulation model, where lots will enter according to the process machine.

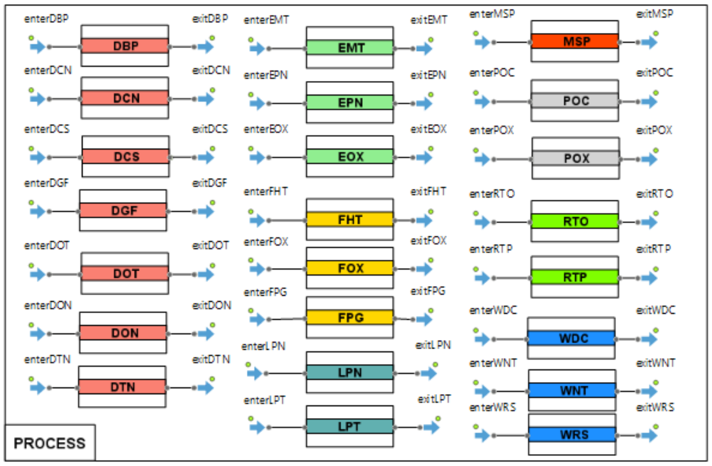



**Figure 5.12 –** *Process equipment area.*

In every process equipment group, see Figure 5.13, the lot enters in the associated process machine and the $W@R$ indicator of the process machine will increase with the number of processed wafers. The process machines have another indicator called LPM, that registers the processing order of the last measured lot that was processed. This indicator is used by the metrology area. Finally, the information of the lots processed by every group is gathered in a collection called "Lots" and, after the lot leaves the process equipment area, a new row of the output data is registered.

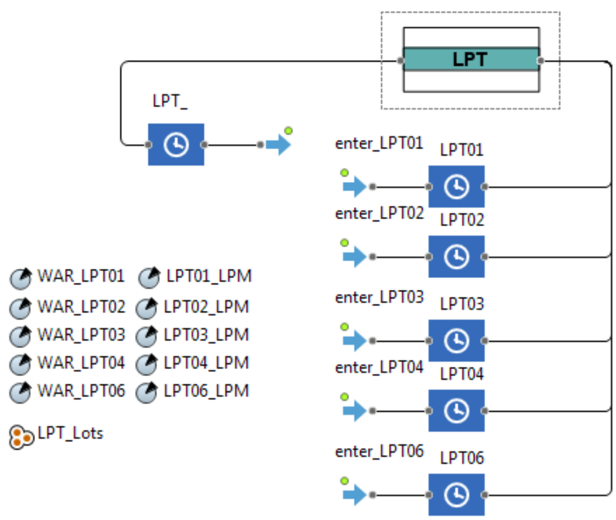



**Figure 5.13 –** *Process equipment group distribution.*

#### 5.3.2.3 Lot transport and waiting times

This part of the simulation model is in charge of sending the flagged lots from the process area to the metrology area. It is reserved only for the FIFO, LIFO and Sampling modes, since the normal mode directly distributes lots to the metrology area at the exact time the lots were measured. Figure 5.14 shows a screenshot of the simulation model with the transportation part.

The time spent to move a lot from a machine to the racks of the metrology area, taking into account the machine allocation in the fab and their proximity to the metrology tools, follows a triangular distribution.

The time spent by lots waiting in the machine after the process operation is also considered. Lots will be held during a certain time before leaving the process equipment area and being transported to the metrology area.

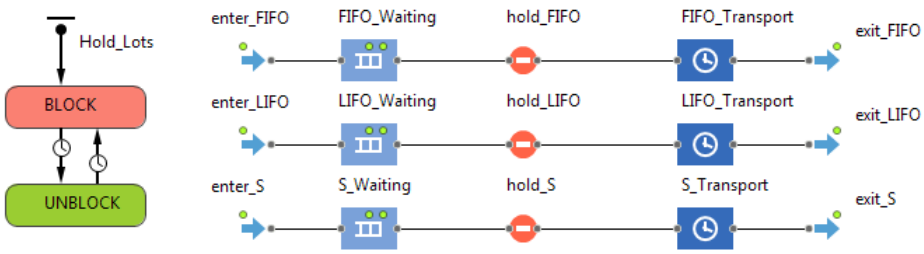



**Figure 5.14 –** *Lot transport.*

#### 5.3.2.4 Metrology area: Sampling and dispatching policies

The metrology area of the simulation model manages the lots that have to be measured, and that depends on the selected mode. A **collection** gathers all the information on the lots that enter in this area (the machine which processed the lot, number of wafers, processing order, and so on). Two input files are needed to manage lots:

1. **Metrology qualifications**: This table provides by process machine and process operation, the metrology tools that are qualified to perform the measure.

2. **Measurement times**: Every metrology tool takes a different measurement time depending on the machine that processed the lot, this table contains all measurement times by metrology tool according to the process machine.

The assignment system of the metrology tool and the measurement time for lots are shown in Figure 5.15. When a lot arrives at this element of the path, the system is triggered and the lot waits until the system finishes, then the lot continues its path. The dynamic flowchart waits until a new lot enters in the metrology area (1) to start the procedure. The input files are read by going over the rows of the tables by using row indicators that at the end are reset to zero (8). When the simulation model is ran in normal mode, the metrology tool is already selected (because the lot will go to the metrology tool which was used in reality), thus the system will jump to the measure time assignment (5). For other modes, in the first place, if the lot skips the measure, the lot will be marked to avoid the measure (2), otherwise a search will start (3) through the input file of metrology qualifications to find the process machine in the data table. In (4), the availability of all metrology tools qualified to measure the lot performed on the given process machine is checked, and the metrology tool which is available is assigned to the lot.

After assigning the metrology tool, the measurement time search starts (5). Again, the system goes over the rows of the second input table to find the row corresponding to the process machine and metrology tool. If a metrology tool is not found in the table (that could happen if this file is not updated), a fixed measurement time of 300 seconds (the measurement time average of all metrology tools) is assigned to the lot in (6). As a final step, the measurement time (the time that the lot will spend in the metrology tool) is assigned to the lot in (7).

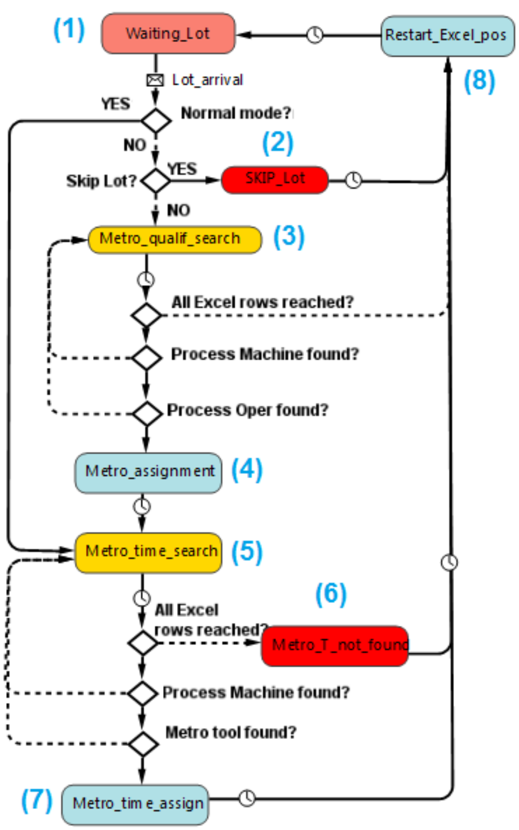



**Figure 5.15 –** *Assignment system of metrology tools and measurement times.*

The metrology area behaves differently according to the selected simulation mode, each one has its own structure that provides different results at the end of the simulation.

In the **normal mode**, see Figure 5.16, lots go through a skipping step, then another step which triggers the assignment system (lots wait few milliseconds while the assignment flowchart finishes the procedure) and finally lots are sent to the assigned metrology tools. For this mode, even if the assignment system finds the measurement time for the metrology tool, lots will not spend this time inside the metrology tool because actual end times of measurement operations are used. Anyway, this measurement time is kept inside lots to do the parameter calculations later.

Activating a skipping option is also possible. This option allows the evaluation of improvements that would be obtained if some skipping rules were added to the current running of the metrology workshop. If the skipping option is activated, every time a lot enters the skipping step, the LPM (the processing order of the last measured lot) associated with the same process machine of the current lot is searched in the metrology collection. If the LPM is larger than the processing order of the current lot, the lot will skipped (for example, if the current processing order is 10 and the LPM of the process machine is 15), otherwise it will be measured. The lots are skipped according to the processing order because in terms of risk it makes no sense to measure an older lot when a more recent lot was already measured.

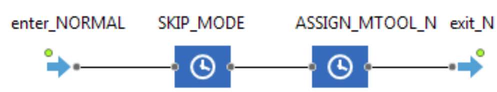



**Figure 5.16 –** *Normal mode structure in the metrology area.*

For the other modes of the simulation model, the metrology queue plays an important role. A **holding system** keeps lots in the metrology queue until they are dispatched by an operator. The frequency at which operators go to the metrology racks to pick up lots and put them on metrology tools, is governed by a holding time and the capacity levels of the metrology queue. Figure 5.17 shows the holding system that blocks and unblocks the path of the lots waiting in the metrology queue.

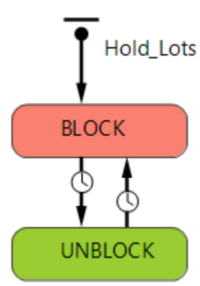



**Figure 5.17 –** *Holding system in the metrology area.*

The following modes manage the lots that are waiting in the metrology queue in different ways. The transport time from the queue to the metrology tools is the same for each mode and the number of operators as well. In contrast to the normal mode, lots will stay in the metrology tool the assigned measurement time. The developed modes are divided in two groups of different policies: Sampling and Dispatching.

Two sampling policies were developed:

1. **Skipping sampling policy (SSP)** considers a skipping policy in the queue. As soon as lots are arriving in the metrology queue, the skipping algorithm is executed and those lots that coincide with another more recent lot of the same process machine will be skipped. Lots, in the same order that they arrived, will be transported to the assigned metrology tools or skipped.

2. **Skipping and priority sampling policy (SPSP)** combines a skipping policy and a priority ordering policy. Firstly, lots that must be skipped are found. Secondly, after the holding system, the priority ordering system is executed in a new queue. Lots that are skipped leave the metrology area. The remaining lots are ordered from the one that will reduce the risk the most to the one that will reduce the risk the least, to be transported and sent to the metrology tools.

Figure 5.18 shows at the top of the picture the path followed by lots in SSP and at the bottom of the picture the path for SPSP.

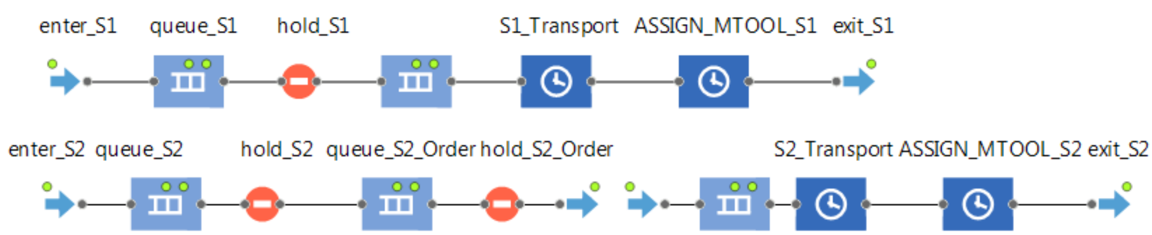



**Figure 5.18 –** *Sampling elements of the simulation model for SSP (above) and SPSP (below).*

For SSP and SPSP, in order to select the lots that provides the best results in terms of risk reduction and capacity gain, two policies were developed in the simulation model.

A **Skipping policy** is integrated in the metrology queue that is applied every time a lot is placed in the queue. The objective is to skip lots from the queue to avoid the measurement operation and to leave the workshop by only measuring the lots that were processed the most recently by the process machines. Figure 5.19 illustrates a comparison between a normal lot order and the skipping policy.

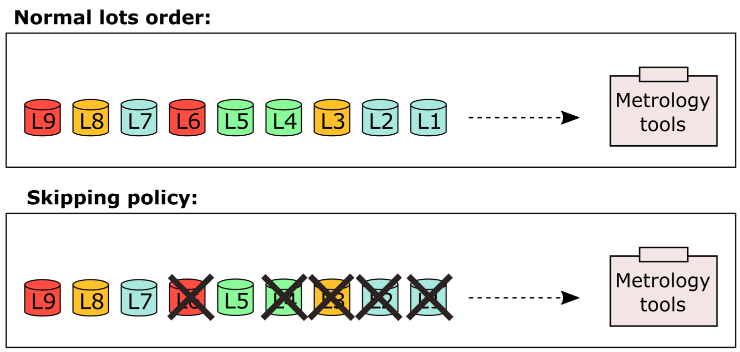



**Figure 5.19 –** *Skipping policy of the sampling modes.*

The policy in this section of the simulation model compares the newly arrived lot with the other ones in the queue. A boolean attribute called "SKIP" for every lot that is by default *false* is created. If the lot should be skipped the attribute is changed to *true*.

The second policy is a **priority ordering** of lots. This policy reorders all lots waiting in the metrology queue. Firstly, lots that were previously skipped are released. Secondly, the remaining lots are reordered by decreasing order of risk (from the largest to the lowest). Figure 5.20 presents an example of the priority risk ordering.

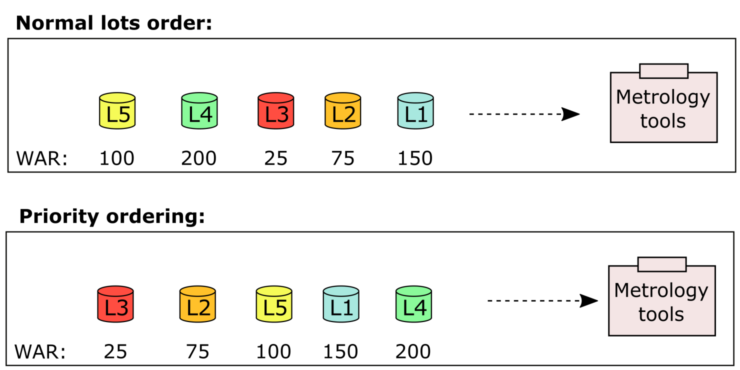



**Figure 5.20 –** *Priority ordering policy of the sampling modes.*

The policy starts by removing all the lots to skip, and reorder the remaining lots by their $W@R$ value. The size of the metrology queue changes until all lots to be skipped are definitely skipped.

The dispatching modes follow different policies:

1. **FIFO policy:** Lots are stored in the metrology queue as they arrive from the process step. When the holding system releases the path, lots leave the queue following the arrival order, i.e. the first lot to arrive is the first lot to leave the queue. The transportation time to the metrology tool is taken into account for every lot. The assignment system associates to every lot the metrology tool and the measurement time, and the lot is finally sent to the corresponding metrology tool. The path followed by lots is presented in Figure 5.21.

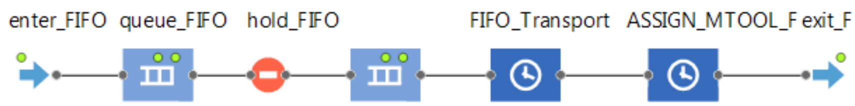



**Figure 5.21 –** *FIFO mode in the metrology area.*

2. **LIFO policy:** This mode is similar to FIFO, except that, when the holding system unblocks the path, the last lot to arrive in the metrology queue is the first lot to leave. Figure 5.22 shows the LIFO path for lots.

