A Literature Review on Sampling Techniques in Semiconductor Manufacturing

Justin Nduhura-Munga, Gloria Rodriguez-Verjan, Stéphane Dauzère-Pérès, Claude Yugma, Philippe Vialletelle, and Jacques Pinaton

***Abstract*—This paper reviews sampling techniques for inspection in semiconductor manufacturing. We discuss the strengths and weaknesses of techniques developed in the last last 20 years for excursion monitoring (when a process or machine falls out of specifications) and control. Sampling techniques are classified into three main groups: static, adaptive, and dynamic. For each group, a classification is performed per year, approach, and industrial deployment. A comparison between the groups indicates a complementarity strongly linked to the semiconductor environment. Benefits and drawbacks of each group are discussed, showing significant improvements from static to dynamic through adaptive sampling techniques. Dynamic sampling seems to be more appropriate for modern semiconductor plants.**

***Index Terms*—Control, sampling, semiconductor.**

I. Introduction

SeMICONDUCTOR manufacturing is characterized by more than 700 processing steps, resulting in a significant cycle time of more than two months. With the reduction in device sizes, re-entrant flows (repetition of similar processing steps), and the variety of products to be manufactured, the complexity has strongly increased these recent years. This complexity, combined with the strong competition, forces semiconductor manufacturers to introduce several layers of controls in order to guarantee high yield [1]. Therefore, after some process steps, control operations are introduced at different levels (product, process, and equipment) in order to verify that the process is still under control and the product within specifications. However, due to the cost of an inspection [2] and the related impact on cycle times [3], each time a control operation is introduced, a sampling strategy has to be defined in order to find the trade-off between yield and cycle time [4] [5].

A sampling strategy consists in selecting some lots to measure depending on the available capacity in metrology [6]. This selection of lots to measure aims at minimizing risk in production. Not selecting the best lots to measure can lead to significant losses, especially when a problem occurs. Depending on the production environment, different sampling strategies have to be developed. This is not recent in semiconductor manufacturing [7] [8] [9]. One of the main reasons is that a 100-percent inspection does not provide 100-percent quality since, in semiconductor manufacturing, the inspection is never totally reliable and can easily introduce an error of almost the same order as the fraction of defectives [10]. Significant improvements have been observed in sampling techniques and, today, new challenges are being faced. Complex sampling techniques can now be deployed for real time analysis.

This paper surveys sampling techniques that have been developed in the past twenty years for metrology steps (defect inspection, critical dimensions, overlay, thickness, or step height measurements) in semiconductor manufacturing. We will discuss the trade-off between the cost of inspection and the related cost in term of risk reduction, the development of an effective sampling technique, and future challenges. Articles from the literature are reviewed through statements, critical analysis, and also discussions on industrial deployments in semiconductor plants.

In semiconductor manufacturing, sampling techniques are grouped into three main groups [11]:

1) **Excursion monitoring and control** aim at frequently monitoring the process so that any process deviations are caught and the causes for the process excursion are fixed.
2) **Process integration and yield improvement** aim at adjusting the percentage of lots flagged at the start of their production (baseline lots) in order to identify the main detractors for a given technology and eliminate them. For low-mix semiconductor plants, the percentage of lots flagged at the start of production is adjusted in order to compensate the potential loss based on measurement results.
3) **Defect detection and learning** aim at learning on different defect types and their casual mechanisms: Killer rates. The rate of sampling has to enable defect detection

Manuscript received May 15, 2012; revised March 12, 2013; accepted March 30, 2013. Date of publication April 4, 2013; date of current version May 1, 2013.

J. Nduhura-Munga is with the Ecole des Mines de Saint-Étienne, Centre Microélectronique de Provence, Gardanne 13541, France, and also with STMicroelectronics Crolles, Crolles 38926, France (e-mail: nduhura@emse.fr).

G. Rodriguez-Verjan is with the Ecole des Mines de Saint-Étienne, Centre Microélectronique de Provence, Gardanne 13541, France, and also with STMicroelectronics Rousset, Rousset 13106, France (e-mail: rodriguez@emse.fr).

S. Dauzère-Pérès and C. Yugma are with the Ecole des Mines de Saint-Étienne, Centre Microélectronique de Provence, Gardanne 13541, France (e-mail: dauzere-peres@emse.fr; yugma@emse.fr).

P. Vialletelle is with STMicroelectronics Crolles, Crolles 38926, France (e-mail: philippe.vialletelle@st.com).

J. Pinaton is with STMicroelectronics Rousset, Rousset 13106, France (e-mail: jacques.pinaton@st.com).

Digital Object Identifier 10.1109/TSM.2013.2256943

0894-6507/$31.00 © 2013 IEEE

at a rate that is matched to the one of root-cause analysis and problem fixing.

In this survey, we focus on the first group of sampling techniques: **Excursion monitoring and control**. The objective is twofold: Reduce the number of measurements without increasing the risk in production, and detect as quickly as possible potential excursions. Missing these two objectives may lead to significant losses. Indeed, if the focus is only on the reduction of measurements, the danger is to miss the detection of potential excursions. For critical layers [12] or when a process is likely to be out of control, increasing the number of measurements can help to detect as quickly as possible excursions. Similarly, if the focus is only on excursion monitoring, the danger is to increase the number of measurements leading to increased cycle times, and therefore increased product costs. However, the application of excursion monitoring and control may vary depending on the product life cycle. When a new product is introduced, the number of measurements must be increased because of the high risk of excursion. In the ramp-up phase, the number of measurements will progressively be reduced before being stabilized during the mature phase, and finally strongly reduced in the end-of-life phase [13]. The goal is to better use the available metrology capacity.

