Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

A Sampling Decision System for Semiconductor Manufacturing - Relying on Virtual Metrology and Actual Measurements

Daniel Kurz
Jürgen Pilz

Department of Statistics
University of Klagenfurt
Universitaetsstraße 65-67
9020 Klagenfurt, AUSTRIA

Andrea Schirru
Simone Pampuri

Department of Computer Science Engineering
University of Pavia
Strada Nuova 65
27100 Pavia, ITALY

Cristina De Luca

Infineon Technologies Austria AG
Siemensstraße 2
9500 Villach, AUSTRIA

ABSTRACT

In semiconductor manufacturing, optimization of the sampling measurement plan through production steps is key to maximize productive performances. The measurement plan must guarantee high quality and compliance to wafer specifications limits. In this article, the relationships between virtual metrology (VM) and actual measurements are investigated with respect to a sampling decision system (SDS); specifically, a multilevel VM strategy is relied on to provide predictive information. Such virtual measurements serve as input for the sampling decision system, which in turn suggests the optimal measurement strategy. Two approaches relying on decision-theoretical concepts are discussed: the expected value of measurement information (EVofMI) and a two stage sampling decision model. The basic assumption of the SDS-VM system is that it is not necessary to perform a real measurement until it is strictly needed. The two methodologies are then validated relying on simulation studies and actual chemical vapor deposition (CVD) process and measurement data. The ability of the proposed system to sample dynamically the wafer measurements in dependence of the calculated risk is then evaluated and discussed.

1 INTRODUCTION

During the last 20 years, semiconductor manufacturers have focused on advanced process control topics, including fault detection and classification (FDC) and Run-to-Run controllers. After early deployments on specific processes such as chemical vapor deposition (CVD) (Campbell 1999), (Campbell, Firth, Toprac, and Edgar 2002), lithography, diffusion and etching, the efforts were concentrated in the realization of a fab-wide solution (Moyne 2004). The driving factor, especially in the last ten years, has been the need to improve yield (Tobin and Neiberg 2001) and quality without impacting productive throughput; such topics are even more important for the industries facing high product mix and very rapid product development phases. Some researches focused on sampling techniques: while on one side this would allow to reduce costs by avoiding unnecessary measurement operations, this posed sensitive problems to Run-to-Run controllers who need a constant flow of real measurements to work. Constant challenges include the costs and reliability of equipment integration for data collection, and interaction dynamics with

existing software (such as Run-to-Run controllers, FDC, in-situ-sensors, yield management systems). The theory of a fab-wide solution that goes from APC to yield and integrates components such as scheduling and dispatching represents the next step in the semiconductor industrial roadmaps. In this article, the interactions between virtual metrology (Weber 2007) and sampling decision systems are explored. Virtual (or soft) sensors can be used to calculate a virtual measurement with the goal of reducing measurement costs and the number of time-consuming measurement steps. In the design of virtual sensors, the main goal is to find and exploit the complex, possibly non linear relations between collectible data and real measurements, and assess prediction uncertainty in a meaningful way. In this article, the multilevel virtual metrology approach first described in (Schirru, Pampuri, DeLuca, and Nicolao 2011) is used as information provider. In order to exploit such information to provide optimal sampling and measurement plans, statistical decision theory is jointly employed with multi-stage finite decisions problems that are solved by means of dynamic programming. The proposed approach is validated on chemical vapor deposition (CVD) data.

2 PROCESS FLOW AND VIRTUAL METROLOGY

Virtual metrology methodologies aim to link statistically significant process equipment parameters to measurement results. A VM prediction model should be learned using training data sets

$$\mathscr{S} = \{x_i \in \mathscr{X}, y_i \in \mathscr{Y}, i = 1, \ldots, N\}, \tag{1}$$

where $\mathscr{X}$ and $\mathscr{Y}$ are referred to as the input and output space. To be more concrete, the goal is to find a map $f : \mathscr{X} \to \mathscr{Y}$ that is then extended to new input data $x_{new} \in \mathscr{X}$ in order to predict the unknown measurement result $y_{new} \in \mathscr{Y}$. The obtained prediction $f(x_{new})$ is referred to as the virtual measurement.

2.1 CVD Process Flow

The analysis of the VM-SDS system considers the chemical vapor deposition (CVD) process aiming to form a thin film on the wafer's surface with appropriate purity and uniformity. In other words, *CVD refers to the formation of a non-volatile solid film on a substrate from the reaction of vapor phase chemical reactants containing the right constituents*. Within the reaction chamber employed in this process, the reactant gases are introduced to mix and react with the substrate with the goal of forming the required film. Depending on the needed material, many types of different CVD processes are used in semiconductor industry (low pressure - LPCVD, plasma enhanced -PECVD, high density plasma - HDCVD). A BPSG process (running on a LPCVD equipment) exploits silicate glasses with different additives to improve output reaction and quality; boron and phosphorus are added to the silicate glass. Figure 1 shows the material flow in connection with the VM-SDS system. Lots consisting of 25 wafers are processed. With the end of processing a specific wafer, equipment parameters are collected and sent to the VM-SDS system. After testing data quality, the VM system provides the SDS with a virtual measurement for wafer sampling evaluation. The SDS system takes a decision if a real metrology operation is required or the entire lot is allowed to be skipped and moved to the next production step.

