The Dissertation Committee for Huidong Zhang certifies that this is the approved version of the following Dissertation:

Integrated Operational Decision-Making in Flexible Manufacturing Systems with Considerations of Quality and Reliability

Committee:

Dragan Djurdjanovic, Supervisor

J. Eric Bickel

Grani A. Hanasusanto

Maria Chiara Magnanini

S.V. Sreenivasan

Integrated Operational Decision-Making in Flexible Manufacturing Systems with Considerations of Quality and Reliability

by

Huidong Zhang

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2021

Dedicated to my family.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Dr. Dragan Djurdjanovic, who has guided me, encouraged me, and supported me through this journey. In countless communications through presentations, text messages, and phone calls, he imperceptibly helped me build up my academic prowess with his wisdom, understanding, and patience. He is not only my academic mentor but also a role model who has a profound influence on my life.

I would like to thank the other members in my committee: Dr. Eric Bickel for educating me in decision analysis and guiding me into the field of operations research as my graduate research supervisor when I was pursuing my master degree; Dr. Grani Hanasusanto, who has taught me so much about optimization under uncertainty, offered me great advice to solve complex problems and provided me feedbacks to improve the writing; Dr. Maria Chiara Magnanini and Dr. S.V. Sreenivasan for pleasant communications and meaningful insights that boosted maturity in this Ph.D. research. This dissertation would not have been possible without all of your contributions and support.

In this journey, my colleagues in Dr. Djurdjanovic's lab have offered me helps in various forms. I would like to extend my gratitude to Dr. Asad UI Haq, Dr. Deyi Zhang, Ramin Sabbagh, Zicheng Cai, Kaiwen Yang, Roberto Daily, Kevin Helm, Noah Graff,

Haihua Ou. I have benefited a lot from discussions with each of you. The productive and supportive lab environment you have created has allowed me to enjoy my time in the lab.

On a personal note, I am grateful for my friends from UT Austin, including Qianru Zhu, Chuwen Zhang, Weiming Zhang, Jia Guo, Weiyan Sun, Qiang Xie, Weigu Li. Each of you has accompanied, motivated, or helped me in a certain period in this journey, which made my life at UT wonderful and unforgettable. Outside of Austin, I would like to thank my bosom friend Qiuyu Chen, who has always been there accompanying me over the years in spite of being so far away.

Finally and ultimately, I would like to express my heartfelt gratitude to my family. My parents, Zhenxin Zhang and Yuxiang Bao, always give me their best. They care for me in every possible way and give me endless spiritual encouragement and comfort. I would not go this far without their unconditional support and love. Special thanks to my husband, Tianheng Feng. Ten years ago, we first met in SJTU. After overcoming so many difficulties, we got married in Austin while pursuing our doctoral degrees at UT. Thank you for always being there, offering constant love, encouraging me to overcome challenges, and cheering me up, while being busy as a doctoral student and a newbie at work. It is lucky to have you in my life.

Abstract

Integrated Operational Decision-Making in Flexible Manufacturing Systems with Considerations of Quality and Reliability

Huidong Zhang

The University of Texas at Austin, 2021

Supervisor: Dragan Djurdjanovic

Inherent interactions between operational decisions and their impact on the product quality and equipment reliability necessitate those decisions being optimized concurrently, especially in highly flexible manufacturing systems, such as semiconductor manufacturing fabs, where those interactions are even stronger and where flexibility allows for such decisions to be implemented. This dissertation addresses multiple challenges associated with such integrated operational decisions in the manufacturing process with considerations of quality and reliability.

The first part of this dissertation proposes an integrated decision-making policy for production and maintenance operations on a single machine under uncertain demand, with concurrent considerations and learning of yield dependencies on the equipment conditions

and production rates. This policy is obtained through a two-stage stochastic programming model, which considers the variable demand, machine degradation, and maintenance times. This model incorporates outsourcing decisions and operational decisions regarding reworking, scraping of imperfect products to ensure the demand is adequately met. A closed-form reinforcement learning method is utilized to learn yield dependencies. Simulations confirm the necessity of yield learning and show the proposed method outperforms the traditional, fragmented approaches where the effects of production rates and machine conditions on the resulting yield rates are not considered.

In the second part of this dissertation, a novel optimization framework that couples a recently introduced approach for robust control of overlay errors in photolithography processes with a strategic selection of overlay measurement markers to enable improved control of overlay errors using a reduced number of measurements. Application of this method to the data and models from an industrial-scale semiconductor lithography process shows that the newly proposed combination of the robust overlay control paradigm and optimized marker selection enables improved overlay control, even with a significantly reduced number of markers. Thus, the new methodology enables the reduction of measurement times and subsequent overall cycle time, without deteriorating the outgoing product quality.

The measurement selection method suggested in the second part of this dissertation pursues optimality from a purely quality control aspect. Since any measurement down-selection procedure directly affects cycle-times of the resulting process and the

understanding of yield rate behaviors, the final portion of this dissertation tackles the task of developing an optimization framework from a more system-level operational aspect. In this method, an optimal number of markers is decided that maximizes the profit considering revenue earned from perfect layer patterns, cost of misidentified bad layers, as well as production and measurement cost. At the same time, the distribution of those markers is optimized considering one's ability to estimate actuator uncertainties and to understand the yield rate behavior. Application of this method to the data and models from an industrial-scale semiconductor lithography process, as well as sensitivity analyses, are presented to illustrate the proposed method and evaluate the effects of a variety of relevant parameters on the profit of the system.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1. Motivation and Background

Flexible manufacturing systems (FMS) are designed to adapt to changes in the type and quantity of the product being manufactured to get improved productivity and quality of the product [1]. Problems in an FMS life cycle can be broadly classified into two areas: design problems and operation problems. At the design stage, one is interested in specifying the system so that the desired performance goals are achieved. On the other hand, the operation problems are aimed at making decisions related to planning, scheduling, and control of a given FMS [2].

To improve performance of manufacturing systems, operational decisions are usually made in terms of a variety of areas, such as quality management, process, and capacity design, location and logistic design, inventory control, scheduling, maintenance, etc. [3]. In FMS, operational decisions are considerably harder to make, as high flexibility and complexity of the system lead to a dramatically bigger and more complex decision spaces [4].

For example, in such systems, the machines have the capability of conducting different manufacturing operations or producing at various speeds, while degradation of machine depends highly on these operations performed on that machine. Degradation of machine conditions not only affects the system reliability and machine availability, in some

cases, it also affects the quality of the product. For example, in the etch processes in semiconductor manufacturing, the likelihood of a defect can significantly increase with both the increasing processing speed and a deteriorating equipment condition [5][6]. On the other hand, the quality of the product has direct impact on the ability of the decision-maker to fulfil the production demands, which is closely related to the inventory problem. At the same time, to deal with machine degradation, maintenance actions, especially preventive maintenance actions, interrupt production process and change the machine condition, which directly impacts the system reliability, machine availability, and product quality. This in turn affects decisions as to what operations and producing speed should be chosen.

For the quality control aspect point of view, as the FMS are designed to adjust to various type of products, the inspection equipment, which is one of the essential components of FMS, needs to be able to carry out product inspection in the most efficient way. Analysis of sensing and metrology data, which provides information about product quality and machine reliability, is essential for control, optimization, and mathematical parameters estimation in the decision-making process of manufacturing systems [7], especially during the launch and ramp-up period of manufacturing lines. Different sets of measurements carry different amounts of information about the underlying product and process. Therefore, the quality of operational decisions depends strongly on the measurement schemes, which places a great significance on finding the most informative set of measurements. Techniques for optimal selection of measurement or sensor positions

have been studied in many areas, such as machine vision, Coordinate Measurement Machine (CMM) measurements and multi-station machining systems [8].

In summary, inherent interactions between operational decisions and their impact on the product quality and equipment reliability necessitate concurrent optimization of those decisions. However, in existing literature, the above-mentioned problems have not been sufficiently addressed, and there is a need for an integrated decision support tool to jointly consider its impact on product quality and equipment reliability, especially in highly flexible manufacturing systems, such as semiconductor manufacturing fabs, where those interactions are even more substantial and where flexibility allows for such decisions to be implemented.

1.2. **Objective of this Research**

The main objective of this doctoral research is to devise integrated operational decision-making policies in the FMS with considerations of impacts of those decisions on product quality and machine reliability. Those operational decisions will be pursued under model or parameter uncertainties, with the objectives of maximizing a measure of product quality or a customizable profit function with respect to operation-dependent degradation models and production targets.

The contributions of this work can be summarized as follows:

1. An integrated decision-making policy in terms of production and maintenance planning with concurrent learning of dependencies of the yield rate on the underlying equipment conditions and production rates. The novel decision-making policy addresses demand uncertainties by optimizing decisions regarding excess production above the observed demand, or meeting demand via outsourcing or obtaining additional products from the inventory.

2. A combinatorial decision-making problem which aims to find the optimal layout of measurement markers across the wafer for the robust control of both layer-to-layer and stack-up overlay errors in photolithography processes. It can deal with non-Gaussian uncertainties and uncertainties whose characteristics cannot be accurately modeled.

3. An optimization framework for the dynamic decision-making on the number and selection of measurement markers in photolithography processes. It aims to maximize the profit considering revenue earned from perfect products, cost of misidentified imperfect products, as well as production and measurement cost. At the same time, the distribution of those markers is optimized considering one's ability to estimate actuator uncertainties for the control of overlay errors and to understand the yield rate behavior.

The challenges in achieving the corresponding contributions can be summarized as follows:

1. With the target of dealing with realistic systems, the decision-making policy should be designed generically enough to meticulously consider the stochastic effects associated with operational decisions and one's ability to fulfill the stochastic variable demand. As the machine conditions are not observable during the production period, the machine degradation processes need to be modeled carefully and adequately to enable learning of the dependence of yield rate on production rates and machine conditions through periodic review of the machine condition.

2. For a given number of measurement markers, to find their optimal layout that facilitates control of lithography errors across the whole wafer, the set of selected measurement markers and the resulting control algorithm needs to be able to control the performance of lithography errors across the whole wafer. When uncertainties in the model of overlay errors are non-Gaussian or cannot be accurately modeled, the worst-case performance needs to be considered, which adds another level of complexity in solving the decision-making problem.

3. In addition to the quality control, the number and layout of measurement markers directly affect the cycle time of the resulting process and the understanding of yield rate behaviors. This adds a new level of decision and more interactions to the decision-making process. Therefore, there is a tremendous challenge in developing a mathematical model that can quantify the influence of the number and layout of measurement markers on the profit of the manufacturing system.

1.3. **Outline of This Dissertation**

The rest of this doctoral thesis is organized as follows. Chapter 2 presents a review of the literature relevant to the proposed research. In Chapter 3, a two-stage stochastic programming problem is established for integrated production and maintenance planning under uncertain demand. Chapter 4 describes a novel optimization framework that couples a recently introduced approach for robust control of overlay errors in photolithography processes with a strategic selection of overlay measurement markers to enable improved control of overlay errors using a reduced number of measurements. In Chapter 5, a novel optimization framework is proposed for the dynamic decision-making on the number and selection of measurement markers in photolithography processes to maximize the profit of the manufacturing system considering one's ability to estimate the yield rate behavior and actuator uncertainties for the control of overlay errors. Finally, Chapter 6 details the conclusions, scientific contributions of the doctoral research, along with past and foreseen publications expected to emerge from this doctoral research.

Chapter 2. Literature Review

In the first part of this dissertation, an integrated decision-making policy will be proposed for production and maintenance operations on a single machine under uncertain demand, with concurrent considerations and learning of yield dependencies on the equipment conditions and production rates. The first objective is to study integrated maintenance and production planning with considerations of yield rate. Therefore a review of related literature is given in Section 2.1. The second objective is to illustrate methods for the learning of yield rate. To address this, Thompson Sampling and alternative algorithms for online decision problems are reviewed in Section 2.2.

In the second and third parts of this dissertation, optimal selection of measurement markers for the control of overlay errors will be studied. Therefore, literature on the optimal layout of measurements of quality characteristics of the product in manufacturing systems is reviewed in Section 2.3. Specifically for the control of overlay errors in photolithography process, existing literature for the optimal layout of measurement markers is reviewed in Section 2.4.

2.1. **Integrated Maintenance and Production Planning with Considerations of Yield Rate**

Joint considerations of yield dependencies on both the equipment conditions and production rates are essential for improving the efficiency of manufacturing operations. Although recent research recognizes that integrated production and maintenance decisions can provide benefits compared to the traditional, fragmented approaches [9], most models available in the literature do not link the effects of the integrated production and maintenance decisions to yield rates.

The limited literature that considers yield in that respect mainly focuses on the dependences from only one perspective, i.e. production rate only, or equipment condition only. For example, in [9] and [11], only the adverse impact of the increasing production rates on yield rates was considered, while the impact of machine condition on the yield was not considered. Similarly, in [12] and [13] the authors only consider models that solely take into account the effects of equipment condition on the yield rates, without concurrently considering the effects of production rates.

Research that jointly considers the impact of production rates and equipment conditions on the yield rates is rare and can be found in [14] and [15], where a Markov and a semi-Markov decision process are respectively developed, with the transition probabilities being modeled as functions of yield rates that depend on the machine's degradation state and product type. Furthermore, in [16], Batun and Maillart reassessed

these models and presented further evidence of significant potential benefits that could be obtained through integrated maintenance and production decision-making that concurrently considers the impact those maintenance and production decisions have on yield rates.

Within the relatively limited literature that describes integrated maintenance and production decision-making frameworks with joint considerations of the impacts of maintenance and production decisions on the yield rates, research that addresses the uncertain nature of the underlying demand is even harder to find. In [17], a Markov decision process based approach is utilized to formulate a production and maintenance model of a manufacturing system for which the underlying demand is random and in which periodic reviews of the equipment condition and yield in the system take place. It assumes that repairing the machine instantaneously returns it to the perfect working condition, where the maintenance time is not considered. The other limitation of this model is that it is only suitable for a small number of decision alternatives because of the curse of dimensionality. In addition, it only considers yield dependencies on the equipment conditions and does not consider the yield rate dependency on production rates.

In order to deal with problems involving a large number of decision alternatives, a novel two-stage stochastic program model with recourse is proposed in [5]. Nevertheless, it also oversimplifies the production process model in the sense that the maintenance time is assumed to be negligible. On the other hand, in many real-world production environments, maintenance time is not negligible, as maintenance downtime may result in

shutdowns that interrupt the entire production process, which in turn incurs inevitable costs in efficiency. In addition, maintenance activities are usually unique such that they commonly involve unexpected needs that make their duration highly uncertain. Therefore, an integrated decision-making model that is able to consider uncertain demands and deal with a large number of decision alternatives while acknowledging equipment degradation and uncertain maintenance times is needed.

2.2. **Thompson Sampling and Alternative Approaches for Online Decision Problems**

The issue of learning of yield rate distribution from the data collected in the decision-making process is similar to multi-armed bandit problem with a finite number of independent actions, where one cannot observe the random variables directly but can only observe them depending on the previous decisions.

Bandit problems were introduced by William R. Thompson in an article published in 1933 [18], where Thompon Sampling - also known as posterior sampling and probability matching - was first proposed. It is an algorithm for online decision problems where actions are taken sequentially in a manner that must balance between exploiting what is known to maximize immediate performance and investing to accumulate new information that may improve future performance [19]. As information is gathered, beliefs about action rewards are carefully tracked. By sampling actions according to the posterior probability that they

are optimal, the algorithm continues to sample all actions that could plausibly be optimal while shifting sampling away from those that are unlikely to be optimal. Roughly speaking, the algorithm tries all promising actions, while gradually discarding those that are believed to underperform. The intuition is formalized in recent theoretical analyses reviewed in [20].

Thompson Sampling algorithm was largely ignored in the academic literature until recently, because of the need for highly sophisticated computer technology and data processing capabilities which become available only in recent years. This was spurred partly by two influential articles [21][22] that displayed the algorithm's strong empirical performance. In the subsequent seven years, the literature on Thompson Sampling has grown rapidly. Adaptations of Thompson sampling have now been successfully applied in a wide variety of domains, including revenue management, marketing, website optimization, Monte Carlo tree search, internet advertising, etc. [20].

For multi-armed bandit problems with a finite number of independent actions, when the objective is to maximize the expected discounted reward, the Gittins index theorem [23][25] characterizes an optimal strategy. It is implemented by solving a dynamic program for action in each period, as explained in [24]. Therefore, it is computationally onerous relative to Thompson Sampling, especially for complicated problems [25].

Upper-confidence-bound algorithms are similar to Thompson Sampling, where they continue sampling promising actions while gradually discarding those that underperform [27]. However, it is often challenging to design an upper confidence bounds algorithm that simultaneously accommodates both statistical and computational efficiency, which leads

to lower statistical efficiency as compared with Thompson Sampling [28][29]. The other direction of approaches in current literature is to focus on carefully assessing the value of information, for example, Information-directed Sampling [30] and Knowledge Gradient Algorithm [31][32], which we will not review here and a thorough coverage can be seen in [20].

2.3. **Optimal Layout of Quality Measurements in Manufacturing systems**

In manufacturing systems, sensors and inspection systems are widely implemented for root cause identification and product quality control. Those sensors and inspection systems obtain data of critical indicators of system status, operation performance, and product quality [33]. Different types, numbers, and spatial combinations of sensor networks provide dense and sufficient data, which produces a comprehensive description of the dynamic changes of manufacturing systems. Due to resource and space limitations, sensors cannot be installed to monitor every feature of the product [34].

Although adding a large number of sensors to a manufacturing system is helpful in terms of improving the diagnostic and monitoring capacity of the system and can effectively reduce the loss of information, adding sensors and inspection increases the inspection costs and duration [35][36]. Furthermore, evidence shows that a highly redundant sensor distribution is not conducive to improving the monitoring and diagnostic

capabilities of the system [37]-[39]. In addition, massive data curation and transmission demands higher transmission bandwidth and significantly increases the cost of data analysis and processing, especially for remote control and wireless sensor networks [40][41].

Therefore, optimal layout of informative measurements is crucial for enabling test and accurate identification of root course of quality problems in many systems and their subsequent removal, as poor distribution of measurements often produces large amounts of conflicting and vague data [36]. In addition, in a flexible manufacturing system (FMS), the system is designed to adjust to various types of products. Therefore, the root course identification and removal need to be done in the most efficient way for different types of products using sensors and measurements which enable these functions in the most cost-effective and timely manner.

The problem pertaining to the optimal layout of quality measurements in manufacturing systems are extensively studied in the current literature [42]-[49], which generally consists of the following four aspects [37][38][39]:

(1) Model the cause-effect relationship of fault variations on sensor measurements.

(2) Set up the objective function for sensor deployment based on the cause-effect relationship.

(3) Find approaches to optimize the measurement layout strategy.

(4) Evaluate the optimized strategy.

Based on this, in the remainder of this section, a literature review will be presented from two perspectives, the modeling of the cause-effect relationship of process variations on measurements and the optimization problem for the selection of measurements.

2.3.1. **Model the Cause-Effect Relationships between Process Variations and Measurements**

The foundation of the optimal measurement layout problem is the model of cause-effect relationships between process variations and measurements [51]. Many mathematical models have been developed by researchers in various application areas. In order to establish an analytical connection between manufacturing process parameters and errors in product quality in autobody assembly [52][53] and machining of prismatic parts [54]-[57], significant research efforts were directed towards devising explicit models of the flow of dimensional errors in the mid- and late 1990s. These models are commonly referred to as Stem of Variation (SoV) models, which have been used extensively for isolating and identifying the sources of dimensional errors in manufacturing systems based on the distributed measurements of the product [58]-[61]. Based on the amount of information measurements carry about the process-level root causes of quality problems, the SoV models enable the evaluation of the “informative-ness” of any set of measurements [61]-

[65]. Using this ability, research was conducted on optimal allocation of measurements based on the maximization of the information contained in these measurements [74][75].

In addition, the development of networked control in manufacturing [76] and controllable tolling with programable parameters [77][78] signifies that systematic methods are needed to improve product quality through strategical utilization of sensing and actuation capabilities. This opportunity has been noticed by several researchers. In [79], the author proposed a run-to-run control strategy based on the well-known dead-band adjustment concept from automatic process control. A state-space model was used to decompose the necessary process action onto a set of equivalent fixture parameters. Case studies were conducted to demonstrate the model advantage in measurement reduction for root cause identification. In [80], a distributed stochastic feed-forward control method is devised to optimally reduce the variations in dimensional workpiece quality with a limited number of controllable tooling components and measurements distributed across a multistage manufacturing process. A reactive tabu search algorithm is proposed to enable joint optimal allocation of measurement points of controllable tooling devices.

2.3.2. **Optimization Problem for Spatial Selection of Measurement Markers**

In different manufacturing systems, the purpose of quality measurements varies and considerations when determining the layout of those measurements are also different.

Therefore, the objective and constraints of optimization problems formulated for the selection of measurement markers also shows diversity.

In initial optimal measurement layout researches, authors mainly focus on methods with a single objective function, which primarily aim at optimizing the cost of the measurement sensors, the system reliability, the fault unobservability, and system diagnosability [42], [66]-[69]. Bhushan et al. [66] proposed a reliability formulation that aims to maximize the minimum reliability of the system while considering other quantitative information in the constraints, including sensor costs, fault occurrence probabilities, and sensor failure probabilities. Yang et al. [67] proposed an optimization problem, where the objective is to minimize all the fault undetectability in the system, and the constraints are the false alarm probability and the cost limit. Khan et al. proposed [68] optimization frameworks that aims to maximize the diagnosability, which is quantified using Euclidean distances between pairs of diagnostic vectors.

The optimal measurement layout for hybrid sensor networks has been investigated, as in some manufacturing process, optimal hybrid sensor networks can be cheaper, more robust and accurate, as compared with networks with only homogeneous sensors. Costiner et al. [70] described an approach that aims at finding the optimal number and location of different types of sensors by maximizing their combined probability of detecting a set of cracks. This approach can be extended to any number or type of sensors, where the probability of detection of the system needs to be optimized.

In order to consider more aspects of manufacturing systems, optimization problems are investigated with considerations of multiple objective functions simultaneously or sequentially [50], [71]-[73]. In [50], Bhushan et al. proposed several formulations for the optimal measurement layout problem based on various diagnosability and reliability criteria. Firstly, a maximum-reliability model and a minimum-cost model are proposed respectively for an optimal sensor network design. Then, a one-step optimization formulation is proposed with reliability as the primary objective and cost as the secondary objective. In addition, the proposed formulations are also able to handle various constraints regarding fault diagnosis and variable measurements, for example, different types of sensors may have different availabilities, a particular failure may need to be observed by more than a given number of sensors. This research addresses the necessity of developing methods that is able to seamlessly incorporate various criteria such that the location of sensors is optimized in a broader sense.

Although objective functions vary according to their different interests, nearly all the current literature for the problem of optimal measurement layout has integer decision variables in the formulation of their optimization problems. Therefore, most of the formulated problems are attributed to integer or mixed-integer programming problems, which are usually NP-hard problems. A variety of optimization algorithms have been utilized, aiming to solve those problems efficiently.

In [81]-[83], the authors proposed to use finite element method based simulation methods to solve their nonlinear and integer programming problems effectively. Exchange

algorithms [44][84][85] and gradient-based search methods [42][68] have been utilized to search for the optimal measurement layout. However, the aforementioned methods sometimes cannot be easily implemented in the actual manufacturing process and suffer great computational complexity. Therefore, heuristic algorithms are widely utilized for the optimal measurement layout problem, although they might not always reach the optimal solution. In [46], Li and Jin proposed an integrated algorithm based on the greedy search algorithm to search for the optimal measurement layout that maximizes the abnormalities detection. Although the greedy search algorithm is good at effectively discovering good feasible solutions, its capability of solving large-scale problems is limited. At the same time, in actual manufacturing systems, the number of candidate measurements may be vast and the computational efficiency determines whether a selection strategy can be effectively applied. Therefore, compared with greedy search algorithms, various types of evolutionary algorithms are more widely deployed to achieve a better solution more efficiently, as they have a strong ability to optimize in a high-dimensional search space. Among them, the most commonly used methodological paradigm is the genetic algorithm (GA), especially for complex manufacturing systems [86]-[90]. In [45], an integrated data-mining guided GA has been investigated to solve the optimal sensor distribution problem with the objective of maximizing the variance detection capability. In [92], an improved particle swarm optimization (PSO) algorithm has been proposed to search for the optimal acoustic emission location.

Although extensive research has been done to determine the optimal layout of quality measurements, there are still opportunities for improvement. Firstly, in the most current literature, spatial layout of measurements is usually predetermined. An adaptive strategy that is able to adapt the selection of measurement locations based on the current system state is needed.

Secondly, in many manufacturing systems, the quantity of measurements of product quality directly affects cycle-times of the manufacturing process. There is a trade-off between measurement effort and efficiency of the manufacturing system and in current literature, strategies for the optimal layout of measurements that are obtained from a system-level perspective have rarely been explored. Explorations and optimization of these trade-offs could bring tremendous benefits, especially in highly complex FMS, such as photolithography.

2.4. **Optimal Layout of Measurement Markers for Overlay Errors in Photolithography**

In the second and third parts of this dissertation, problems of optimal layout of quality measurements for the control of overlay errors in lithography process will be studied. Therefore, a review is given to the literature on optimal distribution of measurement markers for this specific area.

For the measure of overlay errors in lithography process, related literature has proposed several strategies for the sampling of measurement markers. Early studies obtained the layout of measurement points based on expert knowledge. In [93], the authors looked at nine different sampling plans and demonstrated that the layout of measurement points could have a significant effect on overlay errors. In [94], inter-field and intra-field sampling plans are generated to respectively consider factors that affect the inter-field and intra-field overlay errors. In [95], four types of geometry-based layouts - radial, circular, symmetrically random, and spatially optimized distributions - were developed and compared. They found that the layout with spatially optimized distribution, which intends to maximize the distance between any two selected fields, can offer the best control efficiency.

Therefore, rule-based approaches, which aim to distribute measurement points evenly across the wafer, were preferred as an alternative. In [96], a cost-effective rule-based sparse sampling method is proposed that satisfies three rules: (i) homogeneous distribution of measurement points, (ii) equal number of points in scan up and scan down condition and (iii) equal number of points in each field. In [97], a smart sampling scheme is developed, which primarily uses a rule-based field-by-filed sampling approach to achieve maximized spatial uniformity.

However, none of these methods use a systematic search method to maximize the information content in the measurements. Recent research [98] is the first to utilize a metaheuristic-based optimization method to improve the layout of measurement points.

The authors proposed a sparse particle swarm optimization algorithm to search for the layout of measurement points [99], which is a popular metaheuristic method in semiconductor manufacturing [100][101]. The algorithm was designed to consider the positions of points and maximize the sparsity during the iterative solution generation process. However, it did not consider the stochastic nature of the overlay model and the authors assumed all fields of the wafer have the same layout to make sure measurement points are evenly distributed in the entire wafer. Additionally, their selection method did not consider the control of stack-up errors. Therefore, there is still a great need for an optimization method for strategic selection of overlay measurement points across the wafer, which will enable optimal control of both layer-to-layer and stack-up overlay errors, with considerations of the stochastic nature of the overlay phenomenon and the inherently uncertain character of the underlying model parameters.

Chapter 3. Integrated Production and Maintenance Planning under Uncertain Demand with Concurrent Learning of Yield Rate[1]

3.1. Introduction

In many manufacturing systems, the machines have the capability of conducting different manufacturing operations or producing at various speeds, while machine degradation depends highly on these operations performed on that machine. Degradation of machine conditions not only affects the system reliability and machine availability, in some cases, it also affects the quality of the product. On the other hand, the quality of the product has a direct impact on the ability of the decision-maker to fulfill the production demands, which is closely related to the inventory problem. At the same time, in order to deal with machine degradation, maintenance actions, including preventive and reactive maintenance actions, interrupt production process and change the machine condition, which has a direct impact on the system reliability, machine availability, and product quality. This in turn affects decisions as to what operations and producing speed should be

[1] The work presented in this chapter is based on the publication: Zhang, H. and Djurdjanovic, D., 2021. "Integrated production and maintenance planning under uncertain demand with concurrent learning of yield rate," *Flexible Services and Manufacturing Journal*, pp.1-22.
Huidong Zhang wrote this publication under the supervision of Dragan Djurdjanovic.

chosen. Therefore, it is necessary and challenging to use information about a machine's degradation level to make integrated production and maintenance decisions.

Yield rate is one of the most important measures of manufacturing performance in many industries [14][102]. In terms of the operational decision-making in manufacturing systems, it has been well recognized that yield rate can be affected by deteriorating machine conditions [12][13]. However, in many manufacturing environments, yield rate can also be adversely affected by increasing production rates. For example, in the etch processes in semiconductor manufacturing, the likelihood of a defect can significantly increase with both the increasing processing speed and a deteriorating equipment condition [5][6]. Consequently, joint considerations of yield dependencies on both the equipment conditions and production rates are essential for improving the efficiency of manufacturing operations. Although recent research recognizes that integrated production and maintenance decisions can provide benefits compared to the traditional, fragmented approaches [9], most models available in the literature do not link the effects of the integrated production and maintenance decisions to yield rates. Research that jointly considers the impact of production rates and equipment conditions on the yield rates is rare [14].

One of the most direct impacts of understanding the behavior of yield rates within an integrated maintenance and production decision-making framework is improving the ability of the decision-maker to fulfil the production demands. This is especially important if one takes into consideration the fact that the production demand is often hard to predict

and is inherently uncertain. Within the relatively limited literature that describes integrated maintenance and production decision-making frameworks with joint considerations of the impacts of maintenance and production decisions on the yield rates, research that addresses the uncertain nature of the underlying demand, without oversimplifying the production process model, is even harder to find [17][5]. Therefore, an integrated decision-making model that is able to consider uncertain demands and deal with a large number of decision alternatives, while acknowledging equipment degradation and non-negligible maintenance times is needed.

The model relating the equipment conditions and production rates with the outgoing yield is inherently unknown and needs to be learned during the production process. This is especially impacting during the ramp-up period of production, when very little historical information about the underlying process exists. Consequently, the need for understanding of the yield rate behaviors, especially during the ramp up production periods, is widely accepted in the existing research [17][104]. However, most of it focuses on modeling of yield dynamics over time rather than possibilities to learn yield models within the very process of decision-making.

In summary, there is a great need to develop a formal and systematic method to incorporate the learning of yield rates into the integrated production and maintenance decision-making that takes into account the decision-maker's ability to fulfil a stochastically variable demand. This model should be able to consider stochastic nature of the machine degradation processes, maintenance events and production demand, along

with learning and updating within the decision-making process itself of the knowledge of yield rates as they depend on the equipment conditions and production rates. In order to solve the above-described problems, this chapter develops a two-stage stochastic programming model for the integrated decision-making problem that considers operational decisions in terms of preventive and reactive repairs, production rates, reworking or scraping of imperfect products, as well as decisions regarding overproduction, outsourcing to obtain extra products from the inventory to satisfy the demand, while in parallel learning online the dependency of yield rates on the underlying equipment conditions and production rates.

The remainder of this chapter is organized as follows. Section 3.2 presents the mathematical formulation of the two-stage stochastic programming model, a Markov-Chain-Monte-Carlo-simulation-based approach that is used for solving this model in an efficient way, as well as a closed-form Beta-Bernoulli model-based method for reinforcement learning of yield rates within the decision-making process. In Section 3.3, results of numerical experiments based on discrete-event simulations of a manufacturing system are presented and discussed. Finally, a summary of conclusions drawn from the work presented in this chapter and possible directions for future work can be found in Section 3.4.

3.2. **Proposed Decision-Making and Learning Framework**

To achieve the maximum total expected profit for each decision epoch of the integrated production and maintenance decision-making under uncertain demand, we propose a two-stage stochastic programming model, which can be seen as an extension of the model proposed in [5]. At the beginning of each decision epoch, first-stage decisions regarding the production rate and maintenance are made based on the observed machine condition. At the end of the epoch, i.e. after the production for this epoch is finished and its results could be observed, second-stage decisions regarding the production outsourcing or storage of excess production are made based on the actual demand for the product during this epoch, and the observed number of good and faulty products produced during that epoch.

Relevant notation, formal description of the above described two-stage optimization model and the approach used to solve this problem are respectively provided in Sections 3.1-3.3. Section 3.4 describes a simplified benchmark model in which all uncertainties are modeled as their expected values. Performance of this simplified model will be compared to results obtained using the newly proposed approach described in Sections 3.1-3.3. Section 3.5 presents the method for learning of yield rates with the proposed decision-making model embedded in it.

3.2.1. **Notations**

Throughout this chapter, we use bold letters to denote vectors and matrices. For any natural number i, we define $[i]$ as the set $\{1,...,i\}$, and we use $\mathbb{Z}_2$ to denote binary set $\{0,1\}$. All random variables will be denoted by a symbol ξ with appropriate subscripts. We will now define below the main portions of the notation used in the chapter, while some additional parameters and variables will be introduced as needed.

Parameters

T	Number of time units for production in one decision epoch
N_w	Number of degradation levels for the machine condition
N_p	Upper bound of the production rate, i.e. maximum number of products produced per production time unit.
N_m	Upper bound of the uncertain maintenance times
c	(with various subscripts) unit costs corresponding to each product, maintenance action and time unit of production
p	(with various subscripts) revenues corresponding to each product

Decision Variables

$\mathbf{x}$	Vector of binary decision variables $\mathbf{x} \in \mathbb{Z}_2^{N_p}$, where each entry $x_q, \forall q \in [N_p]$ denotes if the production rate q is chosen, i.e. $x_q = 1$, or not.
x_{PM}	Binary decision variable denoting whether or not to conduct preventive maintenance (PM) at the beginning of the decision epoch. It is assumed that machine is restored to the perfect condition after each PM.

x_{RMO}	Binary decision variable denoting whether to conduct a perfect or a minimum reactive maintenance if the machine fails. Specifically, $x_{RMO}=1$ indicates that a perfect reactive maintenance (PRM) option is chosen, which means that RM operation restores the machine to the perfect condition after the repair, while $x_{RMO}=0$ indicates that a minimal reactive maintenance (MRM) is chosen, which means that the RM operation brings the machine condition to the condition observed at the beginning of the relevant decision epoch.
y_{out}	Number of products that need to be bought from outside[2] (or can be provided from the inventory) to meet the demand
$y_{backorder}$	Number of products on backorder
y_{over}	Overproduction that can be sold to others (or can be stored to the inventory)
y_{wasted}	Number of products that are wasted

3.2.2. **Modeling of Machine Deterioration and Maintenance Time using Discrete Time Markov Chain (DTMC)**

In this chapter, machine degradation will be modeled as a unidirectional discrete time Markov chain (DTMC), while repair times will be assumed to be random and following a

[2] I.e. outsourced

known discrete distribution on the support set $[N_m]$. There is abundant research that deals with systems undergoing Markovian deterioration [105]. However, a few studies have considered optimal production and maintenance decisions, with the uncertain maintenance times also modeled using the Markovian framework [13]. In order to model the randomness of maintenance times, we augment the state-space of machine degradation levels $\Xi_w = \{W_1, W_2, ..., W_{N_w}\}$ with the set of maintenance states $\Xi_m = \{M_{N_m}, M_{N_m-1}, ..., M_2, M_1\}$, where the state $W_n, n \in [N_w]$ means the machine is in a working condition and the degradation level is n (W_1 indicates the perfect working condition), while the state $M_n, n \in [N_m]$ means the machine is being maintained and the remaining maintenance time on it is n.

In this framework, let S_t be the state of the machine at production time unit t. The sequence of machine states $\{S_t, t \geq 0\}$ can be seen as a DTMC on the state-space $\Xi_S = \{W_1, W_2, ..., W_{N_w}, M_{N_m}, ..., M_2, M_1\}$ with the transition probability matrix

$$\mathbf{P_{trans}}(\mathbf{x}, x_{PM}, x_{RMO}) = \left[\begin{array}{c|c} \mathbf{P_D}(\mathbf{x}) & \mathbf{P_F}(\mathbf{x}, x_{RMO}) \\ \hdashline \mathbf{P_R}(x_{PM}, x_{RMO}) & \mathbf{P_M} \end{array}\right]$$

consisting of four portions $\mathbf{P_D}(\mathbf{x})$, $\mathbf{P_F}(\mathbf{x}, x_{RMO})$, $\mathbf{P_M}$ and $\mathbf{P_R}(x_{PM}, x_{RMO})$ described below.