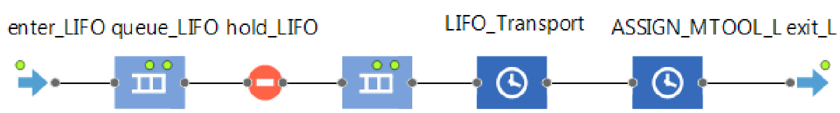



**Figure 5.22 –** *LIFO mode in the metrology area.*

### 5.3.2.5 Parameter update and model outputs

At the end of the simulation, three data tables are generated as inputs, the two first tables provide the activity of lots; one registers the process and measurement operations and the other table only the measurement operations. A new row is added each time a lot is processed or measured. The information of lots is ordered by time in each row. They are the same fields than the input file (lot ID, quantity of wafers, etc.) and also new parameters are added: The $W@R$ value associated of lot, the transport time, the measurement time and the queue time.

The third data table registers global parameters of the whole metrology workshop for the simulation mode selected by each day of the input data and the mean of the total time period.

These global parameters are calculated while the simulation model is running and are the following:

- The **W@R values** of process machines are updated when a new lot is processed by a machine, increasing its risk. $W@R_{Average}$ sums the $W@R$ value of all process machines during the simulation, and $W@R_{Average}^{machine}$ corresponds to $W@R_{Average}$ divided by the number of machines. The $W@R_{Average}^{Max}$ indicator provides the average of the $W@R$ values reached when the measures are performed. If the current $W@R$ value is 0 when a new measure is done, the $W@R_{Average}^{Max}$ indicator is not updated.
- **The queue time (QT)** spent by a lot is calculated right after its measure. The model runs during a simulated time and to calculate $QT$, several time instants are registered: When a lot was processed by the machine and when the same lot finishes its measurement operation. The formula used to calculate QT is (5.4).

$$QT = Time(\text{current}) - [Time(\text{Measure done}) + Time(\text{Process end})] \tag{5.4}$$

An average of $QT$ for all lots is calculated, and a limit called $QT^{Max}$ is considered to discard high $QT$ values related to exceptions.

- The number of **useless measurements** in terms of risk is counted. This is when a lot processed by a machine is measured and there is no risk reduction. There is a comparison between the processing order of the current lot and the LPM indicator of the machine that processed the lot.

## 5.3.3 Numerical experiments

### 5.3.3.1 Setting parameters

In addition to the input data that include the table of processed and measured lots, some parameters must be set to simulate the behavior of lots in the workshop. This simulation model is considered as non-deterministic, and some stochastic parameters are defined by probability distributions.

Two probability distributions are used. The triangular distribution which generates a random value that lies between a minimum and a maximum values that are most likely close to the mode. And the discrete uniform distribution between a minimum and maximum values. Figure 5.23 presents a schematic representation of the selected probability distributions.

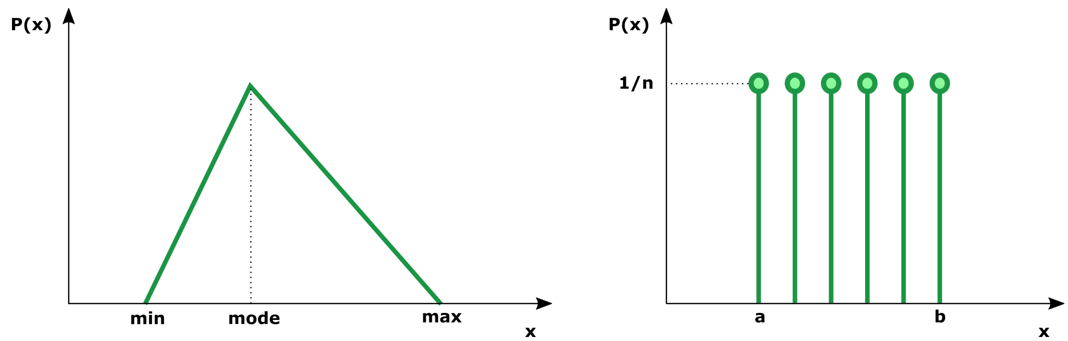



**Figure 5.23 –** *Triangular and discrete uniform distributions.*

The data setting used to obtain the numerical results is the following:

- Number of process machines: 259.
- Number of days of the simulation: 5.
- Waiting time of lots in process machine ports: The time spent by lots waiting to be transported to the metrology area follows a discrete uniform distribution between 60 and 300 seconds.
- Lot transport time: The time needed for an operator to move a lot to the metrology area is generated by a triangular distribution. The considered times for minimum, mode and maximum values are 30, 90 and 300 respectively.
- Number of operators to transport lots: 3.
- Frequency to review metrology racks: The holding system simulates the frequency that an operator goes to the metrology racks to pick up lots. It keeps lots in the metrology queue a given time determined by a discrete uniform distribution between 600 and 1500 seconds.
- Number of operators to measure lots: 1.
- Lot transport time in metrology area: The time spent for operators to move lots from racks to the metrology tools is generated through a triangular distribution, with a minimum time of 10 seconds, a mode value of time of 20 seconds and a maximum value of 30 seconds. Racks are in front of metrology tools and few seconds to start metrology tools is also considered.
- $QT^{Max}$: The maximum queue time considered is 7200 seconds. Queue time values larger than this limit would be excluded from the calculation of the global average of queue time, since these lots are considered off-peak lots that reach exceptionally large queue time values that degrade the significance of average of queue time value.

#### 5.3.3.2 Analysis

The final results of the numerical experiments are discussed by parameter. Different sampling strategies are compared to see the possible improvement margin for each of them. Only flagged lots (those that in reality were measured) will experience the different changes for each simulation mode. All the sampling strategies are proposed to improve the indicators compared to the current situation (normal mode). Maybe some elements should be added in the simulation modes to make the simulation model more accurate. However, it is possible to compare the new sampling strategies to know under the same conditions which one is the best.

The results obtained from the QT values show that by only adding skipping policies in the current situation (normal mode), it is possible to save 80 seconds on average. Among the different sampling strategies, SPSP provides the best value. It is surprising to see how LIFO brings better results than FIFO. This is because last lots lead to smaller QT values and the global QT is reduced, and the QT difference calculated for the first lots is not so large. Table 5.3 presents the values of queue times.

**Table 5.3 –** *Queue times by simulation mode.*

| | **Average of QT (seconds)** | | | | | |
|---|---|---|---|---|---|---|
| Mode | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Total |
| **Normal** | 1483 | 1436 | 1654 | 1463 | 1705 | 1549 |
| **Normal with skipping** | 1405 | 1350 | 1584 | 1408 | 1595 | 1469 |
| **FIFO** | 1216 | 1307 | 1319 | 1273 | 1272 | 1278 |
| **LIFO** | 1176 | 1276 | 1293 | 1224 | 1225 | 1240 |
| **SSP** | 1176 | 1281 | 1288 | 1221 | 1228 | 1240 |
| **SPSP** | 1171 | 1268 | 1275 | 1208 | 1220 | 1229 |

For $W@R^{machine}_{Average}$ values, SPSP and FIFO reach the best result with 57. SSP and LIFO achieve a value of 58. As expected, the normal mode, with and without considering the skipping option, has the same $W@R$ values because the unnecessary measures performed by the normal mode do not reduce the risk values. Table 5.4 shows the different $W@R^{machine}_{Average}$ values depending on the approach.

**Table 5.4 –** $W@R^{machine}_{Average}$ *by simulation mode.*

| | $W@R^{machine}_{Average}$ | | | | | |
|---|---|---|---|---|---|---|
| Mode | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Total |
| **Normal** | 35 | 61 | 65 | 66 | 72 | 60 |
| **Normal with skipping** | 35 | 61 | 65 | 66 | 72 | 60 |
| **FIFO** | 33 | 59 | 61 | 64 | 70 | 57 |
| **LIFO** | 33 | 59 | 62 | 65 | 70 | 58 |
| **SSP** | 33 | 60 | 62 | 65 | 70 | 58 |
| **SPSP** | 33 | 59 | 61 | 65 | 70 | 57 |

$W@R^{Max}_{Average}$ values for SSP, SPSP, FIFO and LIFO are the same, as shown in Table 5.5. Again, the normal mode and the normal with skipping mode provide the same results because unnecessary measures do not improve the $W@R$ values.

**Table 5.5 –** $W@R^{Max}_{Average}$ *by simulation mode.*

| | $W@R^{Max}_{Average}$ | | | | | |
|---|---|---|---|---|---|---|
| Mode | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Total |
| **Normal** | 37 | 43 | 46 | 43 | 46 | 43 |
| **Normal with skipping** | 37 | 43 | 46 | 43 | 46 | 43 |
| **FIFO** | 31 | 37 | 37 | 37 | 38 | 36 |
| **LIFO** | 31 | 37 | 37 | 37 | 38 | 36 |
| **SSP** | 30 | 36 | 37 | 36 | 38 | 36 |
| **SPSP** | 30 | 36 | 37 | 37 | 39 | 36 |

Looking at the useless measures, SSP and SPSP do not perform measures when the risk is not reduced, thus no useless measure is performed, leading to a gain of metrology capacity. The normal with skipping mode avoids 2179 useless measures. LIFO performs 468 measures that are not necessary because old lots are measured even if lots processed not long ago by the same machine were measured before. FIFO does 22 useless measures, when actually it should be 0. The explanation is that these lots were sent in the correct order, but lots which arrived later to the metrology area went to a faster metrology tool and when the first lot finished the measure operation, they were considered useless. Table 5.6 shows the number of useless measures.

**Table 5.6 –** *Number of useless measures by simulation mode.*

| | Useless measures | | | | | |
|---|---|---|---|---|---|---|
| Mode | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Total |
| **Normal** | 362 | 495 | 514 | 399 | 409 | 2179 |
| **Normal with skipping** | 0 | 0 | 0 | 0 | 0 | 0 |
| **FIFO** | 3 | 5 | 5 | 4 | 5 | 22 |
| **LIFO** | 79 | 78 | 117 | 103 | 91 | 468 |
| **SSP** | 0 | 0 | 0 | 0 | 0 | 0 |
| **SPSP** | 0 | 0 | 0 | 0 | 0 | 0 |

In Table 5.7 the number of skipped lots in the different modes is shown. The normal with skipping mode skips the same number of lots which were considered as useless measures for normal mode in Table 5.6. SSP and SPSP have the same number of skipped lots (486) because they have the same method to decide which lots should be removed from the metrology queue.

**Table 5.7 –** *Number of skipped lots by simulation mode.*

| | Skipped lots | | | | | |
|---|---|---|---|---|---|---|
| Mode | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Total |
| **Normal** | 0 | 0 | 0 | 0 | 0 | 0 |
| **Normal with skipping** | 362 | 495 | 514 | 399 | 409 | 2179 |
| **FIFO** | 0 | 0 | 0 | 0 | 0 | 0 |
| **LIFO** | 0 | 0 | 0 | 0 | 0 | 0 |
| **SSP** | 83 | 84 | 119 | 107 | 93 | 486 |
| **SPSP** | 83 | 79 | 124 | 107 | 93 | 486 |

#### 5.3.3.3 Summary

Taking into account all parameters, we conclude that SPSP is the best sampling policy because lots spend less time in the metrology racks and are best sampled with an average $QT$ of 1229 seconds. Besides, SPSP provides the best $W@R^{machine}_{Average}$ value (57), as FIFO, but it also saves 486 measures by skipping lots, and no useless measures are performed allowing to gain metrology capacity. SSP is the second best choice with a $QT$ of 1240 seconds, a $W@R^{machine}_{Average}$

value of 58 and 486 skipped lots. The simulation model shows that improvements of queue times can be achieved by adding simple skipping policies.

## 5.4 Conclusions

In this chapter, two simulation model for a real workshops in semiconductor manufacturing have been proposed and analyzed. It has led us to better understand the current risk level, to detect several points to improve the current metrology process and to propose new risk reduction strategies.

The first simulation model corresponding to the Ion Implantation workshops has shown that qualifying new relevant products leads to significant risk reduction. The values of the Maximum $W@R$ and the Average $W@R$ are reduced, and the gain is larger when several products are qualified. Our current research aims at finding a better method to select the products to qualify. We also want to implement new sampling techniques in the simulation model that select lots to measure at the right time that bring information, neither waiting too long before measuring a lot for a process machine nor performing a measure too early. An interesting perspective is to propose new scheduling and dispatching rules to assign lots to process machines with high risk.

The second simulation model led us to calculate queue times of lots and risk values of process machines in a metrology workshop for thickness measurement. New dispatching and sampling strategies have been developed. By comparing indicators such as queue times, $W@R$ values and number of useless measurements, it was shown that a sampling strategy that skips lots with redundant information and order them in decreasing risk is preferable than a FIFO strategy and much more than a LIFO strategy.

# General Conclusion and Perspectives

## General Conclusion

This thesis has been carried out in an academic and industrial framework through a collaboration between the *École des Mines de Saint-Étienne* and *STMicroelectronics*. An important objective was to classify the characteristics of the main metrology workshops in semiconductor manufacturing and to calculate their risk level in order to propose new strategies for reducing the risk. The risk measure in this thesis is the Wafers at Risk (W@R) on process machines.

The specific tables which compose the metrology workshop classification leads to better understand the weight of each metrology workshop along the production and helps to find the explanation of W@R values by showing the details of metrology workshop composition that can be modified in the search of a W@R improvement. Two important key points were found. First, the way lots are dispatched in the workshop impacts the final queue times which will increase the risk values. Depending on the dispatching policy, a W@R reduction can be achieved by switching to another dispatching strategy. Second, the link between the nature of the metrology system and the sampling strategy. Depending on the property of the measured wafer, some sampling strategies are preferable than others. Results showed that for a type of metrology which measures superficial defects on wafers, it is not necessary to follow a product-based sampling policy because there were not product recipes more critical than others, and a sampling by process equipment approach is enough. This led to a constant W@R value by machine, to reduce the risk variability, to save measures, and to decrease the cycle time which improves the yield.

The development of the novel approach to optimize sampling rates showed that, when a sampling strategy change is needed, it is important to consider some factors in order to fully use the metrology capacity. When a group of process machines are covered by a group of metrology tools, it is important to discriminate which are among the machines more critical. By assigning for each machine its probability of processing bad wafers and its process throughput, the approach assigns a sampling rate to each process machine. The numerical experiments showed how the proposed heuristics bring good results no matter what kind of metrology workshop (a unique metrology tool, identical metrology tools or different metrology tools).

The possibility of developing simulation models for semiconductor metrology workshops opens a door that allows the conditions of a workshop to be analyzed more easily and helps to evaluate the impact of potential improvements. Through the models which were developed, two conclusions were reached. Firstly, the importance of product qualification is significant. As soon as a workshop only has few products (certain groups of lots) that can be measured (and thus reduce the risk) the management of W@R becomes uncontrollable and the risk levels depend on the product mix and lot distribution in the workshop. The most convenient situation is to have all products qualified for measurement or at least the right qualifications, otherwise high W@R will be reached. Secondly, some sampling techniques or dispatching rules are more suitable than others in the same conditions. Even if some strategies such as FIFO seem logical, the workshop characteristics and queue times are decisive and can provide unexpected results. Using simple sampling strategies that select the lots which reduce the risk the most among all candidate lots and that skip other lots always provides the best results in terms of W@R and queue times.

## Perspectives

The work developed throughout this thesis led us to identify research perspectives.