2.2 VM with Multi Kernel Ridge Regression

In this article, a regularized machine learning setting is employed to introduce and test the proposed methodology: the estimator f is found by minimizing some *loss function* $\mathscr{J}(\theta)$ with respect to some unknown parameter θ. Such loss function is usually the sum of a *loss term* $\mathscr{L}$ and a *regularization term* $\mathscr{R}$,

$$\mathscr{J}(\theta) = \mathscr{L}(\theta) + \lambda \mathscr{R}(\theta). \tag{2}$$

In this framework, given a model specified by θ, $\mathscr{L}$ measures the quality of approximation on the training set $\mathscr{S}$ and $\mathscr{R}$ is a measure of the complexity of the model. Intuitively, the coexistence of $\mathscr{L}$ and $\mathscr{R}$ relates to a trade-off between model regularity and performances on $\mathscr{S}$. The *regularization parameter* $\lambda \in \mathbb{R}^+$

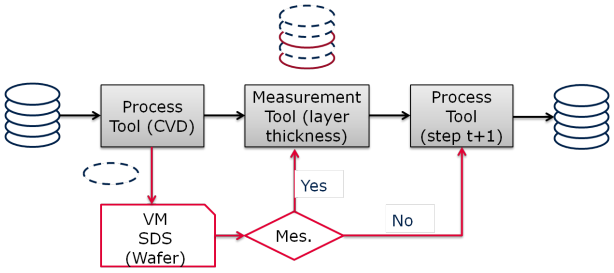

Figure 1: Graphical representation of the material flow and VM-SDS system.

acts as a tuning knob for such trade-off: as λ grows, the order of the selected model gets lower and lower. In this paradigm, a learning algorithm is entirely specified by **(i)** the loss term $\mathscr{L}(\theta)$, **(ii)** the regularization term $\mathscr{R}(\theta)$ and **(iii)** the structure of the estimator $f(x;\theta)$. Remarkably, this structure assumes that the prediction of a generic y_i can be obtained, at best, up to a random uncertainty (depending on $\mathscr{L}$). In other words, adopting an additive error paradigm, it is implied that

$$y_i = f(x_i) + \varepsilon_i, \tag{3}$$

where ε_i is a random variable whose distribution depends on θ. We here assume that $\mathscr{X} \equiv \mathbb{R}^p$ and $\mathscr{Y} \equiv \mathbb{R}$. The goal is to build a map $f : \mathbb{R}^p \to \mathbb{R}$ of the relationship between an input dataset $X \in \mathbb{R}^{N\times p}$ and an array of target observations $Y \in \mathbb{R}^N$. Furthermore, let x_i be the i-th row of X, and y_i be the i-th entry of Y.

The basic machine learning technique, Ordinary Least Squares (OLS), looks for a linear relationship $f(x) = x'w$ by minimizing the following sum of squared residuals with respect to $w \in \mathbb{R}^p$:

$$J_{OLS}(w) := ||Y - Xw||^2 = \sum_{i=1}^{n} (y_i - x_i'w)^2. \tag{4}$$

Under a statistical framework, this amounts to maximizing the conditional probability $p(Y|X)$ when assuming $Y|X \sim N(Xw, \sigma^2 I)$ or, equivalently, $Y = Xw + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2 I)$ (that is, i.i.d. Gaussian noise). The optimal coefficient vector w^{OLS} is just

$$w^{OLS} = (X'X)^{-1}X'Y \tag{5}$$

and does not depend on σ^2. This scheme suffers from two main drawbacks: **(i)** for small datasets ($n \simeq p$) the estimated $f(x)$ may overfit the noisy data or even interpolate the training examples, and **(ii)** the matrix $X'X$ may be ill-conditioned or even singular. To overcome these drawbacks, Ridge Regression (RR), was proposed in the early seventies of the last century. RR is obtained by minimizing the loss function

$$J_{RR}(w) := ||Y - Xw||^2 + \lambda w'w = J_{OLS}(w) + \lambda w'w, \tag{6}$$

where $\lambda \in \mathbb{R}^+$ is the regularization (hyper)parameter. Under a Bayesian framework, J_{RR} is a logposterior distribution, and the term $\lambda w'w$ in (6) is related to the prior distribution $p(w)$ of w, assuming $w \sim N(0, \lambda^{-1} I)$. The larger λ, the smaller the variance of the estimator, at the cost of introducing some bias; in practical applications, λ is often used as a "tuning knob" controlling the bias/variance tradeoff, which is typically tuned either via cross validation. The optimal coefficient vector w^{RR} and the estimator $f_{RR}(x)$ are

$$w^{RR} = (X'X + \lambda I)^{-1}X'Y, \tag{7}$$

$$f_{RR}(x) = x'(X'X + \lambda I)^{-1}X'Y. \tag{8}$$

The numerical stability problems in (5) are now avoided, because $(X'X + \lambda I)$ has full rank for any $\lambda > 0$.