We classify sampling techniques into three main groups: **Static**, **adaptive**, and **dynamic**. Static or start sampling techniques are based on fixed rules that are not changed throughout production. Adaptive sampling techniques consist in adapting sampling rules defined at the start of production. Depending on information brought by other types of controls (statistical analysis, process variations, maintenance, etc.), rules defined at the start of production are adjusted in order to prevent potential drifts or reduce the material at risk (number of lots processed on a production tool between two controls). Dynamic sampling techniques consist in selecting in real time the best lots or wafers to measure depending on the inspection capacity and the actual situation. No rule is defined at the start of production and the decision to sample or not a lot is taken before the inspection step, and based on the information brought by the lot.

For each group (static, adaptive, and dynamic sampling techniques), we define six indicators: Year, mathematical technique, rule-based technique, industrial deployment, simulation, and comparison with other techniques.

The paper is structured as follows. Section II presents and discusses **Static** sampling techniques. Section III and Section IV describe **Adaptive** and **Dynamic** sampling techniques respectively. Section V concludes the paper and provides avenues for further research.

TABLE I
SURVEY ON STATIC OR START SAMPLING

	Year	Mathematical technique	Rule-based technique	Industrial deployment	Simulation	Comparison with other techniques
Lazaroff *et al.* [17]	1991		*		*	
Nurani *et al.* [18]	1994		*		*	
Nurani *et al.* [19]	1996	*				
McIntyre *et al.* [20]	1996		*			*
Tomlinson *et al.* [21]	1997		*	*		
Scanlan *et al.* [22]	1998		*			
Elliott *et al.* [23]	1999		*			
Chien *et al.* [24]	2000	*				
Lee *et al.* [11]	2001	*				
Chien *et al.* [25]	2001		*			*
Shumaker *et al.* [26]	2003		*	*		
Xumei *et al.* [27]	2003		*		*	
Wu and Pearn [28]	2006	*			*	
Kwang and Chin [29]	2008		*	*		

II. STATIC OR START SAMPLING

Static or start sampling technique consists in determining a fixed number of lots to measure at different manufacturing stages. The number of lots to measure depends on the available inspection capacity, the maturity of the technology, and the process step criticality [14]. The frequency and the sensitivity of the measurement are selected in advance, at the start of production. The objective is to monitor and detect process drifts and limit the material at risk [15] between controls. For example, if the sampling plan is to control one lot every five lots, the objective is to limit the material at risk to not more than five. By always measuring the same lots or wafers, the technique enables the identification of the added defect density between sequential inspection steps [16]. Other advantages of a static sampling technique are the simplicity of implementation and adequate resourcing [14].

The technique has been widely used in the 1990's in most semiconductor plants. Nowadays, it does not fit high-mix semiconductor plants because of its main drawbacks of not taking into account the factory dynamics and variability. For the considered lots, there is a strong impact on the cycle time and an increased risk of yield losses due a higher number of steps and the significant time spent in front of each inspection step. However, the technique is still used during the phase of integration for some specific products. In some semiconductor companies, especially in a low-mix context, where a production tool can be qualified to process only a specific type of product, the technique remains valid and some optimized solutions can be designed. The practice is to use tool matching and application of six sigma quality to optimize sampling.

Among papers surveyed in Table I, note that, even if all papers are applied to a case study of a semiconductor plant, very few provide industrial deployments [21] [26] [29]. In [21], the study performed in an IBM plant to determine the optimal sampling plan for the poly etch module is described. The goal is to minimize both the risk for the product and the cost of inspection. Three decision variables are considered in

TABLE II
MATHEMATICAL TECHNIQUES OR APPROACHES FOR STATIC OR START SAMPLING TECHNIQUES

	Algorithms or Mathematical Techniques
Nurani *et al.* [19]	Heuristic approach
Chien *et al.* [24]	Bayes' theorem
Lee *et al.* [11]	Self-Organizing Feature Map (SOFM) network
Wu and Pearn [28]	Process capability index **Cpmk**

the study: Lot sampling interval, number of wafers per lot, and process control limits. Results indicate that an optimal sampling plan may require additional inspection capacity whose cost is much lower than the benefits. In [26], a sampling method developed at Motorola is discussed. The method is based on two steps: The first step consists in determining products that are good candidates for sampling, and the second step performs analysis to determine the break-even operating constraints. The method is developed and validated against historical data. Results indicate a reduction of wafer test costs by a factor of 10. Kwang and Chin [29] worked on data management. They present an industrial deployment of an automatic push-pull sampling methodology. The methodology consists in the transition from manual to automated sampling controls in order to propagate the correct sampling data to the operators and reduce sampling errors due to human interventions. Results indicate an increase of two percent in productivity.

The efficiency of an algorithm depends on its application. This is the case in semiconductor manufacturing, where the environment may completely change from one factory to another and the degree of complexity is not always the same. Different mathematical techniques have been proposed but none of them has been really deployed. Table II presents mathematical techniques and approaches surveyed in the literature.

The complexity is such that most of static or start sampling techniques are rule-based and take into account some observations within the fab, personal experiences, and statistical analysis. Lazaroff *et al.* [17] present an evaluation of different defect sampling techniques using linear regression. A discussion on the strengths and weaknesses of various sampling techniques for Critical Dimension (CD) measurement is presented by Elliott *et al.* [23]. Chien *et al.* [25] and Xumei *et al.* [27] worked on optimizing sampling techniques for overlay measurements and validated their experiments through simulations using historical data from semiconductor plants. Nurani *et al.* [18] present an economic model for optimizing a sampling plan. The model aims at specifying the number of lots to inspect, the number of wafers within a lot, and the number of dies per wafer. Increasing the cost of inspection (number of lots or wafers to inspect) leads to an increased benefit by detecting excursions very quickly. However, above a certain limit, if the cost of inspection is still increasing, all revenues gained by inspections will be offset by the increased learning and subsequent defect reduction.

Close to the work of Nurani *et al.* [18] are the works of McIntyre *et al.* [20] and Scanlan *et al.* [22]. McIntyre *et al.* [20] discuss key factors that influence the cost of an optimal sampling plan and Scanlan *et al.* [22] identify the use of baseline lots as a key in cost inspection reduction.