Concerning the metrology workshop analysis, new characteristics can be added such as the percentage of products qualified to be measured, the cost of performing the measure, the waiting times of lots after measure and so on. Besides providing the information for every field of the properties of the workshops, new correlations between parameters can also be considered. Through these new correlations, a procedure to regularly analyze all metrology workshops from a global point of view in order to find the one with the largest improvement potential can be developed.

A second perspective refers to the approach for changing sampling strategy. As our results showed, a metrology workshop that uses a sampling by product and operation obtains variable sampling ratios instead of constant ratios for process machines, which depending on the lot distribution waiting to be measured may cause long periods without covering machines. Hence, for metrology workshops that control specific properties of the wafer that change depending on the product and process operation and when the use of a sampling policy by product and operation is mandatory, a combined sampling that would consider the sampling by equipment is proposed. This combined sampling would give priority to the sampling by product and operation to send lots to measure, and the sampling by equipment would cover the process machines to satisfy their sampling rates by sending lots to measure if the the previous sampling policy did not trigger the order.

A third perspective related to the optimization of the sampling rates can be the development of new heuristics in order to assign sampling rates and metrology tools to process machines in a different way. Another idea can be to consider the product qualification of metrology tools, for example by taking into account the number of products processed by machine and by associating it to the metrology tool which has the largest number of qualified products in common.

The fourth perspective concerns risk management in simulation models. The idea is to give importance to the control of the products and not only to the process machines. This is interesting for the metrology systems that measure more complex wafer properties (e.g. critical dimension or thickness) because due to the difficulty of these process steps for each product specification, it is also necessary to pay attention to products and the fact that some products need more control than others. W@R counters by product can be added and also a new indicator that combines the W@R levels of process machines and products. A new system which selects the best lots to measure in order to reduce the W@R in the most risky process machines and products according to the indicator should be developed.

# Appendix A

# Résumé en français

## A.1 Introduction

La complexité des processus de fabrication des semi-conducteurs est en augmentation constante. C'est pour cela que des étapes de contrôle sont ajoutées afin de garantir une qualité de produit élevée. Pour fabriquer un circuit intégré (CI), de nombreuses étapes de production sont nécessaires et la plupart d'entre elles sont répétées, conduisant à des temps de cycle de plus de deux mois. La réduction continue de la taille des circuits intégrés et la variété des produits fabriqués augmentent encore plus cette complexité. Malgré la nécessité des mesures de contrôle, la plupart d'entre elles peuvent être considérées comme inutiles et donc comme des opérations sans valeur ajoutée qui augmentent les coûts du produit final. L'objectif des fabricants de semi-conducteurs est de trouver un équilibre entre l'exécution du bon nombre d'opérations de contrôle et l'atteinte d'un rendement (pourcentage de produits finis respectant les contraintes de qualité) élevé.

Cette thèse vise à analyser les différentes propriétés des ateliers de métrologie, à proposer de nouvelles approches pour optimiser les taux d'échantillonnage et à développer de nouvelles stratégies dynamiques de réduction des risques en fabrication des semi-conducteurs.

L'industrie des semi-conducteurs est basée sur la production de CI qui sont des dispositifs fabriqués en connectant une grande quantité de composants électroniques tels que des transistors, des diodes, des résistances, sur une surface de silicium. Le silicium est un matériau semi-conducteur ayant la propriété de conduire l'électricité ou non en fonction des traitements effectués.

Depuis sa naissance, cette industrie a expérimenté une croissance énorme, devenant l'une des plus grandes industries dans le monde. Le chiffre d'affaires du marché des semi-conducteurs pour 2015 a été estimé à environ 341 milliards de dollars américains et, depuis 1990, maintient un taux de croissance annuel de 7.6% [4].

L'incroyable évolution de l'industrie des semi-conducteurs était liée à sa capacité à doubler le nombre de transistors sur un circuit intégré par an et à réduire le coût par transistor, conformément à la fameuse **Loi de Moore** [68] formulée en 1965. Dix ans plus tard, Gordon E. Moore a reformulé sa loi en affirmant que le nombre de transistors par CI doublera tous les 18 mois. Actuellement, en raison de limitations physiques, cette loi n'est plus valide et, pour poursuivre cette tendance, les scientifiques et les ingénieurs font face à de nouveaux défis technologiques [52].

## Les étapes de fabrication des semi-conducteurs

La fabrication des circuits intégrés (CI) commence avec la phase préliminaire de la préparation des plaquettes en découpant un lingot de silicium. Il est découpé en plaquettes de 0.75 mm d'épaisseur en conservant une forme circulaire pour minimiser les pertes dues à la manipulation des plaquettes dans les usines. Les plaquettes tranchées sont polies pour avoir une surface aussi plate que possible. Les étapes de fabrication de circuits intégrés sont généralement divisées en deux étapes principales :

1. **Front-End** : Cette étape correspond à la fabrication de la plaquette. Elle est divisée en deux étapes majeures : Front-End Of Line (FEOL), où les parties actives et passives des composants sont fabriqués (transistors, condensateurs, etc.) et le Back-End Of Line (BEOL), où les composants métalliques connectent les dispositifs et les différentes couches [103]. Cette étape consiste en plusieurs étapes de production qui sont continuellement répétées dans la route de production.

2. **Back-End** : Au cours de cette deuxième étape, chaque dé de la plaquette est soumis à des essais électroniques, coupés et séparés. Les mauvais dés seront rejetés lors de la prochaine étape de l'emballage. Les bons dés sont sciés et collés à l'armature du boîtier final, leurs connexions reliées au fil et un test final sera effectué pour valider les spécifications.

Le travail de cette thèse est exclusivement axé sur le traitement du Front-End. La surface de la plaquette est recouverte couche par couche jusqu'à ce que le CI soit terminé. Le nombre de couches dépend du type de technologie du produit, les technologies plus complexes peuvent atteindre jusqu'à 40 couches, ce qui équivaut à plus de 400 opérations séquentielles. Les principales étapes de fabrication sont :

- *Oxydation* : Les plaquettes sont chauffées à températures élevées en présence de $O_2$, qui produit la croissance d'une couche de dioxyde de silicium ($SiO_2$).

- *Déposition* : Une multitude de couches minces de différents matériaux sont déposées sur la surface de la plaquette au moyen de plusieurs production. Les deux plus importants sont le dépôt chimique en phase vapeur (CVD) et le dépôt physique en phase vapeur (PVD). Il existe également le dépôt chimique en phase vapeur à plasma renforcée (PECVD) qui utilise des techniques de plasma pour produire les réactions chimiques, et plus récemment, le dépôt de couche atomique (ALD).

- *Photolithographie* : Il est chargé de marquer les motifs sur la plaquette. La plaquette est complètement revêtue d'un film de polymère photosensible. Ensuite, un masque photo ou un réticule est placé juste au-dessus de la plaquette et une lumière ultraviolette traverse le masque polymérisant les sections non masquées de la surface de la plaquette. En dernier lieu, ces sections sont retirées de la surface pour continuer à développer le motif.

- *Gravure* : Dans cette étape, les sections de la couche et la surface de la plaquette non couverte par le motif restant de l'étape de photolithographie sont supprimées. Il existe deux types de gravure fondamentaux : Gravure humide qui utilise des produits chimiques liquides et gravure sèche qui utilise des techniques de plasma.

- *Aplanissement* : Pour obtenir une couche plane et éliminer les substances indésirables, la surface de la plaquette est polie. La technique d'aplanissement la plus courante est le Polissage mécanique et chimique (CMP).

- *Implantation ionique* : Afin de modifier les propriétés électriques de certaines régions de la plaquette, un bombardement ionique est effectué pour modifier les conditions de charge du matériau. Différents types de dopants, d'énergie et de paramètres de dose sont utilisés.

La Figure A.1 montre les principales étapes des production du Front-End. Le flux de production n'est pas linéaire mais réentrant. Les lots reviennent de nombreuses fois aux différentes étapes de fabrication, répétant certaines opérations de traitement couche après couche sur la plaquette.

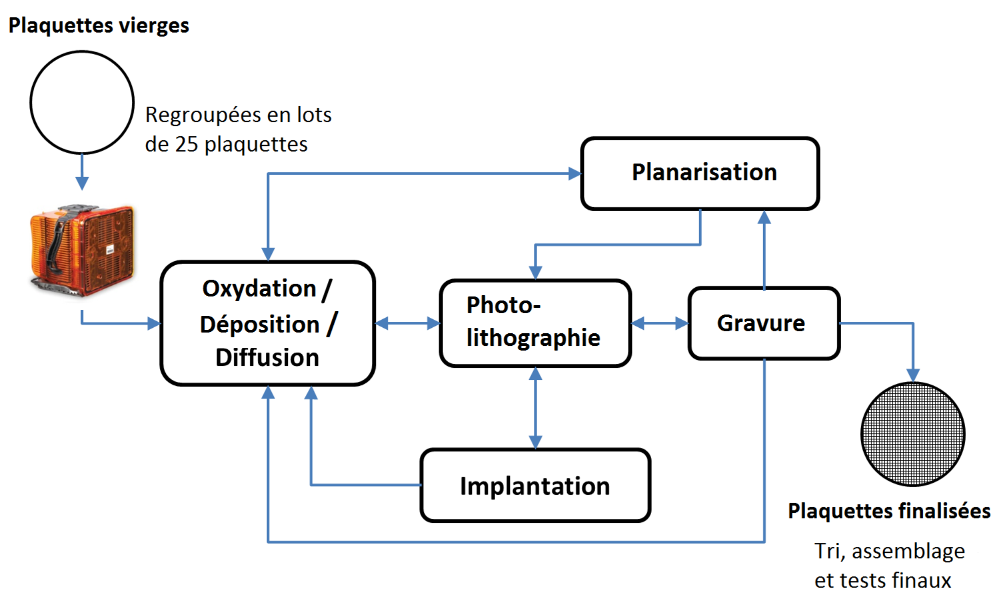



**Figure A.1 –** *Les étapes des production du Front-end. [51].*

## Méthodes de contrôle dans la fabrication de semi-conducteurs

En raison du temps investi pour compléter toute la phase de fabrication de la puce électronique, l'utilisation de contrôles est nécessaire et chaque année la source de défauts diminue progressivement.

Pour assurer leur qualité, il est nécessaire de contrôler l'état des produits (propriétés physiques et électriques) une fois l'opération de production terminé, en validant que le résultat de la mesure du produit correspond bien à l'attendu et ainsi le bon état de la machine de production.

La discipline scientifique responsable de la mesure est appelée **Métrologie**.

Des techniques de traitement de données multiples, qui traitent les informations obtenues à partir de la mesure des plaquettes et celles provenant de l'équipement de traitement, conduisent à une amélioration du rendement.

Ces techniques sont connues sous le nom de techniques de **Contrôle Avancé des Procédés** et elles conduisent à la diminution du nombre d'excursions, à la correction des dérives du processus de production et à la réduction de la variabilité.

Les principales techniques de contrôle avancé des procédés sont résumées et brièvement expliquées :

1. **Maîtrise Statistique des Procédés (SPC en anglais)** est une méthode qui assure la stabilité des procédés grâce à des outils statistiques. En fonction de l'état du procédé donné par un SPC, plusieurs actions peuvent être prises depuis l'ajustement des paramètres du procédé jusqu'à l'arrêt de la machine de production. La SPC a continuellement évolué au fil des ans avec des changements de technologie de fabrication [86].

2. **Détection de défauts et classification (FDC en anglais)** consiste à suivre les variations du procédé au moyen de modèles statistiques en utilisant les paramètres des machines de production (température, pression, flux de gaz, etc.) collectés en temps réel. Dès qu'une variation inattendue (défaut) est détectée, la machine de production (aussi appelée process par la suite) est interrompue et des actions correctives sont prises.

3. **Run-to-run (R2R)** est une technique de contrôle utilisant des boucles de régulation qui rectifient l'écart du procédé par rapport à la cible définie. En utilisant les informations sur l'équipement et le lot obtenues à la suite d'une analyse, le contrôleur modifie les paramètres de recette pour la prochaine exécution [54]. Il y a deux modes pour les boucles de contrôle : *Feed-Forward* et *Feed-Back*. La première ajuste les paramètres en fonction des résultats d'une étape précédente, et la seconde ajuste les paramètres en utilisant les résultats de l'exécution précédente.

4. **Métrologie virtuelle (VM en anglais)** est une technique basée sur la prédiction des mesures de la plaquette. Il s'agit de modèles prédictifs générés à l'aide de paramètres collectés par des capteurs tels que la température, la pression, etc. qui permettent de prévoir les paramètres électriques et physiques de la plaquette [13]. L'objectif est de

réduire le nombre de mesures directes et de les remplacer par des mesures virtuelles qui peuvent avertir plus tôt lorsqu'un procédé commence à dériver.

## Description du problème

Au fil des années, les usines de semi-conducteurs ont appliqué les meilleures pratiques pour améliorer leur rendement de fabrication [20] et en particulier en utilisant des méthodes de contrôle des procédés. Il est important de préserver les machines de production dans de bonnes conditions, parce que si une machine commence à être défectueuse, toutes les plaquettes produites seront affectées, générant des excursions possibles et augmentant ainsi le nombre de plaquettes rejetées.

Il y a deux aspects importants à prendre en compte pour assurer la couverture complète des machines de production :

- La mesurabilité des lots traités, c'est-à-dire le pourcentage de lots traités qui sont autorisés à être mesurés.
- La stratégie d'échantillonnage utilisée. L'échantillonnage est la sélection d'une partie de la population pour déterminer certaines caractéristiques. Dans ce cas d'importance est la fréquence des lots de mesure (taux d'échantillonnage) et la stratégie étant : quels lots sont sélectionnés ?

Pour certains systèmes de métrologie, il est difficile de qualifier de nouvelles opérations de production souvent parce que les recettes sont composées d'un grand nombre de paramètres [81] ou parce que le logiciel utilisé est assez ancien et prend beaucoup de temps.

Par conséquent, juste un pourcentage des lots traités appartient au groupe de lots **mesurables** et ceux qui ne sont pas qualifiés pour être mesurés sont appelés lots **non mesurables**. Cette particularité devrait être prise en compte pour tous les domaines de métrologie puisque, selon la nature de l'atelier, tous les lots ne peuvent pas être candidats à la mesure et donc à réduire le risque.

En ce qui concerne les techniques d'échantillonnage, elles peuvent être classées en trois classes principales : statique, adaptative et dynamique [60].

- La technique d'échantillonnage historique ou **statique** définit les lots à mesurer au début de la production sans aucun changement possible.
- L'échantillonnage **adaptif**, comme l'échantillonnage statique, spécifie aussi les lots à mesurer au début de la production par des règles d'échantillonnage, mais avec la possibilité d'effectuer finalement la mesure ou non en fonction des informations recueillies en production.
- Le troisième type, l'échantillonnage **dynamique** sélectionne en temps réel les meilleurs lots ou plaquettes à mesurer en fonction de la capacité de métrologie et de la situation actuelle.

Des inconvénients ont été observés pour l'échantillonnage statique [59], où certaines machines de production n'ont jamais été couvertes nécessitant l'utilisation de politiques dynamiques.

Pour contrôler les machines de production ou les produits, l'échantillonnage statique est caractérisé par un taux d'échantillonnage donné "1/N", N étant le nombre de lots traités avec des plaquettes à risque, ce qui signifie que pour un taux de 1/3, un lot va être mesurée après 3 lots traités et si chaque lot contient 25 plaquettes, il y aura 75 plaquettes à risque.