It is obviously desirable to extend the possible estimators $f(x)$ to include also non-linear relationships: instead of considering $y = x'w$ as model, we consider $y = \phi(x)'\overline{w}$, where $\phi : \mathbb{R}^p \to \mathbb{R}^{\overline{p}}$ is a mapping function and $\overline{w} \in \mathbb{R}^{\overline{p}}$ is a new coefficient vector. For example, let $p = 2$ and $\phi(x^{(1)}, x^{(2)}) = [x^{(1)}, x^{(2)}, x^{(1)}x^{(2)}, (x^{(1)})^2, (x^{(2)})^2]$: then, the new estimator includes all possible polynomials up to second degree. Unfortunately, the number $\overline{p}$ of coefficients of a basis expanded to the d-th polynomial order would be

$$\overline{p} = (d-1)p + \sum_{i=1}^{d} \binom{p}{i}. \tag{9}$$

To give some numerical example, for $p = 100$ and $d = 2$, $\overline{p} = 5150$; for $d = 3$, $\overline{p} = 166950$. It is apparent that (7) would be computationally intractable even for relatively low degree polynomials ($\overline{p}$-th order matrix inversion is required). To overcome this, we employ the so-called *kernel trick*: by using the matrix identity

$$(X'X + \lambda I)^{-1}X' = X'(XX' + \lambda I)^{-1}, \tag{10}$$

the optimal RR estimator (8) can be rewritten in *dual form*

$$f_{RR}(x) = x'X'(XX' + \lambda I)^{-1}Y = \langle x, X' \rangle (\langle X', X' \rangle + \lambda I)^{-1}Y, \tag{11}$$

where $\langle u, v \rangle = u'v$ for column vectors u and v of equal size denotes the inner product between u and v. Essentially, equation (11) depends only on inner products, and the Gram matrix $\mathbf{K} = \langle X, X \rangle$ is usually referred to as the *kernel*. This result allows to solve a non-linear learning problem by embedding it in a linear framework. Notably, if there exists a kernel function K such that

$$K(u, v) := \langle \phi(u), \phi(v) \rangle, \tag{12}$$

it is possible to compute $f_{RR}(x)$ without explicitly evaluating $\phi(X)$. Indeed, it is sufficient to compute the Gram matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ such that the element (i, j) is $\mathbf{K}[i, j] = K(x_i, x_j)$ and the vector function $\mathbf{k}(x) \in \mathbb{R}^{1 \times n}$ such that $\mathbf{k}(x) = (K(x, x_1), ..., K(x, x_n))^T$. Then, (11) is rewritten as

$$f_{RR}(x) = \mathbf{k}(x)'(\mathbf{K} + \lambda I)^{-1}Y = \mathbf{k}(x)'c^*, \tag{13}$$

where c^* is the solution of the following *Kernel Ridge Regression Problem*:

$$c^* = \arg\min_c \; ||Y - \mathbf{K}c||^2 + \lambda c'\mathbf{K}c. \tag{14}$$

This is an unconstrained quadratic optimization problem with respect to c, the minimizer is

$$c^* = (\mathbf{K} + \lambda I)^{-1}Y. \tag{15}$$

Notable examples of kernel functions are the polynomial (homogeneous and inhomogeneous) kernels of degree d, i.e. $K(u, v) = (\langle u, v \rangle)^d$ and $K(u, v) = (1 + \langle u, v \rangle)^d$, and and the radial basis kernel represented by the kernel function $K(u, v) = exp(-||u - v||^2/2\sigma^2)$.

For an extension to multi-level virtual metrology, see (Schirru, Pampuri, DeLuca, and Nicolao 2011).

3 SAMPLING DECISION SYSTEM USING VM

Sampling plans for metrology operations are key in semiconductor manufacturing. For reasons of destructive tests, intractably high costs, technologically unfeasible inspections and excellent process quality history, there is no need of inspecting the whole production. Concepts of statistical process control (SPC) provide methods for acceptance sampling that have been followed by semiconductor companies for years (Montgomery 2001). Real metrology information is obtained with certain time delay after wafer processing which often

leads to a delayed recognition of process abnormalities. By means of the VM-SDS system, we are able to overcome these deficiencies. Whenever it is possible, the virtual measurement should be used instead of the real one in order to decide on the status of some wafer or process. Reliability of virtual measurements is assessed by means of the equipment health factor (EHF). The VM trust factor adjusts virtual measurements uncertainty comparing real measurement outcomes with corresponding virtual ones. Based on the evaluation of in control (IC) and out of control (OOC) wafer risks, the VM-SDS system results in a dynamic sampling plan fixing wafer fine metrology operations.

3.1 Equipment Health Factor

Virtual measurements are computed with respect to a certain selection of significant equipment parameters. However, process outputs depend on several equipment variables. Thus, the reliability of obtained equipment data has to be assessed. This can be reached by means of the equipment health factor (EHF), which is the first important input of the VM-SDS system. Within the ENIAC IMPROVE (Implementing Manufacturing science solutions to increase equiPment pROductiVity and fab pErformance) project, the semiconductor community defines it as *the result of a probability calculation used to forecast equipment behaviour and optimize control plan, based on the analysis of equipment parameters, maintenance and product historical data*. In this article, it is a quantitative equipment status index. A principal component analysis (PCA) of all equipment parameter streams filters the insignificant information. Feature space dimension is reduced so that 95% of total variability is retained. Based on the PCA filtered data streams, the status of the equipment is assessed by means of Hotelling's T^2. An EHF alarm signal should be an indicator of unreliable predictions and should therefore lead to real measurements. Since the EHF will not always prevent the VM-SDS system from using unreliable predictions, the VM trust factor t_{VM} should be considered in addition.