In the papers on static or start sampling, the authors try to find the best trade-off between the cost of inspections and the cost related to the material at risk. However, decisions are only taken at the start of production and do not consider unexpected events that may occur during production. When for example, the process is likely to be out-of-control, it could be more interesting to sample more lots or wafers in order to detect potential drifts as quickly as possible. When the process is within control, metrology capacity could be saved by reducing the number of sampled lots. These main drawbacks of static sampling led to the introduction and development of adaptive sampling strategies.

III. ADAPTIVE SAMPLING

Adaptive sampling consists in adjusting sampling decisions defined at the start of production. The technique is based on the start/static sampling technique but the main difference is that the number of lots or wafers to select is adjusted throughout production depending on the process state. Most feedback and feed-forward process control techniques are used in conjunction with Statistical Process Control (SPC) techniques to improve sampling efficiency and effectiveness [30] [31]. Table III presents a survey of adaptive sampling techniques in semiconductor manufacturing.

The transition from static to adaptive sampling started in the second part of the 1990's [32] and a significant contribution can be noticed between 1995 and 2005. First industrial deployments can be observed in the beginning of the 2000's [36] [37] [41]. However, among twenty-seven papers browsed in this review (Table III), only eight indicate an industrial deployment. Moreover, among these eight papers, no indication or comparison with other techniques or technologies is given. This shows the complexity and the particularity of the semiconductor environment. Depending on the amount of data to handle, and the strategies in semiconductor plant, a solution can be efficient when simulated but be impracticable because of unexpected events or factory dynamics. The specificity of each factory is such that a given solution can be efficient in one factory and be completely impracticable in another. This explains why no comparison is presented in the literature. Moreover, strong competition and confidentiality reasons explain why many works are not published. Most of the works published or patented do not detail the technical aspects, and actual performances are never published.

Among papers that indicate industrial deployments, Williams *et al.* [36] [37] present the results of a joint research project between Intel Corporation and KLA-Tencor. The project consists in evaluating and optimizing the defect inspection sampling plan for an advanced semiconductor manufacturing process. A Sample Planner is developed by KLA-Tencor to assist in the development of cost-effective defect inspection sampling strategies, and to provide an accurate assessment of whether monitor reduction and/or elimination should be pursued for cost savings. The results of the project indicate that the costs due to defect excursions could com-

TABLE III
SURVEY ON ADAPTIVE SAMPLING

	Year	Mathematical techniques	Rule-based	Industrial deployment	Simulation	Comparison with other techniques
Prahbu *et al.* [32]	1994	*			*	*
Nurani *et al.* [33]	1995		*		*	
Kuo *et al.* [34]	1996		*		*	*
Kuo *et al.* [12]	1997	*				*
Babikian and Engelhard [35]	1998		*			
Williams *et al.* [36]	1999		*	*		
Williams *et al.* [37]	1999		*	*		
Langford *et al.* [38]	2000		*		*	
Nurani and Shantikumar [39]	2000	*			*	*
Lee *et al.* [40]	2001	*			*	
Wootton *et al.* [41]	2001		*	*		
Allebé *et al.* [42]	2002	*			*	
Lee [43]	2002	*			*	
Song-Bor *et al.* [44]	2003		*	*		
Sullivan *et al.* [45]	2004	*		*		
Moon *et al.* [46]	2005	*			*	
Boussetta and Cross [14]	2005		*	*		
Mouli [13]	2005		*			
Shantikumar [47]	2007	*				
Mouli *et al.* [48]	2007	*		*		
Bunday *et al.* [49]	2008		*			
Veetil *et al.* [50]	2009				*	*
Chen *et al.* [51]	2009	*			*	
Sahnoun *et al.* [52]	2010	*			*	
Sahnoun *et al.* [53]	2010	*			*	*
Good *et al.* [54]	2010	*			*	
Nduhura Munga *et al.* [55]	2011	*		*		

pletely eradicate any projected savings from monitor reduction activities, due to the additional defect excursions that would be missed by the reduced inspection sampling plan.

Wootton *et al.* [41] present a study performed between KLA-Tencor and Motorola. The study consists in finding the best sample criteria providing the best representation of existing problems in the inspected wafers. The main drawbacks of random selection are presented and the proposed solution consists in adapting the sample size based on in-line information and priority rules (defect size). Results indicate an improvement of yield, analysis time, and sampling resolution at Motorola.

Boussetta and Cross [14] analyze the key parameters that have to be monitored for an efficient adaptive sampling plan. Their results indicate three key parameters: The variance ratio, the excursion frequency, and the normalized mean shift. They propose a general adaptive sampling plan and recommend a fab-wide strategy, a very good understanding of inspection requirements, and capacity constraints for an efficient adaptive sampling plan.

Song-Bor *et al.* [44], Sullivan *et al.* [45], Mouli *et al.* [48], and Nduhura Munga *et al.* [55] present industrial deployments of adaptive sampling plans in four different semiconductor companies: TSMC, IBM Microelectronics, Intel Corporation, and STMicroelectronics respectively.

Song-Bor *et al.* [44] at TSMC present a capacity-dependence sampling strategy, based on the utilization rate of the defect inspection tools capacity and on the WIP (Work-In-Progress) management. If the utilization of defect detection rises too high, then an automatic function that allows WIP executing defect inspection is turned off temporarily and another function that allows WIP skipping is turned on until the utilization drops to the expected threshold pre-settled by users. If the utilization of defect detection drops too low, the function to force the WIP executing defect inspection is turned on to bring back the utilization level up to the threshold. Results indicate 10% enhancement in tool utilization compared to the previous static sampling plan.

Sullivan *et al.* [45] present an adaptive sampling technique for overlay measurements. The technique is based on a sampling capability ratio (CsK) analogous to the traditional CpK index[1]. The difference between the process capability (CpK) and the proposed CsK is in the selection of historical data. The CpK is the process performance whereas the CsK only considers data from lots that would have been available for skipping through metrology. A skip lot sample plan is implemented based on the results of the CsK. Results indicate significant cost savings. However, authors do not give percentage enhancement.