- *Degradation Transition Submatrix* $\mathbf{P_D}(\mathbf{x})$ *:* If the machine is at degradation level $W_i, i \in [N_w]$ and the production rate q is chosen (in the binary decision vector $\mathbf{x}$,

the entry $x_q = 1$, while others are zero), the probability that it transits to a degradation level W_j, $j \in \{i, i+1, ..., N_w\}$ in the next production time unit[3] is assumed to be known and denoted by $p_{i,j}^{\deg}(\mathbf{x})$. Thus, the portion of the Markovian transition probability matrix that describes machine degradation is given by

$$\left[\mathbf{P_D}(\mathbf{x})\right]_{i,j} = \Pr(S_t = W_j \mid S_{t-1} = W_i, \mathbf{x}) = \begin{cases} p_{i,j}^{\deg}(\mathbf{x}), i \in [N_w],\ j \in \{i, i+1, ..., N_w\} \\ 0 \quad\quad , i \in \{2, 3, ..., N_w\},\ j \in [i-1] \end{cases}$$

Please note that we assume $\sum_{j=k}^{N_w} p_{i,j}^{\deg}(\mathbf{x}) \le \sum_{j=k}^{N_w} p_{i+1,j}^{\deg}(\mathbf{x}), i \in [N_w - 1], k \in \{i, i+1, ..., N_w\}$ for all $\mathbf{x}$, which implies intuitive notion of increasing failure rate (IFR) [106]. Also, it is intuitive to assume that probabilities that the machine transits to worse degradation levels increase as the production rate increases.

- *Failure Transition Submatrix* $\mathbf{P_F}(\mathbf{x}, x_{RMO})$*:* Let $F_i(\mathbf{x})$ be the probability that the machine breaks down at working condition W_i, when it is working under the production rate defined by $\mathbf{x}$. Let us also assume that maintenance times are described by a known probability mass function $p_j^M(x_{RMO})$, $j \in [N_m]$, which is affected by the decision x_{RMO} in the sense that maintenance times for the PRM and MRM follow

[3] Note, degradation is modeled as a unidirectional process and hence working condition of a machine can only remain the same or deteriorate.

different distributions p_j^{PRM} and p_j^{MRM}, $j \in [N_m]$, respectively. Then, the Failure Transition Submatrix $\mathbf{P_F}(\mathbf{x}, x_{RMO})$ in the Markovian transition matrix $\mathbf{P_{trans}}(\mathbf{x}, x_{PM}, x_{RMO})$, which consists of the probabilities that the machine transits from a working condition W_i to a maintenance state M_j, can be described as

$$\left[\mathbf{P_F}(\mathbf{x}, x_{RMO})\right]_{i,j} = \Pr(S_t = M_j \mid S_{t-1} = W_i, \mathbf{x}, x_{RMO}) = F_i(\mathbf{x}) p_j^M(x_{RMO}), i \in [N_w], j \in [N_m]$$

Please note that we must have

$$F_i(\mathbf{x}) + \sum_{j=i}^{j=N_w} p_{i,j}^{\deg}(\mathbf{x}) = 1, i \in [N_w]$$

for all $\mathbf{x}$, which ensures that without maintenance actions, the machine can only remain in its current degradation level, deteriorate, or break down.

- *Maintenance Transition Submatrix* $\mathbf{P_M}$: Once the machine enters a maintenance state, the remaining maintenance times will decrease as time passes. Thus, the maintenance portion of the Markovian probability transition matrix is given by

$$\left[\mathbf{P_M}\right]_{i,j} = \Pr(S_t = M_j \mid S_{t-1} = M_i) = \begin{cases} 1, & j = i-1 \\ 0, & otherwise \end{cases}, i \in [N_m], j \in [N_m]$$

- *Recovery Transition Submatrix* $\mathbf{P_R}(x_{PM}, x_{RMO})$ *:* When the remaining maintenance time reaches 1, i.e. after the maintenance operation is finished, the machine is restored in the next production time unit to the working condition determined by decision variables x_{PM} and x_{RMO}. If a decision to perform PM is selected, i.e. if $x_{PM} = 1$, the

decision period will start in a maintenance state, where the maintenance time ξ_{PM} follows the corresponding discrete distribution of PM times. After the PM is finished, machine will be restored to the perfect working condition before the machine starts working[4]. If the decision is not to perform PM, i.e. if $x_{PM} = 0$, machine will start working, with its condition being the observed initial degradation level S_0. In that case, if the machine breaks down and if MRM option is chosen (i.e. $x_{RMO} = 0$), the machine will be restored to S_0, while if PRM is chosen, machine will be restored to the perfect working condition. Thus, the Recovery Transition Submatrix $\mathbf{P_R}(x_{PM}, x_{RMO})$ can be described as

$$\left[\mathbf{P_R}(x_{PM}, x_{RMO})\right]_{i,j} = \Pr(S_t = W_j \mid S_{t-1} = M_i, x_{PM}, x_{RMO}) = \begin{cases} 1, \ i = 1, j = \begin{cases} 1 \ , \text{if } x_{PM} = 1 \text{ or } x_{RMO} = 1 \\ S_0, \text{ if } x_{PM} = x_{RMO} = 0 \end{cases} \\ 0, \text{otherwise} \end{cases} \quad , i \in [N_m], j = [N_w]$$

In summary, given the first-stage decisions $\mathbf{x}, x_{PM}, x_{RMO}$ and the initial observed degradation level of the machine, the corresponding Markovian transition probability matrix $\mathbf{P_{trans}}(\mathbf{x}, x_{PM}, x_{RMO})$ is completely described.

---

[4] I.e., in that case, no matter what the decision x_{RMO} is, machine will be restored to the perfect working condition after RM.

Subsequently, the history of states of the machine over time $[T]$ can be represented by a binary matrix $\mathbf{s} \in \mathbb{Z}_2^{(N_w+N_m)\times T}$, with each column $\mathbf{s}^t \in \mathbb{Z}_2^{N_w+N_m}$ indicating the machine state at time t in which only one element is 1, and others are zero.

Based on the above described DTMC model, one can use simulations to obtain realizations of machine states $\mathbf{s} \in \mathbb{Z}_2^{(N_w+N_m)\times T}$ and for each realization, one can also obtain the auxiliary variable $z_{RM} \in \mathbb{Z}$ denoting the total number of RMs conducted during this decision epoch. Output of each simulation will be denoted by the function

$$\left(\mathbf{s}, z_{RM}\right) = F_{trans}\left(\mathbf{P}_{\mathbf{trans}}(\mathbf{x}, x_{PM}, x_{RMO}), \mathbf{s^0}, \xi_{PM}, \xi_{MRM}, \xi_{PRM}\right)$$

where $\mathbf{s^0}$ is a binary vector denoting the initial observed degradation level of the machine condition, while ξ_{PM}, ξ_{MRM} and ξ_{PRM} are respectively the uncertain times for PM, minimum maintenance and perfect RM. This DTMC model also enables the simulation process to be formulated by a block of constraints in a mathematical programming problem, which will be shown in Section 3.2.3.

3.2.3. **The Two-stage Stochastic Programming Model**

Following [5], in order to achieve the maximum total expected profit for each decision epoch, we formulate the integrated production and maintenance planning problem as the following two-stage stochastic programming problem:

$$V^* = \max_{\mathbf{x}, x_{PM}, x_{RMO}} -f_{PM} + \mathrm{E}\left[Q^*(\mathbf{x}, x_{PM}, x_{RMO}, \xi_{PM}, \xi_{MRM}, \xi_{PRM}, \xi_{demand})\right] \quad (3.1)$$

$$\textit{subject to} \quad \sum_{q=1}^{N_p} x_q = 1, x_q \in \mathbb{Z}_2, \forall q \in [N_p], x_{PM} \in \mathbb{Z}_2, x_{RMO} \in \mathbb{Z}_2 \quad (3.2)$$

$$\textit{where } Q^*(\mathbf{x}, x_{PM}, x_{RMO}, \xi_{PM}, \xi_{MRM}, \xi_{PRM}, \xi_{demand})$$
$$= \max_{\substack{y_{out}, y_{over} \\ y_{backorder}, y_{wasted}}} f_{sell} + f_{scrap} + f_{over} + f_{wasted} - f_{out} - f_{backorder} - f_{pro} - f_{RM} - f_{extraMT}$$

(3.3)

$$\textit{subject to} \quad (\mathbf{s}, z_{RM}) = F_{trans}\left(\mathbf{P}_{\mathbf{trans}}(\mathbf{x}, x_{PM}, x_{RMO}), \mathbf{s^0}, \xi_{PM}, \xi_{MRM}, \xi_{PRM}\right) \quad (3.4)$$

$$\xi_{demand} - 1 \le \sum_{t=1}^{T} \sum_{i=1}^{N_w} \left(s_i^t \sum_{q=1}^{N_p} q\theta_{i,q} x_q\right) + y_{out} + y_{back} - y_{over} - y_{wasted} \le \xi_{demand} \quad (3.5)$$

$$y_{out} \le M_{out}, y_{over} \le M_{over} \quad (3.6)$$

$$y_{out}, y_{over}, y_{back}, y_{wasted} \in \mathbb{Z}_{\ge 0} \quad (3.7)$$

Eq. (3.1) and (3.2) are respectively the objective function and constraints of the first-stage problem, where the preventive maintenance cost is $f_{PM} = c_{PM} x_{PM}$, while $\mathrm{E}\left[Q^*(\mathbf{x}, x_{PM}, x_{RMO}, \xi_{PM}, \xi_{MRM}, \xi_{PRM}, \xi_{demand})\right]$ is the expected profit of the optimal recourse decisions $y_{out}, y_{backorder}, y_{over}, y_{wasted}$ obtained by solving the second-stage optimization problem (3.3)-(3.7), for any given first-stage decisions $\mathbf{x}, x_{PM}, x_{RMO}$ and the corresponding simulated realizations of uncertainties. Components in the second-stage objective function (3.3) are:

- *Cost of production*: expressed by the term

$$f_{pro} = c_{pro}\left(\sum_{q=1}^{N_p} qx_q\right)\left(\sum_{t=1}^{T}\sum_{i=1}^{N_w} s_i^t\right),$$

where c_{pro} is the unit cost per product.

- *Revenue from meeting the demand*: expressed by the term

$$f_{sell} = p_{sell}\xi_{demand},$$

where p_{sell} is the price of a good product and ξ_{demand} is the realized demand during the decision epoch[5].

- *Cost of outsourcing*: expressed by the term

$$f_{out} = c_{out}y_{out},$$

where c_{out} is the unit outsourcing cost per product.

- *Cost of backorder*: given by the term

$$f_{backorder} = c_{backorder}y_{backorder},$$

where $c_{backorder}$ is the unit backorder cost per product. Note that this term appears if the number of good and outsourcing products is still insufficient after outsourcing, and backorder takes place (i.e., if $y_{backorder} > 0$).

- *Revenue from over-production*: given by the term

$$f_{over} = p_{over}y_{over},$$

[5] Notice that demand is modeled as a random variable with a known discrete distribution.

where if the number of good products is larger than the demand, certain amount of good products can be sold at a lower price[6] p_{over}.

- *Revenue of wasted products*: described by the term

$$f_{wasted} = p_{wasted} y_{wasted}$$

which exists if after selling the excess good products at a lower unit price p_{over}, there are still y_{wasted} remaining good products that can be recycled or disposed of at an even lower price p_{wasted}.

- *Revenue of imperfect Products*: expressed by the term

$$f_{scrap} = p_{scrap} \sum_{t=1}^{T} \sum_{i=1}^{N_w} \left(s_i^t \sum_{q=1}^{N_p} q(1-\theta_{i,q}) x_q \right),$$

where p_{scrap} is the unit cost associated with a faulty product that needs to be scraped or reproduced, and $\theta_{i,q}$ is the yield rate[7] that depends on the machine condition $i \in [N_w]$ and production rate $q \in [N_p]$. The unit cost p_{scrap} is chosen in a way that

$$c_{backorder} \geq c_{out} \geq p_{sell} \geq c_{pro} \geq p_{over} \geq p_{wasted} \geq p_{scrap}.$$

- *Cost of maintenance*: expressed by the term

$$f_{RM} = c_{PM} x_{PM} + \left[c_{MRM} (1 - x_{RMO}) + c_{PRM} x_{RMO} \right] z_{RM},$$

---

[6] Excess products can also be put into inventory, but because of the holding cost this would incur, this can also be seen as equivalent to being sold at a lower price.

[7] The model of yield rates and its learning process are explained in Section 3.2.5.

where c_{PM}, c_{PRM} and c_{MRM}, respectively, are the unit cost of each PM, PRM and MRM operation.

- *Cost of extra-time*: given by the term

$$f_{extraMT} = c_{extraMT} \sum_{i=1}^{N_m} s_{N_w+i}^T ,$$

where $c_{extraMT}$ is the cost assigned to the situation in which the machine fails near the end of the decision epoch and extra time may be needed to finish the corresponding RM[8].

In constraint (3.4), the function $F_{trans}(\bullet)$ represents a block of constraints which are formulated based on the DTMC degradation model in Section 3.2.2. Let us now show and explain those constraints. We use binary vectors $\boldsymbol{\xi}_{\mathbf{PRM}}^{\mathbf{t}}, \boldsymbol{\xi}_{\mathbf{MRM}}^{\mathbf{t}}, \boldsymbol{\xi}_{\mathbf{PM}}^{\mathbf{t}} \in \mathbb{Z}_2^{N_w+N_m}$ to respectively denote the uncertain perfect reactive maintenance, minimum reactive maintenance, and preventive maintenance time with elements satisfying $\sum_{j=1}^{N_w} \xi_{*M}^{t,j} = 0$ and $\sum_{j=N_w+1}^{N_w+N_m} \xi_{*M}^{t,j} = 1$. Element $\xi_{*M}^{t,j} = 1$ indicates that the realization of maintenance time is $(N_m + 1 - j)$. For example, in the following equation, vector $\boldsymbol{\xi}_{\mathbf{PRM}}^{\mathbf{t}}$ indicates that if perfect reactive maintenance starts to be conducted at time t, the realization of maintenance time is 2.

[8] This cost parameter is taken to be big in order to penalize such events and prevent them from happening.

$$\begin{array}{l} States: W_1, W_2, ..., W_{N_w} \;\vdots\; M_{N_m} ... M_2, M_1 \\ \boldsymbol{\xi}^{\mathbf{t}}_{\mathbf{PRM}} = \begin{bmatrix} 0 & 0 & \cdots & 0 \;\vdots\; 0 & \cdots & 1 & 0 \end{bmatrix}^T \end{array} \tag{3.8}$$

We use $\Pr^t_{i,j}$ to denote the element of the realization of the Markovian transition matrix $\mathbf{P}_{\mathbf{trans}}(\mathbf{x}, x_{PM}, x_{RMO})$ at time t. Following are constraints that define submatrices of the transition matrix.

$$\Pr^t_{i,N_w+j} = \left(\sum_{q=1}^{N_P} x_q F_{qi} \right) \left[\xi^{t,N_w+j}_{PRM} x_{RMO} + \xi^{t,N_w+j}_{MRM} (1 - x_{RMO}) \right], \forall i \in [N_w], \forall j \in [N_m], \forall t \in [T] \tag{3.9}$$

$$\Pr^t_{N_w+N_m,1} = x_{RMO} + (1 - x_{RMO}) \left[\Pr^{t-1}_{N_w+N_m,1} \left(1 - \sum_{i'=1}^{N_w} s^t_{i'} \right) + s^t_1 \sum_{i'=1}^{N_w} s^t_{i'} \right], \forall t \in [T] / \{1\} \tag{3.10}$$

$$\Pr^t_{N_w+N_m,j} = \Pr^{t-1}_{N_w+N_m,1} \left(1 - \sum_{i'=1}^{N_w} s^t_{i'} \right) + s^t_1 \sum_{i'=1}^{N_w} s^t_{i'}, \forall j \in [N_w] / \{1\}, \forall t \in [T] / \{1\} \tag{3.11}$$

$$\Pr^1_{N_w+N_m,1} = s^0_1 (1 - x_{PM}) + x_{PM} \tag{3.12}$$

$$\Pr^1_{N_w+N_m,j} = s^0_j, \; j \in [N_w] / \{1\} \tag{3.13}$$

Constraint (3.9) defines the *Failure Transition Submatrix* $\mathbf{P}_{\mathbf{F}}(\mathbf{x}, x_{RMO})$ in the Markovian transition matrix, where F_{qi} is the failure probability of the machine when it is working under production rate q and degradation level i. This constraint aligns the vector of failure probability with the column determined by the realization of maintenance time. Other columns in the submatrix are zeros. The block of constraints (3.10)-(3.13) defines the *Recovery Transition Submatrix* $\mathbf{P}_{\mathbf{R}}(x_{PM}, x_{RMO})$. Constraint (3.10) defines the

probability of restoring to the perfect working condition at time t. If we decide to perform PRM, the probability is one. In the case of MRM, if the machine is working, the probability is $\Pr^{t-1}_{N_w+N_m,1}$, if not, it equals to s^t_1. Constraint (3.11) defines the probability of restoring to other working conditions at time t. Constraints (3.12) and (3.13) define the initial probability of restoring to perfect and other working conditions (i.e. $t=1$).

The second block of constraints performs the transition of states to obtain the history of states of the machine over time $[T]$ represented by a binary matrix $\mathbf{s} \in \mathbb{Z}_2^{(N_w+N_m)\times T}$.

$$\mathbf{s^1} = \boldsymbol{\xi}^{\mathbf{t}}_{\mathbf{PM}}\, x_{PM} + \mathbf{s^0}(1-x_{PM}) \tag{3.14}$$

$$s^t_i\left(\xi^t - \sum_{j'=1}^{j-1} \Pr^t_{i,j'}\right) \le a^t_{i,j} M_{large}, \forall i \in [N_w+N_m], \forall j \in [N_w+N_m]/\{1\}, \forall t \in [T] \tag{3.15}$$

$$(a^t_{i,j}-1)M_{large} \le s^t_i\left(\xi^t - \sum_{j'=1}^{j-1} \Pr^t_{i,j'}\right), \forall i \in [N_w+N_m], \forall j \in [N_w+N_m]/\{1\}, \forall t \in [T] \tag{3.16}$$

$$a^t_{i,1} = s^t_i, \forall i \in [N_w+N_m], \forall t \in [T] \tag{3.17}$$

$$s^t_i\left(\sum_{j'=1}^{j} \Pr^t_{i,j'} - \xi^t\right) \le b^t_j M_{large}, \forall i, j \in [N_w+N_m],\ \forall t \in [T] \tag{3.18}$$

$$(b^t_j-1)M_{large} \le s^t_i\left(\sum_{j'=1}^{j} \Pr^t_{i,j'} - \xi^t\right), \forall i, j \in [N_w+N_m], \forall t \in [T] \tag{3.19}$$

$$s^t_j = a^{t-1}_{i,j}\, b^{t-1}_{i,j}, \forall i, j \in [N_w+N_m], \forall t \in [T]/\{1\} \tag{3.20}$$

Constraint (3.14) defines the initial vector of state according to the decision for PM. Constraint (3.15) - (3.20) formulate the state transition over time $[T]$ using random variables ξ^t as a trigger of transition at each time t. $\mathbf{a^t} \in \mathbb{Z}_2^{(N_w+N_m)\times(N_w+N_m)}$ is a binary vector where only one element has value 1. The element $a_{i,j}^t = 1$ indicates that at time t, the machine is in state i and transition trigger is larger than $\sum_{j'=1}^{j-1} \Pr_{i,j'}^t, \forall j \in [N_w + N_m] / \{1\}$ or larger than 0 for $j = 1$. $\mathbf{b^t} \in \mathbb{Z}_2^{(N_w+N_m)\times(N_w+N_m)}$ is a binary vector where only one element has value 1. The element $b_{i,j}^t = 1$ indicates that at time t, the machine is in state i and transition trigger is smaller than or equal to $\sum_{j'=1}^{j} p_{i,j'}^t, \forall j \in [N_w + N_m]$. M_{large} is a large positive number.

Constraint (3.5) pertains to recourse decisions and ensures that the demand is fulfilled, while constraint (3.6) limits quantities of outsourced products and overproduction to upper-bounds M_{out} and M_{over}, respectively.

A closed form solution for the proposed model cannot be obtained [107] and hence, we approach this problem by pursuing a simulation-based optimization paradigm. A variety of simulation algorithms are proposed to solve two-stage stochastic programming problems with continuous distributions [108]. In order to overcome the challenge of exponential growth of the search space with the number of decision variables [108], we propose to use the Augmented Probability Simulation (APS) approach introduced in [109].

It overcomes the main drawbacks of the widely used Sample Average Approximation (SAA) method [110], by proposing an augmented probability model, where both decision variables and uncertainties are treated as being stochastic. To properly utilize the APS approach, a large enough constant value needs to be added to the objective function to shift it to the positive region, such that proper probability densities are ensured without changing the structure of the probability density function. Because this number should not be too large, the minimum possible objective function value can help determine its value. Applications and discussions of the APS approach in similar problems can be found in recent research [5][111][112].

3.2.4. **Benchmark Model: Expected Value Problem**

To demonstrate the value of the newly proposed method for optimization of manufacturing operations that considers uncertainties in the demand, machine degradation and maintenance times, we introduce here a benchmark model where all uncertainties are modeled as their respective expected values. The corresponding optimization problem is formulated as follows:

$$\max_{\substack{\mathbf{x},x_{PM},x_{PRM} \\ y_{out},y_{over} \\ y_{backorder},y_{wasted}}} \; -f_{PM}+f_{sell}+f_{scrap}+f_{over}+f_{wasted}-f_{out}-f_{backorder}-f_{pro}-f_{RM}-f_{extraMT} \tag{3.21}$$

$$\textit{subject to} \;\; \sum_{q=1}^{N_p} x_q = 1, x_q \in \mathbb{Z}_2, \forall q \in [N_p], x_{PM} \in \mathbb{Z}_2, x_{RMO} \in \mathbb{Z}_2 \tag{3.22}$$

$$(\mathbf{s}, z_{RM}) = F'_{trans}\left(\mathbf{P}_{\mathbf{trans}}(\mathbf{x}, x_{PM}, x_{RMO}), \mathbf{s^0}, \mathbf{E}[\xi_{PM}], \mathbf{E}[\xi_{MRM}], \mathbf{E}[\xi_{PRM}], \mathbf{E}[\xi_{demand}]\right)$$

(3.23)

$$\mathbf{E}[\xi_{demand} - 1] \le \sum_{t=1}^{T}\sum_{i=1}^{N_w}\left(s_i^t \sum_{q=1}^{N_p} q\theta_{i,q}x_q\right) + y_{out} + y_{back} - y_{over} - y_{wasted} \le \mathbf{E}[\xi_{demand}]$$

(3.24)

$$y_{out} \le M_{out}, y_{over} \le M_{over} \tag{3.25}$$

$$y_{out}, y_{over}, y_{back}, y_{wasted} \in \mathbb{Z}_{\ge 0} \tag{3.26}$$

with the function $F'_{trans}(\cdot)$ in Eq.(3.23) being formulated based on the deterministic machine degradation model generated from the DTMC model proposed in Section 3.2.2.

3.2.5. **Learning of Yield Rate Dependences on Production Rates and Machine Conditions**

Based on the goal of maximizing the total profit within the planning horizon, the planning horizon is divided into multiple decisions epochs so that machine monitoring, and subsequent decision-making can be implemented periodically.

One can then see each decision epoch as an opportunity to learn yield rates and how they depend on the production rates and equipment conditions. To that end, we propose to use the Bayesian framework for modeling and updating the belief model of yield rates using the novel two-stage optimization process described in Section 3.2.3.

Based on the definition of a yield rate, it is widely accepted to assume that conformance of quality for a sequence of product units follows a Bernoulli process [113],where a good (quality conforming) product would constitute a success. Given such Bernoulli likelihood, we choose the Beta distribution as the prior belief model of yield rates, since it is a conjugate prior for a Bernoulli distribution [114].

The above-described model enables the issue of learning of yield rate from the data collected in the decision-making process to be solved as a multi-armed bandit problem [115], where a player is faced with a finite number of independent slot machines or bandits, each with a different stationary reward distribution, and the player's objective is to maximize the expected cumulative reward over some given number of trails. Bandit problems are classic and simplified reinforcement learning problems and avoid much of the complexity of the full reinforcement learning problem, because in this class of problems, decisions do not alter the environment and making a decision does not restrict future decisions [116]. Bandit problems were introduced by William R. Thompson in an article published in 1933 [18], where Thompon Sampling - also known as posterior sampling and probability matching - was first proposed. It is an algorithm where actions are taken sequentially in a manner that must balance between exploiting what is known to maximize immediate performance and investing to accumulate new information that may improve future performance [19].

Following [20], the General Thompson sampling (GTS) method is proposed to facilitate learning of yield rate dependencies, as summarized in Figure 1.

Algorithm 1: Beta-Bernoulli TS yield rate learning algorithm with the two-stage decision-making model embedded in it

Index notations: l index of production rates
n index of machine working conditions
t index of decision epochs
Input: number of production rates L
number of machine conditions N_w
number of decision epochs T
initial prior distribution of yield rate $Beta(S_{l,n}(0), F_{l,n}(0))$ for $\forall l$ in $[1, L]$ and $\forall n$ in $[1, N_w]$
observe index of the machine condition n'
Output: V_t: revenue earned during decision epoch t, for $\forall t$ in $[1, T]$
posterior distribution of yield rate $Beta(S_{l,n}(T), F_{l,n}(T))$ for $\forall l$ in $[1, L]$ and $\forall n$ in $[1, N_w]$
for each decision epoch t **in** $[1, T]$ **do**
  for each production rate l **in** $[1, L]$ **do**
    for each machine working condition n **in** $[1, N_w]$ **do**
      sample yield rate: $\theta_{l,n}(t) \sim Beta(S_{l,n}(t), F_{l,n}(t))$
    end
  end
  make decisions for production rate, preventive maintenance or not, perfect reactive maintenance or minimum reactive maintenance using the two-stage stochastic programming model

$$x_1(t), x_{PM}(t), x_{PRM}(t) = \underset{x_1, x_{PM}, x_{PRM}}{\operatorname{argmax}} \; V(\theta_{l,n}(t), n')$$

  observe the machine condition at the end of this decision epoch: n'
  observe revenue earned during this epoch: V_t
  estimate number of perfect products s and imperfect products f produced under respective production rate: l and machine condition: n based on the proposed DTMC model
  update: $S_{lm}(t) := S_{lm}(t) + s$ and $F_{lm}(t) := F_{lm}(t) + f$, for $\forall l$ in $[1, L]$ and $\forall n$ in $[1, N_w]$
end

Figure 1. Proposed Beta-Bernoulli TS yield rate learning algorithm with the two-stage decision-making model embedded in it

3.3. **Numerical Experiments and Results**

In this section, we present results of a set of simulations illustrating the newly introduced decision-making and yield rate learning process. We simulate operations of a single machine that allows ten options of production rates and could undergo three degradation states, with simulation parameters summarized in Table 1. The initial assumption about the yield rate was that it is one for all the machine conditions and production rates, which is a common assumption when no prior knowledge about yield behavior is available. The performance is evaluated through the cumulative sum of the first-stage objective function for each decision-making epoch, as defined by Eq.(3.1), with the term $\mathrm{E}\left[Q^*(x_1, x_{PM}, x_{RMO}, \xi_{PM}, \xi_{MRM}, \xi_{PRM}, \xi_{demand})\right]$ obtained through sample average approximation.

Table 1. Summary of simulation parameters

Basic	Production time epochs in each decision epoch (time unit)	$T = 100$
	Machine working conditions	$\{1, 2, 3\}$
	Initial machine working condition	1 (perfect condition)
Constraint	Production rate	$x_1 \in \{1, 2, ..., 10\}$
	Maximum allowed over production	$M_{over} = 100$
	Maximum allowed outsourced products	$M_{out} = 100$
	Upper bound of maintenance time (time unit)	$M_{MT} = 12$

Cost	Unit production cost		$c_{pro} = 22$
(monetary unit)	Unit PM cost		$c_{PM} = 200$
	Unit PRM cost		$c_{PRM} = 240$
	Unit MRM cost		$c_{MRM} = 220$
	Cost for extra time of maintenance		$c_{extra} = 300$
	Unit outsourcing cost		$c_{out} = 70$
	Unit backorder cost		$c_{backorder} = 80$
Revenue	Unit selling price		$p_{sell} = 60$
(monetary unit)	Unit selling price for overproduction		$p_{over} = 20$
	Unit scrapping price		$p_{scrap} = 10$
	Unit price of wasted products		$p_{wasted} = 15$
Uncertainty	Probability matrix of transition between machine working conditions	$\mathbf{x} = [1,0,0]$	$\mathbf{P_D(x)} = \begin{bmatrix} 0.9 & 0.06 & 0.03 \\ 0 & 0.85 & 0.1 \\ 0 & 0 & 0.7 \end{bmatrix}$
		$\mathbf{x} = [0,1,0]$	$\mathbf{P_D(x)} = \begin{bmatrix} 0.85 & 0.12 & 0.08 \\ 0 & 0.65 & 0.2 \\ 0 & 0 & 0.6 \end{bmatrix}$
		$\mathbf{x} = [0,0,1]$	$\mathbf{P_D(x)} = \begin{bmatrix} 0.55 & 0.2 & 0.15 \\ 0 & 0.45 & 0.3 \\ 0 & 0 & 0.4 \end{bmatrix}$

Probability that machine fails under each machine working condition	$\mathbf{x} = [1,0,0], x_{RMO} = 0$	$\mathbf{P_F}(\mathbf{x}, x_{RMO}) = [0.01 \quad 0.05 \quad 0.3]$
	$\mathbf{x} = [0,1,0], x_{RMO} = 0$	$\mathbf{P_F}(\mathbf{x}, x_{RMO}) = [0.05 \quad 0.15 \quad 0.4]$
	$\mathbf{x} = [0,0,1], x_{RMO} = 0$	$\mathbf{P_F}(\mathbf{x}, x_{RMO}) = [0.1 \quad 0.25 \quad 0.6]$
PM time (time unit)		Discrete Truncated Normal ($\mu = 2, \sigma = 1, a = 1, b = 3$)
PRM time (time unit)		Discrete Truncated Normal ($\mu = 6, \sigma = 2, a = 1, b = 11$)
MRM time (time unit)		Discrete Truncated Normal ($\mu = 3, \sigma = 1, a = 1, b = 5$)
Demand in each decision epoch		Truncated Normal ($\mu = 4, \sigma = 2, a = 0, b = 8$)

3.3.1. **Impact of the Learning of Yield Rate in the Decision-Making Process**

In order to illustrate the importance of online learning of yield rates, we will compare the newly proposed methodology to two benchmark policies. In the first policy, we will assume the ideal situation in which decisions are made using the perfect information of the

yield rates[9], i.e. situation when the actual (true) yield rates are known[10]. In the second policy, which we will refer to as the "policy without learning", we will assume that the initial knowledge about yield rates is true and no learning (updating) takes place. The cumulative value of the first-stage objective function (3.1) obtained by the two policies, and the newly proposed policy with Bayesian learning of yields are shown in the box plots in Figure 2. As can be observed, the newly proposed policy consistently results in a higher cumulative value of the first-stage objective and a narrower likely range of variation, as compared to the "policy without learning".

[9] As will be seen in the simulation results, highest values of the first-stage objective function in Eq.(3.1) can be achieved under that assumption.

[10] This, of course, is ideal and unrealistic.

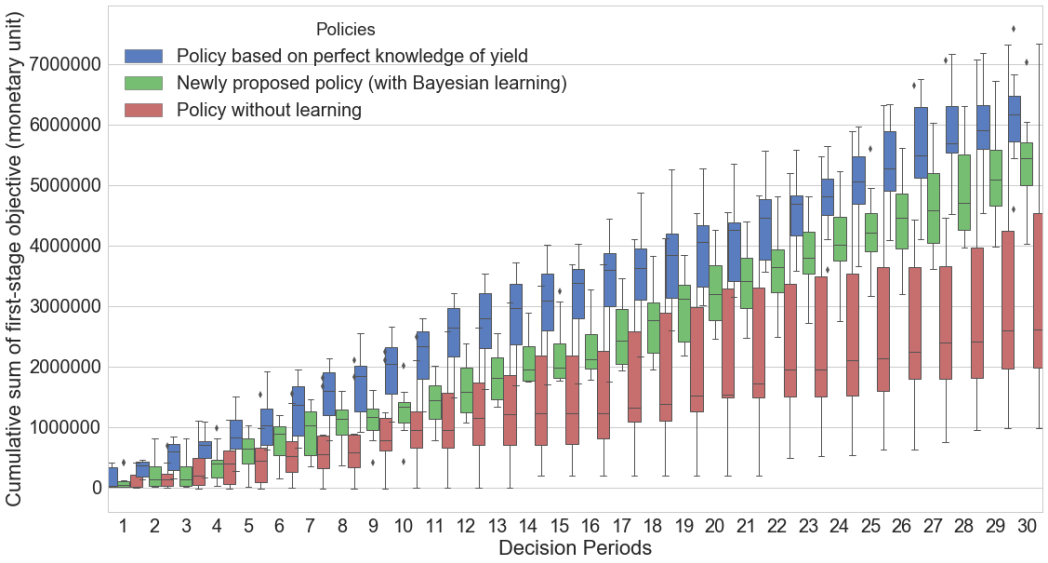

Figure 2. Comparison of cumulative sums of first-stage objective functions obtained using different decision-making policies.

3.3.2. **Effects of Jointly Modeling the Dependence of Yield Rates on the Machine Condition & Production Rate**

In the following study, we aim to illustrate the benefits of considering the concurrent dependence of yield rates on the underlying machine conditions and production rates. We compare results obtained when this joint dependency is taken into account, as proposed in this chapter, to those obtained when yield is assumed to only depend on the machine

condition, as well as to those obtained when yield was assumed to be constant[11]. One can observe from Figure 3 that incorporating more comprehensive considerations of the yield dependencies into the decision-making consistently clearly brings benefits over the benchmark policies.

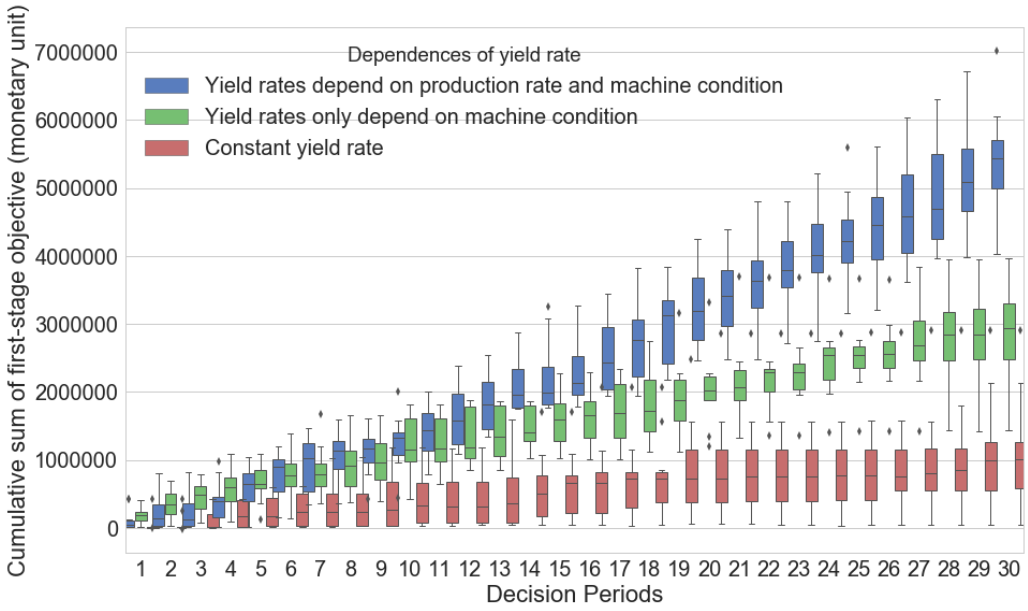

Figure 3. Comparison of cumulative sums of first-stage objective functions for different models of dependences of yield rates.

[11] I.e. when it does not change with the underlying machine condition, nor with the production rates.