3.2 Trust factors for virtual measurements

A virtual measurement provides a predictive probability distribution indicating the more probable measurement outcomes of some wafer k. In the univariate case, the virtual measurement defines a normal prior for the unknown mean value of $p = 9$ wafer sites

$$\mu_k \sim N\left(\mu_{VM_k}, \sigma^2_{VM_k}\right). \tag{16}$$

Multivariate virtual measurements provide expected values of all the $p = 9$ points on the wafer, summarized by means of a VM mean vector $\mu_{VM_k} \in \mathbb{R}^p$ and a corresponding covariance matrix $\Sigma_{VM_k} \in \mathbb{R}^{p\times p}$. It can be seen as p-variate normal distribution covering the more probable measurement vectors $\mathbf{X}_k \in \mathbb{R}^p$.

In order to guarantee reliability of virtual measurements, prediction precision should be updated whenever real measurements become available. By means of a VM trust factor t_{VM}, uncertainty of virtual measurements is adjusted without changing covariance structure. Initially, we have $t_{VM} = 1$. On the one hand, an update of the VM trust factor is performed whenever a real measurement $\mathbf{x_k}$ has a significant squared Mahalanobis distance with respect to the the virtual one; that is if

$$\mathbf{d}_k^T \Sigma_{VM_k}^{-1} \mathbf{d}_k > \chi^2_{p,1-\alpha}, \tag{17}$$

where $\mathbf{d}_k := \mathbf{x}_k - \mu_{VM_k}$. The idea is then to minimally enlarge the determinant of VM covariance matrix Σ_{VM} by means of t_{VM} such that the real measurement is again contained in the VM confidence region. So,

$$t_{VM} = \min\left\{\arg\min_{t\geq 1}\left\{\mathbf{d}_k^T \left(t\Sigma_{VM_k}\right)^{-1}\mathbf{d}_k \leq \chi^2_{p,1-\alpha}\right\}, t^*_{VM}\right\}. \tag{18}$$

A consideration of t^*_{VM} should avoid implausible covariance matrices. In case of p-values (based on the distribution of the squared Mahalanobis distance) within $[\alpha, 1-\alpha]$, VM predictions are then weighted

using the current value of t_{VM}. A p-value greater than $1-\alpha$ indicates t_{VM} to be too large. In this case the VM trust factor should be decreased according to

$$t_{VM}^{(l+1)} = \max\left\{1, \arg\max_{t \le t_{VM}^{(l)}} \left\{ \mathbf{d}_k^T \left(t\Sigma_{VM_l}\right)^{-1} \mathbf{d}_k \ge \chi^2_{p,\alpha} \right\} \right\}. \tag{19}$$

If there is a significant increase of the VM trust factor, the virtual measurement seems to be not reliable any more and real measurements are required. In this way, also mismeasurements might be detected. An example is given in Figure 2, where wafer #17 shows a bad virtual measurements. Figure 3 indicates that the trust factor reacts properly jumping to a higher value. We are more uncertain about virtual measurements as long as the trust factor returns to a normal stage. For more details, see (Kurz, DeLuca, and Pilz 2013).

4 TWO APPROACHES FOR SDS

The proposed sampling decision system is based on the assumption that it is not necessary to perform a real measurement until it is strictly needed. The real measurement leads to an additional quality information about the lot/wafer/process equipment. In order to reduce uncertainty and risk of faults, sometimes real measurements are required. Two similar decision methodologies have been introduced in (Kurz, DeLuca, and Pilz 2012): the first is based on the expected value of measurement information (EVofMI) and the second one concerns running a two-stage sampling decision model. Both decision approaches, are discussed for the mean value of the $p = 9$ wafer points.

4.1 The expected value of measurement information (EVofMI)

In semiconductor manufacturing, it is a standard practice to control process outputs by means of (univariate) control charts (Montgomery 2001). These charts are most often based on the assumption of normally distributed process observations. Since we focus on $\overline{X}$-charts, we assume

$$X_{ki}|\mu_k \sim N(\mu_k, \sigma^2), \tag{20}$$

where σ^2 provides information on process variation. It might be either estimated from data or derived by means of process capability indices C_p or C_{pk}. Assumption (20) is exchangeable. The sample average

$$\bar{x}_k = \frac{1}{p}\sum_{i=1}^{p} x_{kp} \tag{21}$$

is considered in control (IC) if $\bar{x}_k \in [LCL, UCL]$ or out of control (OOC) if $\bar{x}_k \notin [LCL, UCL]$. Then the sampling decision system, requires to consider IC-OOC-decision making in a decision-theoretical framework.