Mouli *et al.* [48] present an Adaptive Metrology Sampling (AMS) based on a risk score evaluation. The concept consists in weighting each lot and wafer within a lot to make metrology sampling decisions and processing sequence (or priority) on metrology tools. The score varies between 0 and 1 and it is calculated based on Advanced Process Control (APC) and Statistical Process Control (SPC) analysis and observations. Results indicate a reduction of 30% of excursions without increasing tool capacity or sampling rates.

Nduhura Munga *et al.* [55] present an adaptive sampling strategy based on the real time computation of the material at risk. In order to optimize the computational time, a Permanent Index per Context (*IPC*) is developed to reduce risk computation by simple subtractions or additions. Results indicate a risk reduction of more than 30% of material at risk compared to the previous static sampling strategy.

Concerning the technical aspects of proposed solutions, some papers are only rule-based while others are mathematical based. Table IV summarizes different mathematical techniques or approaches browsed in this review for the last ten years.

An important point to note is that, among all papers that present a mathematical technique or an algorithm, only three indicate industrial deployments [45] [48] [55]. Most of the

[1]The CpK index is the process capability index. CpK takes into account both accuracy (centering) and precision (dispersion) and helps to determine the cause of failures and the need for changes in the product design, tooling, or the manufacturing process. The larger CpK value, the greater the indication that the process is consistently under control (is within specification limits) [56].

TABLE IV
MATHEMATICAL TECHNIQUES OR APPROACHES FOR ADAPTIVE SAMPLING TECHNIQUES

	Algorithms or Mathematical Techniques
Babikian and Engelhard [35]	Skip-Lot algorithm (CpK)
Nurani and Shantikumar [39]	Explicit Search algorithm
Lee *et al.* [40]	Self-Organizing Feature (SOFM) network
Lee [43]	Artificial Neural Network (ANN)
Sullivan *et al.* [45]	Skip-lot algorithm
Mouli *et al.* [48]	Risk-Score evaluation algorithm
Chen *et al.* [51]	Integer Linear Programming
Sahnoun *et al.* [52]	Skip-Lot algorithm (risk reduction)
Sahnoun *et al.* [53]	Skip-Lot algorithm (risk reduction)
Good *et al.* [54]	Sampling Compensation Algorithm (SCA)
Nduhura Munga *et al.* [55]	Permanent Index per Context (*IPC*)

papers only use simulations to validate models [12] [50] but very few are industrialized.

Through papers surveyed for adaptive sampling strategies, the specificities of semiconductor plants can be highlighted. Most of the sampling techniques browsed in this review are different. This is because of the specificity of each factory: Lot or wafer management, data storage, production tool management or qualifications, IT infrastructure, expert knowledge, company culture, etc. Therefore, the efficiency of a sampling technique varies depending on its application [14] [33].

Compared to static sampling strategies, adaptive sampling strategies offer two main advantages which lead to an increase in yield. The first advantage is the quick response to process variation by an increase of the number of lots to inspect when the process is likely to be out-of-control. The second advantage is a better use of metrology capacity by the reduction of the number of lots to inspect when the risk reduction is not significant or when the process is really under control. However, some drawbacks can be pointed out regarding the management of resources, the complexity of algorithms, and industrial deployment. By modifying the number of lots to sample (increasing or reducing this number depending on the process state), the workload in metrology is no longer the same throughout production. The complexity of algorithms is such that the validation is most of the time performed through simulation and algorithms are never industrialized. To tackle the problems faced by adaptive sampling strategies, dynamic or smart sampling strategies have been introduced.

IV. DYNAMIC SAMPLING

Dynamic sampling consists in selecting in real time the best lot or wafer to measure depending on the production state, metrology capacity, and the factory dynamics. The main difference with adaptive sampling is that no rule is defined at the start of production and the decision to sample or not a lot is taken when the lot can be selected for metrology. The metrology workload remains balanced contrary to adaptive sampling. The objective is to measure the lot that brings as much information as possible on both risk reduction and process variation. In high-mix semiconductor plants, where more than 200 products can be run concurrently, dynamic sampling techniques are seen as more suitable. Table V presents a survey of dynamic sampling technique in semiconductor manufacturing.

TABLE V
SURVEY ON DYNAMIC SAMPLING

	Year	Mathematical techniques	Rule-based	Industrial deployment	Simulation	Comparison with other techniques
Purdy *et al.* [57]	2005	*		*		
Lensing and Stirton [58]	2007		*	*		
Holfeld *et al.* [59]	2007		*	*		
Good and Purdy [60]	2007	*		*	*	
Purdy *et al.* [61]	2007	*		*		
Kaga *et al.* [62]	2008		*	*		*
Jansen *et al.* [63]	2008		*	*		*
Hyung [64]	2008	*				*
Sun *et al.* [65]	2008	*				
Lin *et al.* [66]	2010		*			
Dauzère-pérès *et al.* [67]	2010	*			*	

The first research works have been published in 2005 and a pioneer is this domain is M. A. Purdy who has authored or co-authored most of the papers found in the literature. His works include industrial deployments [57] [59] [60] [61] and a patent can be found in [68]. Compared to adaptive sampling, dynamic sampling is mainly mathematically-based because of decision levels. Industrial deployments in semiconductor plants have been achieved thanks to the computing power that has strongly increased.

Among papers that indicate industrial deployments, Purdy *et al.* [57] present a Dynamic Sampling System (DSS) that combines a number of separate sampling rules into a single sampling decision. The first step consists in removing all sampling rates, i.e. making all lots measurable. For that, some defect inspection operations are defined so that all lots can enter the metrology queue. The next step consists in selecting lots to add in the metrology queue and lots to skip depending on the metrology capacity and on the information brought by each lot. The selection of lots to introduce in the metrology is performed based on an algorithm that analyses all rules (for example metal etchers at 30%, plasma etch at 10%, and a given product at 25%) and tries to ensure that each rule is satisfied with the minimum overall sampling rate when there are overlapping rules. The Last-In-First-Out (LIFO) principle is also used to ensure that the lots most recently added to the queue will be measured first. The aim is to give the greatest probability that the measurement of the current lot will allow for one or more other lots to be removed from the queue. Results indicate that the DSS has been rapidly adopted within the AMD company and only a small percentage of lots that entered the metrology queue were removed.