3.3.3. **Benefits of Considering Demand and Production Process Stochastically**

In this section, we aim to show the benefits of considering the stochasticity of the demand, degradation processes and maintenance operations within the newly proposed decision-making process. We do so by comparing the novel method with the benchmark model described in Section 3.2.4, where all random quantities are deterministically modeled as their respective expected values. The two approaches can be compared using cumulative values of their respective objective functions (3.1) and (3.21), which are concurrently shown in Figure 4. One can observe from Figure 4 that considerations of stochasticities within the newly proposed method yield lower cumulative values of the objective function at the beginning of the learning process, but over time, the newly proposed method gradually outperforms the purely deterministic benchmark approach. Furthermore, from the interquartile ranges of the box plots, one can observe that the likely ranges of variations yielded by the stochastic optimization proposed in this chapter are narrower than the ones obtained using the expected-value-based benchmark model, especially in later stages of the learning process (later decision-making epochs). Therefore, the value of considering uncertainties associated with the demand, degradation and maintenance process is evident.

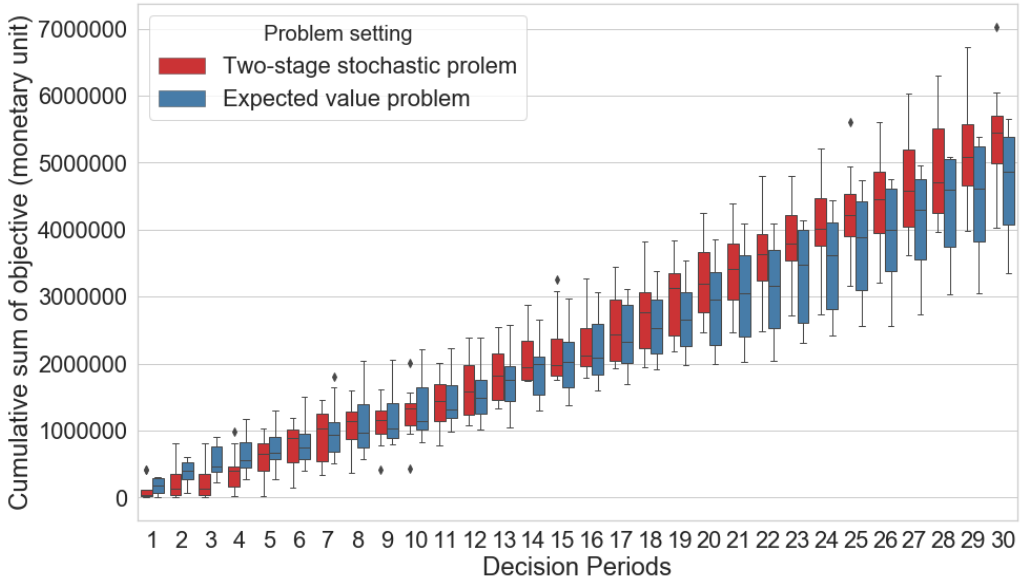

Figure 4. Comparison of cumulative sum of objective for stochastic and deterministic problem settings

3.4. **Conclusions**

This chapter presents an integrated decision-making policy in terms of production and maintenance operations with concurrent considerations and learning of dependencies of the yield rates on the underlying equipment conditions and production rates. Furthermore, the novel decision-making policy addresses demand uncertainties by optimizing decisions regarding excess production above the observed demand, or meeting demand via outsourcing or obtaining additional products from the inventory. The large-scale optimization problem is modeled as a two-stage stochastic programming problem solved using MCMC simulations, while closed form dependencies of the yield rates were

learned and updated within the simulations using Beta-Bernoulli Thompson Sampling method.

The newly proposed methodology was evaluated and benchmarked against the traditional approaches using an example of a single manufacturing machine that allows ten options of production rates and could undergo three degradation states. Numerical experiments show clear and consistent benefits of the newly proposed online learning of integrated yield rate dependencies, compared to what is obtained without learning of yield rates, or assuming that yield is constant or depends on the equipment condition only. Furthermore, the benefit of considering stochasticities of the demand, degradation processes and maintenance operations within the newly proposed method were demonstrated via comparisons with the decisions made using a deterministic model solely based on the expected values of the demand, machine uptimes and maintenance times.

There are various possibilities for extending the work presented in this chapter. Methodologically, robustness to uncertainties in model parameters and structure can be considered within the decision-making process. Namely, parameters of degradation dynamics are never perfectly known and degradation states are rarely directly observable, which can be addressed e.g. by modeling the degradation process using Hidden Markov Model (HMM) with unknown parameters, as was done in [117]. From the practical point of view, the new approach should be extended to more realistic manufacturing systems, involving multiple machines and multiple products, as well as more realistic logistic and organizational constraints.

Chapter 4. Dynamic Down-Selection of Measurement Markers for Optimized Robust Control of Overlay Errors in Photolithography[12]

4.1. Introduction

In semiconductor manufacturing, photolithography steps are used repeatedly to successively transfer patterns from a series of masks onto the wafer surface in order to facilitate building of interconnections between complex structures that make up functioning memory and logic circuits [118]. In order to ensure functionality and performance characteristics of the final product, patterns printed in any given layer need to be precisely aligned to the patterns formed in its adjoining layers [119], with overlay errors describing the misalignment between neighboring pattern layers [118].

Overlay errors can be seen as one of the key factors that limit the semiconductor device's performance, reliability and yield [120][9], which led to significant research efforts to enable controlling of those errors within increasingly narrow margins. Crucial to those efforts were explorations of models that relate lithography overlay errors to various

[12] This chapter is based on a submitted journal paper: Zhang, H., Feng, T. and Djurdjanovic, D., "Dynamic down-selection of measurement markers for optimized robust control of overlay errors in photolithography," under review, *IEEE Transactions on Semiconductor Manufacturing*, Manuscript ID TSM-21-0179. Huidong Zhang wrote this publication under the supervision of Dragan Djurdjanovic.

parameters of the tool on which the process is performed, with a number of those tool parameters being controllable and thus enabling reduction of the resulting overlay errors [121][122]. Traditional industry practice for control of photolithography overlay errors relies on coupling of those overlay error models with various prediction strategies, resulting in the so-called run-to-run (R2R) paradigm of overlay control [123][124]. Specifically, overlay models are used to estimate bias terms relevant to a given pattern layer across the history of previously produced wafers, based on which predictions of bias terms for that pattern in the next wafer, or lot of wafers can be made and compensated for using control commands on the relevant tool. A thorough review of various R2R strategies for control of semiconductor manufacturing processes can be found in [124] and references therein.

The above-described R2R control strategies effectively control overlay errors between neighboring patterns. However, device performance is also influenced by the so-called stack-up overlay errors, which describe misalignments across non-neighboring layers [125]. Therefore, both the layer-to-layer, as well as stack-up overlay errors need to be concurrently controlled, as acknowledged in recent studies [126][127]. In [126], the authors proposed a stochastic optimal control strategy that integrates the R2R controller into a multilayer framework, simultaneously considering both the layer-to-layer and stack-up overlay errors. Though this method outperformed the use of purely R2R control approaches, it relied upon Bayesian assumptions of perfectly known overlay model parameters, along with Gaussianity and independence of all noise terms in the overlay model. These are very strong assumptions, especially when it comes to models of

photolithography overlay errors in semiconductor manufacturing, where model noise terms are known to often be non-Gaussian and structured, both temporally and spatially [128]. This was recently addressed in [127], where a methodology for robust control of lithography errors is proposed, bypassing the aforementioned Bayesian assumptions and pursuing control that optimizes a measure of overlay and stack-up overlay errors for the worst-case scenario regarding the modeling noise and model-parametric uncertainties. Those uncertain terms are assumed to lie within some known bounds, without any assumptions regarding their distributions, or temporal and spatial structure. Simulations based on the model and data from an actual industrial facility demonstrated clear advantages of the robust control method over R2R approaches, as well as the purely stochastic multi-layer overlay control method from [126], especially with larger levels of model parametric uncertainties, or in the presence of non-Gaussian or structured noise terms.

The robust overlay control method introduced in [127] opens the door for quantitative understanding and manipulation of the interplay between one's ability to control overlay errors and one's need to measure those errors in order to facilitate that control. Namely, intuitively speaking, increasing the number of markers in which overlay measurements are taken should lead to better estimation and prediction of relevant parameters, thus improving one's ability to control the overlay errors. However, those gains come with simultaneous increases in metrology times, which will negatively impact the overall throughput. Conversely, reducing the number of markers for overlay measurements

should lead to improvements in the cycle-time, but at the expense of increased uncertainties in the estimation and prediction of bias terms, thus deteriorating the control process. The work reported in [127] facilitates characterization of the aforementioned interplay between the overlay control process and changes in overlay measurement schemes. That opportunity will be exploited in this chapter to facilitate strategic selection of overlay markers on the wafer in a way that maximal reduction of overlay errors can be achieved with minimum measurement effort.

Available literature has proposed several strategies for spatial allocation and sampling of overlay measurement markers. Earliest studies pursued layouts of measurement markers based on expert knowledge [93]-[95]. They found that the layout with spatially optimized distribution, which intends to maximize distance between any two selected fields, can offer the best control efficiency. Therefore, rule-based approaches with the aim of distributing measurement markers evenly across the wafer were preferred [96][97]. Recent research reported in [98] can be seen as the first work that employed a metaheuristic-based optimization method to improve the layout of markers for measurement of overlay errors. However, it did not consider the stochastic nature of overlay phenomena, nor were the stack-up overlay errors considered. Furthermore, the authors assumed that all fields on the wafer had the same layout to make sure the measurement markers are evenly distributed across the entire wafer.

To fill this void, this chapter proposes an optimization method which, given a target number of overlay markers that should be used for measurement, dynamically utilizes the

overlay measurements available before processing of a given wafer to strategically down-select overlay measurement markers from the exhaustive set of markers in a way that, under the measurement constraints, facilitates best possible robust control of overlay and stack-up overlay errors on the next wafer. The remainder of this chapter is organized as follows. Section 4.2 presents the formulation and a solution of the optimization problem yielding a selection of overlay measurement markers which, under the constraints of the desired number of overlay markers in which measurements should be taken, optimizes the performance of the resulting robust overlay control algorithm over the entire set of markers, including those that were not included in the measurement scheme. Section 4.3 presents results of application of the newly proposed method to the overlay error model and data from a 4-layer industrial photolithography process used in a major 300 mm fab. Finally, Section 4.4 summarizes the conclusions of this research and provides directions for future work.

4.2. **Methodology: Mathematical Formulation and Solution of the Robust Measurement Marker Selection Problem**

4.2.1. **Robust Control of Overlay Errors**

For layer pattern *k* of wafer *t*, let $\boldsymbol{o}_{t,k}^{x}$ and $\boldsymbol{o}_{t,k}^{y}$ denote vectors formed by overlay errors in all available measurement markers on the wafer, with $\boldsymbol{o}_{t,k}^{x}$ denoting errors in the

x direction and $\boldsymbol{o}_{t,k}^{y}$ denoting overlay errors in the y direction on the wafer[13]. In the foundations of control of photolithography overlay errors are Zernike polynomial [128] based models which relate overlay errors to controllable parameters on the photolithography tool via the form

$$\begin{cases} \boldsymbol{o}_{t,k}^{x} = \boldsymbol{D}^{cx}\boldsymbol{u}_{t,k}^{x} + \boldsymbol{r}_{t,k}^{x} \\ \boldsymbol{o}_{t,k}^{y} = \boldsymbol{D}^{cy}\boldsymbol{u}_{t,k}^{y} + \boldsymbol{r}_{t,k}^{y} \end{cases} \tag{4.1}$$

where vectors $\boldsymbol{u}_{t,k}^{x}$ and $\boldsymbol{u}_{t,k}^{y}$ consist of controllable tool parameters affecting overlay errors in the x and y directions on the wafer, regression matrices $\boldsymbol{D}^{cx}$ and $\boldsymbol{D}^{cy}$ are fully defined by locations of the overlay measurement markers on the wafer, while residual vector terms $\boldsymbol{r}_{t,k}^{x}$ and $\boldsymbol{r}_{t,k}^{y}$ account for unmodeled effects and process noise.

In general, it is recognized that control of photolithography tools is inherently subject to stochastic actuator uncertainties modeled as

$$\begin{cases} \boldsymbol{u}_{t,k}^{x} = \bar{\boldsymbol{u}}_{t,k}^{x} + \boldsymbol{c}_{t,k}^{x} \\ \boldsymbol{u}_{t,k}^{y} = \bar{\boldsymbol{u}}_{t,k}^{y} + \boldsymbol{c}_{t,k}^{y} \end{cases} \tag{4.2}$$

where $\bar{\boldsymbol{u}}_{t,k}^{x}$ and $\bar{\boldsymbol{u}}_{t,k}^{y}$ are vectors of control commands given to the tool, while vectors

[13]Directions of axes on the wafer are determined based on the notch pre-fabricated on the periphery of each wafer.

$\boldsymbol{c}_{t,k}^{x}$ and $\boldsymbol{c}_{t,k}^{y}$ model those stochastic actuator uncertainties and are commonly referred to as vectors of process bias terms. Namely, due to exceptionally small scales in which controllable parameters of a photolithography tool reside[14], unmodeled process dynamics and external noise sources are inevitably significant and cause process bias terms to always be present and continuously change from one wafer to another. A common practice in the industry is to utilize overlay measurements from historical records of previously manufactured wafers to make predictions $\tilde{\hat{\boldsymbol{c}}}_{t,k}^{x}$ and $\tilde{\hat{\boldsymbol{c}}}_{t,k}^{y}$ of bias vector term in layer *k* of wafer *t* prior to the actual lithography exposure, based on which the relevant control commands $\bar{\boldsymbol{u}}_{t,k}^{x}$ and $\bar{\boldsymbol{u}}_{t,k}^{y}$ during the actual exposure can be set to

$$\begin{cases} \bar{\boldsymbol{u}}_{t,k}^{x} = -\tilde{\hat{\boldsymbol{c}}}_{t,k}^{x} \\ \bar{\boldsymbol{u}}_{t,k}^{y} = -\tilde{\hat{\boldsymbol{c}}}_{t,k}^{y} \end{cases} \tag{4.3}$$

in order to counteract those bias terms. Equation (4.3) describes the so-called *run-to-run* (R2R) paradigm for control of photolithography overlay errors and in its foundations are various methods for dynamic modeling and prediction of bias vector terms, such as Kalman filter-based prediction, various forms of the Exponentially Weighted Moving Average (EWMA) based modeling and prediction methods, or more recently dynamic neural networks [129]and Gaussian Process Regression [130]based approaches. As mentioned

[14]Nanometer, or even sub-nanometer scales.

earlier, a thorough survey of R2R approaches for control of semiconductor manufacturing processes can be found in [124]and references therein.

In this chapter, R2R paradigm will be augmented with considerations of the multistage character of errors in alignment across non-neighboring pattern layers, as well as considerations of robustness of control algorithm performance in the presence of uncertain stochastic terms in the overlay models, as suggested in [127]. Specifically, as we are about to produce pattern layer *k* on wafer *t*, let us assume that we have R2R predictions $\tilde{\hat{\boldsymbol{c}}}_{t,k}^{x}$ and $\tilde{\hat{\boldsymbol{c}}}_{t,k}^{y}$ of the corresponding bias vector terms, as well as observations of overlay errors of layers 1 through *k*-1 of wafer *t* measured in a subset of markers defined by a binary vector $\boldsymbol{F}_t$, which is of the same dimensionality *P* as the total number of available markers, with one or zero entries in it respectively denoting presence or absence of the corresponding marker in the measurement scheme $\boldsymbol{F}_t$. Furthermore, following [127], let us assume that stochastic terms $\boldsymbol{r}_{t,k}^{x}, \boldsymbol{r}_{t,k}^{y}$, $\boldsymbol{c}_{t,k}^{x}$ and $\boldsymbol{c}_{t,k}^{y}$ for layer *k* of wafer *t* reside within some upper and lower bounds, which are assumed to be known prior to actual exposure. More precisely, let $\boldsymbol{r}_{t,k}^{ub,x}$ and $\boldsymbol{r}_{t,k}^{ub,y}$ denote upper bounds corresponding on the vector terms $\boldsymbol{r}_{t,k}^{x}$ and $\boldsymbol{r}_{t,k}^{y}$ of residuals in the layer-level models (4.1) and let $\boldsymbol{r}_{t,k}^{lb,x}$ and $\boldsymbol{r}_{t,k}^{lb,y}$ denote the corresponding lower bounds. In Section 4.2.4 of this chapter, we will propose a statistics-inspired approach for determining these boundaries based on historical records of overlay errors measured on previously produced pattern layers and wafers, though authors wish to acknowledge that other, perhaps less formal methods can be used

as well[15]. Furthermore, let $\boldsymbol{c}_{t,k}^{ub,x}$ and $\boldsymbol{c}_{t,k}^{ub,y}$ denote upper bounds corresponding on the bias vector terms $\boldsymbol{c}_{t,k}^{x}$ and $\boldsymbol{c}_{t,k}^{y}$, respectively, and let $\boldsymbol{c}_{t,k}^{lb,x}$ and $\boldsymbol{c}_{t,k}^{lb,y}$ respectively denote the corresponding lower bounds. These bounds can be adjusted based on R2R predictions $\tilde{\tilde{\boldsymbol{c}}}_{t,k}^{x}$ and $\tilde{\tilde{\boldsymbol{c}}}_{t,k}^{y}$ of bias vector terms $\boldsymbol{c}_{t,k}^{x}$ and $\boldsymbol{c}_{t,k}^{y}$ relevant to pattern *k* on wafer *t*, and the corresponding prediction uncertainties, as will be described in more details in Section 4.2.4 of this chapter, though, once again, other approaches can be employed for this purpose as well.

For measurement scheme $\boldsymbol{F}_t$, let $\boldsymbol{o'}_{t,k}^{x}$ and $\boldsymbol{o'}_{t,k}^{y}$ denote vectors of overlay errors measured in markers $\boldsymbol{F}_t$ on pattern layer *k* of wafer *t***.** Similarly, let $\boldsymbol{s'}_{t,k}^{x}$ and $\boldsymbol{s'}_{t,k}^{y}$ denote vectors of stack-up overlay errors measured in markers $\boldsymbol{F}_t$ on pattern layer *k* of wafer *t***.** We will denote this down-selection of overlay and stack-up overlay error measurements as

$$\boldsymbol{o'}_{t,k}^{x} = \boldsymbol{F}_t \circ \boldsymbol{o}_{t,k}^{x};\ \boldsymbol{o'}_{t,k}^{y} = \boldsymbol{F}_t \circ \boldsymbol{o}_{t,k}^{y}$$

and

$$\boldsymbol{s'}_{t,k}^{x} = \boldsymbol{F}_t \circ \boldsymbol{s}_{t,k}^{x};\ \boldsymbol{s'}_{t,k}^{y} = \boldsymbol{F}_t \circ \boldsymbol{s}_{t,k}^{y}$$

[15]E.g., those boundaries could be set using expert knowledge, or recommendations from the tool supplier, which would be informed by the actual tool design and actuator uncertainties.

where $\boldsymbol{o}_{t,k}^{x}$ and $\boldsymbol{o}_{t,k}^{y}$ denote vectors of overlay errors in all markers on pattern layer *k* of wafer *t*, while $\boldsymbol{s}_{t,k}^{x}$ and $\boldsymbol{s}_{t,k}^{y}$ denote vectors of overlay errors in all markers on pattern layer *k* of wafer *t*. Then, following [127], for pattern layer k on wafer *t*, one can use measurements of stack-up overlay errors obtained from markers $\boldsymbol{F}_t$ in the already produced pattern layers 1,2,…,*k*-1 of wafer *t*, as well as lower and upper bounds on the uncertain overlay model terms[16] $\boldsymbol{r}_{t,k}^{lb,x}$, $\boldsymbol{r}_{t,k}^{lb,y}$, $\boldsymbol{c}_{t,k}^{lb,x}$, $\boldsymbol{c}_{t,k}^{lb,y}$, $\boldsymbol{r}_{t,k}^{ub,x}$, $\boldsymbol{r}_{t,k}^{ub,y}$, $\boldsymbol{c}_{t,k}^{ub,x}$, and $\boldsymbol{c}_{t,k}^{ub,y}$ to pursue control commands $\boldsymbol{u}_{t,k}^{x*}$ and $\boldsymbol{u}_{t,k}^{y*}$ which robustly minimize a measure of overlay and stack-up overlay errors measured in markers $\boldsymbol{F}_t$. More precisely, control commands $\boldsymbol{u}_{t,k}^{x*}$ and $\boldsymbol{u}_{t,k}^{y*}$ will be obtained by solving the following optimization problem

$$\left(\boldsymbol{u}_{t,k}^{x*}, \boldsymbol{u}_{t,k}^{y*}\right) = \underset{\substack{\boldsymbol{u}_{t,k}^{x}\in\mathbb{R}^{N_x} \\ \boldsymbol{u}_{t,k}^{y}\in\mathbb{R}^{N_y}}}{\operatorname{argmin}} \max_{\substack{\boldsymbol{r}_{t,k}^{x},\boldsymbol{r}_{t,k}^{y} \\ \boldsymbol{c}_{t,k}^{x},\boldsymbol{c}_{t,k}^{y}}} \begin{matrix} \lambda^x \left\| \boldsymbol{o'}_{t,k}^{x} \right\|^2 + \lambda^y \left\| \boldsymbol{o'}_{t,k}^{y} \right\|^2 \\ +\alpha^x \left\| \boldsymbol{F}_t \circ \boldsymbol{s}_{t,k-1}^{x} + \boldsymbol{o'}_{t,k}^{x} \right\|^2 \\ +\alpha^y \left\| \boldsymbol{F}_t \circ \boldsymbol{s}_{t,k-1}^{y} + \boldsymbol{o'}_{t,k}^{y} \right\|^2 \end{matrix} \tag{4.4}$$

$$subject\ \ to: \quad \begin{cases} \boldsymbol{o'}_{t,k}^{x} = \boldsymbol{F}_t \circ \left[\boldsymbol{D}^{cx}\left(\boldsymbol{u}_{t,k}^{x} + \boldsymbol{c}_{t,k}^{x}\right) + \boldsymbol{r}_{t,k}^{x} \right] \\ \boldsymbol{o'}_{t,k}^{y} = \boldsymbol{F}_t \circ \left[\boldsymbol{D}^{cy}\left(\boldsymbol{u}_{t,k}^{y} + \boldsymbol{c}_{t,k}^{y}\right) + \boldsymbol{r}_{t,k}^{y} \right] \end{cases} \tag{4.5}$$

[16] As mentioned above, these bounds can be obtained from historical records of overlay errors from previous wafers and layer patterns. Section 4.2.4 will offer one statistics-inspired method how such updates can be done.

$$\begin{cases} \boldsymbol{r}_{t,k}^{lb,x} \le \boldsymbol{r}_{t,k}^{x} \le \boldsymbol{r}_{t,k}^{ub,x} \\ \boldsymbol{r}_{t,k}^{lb,y} \le \boldsymbol{r}_{t,k}^{y} \le \boldsymbol{r}_{t,k}^{ub,y} \end{cases} \tag{4.6}$$

$$\begin{cases} \boldsymbol{c}_{t,k}^{lb,x} \le \boldsymbol{c}_{t,k}^{x} \le \boldsymbol{c}_{t,k}^{ub,x} \\ \boldsymbol{c}_{t,k}^{lb,y} \le \boldsymbol{c}_{t,k}^{y} \le \boldsymbol{c}_{t,k}^{ub,y} \end{cases} \tag{4.7}$$

One can note that in the objective function (4.4), terms $\boldsymbol{F}_t \circ \boldsymbol{s}_{t,k-1}^{x} + \boldsymbol{o'}_{t,k}^{x}$ and $\boldsymbol{F}_t \circ \boldsymbol{s}_{t,k-1}^{y} + \boldsymbol{o'}_{t,k}^{y}$ describe stack-up overlay errors obtained from markers $\boldsymbol{F}_t$ on pattern layer *k* of wafer *t*, while constants $\lambda^x, \lambda^y, \alpha^x$ and α^y are weighting factors which can be used to denote the relative importance of each term in (4.4). Constraint (4.5) expresses the Zernike polynomial-based model (4.1)-(4.2) for overlay errors $\boldsymbol{o'}_{t,k}^{x}$ and $\boldsymbol{o'}_{t,k}^{y}$ in measurement markers $\boldsymbol{F}_t$ on pattern layer *k* of wafer *t*, while constraints (4.6) and (4.7) denote that vector terms $\boldsymbol{r}_{t,k}^{x}, \boldsymbol{r}_{t,k}^{y}, \boldsymbol{c}_{t,k}^{x}$ and $\boldsymbol{c}_{t,k}^{y}$ in the model (4.5) are unknown, but reside within some known lower and upper bounds.

Control algorithm defined by optimization (4.4)-(4.7) pursues minimization of the worst-case of objective (4.4) which incorporates a measure of overlay and stack-up overlay errors in markers $\boldsymbol{F}_t$ on layer *k* of wafer *t*. The worst case of objective (4.4) is assessed regarding uncertain model terms $\boldsymbol{r}_{t,k}^{x}, \boldsymbol{r}_{t,k}^{y}, \boldsymbol{c}_{t,k}^{x}$ and $\boldsymbol{c}_{t,k}^{y}$ in the overlay errors model (4.5) and an effective method for solving (4.4)-(4.7) was described in [127]. In the remainder of this section, this ability to solve (4.4)-(4.7) for any given measurement scheme $\boldsymbol{F}_t$ will be used to obtain measurement schemes which enable best possible performance of the control

algorithm (4.4)-(4.7) under different constraints imposed on the number of measurement markers that are allowed to be retained in the measurement scheme $\boldsymbol{F}_t$.

4.2.2. **Problem Formulation**

Given any set of selected overlay measurement markers $\boldsymbol{F}_t$, we will describe the corresponding control performance $f_{t,k}(\boldsymbol{F}_t)$ at layer k of wafer t using the weighted sum of L^2-norms of overlay and stack-up errors *in all the candidate markers*, as shown below.

$$f_{t,k}(\boldsymbol{F}_t) = \lambda^x \left\| \boldsymbol{o}_{t,k}^x \right\|^2 + \lambda^y \left\| \boldsymbol{o}_{t,k}^y \right\|^2 + \alpha^x \left\| \boldsymbol{s}_{t,k}^x \right\|^2 + \alpha^y \left\| \boldsymbol{s}_{t,k}^y \right\|^2 \quad (4.8)$$

Given the robust control commands $\boldsymbol{u}_{t,k}^{x*}$, $\boldsymbol{u}_{t,k}^{y*}$ obtained for the selected markers $\boldsymbol{F}_t$ using procedure (4.4)-(4.7), the worst-case performance of the metric (4.8) can be obtained by solving the following optimization problem:

$$J_{t,k}(\boldsymbol{F}_t) = \max_{\substack{\boldsymbol{c}_{t,k}^x, \boldsymbol{c}_{t,k}^y \\ \boldsymbol{r}_{t,k}^x, \boldsymbol{r}_{t,k}^y}} f_{t,k}(\boldsymbol{F}_t) \quad (4.9)$$

$$subject\ to: \quad \begin{cases} \boldsymbol{o}_{t,k}^x = \boldsymbol{D}^{cx}\left(\boldsymbol{u}_{t,k}^{x*} + \boldsymbol{c}_{t,k}^x\right) + \boldsymbol{r}_{t,k}^x \\ \boldsymbol{o}_{t,k}^y = \boldsymbol{D}^{cy}\left(\boldsymbol{u}_{t,k}^{y*} + \boldsymbol{c}_{t,k}^y\right) + \boldsymbol{r}_{t,k}^y \end{cases} \quad (4.10)$$

$$\begin{cases} \boldsymbol{s}_{t,k}^x = \boldsymbol{s}_{t,k-1}^x + \boldsymbol{o}_{t,k}^x \\ \boldsymbol{s}_{t,k}^y = \boldsymbol{s}_{t,k-1}^y + \boldsymbol{o}_{t,k}^y \end{cases} \quad \begin{cases} \boldsymbol{s}_{t,0}^x = 0 \\ \boldsymbol{s}_{t,0}^y = 0 \end{cases} \quad (4.11)$$

$$\begin{cases} \boldsymbol{r}_{t,k}^{lb,x} \leq \boldsymbol{r}_{t,k}^{x} \leq \boldsymbol{r}_{t,k}^{ub,x} \\ \boldsymbol{r}_{t,k}^{lb,y} \leq \boldsymbol{r}_{t,k}^{y} \leq \boldsymbol{r}_{t,k}^{ub,y} \end{cases} \tag{4.12}$$

$$\begin{cases} \boldsymbol{c}_{t,k}^{lb,x} \leq \boldsymbol{c}_{t,k}^{x} \leq \boldsymbol{c}_{t,k}^{ub,x} \\ \boldsymbol{c}_{t,k}^{lb,y} \leq \boldsymbol{c}_{t,k}^{y} \leq \boldsymbol{c}_{t,k}^{ub,y} \end{cases} \tag{4.13}$$

where constraints (4.10) and (4.11) respectively calculate overlay and stack-up errors in all the candidate markers, while constraints (4.12) and (4.13) provide lower and upper bounds for the uncertain process bias and residual vector terms in the overlay models.

With the worst-case overlay control objective $J_{t,k}(\boldsymbol{F}_t)$ describing performance of any given measurement scheme $\boldsymbol{F}_t$ on layer *k* of wafer *t*, and given a constraint that one wishes to use only P_{obj} markers on wafer *t*, the best-performing set of measurement markers $\boldsymbol{F}_t^*$ on wafer *t* can be pursued by solving the following optimization problem:

$$\boldsymbol{F}_t^* = \underset{\boldsymbol{F}_t \in \mathbb{Z}_2^P}{argmin} \sum_{k=1}^{K} J_{t,k}(\boldsymbol{F}_t) \tag{4.14}$$

$$subject\ \ to: \quad \sum_{i=1}^{P} F_{t,i} = P_{obj} \tag{4.15}$$

One can observe that objective function (4.14) minimizes the sum of worst-case control performance evaluated on all available candidate markers, across all layers of wafer *t*, while constraint (4.15) restricts the number of selected markers on wafer *t* to be P_{obj}.

Based on the problem formulation proposed above, let us now express the process of evaluating the objective function in (4.14) for each candidate measurement selection $\boldsymbol{F}_t$. Given boundaries on the uncertain, stochastic terms for wafer *t*, the following procedure is conducted, starting with the first layer $k = 1$.

Step 1. *For layer k, solve the robust control problem (4.4)-(4.7) to obtain control commands $\boldsymbol{u}_{t,k}^{x*}$ and $\boldsymbol{u}_{t,k}^{y*}$.*

Control problem (4.4)-(4.7) is a robust least-squares optimization problem, which can be solved efficiently, as described in [131]. Solution of the robust control problem[17] (4.4)-(4.7) is based on reformulating it into a copositive programming that admits a conservative semidefinite programming approximation, which can then be efficiently solved using commercial solvers, such as MOSEK [132].

Step 2. *For layer k, solve optimization problem (4.9)-(4.13) and thus characterize worst-case performance of the resulting robust control law, as evaluated on all markers in layer k of wafer t.*

Given the control commands $\boldsymbol{u}_{t,k}^{x*}$ and $\boldsymbol{u}_{t,k}^{y*}$ obtained in *Step 1*, the worst-case objective function $J_{t,k}(\boldsymbol{F}_t)$ is obtained by solving the optimization problem (4.9)-

[17] As mentioned earlier, recent research already employed the use of those methodologies for robust control of photolithography overlay errors [127].

(4.13). This is a concave quadratic programming problem, with numerous approaches available to tackle it [133][134][135][136], including implementations in commercial solvers, such as CPLEX.

Step 3. *If* $k < K$*, obtain the values for stack-up overlays* $\boldsymbol{s}_{t,k}^{x}$ *and* $\boldsymbol{s}_{t,k}^{y}$ *corresponding to the worst-case performance* $J_{t,k}(\boldsymbol{F}_t)$ *and feed them into the next layer* $k+1$*. Then, let* $k \leftarrow k+1$ *and go the Step 1. If* $k = K$*, add up layer-specific objectives* $J_{t,k}(\boldsymbol{F}_t)$ *for all layers and thus obtain the value for the objective function (4.14).*

4.2.3. **Genetic Algorithm based Optimization Framework**

A genetic algorithm (GA) based approach is pursued in this chapter to solve the combinatorial optimization problem (4.14)-(4.15) and thus obtain the best subset of overlay measurement markers for wafer *t*.

The GA starts from an initial population consisting of N_G randomly generated feasible candidate solutions $\boldsymbol{F}_t$ to the optimization problem (4.14)-(4.15). Performance of each candidate solution is expressed as the sum across all layers *k* of the worst-case layer-specific objective functions $J_{t,k}$, as formulated in (4.14), with better performing candidates yielding smaller objective functions (4.14). Starting with the initial generation of candidate solutions, successive selection, chromosome crossover, repair and mutation operators are performed following the typical GA heuristics of natural selection and survival of the

fittest, resulting in successive generations of increasingly better performing candidates. More detailed description of how these GA operators were implemented in this chapter is given below, with Figure 1 providing an illustrative example of how GA crossover, repair, and mutation operators were conducted.

1) Selection operator

Through this operator, candidate solutions from the current generation of solutions are selected for reproduction and creation of the next generation of solutions, with the heuristic that better performing individuals are selected with higher probability. In this chapter, the so-called tournament selection is implemented, since it is considered to work well with a wide array of fitness functions and generally leads to a higher convergence rate than the purely proportional selection method [137]. In this method, n out of N_G individuals are randomly picked with an equal chance and the best of them is selected. The selection procedure is repeated N_S times, yielding N_S individuals which are selected for crossover.

2) Crossover operator

A single-point crossover operator is applied to pairs of selected chromosomes [138][139]. Assume the parent chromosomes are denoted by $[a_1, a_2, \dots, a_P]$ and $[b_1, b_2, \dots, b_P]$. Then, the offspring chromosomes after crossover can be described as

$$offspring_1 = [a_1, a_2, \dots, a_l, b_{l+1}, b_{l+2}, \dots, b_P]$$
$$offspring_2 = [b_1, b_2, \dots, b_l, a_{l+1}, a_{l+2}, \dots, a_P]$$

where point $l \in \{1,2,\ldots,P\}$ is randomly selected to break the parent chromosomes.

3) Repair operator

Through this operator, offspring solutions are repaired so that the number of selected markers is fixed in order to satisfy the constraint (4.15). As candidate solutions are represented as binary chromosome vectors, if the number of ones in the vector is smaller than P_{obj}, appropriate number of zeros in the chromosome vector are randomly selected and changed to ones in order to meet the constraint (4.15), and *vice versa.*

4) Mutation operator

Through this operator, ones and zeros in each offspring solution are randomly switched with a small probability to increase diversity of candidate solutions and help the GA jump out of local optima [140][141]. The mutation rates are usually selected to be small to enable steady convergence.

Finally, let us note that in order to increase the convergence rate and stability of the GA, elitism is invoked by preserving the best performing solution from the previous generation, if it is better performing than the best candidate solution in the new generation of solutions [142][143].

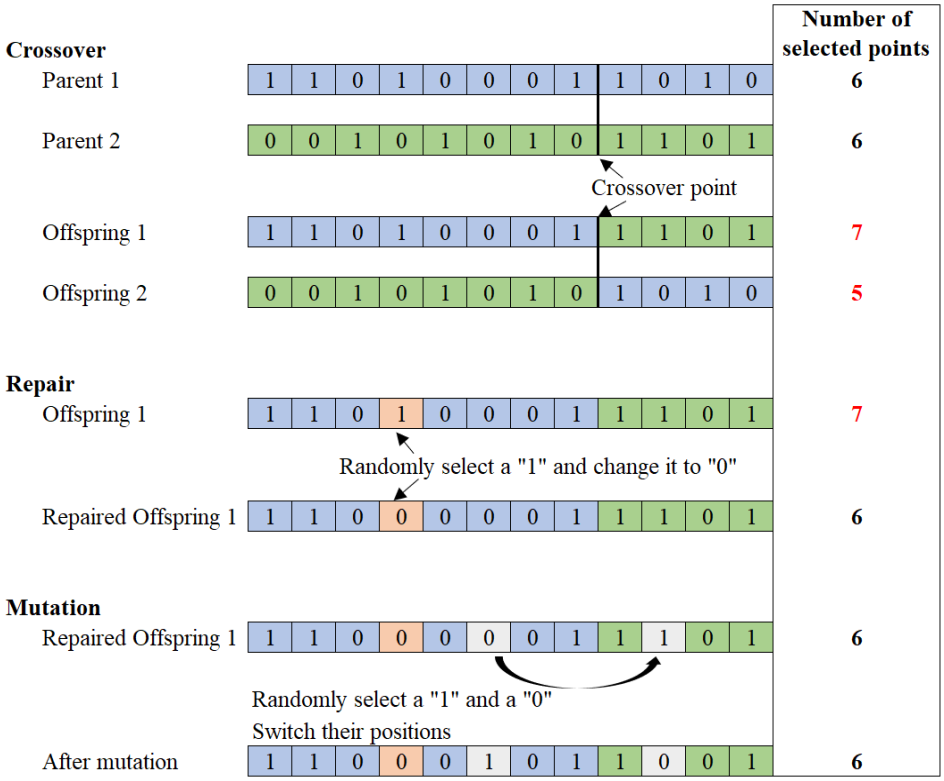

Figure 5. Top of the figure gives an example of the crossover operator applied on two parent candidate solutions. Middle and bottom of the figure illustrate repair and mutation operators on one of their offspring, with the constraint on the number of selected measurement markers being 6.