First of all, we specify actions $a_i \in \mathscr{A}$, $i \in \{0,1\}$, where

$$a_0 = \text{"accept IC"}, \qquad a_1 = \text{"accept OOC"} \tag{22}$$

and states of nature $\theta_j \in \Theta$, $j \in \{0,1\}$, using

$$\theta_0 = IC \Leftrightarrow \mu_k \in [LCL, UCL], \qquad \theta_1 = OOC \Leftrightarrow \mu_k \notin [LCL, UCL]. \tag{23}$$

Consequences of wrongly classifying a wafer to be IC and OOC, respectively, are quantified by means of a loss function $L : \Theta \times \mathscr{A} \to \mathbb{R}$ defined as

$$L(\theta, a_0) = \begin{cases} 0 & \theta = \theta_0 \\ l_{01} & \theta = \theta_1 \end{cases}, \qquad L(\theta, a_1) = \begin{cases} l_{10} & \theta = \theta_0 \\ 0 & \theta = \theta_1 \end{cases} \tag{24}$$

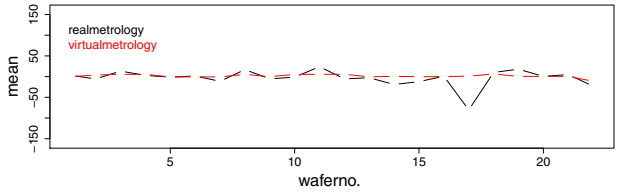

Figure 2: Real and virtual thickness measurements - normalized mean values from CVD process.

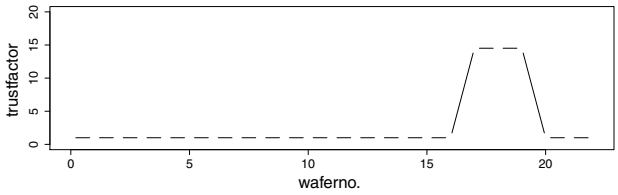

Figure 3: Computed trust factors for CVD thickness measurements.

and $l_{01}, l_{10} > 0$. The optimal prior decision is then found by means of prior expected loss. Therefore, we first compute prior probabilities for the k-th wafer of being IC and OOC using the trust factor weighted virtual measurement, see Section 3.2. Thus,

$$P(\theta = IC) = P(\mu_k \in [LCL, UCL]), \qquad P(\theta = OOC) = 1 - P(\theta = IC). \tag{25}$$

Afterwards, we determine the optimal prior action $\tilde{a}$ by calculating prior expected losses (prior risks) of actions a_0 and a_1, i.e.

$$E_\theta L(\theta, a_0) = l_{01} P(\theta = OOC), \qquad E_\theta L(\theta, a_1) = l_{10} P(\theta = IC). \tag{26}$$

The action with the smaller prior risk is chosen as the optimal prior one. The risk taken from this decision (prior decision value) is (Parmigiani and Inoue 2009)

$$V_\theta = \min(E_\theta L(\theta, a_0), E_\theta L(\theta, a_1)). \tag{27}$$

In case of a precise virtual measurement predicting a mean near the target, we obtain a high probability for the wafer of being IC. This results in a prior decision in favor of a_0 with decision risk $V_\theta \approx 0$.

If a sample mean $\bar{x}_k$ of p measured wafer sites becomes available, the posterior distribution of $\mu_k|\bar{x}_k$ can be assessed applying

$$f(\mu_k|\bar{x}_k) = f(\bar{x}_k|\mu_k) f(\mu_k) / f(\bar{x}_k). \tag{28}$$

In our case, the posterior distribution is given as $\mu_k|\bar{x}_k \sim N(\mu_k^*, \sigma_k^{2*})$ with (Bernardo and Smith 2000)

$$\mu_k^* = \sigma_k^{2*} \left(\frac{1}{\sigma^2/p} \bar{x}_k + \frac{1}{\sigma_{VM_k}^2} \mu_{VM_k} \right), \qquad \frac{1}{\sigma_k^{2*}} = \frac{1}{\sigma^2/p} + \frac{1}{\sigma_{VM_k}^2}. \tag{29}$$

The optimal action after observing $\bar{x}_k$ is then derived by means of minimum posterior expected loss (DeGroot 1970). Therefore, we compute posterior probabilities for the wafer of being IC and OOC according to

$$P(\theta = IC|\bar{x}_k) = P(\mu_k|\bar{x}_k \in [LCL, UCL]), \qquad P(\theta = OOC|\bar{x}_k) = 1 - P(\theta = IC|\bar{x}_k) \tag{30}$$

and assess posterior expected losses of actions a_0 and a_1

$$E_{\theta|\bar{x}_k}L(\theta,a_0) = l_{01}P(\theta = OOC|\bar{x}_k), \qquad E_{\theta|\bar{x}_k}L(\theta,a_1) = l_{10}P(\theta = IC|\bar{x}_k). \tag{31}$$

The Bayes action a^* has minimum posterior expected loss and minimum Bayes risk, respectively. The risk that is taken with respect to a^* (posterior decision value) is

$$V_{\theta|\bar{x}_k} = \min(E_{\theta|\bar{x}_k}L(\theta,a_0), E_{\theta|\bar{x}_k}L(\theta,a_1)). \tag{32}$$

The information gained from observing $\bar{x}_k$ is assessed by means of the value of measurement information $V_{\bar{x}_k}$ (Parmigiani and Inoue 2009). It is defined as the difference between the posterior decision value $V_{\theta|\bar{x}_k}$ and the posterior expected loss of the optimal prior action $\tilde{a}$. Hence,

$$V_{\bar{x}_k} = V_{\theta|\bar{x}_k} - E_{\theta|\bar{x}_k}L(\theta,\tilde{a}) \geq 0. \tag{33}$$

The value of measurement information is 0 in case of an equal prior and posterior decision. In this case, the measurement information would actually not have been needed. However, we receive a high informational gain from the measurement if $\bar{x}_k$ revises our decision with respect to the virtual measurement and results in a small posterior decision risk.