W@R (*wafers at risk* en anglais) est un indicateur qui permet de gérer le niveau de risque d'une machine de production, d'un produit, etc. Il indique le nombre de plaquettes traitées qui pourraient être à risque entre deux mesures.

Il y a des ateliers de métrologie, couvrant un groupe de machines de production, qui suivent une stratégie d'échantillonnage statique par produit et opération et qui ont comme inconvénient de ne jamais avoir un taux constant pour couvrir les machines de production, et donc leurs taux d'échantillonnage varient en fonction de la distribution des lots à traiter, et par conséquent les plaquettes à risque ne sont pas contrôlées.

En outre, il a été constaté que certains ateliers de métrologie ne sont pas dépendants du produit, mais dépendent de la machine de production (par exemple sur certaines mesures optiques) et un échantillonnage par équipement est requis. La Figure A.2 montre un exemple des différences entre échantillonnage par produit et opération et échantillonnage par équipement.

Cette figure montre pour trois produits et opérations différents (A, B et C) avec différents taux d'échantillonnage (1/3, 1/4 et 1/5) comment, à partir de cette distribution de lots, aucun lot ne sera mesuré jusqu'au neuvième (1/9). Les valeurs de risque peuvent varier sans contrôle, produisant des périodes de manque de contrôle et même de surcontrôle (mesure de lots consécutifs).

À l'inverse, quel que soit le type de produit ou d'opération, l'échantillonnage par équipement enverra un lot à mesurer suivant le taux d'échantillonnage établi, pour notre exemple ce sera toujours 1/4, et le risque sera sous contrôle avec un taux constant.

Pour les ateliers de métrologie qui ne peuvent passer complètement à l'échantillonnage par machine parce qu'ils dépendent du produit en raison de leur nature (par exemple la mesure d'épaisseur), une nouvelle stratégie d'échantillonnage est nécessaire, qui combine la couverture des machines de production.

Nos premiers travaux consistent à catégoriser les types généraux d'ateliers de métrologie. En premier lieu, les politiques d'échantillonnage utilisées pour adapter les ateliers à l'échantillonnage par équipement, par produit et par opération ou sous une forme combinée, devraient être prises en compte pour créer une procédure générale de changement de stratégie d'échantillonnage.

De plus, il est important de choisir méticuleusement les caractéristiques importantes pour la classification des ateliers de métrologie en termes de règles des lots à ne pas mesurer (*skipping*), leur localisation dans l'usine par rapport aux machines de production contrôlées, la difficulté de créer de nouvelles qualifications, les temps d'attente de la mesure après les

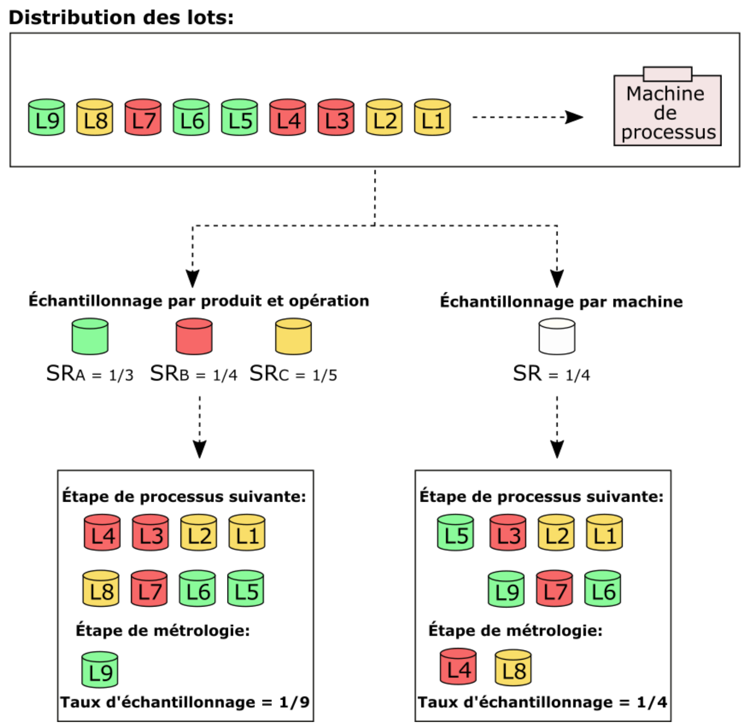



**Figure A.2 –** *Échantillonnage par produit et opération et échantillonnage par machine.*

opérations de traitement et les valeurs actuelles des plaquettes à risque. Ces caractéristiques nous fourniront des informations pour concevoir de nouvelles politiques d'échantillonnage [21], diminuant ainsi les temps d'attente, réduisant le temps de cycle de fabrication, réduisant l'exposition des plaquettes en production à un mauvais traitement et améliorant ainsi le rendement [43].

Comme mentionné précédemment, lorsqu'un changement de stratégie d'échantillonnage nécessite un passage de l'échantillonnage par produit et opération à l'échantillonnage par équipement, le calcul précis des taux d'échantillonnage associé aux machines de production est exigé, donc l'une des clés est de considérer l'optimisation des taux d'échantillonnage une fois le changement d'échantillonnage effectué. Un autre axe de recherche, pour un atelier de métrologie donné avec de sérieux problèmes de qualification des produits sur mesure, avec seulement un faible pourcentage de lots mesurables, est de savoir comment choisir les meilleurs produits à qualifier.

## A.2 Analyse des ateliers de métrologie et amélioration de l'échantillonnage

Dans ce chapitre, les principaux ateliers de métrologie en fabrication de semi-conducteurs sont analysés, en soulignant leurs principales propriétés et leur comportement en fonction de la nature de la mesure. Une méthode de sélection des règles de répartition ou d'échantillonnage en fonction des valeurs de risque et des temps d'attente de l'atelier de métrologie est proposée. Une méthode de changement de stratégie d'échantillonnage a été développée afin de définir les étapes correctes pour identifier les caractéristiques de la stratégie d'échantillonnage actuelle pour un certain atelier de métrologie, vérifier s'il existe une marge d'amélioration en termes de risque, nombre de mesures inutiles, et la phase de suivi correspondante.

### Analyse des ateliers de métrologie

Les caractéristiques de l'atelier de métrologie sont séparées en quatre sections.

Un premier domaine est lié à la nature de la mesure, sa localisation dans l'usine, sa difficulté à être qualifiée, les règles de sélection utilisées qui apportent des informations qualitatives pour établir le lien avec des valeurs à haut risque.

Un deuxième domaine est axé sur les temps de mesure et d'attente et les politiques d'échantillonnage, qui permettent d'associer les niveaux de risque au temps passé pour effectuer une mesure ou les temps d'attente des lots pour proposer de nouvelles approches d'échantillonnage.

Un troisième axe concerne les taux d'échantillonnage et le pourcentage global de lots mesurés par atelier, pour vérifier d'une part si les taux d'échantillonnage sont corrects par rapport aux niveaux de risque, et d'autre part renseigner sur quels ateliers de métrologie sont les plus chargés.

Finalement, un dernier axe contient des données sur les risques qui nous permettent d'identifier les ateliers de métrologie qui ne contrôlent pas régulièrement leurs machines de production.

Les différents axes à analyser, comme le montre la Figure A.3, sont les suivants :

1. **Table des propriétés** : propriétés à vérifier de la plaquette, système de métrologie associé, étapes de production contrôlées, nombre de machines de production contrôlées, difficulté à qualifier les machines de métrologie pour effectuer la mesure, localisation dans l'unité de fabrication et ainsi de suite.

2. **Temps et politiques d'échantillonnage** : temps moyens de mesure et de file d'attente pour chaque atelier de métrologie, nombre d'opérations de métrologie précédentes exécutées et politiques d'échantillonnage actuelles.

3. **Données des mesures** : valeurs des taux d'échantillonnage, moyenne des lots mesurés par semaine et pourcentage correspondant de lots mesurés pour chaque atelier de métrologie.

4. **Valeurs du W@R** : valeurs en termes de *Wafer at Risk* (W@R), moyennes, moyennes des valeurs maximales du W@R. L'indicateur W@R indique le nombre de plaquettes traitées sur une machine de production depuis le dernier contrôle effectué.

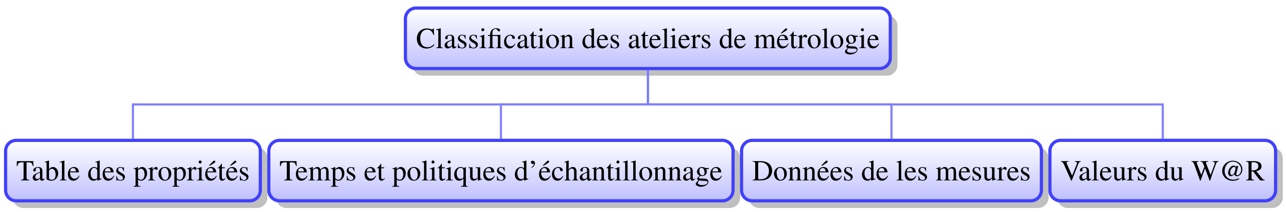



**Figure A.3 –** *Organigramme de classification des ateliers de métrologie.*

## Méthode d'amélioration de l'échantillonnage

Afin de détecter les points d'amélioration possibles dans un atelier de métrologie chargé de contrôler plusieurs machines de production, une procédure de vérification de l'état des stratégies d'échantillonnage actuelles a été développée. L'objectif est d'évaluer le fonctionnement actuel d'un atelier de métrologie, de vérifier s'il ne fonctionne pas comme prévu et de guider vers la réalisation du comportement souhaité. L'utilisation de cette approche doit être fréquente pour vérifier l'utilité de l'utilisation d'un système de métrologie donné et pour corriger la politique d'échantillonnage actuelle.

La procédure a été divisée en quatre étapes principales, comme le montre la Figure A.4 :

1. Identification des propriétés du système de métrologie.

2. Analyse du lien entre le groupe de machines de production et les machines de métrologie.

3. Vérification de l'écart possible à améliorer et sélection des indicateurs de performance clés.

4. Identification de la stratégie d'échantillonnage actuelle, analyse de la mise en œuvre d'une nouvelle stratégie d'échantillonnage et calcul de nouveaux taux d'échantillonnage.

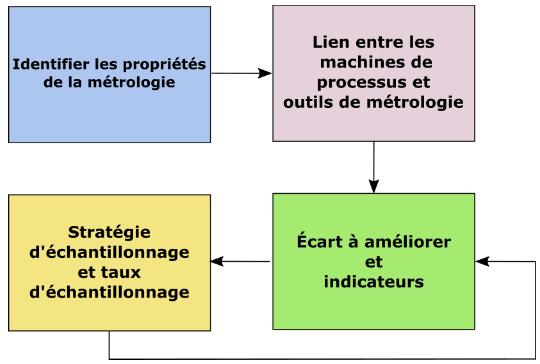



**Figure A.4 –** *Principales étapes de l'approche pour changer la stratégie d'échantillonnage.*

L'approche commence par identifier les propriétés de la machine de métrologie, puis le lien entre les machines de production et les machines de métrologie. La troisième étape est chargée de définir les bons indicateurs pour analyser la marge d'amélioration possible. Si les indicateurs sont bons, l'échantillonnage actuel est conservé. Sinon, l'approche passe à la quatrième étape et une nouvelle stratégie d'échantillonnage doit être mise en œuvre. Après avoir adopté une nouvelle stratégie d'échantillonnage, les indicateurs sont à nouveau calculés pour assurer une tendance appropriée au changement.

## Résultats industriels

La procédure proposée ici a été appliquée pour effectuer un changement de stratégie d'échantillonnage pour un atelier de métrologie sur le site de Rousset à STMicroelectronics. Cette mesure vérifie la surface de la plaquette à la recherche de défauts à l'échelle macroscopique. La politique d'échantillonnage utilisée pour mesurer les lots consistait à attribuer des taux d'échantillonnage par produit et par opération, et donc les taux d'échantillonnage par machine de production n'étaient pas constants.

Pendant une période de trois mois, le changement à une stratégie d'échantillonnage par machine pour les machines de traitement associées à la mesure a été effectué pour certaines machines de production seulement.

La Figure A.6 montre l'évolution des valeurs du W@R avec des graphiques de boîtes à moustaches pour trois machines de production. Les données pour les périodes *Avant* et *Après* consistent en un mois de production, avec le même nombre de jours. Le passage à une stratégie d'échantillonnage par machine permet d'obtenir un taux d'échantillonnage constant qui assure le contrôle attendu de la machine.

En dehors des valeurs du **W@R**, deux autres indicateurs de performance clés ont permis de vérifier s'il y avait une amélioration avec le changement d'échantillonnage : **Variabilité du risque** et **$\Delta$CT** (différence du temps de cycle).

La variabilité globale du risque pour le type de métrologie a été réduite. Même si toutes les machines de production n'étaient pas basculées vers une stratégie d'échantillonnage par équipement, ce premier changement a contribué à réduire l'écart type, $\sigma$, des valeurs du W@R de 91 à 59.

L'impact de la modification de la stratégie d'échantillonnage sur la variabilité globale est représenté sur la Figure A.5.

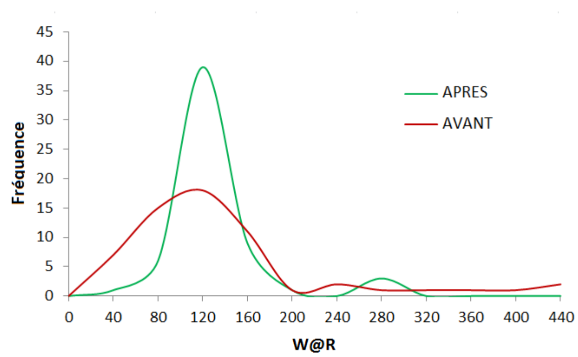
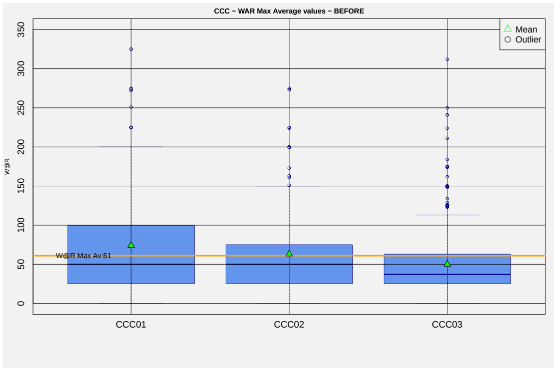
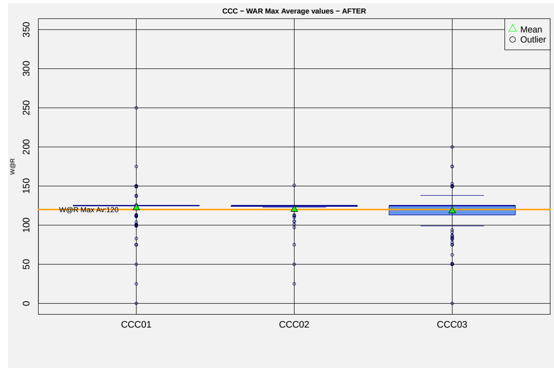