4.2.4. **Establishment and Maintenance of Boundaries on the Uncertain Terms in Overlay Error Models**

This section describes a statistically inspired procedure based on which boundaries on the modeling residuals and bias terms in constraints (4.6) and (4.7) can be established and maintained.

Let us assume that we have control signals and overlay errors measured at all the available markers collected from a set of T_0 historical wafers. Results of such a pilot production run can be used to initialize the boundaries in constraints (4.6) and (4.7) for the next wafer $T_0 + 1$ via the following procedure.

Zernike polynomial-based overlay models (4.1) and overlay errors $\boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}^{Tot})$ and $\boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}^{Tot})$ measured in all available markers $\boldsymbol{F}^{Tot}$ on wafers $t \in \{1,2,\dots,T_0\}$ and pattern layers $k \in \{1,2,\dots,K\}$ can be used to obtain least-squares based estimates $\widehat{\boldsymbol{u}}_{t,k}^{x}(\boldsymbol{F}^{Tot})$ and $\widehat{\boldsymbol{u}}_{t,k}^{y}(\boldsymbol{F}^{Tot})$ of the corresponding control commands $\boldsymbol{u}_{t,k}^{x}$ and $\boldsymbol{u}_{t,k}^{y}$ as

$$\begin{cases} \widehat{\boldsymbol{u}}_{t,k}^{x}(\boldsymbol{F}^{Tot}) = \left(\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{xT}\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}\right)^{-1}\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{xT}\boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}^{Tot}) \\ \widehat{\boldsymbol{u}}_{t,k}^{y}(\boldsymbol{F}^{Tot}) = \left(\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{yT}\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{y}\right)^{-1}\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{yT}\boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}^{Tot}) \end{cases} \tag{4.16}$$

where $\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}$ and $\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{y}$ are regression matrices of the overlay error models (4.1) corresponding to the use of all available markers on the wafer. The corresponding overlay model residuals for all historical wafers $t \in \{1,2,\dots,T_0\}$ and all pattern layers $k \in \{1,2,\dots,K\}$ can be obtained as

$$\begin{cases} \boldsymbol{r}_{t,k}^{x}(\boldsymbol{F}^{Tot}) = \boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}^{Tot}) - \boldsymbol{D}_{\boldsymbol{F}_t^*}^{x}\widehat{\boldsymbol{u}}_{t,k}^{x}(\boldsymbol{F}^{Tot}) \\ \boldsymbol{r}_{t,k}^{y}(\boldsymbol{F}^{Tot}) = \boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}^{Tot}) - \boldsymbol{D}_{\boldsymbol{F}_t^*}^{y}\widehat{\boldsymbol{u}}_{t,k}^{y}(\boldsymbol{F}^{Tot}) \end{cases} \tag{4.17}$$

based on which boundaries on modeling residual constraints (4.6) for wafer $T_0 + 1$ can be established using ranges within which those historical residuals reside in each marker location on the wafer. Furthermore, residuals (4.17) can be also used to estimate variances

$\left[\sigma_{r_{t,k}}^{x}\right]^2$ and $\left[\sigma_{r_{t,k}}^{y}\right]^2$ associated with each wafer and each pattern layer[18], based on which variance-covariance matrices describing errors associated with the estimation of actual control commands in each layer of each historical wafer can be calculated as [144]

$$\begin{cases}\left[\boldsymbol{\sigma}_{\hat{u}_{t,k}}^{x}(\boldsymbol{F}^{Tot})\right]^2 = \left(\left[\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}\right]^T \cdot \boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}\right)^{-1}\left[\sigma_{r_{t,k}}^{x}\right]^2 \\ \left[\boldsymbol{\sigma}_{\hat{u}_{t,k}}^{y}(\boldsymbol{F}^{Tot})\right]^2 = \left(\left[\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}\right]^T \cdot \boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{y}\right)^{-1}\left[\sigma_{r_{t,k}}^{y}\right]^2\end{cases} \quad (4.18)$$

Since, as stated in (4.2), bias vector terms represent a difference between the actually realized control commands, which are stochastic, and the desired control signals, which consist of constants calculated by the corresponding control algorithm, variance-covariance matrices $\left[\boldsymbol{\sigma}_{\hat{c}_{t,k}}^{x}(\boldsymbol{F}^{Tot})\right]^2$ and $\left[\boldsymbol{\sigma}_{\hat{c}_{t,k}}^{y}(\boldsymbol{F}^{Tot})\right]^2$ describing errors in the estimation of process bias vectors will be the same as variance-covariance matrices $\left[\boldsymbol{\sigma}_{\hat{u}_{t,k}}^{x}(\boldsymbol{F}^{Tot})\right]^2$ and $\left[\boldsymbol{\sigma}_{\hat{u}_{t,k}}^{y}(\boldsymbol{F}^{Tot})\right]^2$ associated with the error of estimation of control signals, i.e.

$$\begin{cases}\left[\boldsymbol{\sigma}_{\hat{c}_{t,k}}^{x}(\boldsymbol{F}^{Tot})\right]^2 = \left(\left[\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}\right]^T \cdot \boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}\right)^{-1}\left[\sigma_{r_{t,k}}^{x}\right]^2 \\ \left[\boldsymbol{\sigma}_{\hat{c}_{t,k}}^{y}(\boldsymbol{F}^{Tot})\right]^2 = \left(\left[\boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{x}\right]^T \cdot \boldsymbol{D}_{\boldsymbol{F}^{Tot}}^{y}\right)^{-1}\left[\sigma_{r_{t,k}}^{y}\right]^2\end{cases} \quad (4.19)$$

[18]Please note that those variance estimates can also be used to establish boundaries (6) on the modeling residuals as some multiple of the corresponding variance.

Estimated bias terms $\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot})$ and $\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot})$, as well as variance-covariance matrices $\left[\boldsymbol{\sigma}_{\hat{c}_{t,k}}^{x}(\boldsymbol{F}^{Tot})\right]^2$ and $\left[\boldsymbol{\sigma}_{\hat{c}_{t,k}}^{y}(\boldsymbol{F}^{Tot})\right]^2$ associated with the corresponding estimation errors for all historical wafers $t \in \{1,2,\dots,T_0\}$ and all pattern layers $k \in \{1,2,\dots,K\}$ can be used to establish boundaries for the bias terms on the next wafer T_0+1.

Given that the procedure described above naturally gives estimates, as well as estimation uncertainties associated with bias vector terms, in this chapter, we pursued establishing of those boundaries by building a heteroscedastic Gaussian process regression (GPR) based R2R prediction model [145], yielding the predicted process bias vectors $\tilde{\hat{\boldsymbol{c}}}_{T_0+1,k}^{x}$ and variance-covariance matrices $\left[\boldsymbol{\sigma}_{c_{T_0+1,k}}^{x}\right]^2$ characterizing the corresponding prediction errors for all pattern layers $k=1,2,\dots,K$ on wafer T_0+1. This prediction model also gives a history of prediction errors on the historical set of wafers $t \in \{1,2,\dots,T_0\}$, based on which vectors of ranges $\boldsymbol{R}_{c,1\dots T_0,k}^{x}$ and $\boldsymbol{R}_{c,1\dots T_0,k}^{y}$ can be established to bracket how far from predicted bias vector terms the corresponding biases estimated from the full measurement scheme on wafers $t \in \{1,2,\dots,T_0\}$ actually fell[19]. Since bias vectors for wafer T_0+1 are predicted from a full set of markers, just like they were on the previous wafers $t \in \{1,2,\dots,T_0\}$, ranges on the actual bias vectors would then be set

[19]This being estimates of actual bias vector terms with highest confidence because they are obtained from the full set of markers $\boldsymbol{F}^{Tot}$.

to

$$\begin{cases} \boldsymbol{c}_{T_0+1,k}^{lb,x} = \tilde{\hat{\boldsymbol{c}}}_{T_0+1,k}^{x} - \boldsymbol{R}_{c,1\dots T_0,k}^{x} \\ \boldsymbol{c}_{T_0+1,k}^{lb,y} = \tilde{\hat{\boldsymbol{c}}}_{T_0+1,k}^{y} - \boldsymbol{R}_{c,1\dots T_0,k}^{y} \\ \boldsymbol{c}_{T_0+1,k}^{ub,x} = \tilde{\hat{\boldsymbol{c}}}_{T_0+1,k}^{x} + \boldsymbol{R}_{c,1\dots T_0,k}^{x} \\ \boldsymbol{c}_{T_0+1,k}^{ub,y} = \tilde{\hat{\boldsymbol{c}}}_{T_0+1,k}^{y} + \boldsymbol{R}_{c,1\dots T_0,k}^{y} \end{cases} \quad (4.20)$$

Given the boundaries of stochastic terms for wafer $T_0 + 1$, the novel method described in Sec. 4.2.1-4.2.3 can be used to find the optimal set of markers $\boldsymbol{F}_{T_0+1}^{*}$ for the wafer $T_0 + 1$. Once that wafer is manufactured, actual overlay errors $\boldsymbol{o}_{T_0+1,k}^{x}(\boldsymbol{F}_{T_0+1}^{*})$ and $\boldsymbol{o}_{T_0+1,k}^{y}(\boldsymbol{F}_{T_0+1}^{*})$, as well as stack-up overlay errors $\boldsymbol{s}_{T_0+1,k}^{x}(\boldsymbol{F}_{T_0+1}^{*})$ and $\boldsymbol{s}_{T_0+1,k}^{y}(\boldsymbol{F}_{T_0+1}^{*})$ can be observed at selected markers $\boldsymbol{F}_{T_0+1}^{*}$ in all pattern layers $k = 1,2,\dots,K$ of the wafer $T_0 + 1$. Furthermore, all actually commanded control parameters on the photolithography tool are also known. Then, estimates of actual process bias vectors $\hat{\boldsymbol{c}}_{T_0+1,k}^{x}(\boldsymbol{F}_{T_0+1}^{*})$ and $\hat{\boldsymbol{c}}_{T_0+1,k}^{y}(\boldsymbol{F}_{T_0+1}^{*})$, as well as the variance-covariance matrices $\left[\boldsymbol{\sigma}_{\hat{c}_{T_0+1,k}}^{x}(\boldsymbol{F}_{T_0+1}^{*})\right]^2$ and $\left[\boldsymbol{\sigma}_{\hat{c}_{T_0+1,k}}^{y}(\boldsymbol{F}_{T_0+1}^{*})\right]^2$ characterizing the corresponding estimation errors can be obtained for all layers $k = 1,2,\dots,K$ of the wafer $T_0 + 1$ following the procedure described by (4.16)-(4.19) for historical wafers $t = 1,2,\dots,T_0$. These estimates, as well as estimates of the corresponding modeling residuals can then be added to the historical records, which now consist of wafers $t \in \{1,2,\dots,T_0 + 1\}$. Furthermore, one can now calculate new predictions of bias vectors $\tilde{\hat{\boldsymbol{c}}}_{T_0+2,k}^{x}$ and $\tilde{\hat{\boldsymbol{c}}}_{T_0+2,k}^{y}$, as well as the corresponding variance-covariance matrices of prediction errors $\left[\boldsymbol{\sigma}_{c_{T_0+2,k}}^{x}(\boldsymbol{F}_{T_0+1}^{*})\right]^2$ and

$\left[\boldsymbol{\sigma}^{y}_{c_{T_0+2,k}}(\boldsymbol{F}^{*}_{T_0+1})\right]^2$ using the heteroscedastic GPR-based R2R prediction. Boundaries on the modeling residuals for the future wafer T_0+2 can be obtained by e.g. updating the corresponding ranges based on the newly observed residuals from wafer T_0+1, while boundaries on the bias vector terms can be established as

$$\begin{cases} \boldsymbol{c}^{lb,x}_{T_0+2,k} = \tilde{\tilde{\boldsymbol{c}}}^{x}_{T_0+2,k} - \boldsymbol{R}^{x}_{c,T_0+1,k} \\ \boldsymbol{c}^{lb,y}_{T_0+2,k} = \tilde{\tilde{\boldsymbol{c}}}^{y}_{T_0+2,k} - \boldsymbol{R}^{y}_{c,T_0+1,k} \\ \boldsymbol{c}^{ub,x}_{T_0+2,k} = \tilde{\tilde{\boldsymbol{c}}}^{x}_{T_0+2,k} + \boldsymbol{R}^{x}_{c,T_0+1,k} \\ \boldsymbol{c}^{ub,y}_{T_0+1,k} = \tilde{\tilde{\boldsymbol{c}}}^{y}_{T_0+1,k} + \boldsymbol{R}^{y}_{c,T_0+1,k} \end{cases} \tag{4.21}$$

where ranges $\boldsymbol{R}^{x}_{c,T_0+1,k}$ and $\boldsymbol{R}^{y}_{c,T_0+1,k}$ can be obtained by scaling historical ranges $\boldsymbol{R}^{x}_{c,1\ldots T_0,k}$ and $\boldsymbol{R}^{y}_{c,1\ldots T_0,k}$ obtained from the first T_0 wafers according to newly obtained prediction uncertainty matrices $\left[\boldsymbol{\sigma}^{x}_{c_{T_0+2,k}}(\boldsymbol{F}^{*}_{T_0+1})\right]^2$ and $\left[\boldsymbol{\sigma}^{y}_{c_{T_0+2,k}}(\boldsymbol{F}^{*}_{T_0+1})\right]^2$. Namely, i^{th} element in $\boldsymbol{R}^{x}_{c,T_0+1,k}$ would be a scaled version of the i^{th} element in $\boldsymbol{R}^{x}_{c,1\ldots T_0,k}$ as

$$\left[\boldsymbol{R}^{x}_{c,T_0+1,k}\right]_i = \frac{\left[\boldsymbol{\sigma}^{x}_{c_{T_0+2,k}}\left(\boldsymbol{F}^{*}_{T_0+1}\right)\right]_{i,i}}{\left[\boldsymbol{\sigma}^{x}_{c_{1\ldots T_0,k}}(\boldsymbol{F}^{Tot})\right]_{i,i}} \cdot \left[\boldsymbol{R}^{x}_{c,1\ldots T_0,k}\right]_i \tag{4.22}$$

where $\left[\boldsymbol{\sigma}^{x}_{c_{T_0+2,k}}\left(\boldsymbol{F}^{*}_{T_0+1}\right)\right]_{i,i}$ denotes standard deviation of prediction errors for the i^{th} component in the bias vector for wafer T_0+2, $\left[\boldsymbol{\sigma}^{x}_{c_{1\ldots T_0,k}}(\boldsymbol{F}^{Tot})\right]_{i,i}$ denotes the corresponding component for bias vector predictions on historical wafers 1 through T_0. Equation (4.22) essentially models the influence of measurement selection on the

boundaries of bias vector predictions and would essentially lead to expansion of corresponding boundaries as one selects fewer measurement markers, where that expansion is proportional to the corresponding prediction uncertainties.

With updated boundaries on the modeling residuals and bias predictions for wafer $T_0 + 2$, one can now pursue optimal set of markers $\boldsymbol{F}^*_{T_0+2}$ for wafer $T_0 + 2$ following procedures described in Sec. 4.2.1- 4.2.3, after which the above-described procedure of predictions and boundary updates continues as production progresses.

4.3. **Simulation Process and Experimental Results**

The newly proposed method is evaluated using the lithography overlay error model and data corresponding to a 4-layer industrial photolithography process used in a major 300 mm fab. Overlay errors and commanded control inputs for 80 consecutive wafers from the actual production process are used as historical data to initialize the simulation process. The weight parameters in the objective functions (4.8) are set to $\lambda^x = \lambda^y = \alpha^x = \alpha^y = 1$. Please note that these weight factors do not reflect the actual relative importance of layers in the process, which needed to be concealed due to proprietary nature of the process.

4.3.1. **Simulation Results for the First Wafer Outside the Initial Wafer Set (wafer no. 81)**

Control signals and overlay measurements from all available markers collected from the 80 historical wafers were used to initialize boundaries on the modeling residuals and bias terms for the 81st wafer, using (4.20) and (4.21) in the procedure described in Section 4.2.4. Given these boundaries, the newly proposed procedure described in Sec. 4.2.1-4.2.3 was used to obtain optimal sets of markers for the 81st wafer under varying constraints on the number of markers that need to be retained.

Figure 6 shows the layer-specific (a) worst-case objective function $J_{81,k}(\boldsymbol{F}_{81}^{*})$, (b) worst-case overlay error magnitudes, and (c) worst-case stack-up overlay error magnitudes for wafer 81, obtained under increasingly stringent constraints on the number of overlay measurement markers that were to be kept in the measurement scheme $\boldsymbol{F}_{81}^{*}$. The results clearly show that when compared to the traditional R2R method, which one should note uses all available markers all the time, the newly proposed method combining robust overlay control with optimized marker selection performs better in terms of all 3 criteria shown in Figure 6, even when the percentage of selected markers is decreased to as far down as 60% of available markers. Only when we go down to selecting just 50% of the available makers can we see a situation in which the worst-case overlay and stack-up overlay error magnitudes at layer 1 become greater than what is obtained using the R2R

control method[20].

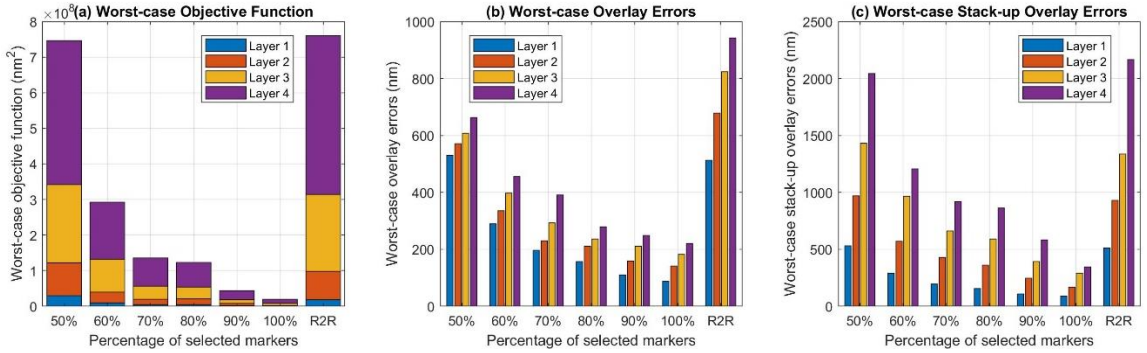

Figure 6. Comparison of the worst-case performance between the traditional R2R method and the newly proposed robust selection method with various percentages of selected markers for wafer 81. These bar plots present (a) worst-case outcome of objective function, (b) worst-case overlay error magnitudes, and (c) worst-case stack-up overlay error magnitudes.

In order to observe details of the distributions of performance metrics of the robust control algorithm under various levels of measurement constraints, for each optimized measurement scheme $\boldsymbol{F}_{81}^{*}$ we conducted 200 Monte Carlo simulations of the wafer no. 81, with residuals and process bias terms simulated using various distributions supported by the corresponding boundaries.

For the case of uniform distributions of uncertain model terms, distributions of layer-specific objective functions $f_{81,k}(\boldsymbol{F}_{81}^{*})$ under increasingly stringent constraints for the

[20]Once again, let us repeat that unlike our proposed method, traditional R2R overlay control paradigm uses all available markers, all the time.

measurement marker selection are illustrated by box-and-whisker plots shown in Figure 7, while the corresponding means and standard deviations are shown in Figure 8. In both cases, it can be seen that one can select to as few as 60% of markers, and the newly proposed approach combining robust overlay control with optimal down-selection of measurement markers still outperforms the traditional R2R method, which, as mentioned earlier, uses all available markers all the time.

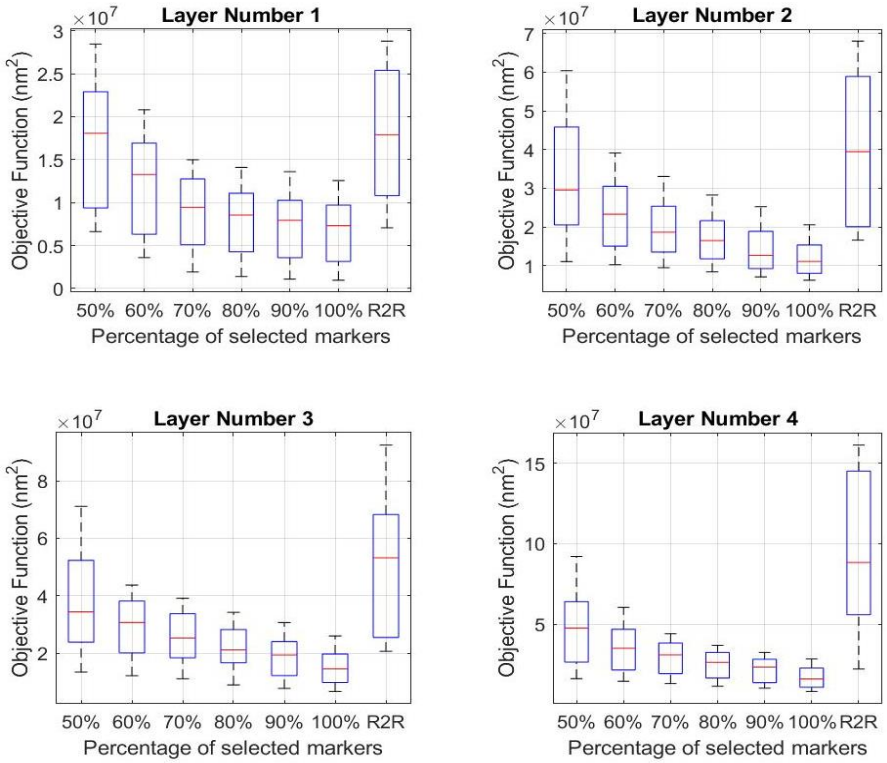

Figure 7. Box-and-whisker plots describing distributions of layer-specific simulated objective function $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 81 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

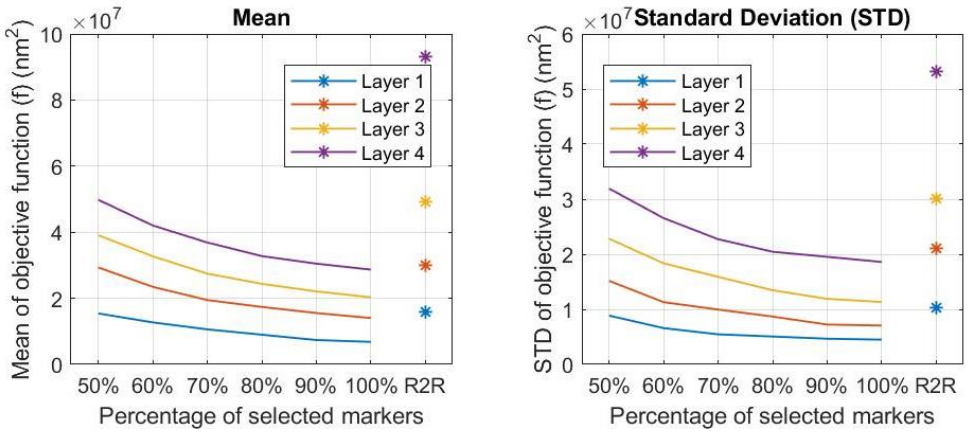

Figure 8. Plots of layer-specific mean and standard deviation of simulated objective functions $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 81 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

More details regarding performance metrics of the two overlay control algorithms can be seen in Figure 9 and Figure 10, with Figure 9 showing percentiles of the distributions of layer-specific overlay error magnitudes, while Figure 10 shows the corresponding means and standard deviations. Yet again, the use of robust overlay control algorithm and optimal selection of overlay measurement markers allows one to down-select measurement as few as 60% of available markers and still outperform the R2R method in terms of all metrics, and it does so in all layers.

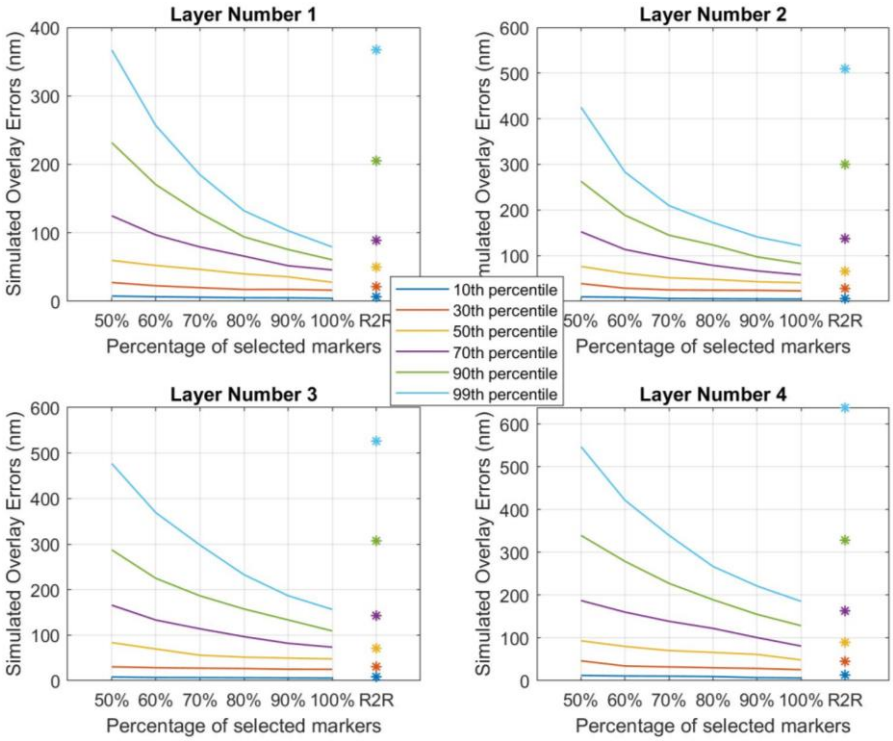

Figure 9. Plots of percentiles describing distributions of layer-specific simulated overlay error magnitudes obtained from the 200 simulations of the wafer 81 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

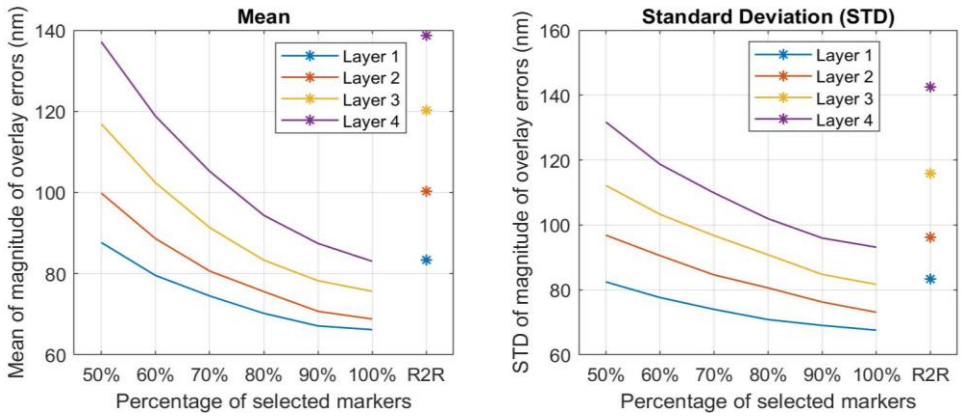

Figure 10. Plots of layer-specific mean and standard deviation of simulated overlay error magnitude obtained from the 200 simulations of the wafer 81 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

Behavior of stack-up overlay errors is analyzed in Figure 11 and Figure 12, with Figure 11 showing percentiles of distributions describing magnitudes of layer-specific stack-up overlay errors, while Figure 12 shows the corresponding means and standard deviations. Once again, one can observe results consistent with what is visible in earlier Figures, in the sense that one can remove as much as 40% of measurement markers and the robust control algorithm still outperforms the traditional R2R method.

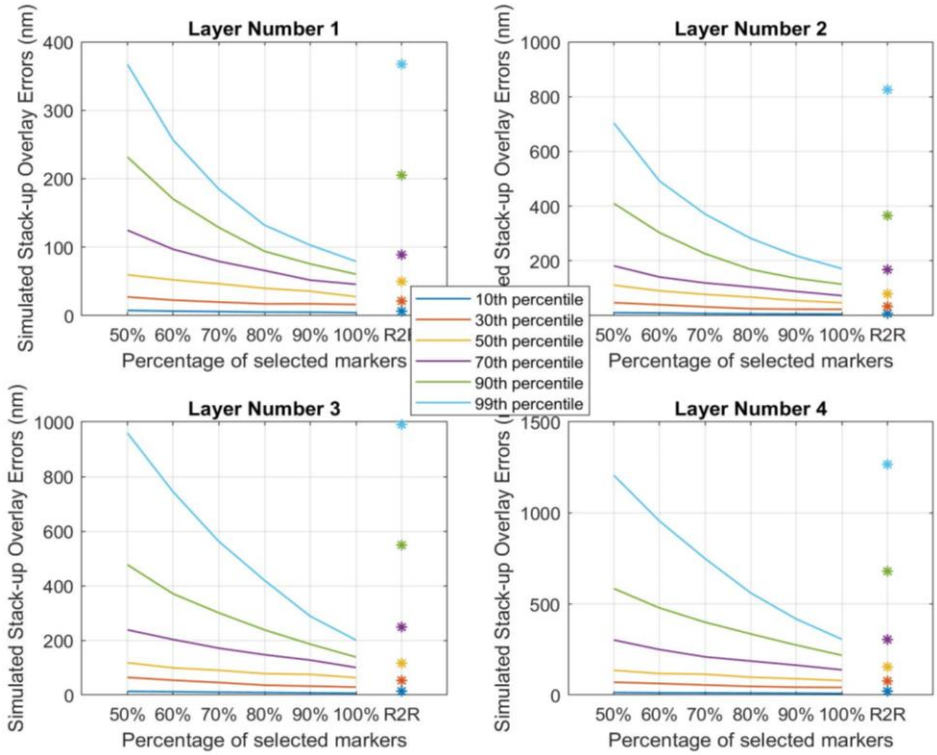

Figure 11. Plots of percentiles describing distributions of layer-specific simulated stack-up overlay error magnitudes obtained from the 200 simulations of the wafer 81 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

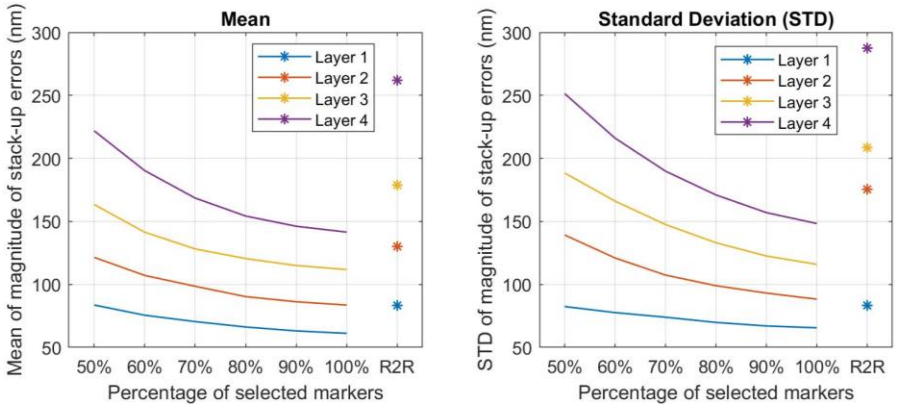

Figure 12. Plots of layer-specific mean and standard deviation of simulated stack-up overlay error magnitude obtained from the 200 simulations of the wafer 81 with uniformly distributed uncertainties Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

For the case of truncated normal distributions of uncertain model terms, we assume that variances of modeling residuals are estimated from the 80 historical wafers, while the means and variances of process bias terms are set to be their predicted value and prediction variance, with these normal distributions being truncated at corresponding boundaries of the stochastic terms. Distributions of layer-specific objective functions $f_{81,k}(\boldsymbol{F}_{81}^{*})$ under increasingly stringent constraints for the measurement marker selection are illustrated by box-and-whisker plots shown in Figure 13, while the corresponding means and standard deviations are shown in Figure 14.

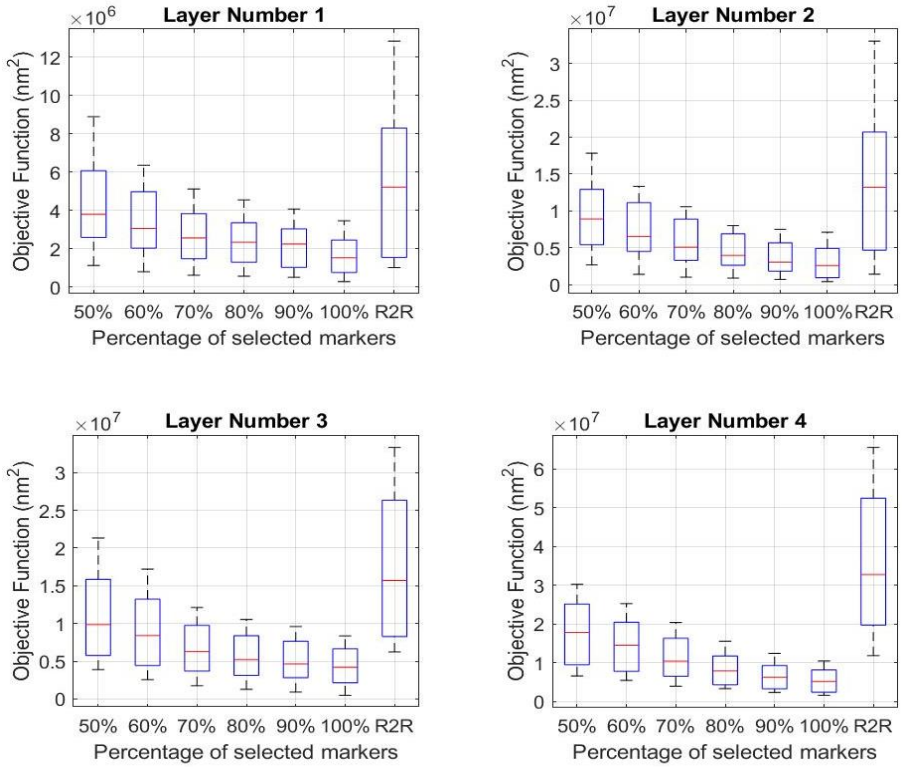

Figure 13. Box-and-whisker plots describing distributions of layer-specific simulated objective function $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 81 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

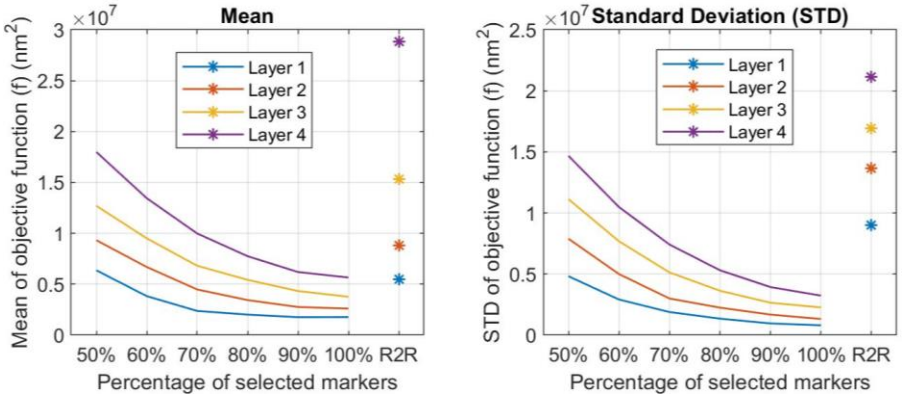

Figure 14. Plots of layer-specific mean and standard deviation of simulated objective functions $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 81 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

Behavior of overlay and stack-up over errors is analyzed in Figure 15 to Figure 18, where Figure 15 and Figure 16 show percentiles, means and standard deviations of layer-specific overlay error magnitudes, while Figure 17 and Figure 18 show percentiles, means and standard deviations of layer-specific stack-up overlay error magnitudes. One can observe that even though the magnitudes of the objective function, overlay and stack-up overlay error become much smaller with the Gaussianity assumption for both control methods analyzed in this chapter, the newly proposed method combining robust overlay control with the optimal down-selection of overlay measurement markers still allows one to utilize as few as 60% of available markers and still outperform the R2R method in terms of all metrics, in all layers of the wafer.