The marginal distribution of $\bar{x}_k$ (prior predictive distribution) (Bernardo and Smith 2000),

$$\overline{X}_k \sim N\left(\mu_{VM_k}, {}^{\sigma^2}/_{p} + \sigma^2_{VM_k}\right) \tag{34}$$

allows to compute the expected value of measurement information (EVofMI) of real measurement M_k (Parmigiani and Inoue 2009):

$$V_{M_k} = E_{\overline{X}_k}V_{\bar{x}_k} = E_{\overline{X}_k}V_{\theta|\overline{X}_k} - V_\theta \geq 0. \tag{35}$$

It is the difference between preposterior decision value $E_{\overline{X}_k}V_{\theta|\overline{X}_k}$ and prior decision value V_θ. Setting a constant measurement loss factor $l_M \geq 0$ in relation to l_{01} and l_{10}, we receive the following sampling rule:

$$\text{"measure"} \Leftrightarrow V_{M_k} > l_M. \tag{36}$$

4.2 The two-stage sampling decision model

The second approach in order to obtain optimal sampling strategies is based on a two-stage sampling decision model. It can be viewed as a generalization of the EVofMI-approach. Within the first stage s_0, a sampling decision a_0 has to be made. Therefore, we specify first stage actions

$$a_{00} = \text{"measurement"}, \qquad a_{01} = \text{"no measurement"}. \tag{37}$$

In the second stage s_1, a decision a_1 with regard to the status of some wafer k in the control chart after processing has to be taken. Possible second stage actions are a_{10} and a_{11} as stated in (22). Wafer status is again treated as discrete random variable with states from (23). If $a_0 = a_{00}$, a sample mean $\bar{x}_k \in \mathbb{R}$ is collected. Since the outcome of the measurement might influence our belief in the state of the wafer, the second stage decision problem depends on the action taken at stage s_0. Since we are not just focused on making a single decision, we now have to evaluate several possible decision policies $p_{ij} = (a_{0i}, a_{1j}) \in \mathscr{P}$ with $i,j \in \{0,1\}$ using a loss function $L : \Theta \times \mathscr{P} \to \mathbb{R}$. Losses $l_{ij,k} = L(\theta_k, p_{ij})$, $i,j \in \{0,1\}$, $k \in \{0,1\}$ are summarized by means of a loss matrix $L = \left(l_{ij,k}\right) \in \mathbb{R}^{4\times 2}$ which looks as follows:

$$L = \begin{pmatrix} l_M & l_{01}+l_M \\ l_{10}+l_M & l_M \\ 0 & l_{01} \\ l_{10} & 0 \end{pmatrix}. \tag{38}$$

The sampling decision problem is visualized using a two-stage decision tree, see Figure 4.

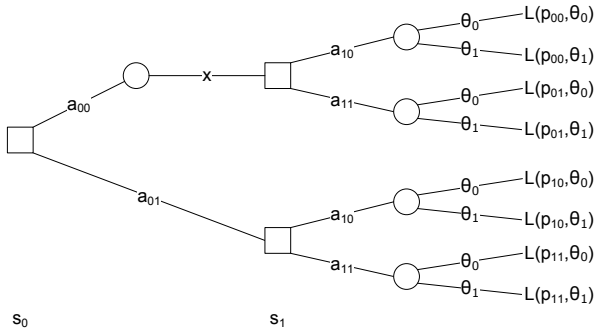

Figure 4: Decision tree of sampling decision model.

The generic multistage decision model is solved by means of backwards induction; it says that a multistage decision model has to be solved from the future back into the past (Bellman 1957). In our case, the backwards induction algorithm results in a comparison of prior decision value V_θ and preposterior decision value $E_{\overline{X}_k} V_{\theta|\overline{X}_k}$ which is now computed with respect to the loss matrix in (38). Finally, we obtain the following sampling rule:

$$\text{"measure"} \Leftrightarrow E_{\overline{X}_k} V_{\theta|\overline{X}_k} < V_\theta. \tag{39}$$

In case of $l_M = 0$, the sampling rule based on the two-stage decision model is equivalent to the sampling rule in (36). Otherwise $l_M > 0$ is directly taken into account when computing posterior decision values. In case of a posterior risk that is expected to compensate l_M in comparison to the prior risk, it is decided to take a measurement. The stronger we penalize wrong decisions, the more often a real wafer measurement is triggered.