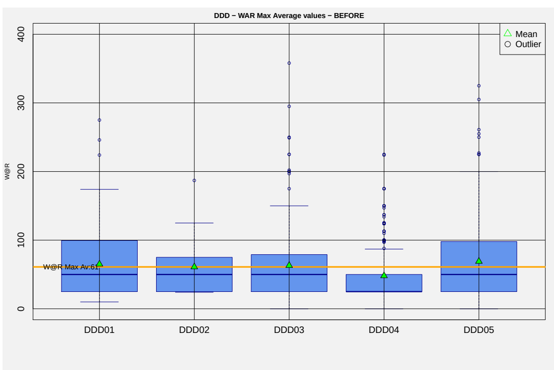

**Figure A.5 –** *Réduction de la variabilité du risque après la mise en œuvre de l'échantillonnage par équipement.*



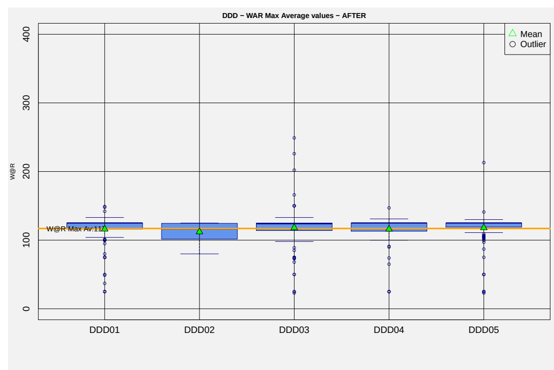

**(a)** *Machines de production CCC - Avant*



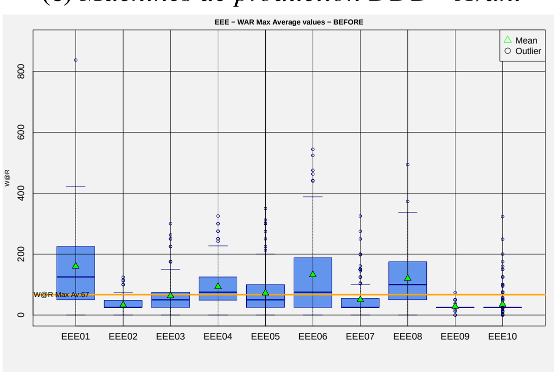

**(b)** *Machines de production CCC - Après*



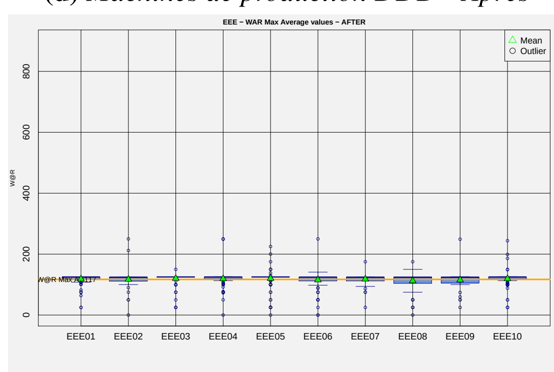

**(c)** *Machines de production DDD - Avant*




**(d)** *Machines de production DDD - Après*




**(e)** *Machines de production EEE - Avant*




**(f)** *Machines de production EEE - Après*

**Figure A.6 –** *Boîtes à moustaches des moyennes maximales du W@R de l'atelier de métrologie*

## A.3 Optimisation du taux d'échantillonnage statique

Selon le type de métrologie, différentes propriétés de la plaquette sont contrôlées, telles que les défauts éventuels de sa surface, ses propriétés mécaniques, l'épaisseur des couches déposées, la dose d'implant, etc. En dehors de leurs avantages, les opérations de métrologie affectent également d'autres aspects du flux de production. Le WIP est augmenté, des coûts supplémentaires sont requis, le temps de cycle est augmenté et les décisions d'ordonnancement sont perturbées.

Dans ce chapitre, une nouvelle approche pour optimiser les taux d'échantillonnage pour un groupe de machines de production couvertes par un atelier de métrologie est proposée. Pour déterminer les taux d'échantillonnage, une capacité de métrologie limitée est prise en compte et les taux de défaillance des machines de métrologie et des machines de production sont pris en compte. Cette approche est axée sur les types d'ateliers de métrologie dans lesquels les lots échantillonnés sont mesurés juste après l'opération de production afin de vérifier les bonnes conditions des machines de production.

### Différents cas selon le type de machine de métrologie

Plusieurs machines de métrologie $t = \{1, \ldots, T\}$ mesurent les produits de plusieurs machines de production $r = \{1, \ldots, R\}$. Les machines de production sont modélisées comme des expériences de Bernoulli et sont différenciées par leur probabilité de défaillance $\boldsymbol{p_r}$.

Les taux de rendement sont indiqués par $\boldsymbol{TP_r}$ pour les machines de production $r$ et par $\boldsymbol{TM_r^t}$ pour la machine de métrologie $t$ qui mesure une plaquette traitée sur la machine de production $r$. La période d'échantillonnage, c'est-à-dire le nombre de cycles de production sur une machine de production $r$ entre deux mesures consécutives, est désignée par $\boldsymbol{SP_r}$.

La valeur de fréquence d'échantillonnage maximale sur laquelle le contrôle qualité est inacceptable est définie par $\boldsymbol{SP^{max}}$. Nous supposons que la production d'une machine défectueuse est entièrement mise au rebut et qu'il n'y a pas de différence entre la valeur des plaquettes sur les différentes machines de production.

Trois cas sont considérés en fonction des machines de métrologie :

1. **Machine de métrologie unique** : Ce premier modèle est développé en considérant une machine de métrologie unique $t$, avec un taux de mesure $\boldsymbol{TM_r}$, chargé de couvrir un groupe de machines de production en mesurant les plaquettes traitées sur celles-ci. Les variables de décision sont les périodes d'échantillonnage $SP_r$, $\forall r = \{1, \ldots, R\}$. Une fois que les varisbles $SP_r$ sont fixées, le débit des plaquettes défectueuses et la consommation de la capacité de la machine de métrologie, $g_r(SP_r)$, sont déterminés.

2. **Machines de métrologie identiques** : Plusieurs machines de métrologie identiques $t = \{1, \ldots, T\}$ contrôlent un groupe de machines de production $r = \{1, \ldots, R\}$. Le débit des machines de métrologie lors de la mesure des lots traités sur la machine de production $r$, $\boldsymbol{TM_r}$, est le même pour toutes les machines de métrologie. Une et une

seule machine de métrologie est assignée pour contrôler la totalité de la production de chaque machine de production, donc $T \leq R$.

3. **Machines de métrologie différentes** : Dans ce cas, nous supposons que différentes machines de métrologie $t = \{1, \ldots, T\}$ contrôlent un groupe de machines de production $r = \{1, \ldots, R\}$. Le débit d'une machine de métrologie $t$ lors de la mesure de lots traités sur la machine $r$ est noté par $\boldsymbol{TM_r^t}$. L'opération de mesure est considérée comme imparfaite et renvoie avec une probabilité $\boldsymbol{\alpha_r^t}$ un faux négatif (le résultat de la mesure est bon alors que $r$ ne fonctionne pas correctement).

Si une défaillance survient dans un premier cycle de production, toutes les plaquettes $SP_r$ suivantes seront mises au rebut, si elles se produisent dans un deuxième cycle de production $SP_r - 1$ plaquettes seront mises au rebut, et ainsi de suite. Nous supposons que la production d'une machine en bon état est parfaite. En ajoutant $\frac{TP_r}{SP_r}$, la fréquence à laquelle la mesure est effectuée à la station, $\boldsymbol{WL_r(SP_r)}$ est le taux global attendu de plaquettes défectueuses produites par une machine $r$ couverte par une période d'échantillonnage $SP_r$, appelée **Wafer Loss**, et est déterminée par (A.1) :

$$WL_r(SP_r) = \frac{p_r TP_r}{SP_r} \sum_{i=0}^{SP_r-1} (SP_r - i)(1 - p_r)^i \tag{A.1}$$

La partie de la capacité de métrologie consommée par la machine de production $r$ pour une période d'échantillonnage donnée $SP_r$, $g_r$, peut s'écrire :

$$g_r(SP_r) = \frac{TP_r}{SP_r TM_r} \tag{A.2}$$

Le problème d'optimisation (P) peut être formulé comme :

$$\min \sum_{r=1}^{R} WL_r(SP_r) \tag{A.3}$$

$$s.t.$$

$$\sum_{r=1}^{R} g_r(SP_r) \leq 1, \tag{A.4}$$

$$SP_r = \{1, \ldots, SP^{max}\}, \quad r = \{1, \ldots, R\}. \tag{A.5}$$

En linéarisant les contraintes (A.3) et (A.4) dans lesquelles les variables de décision $SP_r$ sont au dénominateur, le problème (P) peut être réécrit comme un programme linéaire en nombres entiers (PLNE). Définissons $s$ comme l'indice de la période d'échantillonnage, soit $s = \{1, \ldots, SP^{max}\}$ et la variable binaire $u_r^s$ qui est égale à 1 si la période d'échantillonnage de la machine de production $r$ est $s$, c'est-à-dire $SP_r = s$ et 0 sinon. Par conséquent, (P) peut être reformulé comme un PLNE, désigné par (PI1).

$$\min \sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} WL_r(s)u_r^s \tag{A.6}$$

$$s.t.$$

$$\sum_{r=1}^{R} \sum_{s=1}^{SP^{max}} g_r(s)u_r^s \leq 1, \tag{A.7}$$

$$\sum_{s=1}^{SP^{max}} u_r^s = 1, \quad r = \{1, \ldots, R\}, \tag{A.8}$$

$$u_r^s \in \{0, 1\}, \quad r = \{1, \ldots, R\}; s = \{1, \ldots, SP^{max}\}. \tag{A.9}$$

Ce problème est un problème du sac à dos à choix multiples (Multiple-Choice Knapsack Problem, MCKP, en anglais) tel que présenté par Sinha et Zoltners [85], ce problème a été étudié au cours des années [115] [66], il s'agit essentiellement d'objets organisés en classes, et un seul objet doit être sélectionné par classe. Dans notre problème, les périodes d'échantillonnage sont des objets et les machines de production sont des classes.

Nous avons développé des heuristiques rapides et simples pour chaque cas en fonction du type de machine de métrologie, faciles à mettre en œuvre et compétitives en termes de qualité de solution avec une approche exacte.

## Résultats industriels

L'atelier de métrologie étudié vérifie les propriétés des plaquettes au niveau macro (rayures sur surface, particules, etc.), il consiste en deux machines de métrologie avec une moyenne globale de performance de mesure pour toutes les machines de production de $TM_r^1 = 298.89$ secondes et $TM_r^2 = 297.95$ secondes. Ainsi, elles sont considérées comme des machines de métrologie identiques. Les systèmes de métrologie couvrent 26 machines de production avec des rendements de processus différents, $TP_r$.

La Table A.1 présente les taux d'échantillonnage par machine de production. Les taux d'échantillonnage optimisés sont obtenus par notre approche, et les taux d'échantillonnage réels sont la moyenne des lots traités avant d'envoyer un lot à la mesure sur une période d'un mois.

On peut noter que la stratégie d'échantillonnage utilisée lors de l'extraction des données réelles est une stratégie d'échantillonnage par produit et opération de production. Par conséquent, certaines machines de traitement ne sont pas contrôlées en raison de la distribution des lots comme la machine 6 avec un taux d'échantillonnage réel de 31,23.

Les taux d'échantillonnage optimisés proposés ont une moyenne de 6,11 et la moyenne des taux d'échantillonnage réels est égale 7,04 (à l'exclusion de la machine 6). Les taux d'échantillonnage optimisés sont équilibrés avec un taux de $SP_r = 6$, sauf pour les machines 1, 10, 17 et 25 à 7 car le taux $\frac{TP_r}{TM_r^t}$ permet un contrôle moins strict des lots traités sur ces machines. Le contraire se produit pour la machine 6 qui a besoin d'un taux d'échantillonnage plus petit, $SP_r = 5$.

**Table A.1 –** *Résultats sur les données industrielles pour des machines de métrologie identiques.*

| Machines | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $SP_r$ Réel | 3.3 | 2.5 | 4 | 6 | 5.8 | 31.23 | 14.33 | 13.6 | 5.1 | 4.7 | 4.83 | 6 | 8 |
| $SP_r$ Optimisé | **7** | **6** | **6** | **6** | **6** | **5** | **6** | **6** | **6** | **7** | **6** | **6** | **6** |
| Machines | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| $SP_r$ Réel | 4.9 | 4.9 | 6.8 | 6 | 6.7 | 6.4 | 10 | 9 | 17 | 6 | 8.5 | 5.4 | 6.2 |
| $SP_r$ Optimisé | **6** | **6** | **6** | **7** | **6** | **6** | **6** | **6** | **6** | **6** | **6** | **7** | **6** |

## A.4 Gestion dynamique des risques dans la fabrication de semi-conducteurs

Les stratégies des fabricants de semi-conducteurs pour améliorer leurs performances sont basées sur des mesures de performance telles que le rendement, le débit des machines et le temps de cycle des produits. Pour atteindre ces objectifs, des plans de contrôle sont utilisés pour superviser le comportement des machines de production. En raison de la complexité de cette industrie, une révision permanente pour vérifier l'efficacité des étapes de contrôle est nécessaire [5].

Pour fabriquer un circuit intégré, les plaquettes passent par de nombreuses étapes de procédés telles que la photolitographie, la gravure, le dépôt chimique, etc. Comme représenté sur la Figure A.7, ce chapitre présente et analyse les résultats de modèles de simulation mis en œuvre pour étudier le comportement de certains ateliers dans des usines de fabrication de semi-conducteurs dans le but de contrôler le risque sur les équipements de production.

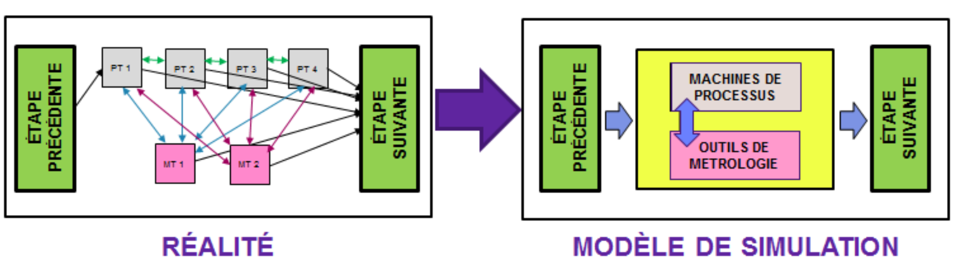



**Figure A.7 –** *Représentation d'un atelier réel dans un modèle de simulation*

La première partie du chapitre se concentre sur une étape particulière de la fabrication de semi-conducteurs : l'implantation ionique. L'objectif est d'analyser le comportement des machines de production dans cet atelier et les niveaux de risque à travers un modèle de simulation, et de proposer et valider des approches pour réduire le risque. Dans la deuxième partie du chapitre, pour un atelier de métrologie en charge de la mesure de l'épaisseur, plusieurs politiques de distribution de lots (FIFO, LIFO, etc.) et des politiques d'échantillonnage sont comparées afin d'obtenir la meilleure stratégie pour obtenir des valeurs de risques plus basses. Une méthode d'échantillonnage pour sélectionner les meilleurs lots afin de réduire les risques et qui apporte des gains en termes de temps de cycle et de temps d'attente est proposée.

## Atelier d'implantation ionique

L'implantation ionique en fabrication de semi-conducteurs est une technique de dopage. Des régions spécifiques peuvent être implantées avec un contrôle précis des niveaux de dopage modifiant la conductivité du semi-conducteur. Les impacts d'ions modifient la composition élémentaire de la plaquette, sinon chaque ion individuel peut provoquer des défauts ponctuels sur la surface de la plaquette. Lorsqu'un lot est mesuré pour vérifier que la machine de production n'est pas défectueuse, les dommages cristallographiques sont vérifiés.