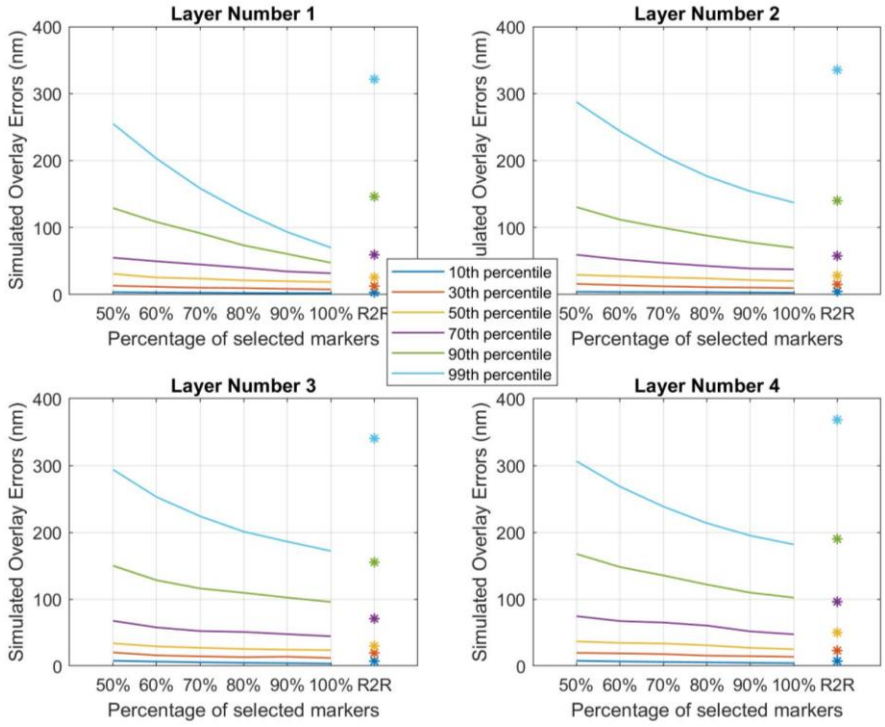

Figure 15. Plots of percentiles describing distributions of layer-specific simulated overlay error magnitudes obtained from the 200 simulations of the wafer 81 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

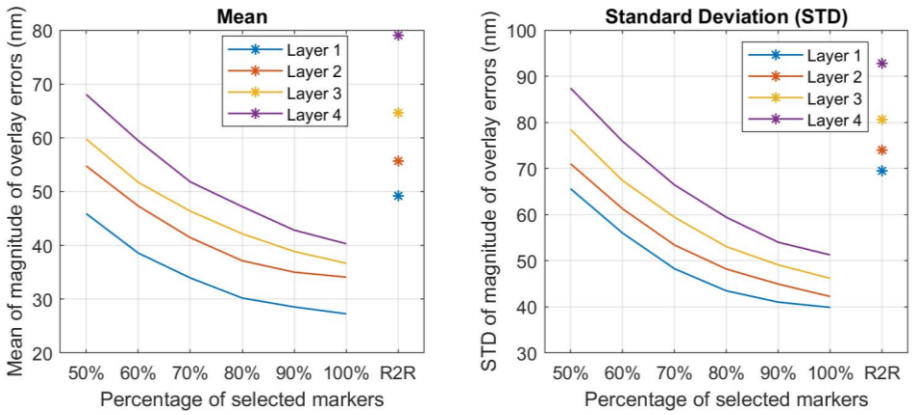

Figure 16. Plots of layer-specific mean and standard deviation of simulated overlay error magnitude obtained from the 200 simulations of the wafer 81 where uncertainties follow

truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

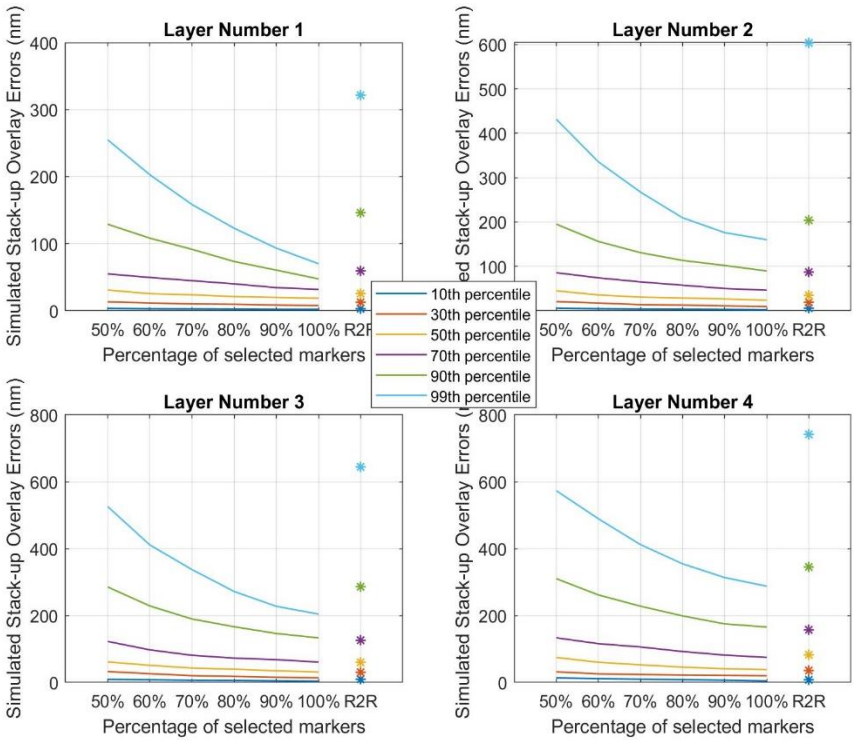

Figure 17. Plots of percentiles describing distributions of layer-specific simulated stack-up overlay error magnitudes obtained from the 200 simulations of the wafer 81 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and are compared against the traditional R2R method which uses all the available markers.

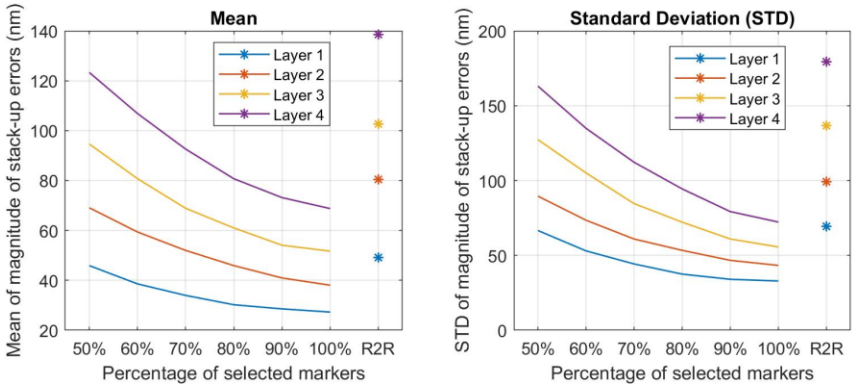

Figure 18. Plots of layer-specific means and standard deviations of simulated stack-up overlay error magnitude obtained from the 200 simulations of the wafer 81 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and are compared against the traditional R2R method which uses all the available markers.

4.3.2. **Simulation Results for the 20th Wafer Outside the Initial Wafer Set (wafer no. 100)**

Model and data from the initial 80 wafers were used to conduct simulations and evaluate overlay control performance in wafers beyond the 81st wafer, following the process described in Section 4.2.4. Namely, after finding the optimal set of markers $\boldsymbol{F}_{81}^{*}$ for wafer 81, we sampled one realization of uncertain, stochastic terms from uniform distributions supported by the relevant boundaries. This realization was used to simulate overlay errors in selected markers $\boldsymbol{F}_{81}^{*}$ on the 81st wafer, which were then used to obtain estimation of actual process bias terms on the 81st wafer, as well as the variance of the corresponding estimation error, as per (4.18) and (4.19). Estimated bias and variance of estimation errors on wafer 81 were then used to obtain predictions of bias vector terms and

the corresponding prediction error variance for the 82nd wafer using heteroscedastic GPR regression, as well as to update upper and lower bounds on the vectors of bias terms and residuals for wafer 82, as described in Section 4.2.4. This procedure was repeated for 20 consecutive wafers (wafer no. 81 through wafer no. 100) and for brevity, only performance metrics for wafer no. 100 will be shown here, with an important note that similar observations could be made regarding all of the simulated wafers.

Figure 19 shows the layer-specific (a) worst-case objective function $J_{100,k}(\boldsymbol{F}_{100}^{*})$, (b) worst-case overlay error magnitudes, and (c) worst-case stack-up overlay error magnitudes for wafer 100, obtained under increasingly stringent constraints on the number of overlay measurement markers that are to be kept in the measurement scheme $\boldsymbol{F}_{100}^{*}$. Similar to what we observed for the 81st wafer, the results clearly show that when compared to the traditional R2R method, newly proposed method combining robust overlay control with optimized marker selection performs better in terms of all 3 criteria shown in Figure 19 even when the percentage of selected markers is decreased to as far down as 60% of available markers. Only when we go down to selecting just 50% of the available makers can we see a situation in which the worst-case overlay and stack-up overlay error magnitudes at layer 1 become greater than what is obtained using the R2R control method.

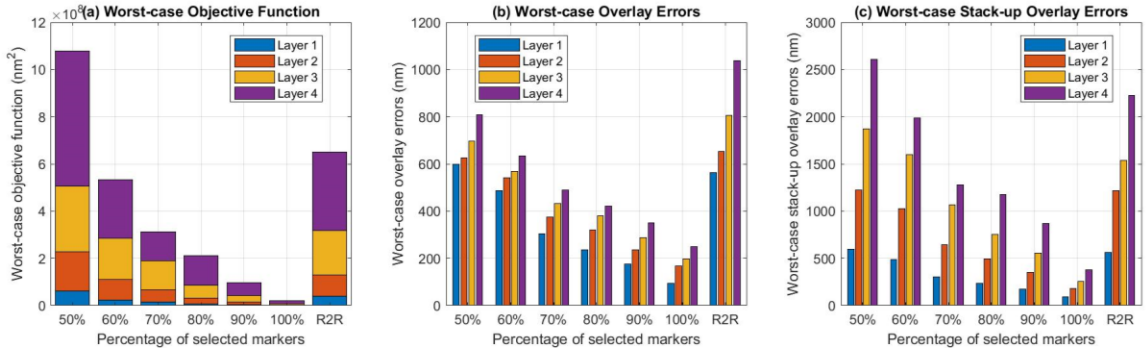

Figure 19. Comparison of the worst-case performance between the traditional R2R method and the newly proposed robust selection method with various percentage of selected markers for wafer 100. These bar plots present layer-specific (a) worst-case outcome of objective function, (b) worst-case overlay error magnitudes, and (c) worst-case stack-up overlay error magnitudes

Similar to what we did when analyzing overlay control performance associated with wafer no. 81, in order to observe details of the distributions of metrics describing performance of the robust control algorithm under various levels of measurement constraints, for each optimized measurement scheme $\boldsymbol{F}_{100}^{*}$ we conducted 200 Monte Carlo simulations of the wafer no.100, with residuals and process bias terms simulated using uniform and truncated normal distributions supported by the corresponding boundaries.

For the case of uniform distributions of uncertain model terms, distributions of layer-specific objective functions $f_{100,k}(\boldsymbol{F}_{100}^{*})$ under increasingly stringent constraints for the measurement marker selection are illustrated by box-and-whisker plots shown in Figure 20, while the corresponding means and standard deviations are shown in Figure 21.

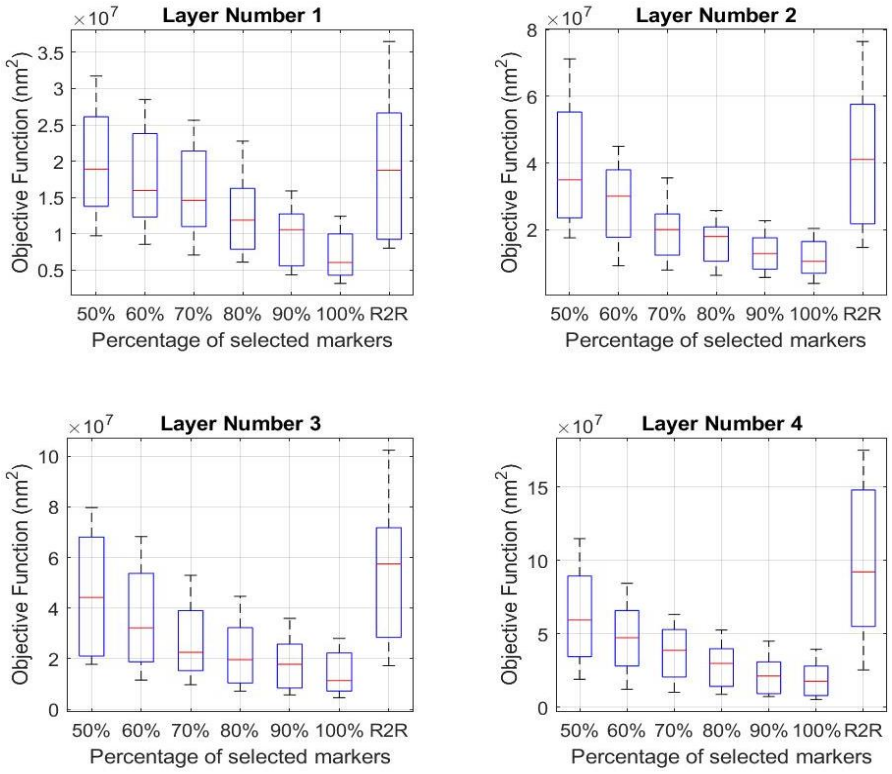

Figure 20. Box-and-whisker plots describing distributions of layer-specific simulated objective function $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 100 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

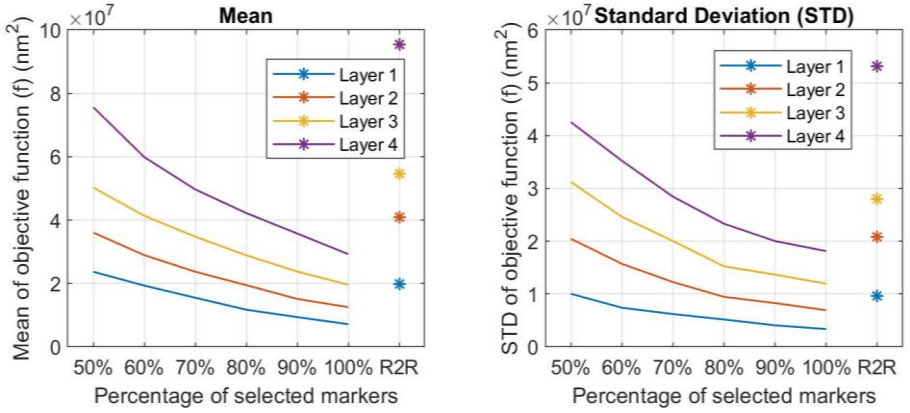

Figure 21. Plots of layer-specific mean and standard deviation of simulated objective functions $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 100 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

More details regarding the performance metrics of the two overlay control algorithms can be seen in Figure 22 and Figure 23, with Figure 22 showing percentiles of the distributions of layer-specific overlay error magnitudes, while Figure 23 shows the corresponding means and standard deviations. Behavior of stack-up overlay errors is analyzed in Figure 24 and Figure 25, with Figure 24 showing percentiles of distributions describing magnitudes of layer-specific stack-up overlay errors, while Figure 25 shows the corresponding means and standard deviations. It can be seen that one can select as few as 60% of markers, and the newly proposed approach combining robust overlay control with optimal down-selection of measurement markers still outperforms the traditional R2R method, which, as mentioned earlier, uses all available markers all the time.

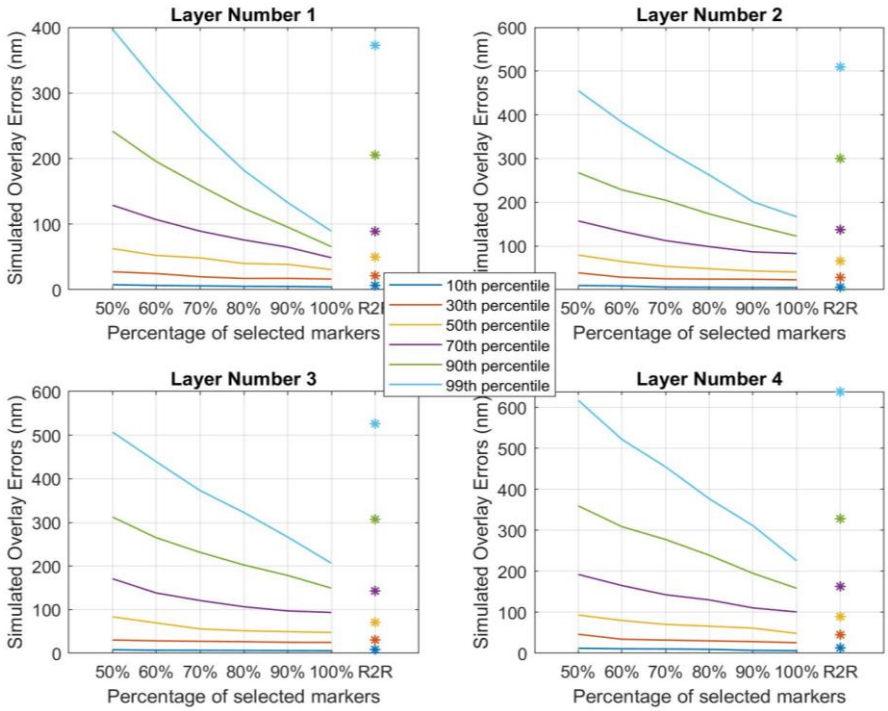

Figure 22. Plots of percentiles describing distributions of layer-specific simulated overlay error magnitudes obtained from the 200 simulations of the wafer 100 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

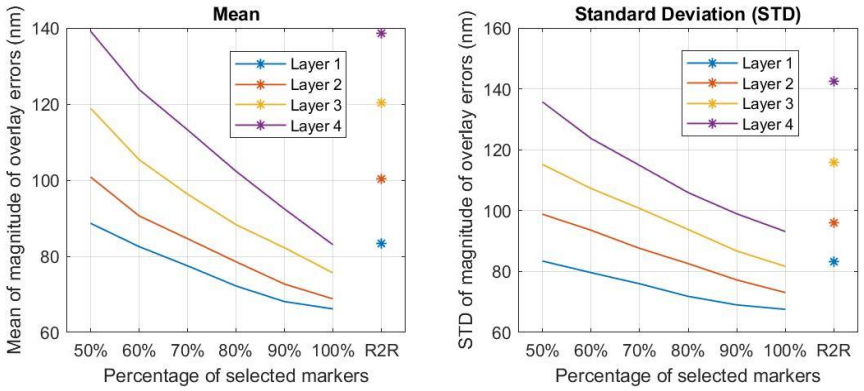

Figure 23. Plots of layer-specific mean and standard deviation of simulated overlay error magnitude obtained from the 200 simulations of the wafer 100 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with

various percentages of selected markers and the traditional R2R method using all the markers.

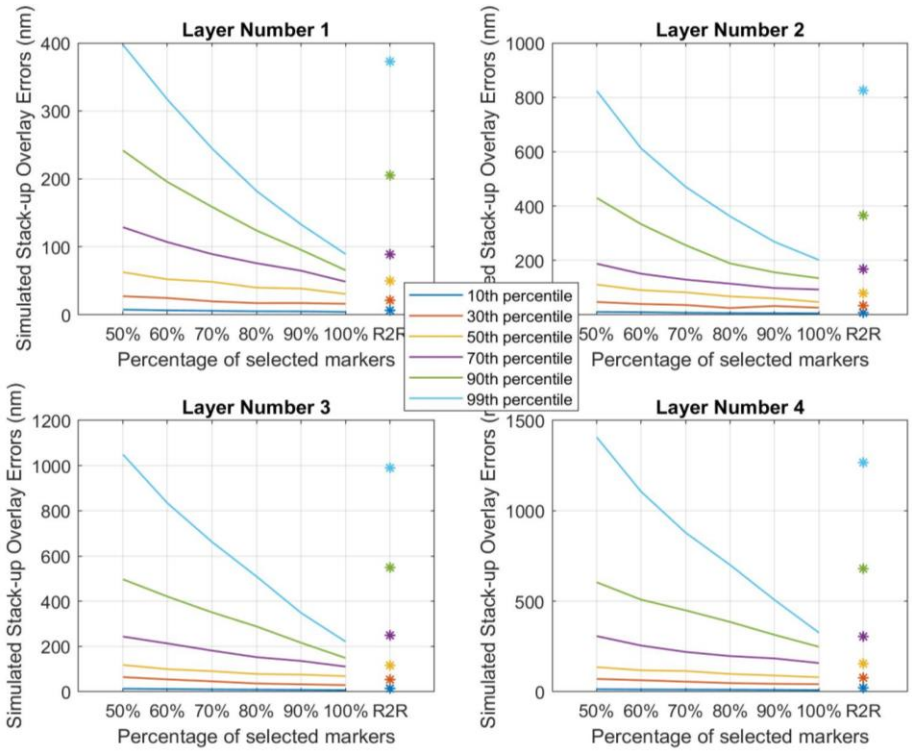

Figure 24. Plots of percentiles describing distributions of layer-specific simulated stack-up overlay error magnitudes obtained from the 200 simulations of the wafer 100 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

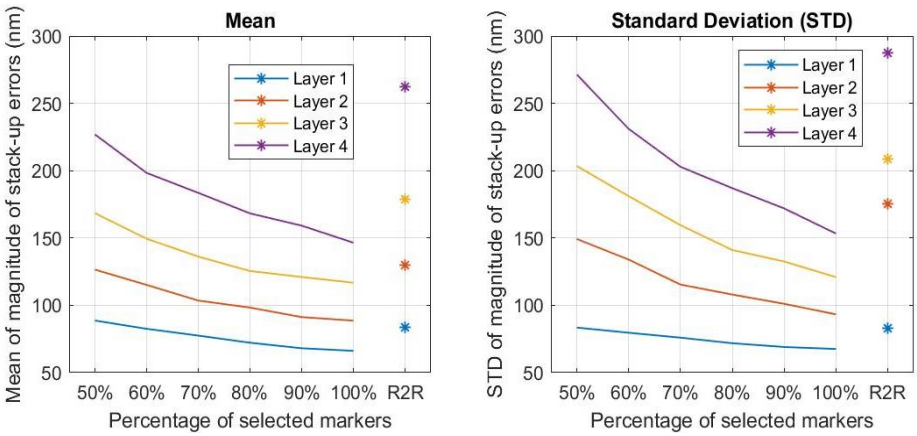

Figure 25. Plots of layer-specific mean and standard deviation of simulated stack-up overlay error magnitude obtained from the 200 simulations of wafer 100 with uniformly distributed uncertainties. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

Finally, let us summarize results associated with simulations using of truncated normal distributions for uncertain model terms. Distributions of layer-specific objective functions $f_{100,k}(\boldsymbol{F}_{100}^{*})$ under increasingly stringent constraints for the measurement marker selection are illustrated by box-and- whisker plots shown in Figure 26, while the corresponding means and standard deviations are shown in Figure 27. Behavior of overlay and stack-up over errors is analyzed in Figure 28 to Figure 31, where Figure 28 and Figure 29 show percentiles, means and standard deviations of layer-specific overlay error magnitudes, while Figure 30 and Figure 31 show percentiles, means and standard deviations of layer-specific stack-up overlay error magnitudes. Once again, one can see that the newly proposed method which combines robust overlay control with the optimal down-selection of overlay measurement markers allows the use of a dramatically lower number of overlay markers, while still outperforming the R2R method in terms of all

metrics, in all layers of the wafer.

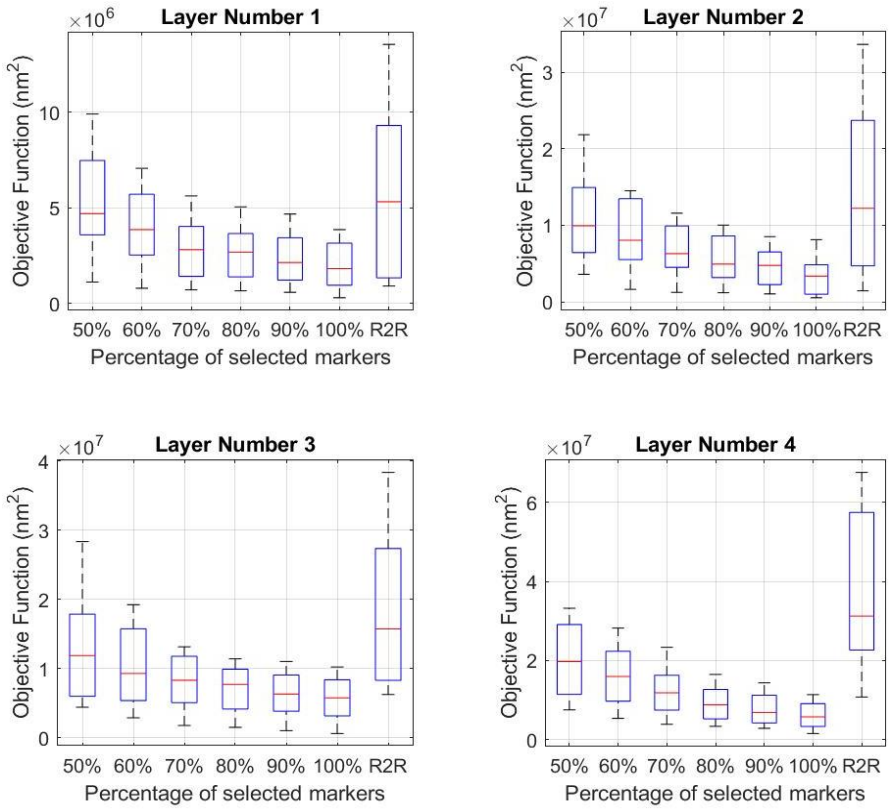

Figure 26. Box-and-whisker plots describing simulated layer-specific objective function $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 100, where uncertainties follow truncated normal distributions. Results are presented for the newly proposed measurement selection method with various percentages of selected markers and the traditional R2R method using all the markers.

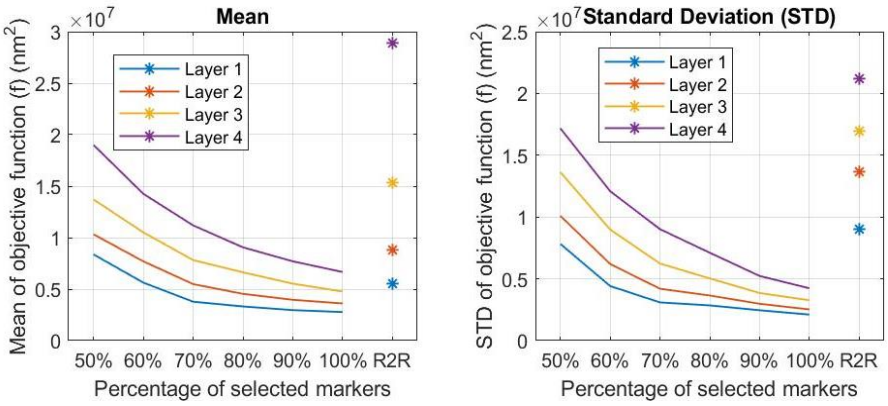

Figure 27. Plots of layer-specific mean and standard deviation of simulated objective functions $f_{t,k}(F_t^*)$ obtained from the 200 simulations of the wafer 100 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

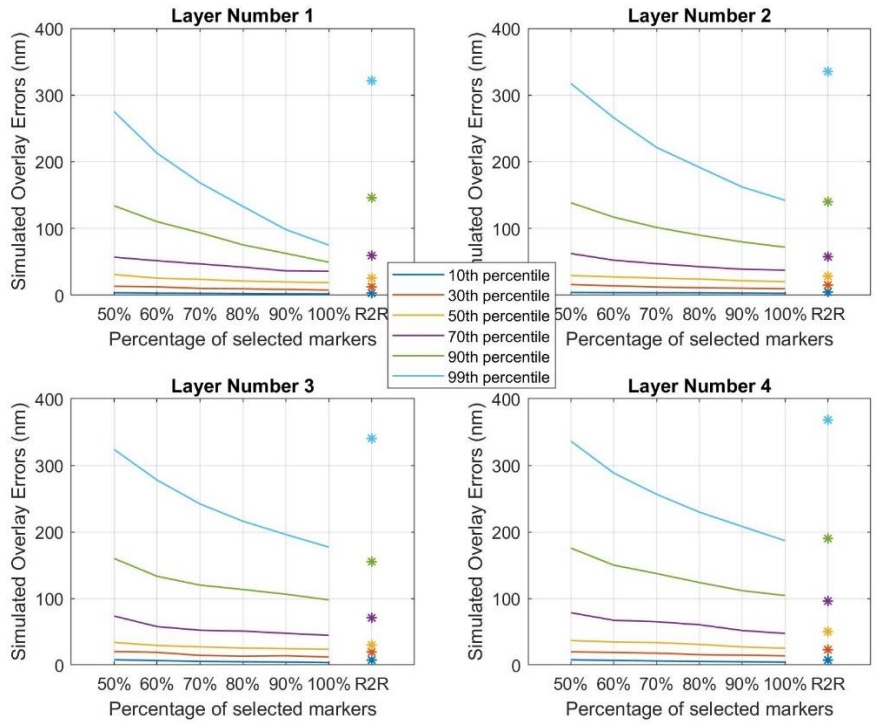

Figure 28. Plots of percentiles describing distributions of layer-specific simulated overlay error magnitudes obtained from 200 simulations of the wafer 100, where uncertainties

follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

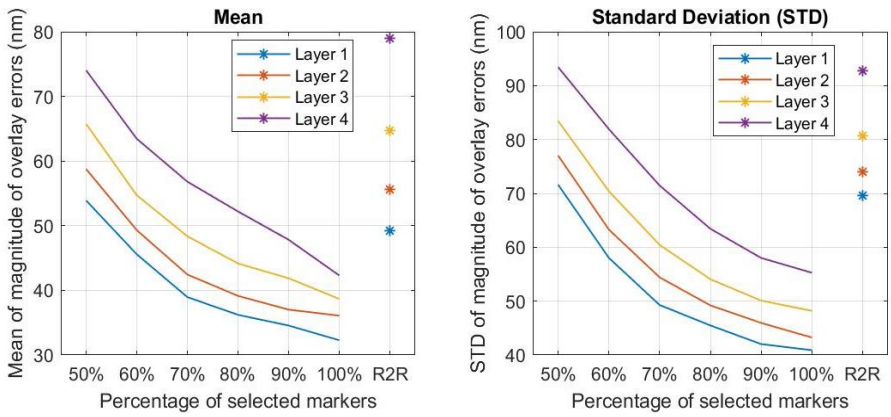

Figure 29. Plots of layer-specific mean and standard deviation of simulated overlay error magnitude obtained from the 200 simulations of the wafer 100 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and the traditional R2R method using all the markers.

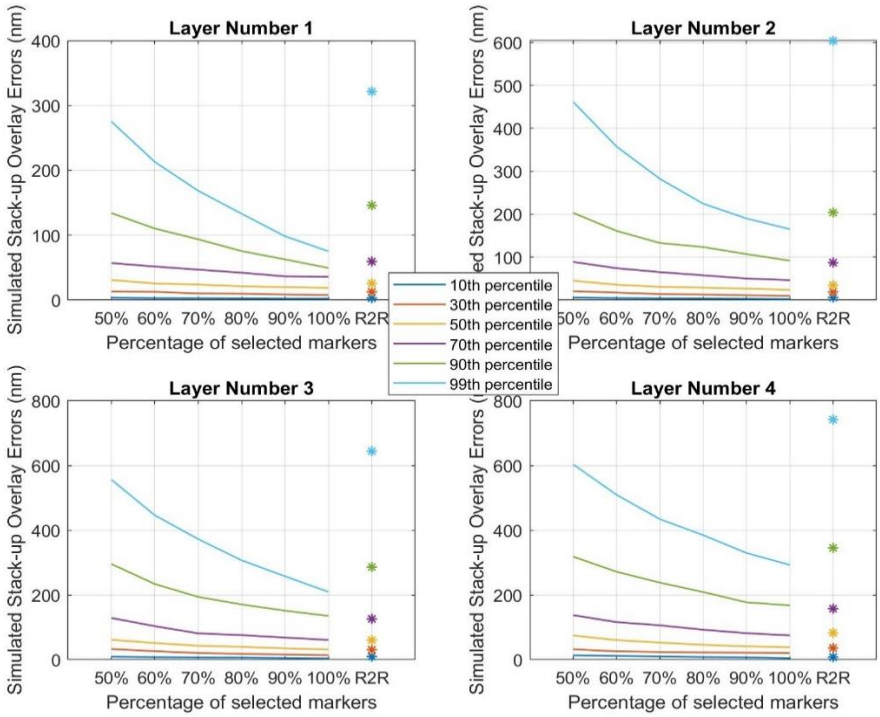

Figure 30. Plots of percentiles describing distributions of layer-specific simulated stack-up overlay error magnitudes obtained from the 200 simulations of the wafer 100 where uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and are compared against the traditional R2R method which uses all the available markers.

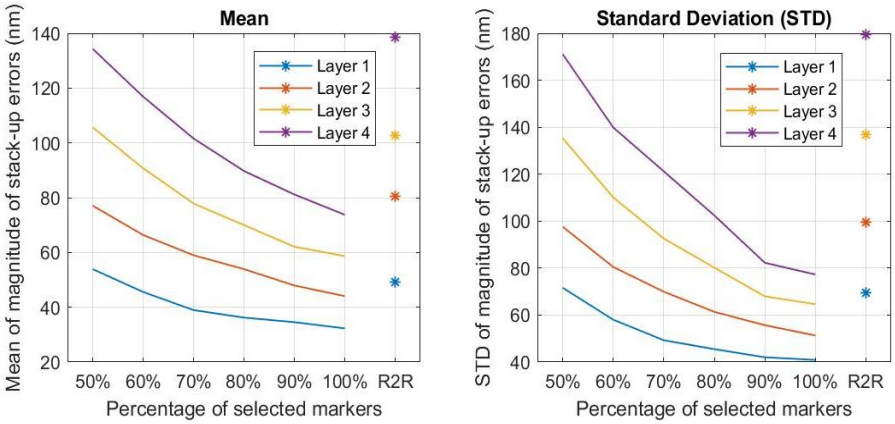

Figure 31. Plots of layer-specific means and standard deviations of simulated stack-up overlay error magnitude obtained from the 200 simulations of wafer 100 where

uncertainties follow truncated normal distributions. Results are presented for the newly proposed robust selection method with various percentages of selected markers and are compared against the traditional R2R method which uses all the available markers.

4.3.3. **Distribution of Removed Markers**

In order to analyze the spatial distribution of removed and retained overlay measurement markers across the wafer, we divided the wafer into three regions of equal area - central, middle, and peripheral ring. Inside each of those areas, we noted percentages of markers selected by our algorithm under different constraints on the measurement selection versus the total number of available markers in that region. Figure 32 shows the averages of those percentages over the 20 simulated wafers and from it, one can observe that the measurement selection algorithm introduced in this chapter consistently tends to remove measurement markers more aggressively from the inner part of the wafer, and less aggressively as one moves towards the periphery. This phenomenon was observed for each individual wafer in our simulations and is in accordance with the well-established engineering knowledge that larger levels of quality control uncertainties uncertainty and majority yield losses in semiconductor manufacturing happen to be associated with peripheral areas of the wafer [128][146].