5 EXPERIMENTAL RESULTS

Let us now test the two approaches for the SDS by extending them to really obtained and simulated virtual measurements. The considered process currently shows standardized control limits UCL$= 33.5$, LCL$= -33.5$ and $C_{pk} = 1.5$. Figure 5 shows VM expected values and real metrology means. Current VM prediction variance is $\sigma^2_{VM} = 119.52$. Health factors do not indicate bad VM prediction quality. The SDS performs optimally if it allows skipping a better number of wafer measurements within stable process phases, and suggests a real measurement operation whenever a virtual measurement indicates an abnormal process output.

By default, we use unit loss and measurement loss factor $l_M = 0.05$. Figure 6 shows results of computing expected values of measurement information. Based on the virtual measurement of wafer #17, we are not able to decide if the wafer is IC or OOC. Hence, a real measurement is required. Apart from that, other wafer predictions indicate process stability; the EVofMI-approach suggests to skip corresponding metrology operations.

A comparison of prior and preposterior decision values based on the two-stage sampling decision model is done within Figure 7. Corresponding decision strategies are summarized in Table 1. Both approaches

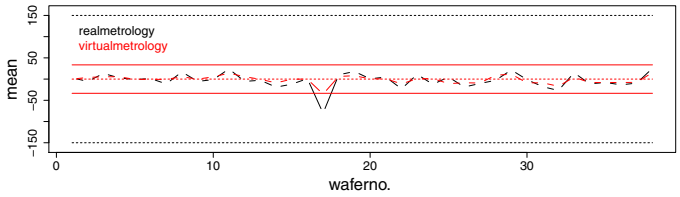

Figure 5: Virtual and real CVD thickness mean values.

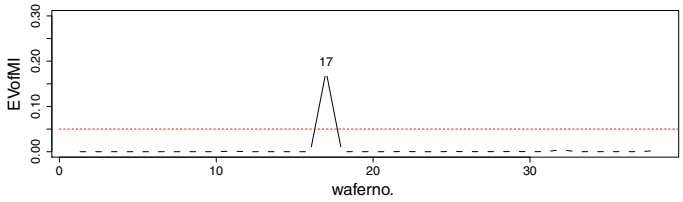

Figure 6: EVofMI for virtual measurements of Figure 5.

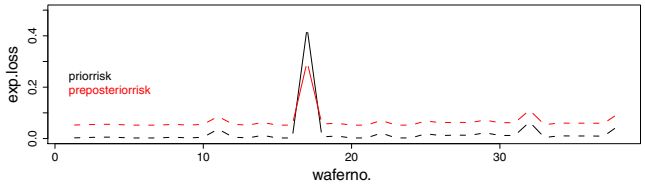

Figure 7: Prior and preposterior decision risks for virtual measurements of Figure 5.

wafer	sampling decision	prior wafer status
1	no measurement	accept IC
...	...	...
16	no measurement	accept IC
17	measurement	-
18	no measurement	accept IC
...	...	...
38	no measurement	accept IC

Table 1: Optimal decision strategies for virtual measurements of Figure 5.

provide equal results. However, the real measurement of wafer #17 (observed with some delay) shows a bad VM prediction. The updated VM trust factor might affect future decisions.

Let us now test performance of the two SDS approaches with respect to a simulated process drift as shown in Figure 8 that is captured within the virtual measurements. Expected values of measurement information and decision values based on the two-stage decision model are plotted in Figures 9 and 10, accordingly. Additionally, Table 2 provides optimal decision strategies. One can realize that both approaches react with the simulated process drift and suggest real measurements of wafers #49 and #50. The process might be OOC.

6 CONCLUSIONS

In this article, a first study of interaction between Virtual Metrology (VM) and sampling decision systems (SDS) has been presented. Considering virtual measurements within a $\overline{X}$-chart, the proposed SDS approaches are able to provide an optimal sampling action in dependence of the information coming from the VM system. In other words, depending on whether the virtual measurement makes it possible to decide on the status of the wafer, a real measurement operation is suggested. If some process drifts are observable, the sampling decision system will (with a time constant that depends on preset SDS parameters) request real wafer measurements. A large number of real wafer measurements might be skipped in case of reliable virtual measurements indicating good process performance. In conclusion, we show how a sampling decision

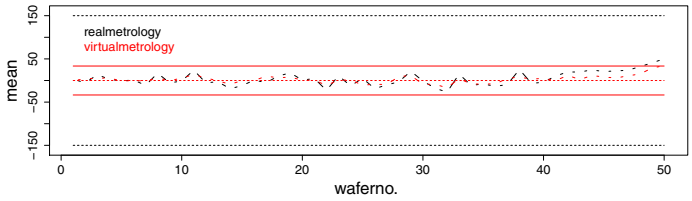

Figure 8: Data with simulated process drift.

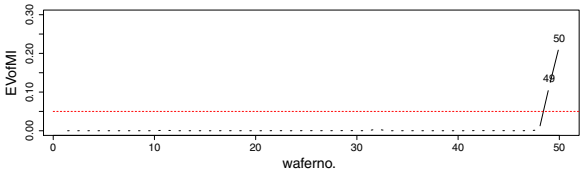

Figure 9: EVofMI for virtual measurements of Figure 8.