Lensing and Stirton [58], Holfeld *et al.* [59], and Purdy *et al.* [61] present and discuss the fab-wide APC sampling deployed within an AMD's fab. This APC sampling system is based on the algorithm introduced by Good and Purdy [60]. The algorithm aims at selecting the best wafers to measure given a set of sampling rules that can be infeasible, by assigning a penalty to each rule that is violated. This penalty is chosen such that it is larger for critical rules. The problem is written as a Mixed-Integer Linear Program (MILP) and the best wafers to measure correspond to the set that minimizes the sum of penalties. Results indicate rapid deployments within the fab and increased product yields. However, authors do not give comparisons with the previous system and percentage enhancement in terms of risk or cycle time reduction.

Kaga *et al.* [62] and Jansen *et al.* [63] discuss the use of design information to dynamically improve sampling for defect review. Lin *et al.* [66] discuss the benefit of developing a dynamic and intelligent sampling system in a semiconductor manufacturing. Based on their experience, they point out three main benefits of a dynamic sampling system: Sampling stability, satisfactory coverage of in-line products, and comprehensive inclusion of process tools. Hyung [64] presents a model that combines the cost of sampling with the performance of control in terms of yield and cycle time. Tests are performed on different areas such as CVD (Chemical Vapor Deposition), PVD (Physical Vapor Deposition), and Photo-Lithography. Results show that the performance of dynamic sampling depends on the characteristics of the process. When the process is very stable, dynamic sampling has no effects, whereas it is effective when data have large step disturbances.

Sun *et al.* [65] present a scoring algorithm based on weighted objectives to determine the optimal wafer sampling for maximum coverage. The algorithm is a multi-stage approach. The first stage consists in setting up various numbers of wafer samples and various numbers of equipment units (chambers, chuck, bowl, etc.). The aim is to ensure that all possible, but not redundant combinations of wafers are captured. The second stage consists in using the scoring algorithm to evaluate and determine the preferred wafer sample based on pre-defined objectives and weighting factors. The score is calculated by multiplying individual normalized scores by associative weighted factors and summarizing them. The last stage uses the second stage results and designs a set of algorithms based on the number of experimental design groups. This set of algorithms is used to select wafers in each group. No industrial assessment is mentioned.

Dauzère-pérès *et al.* [67] present a sampling, scheduling, and skipping algorithm to minimize risk dynamically. The algorithm is based on a Global Sampling Indicator (GSI) that gives a weight to each lot arriving at the measurement step i.e. in front of metrology. This weight is computed based on the lot history and on two key parameters, called Warning Limit (WL) and Inhibit Limit (IL). The WL indicates when the situation starts to become critical, and the IL corresponds to the maximum risk that can be tolerated for each production tool regarding the metrology capacity and production state. An Integer Linear Programming is provided in [69], and helps to compute the values of WL and IL depending on the production state. The sampling, scheduling, and skipping algorithm has been embedded in a prototype and simulated with actual data from STMicrolectronics. Results indicate that the risk can be strongly reduced while keeping a limited number of measures.

TABLE VI

MATHEMATICAL TECHNIQUES OR APPROACHES FOR DYNAMIC SAMPLING TECHNIQUES

	Algorithms or Mathematical Techniques
Good and Purdy [60]	Mix Integer Linear Programming (MILP)
Sun *et al.* [65]	Risk Scoring Algorithm
Dauzère-pérès *et al.* [67]	Global Sampling Indicator (GSI) algorithm

Table VI summarizes the main mathematical techniques, approaches or algorithms developed for dynamic sampling.

If the development of dynamic sampling is still recent in semiconductor manufacturing, significant improvements are reported compared to static and adaptive sampling. However, dynamic sampling has some limitations. Depending on the information brought by each lot, the throughput of different production tools, and the availability of metrology tools, it could be interesting to anticipate the arrival of lots. This means selecting a lot before it is actually available for sampling, e.g. lots being processed on production tools whose next step is metrology. Such "predictive" sampling strategies will require additional information on production flows.

V. CONCLUSION AND PERSPECTIVES

In this paper, we surveyed the literature for sampling techniques in semiconductor manufacturing. We discussed the strengths and weaknesses of the techniques developed the past twenty years for excursion monitoring and control. We focus on the trade-off between yield and cycle time, i.e. reducing the number of inspections without increasing the risk, and detecting as quickly as possible potential excursions.

When comparing with other techniques, significant improvements from static to dynamic through adaptive sampling techniques are noticed. Dynamic sampling takes into account the factory dynamics and variability, and integrate the available metrology capacity. This latter technique is thus more suitable for modern and high-mix semiconductor manufacturing [57] [59] [60] [61] [67], and is one of the keys to increase yield without impacting cycle times. The challenge for the future is in the development of "predictive" sampling techniques that anticipate the arrival of lots. Lots could be accelerated or prioritized. Various research avenues can be explored regarding the delay between a process operation and the next inspection step [70], the cycle time between the process equipment and the inspection equipment, the number of queues in front of inspection steps depending on the recipe to be used, or the time of inspection depending on the technology, the capability, or the process criticality.

ACKNOWLEDGMENT

This work has been done within the framework of a joint collaboration between STMicroelectronics and Ecole

des Mines de Saint-Etienne (Centre Microélectronique de Provence - Georges Charpak) in Gardanne, France. It has also been written as a part of the European project IMPROVE (Implementing Manufacturing science solutions to increase equiPment pROductiVity and fab pErformance).