L'atelier est divisé en différents groupes de machines de production avec différentes propriétés adaptées à chaque traitement possible. Il y a une zone de métrologie à proximité où certains lots sont mesurés après avoir été traités. Habituellement, beaucoup de lots ne sont traités que sur une machine de métrologie dans l'atelier avant d'être mesurés ou de continuer directement vers d'autres ateliers. Dans certains cas, un lot peut devoir passer par deux ou trois opérations consécutives d'implantation ionique. Dans ces cas, le lot peut toujours être mesuré après la première opération et parfois après les opérations suivantes en fonction des caractéristiques de la gamme opératoire du lot. Dans notre étude, les lots appartenant à un produit mesurable sont marqués avant d'entrer dans l'atelier pour la première fois, et ils sont mesurés chaque fois qu'ils visitent l'atelier juste après leur traitement.

Dans cette étude, le risque est évalué comme le nombre de plaquettes traitées sur une machine de production depuis le dernier contrôle effectué pour cette machine de métrologie. L'indicateur est appelé $W@R$ (Wafers at Risk en anglais) et correspond à la perte possible de nombre de plaquettes en cas de dysfonctionnement de la machine pendant la production.

Soit $NW(l)$ indique le nombre de plaquettes dans le lot $l$, $W@R_m$ indique les plaquettes actuelles à risque de la machine de production $m$ ($W@R_m$ évolue dynamiquement) et $W@R_m(l)$ indique les plaquettes à risque lorsque le lot $l$ est terminé sur $m$. Ensuite, $W@R_m$ et $W@R_m(l)$ sont mis à jour comme suit quand le lot $l$ est terminé sur $m$ : $W@R_m = W@R_m + NW(l)$ and $W@R_m(l) = W@R_m$.

Si le lot $l$ qui a été traité sur la machine de production $m$ est mesuré, $W@R_m$ est mis à jour (c'est-à-dire diminué) en réduisant sa valeur par le nombre de plaquettes traitées depuis la dernière opération de mesure. Pour l'atelier d'implantation ionique, un lot est mesuré juste après avoir été traité, donc $W@R_m := 0$ quand un lot traité sur $m$ est mesuré.

Cette étude se concentre sur la construction d'un modèle de simulation d'implantation ionique, avec toutes les machines de production et de métrologie, basé sur le comportement réel et les données dans une usine. L'objectif est de superviser le risque de chaque machine de production en termes de valeurs du $W@R$.

Le modèle de simulation est divisé en quatre parties :

1. **Chargement de données réelles.** Il est essentiel pour nous de simuler avec précision ce qui se passe dans la réalité pour valider les améliorations potentielles. Pour ce faire, les données industrielles sont utilisées comme entrées pour injecter des lots dans le modèle. Les données proviennent de rapports qui contiennent ce qui s'est réellement passé dans l'usine. Un tableau d'états gère ce flux de lots. Les lots traités et mesurés sont enregistrés dans un tableau trié par dates.

2. **Corps du modèle de simulation.** L'atelier comprend 17 machines de production réparties en 4 groupes de propriétés différentes. Chaque groupe a son propre comportement. Par conséquent, si les valeurs du $W@R$ des machines du même groupe sont comparées, elles devraient être du même ordre de grandeur, mais les valeurs de $W@R$ de machines de production de groupes différents pourraient être différentes. Il existe deux machines de métrologie, toutes deux avec les mêmes spécifications, sauf qu'une est plus rapide que l'autre.

3. **Modélisation des qualifications des produits.** La qualification d'un produit à mesurer prend beaucoup de temps pour les ingénieurs, puisqu'ils doivent élaborer la recette pour l'équipement de métrologie. Par conséquent, un pourcentage limité de lots est disponible pour la mesure parmi tous les lots traités. La qualification de plus de produits a été testée en utilisant le modèle de simulation pour évaluer les avantages de cette stratégie.

4. **Sorties du modèle.** Notre objectif est de gérer le risque en termes du $W@R$ pour chaque machine de production. Avoir des données pertinentes et des tableaux de sortie pour effectuer l'analyse sont des points clés du modèle de simulation. Grâce aux graphiques, nous pouvons détecter si une machine de traitement est sur ou sous contrôlée. Chaque machine de traitement $m$ a deux paramètres qui sont mis à jour dynamiquement : $W@R_m$ et $W@R^m_{max}$. $W@R^m_{max}$ est mis à jour comme suit : $W@R^m_{max} := \max(W@R^m_{max}, W@R_m)$, c'est-à-dire que $W@R^m_{max}$ augmente seulement avec le temps.

Les diagrammes $W@R$ obtenus après l'exécution du modèle de simulation aident à trouver rapidement sur quelles machines de production nous devons agir afin d'améliorer leur niveau de risque, voir par exemple la figure A.8 où le comportement de deux machines appartenant au groupe A sont montrées. Dans les graphiques, chaque fois qu'un lot est traité, le $W@R$ de la machine de production correspondante augmente, et les pics montrent les valeurs du $W@R$ maximales qui sont atteintes avant qu'une mesure ne soit effectuée. Pour des raisons de confidentialité, les axes ont été normalisés.

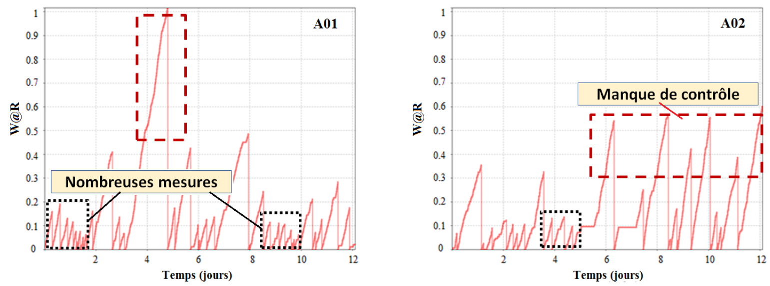



**Figure A.8 –** *Evolution du W@R pour les machines de production A01 et A02 avec des données réelles.*

Fréquemment, il y a des périodes de temps avec un manque de contrôle. L'objectif est de réduire les valeurs maximales du $W@R$ pour rester en dessous des limites du $W@R$ établies et même d'utiliser des limites $W@R$ inférieures. En regardant la Figure A.8 et une limite de 0.5, la valeur du $W@R$ dépasse la limite et une opération de contrôle doit être effectuée dès que possible. Lors de l'analyse de l'efficacité de la mesure pour la machine *A01*, trop de mesures sont parfois effectuées, par exemple avant le début du deuxième jour et entre les jours 8 et 10. C'est une perte de capacité de mesure puisque les opérations de contrôle n'apportent pas suffisamment de valeur ajoutée.

## Atelier de mesure d'épaisseur

Un groupe de machines de production couvertes par un atelier de métrologie peut atteindre de grandes valeurs de risque en fonction de la manière dont les lots sont échantillonnés et envoyés sur les machines de métrologie. Dans cette section, plusieurs méthodes sont étudiées pour échantillonner des lots après les avoir complétés sur un groupe de machines de production, et pour « relâcher » (*skip*) des lots sur des machines de métrologie qui mesurent l'épaisseur des films des plaquettes.

L'atelier de métrologie étudié mesure en moyenne 8 451 lots par semaine, soit environ 22% du total des mesures dans l'usine. Il est composé de 20 machines de métrologie qui mesurent les lots traités par 23 groupes de machines de production différents. Comme dans la section précédente, le risque est géré à l'aide de l'indicateur $W@R$, en considérant la valeur actuelle de la plaquette à risque pour une machine donnée $m$ quand un nouveau lot $l$ avec $NW(l)$ plaquettes est traité : $W@R_m = W@R_m + NW(l)$.

L'objectif principal de cette étude est de rassembler plusieurs indicateurs pour évaluer le comportement normal de l'atelier de métrologie en termes de temps d'attente des lots et de valeurs du risque des machines de production, puis de les comparer avec de nouvelles stratégies simples pour choisir les lots à ne pas mesurer (ou à « relâcher », *skipping*) sur les machines de métrologie et aussi de nouvelles techniques d'échantillonnage priorisant les machines les plus à risque.

De nouvelles règles de répartition et des politiques d'échantillonnage sont proposées :

- Comportement normal avec stratégie de relâchement : La situation actuelle est respectée et seuls les lots qui n'offrent pas de réduction des risques sont relâchés.
- FIFO (*First In, First Out* en anglais) : Les lots plus anciens de la file d'attente de métrologie sont sélectionnés en premier pour être mesurés.
- LIFO (*Last In, First Out* en anglais) : Les lots les plus récemment traités dans la file d'attente de métrologie sont d'abord mesurés.
- Politique d'échantillonnage avec stratégie de relâchement (SSP) : Les lots plus anciens de la file d'attente de métrologie qui n'offrent pas de réduction des risques sont supprimés de la file d'attente et envoyés à l'étape de production suivante.
- Politique de relâchement et d'échantillonnage prioritaire (SPSP) : Les lots sont supprimés de la file d'attente de métrologie comme dans SSP, et le premier lot à mesurer est celui qui a été traité sur la machine de production la plus risquée.

Le modèle de simulation est composé des composantes ci-dessous :

1. **Chargement des données** : Cette partie gère la génération des lots dans le modèle.
2. **Zone des machines de production** : Les lots sont envoyés dans cette zone pour être traités par la même machine que dans la réalité. Les compteurs du $W@R$ par machine sont mis à jour après le traitement des lots.
3. **Transport du lot et temps d'attente** : Le temps passé par un lot de l'étape du production à l'étape de mesure est simulé à travers cette partie du modèle de simulation.
4. **Zone de la métrologie** : Cette zone contient les files d'attente de métrologie, les machines de métrologie et les différentes approches proposées pour simuler les nouvelles stratégies d'échantillonnage.
5. **Mise à jour des paramètres et sorties du modèle** : Avant que les lots quittent l'atelier de métrologie, certains paramètres sont mis à jour et rassemblés pour analyse.

Les résultats finaux des expérimentations numériques sont discutés par paramètre.

Différentes stratégies d'échantillonnage sont comparées pour voir la marge d'amélioration possible pour chacune d'elles.

Seuls les lots marqués (ceux qui ont été mesurés en réalité) vont permettre d'évaluer les différents changements pour chaque mode de simulation.

Toutes les stratégies d'échantillonnage sont proposées pour améliorer les indicateurs par rapport à la situation actuelle (mode normal). Peut-être que certains éléments devraient être ajoutés dans les modes de simulation pour rendre le modèle de simulation plus précis. Cependant, il est possible de comparer les nouvelles stratégies d'échantillonnage pour savoir dans les mêmes conditions quelle stratégie est la meilleure.

## A.5 Conclusions et Perspectives

Cette thèse a été réalisée dans un cadre académique et industriel grâce à une collaboration entre l'*École des Mines de Saint-Étienne* et *STMicroelectronics*.

### Conclusions

Un objectif important était de classer les caractéristiques des principaux ateliers de métrologie en fabrication des semi-conducteurs et de calculer leur niveau de risque afin de proposer de nouvelles stratégies de réduction des risques.

Le risque dans cette thèse est mesuré avec l'indicateur du *Wafers at Risk* (W@R) sur les machines de production.

Les tableaux spécifiques qui composent la classification des ateliers de métrologie permettent de mieux comprendre le poids de chaque atelier de métrologie tout au long de la production, et aident à trouver l'explication des valeurs du W@R en montrant les détails de la composition de l'atelier de métrologie que peuvent être modifiables dans la recherche d'une amélioration du W@R. Deux points clés importants ont été trouvés.

Premièrement, la façon dont les lots sont distribués dans l'atelier a une incidence sur les temps d'attente finaux qui augmenteront les valeurs de risque. Selon la politique de répartition, une réduction du W@R peut être obtenue en passant à une autre stratégie de répartition des lots.

Deuxièmement, le lien entre la nature du système de métrologie et la stratégie d'échantillonnage.

Selon la propriété de la plaquette mesurée, certaines stratégies d'échantillonnage sont préférables à d'autres.

Les résultats ont montré que, pour un type de métrologie qui mesure les défauts superficiels sur les plaquettes, il n'est pas nécessaire de suivre une politique d'échantillonnage basée sur le produit car les recettes ne sont pas plus critiques que d'autres. Cela a conduit à une valeur constante du W@R par machine, pour réduire la variabilité du risque, pour économiser des mesures, et pour diminuer le temps de cycle qui améliore le rendement.

Le développement de la nouvelle approche pour optimiser les taux d'échantillonnage a montré que, lorsqu'un changement de stratégie d'échantillonnage est nécessaire, il est important de prendre en compte certains facteurs afin d'utiliser pleinement la capacité de métrologie.

Lorsqu'un groupe de machines de production est couvert par un groupe de machines de métrologie, il est important de distinguer les machines parmi les plus critiques.

En assignant à chaque machine sa probabilité de traiter les mauvaises plaquettes et sa performance, l'approche affecte une fréquence d'échantillonnage à chaque machine de production.

Les expériences numériques ont montré comment les heuristiques proposées apportent de très bons résultats quel que soit le type d'atelier de métrologie (une machine de métrologie unique, des machines de métrologie identiques ou des machines de métrologie différents).

La possibilité de développer des modèles de simulation pour les ateliers de métrologie en fabrication de semi-conducteurs ouvre une porte qui permet d'analyser plus facilement les conditions d'un atelier et d'évaluer l'impact d'améliorations potentielles.

À travers les modèles qui ont été développés, deux conclusions ont été atteintes.

Premièrement, l'importance des qualifications des produits est significative. Dès qu'un atelier ne dispose que de peu de produits (certains groupes de lots) pouvant être mesurés (et donc réduire le risque), la gestion du W@R devient incontrôlable et les niveaux de risque dépendent de la gamme de produits et de la distribution des lots dans l'atelier.

Deuxièmement, certaines techniques d'échantillonnage ou règles de répartition sont plus appropriées que d'autres dans les mêmes conditions. Même si certaines stratégies telles que FIFO semblent logiques, les caractéristiques de l'atelier et les temps d'attente sont déterminants et peuvent donner des résultats inattendus. L'utilisation de stratégies d'échantillonnage simples qui sélectionnent les lots réduisant le plus le risque parmi tous les lots candidats et qui rejettent d'autres lots donne toujours les meilleurs résultats en termes de W@R et de temps d'attente.

## Perspectives

Le travail développé tout au long de cette thèse nous a conduit à identifier des perspectives de recherche.

Concernant l'analyse des ateliers de métrologie, de nouvelles caractéristiques peuvent être ajoutées telles que le pourcentage de produits qualifiés à mesurer, le coût d'exécution de la mesure et les temps d'attente des lots après mesure. En plus de fournir des informations pour chaque domaine des propriétés des ateliers, de nouvelles corrélations entre paramètres peuvent également être envisagées. Grâce à ces nouvelles corrélations, une nouvelle procédure peut être développée qui permet d'analyser régulièrement tous les ateliers de métrologie d'un point de vue global afin de trouver celui qui présente le plus grand potentiel d'amélioration.

Une deuxième perspective est associée à l'approche pour changer la stratégie d'échantillonnage. Comme nos résultats le montrent, un atelier de métrologie qui utilise un échantillonnage par produit et par opération obtient des taux d'échantillonnage variables au lieu de constants pour les machines de production qui, en fonction de la distribution du lot à mesurer, peuvent provoquer de longues périodes sans que les machines soient couvertes.