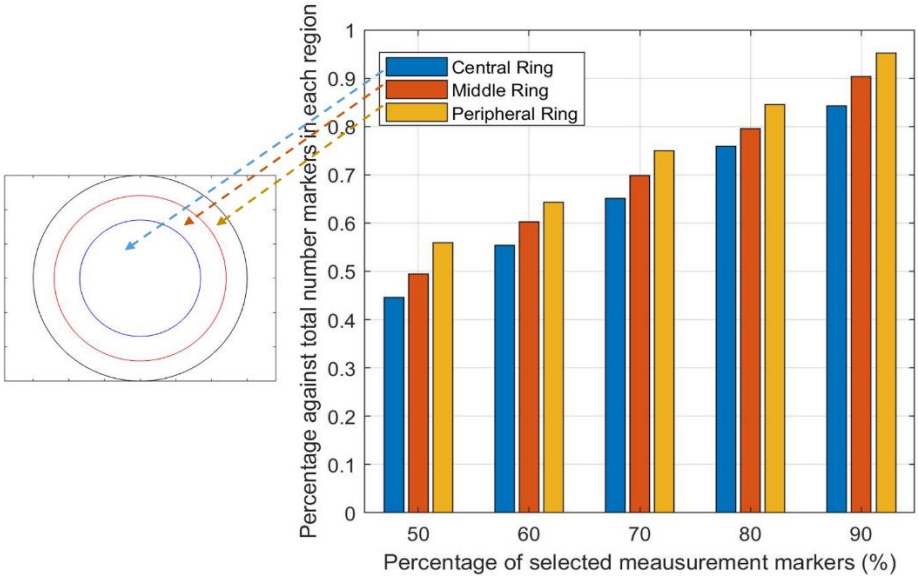

Figure 32. Average percentages of selected overlay measurement markers in each of the three regions of the wafer against the total number of markers in each region over the 20 simulated wafers.

4.4. **Conclusions and Future Work**

This chapter proposes a novel method for optimal down-selection of overlay measurement markers from an exhaustive set of available measurement markers in photolithography processes. Given a history of overlay measurements from previously manufactured wafers and a constraint in the form of the desired number of markers from which overlay measurements are to be taken on the next wafer, the method adaptively selects a subset of available overlay measurement markers from one wafer to another in a way that the selected markers facilitate best possible performance of the robust control algorithm for control of lithography overlay errors. The marker selection problem is

formulated as a bi-level robust optimization problem with the objective of minimizing the worst-case outcomes of the weighted sum of squared overlay and stack-up errors, where the worst-case performance is evaluated over uncertainties in the modelling noise and process bias terms. The underlying optimization framework is proposed in such a way that it can be efficiently solved using a GA-based nested evolutionary algorithm and commercial solvers, such as CPLEX.

The concepts and methods introduced in this chapter were evaluated using a series of simulations based on the data and overlay models from a photolithography process utilized in a major semiconductor fab processing 300mm diameter wafers. Results clearly indicate that, compared to the traditional run-to-run overlay control, the newly proposed coupling of the robust approach to controlling of overlay errors with strategic down-selection of overlay measurement markers enables improvements in the performance of the overlay control process, while using fewer measurement markers. This is of tremendous significance because those results show the opportunity to improve overlay quality, while reducing process cycle times, both of which are highly beneficial for the overall manufacturing system performance. Further evidence of the efficacy of the research introduced in this chapter can be found in the fact that the newly proposed measurement selection method consistently removed smaller portions of markers from peripheral areas of the wafer, compared to the marker removals in the central areas of the wafer. This is in accordance with the engineering expertise and intuition in the industry, where it is

commonly known that most uncertainties and yield losses occur in peripheral areas of the wafer, which implies the need to measure more intensely in those areas.

There are several directions for future research that could build on the results presented in this chapter. One obvious opportunity for future work is industrial-scale implementation of the concepts and methods introduced this manuscript. In addition, the work presented in this chapter revolves around optimizing a purely quality-loss based objective function. Since any measurement down-selection procedure, such as the one proposed in this chapter, directly affects cycle-times of the resulting process, future research should explore possibilities to explicitly include cycle-time related metrics into the optimization process and thus pursue optimality from a more system-level operational aspect, rather than just quality control aspect, as was done in this chapter.

Chapter 5. Dynamic Decision-Making on Number and Selection of Measurement Markers for Stochastic Control of Overlay Errors in Photolithography

5.1. Introduction

The measurement selection method suggested in Chapter 4 of this dissertation pursues optimality from a purely quality control aspect. However, any measurement down-selection procedure directly affects cycle-times of the resulting process and the knowledge of yield rate behaviors. A larger number of measurement markers enables a better observation of overlay errors, a better estimation of actuator uncertainties and thus, a better estimation of unobservable overlay errors, those gains come with simultaneous increases in the metrology times, which negatively impacts throughput. This requires the number of measurement markers to be determined from a system-level perspective that is able to optimize the efficiency of the manufacturing process.

This chapter aims to propose a novel optimization framework to make dynamic decisions about the number and selection of measurement markers for each layer of each wafer. It considers the determination of the optimal number of measurement makers that maximizes the profit of each unit time considering revenue earned from perfect layer patterns, cost of misidentified bad layers, as well as production and measurement cost. For any given number of markers, an optimization method for the optimal selection of

measurement markers is proposed to maximize ones' ability to estimate actuator uncertainties for the stochastic control of overlay errors. The remainder of this chapter is organized as follows. In Section 5.2, we introduce a novel optimization framework with problem formulations for the determination of the optimal number of measurement markers, the optimal layout of markers, and the optimal control commands for the stochastic control of overlay errors. In Section 5.3, we apply the newly proposed method to the overlay error model and data from a 4-layer industrial photolithography process used in a major 300 mm fab. Results of baseline settings and sensitivity analyses are presented to illustrate the newly proposed method. Section 5.4 summarizes the conclusions of this chapter and provides directions for future work.

5.2. **Methodology**

5.2.1. **Stochastic Control of Overlay Errors**

The foundations of control of photolithography overlay errors are Zernike polynomial based models, which relate overlay errors to controllable parameters on the photolithography tool. Let $\boldsymbol{o}_{t,k}^{x}$ and $\boldsymbol{o}_{t,k}^{y}$ denote vectors formed by overlay errors in all available measurement markers on layer pattern *k* of wafer *t*, with $\boldsymbol{o}_{t,k}^{x}$ denoting errors in

the x direction and $\boldsymbol{o}_{t,k}^{y}$ denoting overlay errors in the y direction on the wafer[21]. The model of overlay errors can be modeled as

$$\begin{cases} \boldsymbol{o}_{t,k}^{x} = \boldsymbol{D}^{x}(\boldsymbol{u}_{t,k}^{x} + \boldsymbol{c}_{t,k}^{x}) + \boldsymbol{r}_{t,k}^{x} \\ \boldsymbol{o}_{t,k}^{y} = \boldsymbol{D}^{y}(\boldsymbol{u}_{t,k}^{y} + \boldsymbol{c}_{t,k}^{y}) + \boldsymbol{r}_{t,k}^{y} \end{cases} \tag{5.1}$$

where:

1) $\boldsymbol{u}_{t,k}^{x}$ and $\boldsymbol{u}_{t,k}^{y}$ are vectors of control commands given to the tool consisting of controllable tool parameters affecting overlay errors in the x and y directions on the wafer.

2) $\boldsymbol{c}_{t,k}^{x}$ and $\boldsymbol{c}_{t,k}^{y}$ are commonly referred to as vectors of process bias terms. They model the stochastic actuator uncertainties that control of photolithography tools is inherently subject to.

3) $\boldsymbol{D}^{cx}$ and $\boldsymbol{D}^{cy}$ are regression matrices, fully defined by locations of the overlay measurement markers on the wafer

4) $\boldsymbol{r}_{t,k}^{x}$ and $\boldsymbol{r}_{t,k}^{y}$ are residual vector terms that account for unmodeled effects and process noise.

[21]Directions of axes on the wafer are determined based on the notch pre-fabricated on the periphery of each wafer.

Due to exceptionally small scales in which controllable parameters of a photolithography tool reside[22], unmodeled process dynamics and external noise sources are inevitably significant and cause process bias terms to always be present and continuously change from one wafer to another. A common practice in the industry is to utilize overlay measurements from historical records of previously manufactured wafers to make run-to-run(R2R) predictions $\boldsymbol{\mu}_{t,k}^{cx}$ and $\boldsymbol{\mu}_{t,k}^{cy}$ of bias vector term in layer *k* of wafer *t* prior to the actual lithography exposure, with the prediction variance being $\boldsymbol{\sigma}_{t,k}^{cx\,2}$ and $\boldsymbol{\sigma}_{t,k}^{cy\,2}$. There are various methods for dynamic modeling and prediction of bias vector terms, such as Kalman filter-based prediction, various forms of the Exponentially Weighted Moving Average (EWMA) based modeling and prediction methods, or more recently dynamic neural networks [129] and Gaussian Process Regression [130] based approaches. Following [126], we assume that the process bias of each controllable component follows Gaussian distribution whose mean $\boldsymbol{\mu}_{t,k}^{cx}, \boldsymbol{\mu}_{t,k}^{cy}$ and variance $\boldsymbol{\sigma}_{t,k}^{cx\,2}, \boldsymbol{\sigma}_{t,k}^{cy\,2}$ are obtained from prediction results, and the residuals at each marker i follow Gaussian distribution with 0 mean and known variance $\sigma_{t,k,i}^{rx\,2}$, which can be obtained from historical records.

For layer k of wafer t, to obtain the optimal control commands, our objective is to minimize the overlay errors as well as the stack-up overlay errors. Given predicted process

[22]Nanometer, or even sub-nanometer scales.

bias $\boldsymbol{\mu}_{t,k}^{cx}, \boldsymbol{\mu}_{t,k}^{cy}$, prediction variance $\boldsymbol{\sigma}_{t,k}^{cx\,2}, \boldsymbol{\sigma}_{t,k}^{cy\,2}$, variance $\boldsymbol{\sigma}_{t,k}^{rx\,2}, \boldsymbol{\sigma}_{t,k}^{ry\,2}$ of residuals at all the markers, as well as mean $\boldsymbol{\mu}_{t,k-1}^{sx}, \boldsymbol{\mu}_{t,k-1}^{sy}$ and variance $\boldsymbol{\sigma}_{t,k-1}^{sx\,2}, \boldsymbol{\sigma}_{t,k-1}^{sy\,2}$ of stack up overlay errors in previous layer $k-1$, the overlay control problem is thus characterized as the following stochastic optimization problem

$$\left(\boldsymbol{u}_{t,k}^{x*}, \boldsymbol{u}_{t,k}^{y*}\right) = \underset{\substack{\boldsymbol{u}_{t,k}^{x} \in \mathbb{R}^{N_x}, \\ \boldsymbol{u}_{t,k}^{y} \in \mathbb{R}^{N_y}}}{\operatorname{argmin}} \; \underset{\substack{\boldsymbol{c}_{t,k}^{x}, \boldsymbol{c}_{t,k}^{y} \\ \boldsymbol{r}_{t,k}^{x}, \boldsymbol{r}_{t,k}^{y} \\ \boldsymbol{s}_{t,k-1}^{x}, \boldsymbol{s}_{t,k-1}^{y}}}{\mathbb{E}} \begin{bmatrix} \lambda^{x}\left\|\boldsymbol{o}_{t,k}^{x}\right\|^{2} + \lambda^{y}\left\|\boldsymbol{o}_{t,k}^{y}\right\|^{2} \\ +\alpha^{x}\left\|\boldsymbol{s}_{t,k-1}^{x} + \boldsymbol{o}_{t,k}^{x}\right\|^{2} + \alpha^{y}\left\|\boldsymbol{s}_{t,k-1}^{y} + \boldsymbol{o}_{t,k}^{y}\right\|^{2} \end{bmatrix} \tag{5.2}$$

$$\text{subject to: (5.1)}$$

Following [126], this problem can be reformulated as a deterministic optimization problem and can be easily solved with an analytical solution.

5.2.2. **Problem Formulation for the Optimal Selection of Measurement Markers**

Given the optimal control commands $\boldsymbol{u}_{t,k}^{x*}, \boldsymbol{u}_{t,k}^{y*}$ obtained using procedure (5.1) and (5.2), overlay errors can be observed at measurement markers. Those measurements are used to estimate the actual process bias terms, which are crucial in the quality control of the pattern layers, as they need to be used for 1) the estimation of overlay errors at other unobserved markers; 2) the R2R prediction of process bias terms for the next wafer. Let P

denote the number of all the available markers and $P_{t,k}^{obj}$ denote the percentage of selected markers. For any given number of markers $P_{t,k}^{obj} \cdot P$, our objective is to find the best set of markers $\boldsymbol{F}_{t,k}^{*}(P_{t,k}^{obj})$ that can provide us with estimations of process bias terms that is as close as possible with the estimations obtained from overlay errors observed at all the candidate markers $\boldsymbol{F}^{Tot}$.

Given predicted process bias $\boldsymbol{\mu}_{t,k}^{cx}, \boldsymbol{\mu}_{t,k}^{cy}$ and prediction variance $\boldsymbol{\sigma}_{t,k}^{cx\,2}, \boldsymbol{\sigma}_{t,k}^{cy\,2}$, variance $\boldsymbol{\sigma}_{t,k}^{rx\,2}, \boldsymbol{\sigma}_{t,k}^{ry\,2}$ of residuals, optimal control commands $\boldsymbol{u}_{t,k}^{x*}, \boldsymbol{u}_{t,k}^{y*}$, we now describe the approach for estimating the process bias terms with overlay errors $\boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}_{t,k})$, $\boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}_{t,k})$ measured at any selected markers $\boldsymbol{F}_{t,k}$.

Based on Zernike polynomial-based overlay models, the least-squares based estimates of the process bias $\hat{\mathbf{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k})$ and $\hat{\mathbf{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k})$ can be obtained as

$$\begin{cases} \hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k}) = \left(\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{xT}\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{x}\right)^{-1}\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{xT}\boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}_{t,k}) - \boldsymbol{u}_{t,k}^{*x} \\ \hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k}) = \left(\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{yT}\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{y}\right)^{-1}\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{yT}\boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}_{t,k}) - \boldsymbol{u}_{t,k}^{*y} \end{cases} \tag{5.3}$$

where $\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{x}$ and $\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{y}$ are regression matrices of the overlay error models (5.1) corresponding to the use of selected markers $\boldsymbol{F}_{t,k}$. Those estimates of process bias are actually stochastic because the measurement of overlay errors $\boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}_{t,k})$, $\boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}_{t,k})$ depend on the realization of stochastic process bias terms and model residuals. Based on

(5.3), the expected value of the estimated bias can be easily derived as follows, which does not depend on the selection of measurement markers.

$$\begin{cases}\mathbb{E}\left[\hat{\boldsymbol{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)\right]=\boldsymbol{\mu}_{t,k}^{cx}\\ \mathbb{E}\left[\hat{\boldsymbol{c}}_{t,k}^{y}\left(\boldsymbol{F}_{t,k}\right)\right]=\boldsymbol{\mu}_{t,k}^{cy}\end{cases}$$

We assume the vector of estimated process bias follows multivariate Gaussian distributions, where we use the expected value $\mathbb{E}\left[\hat{\boldsymbol{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)\right], \mathbb{E}\left[\hat{\boldsymbol{c}}_{t,k}^{y}\left(\boldsymbol{F}_{t,k}\right)\right]$ of the vector of estimated bias and the expected value $\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)}\right], \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}\left(\boldsymbol{F}_{t,k}\right)}\right]$ of the variance-covariance matrix of the estimated bias to describe the mean and variance-covariance of the distribution.

$$\begin{cases}\hat{\boldsymbol{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)\sim\boldsymbol{\mathcal{N}}\left(\boldsymbol{\mu}_{t,k}^{cx},\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)}\right]\right)\\ \hat{\boldsymbol{c}}_{t,k}^{y}\left(\boldsymbol{F}_{t,k}\right)\sim\boldsymbol{\mathcal{N}}\left(\boldsymbol{\mu}_{t,k}^{cy},\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}\left(\boldsymbol{F}_{t,k}\right)}\right]\right)\end{cases} \tag{5.4}$$

One needs to notice that the expected value $\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)}\right], \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}\left(\boldsymbol{F}_{t,k}\right)}\right]$ of the variance-covariance matrix of the estimated bias is affected by the selection of measurement markers. Focusing on the x-axis, the variance-covariance matrix of the least square estimates is given by

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)}=\sigma_{\boldsymbol{r}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)}^{\boldsymbol{2}}\left(\boldsymbol{D}_{\boldsymbol{F}_{t,k}^{x}}^{x\ T}\boldsymbol{D}_{\boldsymbol{F}_{t,k}^{x}}^{x}\right)^{-1} \tag{5.5}$$

where $\sigma_{\boldsymbol{r}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}\right)}^{\boldsymbol{2}}$ is variance of residuals of the regression model fitted with selected markers $\boldsymbol{F}_{t,k}$. It can be estimated with the sum of squared residuals at selected markers as

$$\sigma^2_{\boldsymbol{r}^x_{t,k}(\boldsymbol{F}_{t,k})} = \frac{\boldsymbol{r}^x_{t,k}(\boldsymbol{F}_{t,k})^T \boldsymbol{r}^x_{t,k}(\boldsymbol{F}_{t,k})}{P^{obj}_{t,k} \cdot P - N^{cx} - 1} \tag{5.6}$$

where the vector of residuals at selected markers is calculated as

$$\boldsymbol{r}^x_{t,k}(\boldsymbol{F}_{t,k}) = \boldsymbol{o}^x_{t,k}(\boldsymbol{F}_{t,k}) - \boldsymbol{D}^x_{\boldsymbol{F}_{t,k}} \hat{\boldsymbol{c}}^x_{t,k}(\boldsymbol{F}_{t,k}) \tag{5.7}$$

Based on (5.4) - (5.7), the variance-covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}^x_{t,k}(\boldsymbol{F}_{t,k})}$, $\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}^y_{t,k}(\boldsymbol{F}_{t,k})}$ of estimated process bias can be expressed as

$$\begin{cases} \boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}^x_{t,k}(\boldsymbol{F}_{t,k})} = \dfrac{\boldsymbol{o}^x_{t,k}(\boldsymbol{F}_{t,k})^T \left[1 - \boldsymbol{D}^x_{\boldsymbol{F}_{t,k}} \left(\boldsymbol{D}^{xT}_{\boldsymbol{F}_{t,k}} \boldsymbol{D}^x_{\boldsymbol{F}_{t,k}}\right)^{-1} \boldsymbol{D}^{xT}_{\boldsymbol{F}_{t,k}}\right] \boldsymbol{o}^x_{t,k}(\boldsymbol{F}_{t,k}) \left(\boldsymbol{D}^{xT}_{\boldsymbol{F}_{t,k}} \boldsymbol{D}^x_{\boldsymbol{F}_{t,k}}\right)^{-1}}{P^{obj}_{t,k} \cdot P - N^{cx} - 1} \\ \boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}^y_{t,k}(\boldsymbol{F}_{t,k})} = \dfrac{\boldsymbol{o}^y_{t,k}(\boldsymbol{F}_{t,k})^T \left[1 - \boldsymbol{D}^y_{\boldsymbol{F}_{t,k}} \left(\boldsymbol{D}^{yT}_{\boldsymbol{F}_{t,k}} \boldsymbol{D}^y_{\boldsymbol{F}_{t,k}}\right)^{-1} \boldsymbol{D}^{yT}_{\boldsymbol{F}_{t,k}}\right] \boldsymbol{o}^y_{t,k}(\boldsymbol{F}_{t,k}) \left(\boldsymbol{D}^{yT}_{\boldsymbol{F}_{t,k}} \boldsymbol{D}^y_{\boldsymbol{F}_{t,k}}\right)^{-1}}{P^{obj}_{t,k} \cdot P - N^{cy} - 1} \end{cases} \tag{5.8}$$

Then, their expected values can be derived as

$$\begin{cases} \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}^x_{t,k}(\boldsymbol{F}_{t,k})}\right] \\ = \dfrac{\mathrm{diag}\left(1 - \boldsymbol{D}^x_{\boldsymbol{F}_{t,k}} \left(\boldsymbol{D}^{xT}_{\boldsymbol{F}_{t,k}} \boldsymbol{D}^x_{\boldsymbol{F}_{t,k}}\right)^{-1} \boldsymbol{D}^{xT}_{\boldsymbol{F}_{t,k}}\right)^T \left(\mathbb{E}[\boldsymbol{o}^x_{t,k}(\boldsymbol{F}_{t,k})]^2 + \mathrm{Var}[\boldsymbol{o}^x_{t,k}(\boldsymbol{F}_{t,k})]\right) \left(\boldsymbol{D}^x_{\boldsymbol{F}_{t,k}}{}^T \boldsymbol{D}^x_{\boldsymbol{F}_{t,k}}\right)^{-1}}{P^{obj}_{t,k} \cdot P - N^{cx} - 1} \\ \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}^y_{t,k}(\boldsymbol{F}_{t,k})}\right] \\ = \dfrac{\mathrm{diag}\left(1 - \boldsymbol{D}^y_{\boldsymbol{F}_{t,k}} \left(\boldsymbol{D}^{yT}_{\boldsymbol{F}_{t,k}} \boldsymbol{D}^y_{\boldsymbol{F}_{t,k}}\right)^{-1} \boldsymbol{D}^{yT}_{\boldsymbol{F}_{t,k}}\right)^T \left(\mathbb{E}\left[\boldsymbol{o}^y_{t,k}(\boldsymbol{F}_{t,k})\right]^2 + \mathrm{Var}\left[\boldsymbol{o}^y_{t,k}(\boldsymbol{F}_{t,k})\right]\right) \left(\boldsymbol{D}^y_{\boldsymbol{F}_{t,k}}{}^T \boldsymbol{D}^y_{\boldsymbol{F}_{t,k}}\right)^{-1}}{P^{obj}_{t,k} \cdot P - N^{cy} - 1} \end{cases} \tag{5.9}$$

where the expected value and variance of overlay errors can be calculated as

$$\begin{cases}\mathbb{E}\left[\boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}_{t,k})\right]=\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{x}\left(\boldsymbol{u}_{t,k}^{*x}+\boldsymbol{\mu}_{t,k}^{cx}\right)\\ \mathbb{E}\left[\boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}_{t,k})\right]=\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{y}\left(\boldsymbol{u}_{t,k}^{*y}+\boldsymbol{\mu}_{t,k}^{cy}\right)\end{cases},\begin{cases}\mathrm{Var}\left[\boldsymbol{o}_{t,k}^{x}(\boldsymbol{F}_{t,k})\right]=\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{x\,\circ 2}\boldsymbol{\sigma}_{t,k}^{cx\,2}+\boldsymbol{\sigma}_{t,k}^{rx\,2}\\ \mathrm{Var}\left[\boldsymbol{o}_{t,k}^{y}(\boldsymbol{F}_{t,k})\right]=\boldsymbol{D}_{\boldsymbol{F}_{t,k}}^{y\,\circ 2}\boldsymbol{\sigma}_{t,k}^{cy\,2}+\boldsymbol{\sigma}_{t,k}^{ry\,2}\end{cases}\quad(5.10)$$

Given any number of selected markers $P_{t,k}^{obj}\cdot P$, our objective is to find the best set of markers $\boldsymbol{F}_{t,k}^{*}(P_{t,k}^{obj})$ that minimize the difference between the distribution (5.4) of process bias $\hat{\mathbf{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k}),\hat{\mathbf{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k})$ estimated with observations from selected markers and the distribution of $\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot}),\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot})$ estimated with observations from all the candidate markers. To measure the similarity between two distributions, we propose to use the Wasserstein distance (a.k.a. Earth Mover's Distance) [147]as the objective function. Intuitively speaking, this distance metric is proportional to the minimum amount of work required to change one distribution into the other, where one unit of work means the amount of work necessary to move one unit of weight by one unit of distance. Using the Euclidian norm as the distance function, the $\mathcal{W}_2$ Wasserstein coupling distance between two probability measures ν_1 and ν_2 on $\mathbb{R}^n$ is

$$\mathcal{W}_2(\nu_1;\nu_2)=(\inf\mathbb{E}[\|\boldsymbol{Z_1}-\boldsymbol{Z_2}\|_2^2])^{1/2}$$

where the infimum runs over all random vectors $(\boldsymbol{Z_1},\boldsymbol{Z_2})$ of $\mathbb{R}^n\times\mathbb{R}^n$ with $\boldsymbol{Z_1}\sim\nu_1$ and $\boldsymbol{Z_2}\sim\nu_2$. Based on [148], for two multivariate Gaussian distributions, the 2-Wasserstein distance can be calculated as follows

$$\begin{cases} \mathcal{W}_2\left(\mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cx}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k})}\right]\right); \mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cx}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot})}\right]\right)\right)^2 \\ = tr\left(\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k})}\right] + \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot})}\right] - 2\left(\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k})}\right]\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot})}\right]\right)^{1/2}\right) \\ \mathcal{W}_2\left(\mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cy}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k})}\right]\right); \mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cy}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot})}\right]\right)\right)^2 \\ = tr\left(\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k})}\right] + \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot})}\right] - 2\left(\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k})}\right]\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot})}\right]\right)^{1/2}\right) \end{cases} \tag{5.11}$$

Therefore, given predicted process bias $\boldsymbol{\mu}_{t,k}^{cx}, \boldsymbol{\mu}_{t,k}^{cy}$, prediction variance $\boldsymbol{\sigma}_{t,k}^{cx\,2}, \boldsymbol{\sigma}_{t,k}^{cy\,2}$, variance $\boldsymbol{\sigma}_{t,k}^{rx\,2}, \boldsymbol{\sigma}_{t,k}^{ry\,2}$ of residuals at all the markers, optimal control commands $\boldsymbol{u}_{t,k}^{x*}, \boldsymbol{u}_{t,k}^{y*}$ and distributions of estimated process bias

$$\begin{cases} \hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cx}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot})}\right]\right) \\ \hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot}) \sim \mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cy}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot})}\right]\right) \end{cases}$$

when all the markers $\boldsymbol{F}^{Tot}$ are selected, for any number of selected markers $P_{t,k}^{obj} \cdot P$, the best set of markers $\boldsymbol{F}_{t,k}^{*}(P_{t,k}^{obj})$ is found by solving the following optimization problem

$$\boldsymbol{F}_{t,k}^{*}\left(P_{t,k}^{obj}\right) = \underset{\boldsymbol{F}_{t,k} \in \mathbb{Z}_2^P}{\arg min} \begin{matrix} \mathcal{W}_2\left(\mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cx}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k})}\right]\right); \mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cx}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}^{Tot})}\right]\right)\right)^2 \\ +\mathcal{W}_2\left(\mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cy}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k})}\right]\right); \mathcal{N}\left(\boldsymbol{\mu}_{t,k}^{cy}, \mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}^{Tot})}\right]\right)\right)^2 \end{matrix} \tag{5.12}$$

$$\text{subject to:} \quad \mathbf{1}^T \boldsymbol{F}_{t,k} = P_{t,k}^{obj} \cdot P \tag{5.13}$$

where the objective function (5.12) is calculated following (5.9), (510) and (5.11).

This optimization problem can be solved with the Genetic Algorithm (GA) based approach explained in Section 4.2.3, where operators including selection, crossover, repair, mutation are conducted. In addition, the elitism is invoked to increase the convergence rate and stability of the GA.

5.2.3. **Problem Formulation for the Optimal Number of Measurement Markers**

The quality of each layer pattern is evaluated by checking whether overlay errors measured at selected markers are within their specification limits. Given lower Specification Limits (LSL) $\boldsymbol{o}_{t,k}^{x,LSL}, \boldsymbol{o}_{t,k}^{y,LSL}, \boldsymbol{s}_{t,k}^{x,LSL}, \boldsymbol{s}_{t,k}^{y,LSL}$ and Upper Specification Limits (USL) $\boldsymbol{o}_{t,k}^{x,USL}, \boldsymbol{o}_{t,k}^{y,USL}, \boldsymbol{s}_{t,k}^{x,USL}, \boldsymbol{s}_{t,k}^{y,USL}$, the probability that overlay and stack-up overlay errors, observed at each selected marker $i \in \boldsymbol{F}_{t,k}^{*}(P_{t,k}^{obj})$, are within their specification limits can be calculated as

$$\mathbb{P}(\mathcal{A}_i) = R_i^{o^x} \times R_i^{o^y} \times R_i^{s^x} \times R_i^{s^x}, \forall i \in \boldsymbol{F}_{t,k}^{*}\left(P_{t,k}^{obj}\right) \tag{5.14}$$

where

$$\begin{cases} R_i^{o^x} = \mathbb{P}\left(o_{t,k,i}^{x,LSL} \le o_{t,k,i}^{x} \le o_{t,k,i}^{x,USL}\right) \\ R_i^{o^y} = \mathbb{P}\left(o_{t,k,i}^{y,LSL} \le o_{t,k,i}^{y} \le o_{t,k,i}^{y,USL}\right) \end{cases}, \begin{cases} R_i^{s^x} = \mathbb{P}\left(s_{t,k,i}^{x,LSL} \le s_{t,k,i}^{x} \le s_{t,k,i}^{x,USL}\right) \\ R_i^{s^y} = \mathbb{P}\left(s_{t,k,i}^{y,LSL} \le s_{t,k,i}^{y} \le s_{t,k,i}^{y,USL}\right) \end{cases}$$

One can easily show that overlay and stack-up overlay errors follow Gaussian distributions

$$\begin{cases} o_{t,k,i}^{x} \sim \mathcal{N}\left(\mu_{t,k,i}^{ox}, {\sigma_{t,k,i}^{ox}}^{2}\right) \\ o_{t,k,i}^{y} \sim \mathcal{N}\left(\mu_{t,k,i}^{oy}, {\sigma_{t,k,i}^{oy}}^{2}\right) \end{cases}, \begin{cases} s_{t,k,i}^{x} \sim \mathcal{N}\left(\mu_{t,k,i}^{sx}, {\sigma_{t,k,i}^{sx}}^{2}\right) \\ s_{t,k,i}^{y} \sim \mathcal{N}\left(\mu_{t,k,i}^{sy}, {\sigma_{t,k,i}^{sy}}^{2}\right) \end{cases}$$

where means and variances are calculated as

$$\begin{cases} \mu_{t,k,i}^{ox} = \boldsymbol{D}_{i}^{x}\left(\boldsymbol{u}_{t,k}^{*x} + \boldsymbol{\mu}_{t,k}^{cx}\right) \\ \mu_{t,k,i}^{oy} = \boldsymbol{D}_{i}^{y}\left(\boldsymbol{u}_{t,k}^{*y} + \boldsymbol{\mu}_{t,k}^{cy}\right) \\ {\sigma_{t,k,i}^{ox}}^{2} = \boldsymbol{D}_{i}^{x\circ 2}{\boldsymbol{\sigma}_{t,k}^{cx}}^{2} + {\sigma_{t,k,i}^{rx}}^{2} \\ {\sigma_{t,k,i}^{oy}}^{2} = \boldsymbol{D}_{i}^{y\circ 2}{\boldsymbol{\sigma}_{t,k}^{cy}}^{2} + {\sigma_{t,k,i}^{ry}}^{2} \end{cases}, \begin{cases} \mu_{t,k,i}^{sx} = \mu_{t,k-1,i}^{sx} + \mu_{t,k,i}^{ox} \\ \mu_{t,k,i}^{sy} = \mu_{t,k-1,i}^{sy} + \mu_{t,k,i}^{oy} \\ {\sigma_{t,k,i}^{sx}}^{2} = {\sigma_{t,k-1,i}^{sx}}^{2} + {\sigma_{t,k,i}^{ox}}^{2} \\ {\sigma_{t,k,i}^{sy}}^{2} = {\sigma_{t,k-1,i}^{sy}}^{2} + {\sigma_{t,k,i}^{oy}}^{2} \end{cases} \tag{5.15}$$

The probability that this layer pattern is observed to be a good product can be calculated as

$$\mathbb{P}(\boldsymbol{\mathcal{A}}) = \prod_{i \in \boldsymbol{F}_{t,k}^{*}\left(P_{t,k}^{obj}\right)} \mathbb{P}(\mathcal{A}_i) \tag{5.16}$$

which can be called as the *yield rate based on selected makers*.

However, this evaluation of product quality only focuses on the measured overlay errors. There is still a certain possibility that overlay and stack-up overlay errors at markers that are not measured exceed the specification limits but are not observed. To evaluate this probability, we use the estimation $\hat{\mathbf{c}}_{t,k}^{x}\left(\boldsymbol{F}_{t,k}^{*}\right), \hat{\mathbf{c}}_{t,k}^{y}\left(\boldsymbol{F}_{t,k}^{*}\right)$ of actual realized process bias terms, obtained with overlay errors measured at selected markers $\boldsymbol{F}_{t,k}^{*}\left(P_{t,k}^{obj}\right)$, to establish the estimation $\hat{o}_{t,k,j}^{x}, \hat{o}_{t,k,j}^{y}, \hat{s}_{t,k,j}^{x}, \hat{s}_{t,k,j}^{y}$ of overlay and stack-up overlay errors at each marker $j \in \overline{\boldsymbol{F}}_{t,k}^{*}(P_{t,k}^{obj})$, where $\overline{\boldsymbol{F}}_{t,k}^{*}\left(P_{t,k}^{obj}\right) = 1 - \boldsymbol{F}_{t,k}^{*}(P_{t,k}^{obj})$ is the set of markers that are not selected. The probability that overlay and stack-up overlay errors estimated at each

unselected marker $j \in \overline{\boldsymbol{F}}^{*}_{t,k}(P^{obj}_{t,k})$, are within their specification limits can be calculated as

$$\mathbb{P}(\mathcal{B}_j) = R_j^{\hat{o}^x} \times R_j^{\hat{o}^y} \times R_j^{\hat{s}^x} \times R_j^{\hat{s}^x}, \forall j \in \overline{\boldsymbol{F}}^{*}_{t,k}\left(P^{obj}_{t,k}\right) \tag{5.17}$$

where

$$\begin{cases} R_j^{\hat{o}^x} = \mathbb{P}\left(o_{t,k,j}^{x,LSL} \le \hat{o}_{t,k,j}^{x} \le o_{t,k,j}^{x,USL}\right) \\ R_j^{\hat{o}^y} = \mathbb{P}\left(o_{t,k,j}^{y,LSL} \le \hat{o}_{t,k,j}^{y} \le o_{t,k,j}^{y,USL}\right) \end{cases}, \begin{cases} R_j^{\hat{s}^x} = \mathbb{P}\left(s_{t,k,j}^{x,LSL} \le \hat{s}_{t,k,j}^{x} \le s_{t,k,j}^{x,USL}\right) \\ R_j^{\hat{s}^y} = \mathbb{P}\left(s_{t,k,j}^{y,LSL} \le \hat{s}_{t,k,j}^{y} \le s_{t,k,j}^{y,USL}\right) \end{cases}$$

These estimated overlay and stack-up overlay errors also can be shown to follow Gaussian distributions

$$\begin{cases} \hat{o}_{t,k,j}^{x} \sim \mathcal{N}\left(\mu_{t,k,j}^{\hat{o}x}, {\sigma_{t,k,j}^{\hat{o}x}}^2\right) \\ \hat{o}_{t,k,j}^{y} \sim \mathcal{N}\left(\mu_{t,k,j}^{\hat{o}y}, {\sigma_{t,k,j}^{\hat{o}y}}^2\right) \end{cases}, \begin{cases} \hat{s}_{t,k,j}^{x} \sim \mathcal{N}\left(\mu_{t,k,j}^{\hat{s}x}, {\sigma_{t,k,j}^{\hat{s}x}}^2\right) \\ \hat{s}_{t,k,j}^{y} \sim \mathcal{N}\left(\mu_{t,k,j}^{\hat{s}y}, {\sigma_{t,k,j}^{\hat{s}y}}^2\right) \end{cases}$$

where means and variances are calculated as

$$\begin{cases} \mu_{t,k,j}^{\hat{o}x} = \boldsymbol{D}_j^{x}\left(\boldsymbol{u}_{t,k}^{*x} + \boldsymbol{\mu}_{t,k}^{cx}\right) \\ \mu_{t,k,j}^{\hat{o}y} = \boldsymbol{D}_j^{y}\left(\boldsymbol{u}_{t,k}^{*y} + \boldsymbol{\mu}_{t,k}^{cy}\right) \\ {\sigma_{t,k,j}^{\hat{o}x}}^2 = \boldsymbol{D}_j^{x}\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{x}(\boldsymbol{F}_{t,k})}\right]\boldsymbol{D}_j^{xT} + {\sigma_{t,k,j}^{rx}}^2 \\ {\sigma_{t,k,j}^{\hat{o}y}}^2 = \boldsymbol{D}_j^{y}\mathbb{E}\left[\boldsymbol{\Sigma}_{\hat{\boldsymbol{c}}_{t,k}^{y}(\boldsymbol{F}_{t,k})}\right]\boldsymbol{D}_j^{yT} + {\sigma_{t,k,j}^{rx}}^2 \end{cases}, \begin{cases} \mu_{t,k,j}^{\hat{s}x} = \mu_{t,k-1,j}^{sx} + \mu_{t,k,j}^{\hat{o}x} \\ \mu_{t,k,j}^{\hat{s}y} = \mu_{t,k-1,j}^{sy} + \mu_{t,k,j}^{\hat{o}y} \\ {\sigma_{t,k,j}^{\hat{s}x}}^2 = {\sigma_{t,k-1,j}^{sx}}^2 + {\sigma_{t,k,j}^{\hat{o}x}}^2 \\ {\sigma_{t,k,j}^{\hat{s}y}}^2 = {\sigma_{t,k-1,j}^{sy}}^2 + {\sigma_{t,k,j}^{\hat{o}y}}^2 \end{cases} \tag{5.18}$$

Then, we calculate the probability that this layer pattern is a perfect product at all the unselected markers as

$$\mathbb{P}(\boldsymbol{\mathcal{B}}) = \prod_{j \in \overline{\boldsymbol{F}}^{*}_{t,k}\left(P^{obj}_{t,k}\right)} \mathbb{P}\left(\mathcal{B}_j\right) \tag{5.19}$$

which can be called as the *yield rate based on unselected makers*.