REFERENCES

[1] N. Kumar, K. Kennedy, K. Gildersleeve, R. Abelson, C. M. Mastrangelo, and D. C. Montgomery, "A review of yield modeling techniques for semiconductor manufacturing," *Int. J. Prod. Res.*, vol. 44, no. 23, pp. 5019–5036, 2006.

[2] D. Gudmundsson and J. G. Shantikumar, "Improving the deployment of inspection tools: Tutorial on inspection capacity and sample planning," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2005, pp. 410–413.

[3] W. Jang, E. H. Wang, and R. Akella, "Economic in-line inspection sampling strategy with learning effects," *Int. J. Prod. Res.*, vol. 38, no. 18, pp. 4811–4821, 2000.

[4] D. Gudmundsson, "Inspection and metrology capacity allocation in the full production and ramp phases of semiconductor manufacturing," Ph.D. dissertation, Univ. California, Berkeley, CA, USA, 2005.

[5] S. J. Lee, "Inspection tool selection and capacity allocation for in-line process control," Ph.D. dissertation, Univ. California, Berkeley, CA, USA, 1999.

[6] R. Barlović, A. Holfeld, and J. Räbiger, *Sampling in Semiconductor Manufacturing*. New York, NY, USA: Wiley, 2008.

[7] B. P. Dudding, "Sampling inspection and quality control," *J. Institution Prod. Engineers*, vol. 21, no. 1, pp. 13–34, 1942.

[8] A. Cronshagen, "Analysis of a cumulative results sampling plan for use with sampling tables using zero acceptance numbers," *Trans. IRE Professional Group Reliability Quality Control*, vol. 4, no. 1, pp. 14–31, Dec. 1954.

[9] J. I. S. Hsu, "A cost model for skip-lot destructive sampling," *IEEE Trans. Reliab.*, vol. 26, no. 1, pp. 70–72, Apr. 1977.

[10] L. Pesotchinsky, "Problems associated with quality control sampling in modern IC manufacturing," *IEEE Trans. Compon. Hybrids Manuf. Technol.*, vol. 10, no. 1, pp. 107–110, Mar. 1987.

[11] J. H. Lee, S. J. You, and S. C. Park, "A new intelligent SOFM-based sampling plan for advanced process control," *Expert Syst. Appl.*, vol. 20, no. 2, pp. 133–151, 2001.

[12] W. W. Kuo, R. Akella, and D. Fletcher, "Adaptive sampling for effective multi-layer defect monitoring," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 1997, pp. 289–293.

[13] C. Mouli, "Adaptive sampling technology: The next step to factory efficiency," *EuroAsia Semicond. Mag.*, vol. 1, no. 5, pp. 1–3, May 2005.

[14] A. Boussetta and A. J. Cross, "Adaptive sampling methodology for in-line defect inspection," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2005, pp. 25–31.

[15] J. W. Bean, "Variation reduction in a wafer fabrication line through inspection optimization," M.S. thesis, Massachussets Instit. Technol., Cambridge, MA, USA, 1997.

[16] R. Guldi, "In-line defect reduction from a historical perspective and its implications for future integrated circuit manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 4, pp. 629–640, Nov. 2004.

[17] D. Lazaroff, D. Bakker, and D. R. Granath, "Evaluating different sampling techniques for process control using automated patterned wafer inspection systems," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 1991, pp. 92–97.

[18] R. K. Nurani, R. Akella, A. J. Strojwas, R. Wallace, M. G. McIntyre, J. Shields, and I. Emami, "Development of an optimal sampling strategy for wafer inspection," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 1994, pp. 143–146.

[19] R. K. Nurani, R. Akella, and A. J. Strojwas, "In-line defect sampling methodology in yield management: An integrated framework," *IEEE Trans. Semicond. Manuf.*, vol. 9, no. 4, pp. 506–517, Nov. 1996.

[20] M. McIntyre, R. K. Nurani, and R. Akella, "Key considerations in the development of defect sampling methodologies," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 1996, pp. 81–85.

[21] W. Tomlinson, R. K. Nurani, M. Burns, and J. G. Shanthikumar, "Development of cost effective sampling strategy for in-line monitoring," in *Proc. IEEE/SEMI Adv. Semicond. Manuf. Conf.*, 1997, pp. 8–12.

[22] B. Scanlan, "Defect inspection sampling plans-which one is right for me?" in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 1998, pp. 103–108.

[23] R. C. Elliott, R. K. Nurani, D. Gudmundsson, M. Preil, R. Nasongkhla, and J. G. Shanthikumar, "Critical dimension sample planning for sub-0.25 micron processes," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 1999, pp. 139–142.

[24] C.-F. Chien, S.-C. Hsu, S. Peng, and C.-H. Wu, "A cost-based heuristic for statistically determining sampling frequency in a wafer fab," in *Proc. Semicond. Manuf. Technol. Workshop*, 2000, pp. 217–229.

[25] C.-F. Chien, K.-H. Chang, and C.-P. Chen, "Sampling strategy and model to measure and compensate the overlay errors," in *Metrology, Inspection, and Process Control for Microlithography*, vol. 4344. Santa Clara, CA, USA: International Society for Optical Engineering, 2001, pp. 245–256.

[26] J. Shumaker, A. Phillips, and M. Lauderdale, "Intelligent sample test using cost based methodologies," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2003, pp. 439–442.

[27] C. Xuemei, M. E. Preil, M. Le Gaff-Dussable, and M. Maenhoudt, "Optimizing overlay sampling for higher yields," *Semicond. Int.*, vol. 26, no. 4, pp. 56–60, 2003.

[28] C.-W. Wu and W. L. Pearn, "A variables sampling plan based on Cpmk for product acceptance determination," *Eur. J. Oper. Res.*, vol. 184, no. 2, pp. 549–560, 2006.

[29] C. L. Kwang and O. E. Chin, "A novel push-pull sampling methodology for test production in semiconductor manufacturing industries," in *Proc. IEEE/CPMT Int. Symp. Electron. Manuf. Technol.*, 2008, pp. 1–3.