Par conséquent, pour les ateliers de métrologie qui contrôlent des propriétés spécifiques de la plaquette qui changent en fonction du produit et du processus et lorsque l'utilisation d'une politique d'échantillonnage par produit et opération est obligatoire, un échantillonnage combiné est envisagé.

Cet échantillonnage combiné donnerait la priorité à l'échantillonnage par produit et par opération pour envoyer des lots à la mesure, et l'échantillonnage par équipement couvrirait les machines de production pour satisfaire leurs taux d'échantillonnage en envoyant des lots sur mesure si la politique d'échantillonnage précédente n'a pas déclenché l'ordre.

Une troisième perspective liée à l'optimisation des taux d'échantillonnage peut être le développement de nouvelles heuristiques afin d'affecter des taux d'échantillonnage et des machines de métrologie pour traiter les machines de production d'une manière différente. Une autre idée peut être de considérer la qualification des produits sur les machines de métrologie, par exemple en prenant en compte le nombre de produits traités par machine et en l'associant à la machine de métrologie qui a le plus grand nombre de produits qualifiés en commun.

La quatrième perspective concerne la gestion des risques dans les modèles de simulation.

L'idée est de donner de l'importance au contrôle des produits et pas seulement aux machines de production. Ceci est intéressant pour les systèmes de métrologie qui mesurent des propriétés de plaquettes plus complexes (par exemple dimension critique ou épaisseur). A cause de à la difficulté de ces étapes pour chaque spécification de produit, il est également nécessaire de prêter attention aux produits qui ont besoin de plus de contrôle que les autres. Des compteurs de W@R par produit peuvent être ajoutés, et un nouvel indicateur pourrait combiner les niveaux du W@R des machines de production et des produits.

Un nouveau système qui sélectionne les meilleurs lots à mesurer afin de réduire le W@R pour les machines et les produits les plus risqués selon l'indicateur devrait être développé.

# Glossary

A glossary of the terminology used in this thesis is presented:

- **APC (Advanced Process Control)**: Set of techniques used for controlling machines and processes: SPC, FDC, R2R and VM.
- **Cycle time**: Time spent by a lot in a workshop or in the entire fab.
- **Die**: Individual chip cut from a silicon wafer.
- **Dispatching strategy**: The way that lots are sent to the next route step, it involves the decision of selecting a given process machine or metrology tool among a group of candidates.
- **Excursion**: Out of control specification on a process or machine.
- **Fab**: Semiconductor fabrication plant, where integrated circuits are manufactured.
- **FIFO**: First-In, First-Out, i.e. the first lot arriving in a area is the first one to leave.
- **FDC (Fault Detection and Classification technique)**: Technique which consists in statistically monitoring process variations by analyzing process machine parameters.
- **IC (Integrated circuit)**: Collection of electronic circuits built on a single piece of silicon substrate.
- **LIFO**: Last-In, First-Out, i.e. the last lot arriving in a area is the first one to leave.
- **Lot**: Group of wafers, normally a lot contains 25 wafers.
- **Measurable lot**: A lot that belongs to a product which recipe exists in the metrology tool and thus can be measured.
- **Metrology**: Science of measurement, and the term is related to control operations performed by metrology tools to gather specific properties of wafers.
- **Qualification**: Possibility for a metrology tool to measure a given lot by registering on it the recipe for measuring the product of the lot.
- **Queue time**: Time spent by a lot between the process operation finishes and the measure operation starts.

- **Recipe**: Set of parameters and instructions for a metrology tool to measure a lot that belongs to a given product.
- **Risk**: In this thesis the risk is quantified by the amount of material at risk, specifically by the number of wafers at risk, which indicates the potential loss if a problem occurs in production.
- **Route**: Sequence of process and measure operations required to obtain the final product.
- **Run to run (R2R)**: Loop control technique that rectifies the process deviation from the defined target.
- **Sampling strategy**: How lots to be measured are selected.
- **Sampling rate**: Frequency at which lots are measured after the process operation.
- **Statistical Process Control (SPC)**: Method to control the stability of processes through statistical tools.
- **Throughput rate**: The production speed on a process machine or metrology tool.
- **Virtual Metrology (VM)**: Technique that predicts the measurements of wafers.
- **Wafer**: Circular plate of silicon used to produce integrated circuits.
- **Wafer Loss**: Overall expected rate of defective wafers produced by a process machine covered by a sampling period on a metrology tool.
- **W@R (Wafer at Risk)**: Number of wafers produced on a process machine since the last control performed on the machine.
- **Work in Process (WIP)**: Number of lots waiting to be processed.
- **Workshop**: Set of process machines or metrology tools that are used for operating a given production or metrology step.
- **Yield**: Refers to the number of good produced units over the total produced units, e.g. wafers, dies, etc.

# Bibliography

[1] Chandrajit P Ahire and Anand S Relkar. Correlating failure mode effect analysis (fmea) & overall equipment effectiveness (oee). *Procedia Engineering*, 38:3482–3486, 2012.

[2] Ram Akella, C-H Lin, and Prasanna Chitturi. Sampling methodology for sem-based defect classification: risk, cost, and benefit analysis. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1998. 1998 IEEE/SEMI*, pages 109–114. IEEE, 1998.

[3] Richard Babikian and Curt Engelhard. Statistical methods for measurement reduction in semiconductor manufacturing. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1998. 1998 IEEE/SEMI*, pages 212–215. IEEE, 1998.

[4] Brittany Bagley. Semiconductor financing: Recent trends. In *Power Semiconductor Devices and ICs (ISPSD), 2016 28th International Symposium on*, pages 1–4. IEEE, 2016.

[5] Samuel Bassetto and Ali Siadat. Operational methods for improving manufacturing control plans: case study in a semiconductor industry. *Journal of intelligent manufacturing*, 20(1):55–65, 2009.

[6] Dario L Benavides, James R Duley, and Blake E Johnson. As good as it gets: optimal fab design and deployment. *IEEE Transactions on Semiconductor Manufacturing*, 12(3):281–287, 1999.

[7] Belgacem Bettayeb, Samuel Bassetto, Philippe Vialletelle, and Michel Tollenaere. Quality and exposure control in semiconductor manufacturing. part i: Modelling. *International Journal of Production Research*, 50(23):6835–6851, 2012.

[8] IEC BIPM, ILAC IFCC, IUPAP IUPAC, and OIML ISO. The international vocabulary of metrology - basic and general concepts and associated terms (vim), 3rd edn. jcgm 200: 2012. *JCGM (Joint Committee for Guides in Metrology)*, 2008.

[9] Mohammed Farouk Bouaziz, Eric Zamai, Frederic Duvivier, and Stéphane Hubac. Dependability of complex semiconductor systems: Learning bayesian networks for decision support. In *Dependable control of discrete systems (dcds), 2011 3rd international workshop on*, pages 7–12. IEEE, 2011.

[10] Ali Bousetta and Andrew J Cross. Adaptive sampling methodology for in-line defect inspection. In *Advanced semiconductor manufacturing conference and workshop, 2005 IEEE/SEMI*, pages 25–31. IEEE, 2005.

[11] Benjamin Bunday, Bart Rijpers, Bill Banke, Chas Archie, Ingrid B Peterson, Vladimir Ukraintsev, Thomas Hingst, and Masafumi Asano. Impact of sampling on uncertainty: semiconductor dimensional metrology applications. In *SPIE Advanced Lithography*, pages 69220X–69220X. International Society for Optics and Photonics, 2008.

[12] Alberto Ceselli and Giovanni Righini. A branch-and-price algorithm for the multilevel generalized assignment problem. *Operations Research*, 54(6):1172–1184, 2006.

[13] Yaw-Jen Chang, Yuan Kang, Chih-Liang Hsu, Chi-Tim Chang, and Tat Yan Chan. Virtual metrology technique for semiconductor manufacturing. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 5289–5293. IEEE, 2006.

[14] Argon Chen, Sean Hsueh, and Jakey Blue. Optimum sampling for track peb cd integrated metrology. In *Automation Science and Engineering, 2009. CASE 2009. IEEE International Conference on*, pages 439–442. IEEE, 2009.

[15] Ping-Shun Chen and Ming-Tsung Wu. A modified failure mode and effects analysis method for supplier selection problems in the supply chain risk environment: A case study. *Computers & Industrial Engineering*, 66(4):634–642, 2013.

[16] Chen-Fu Chien, Shao-Chung Hsu, Smith Peng, and Chia-Hung Wu. A cost-based heuristic for statistically determining sampling frequency in a wafer fab. In *Semiconductor Manufacturing Technology Workshop, 2000*, pages 217–229. IEEE, 2000.

[17] Chen-Fu Chien, Hung-Ju Wang, and Min Wang. A unison framework for analyzing alternative strategies of ic final testing for enhancing overall operational effectiveness. *International Journal of Production Economics*, 107(1):20–30, 2007.

[18] Leng Kwang Chow and Eu Chin Ong. A novel push-pull sampling methodology for test production in semiconductor manufacturing industries. In *Electronic Manufacturing Technology Symposium (IEMT), 2008 33rd IEEE/CPMT International*, pages 1–3. IEEE, 2008.

[19] Marcello Colledani. Integrated analysis of quality and production logistics performance in asynchronous manufacturing lines. *IFAC Proceedings Volumes*, 41(2):8368–8374, 2008.

[20] Sean P Cunningham, Costas J Spanos, and Katalin Voros. Semiconductor yield improvement: results and best practices. *IEEE Transactions on Semiconductor Manufacturing*, 8(2):103–109, 1995.

[21] Stéphane Dauzere-Péres, Jean-Loup Rouveyrol, Claude Yugma, and Philippe Vialletelle. A smart sampling algorithm to minimize risk dynamically. In *2010 IEEE/SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, pages 307–310. IEEE, 2010.

[22] Gary DePinto. Managing factory risk to improve customer satisfaction. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1996. ASMC 96 Proceedings. IEEE/SEMI 1996*, pages 224–229. IEEE, 1996.

[23] Alain C Diebold. *Handbook of silicon semiconductor metrology*. CRC Press, 2001.

[24] Martin E Dyer. An o (n) algorithm for the multiple-choice knapsack linear program. *Mathematical programming*, 29(1):57–63, 1984.

[25] RC Elliott, RK Nurani, D Gudmundsson, M Preil, R Nasongkhla, and JG Shanthikumar. Critical dimension sample planning for sub-0.25 micron processes. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1999 IEEE/SEMI*, pages 139–142. IEEE, 1999.

[26] Miri Gilenson, Michael Hassoun, and Liron Yedidsion. Setting defect charts control limits to balance cycle time and yield for a tandem production line. *Computers & Operations Research*, 53:301–308, 2015.

[27] Richard P Good, Detlef Pabst, and James Broc Stirton. Compensating for the initialization and sampling of ewma run-to-run controlled processes. *IEEE Transactions on Semiconductor Manufacturing*, 23(2):168–177, 2010.

[28] Richard P Good and Matthew A Purdy. An milp approach to wafer sampling and selection. *IEEE Transactions on Semiconductor Manufacturing*, 20(4):400–407, 2007.

[29] AndrÉ Holfeld, Robert Barlovi, and Richard P Good. A fab-wide apc sampling application. *IEEE Transactions on Semiconductor Manufacturing*, 20(4):393–399, 2007.

[30] Afsaneh Noori Houshyar, Muriati Muktar, and Riza Sulaiman. Simulating the effect of supply chain risk and disruption:" a malaysian case study". In *Information Technology (ITSim), 2010 International Symposium in*, volume 3, pages 1362–1367. IEEE, 2010.

[31] Sylvain Housseman, Stéphane Dauzère-Pérès, Gloria Rodriguez-Verjan, and Jacques Pinaton. Smart dynamic sampling for wafer at risk reduction in semiconductor manufacturing. In *Automation Science and Engineering (CASE), 2014 IEEE International Conference on*, pages 780–785. IEEE, 2014.

[32] Albert Hu, He Du, Steve Wong, Peter Renteln, and Emmanuel Sachs. Application of run by run controller to the chemical-mechanical planarization process. ii. In *Electronics Manufacturing Technology Symposium, 1994. Low-Cost Manufacturing Technologies for Tomorrow's Global Economy. Proceedings 1994 IEMT Symposium., Sixteenth IEEE/CPMT International*, pages 371–378. IEEE, 1994.

[33] Jacek Izydorczyk and Michael Izydorczyk. Microprocessor scaling: What limits will hold? *Computer*, 43(8):20–26, 2010.

[34] Scott Jansen, Glenn Florence, Alexa Perry, and Steven Fox. Utilizing design layout information to improve efficiency of sem defect review sampling. In *Advanced Semiconductor Manufacturing Conference, 2008. ASMC 2008. IEEE/SEMI*, pages 69–71. IEEE, 2008.

[35] Payman Jula, Costas J Spanos, and Robert C Leachman. Comparing the economic impact of alternative metrology methods in semiconductor manufacturing. *IEEE transactions on semiconductor manufacturing*, 15(4):454–463, 2002.

[36] Yasuhiro Kaga, Yoshiyuki Sato, Yasuyuki Yamada, Yuuichiro Yamazaki, Masami Aoki, Ryota Harukawa, and Ellis Chang. Integrated defect sampling method by using design attribute for high sensitivity inspection in 45nm production environment. In *Semiconductor Manufacturing (ISSM), 2008 International Symposium on*, pages 379–381. IEEE, 2008.

[37] Faisal I Khan and Mahmoud M Haddara. Risk-based maintenance (rbm): a quantitative approach for maintenance/inspection scheduling and planning. *Journal of loss prevention in the process industries*, 16(6):561–573, 2003.

[38] H Kunzmann, T Pfeifer, R Schmitt, H Schwenke, and A Weckenmann. Productive metrology-adding value to manufacture. *CIRP Annals-Manufacturing Technology*, 54(2):155–168, 2005.

[39] William W Kuo, Ram Akella, and David Fletcher. Adaptive sampling for effective multi-layer defect monitoring. In *Semiconductor Manufacturing Conference Proceedings, 1997 IEEE International Symposium on*, pages D1–D4. IEEE, 1997.

[40] Daniel Kurz, Cristina De Luca, and Jürgen Pilz. A sampling decision system for virtual metrology in semiconductor manufacturing. *IEEE Transactions on Automation Science and Engineering*, 12(1):75–83, 2015.

[41] Daniel Kurz, Cristina DeLuca, and Jürgen Pilz. Sampling decision system in semiconductor manufacturing using virtual metrology. In *Automation Science and Engineering (CASE), 2012 IEEE International Conference on*, pages 74–79. IEEE, 2012.

[42] Dennis Lazaroff, David Bakker, and DR Granath. Evaluating different sampling techniques for process control using automated patterned wafer inspection systems. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1991. ASMC 91 Proceedings. IEEE/SEMI 1991*, pages 92–97. IEEE, 1991.

[43] Robert C Leachman and Shengwei Ding. Excursion yield loss and cycle time reduction in semiconductor manufacturing. *IEEE Transactions on Automation science and engineering*, 8(1):112–117, 2011.

[44] Hyung Joo Lee. *Advanced process control and optimal sampling in semiconductor manufacturing*. ProQuest, 2008.

[45] Kevin Lensing and Broc Stirton. Perspectives on integrated metrology and wafer-level control. In *Semiconductor Manufacturing, 2007. ISSM 2007. International Symposium on*, pages 1–5. IEEE, 2007.