Our objective is to find the optimal percentage of measurement markers $P^{*obj}_{t,k}$ that maximizes the profit per unit time. The objective function can be expressed as follows

$$V_{t,k}\left(P^{obj}_{t,k}\right) = \frac{f_{rev} - f_{mis} - f_{pm}}{T_{total}} \tag{5.20}$$

where T_{total} is the total time period we would like to evaluate. Other components in the objective function are:

- *Total revenue earned from perfect layer patterns*

$$f_{rev} = r_{ev} \frac{T_{total}}{T_p + T_m P^{obj}_{t,k}} \mathbb{P}(\boldsymbol{\mathcal{A}}) \tag{5.21}$$

  where r_{ev} is the unit revenue per perfect layer pattern, T_p is the production time per layer pattern, T_m is the measurement time per layer pattern when all markers are selected, and $\frac{T_{total}}{T_p + T_m P^{obj}_{t,k}}$ calculates the total number of layers finished in the total time period T_{total}.

- *Total cost of misidentified bad layers*

$$f_{mis} = c_{ubad} \frac{T_{total}}{T_p + T_m P_{t,k}^{obj}} \mathbb{P}(\boldsymbol{\mathcal{A}}) \big(1 - \mathbb{P}(\boldsymbol{\mathcal{B}})\big) \tag{5.22}$$

where c_{ubad} is the cost per misidentified imperfect layer pattern, which is measured to be perfect but is actually imperfect at unobserved markers.

- *Total production and measurement cost*

$$f_{pm} = \big[c_m P_{t,k}^{obj} + c_p\big] \frac{T_{total}}{T_p + T_m P_{t,k}^{obj}} \tag{5.23}$$

where c_m is the measurement equipment cost per layer pattern when all the markers are selected, c_p is the production cost per layer pattern. It is assumed that the unit cost and revenues are chosen in a way that

$$c_m + c_p \leq r_{ev} \leq c_{ubad}$$

We find the optimal percentage of selected markers $P_{t,k}^{*obj}$ by solving the following optimization problem:

$$P_{t,k}^{*obj} = \underset{P_{t,k}^{obj}}{arg\max} \; V_{t,k}\big(P_{t,k}^{obj}\big) \tag{5.24}$$

$$s.t. \quad P_{t,k}^{obj} \in \{10\%, 20\%, 30\%, \dots, 100\%\} \tag{5.25}$$

where the objective function in (5.22) is calculated using (5.14)-(5.23). The total evaluation time T_{total} can be set as any value, since it is actually canceled in both the denominator and numerator.

5.3. **Experimental Results**

The newly proposed method is evaluated using the lithography overlay error model and data corresponding to a 4-layer industrial photolithography process used in a major 300 mm fab. The weight parameters in the objective functions (5.2) are set to $\lambda^x = \lambda^y = \alpha^x = \alpha^y = 1$. Please note that these weight factors do not reflect the actual relative importance of layers in the process. These factors and other specific details, including the model structure, model parameters, and noise characteristics need to be concealed due to proprietary nature of the process. Baseline parameters of the objective function (5.19) are summarized in Table 2. To gain a better understanding of the newly proposed method, we will further study the sensitivity of objective function (5.19) to fluctuations in those parameters, which are summarized in Table 2 as well.

Table 2. Summary of Baseline Parameters of Objective Function (5.19) and their Alternative Values where Results were Evaluated for Sensitivity Analysis

Category	Illustration	Baseline Value	Fluctuations
Revenue (monetary unit)	revenue per perfect layer pattern	$r_{ev} = 550$	$\{350, 550, 700, 800\}$
Cost (monetary unit)	cost per misidentified imperfect layer pattern	$c_{ubad} = 800$	$\{550, 800, 1200, 1600\}$
	production cost per layer pattern	$c_p = 300$	$\{100, 200, 300, 400\}$
	measurement equipment cost per layer pattern when all the markers are selected	$c_m = 50$	$\{10, 50, 90, 130\}$
Time (time unit)	total evaluation time period	$T_{Total} = 1$	-
	measurement time per layer pattern when all markers are selected	$T_m = 0.3$	$\{0.1, 0.3, 0.5, 0.7\}$
	production time per layer pattern	$T_p = 1 - T_m$	-

Overlay errors and commanded control inputs for 80 consecutive wafers from the actual production process are used as historical data to initialize the experiments. For all historical wafers $t \in \{1,2,\dots,80\}$, the estimated process bias and variance of estimation can be calculated using (5.3) and (5.8) with all the markers being selected. Given those

estimates, in this chapter, we built a heteroscedastic Gaussian process regression (GPR) based R2R prediction model [145], yielding the predicted process bias vectors $\boldsymbol{\mu}_{t,k}^{cx}, \boldsymbol{\mu}_{t,k}^{cy}$ and variance-covariance matrices $\boldsymbol{\sigma}_{t,k}^{cx\,2}, \boldsymbol{\sigma}_{t,k}^{cy\,2}$ for all pattern layers $k = 1,2,\dots,K$ on wafer the next wafer $t = 81$. For all historical wafers $t \in \{1,2,\dots,80\}$, vector of residuals can be calculated using (5.7), which were used to estimate the variance $\boldsymbol{\sigma}_{t,k}^{rx\,2}, \boldsymbol{\sigma}_{t,k}^{ry\,2}$ of residuals at all the markers. Given those mean and variance of stochastic terms, the newly proposed procedure described in Section 5.2 was used to obtain values of the objective function (5.16) under the varying percentage of markers that need to be retained. Based on this, the optimization problem (5.20), (5.21) was solved and the optimal percentage $P_{t,k}^{*obj}$ and selection $\boldsymbol{F}_{t,k}^{*}(P_{t,k}^{*obj})$ of measurement markers were determined.

For every layer $k \in \{1,\dots,K\}$, after overlay errors $o_{t,k,i}^{x}, o_{t,k,i}^{y}$ were observed at selected markers $i \in \boldsymbol{F}_{t,k}^{*}(P_{t,k}^{*obj})$, the means and variance of stack-up overlay errors at all the selected markers $i \in \boldsymbol{F}_{t,k}^{*}(P_{t,k}^{*obj})$ are calculated as

$$\begin{cases} \mu_{t,k,i}^{sx} = \mu_{t,k-1,i}^{sx} + o_{t,k,i}^{x} \\ \mu_{t,k,i}^{sy} = \mu_{t,k-1,i}^{sy} + o_{t,k,i}^{y} \\ \sigma_{t,k,i}^{sx\,2} = \sigma_{t,k-1,i}^{sx\,2} \\ \sigma_{t,k,i}^{sy\,2} = \sigma_{t,k-1,i}^{sy\,2} \end{cases}$$

where the mean $\boldsymbol{\mu}_{t,0}^{sx}, \boldsymbol{\mu}_{t,0}^{sy}$ and variance $\boldsymbol{\sigma}_{t,0}^{sx\,2}, \boldsymbol{\sigma}_{t,0}^{sy\,2}$ of stack-up overlay errors are zeros. For all the unselected markers $j \in \overline{\boldsymbol{F}}_{t,k}^{*}\left(P_{t,k}^{*obj}\right)$, the means and variance of stack-up overlay errors are updated using the estimated overlay errors following (5.17).

5.3.1. **Results for the Baseline Settings**

Optimal values of the summation of 2-Wasserstein distance in objective function (5.12) for various percentages of selected markers are shown in Figure 33. It can be seen that, with a careful selection of measurement markers, decreasing the percentage of selected markers from 100% to 60% has little influence on the estimation of process bias terms, but the distance metric starts to rise rapidly when the percentage of selected markers continue to decrease. This deviation of estimation leads to a decrease in the accuracy of our understanding of yield rate behavior at unselected markers. Figure 34 shows the layer-specific yield rate (a) $\mathbb{P}(\boldsymbol{\mathcal{A}})$ based on selected markers, (b) $\mathbb{P}(\boldsymbol{\mathcal{B}})$ based on unselected markers and (c) $\mathbb{P}(\boldsymbol{\mathcal{A}})\mathbb{P}(\boldsymbol{\mathcal{B}})$ based on all the markers. We observe that when the percentage of selected markers decreases, in (a), the yield rate $\mathbb{P}(\boldsymbol{\mathcal{A}})$ keeps increasing to one, which indicates that we are less likely to identify bad layer patterns through observations. In (b), the yield rate $\mathbb{P}(\boldsymbol{\mathcal{B}})$ continues to decrease, and the amount of change is greater than $\mathbb{P}(\boldsymbol{\mathcal{A}})$, because $\mathbb{P}(\boldsymbol{\mathcal{B}})$ is not only affected by the percentage of unselected markers, but also affected by the reduced accuracy of the understanding of overlay errors at unselected markers. It means that we are becoming less and less confident about our

estimation of overlay errors. Therefore we have to admit that based on our current understanding, the probability of them being bad is increasing. Dominated by $\mathbb{P}(\boldsymbol{\mathcal{B}})$, in (c), the yield rate $\mathbb{P}(\boldsymbol{\mathcal{A}})\mathbb{P}(\boldsymbol{\mathcal{B}})$ at all the markers keeps decreasing as well.

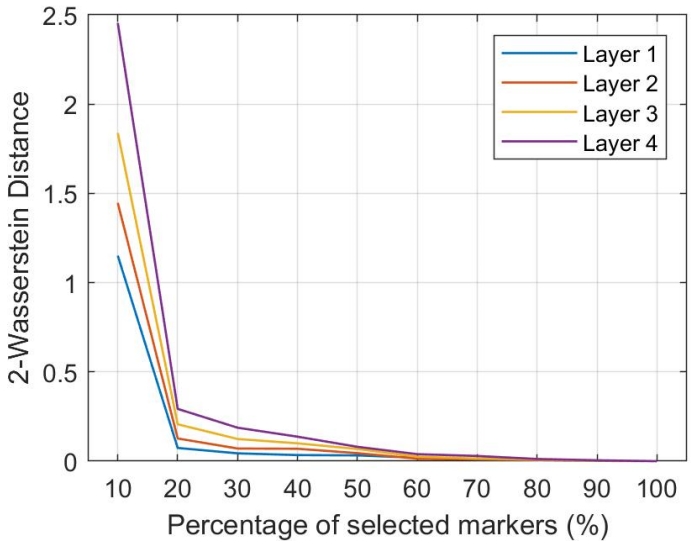

Figure 33. Plots of layer-specific objective function in (5.12) when the optimal set of markers is found for various percentages of selected markers.

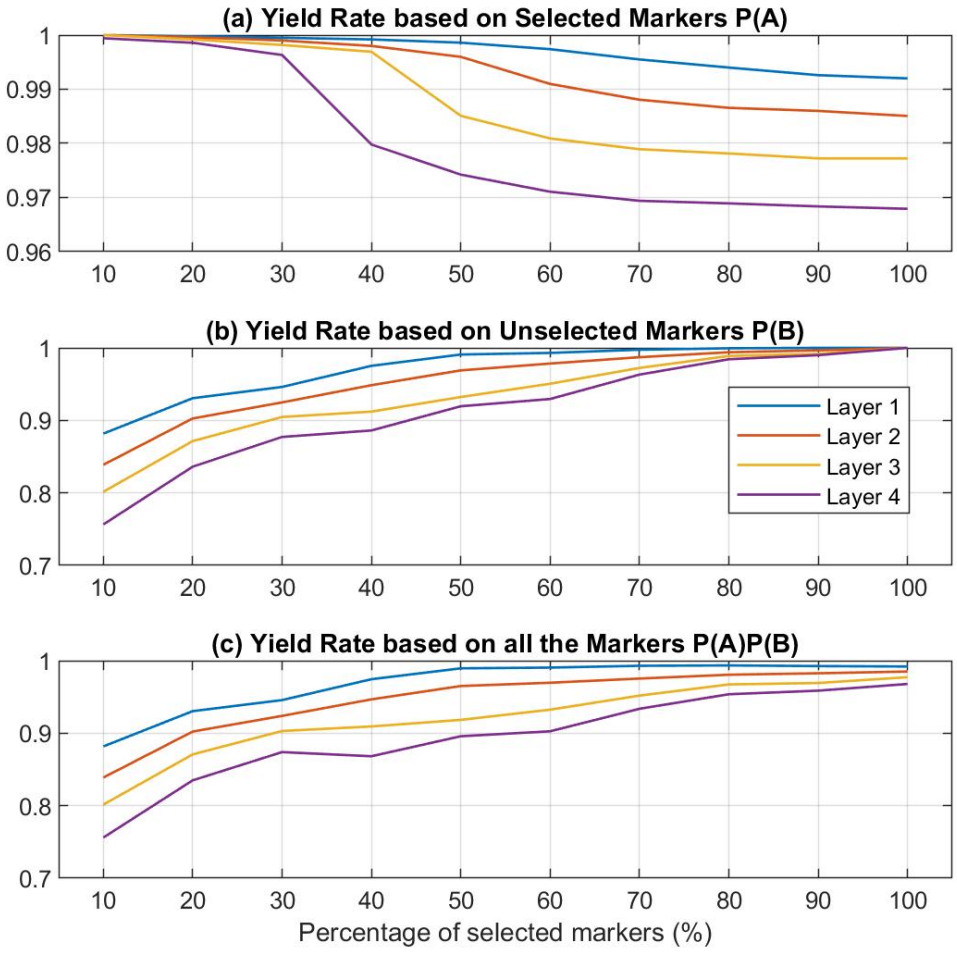

Figure 34. Plots of layer-specific yield rate (a) $\mathbb{P}(\boldsymbol{\mathcal{A}})$ based on selected markers, (b) $\mathbb{P}(\boldsymbol{\mathcal{B}})$ based on unselected markers and (c) $\mathbb{P}(\boldsymbol{\mathcal{A}})\mathbb{P}(\boldsymbol{\mathcal{B}})$ based on all the markers.

Figure 35 shows how the percentage of selected markers affect the objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$: profit per unit time in (5.23) and its components. When the percentage of selected marker decreases, we observe the following. In (a), an optimal percentage of selected markers can be found for each layer pattern, resulting in the maximum profit per unit time. In (b), the revenue earned from perfect layer patterns increases, as the cycle-time decreases, and we are less likely to observe bad layer patterns. In (c), the cost of

misidentified bad layer pattern increases, especially when the percentage is below 50%, which is due to the increase of the percentage of misidentified bad layers $\mathbb{P}(\boldsymbol{\mathcal{A}}) - \mathbb{P}(\boldsymbol{\mathcal{A}})\mathbb{P}(\boldsymbol{\mathcal{B}})$ and reduction of cycle-time. In (d), increases of the production and measurement cost are the same for all the layers, as it is affected by the cycle-time and measurement cost, which both solely depend on the percentage of selected markers.

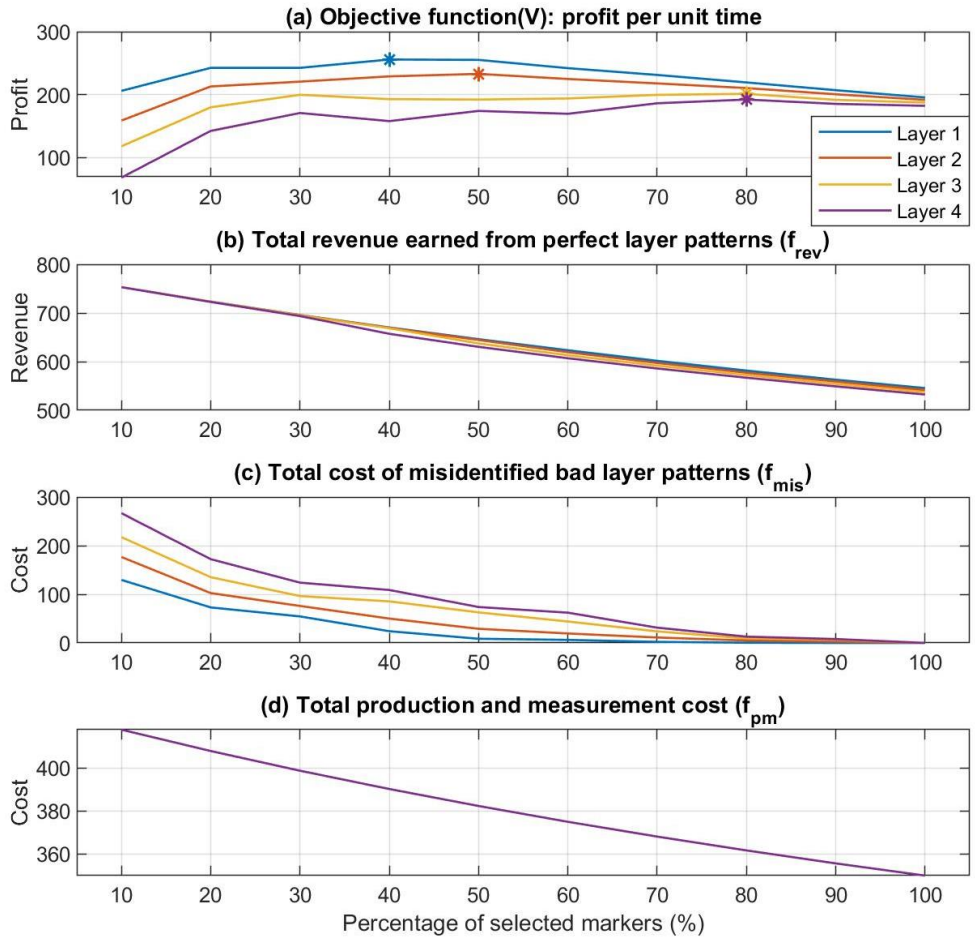

Figure 35. Plots of (a) layer-specific objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum, (b) layer-specific f_{rev}, (c) layer-specific f_{mis} and (d) f_{pm}.

5.3.2. **Influence of the Revenue per Perfect Layer Pattern**

In this section, results of the newly proposed method, including the value of the objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum at each layer, are evaluated with the revenue per perfect layer pattern r_{ev} being different values $\{350, 550, 700, 800\}$ as shown in Figure 36. We can observe that we tend to select more measurement markers, when the value of r_{ev} decreases. It is because of the resulting increase in the ratio of the cost we paid for one misidentified imperfect layer to the revenue we earned from it. Namely, the reduction in the revenue per perfect layer makes us unwilling to bear the cost of selling bad products, leading to an increase in the percentage of selected markers to ensure a better yield rate. In addition, it is worth notice that when the revenue r_{ev} is decreased to 350, which is the same as the total production and measurement cost when all the markers are selected, a positive profit can still be achieved with a proper reduction of the percentage of measurement markers.

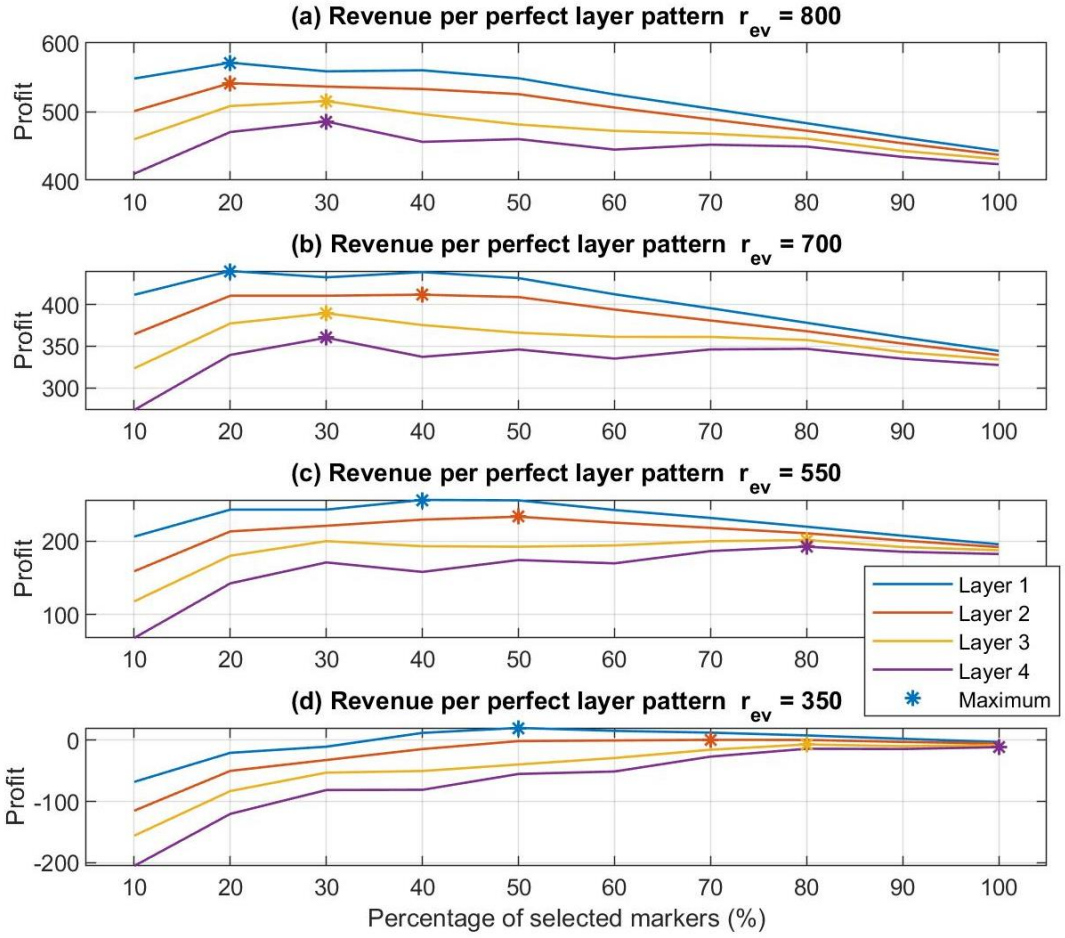

Figure 36. Plots of layer-specific objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum with r_{ev} in $\{350, 550, 700, 800\}$.

5.3.3. **Influence of the Cost per Misidentified Bad Layer Pattern**

In order to illustrate the influence of the cost per misidentified imperfect layer pattern on results of the newly proposed method, the profit $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum at each layer, with c_{ubad} being different values $\{550, 800, 1200, 1600\}$, are shown in Figure 37.

As can be observed, when the value of c_{ubad} increases, the maximum profit tends to be achieved at the point with a larger percentage of selected markers. Namely, when the cost of mistakenly selling a bad product increases, we are willing to measure more markers to increase our understanding of yield rate.

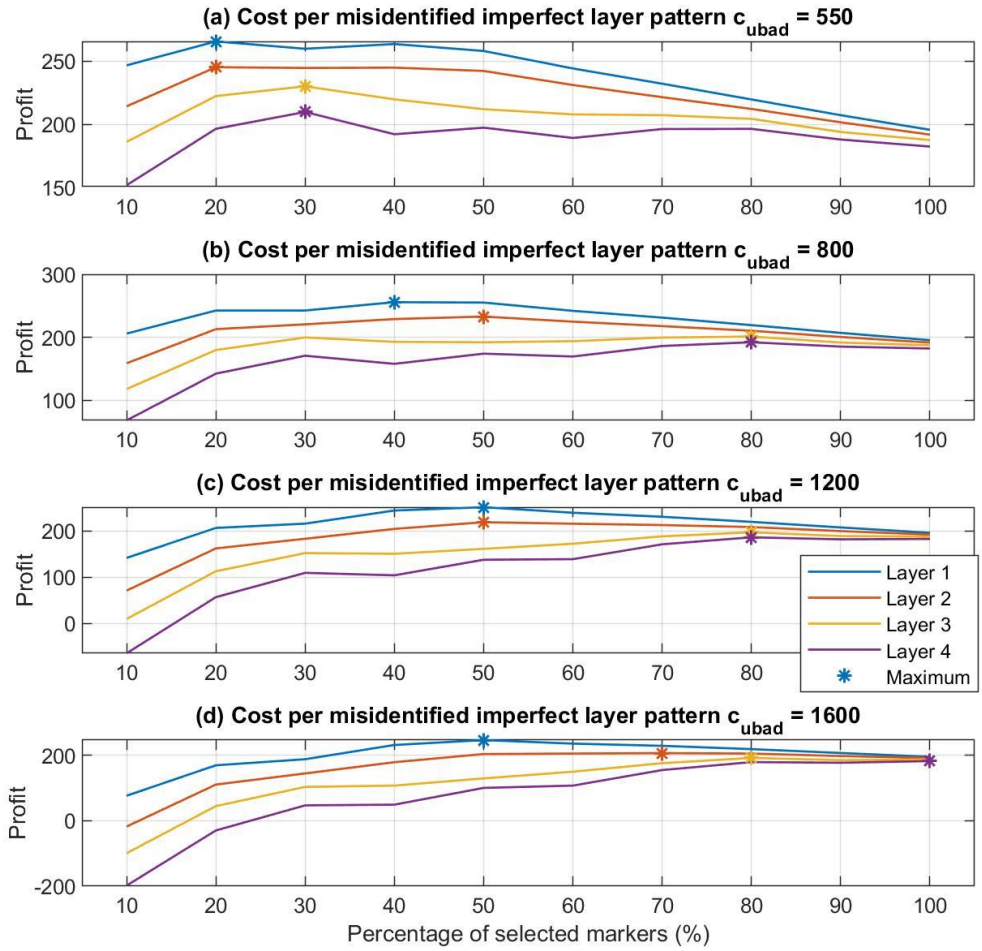

Figure 37. Plots of layer-specific objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum with c_{ubad} in $\{550, 800, 1200, 1600\}$.

5.3.4. **Influence of the Production and Measurement Cost**

In this section, sensitivity analyses are conducted to illustrate the influence of the production and measurement cost on the results of the newly proposed method. Figure 38 shows the objective function $V_{t,k}(P_{t,k}^{obj})$ and its maximum for each layer pattern, with the production cost c_p in $\{100, 200, 300, 400\}$. It can be observed that when the production cost increases, we prefer to select more markers. This is because the profit margin is squeezed and the cost of measurements becomes relatively less important. We cannot afford too much cost resulting from misidentified products and prefer to spend more cost and time measuring more markers.

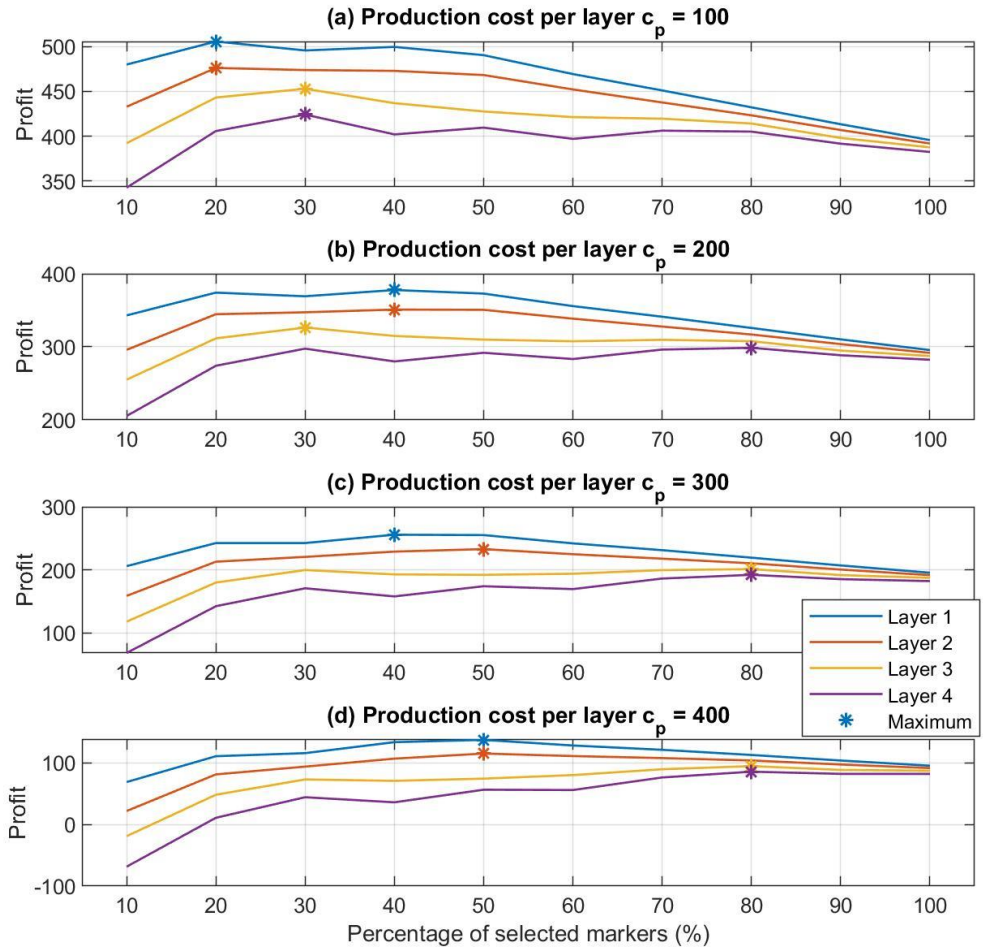

Figure 38. Plots of layer-specific objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum with c_p in $\{100, 200, 300, 400\}$.

Figure 39 shows the objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum for each layer pattern, with measurement cost c_m in $\{10, 50, 90, 130\}$. It can be seen that when the measurement cost decreases, we tend to select more markers, as with the same

measurement cost, we can measure more markers and bring more profits with increased accuracy of estimation.

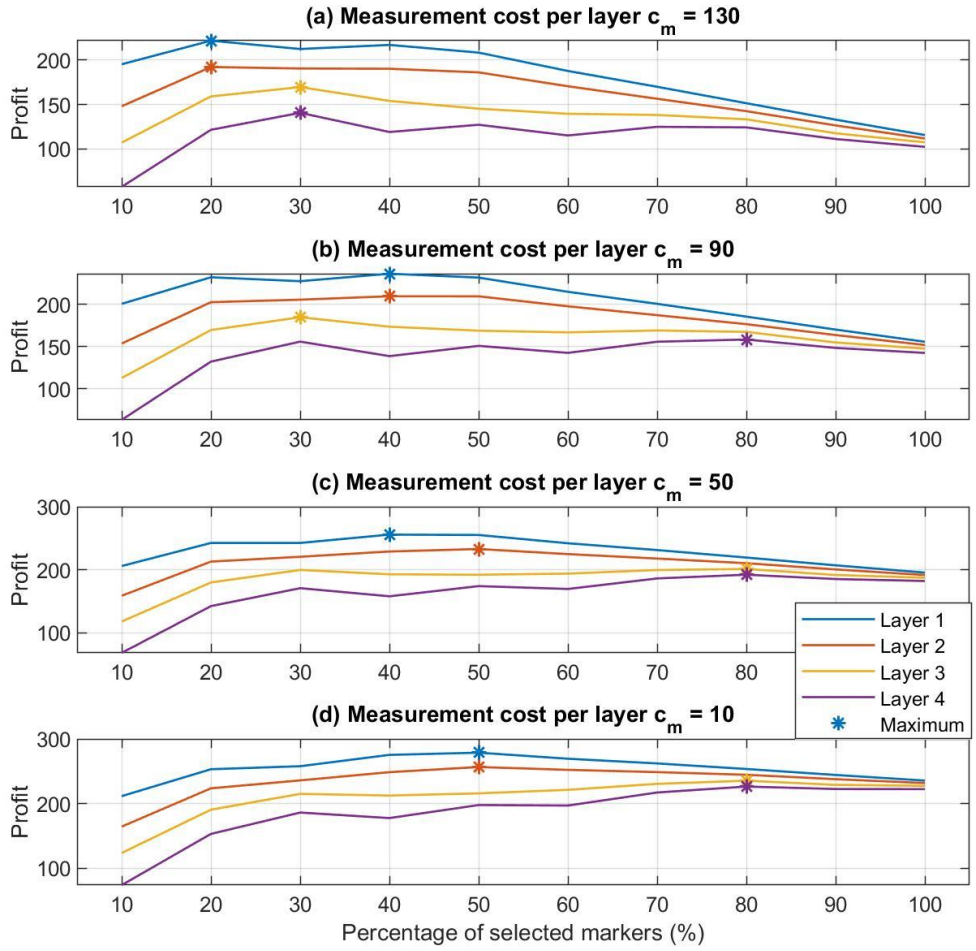

Figure 39. Plots of layer-specific objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum with c_m in $\{10, 50, 90, 130\}$.

5.3.5. **Influence of the Measurement Time**

In order to evaluate the influence of measurement time on the results of the proposed method, the profit $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum at each layer, with the measurement time T_m being different values $\{0.1, 0.3, 0.5, 0.7\}$ are shown in Figure 37, where the production time is calculated with $T_p = 1 - T_m$. Therefore, the value of T_m is actually the percentage of measurement time in the total time of each layer. We can observe that when the percentage of measurement time decreases, we prefer to select more markers, as increasing the percentage of markers now has a reduced impact on the cycle time so that its benefits outweigh the drop in the efficiency.

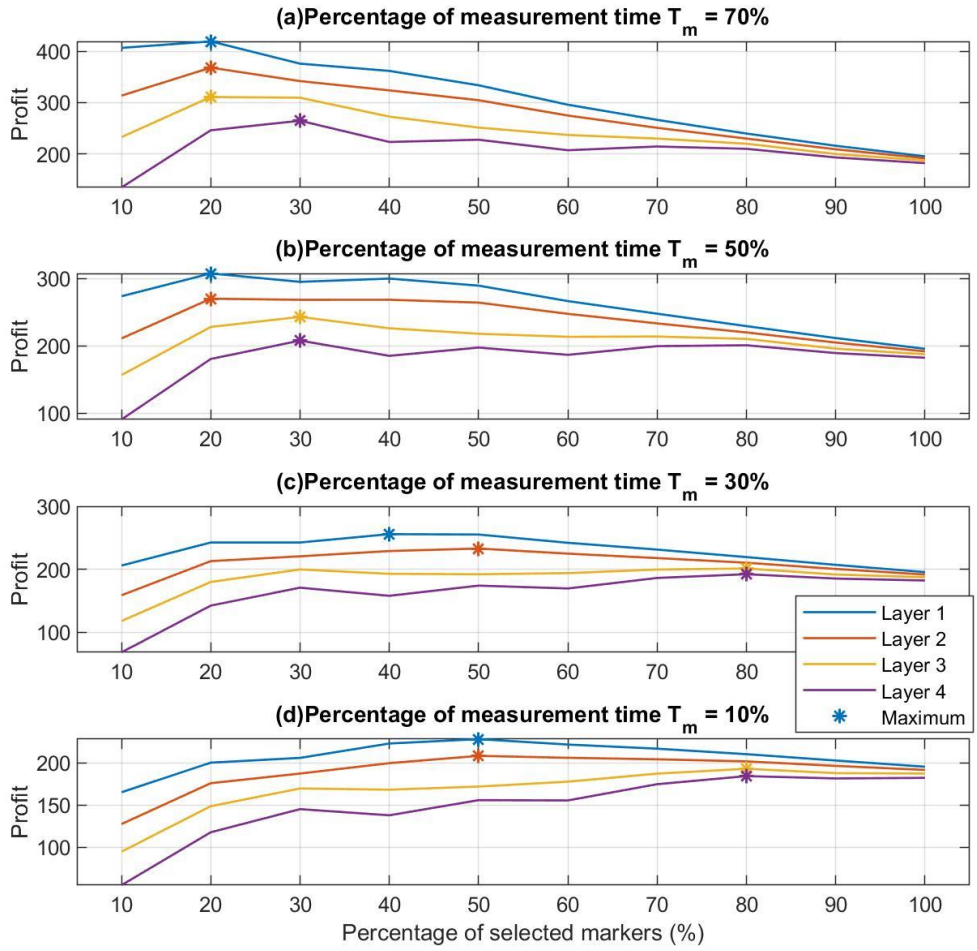

Figure 40. Plots of layer-specific objective function $V_{t,k}\left(P_{t,k}^{obj}\right)$ and its maximum with T_m in $\{0.1, 0.3, 0.5, 0.7\}$, and $T_p = 1 - T_m$.