[30] E. Sachs, A. Hu, and A. Ingolfsson, "Run by Run Process Control: Combining SPC and Feedback Control," *IEEE Trans. Semicond. Manuf.*, vol. 8, no. 1, pp. 26–43, Feb. 1995.

[31] N. Jedidi, P. Sallagoity, A. Roussy, and Dauzère-Pérès, "Feed-forward run-to-run control for reduced parametric transistor variation in CMOS Logic 0.13 um Technology," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 2, pp. 273–279, May 2011.

[32] S. Prabhu, D. C. Montgomery, and G. C. Runger, "A combined adaptive sample size and sampling interval x-bar control scheme," *J. Quality Technol.*, vol. 26, no. 3, pp. 164–176, 1994.

[33] R. K. Nurani, R. Akella, A. J. Strojwas, and R. Wallace, "Role of in-line defect sampling methodology in yield management," in *Proc. Int. Symp. Semicond. Manuf.*, 1995, pp. 243–247.

[34] W. W. Kuo, A.-H. Wang, R. Akella, and R. K. Nurani, "A combined adaptive sampling strategy with limited inspection capacity," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 1996, pp. 235–238.

[35] R. Babikian and C. Engelhard, "Statistical methods for measurement reduction in semiconductor manufacturing," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 1998, pp. 212–215.

[36] R. Williams, D. Gudmundsson, R. Nurani, M. Stoller, A. Chatterjee, S. Seshadri, and J. G. Shanthikumar, "Challenging the paradigm of monitor reduction to achieve lower product costs," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 1999, pp. 420–425.

[37] R. Williams, D. Gudmundsson, K. Monahan, and J. G. Shanthikumar, "Optimized sample planning for wafer defect inspection," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 1999, pp. 43–46.

[38] R. E. Langford, G. Hsu, and C. Sun, "The identification and analysis of systematic yield loss," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2000, pp. 92–95.

[39] R. K. Nurani and J. G. Shantikumar, "Process control for items produced in lots with inter and intra lot variations," *Int. J. Ind. Eng.*, vol. 7, no. 1, pp. 57–66, 2000.

[40] J. H. Lee, S. J. Yu, and S. C. Park, "Design of intelligent data sampling methodology based on data mining," *IEEE Trans. Robot. Autom.*, vol. 17, no. 5, pp. 637–649, Oct. 2001.

[41] P. Wootton, B. Saville, A. Lutz, and J. Oakley, "Review sample shaping through the simultaneous use of multiple classification technologies in IMPACT ADC," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2001, pp. 207–217.

[42] C. Allebe, B. Govaerts, E. Van Kerschaver, S. De Wolf, and J. Szlufcik, "Control charts and efficient sampling methodologies in the field of photovoltaics," in *Proc. IEEE Photovoltaic Specialists Conf.*, 2002, pp. 387–390.

[43] J. H. Lee, "Artificial intelligence-based sampling planning system for dynamic manufacturing process," *Expert Syst. Appl.*, vol. 22, no. 2, pp. 117–133, 2002.

[44] L. Song-bor, R. L. Ta-Yung, L. Janson, and C. Yu-Ching, "A capacity-dependence dynamic sampling strategy," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2003, pp. 312–314.

[45] D. B. Sullivan, E. W. Conrad, and J. S. Smyth, "Overlay metrology sampling capability analysis and implementation in manufacturing," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2004, pp. 208–212.

[46] B. Moon, J. McNames, B. Whitefield, P. Rudolph, and J. Zola, "Wafer sampling by regression for systematic wafer variation detection," in *Proc. Data Anal. Modeling Process Control*, vol. 5755. 2005, pp. 212–221.

[47] J. G. Shanthikumar, "Effects of capture rate and its repeatability on optimal sampling requirements in semiconductor manufacturing," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2007, pp. 1–6.

[48] C. Mouli and M. J. Scott, "Adaptive metrology sampling techniques enabling higher precision in variability detection and control," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2007, pp. 12–17.

[49] B. Bunday, B. Rijpers, B. Banke, C. Archie, I. B. Peterson, V. Ukraintsev, T. Hingst, and M. Asano, "Impact of sampling on uncertainty: Semiconductor dimensional metrology applications," in *Proc. Metrol. Inspection Process Control Microlithography*, vol. 6922. 2008, pp. 1–22.

[50] V. Veetil, D. Sylvester, D. Blaauw, S. Shah, and S. Rochel, "Efficient smart sampling based full-chip leakage analysis for intra-die variation considering state dependence," in *Proc. IEEE Design Autom. Conf.*, 2009, pp. 154–159.

[51] A. Chen, S. Hsueh, and J. Blue, "Optimum sampling for track PEB CD integrated metrology," in *Proc. IEEE Int. Conf. Automation Sci. Eng.*, 2009, pp. 439–442.

[52] M. Sahnoun, P. Vialletelle, S. Bassetto, S. Bastoini, and M. Tollenaere, "Optimizing return on inspection trough defectivity smart sampling," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2010, pp. 1–4.

[53] M. Sahnoun, S. Bassetto, S. Bastoini, and P. Vialletelle, "Optimisation of the process control in a semiconductor company: Model and case study of defectivity sampling," *Int. J. Prod. Econ.*, vol. 49, no. 13, pp. 3873–3890, 2011.

[54] R. P. Good, D. Pabst, and J. B. Stirton, "Compensating for the initialization and sampling of EWMA run-to-run controlled processes," *IEEE Trans. Semicond. Manuf.*, vol. 23, no. 2, pp. 168–177, 2010.

[55] J. Nduhura Munga, S. Dauzère-Pérès, P. Vialletelle, and C. Yugma, "Dynamic management of controls in semiconductor manufacturing," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2011, pp. 1–6.

[56] C. Paccard, "Développement d'outils statistiques pour la mise en place des boucles de régulation," Ph.D. dissertation, Univ. Toulouse III-Paul Sabatier, France, 2008.