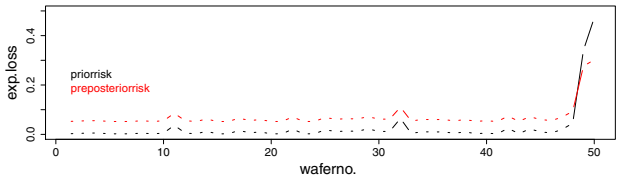

Figure 10: Prior and preposterior decision values for virtual measurements of Figure 8.

system using Virtual Metrology allows to improve performances based on usual statical sampling rules. Of course, a small amount of real measurements will always have to be performed, in order to continuously validate the suggestions of the proposed SDS system.

REFERENCES

Bellman, R. 1957. *Dynamic Programming*. Princeton, NJ: Princeton University Press.

Bernardo, J. M., and A. F. M. Smith. 2000. *Bayesian Theory*. Chichester: John Wiley and Sons Ltd.

Campbell, W. 1999. *Model predictive run-to-run control of chemical mechanical planarization*. Ph.D. thesis, University of Texas at Austin.

Campbell, W., S. Firth, A. Toprac, and T. Edgar. 2002. "A comparison of run-to-run control algorithms". In *IEEE Proceedings of the American Control Conference*, Volume 3, 2150–2155.

Caruana, R. 1997. "Multitask learning". *Machine Learning* 28 (1): 41–75.

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill, Inc.

Kurz, D., C. DeLuca, and J. Pilz. 2012. "Sampling decision system in semiconductor manufacturing using virtual metrology". In *IEEE International Conference on Automation Science and Engineering*, 74–79.

Kurz, D., C. DeLuca, and J. Pilz. 2013. "Monitoring Virtual Metrology Reliability in a Sampling Decision System". In *IEEE International Conference on Automation Science and Engineering*, 20–25.

Montgomery, D. C. 2001. *Introduction to Statistical Process Control*. NY: John Wiley and Sons, Inc.

Moyne, J. 2004. "The Evolution of APC: The Move to Total Factory Control". *Solid State Technology* 47 (9): 47–54.

Parmigiani, G., and L. Inoue. 2009. *Decision Theory: Principles and Approaches*. Chichester: John Wiley and Sons Ltd.

wafer	sampling decision	prior wafer status
1	no measurement	accept IC
2	no measurement	accept IC
...	...	...
48	no measurement	accept IC
49	measurement	-
50	measurement	-

Table 2: Optimal decision strategies for virtual measurements of Figure 8.

Schirru, A., S. Pampuri, C. DeLuca, and G. D. Nicolao. 2011. "Multilevel Kernel Methods for Virtual Metrology in Semiconductor Manufacturing". In *Proc. of the 18th IFAC World Congress*, 11614–11621.

Tobin, K., and L. Neiberg. 2001. "Metrology data management and information systems". In *Handbook of Silicon Semiconductor Metrology*, 537–557.

Weber, A. 2007. "Virtual metrology and your technology watch list: ten things you should know about this emerging technology". *Future Fab International* 22:52–54.

AUTHOR BIOGRAPHIES

DANIEL KURZ received a master degree in technical mathematics with main focus on statistics at the AAU Klagenfurt/Austria in 2012. He has been involved in research topics from the semiconductor industry since 2010 and is currently working toward the PhD degree in statistics at the AAU Klagenfurt/Austria. His email address is daniel.kurz@aau.at.

SIMONE PAMPURI received a PhD degree in computer science engineering at the University of Pavia/Italy in 2012. He has authored several papers in the areas of Virtual Metrology and Advanced Process Control. He is currently working as a Post Doc at the National University of Ireland, Maynooth/Ireland. His email address is simone.pampuri@unipv.it.

ANDREA SCHIRRU received a PhD degree in computer science engineering at the University of Pavia/Italy in 2011. He has authored several publications in the areas of Virtual Metrology and predictive maintenance. He is currently working as a Post Doc researcher in the semiconductor industry. His email address is andrea.schirru@unipv.it.

CRISTINA DE LUCA received a PhD degree in statistics at the Alpen-Adria-University (AAU) of Klagenfurt/Austria in 2003. She is currently working as an Innovation Project Manager at Infineon Technologies Austria AG, Villach/Austria. Her email address is cristina.deluca@infineon.com.

JÜRGEN PILZ is Professor of Applied Statistics at the Alpen-Adria-University (AAU) of Klagenfurt/Austria. He has authored more than 100 publications in international journals and conference proceedings in the areas of Bayesian statistics, spatial statistics, environmental and industrial statistics, statistical quality control and design of experiments. He is the author of six books which appeared in internationally renowned publishing houses such as Wiley, Springer and Chapman and Hall, and authored several book chapters. He obtained his M.Sc., PhD and D.Sc. degrees from Freiberg Technical University in Germany in 1974, 1978 and 1988, respectively, in the areas of mathematics and statistics. Since 2007 he serves as head of the department of Applied Statistics at AAU Klagenfurt, Austria. He is elected member of the International Statistical Institute (ISI) and the Institute of Mathematical Statistics (IMS). He is on editorial boards of several international statistics journals. His email address is juergen.pilz@aau.at.