5.4. **Conclusions and Future Work**

In this chapter, we propose a novel method for the dynamic decision-making on the number and selection of measurement markers from an exhaustive set of available markers

in photolithography processes. Given a history of overlay measurements from previously manufactured wafers, an optimal number of markers is decided that maximizes the profit considering revenue earned from perfect layer patterns, cost of misidentified bad layers, as well as production and measurement cost. At the same time, the distribution of those markers is optimized considering one's ability to estimate actuator uncertainties and to understand the yield rate behavior. The control commands are determined using a stochastic control algorithm for control of overlay errors.

Application of this method to the data and models from an industrial-scale semiconductor lithography process is presented to illustrate the newly proposed method. To evaluate effects of a variety of relevant parameters on the profit of the system, sensitivity analyses were conducted and results showed that we tend to select more measurement markers in any of the following cases, including when the revenue per perfect layer pattern decreases, when the cost per misidentified bad layer pattern increases, when the production cost increases, when the maintenance cost decreases, and when measurement time decreases. The reasons for these results can be summarized in the fact that the newly proposed method can adjust to changes in the systems by pursuing optimization of the number of selection of measurement markers from a system-level operational aspect with considerations of quality control, cycle-time, and understanding of yield rate behavior.

Several research extensions could be built on the work presented in this chapter. One obvious direction is implementing the concepts and methods introduced in this chapter to

an industrial scale problem. In addition, the optimization framework can be further improved to incorporate more system-level production strategies, such as sample size of wafers, optimal buffer capacity, and the dynamics of production speed.

Chapter 6. Conclusions and Future Work

6.1. Summary of the Research

The dissertation studies the problems associated with integrated operational decision-making in flexible manufacturing systems with considerations of quality and reliability. The objective is to develop decision-making frameworks that concurrently consider various aspects of system operations to optimize customizable objective functions that reward the quality of products, while penalizing the utilization of manufacturing resources.

Chapter 2 presents a literature review on the research related to this dissertation. Firstly, a review of integrated maintenance and production planning with considerations of yield rate is given. Then, a review of algorithms for online decision problems is presented to understand methods for online learning of yield rate. These two sections illustrate existing approaches to the problem studied in Chapter 3. For problems of optimal selection of quality measurements studied in Chapter 4 and Chapter 5, a general review of research dealing with this kind of problems in manufacturing systems is presented. Then, more specifically, existing literature with a focus on solving the problem for the control of overlay errors in photolithography processes is reviewed.

Chapter 3 presents a novel method for integrated production and maintenance planning on a single machine with the novel method incorporating concurrent learning of

yield rate under uncertain demand. The newly proposed method was shown to consistently outperform the policy obtained without learning of yield rates, or under assumption that yield depends on the equipment condition only. The benefits of considering stochasticities of the demand, degradation process and maintenance operations were demonstrated via comparisons with the decisions made with a deterministic model solely based on the expected values of the demand, machine uptimes and maintenance times.

Chapter 4 presents a novel method for optimal down-selection of overlay measurement markers in photolithography processes which adaptively selects a subset of available overlay measurement markers from one wafer to another in a way that the selected markers facilitate best possible performance of the recently introduced robust control algorithm for control of lithography overlay errors. The marker selection problem is formulated as a bi-level robust optimization problem with the objective of minimizing the worst-case outcomes of the weighted sum of squared overlay and stack-up errors, where the worst-case performance is evaluated over uncertainties in the modeling noise and process bias terms. The underlying robust optimization problem does not assume anything about the distribution of uncertainties except for the prior knowledge of upper and lower bounds, which are maintained and adjusted for each wafer. The underlying optimization framework is proposed in such a way that it can be efficiently solved using a GA-based nested evolutionary algorithm and commercial solvers. The newly proposed method is offered with validation using data and models from an industrial-scale semiconductor photolithography process. The results show that the newly proposed combination of the

robust overlay control paradigm and optimized marker selection enables improved overlay control, even with a significantly reduced number of markers. Thus, the new methodology enables reduction of measurement times and subsequent overall cycle times, without deteriorating the outgoing product quality.

An extension of the research described in Chapter 4 was pursued in Chapter 5, where the influence of the number of selected markers on cycle-times of the resulting process and the understanding of yield rate behaviors are considered. Optimal number of markers is decided by maximizing the profit which considers revenue earned from perfect layer patterns, cost of misidentified bad layers, as well as production and measurement costs. Given any number of available markers, the spatial layout of those markers is optimized considering one's ability to understand the yield rate behavior and estimate actuator uncertainties for the stochastic control of overlay errors. Application of this method to the data and models from an industrial-scale semiconductor lithography process is presented to illustrate the benefit of the newly proposed method. Sensitivity analyses demonstrate that the newly proposed method can adjust to changes in the underlying cost parameters affecting the system by pursuing optimization of the number and selection of measurement markers from a system-level operational aspect, with considerations of quality control, cycle-times, and knowledge about yield rate.

6.2. **Scientific Contributions**

Several scientific contributions are expected from this doctoral research：

1. A novel decision-making framework for integrated production and maintenance planning under uncertain demand with concurrent learning of dependences of yield rate on production rate and machine conditions. To the best of our knowledge, this is the first integrated decision-making study in a manufacturing system that considers the learning of yield rate.

2. A new optimization method for the selection of measurement markers for the robust control of overlay errors. In the existing literature on the topic of the selection of measurement markers, the proposed work is the first to consider robustness of uncertainties in the model of overlay errors to deal with uncertainties whose characteristics cannot be accurately modeled.

3. A novel optimization framework for decision-making regarding number and spatial allocation of measurement markers that maximizes the profit of the manufacturing process and information collected from measurments. In the existing literature, this dissertation is the first to quantify the influence of the layout of markers on the profit of manufacturing from both the quality control and system-level operational aspect.

6.3. **Publications**

This section presents a list of publications already produced or anticipated to be produced based on this doctoral research.

- Zhang, H. and Djurdjanovic, D., 2021. "Integrated production and maintenance planning under uncertain demand with concurrent learning of yield rate," *Flexible Services and Manufacturing Journal*, pp.1-22.

- Zhang, H., Feng, T. and Djurdjanovic, D., "Dynamic down-selection of measurement markers for optimized robust control of overlay errors in photolithography," under review, *IEEE Transactions on Semiconductor Manufacturing*, Manuscript ID TSM-21-0179.

- Zhang, H., Feng, T. and Djurdjanovic, D., "Dynamic decision-making on number and selection of measurement markers for stochastic control of overlay errors in photolithography," anticipated journal paper based on Chapter 5. This manuscript will be submitted to IIE Transactions.

6.4. **Potential Future Work**

Based on above-completed tasks and the current demands in academia and industry, possible directions for future research are summarized in this section. Firstly, one obvious

opportunity for future work is an industrial-scale implementation of the concepts and methods introduced in this dissertation.

Methodologically, robustness to uncertainties in model parameters and structure can be considered within the decision-making process. Namely, the method proposed in Chapter 3 can be extended by considering the fact that parameters of degradation dynamics are never perfectly known and degradation states are rarely directly observable, which can be addressed e.g. by modeling the degradation process using Hidden Markov Model (HMM) with unknown parameters, as was done in [86] . The work presented in Chapter 5 can be further extended to consider assumptions of uncertainties and their updating mechanisms, as proposed in Chapter 4.

From the practical point of view, the concepts and methodologies proposed in this dissertation can be further extended to more realistic manufacturing systems. The optimization framework proposed in Chapter 3 can be improved by involving multiple machines and multiple products, as well as more realistic logistic and organizational constraints. The decision-making framework proposed in Chapter 5, can be further improved to incorporate more system-level production strategies, such as sample size of wafers, optimal buffer capacity and the dynamics of production speed.

Bibliography

[1] Shivanand, H. K. (2006). Flexible manufacturing system. New Age International.

[2] Basnet, C., & Mize, J. H. (1994). Scheduling and control of flexible manufacturing systems: a critical review. International Journal of Computer Integrated Manufacturing, 7(6), 340-355.

[3] Heizer, J., Render, B., & Munson, C. (2008). Operations management. Prentice-Hall.

[4] Kaushal, A., Vardhan, A., & Rajput, R. S. (2016). Flexible Manufacturing System. A modern approach to manufacturing technology. International Refereed Journal of Engineering and Science, 5(4), 16-23.

[5] Ekin, T. (2018). Integrated maintenance and production planning with endogenous uncertain yield. Reliability Engineering & System Safety, 179, 52-61.

[6] Bearda, T., Mertens, P. W., & Beaudoin, S. P. (2018). Overview of Wafer Contamination and Defectivity. In Handbook of Silicon Wafer Cleaning Technology (Third Edition) (pp. 87-149).

[7] Madron, František, and Vladimír Veverka. "Optimal selection of measuring points in complex plants by linear models." AIChE journal 38.2 (1992): 227-236.

[8] Djurdjanovic, D., & Ni, J. (2004). Measurement scheme synthesis in multi-station machining systems. J. Manuf. Sci. Eng., 126(1), 178-188.

[9] Djurdjanovic, D., Mears, L., Niaki, F. A., Haq, A. U., & Li, L. (2018). State of the art review on process, system, and operations control in modern manufacturing. Journal of Manufacturing Science and Engineering, 140(6), 061010.

[10] Khouja, M., & Mehrez, A. (1994). Economic production lot size model with variable production rate and imperfect quality. Journal of the Operational Research Society, 45(12), 1405-1417.

[11] Iravani, S. M., & Duenyas, I. (2002). Integrated maintenance and production control of a deteriorating production system. Iie Transactions, 34(5), 423-435.

[12] Wang, H. (2002). A survey of maintenance policies of deteriorating systems. European journal of operational research, 139(3), 469-489.

[13] Alaswad, S., & Xiang, Y. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. Reliability Engineering & System Safety, 157, 54-63.

[14] Sloan, T. W., & Shanthikumar, J. G. (2000). Combined production and maintenance scheduling for a multiple‐product, single‐machine production system. Production and Operations Management, 9(4), 379-399.

[15] Sloan, T. (2008). Simultaneous determination of production and maintenance schedules using in ‐ line equipment condition and yield information. Naval Research Logistics (NRL), 55(2), 116-129.

[16] Batun, S., & Maillart, L. M. (2012). Reassessing tradeoffs inherent to simultaneous maintenance and production planning. Production and Operations Management, 21(2), 396-403.

[17] Sloan, T. W. (2004). A periodic review production and maintenance model with random demand, deteriorating equipment, and binomial yield. Journal of the Operational Research Society, 55(6), 647-656.

[18] Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*(3/4), 285-294.

[19] Thompson, W. R. (1935). On the theory of apportionment. *American Journal of Mathematics*, *57*(2), 450-456.

[20] Russo, D., Van Roy, B., Kazerouni, A., & Osband, I. (2017). A Tutorial on Thompson Sampling. arXiv preprint arXiv:1707.02038.

[21] Chapelle, O., & Li, L. (2011). An empirical evaluation of Thompson sampling. In *Advances in neural information processing systems* (pp. 2249-2257).

[22] Scott, S. L. (2010). A modern Bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, *26*(6), 639-658.

[23] Gittins, J. C., & Jones, D. M. (1979). A dynamic allocation index for the discounted multiarmed bandit problem. Biometrika, 66(3), 561-565.

[24] Katehakis, M. N., & Veinott Jr, A. F. (1987). The multi-armed bandit problem: decomposition and computation. Mathematics of Operations Research, 12(2), 262-268.

[25] Gittins, J., Glazebrook, K., & Weber, R. (2011). Multi-armed bandit allocation indices. John Wiley & Sons.

[26] Lai, Tze Leung, and Herbert Robbins. "Asymptotically efficient adaptive allocation rules." Advances in applied mathematics 6.1 (1985): 4-22.

[27] Russo, D., & Van Roy, B. (2014). Learning to optimize via posterior sampling. Mathematics of Operations Research, 39(4), 1221-1243.

[28] Osband, I., & Van Roy, B. (2017). On optimistic versus randomized exploration in reinforcement learning. arXiv preprint arXiv:1706.04241.

[29] Osband, I., & Van Roy, B. (2017, July). Why is posterior sampling better than optimism for reinforcement learning?. In International Conference on Machine Learning (pp. 2701-2710).

[30] Russo, D., & Van Roy, B. (2014). Learning to optimize via information-directed sampling. In Advances in Neural Information Processing Systems (pp. 1583-1591).

[31] Frazier, P. I., Powell, W. B., & Dayanik, S. (2008). A knowledge-gradient policy for sequential information collection. SIAM Journal on Control and Optimization, 47(5), 2410-2439.

[32] Frazier, P., Powell, W., & Dayanik, S. (2009). The knowledge-gradient policy for correlated normal beliefs. INFORMS journal on Computing, 21(4), 599-613.

[33] Shui, H. (2018). Real-Time Monitoring and Fault Diagnostics in Roll-To-Roll Manufacturing Systems (Doctoral dissertation).

[34] Lee, J., Ni, J., Singh, J., Jiang, B., Azamfar, M., & Feng, J. (2020). Intelligent Maintenance Systems and Predictive Manufacturing. Journal of Manufacturing Science and Engineering, 142(11).

[35] Ding, Y., Kim, P., Ceglarek, D., & Jin, J. (2003). Optimal sensor distribution for in multistation assembly processes. IEEE Transactions on Robotics and Automation, 19(4), 543-556.

[36] Shukla, N., Ceglarek, D., & Tiwari, M. K. (2015). Key characteristics-based sensor distribution in multi-station assembly processes. Journal of intelligent manufacturing, 26(1), 43-58.

[37] Wu, Z., Hsieh, S. J., & Li, J. (2013). Sensor deployment based on fuzzy graph considering heterogeneity and multiple-objectives to diagnose manufacturing system. Robotics and Computer-Integrated Manufacturing, 29(1), 192-208.

[38] He, K., Wang, N., & Zhu, L. (2017, November). A Review of Sensor Layout for Condition Monitoring during Discrete-part Manufacturing. In 2017 International Conference on Applied Mathematics, Modeling and Simulation (AMMS 2017) (pp. 444-447). Atlantis Press.

[39] He, K., Jia, M., & Xu, Q. (2015). Optimal sensor deployment for manufacturing process monitoring based on quantitative cause-effect graph. IEEE Transactions on Automation Science and Engineering, 13(2), 963-975.

[40] Lee, J. (1998). Teleservice engineering in manufacturing: challenges and opportunities. International Journal of Machine Tools and Manufacture, 38(8), 901-910.

[41] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. Computer networks, 38(4), 393-422.

[42] Khan, A., & Ceglarek, D. (2000). Sensor optimization for fault diagnosis in multi-fixture assembly systems with distributed sensing. J. Manuf. Sci. Eng., 122(1), 215-226.

[43] Edan, Y., & Nof, S. Y. (2000). Sensor economy principles and selection procedures. Iie Transactions, 32(3), 195-203.

[44] Liu, C. Q., Ding, Y., & Chen, Y. (2005). Optimal coordinate sensor placements for estimating mean and variance components of variation sources. IIE transactions, 37(9), 877-889.

[45] Ren, Y., & Ding, Y. (2009). Optimal sensor distribution in multi-station assembly processes for maximal variance detection capability. IIE transactions, 41(9), 804-818.

[46] Li, J., & Jin, J. (2010). Optimal sensor allocation by integrating causal models and set-covering algorithms. IIE Transactions, 42(8), 564-576.

[47] Cao, H., Niu, L., & He, Z. (2012). Method for vibration response simulation and sensor placement optimization of a machine tool spindle system with a bearing defect. Sensors, 12(7), 8732-8754.

[48] Abellan-Nebot, J. V., Liu, J., & Subirón, F. R. (2012). Quality prediction and compensation in multi-station machining processes using sensor-based fixtures. Robotics and Computer-Integrated Manufacturing, 28(2), 208-219.

[49] Liu, K., & Shi, J. (2013). Objective-oriented optimal sensor allocation strategy for process monitoring and diagnosis by multivariate analysis in a Bayesian network. Iie Transactions, 45(6), 630-643.

[50] Bhushan, M., & Rengaswamy, R. (2002). Comprehensive design of a sensor network for chemical plants based on various diagnosability and reliability criteria. 1. Framework. Industrial & engineering chemistry research, 41(7), 1826-1839.

[51] Wu, Z., Hsieh, S. J., & Li, J. (2013). Sensor deployment based on fuzzy graph considering heterogeneity and multiple-objectives to diagnose manufacturing system. Robotics and Computer-Integrated Manufacturing, 29(1), 192-208.

[52] Hu, S. J., & Koren, Y. (1997). Stream-of-variation theory for automotive body assembly. CIRP Annals, 46(1), 1-6.

[53] Camelio, J., Hu, S. J., & Ceglarek, D. (2003). Modeling variation propagation of multi-station assembly systems with compliant parts. J. Mech. Des., 125(4), 673-681.

[54] Agrawal, R., Lawless, J. F., & Mackay, R. J. (1999). Analysis of variation transmission in manufacturing processes—part II. Journal of Quality Technology, 31(2), 143-154.

[55] Djurdjanovic, D. R. A. G. A. N., & Ni, J. (2001). Linear state space modeling of dimensional machining errors. Transactions-North American Manufacturing Research Institution of SME, 541-548.

[56] Djurdjanovic, D., & Ni, J. (2003). Dimensional errors of fixtures, locating and measurement datum features in the stream of variation modeling in machining. J. Manuf. Sci. Eng., 125(4), 716-730.

[57] Zhou, S., Huang, Q., & Shi, J. (2003). State space modeling of dimensional variation propagation in multistage machining process using differential motion vectors. IEEE Transactions on robotics and automation, 19(2), 296-309.

[58] Ceglarek, D., & Shi, J. (1996). Fixture failure diagnosis for autobody assembly using pattern recognition.

[59] Ceglarek, D., & Shi, J. (1999). Fixture failure diagnosis for sheet metal assembly with consideration of measurement noise.

[60] Huang, Q., Zhou, S., & Shi, J. (2002). Diagnosis of multi-operational machining processes by using virtual machining. Robotics and Computer Integrated Manufacturing, 18, 233-239.

[61] Zhou, S., Ding, Y., Chen, Y., & Shi, J. (2003). Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. Technometrics, 45(4), 312-325.

[62] Djurdjanovic, D., & Ni, J. (2001). Stream of variation based analysis and synthesis of measurement schemes in multi-station machining systems. In Proceedings of the ASME International Mechanical Engineering Congress and Exposition (Vol. 12, pp. 297-304).

[63] Djurdjanovic, D., & Ni, J. (2003). Bayesian approach to measurement scheme analysis in multistation machining systems. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 217(8), 1117-1130.

[64] Djurdjanovic, D., & Ni, J. (2002). Measurement scheme analysis in multi-station machining systems. Trans. ASME, J. Mfg Sci. Engng.

[65] Xiang, L., & Tsung, F. (2008). Statistical monitoring of multi-stage processes based on engineering models. IIE transactions, 40(10), 957-970.

[66] Bhushan, M., & Rengaswamy, R. (2000). Design of sensor location based on various fault diagnostic observability and reliability criteria. Computers & Chemical Engineering, 24(2-7), 735-741.

[67] Yang, F., Xiao, D., & Shah, S. L. (2009). Optimal sensor location design for reliable fault detection in presence of false alarms. Sensors, 9(11), 8579-8592.

[68] Khan, A., Ceglarek, D., & Ni, J. (1998). Sensor location optimization for fault diagnosis in multi-fixture assembly systems.

[69] Khan, A., Ceglarek, D., Shi, J., Ni, J., & Woo, T. C. (1999). Sensor optimization for fault diagnosis in single fixture systems: a methodology.

[70] Costiner, S., Winston, H. A., Gurvich, M. R., Ghoshal, A., Welsh, G. S., Butler, S. L., ... & Bordick, N. (2013). A probabilistic hybrid sensor fusion and optimization approach for aircraft composite components. Journal of intelligent material systems and structures, 24(17), 2110-2134.

[71] Bastani, K., Kong, Z., Huang, W., & Zhou, Y. (2016). Compressive sensing–based optimal sensor placement and fault diagnosis for multi-station assembly processes. IIE Transactions, 48(5), 462-474.

[72] Shukla, N., Tiwari, M. K., & Shankar, R. (2009). Optimal sensor distribution for multi-station assembly process using chaos-embedded fast-simulated annealing. International Journal of Production Research, 47(1), 187-211.

[73] Sun, J. W., Xi, L. F., Pan, E. S., Du, S. C., & Xia, T. B. (2009). Design for diagnosability of multistation manufacturing systems based on sensor allocation optimization. Computers in Industry, 60(7), 501-509.

[74] Ding, Y., Kim, P., Ceglarek, D., & Jin, J. (2003). Optimal sensor distribution for variation diagnosis in multistation assembly processes. IEEE Transactions on Robotics and Automation, 19(4), 543-556.

[75] Djurdjanovic, D., & Ni, J. (2004). Measurement scheme synthesis in multi-station machining systems. J. Manuf. Sci. Eng., 126(1), 178-188.

[76] Lian, F. L., Moyne, J. R., & Tilbury, D. M. (1999). Performance evaluation of control networks for manufacturing systems. In In Proceedings of the ASME International Mechanical Engineering Congress and Exposition (Dynamic Systems and Control Division.

[77] Shen, C. H., Lin, Y. T., Agapiou, J. S., Jones, G. L., Kramarczyk, M. A., & Bandyopadhyay, P. (2003). An innovative reconfigurable and totally automated fixture system for agile machining applications. TECHNICAL PAPERS-SOCIETY OF MANUFACTURING ENGINEERS-ALL SERIES-.

[78] Wang, H., & Huang, Q. (2007). Using error equivalence concept to automatically adjust discrete manufacturing processes for dimensional variation control.

[79] Wang, H., Huang, Q., & Katz, R. (2005). Multi-operational machining processes modeling for sequential root cause identification and measurement reduction.

[80] Jiao, Y., & Djurdjanovic, D. (2010). Joint allocation of measurement points and controllable tooling machines in multistage manufacturing processes. IIE Transactions, 42(10), 703-720.

[81] Ghani, J. A., Jye, P. S., Haron, C. H. C., Rizal, M., & Nuawi, M. Z. (2012). Determination of sensor location for cutting tool deflection using finite element method simulation. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 226(9), 2373-2377.

[82] Bachmann, F., Bergamini, A. E., & Ermanni, P. (2012). Optimum piezoelectric patch positioning: A strain energy–based finite element approach. Journal of intelligent material systems and structures, 23(14), 1575-1591.

[83] Belloli, A., & Ermanni, P. (2007). Optimum placement of piezoelectric ceramic modules for vibration suppression of highly constrained structures. Smart Materials and Structures, 16(5), 1662.

[84] Kim, P., & Ding, Y. (2004). Optimal design of fixture layout in multistation assembly processes. IEEE Transactions on Automation Science and Engineering, 1(2), 133-145.

[85] Camelio, J. A., Hu, S. J., & Yim, H. (2005). Sensor placement for effective diagnosis of multiple faults in fixturing of compliant parts. J. Manuf. Sci. Eng., 127(1), 68-74.

[86] Sen, S., Narasimhan, S., & Deb, K. (1998). Sensor network design of linear processes using genetic algorithms. Computers & chemical engineering, 22(3), 385-390.

[87] Chen, H., Clark, J. A., Tapiador, J. E., Shaikh, S. A., Chivers, H., & Nobles, P. (2009). A multi-objective optimisation approach to IDS sensor placement. In Computational Intelligence in Security for Information Systems (pp. 101-108). Springer, Berlin, Heidelberg.

[88] Worden, K., & Burrows, A. P. (2001). Optimal sensor placement for fault detection. Engineering structures, 23(8), 885-901.

[89] Wang, X., Wang, S., & Jiang, A. (2006, October). Optimized deployment strategy of mobile agents in wireless sensor networks. In Sixth International Conference on Intelligent Systems Design and Applications (Vol. 2, pp. 893-898). IEEE.

[90] Yang, C. C., & Ciarallo, F. W. (2001). Optimized sensor placement for active visual inspection. Journal of Robotic Systems, 18(1), 1-15.

[91] Jung, B. K., Cho, J. R., & Jeong, W. B. (2015). Sensor placement optimization for structural modal identification of flexible structures using genetic algorithm. Journal of Mechanical Science and Technology, 29(7), 2775-2783.

[92] Lu, S., Jiang, M., Sui, Q., Dong, H., Sai, Y., & Jia, L. (2016). Acoustic emission location on aluminum alloy structure by using FBG sensors and PSO method. Journal of Modern Optics, 63(8), 742-749.

[93] Rangarajan, B., Templeton, M. K., Capodieci, L., Subramanian, R., & Scranton, A. B. (1998, June). Optimal sampling strategies for sub-100-nm overlay. In Metrology, Inspection, and Process Control for Microlithography XII (Vol. 3332, pp. 348-359). International Society for Optics and Photonics.

[94] Chien, C. F., Chang, K. H., & Chen, C. P. (2003). Design of a sampling strategy for measuring and compensating for overlay errors in semiconductor manufacturing. International Journal of Production Research, 41(11), 2547-2561.

[95] Chue, C. F., Chiou, T. B., Huang, C. Y., Chen, A. C., & Shih, C. L. (2009, December). Optimization of alignment/overlay sampling and marker layout to improve overlay performance for double patterning technology. In Lithography Asia 2009 (Vol. 7520, p. 75200G). International Society for Optics and Photonics.

[96] Aung, N. L., Chung, W. J., Subramany, L., Hussain, S., Samudrala, P., Gao, H., ... & Gomez, J. M. (2016, March). Overlay optimization for 1x node technology and beyond via rule based sparse sampling. In Metrology, Inspection, and Process Control for Microlithography XXX (Vol. 9778, p. 97782G). International Society for Optics and Photonics.

[97] Subramany, L., Chung, W., Samudrala, P., Gao, H., Aung, N., Gomez, J. M., ... & Yap, L. (2016, March). Advanced overlay: sampling and modeling for optimized run-to-run control. In Metrology, Inspection, and Process Control for Microlithography XXX (Vol. 9778, p. 97782K). International Society for Optics and Photonics.

[98] Lee, K. B., & Kim, C. O. (2019). Marker Layout for Optimizing the Overlay Alignment in a Photolithography Process. IEEE Transactions on Semiconductor Manufacturing, 32(2), 212-219.

[99] Robinson, J., & Rahmat-Samii, Y. (2004). Particle swarm optimization in electromagnetics. IEEE transactions on antennas and propagation, 52(2), 397-407.

[100] Wang, P., Gao, R. X., & Yan, R. (2017). A deep learning-based approach to material removal rate prediction in polishing. CIRP Annals, 66(1), 429-432.

[101] Jamrus, T., Chien, C. F., Gen, M., & Sethanan, K. (2017). Hybrid particle swarm optimization combined with genetic operators for flexible job-shop scheduling under uncertain processing time for semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing, 31(1), 32-41.

[102] Sloan, T. W., & Shanthikumar, J. G. (2002). Using in-line equipment condition and yield information for maintenance scheduling and dispatching in semiconductor wafer fabs. IIE transactions, 34(2), 191-209.

[103] Terwiesch, C., & Bohn, R. E. (2001). Learning and process improvement during production ramp-up. International journal of production economics, 70(1), 1-19.

[104] Terwiesch, C., & Xu, Y. (2004). The copy-exactly ramp-up strategy: Trading-off learning with process change. IEEE Transactions on Engineering Management, 51(1), 70-84.

[105] Aramon Bajestani, M., Banjevic, D., & Beck, J. C. (2014). Integrated maintenance planning and production scheduling with Markovian deteriorating machine conditions. International Journal of Production Research, 52(24), 7377-7400.
[106] Derman, C. (1970). Finite state Markovian decision processes (No. 04; T57. 83, D47.).
[107] Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). Lectures on stochastic programming: modeling and theory. Society for Industrial and Applied Mathematics.
[108] Birge, J. R., & Louveaux, F. (2011). Introduction to stochastic programming. Springer Science & Business Media.
[109] Ekin, T., Polson, N. G., & Soyer, R. (2014). Augmented Markov chain Monte Carlo simulation for two-stage stochastic programs with recourse. Decision Analysis, 11(4), 250-264.
[110] Homem-de-Mello, T., & Bayraksan, G. (2014). Monte Carlo sampling-based methods for stochastic optimization. Surveys in Operations Research and Management Science, 19(1), 56-85.
[111] Aktekin, T., & Ekin, T. (2016). Stochastic call center staffing with uncertain arrival, service and abandonment rates: A Bayesian perspective. Naval Research Logistics (NRL), 63(6), 460-478.
[112] Ekin, T., Polson, N. G., & Soyer, R. (2017). Augmented nested sampling for stochastic programs with recourse and endogenous uncertainty. Naval Research Logistics (NRL), 64(8), 613-627.
[113] Yano, C. A., & Lee, H. L. (1995). Lot sizing with random yields: A review. Operations Research, 43(2), 311-334.
[114] Christensen, R., Johnson, W., Branscum, A., & Hanson, T. E. (2011). Bayesian ideas and data analysis: an introduction for scientists and statisticians. CRC Press.
[115] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.
[116] Lattimore, T., & Szepesvári, C. (2020). Bandit algorithms. Cambridge University Press.
[117] Celen, M. (2016). Joint maintenance and production operations decision making in flexible manufacturing systems (Doctoral dissertation). Retrieved from University of Texas Libraries (OCLC number: 979556469)
[118] Geng, H. (2005). Semiconductor manufacturing handbook. McGraw-Hill, Inc..
[119] Mack, C. (2008). Fundamental principles of optical lithography: the science of microfabrication. John Wiley & Sons.
[120] Ghaida, R. S., Gupta, M., & Gupta, P. (2013). Framework for exploring the

interaction between design rules and overlay control. Journal of Micro/Nanolithography, MEMS, and MOEMS, 12(3), 033014.

[121] Perloff, D. S. (1978). A four-point electrical measurement technique for characterizing mask superposition errors on semiconductor wafers. IEEE Journal of Solid-State Circuits, 13(4), 436-444.

[122] Van Den Brink, M. A., De Mol, C. G. M., & George, R. A. (1988, January). Matching performance for multiple wafer steppers using an advanced metrology procedure. In Integrated Circuit Metrology, Inspection, and Process Control II (Vol. 921, pp. 180-197). International Society for Optics and Photonics.

[123] Chien, C. F., Chen, Y. J., Hsu, C. Y., & Wang, H. K. (2013). Overlay error compensation using advanced process control with dynamically adjusted proportional-integral R2R controller. IEEE Transactions on Automation Science and Engineering, 11(2), 473-484.

[124] Tan, F., Pan, T., Li, Z., & Chen, S. (2015). Survey on run-to-run control algorithms in high-mix semiconductor manufacturing processes. IEEE Transactions on Industrial Informatics, 11(6), 1435-1444.

[125] Amir, N., Shuall, N., Tarshish-Shapir, I., & Leray, P. (2013, April). Multi layer overlay measurement recent developments. In Metrology, Inspection, and Process Control for Microlithography XXVII (Vol. 8681, p. 86812V). International Society for Optics and Photonics.

[126] Jiao, Y., & Djurdjanovic, D. (2011). Stochastic control of multilayer overlay in lithography processes. IEEE transactions on semiconductor manufacturing, 24(3), 404-417.

[127] Haq, A. U., & Djurdjanovic, D. (2019). Robust Control of Overlay Errors in Photolithography Processes. IEEE Transactions on Semiconductor Manufacturing, 32(3), 320-333.

[128] Levinson, H. J. (1999). Lithography process control (Vol. 28). SPIE Press.

[129] Kuo, H. F., & Faricha, A. (2016). Artificial neural network for diffraction based overlay measurement. IEEE Access, 4, 7479-7486.

[130] Wan, J., & McLoone, S. (2017). Gaussian process regression for virtual metrology-enabled run-to-run control in semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing, 31(1), 12-21.

[131] Anjos, M. F., & Lasserre, J. B. (Eds.). (2011). Handbook on semidefinite, conic and polynomial optimization (Vol. 166). Springer Science & Business Media.

[132] ApS, M. (2019). Mosek optimization toolbox for matlab. User's Guide and Reference Manual, Version, 4.

[133] Tuy, H., Hoang, T., Hoang, T., Mathématicien, V. N., Hoang, T., & Mathematician, V. (1998). Convex analysis and global optimization. Dordrecht: Kluwer.
[134] Zamani, M. (2019). A new algorithm for concave quadratic programming. Journal of Global Optimization, 75(3), 655-681.
[135] Sinha, A., Malo, P., & Deb, K. (2017). A review on bilevel optimization: From classical to evolutionary approaches and applications. IEEE Transactions on Evolutionary Computation, 22(2), 276-295.
[136] Lu, Z., Deb, K., & Sinha, A. (2015, May). Handling decision variable uncertainty in bilevel optimization problems. In 2015 IEEE Congress on Evolutionary Computation (CEC) (pp. 1683-1690). IEEE.
[137] Haupt, R. L., & Ellen Haupt, S. (2004). Practical genetic algorithms.
[138] Beasley, J. E., & Chu, P. C. (1996). A genetic algorithm for the set covering problem. European journal of operational research, 94(2), 392-404.
[139] Tanese, R. (1989). Distributed genetic algorithms for function optimization.
[140] Mathieu, R., Pittard, L., & Anandalingam, G. (1994). Genetic algorithm based approach to bi-level linear programming. RAIRO-Operations Research, 28(1), 1-21.
[141] Bhattacharjya, R. K. (2012). Introduction to genetic algorithms. IIT Guwahati, 12.
[142] Feng, T., Bakshi, S., Gu, Q., Yan, Z., & Chen, D. (2019). Design optimization of bottom-hole assembly to reduce drilling vibration. Journal of Petroleum Science and Engineering, 179, 921-929.
[143] Carr, J. (2014). An introduction to genetic algorithms. Senior Project, 1(40), 7.
[144] Friedman, J. H. (2017). The elements of statistical learning: Data mining, inference, and prediction. springer open.
[145] Le, Q. V., Smola, A. J., & Canu, S. (2005, August). Heteroscedastic Gaussian process regression. In Proceedings of the 22nd international conference on Machine learning (pp. 489-496).
[146] Kim, M. S., Won, H. Y., Jeong, J. M., Böcker, P., Vergaij-Huizer, L., Kupers, M., ... & Suh, J. J. (2016, March). Reduction of wafer-edge overlay errors using advanced correction models, optimized for minimal metrology requirements. In Optical Microlithography XXIX (Vol. 9780, p. 97800A). International Society for Optics and Photonics.
[147] Panaretos, V. M., & Zemel, Y. (2019). Statistical aspects of Wasserstein distances. Annual review of statistics and its application, 6, 405-431.
[148] Dowson, D. C., & Landau, B. V. (1982). The Fréchet distance between multivariate normal distributions. Journal of multivariate analysis, 12(3), 450-455.

Vita

Huidong Zhang was born in Jiangsu, China in 1992. She obtained her B.S. degree in Naval Architecture and Ocean Engineering in 2014 from Shanghai Jiao Tong University, China and her M.Sc. degree in Systems and Project Management in 2015 from Nanyang Technological University, Singapore. In 2015, she joined the University of Texas at Austin as a graduate student in Operations Research and Industrial Engineering (ORIE) and obtained her M.S. degree in ORIE in 2017, after which she has gone on to work on a Ph.D.

Email: huidong.zhang@utexas.edu

This dissertation was typed by Huidong Zhang.