[57] M. Purdy, C. Nicksic, and K. Lensing, "Method for efficiently managing metrology queues," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2005, pp. 71–74.

[58] K. Lensing and B. Stirton, "Perspectives on integrated metrology and wafer-level control," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2007, pp. 1–5.

[59] A. Holfeld, R. Barlović, and P. Good, "A fab-wide APC sampling application," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 393–399, Nov. 2007.

[60] R. P. Good and M. A. Purdy, "An MILP approach to wafer sampling and selection," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 400–407, Nov. 2007.

[61] M. Purdy, "Dynamic, weight-based sampling algorithm," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2007, pp. 1–4.

[62] Y. Kaga, Y. Sato, Y. Yamada, Y. Yamazaki, M. Aoki, R. Harukawa, and E. Chang, "Integrated defect sampling method by using design attribute for high sensitivity inspection in 45nm production environment," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2008, pp. 379–381.

[63] S. Jansen, G. Florence, A. Perry, and S. Fox, "Utilizing design layout information to improve efficiency of SEM defect review sampling," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2008, pp. 69–71.

[64] B. Hyung Joo Lee, "Advanced process control and optimal sampling in semiconductor manufacturing," Ph.D. dissertation, Univ. Texas, Austin, TX, USA, 2008.

[65] S. Sun and K. Johnson, "Method and system for determining optimal wafer sampling in real-time inline monitoring and experimental design," in *Proc. IEEE Int. Symp. Semicond. Manuf.*, 2008, pp. 44–47.

[66] C.-T. Lin, C.-C. Huang, C.-Y. Yang, Y.-W. Wu, C.-S. Lu, P.-Y. Tsai, C.-M. Huang, and Y.-L. Wang, "Defect intelligent sampling system," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2010, pp. 162–164.

[67] S. Dauzère-Pérès, J. Rouveyrol, C. Yugma, and P. Vialletelle, "A smart sampling algorithm to minimize risk dynamically," in *Proc. IEEE/SEMI Advanced Semicond. Manuf. Conf.*, 2010, pp. 307–310.

[68] M. A. Purdy, "Dynamic metrology sampling methods, and system for performing same," U.S. Patent 2005/0 033 467, 2005.

[69] J. Nduhura Munga, S. Dauzère-Pérès, C. Yugma, and P. Vialletelle, "A mathematical programming approach for determining control plans in semiconductor manufacturing," in *Proc. Int. Conf. Ind. Eng. Syst. Manage.*, 2011, pp. 1–9.

[70] G. Rodriguez-Verjan, S. Dauzère Pérès, and J. Pinaton, "Impact of control plan design on tool risk management: A simulation study in semiconductor manufacturing," in *Proc. Modeling Anal. Semicond. Manuf. Conf.*, 2011, pp. 1918–1925.

Justin Nduhura-Munga received the Engineering degree in computer science, microelectronics, and automation from the Ecole Polytechnique de Lille, Lille, France, in 2009, and the M.S. degree in microelectronics from the University of Lille, Lille, France, in 2009. He is currently pursuing the Ph.D. degree in industrial engineering at the Ecole des Mines de Saint-Etienne, Gardanne, France. He is at STMicroelectronics, Crolles, France. His works mostly focus on implementing dynamic controls in high-mix semiconductor plants. He was a recipient of the Best Student Paper Award from the IEEE/SEMI Advanced Semiconductor Manufacturing Conference, Saratoga Springs, NY, USA, in 2011.

Gloria Rodriguez-Verjan graduated with highest honors as an Industrial Engineer from the Pontificia Universidad Javeriana, Bogotà, Colombia, and received the M.Sc. degree in industrial engineering from École Nationale Supériere des Mines de Saint-Étienne, France. She is currently pursuing the Ph.D. degree in industrial engineering at the Ecole des Mines de Saint-Etienne, Gardanne, France. She is at STMicroelectronics, Rousset, France. Her current research interests include optimization of semiconductor manufacturing operations.

Stéphane Dauzère-Pérès received the Ph.D. degree from Paul Sabatier University, Toulouse, France, in 1992, and the Habilitation à Diriger des Recherches degree from Pierre and Marie Curie University, Paris, France, in 1998. He is currently a Professor at the Center of Microelectronics in Provence of the Ecole des Mines de Saint-Etienne, where he heads the Manufacturing Sciences and Logistics Department. Since March 2004, he has been a Professor at the Ecole des Mines de Saint-Etienne. He has published more than 40 papers in international journals and 100 communications in conferences. His current research interests include optimization in production and logistics, with applications in planning, scheduling, distribution, and transportation.

Claude Yugma received the Ph.D. degree in computer science and combinatorial optimization from the Grenoble Institute of Technology, Grenoble, France. He is currently an Associate Professor at EMSE. He was involved in the project Rousset 2003–2009 and in the MEDEA+ European Project HYMNE. Currently, he is involved in the ENIAC European Project IMPROVE. His current research interests include modeling an optimizing semiconductor manufacturing operation: global and local scheduling, interactions between advanced process control and scheduling, dynamic sampling, and so on.

Philippe Vialletelle is currently a Manager of the Operations and Methods System Group, STMicroelectronics. After receiving the Engineering degree in physics, he entered the semiconductor industry working on ESD and physical characterization. He finally integrated industrial engineering and is currently responsible for the development of advanced programs for the management of Crolles 300-mm production line. At the European level, he is in charge of the definition and followup of collaborative programs in the field of manufacturing sciences such as HYMNE or IMPROVE.

Jacques Pinaton is currently a Manager of the Process Control System Group, STMicroelectronics Rousset, France. After being an Engineer in metallurgy from Conservatoire National des Arts et metiers d'Aix in Provence, France, he joined STMicroelectronics in 1984. After five years within the Process Engineering Group, he joined the Device Department to implement SPC, process control methodology, and tools. He is leading several Rousset research and development programs on manufacturing science, such as automation, APC, and diagnostic. He is involved in the ENIAC European Project IMPROVE.