[46] Chen-Ting Lin, Chin-Chang Huang, Chin-Yi Yang, Yao-Wen Wu, Chin-Sheng Lu, Po-Yueh Tsai, Chih-Mu Huang, and Ying-Lang Wang. Defect intelligent sampling system. In *Advanced Semiconductor Manufacturing Conference (ASMC), 2010 IEEE/SEMI*, pages 162–164. IEEE, 2010.

[47] Chris A Mack. Fifty years of moore's law. *IEEE Transactions on semiconductor manufacturing*, 24(2):202–207, 2011.

[48] Gary S May and Costas J Spanos. *Fundamentals of semiconductor manufacturing and process control*. John Wiley & Sons, 2006.

[49] Aymen Mili, Samuel Bassetto, Ali Siadat, and Michel Tollenaere. Dynamic risk management unveil productivity improvements. *Journal of Loss Prevention in the Process Industries*, 22(1):25–34, 2009.

[50] A Hamid Mollah. Application of failure mode and effect analysis (fmea) for process risk assessment. *BioProcess International*, 3(10):12–20, 2005.

[51] Lars Mönch, John W Fowler, Stéphane Dauzère-Pérès, Scott J Mason, and Oliver Rose. A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations. *Journal of Scheduling*, 14(6):583–599, 2011.

[52] Gordon E Moore. No exponential is forever: but" forever" can be delayed![semiconductor industry]. In *Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International*, pages 20–23. IEEE, 2003.

[53] Chandra Mouli and Michael J Scott. Adaptive metrology sampling techniques enabling higher precision in variability detection and control. In *Advanced semiconductor manufacturing conference, 2007. ASMC 2007. IEEE/SEMI*, pages 12–17. IEEE, 2007.

[54] James Moyne, Enrique Del Castillo, and Arnon Max Hurwitz. *Run-to-run control in semiconductor manufacturing*, volume 200. Springer, 2001.

[55] Justin Nduhura Munga, Stéphane Dauzère-Pérès, Philippe Vialletelle, and Claude Yugma. Dynamic management of controls in semiconductor manufacturing. In *Advanced Semiconductor Manufacturing Conference (ASMC), 2011 22nd Annual IEEE/SEMI*, pages 1–6. IEEE, 2011.

[56] Justin Nduhura Munga, Stéphane Dauzère-Pérès, Philippe Vialletelle, and Claude Yugma. Industrial implementation of a dynamic sampling algorithm in semiconductor manufacturing: Approach and challenges. In *Simulation Conference (WSC), Proceedings of the 2012 Winter*, pages 1–9. IEEE, 2012.

[57] Justin Nduhura Munga, Stéphane Dauzère-Pérès, Claude Yugma, and Philippe Vialletelle. A mathematical programming approach for optimizing control plans in semiconductor manufacturing. *International Journal of Production Economics*, 160:213–219, 2015.

[58] Justin Nduhura Munga, Philippe Vialletelle, Stéphane Dauzère-Pérès, and Claude Yugma. Optimized management of excursions in semiconductor manufacturing. In *Simulation Conference (WSC), Proceedings of the 2011 Winter*, pages 2100–2107. IEEE, 2011.

[59] Justin Nduhura-Munga. *Implementing and optimizing dynamic control plans in semiconductor manufacturing*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, Gardanne, France, 2012.

[60] Justin Nduhura-Munga, Gloria Rodriguez-Verjan, Stephane Dauzere-Peres, Claude Yugma, Philippe Vialletelle, and Jacques Pinaton. A literature review on sampling techniques in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 26(2):188–195, 2013.

[61] Yoshio Nishi and Robert Doering. *Handbook of Semiconductor Manufacturing Technology*. CRC Press, 2007.

[62] Raman K Nurani, Ram Akella, and Andrzej J Strojwas. In-line defect sampling methodology in yield management: an integrated framework. *IEEE Transactions on Semiconductor Manufacturing*, 9(4):506–517, 1996.

[63] June S Park, Byung Ha Lim, and Youngho Lee. A lagrangian dual-based branch-and-bound algorithm for the generalized multi-assignment problem. *Management Science*, 44(12-part-2):S271–S282, 1998.

[64] R Gary Parker and Ronald L Rardin. Discrete optimization. 1988.

[65] LEON Pesotchinsky. Problems associated with quality control sampling in modern ic manufacturing. *IEEE transactions on components, hybrids, and manufacturing technology*, 10(1):107–110, 1987.

[66] David Pisinger. A minimal algorithm for the multiple-choice knapsack problem. *European Journal of Operational Research*, 83(2):394–410, 1995.

[67] Sharad S Prabhu, Douglas Montgomery, and George Runger. Combined adaptive sample size and sampling interval x control scheme. *Journal of Quality Technology*, 26(3):164–176, 1994.

[68] I PRESENT. Cramming more components onto integrated circuits. *Readings in computer architecture*, 56, 2000.

[69] Matthew Purdy. Dynamic, weight-based sampling algorithm. In *Semiconductor Manufacturing, 2007. ISSM 2007. International Symposium on*, pages 1–4. IEEE, 2007.

[70] Matthew Purdy, Cabe Nicksic, and Kevin Lensing. Method for efficiently managing metrology queues. In *Semiconductor Manufacturing, 2005. ISSM 2005, IEEE International Symposium on*, pages 71–74. IEEE, 2005.

[71] Aditya P Rastogi, John W Fowler, W Matthew Carlyle, Ozgur M Araz, Arnold Maltz, and Burak Büke. Supply network capacity planning for semiconductor manufacturing with uncertain demand and correlation in demand considerations. *International Journal of Production Economics*, 134(2):322–332, 2011.

[72] Tzvi Raz and E Michael. Use and benefits of tools for project risk management. *International journal of project management*, 19(1):9–17, 2001.

[73] Seung J Rhee and Kosuke Ishii. Using cost based fmea to enhance reliability and serviceability. *Advanced Engineering Informatics*, 17(3):179–188, 2003.

[74] ITRS Roadmap. International technology roadmap for semiconductors, 2009 edn. *Executive Summary. Semiconductor Industry Association*, 2009.

[75] Gloria Rodriguez-Verjan, Eric Tartiere, Jacques Pinaton, Stéphane Dauzère-Pérès, and Alexis Thieullen. Dispatching of lots to dynamically reduce the wafers at risk in semiconductor manufacturing. In *Automation Science and Engineering (CASE), 2012 IEEE International Conference on*, pages 920–923. IEEE, 2012.

[76] Gloria Luz Rodriguez-Verjan, Stéphane Dauzère-Pérès, Sylvain Housseman, and Jacques Pinaton. Skipping algorithms for defect inspection using a dynamic control strategy in semiconductor manufacturing. In *Simulation Conference (WSC), 2013 Winter*, pages 3684–3695. IEEE, 2013.

[77] M'hamed Sahnoun, Belgacem Bettayeb, Michel Tollenaere, and Samuel Bassetto. Smart sampling for risk reduction and delay optimisation. In *Systems Conference (SysCon), 2012 IEEE International*, pages 1–6. IEEE, 2012.

[78] Mhamed Sahnoun, Philippe Vialletelle, Samuel Bassetto, Soidri Bastoini, and Michel Tollenaere. Optimizing return on inspection trough defectivity smart sampling. In *Semiconductor Manufacturing (ISSM), 2010 International Symposium on*, pages 1–4. IEEE, 2010.

[79] Dieter K Schroder, R Hockett, A Diebold, L Fabry, and H Shimizu. Analytical and diagnostic techniques for semiconductor materials, devices, and processes 7. In *The Electrochemical Society*, 2007.

[80] Alejandro Sendón, Stéphane Dauzère-Pérès, and Jacques Pinaton. Simulation model to control risk levels on process equipment through metrology in semiconductor manufacturing. In *Proceedings of the 2015 Winter Simulation Conference*, pages 2941–2952. IEEE Press, 2015.

[81] Krishna Seshan. *Handbook of thin film deposition*. William Andrew, 2012.

[82] M Shanoun, Samuel Bassetto, Soidri Bastoini, and Philippe Vialletelle. Optimisation of the process control in a semiconductor company: model and case study of defectivity sampling. *International Journal of Production Research*, 49(13):3873–3890, 2011.

[83] J George Shanthikumar. Effects of capture rate and its repeatability on optimal sampling requirements in semiconductor manufacturing. In *Semiconductor Manufacturing, 2007. ISSM 2007. International Symposium on*, pages 1–6. IEEE, 2007.

[84] Wataru Shindo, Eric H Wang, Ramakrishna Akella, Andrzej J Strojwas, Wanda Tomlinson, and Richard Bartholomew. Effective excursion detection by defect type grouping in in-line inspection and classification. *IEEE Transactions on Semiconductor Manufacturing*, 12(1):3–10, 1999.

[85] Prabhakant Sinha and Andris A Zoltners. The multiple-choice knapsack problem. *Operations Research*, 27(3):503–515, 1979.

[86] Costas J Spanos. Statistical process control in semiconductor manufacturing. *Proceedings of the IEEE*, 80(6):819–830, 1992.

[87] Costas J Spanos, Payman Jula, and Robert C Leachman. The economic impact of choosing off-line, inline or in situ metrology deployment in semiconductor manufacturing. In *Semiconductor Manufacturing Symposium, 2001 IEEE International*, pages 37–40. IEEE, 2001.

[88] Dean H Stamatis. *Failure mode and effect analysis: FMEA from theory to execution*. ASQ Quality Press, 2003.

[89] Mark G Stewart. Reliability-based assessment of ageing bridges using risk ranking and life cycle cost decision analyses. *Reliability Engineering & System Safety*, 74(3):263–273, 2001.

[90] An-Jhih Su, Cheng-Ching Yu, and Babatunde A Ogunnaike. On the interaction between measurement strategy and control performance in semiconductor manufacturing. *Journal of Process Control*, 18(3):266–276, 2008.

[91] Chao-Ton Su and Chia-Jen Chou. A systematic methodology for the creation of six sigma projects: A case study of semiconductor foundry. *Expert Systems with Applications*, 34(4):2693–2703, 2008.

[92] Susan Sun and Kari Johnson. Method and system for determining optimal wafer sampling in real-time inline monitoring and experimental design. In *Semiconductor Manufacturing (ISSM), 2008 International Symposium on*, pages 44–47. IEEE, 2008.

[93] Cher Ming Tan and Teck-Khim Neo. Qfd implementation in a discrete semiconductor industry. In *Reliability and Maintainability Symposium, 2002. Proceedings. Annual*, pages 484–489. IEEE, 2002.

[94] Christopher S Tang. Perspectives in supply chain risk management. *International Journal of Production Economics*, 103(2):451–488, 2006.

[95] Kai Meng Tay, Chee Sing Teh, and David Bong. Development of a fuzzy-logic-based occurrence updating model for process fmea. In *Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on*, pages 796–800. IEEE, 2008.

[96] Christian Terwiesch, Z Justin Ren, Teck H Ho, and Morris A Cohen. An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management science*, 51(2):208–220, 2005.

[97] Israel Tirkel, Noam Reshef, and Gad Rabinowitz. In-line inspection impact on cycle time and yield. *IEEE Transactions on Semiconductor Manufacturing*, 22(4):491–498, 2009.

[98] Ian Tolle, Ankit Jain, Martin Plihal, and Sumanth Kini. Novel methods for spc defect monitoring: Normalizable diversity sampling: Defect inspection. In *SEMI Advanced Semiconductor Manufacturing Conference (ASMC), 2016 27th Annual*, pages 83–86. IEEE, 2016.

[99] W Tomlinson, RK Nurani, M Burns, and JG Shanthikumar. Development of cost effective sampling strategy for in-line monitoring. In *Advanced Semiconductor Manufacturing Conference and Workshop, 1997. IEEE/SEMI*, pages 8–12. IEEE, 1997.

[100] Robert Trahan and Anthony Pollock. Using an inverted fmea to manage change and reduce risk in a fab. In *Semiconductor Manufacturing Conference Proceedings, 1999 IEEE International Symposium on*, pages 15–18. IEEE, 1999.

[101] Steven R Trammell and Brett J Davis. Using a modified hazop/fmea methodology for assessing system risk. In *Engineering Management for Applied Technology, 2001. EMAT 2001. Proceedings. 2nd International Workshop on*, pages 47–53. IEEE, 2001.

[102] Paolo Trucco, Enrico Cagno, Fabrizio Ruggeri, and Ottavio Grande. A bayesian belief network modelling of organisational factors in risk analysis: A case study in maritime transportation. *Reliability Engineering & System Safety*, 93(6):845–856, 2008.

[103] Peter Van Zant and P Chapman. *Microchip fabrication: a practical guide to semiconductor processing*, volume 5. McGraw-Hill New York, 2000.

[104] Vineeth Veetil, Dennis Sylvester, David Blaauw, Saumil Shah, and Steffen Rochel. Efficient smart sampling based full-chip leakage analysis for intra-die variation considering state dependence. In *Design Automation Conference, 2009. DAC'09. 46th ACM/IEEE*, pages 154–159. IEEE, 2009.

[105] Gloria Luz Rodriguez Verjan. *Smart sampling for risk reduction in semiconductor manufacturing*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2014.

[106] Gloria Luz Rodriguez Verjan, Stéphane Dauzére-Pérès, and Jacques Pinaton. Impact of control plan design on tool risk management: A simulation study in semiconductor manufacturing. In *Simulation Conference (WSC), Proceedings of the 2011 Winter*, pages 1913–1920. IEEE, 2011.

[107] Rick Whitcomb and M Rioux. Failure modes and effects analysis (fmea) system deployment in a semiconductor manufacturing environment. In *Advanced Semiconductor Manufacturing Conference and Workshop. 1994 IEEE/SEMI*, pages 136–139. IEEE, 1994.

[108] Andreas Wieland and Carl Marcus Wallenburg. Dealing with supply chain risks: Linking risk management practices and strategies to performance. *International Journal of Physical Distribution & Logistics Management*, 42(10):887–905, 2012.

[109] Randy Williams, Dadi Gudmundsson, Kevin Monahan, and JG Shanthikumar. Optimized sample planning for wafer defect inspection. In *Semiconductor Manufacturing Conference Proceedings, 1999 IEEE International Symposium on*, pages 43–46. IEEE, 1999.

[110] Terry Williams. A classified bibliography of recent research relating to project risk management. *European Journal of Operational Research*, 85(1):18–38, 1995.

[111] Chien-Wei Wu and Wen Lea Pearn. A variables sampling plan based on c pmk for product acceptance determination. *European Journal of Operational Research*, 184(2):549–560, 2008.

[112] Chen Xu, Julie Lee, Vikas Sachan, and Oliver D Patterson. Defect sampling methodology for yield learning during 22nm process development. In *Advanced Semiconductor Manufacturing Conference (ASMC), 2013 24th Annual SEMI*, pages 268–271. IEEE, 2013.

[113] Ascensión Zafra-Cabeza, Miguel A Ridao, and Eduardo F Camacho. Using a risk-based approach to project scheduling: A case illustration from semiconductor manufacturing. *European Journal of Operational Research*, 190(3):708–723, 2008.

[114] Ascensión Zafra-Cabeza, Miguel A Ridao, Eduardo F Camacho, Karl G Kempf, and Daniel E Rivera. Managing risk in semiconductor manufacturing: A stochastic predictive control approach. *Control Engineering Practice*, 15(8):969–984, 2007.

[115] Eitan Zemel. The linear multiple choice knapsack problem. *Operations Research*, 28(6):1412–1423, 1980.

[116] Eitan Zemel. An o (n) algorithm for the linear multiple choice knapsack problem and related problems. *Information Processing Letters*, 18(3):123–128, 1984.