

Effect of Periodic Virtual Metrology Recalibration on Blended Metrology Schemes

Jae Yeon (Claire) Baek

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2012-267

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-267.html>

December 17, 2012



Copyright © 2012, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**Effect of Periodic Virtual Metrology Recalibration on Blended Metrology
Schemes**

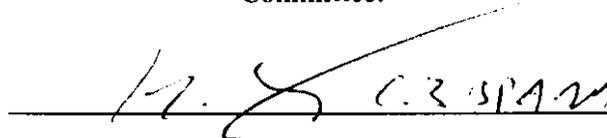
by Jae Yeon Baek

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II.**

Approval for the Report and Comprehensive Examination:

Committee:



Professor Costas J. Spanos
Research Advisor

12/17/2012

Date



Professor Andrew R. Neureuther
Second Reader

December 11, 2012

Date

Effect of Periodic Virtual Metrology Recalibration on the Performance of Blended Metrology Schemes

Jae Yeon Baek
University of California, Berkeley
Electrical Engineering and Computer Sciences

Contents

1	Introduction	1
1.1	Advanced Process Control and Virtual Metrology	1
1.2	Blended Metrology	3
1.3	Higher Level Perspective: Cost Analysis	4
2	Design of Cost Analysis	5
2.1	Defining Total Profit	5
2.2	Costs Associated with Imperfect VM Predictions	5
3	Part I: Preliminary Simulations	8
3.1	Sampling Scenario	8
3.2	Simulation Setup	8
3.3	Results	10
3.4	Use of VM in a Faulty Process and the Dependence of VM R^2 on SEM Proportion	12
4	Part II: Blended Metrology Optimization Using Virtual Metrology Recalibration	14
4.1	Drifting Regression Model	14
4.2	Virtual Metrology Models	15
4.2.1	Ordinary Least Squares	15
4.2.2	Exponentially-weighted Linear Regression	16
4.2.3	Kalman Filter	16
4.3	Results	19
4.3.1	Ordinary Least Squares	19
4.3.2	Exponentially-weighted Linear Regression	24
4.3.3	Kalman Filter	24
4.4	Discussion	25
5	Conclusion and Outlook	27

Abstract

Risk and cost must be balanced in the design of semiconductor processing metrology. More specifically, one needs to balance the cost of operating the metrology tool, and the loss in terms of processing cost and yield due to the limited sampling and the time lapse between the occurrence and the correction of a process fault. In virtual metrology (VM), the real-time data produced by the processing tool (e.g. plasma etching data during isolation trench formation) is used to predict an outcome of the wafer (e.g. critical dimension of the trench) utilizing an empirical model. Although VM prediction quality is not as good as that of conventional metrology, it produces an immediate, low cost prediction for each wafer going through a process. We envision that practical metrology schemes in the future will involve a synergistic blend of VM and actual metrology, the latter being used for the needed periodic recalibration of the VM empirical model. In this work, we have formulated the costs associated with type I and type II errors that result from a blended metrology scheme; the revenue, processing cost, and off-line metrology cost.

In the first part of the work, the net profit as a function of proportion of samples that go through VM prediction and the prediction quality of the VM model was plotted for a given off-line metrology cost. Results showed that the prediction quality of the VM model could be relaxed and still be beneficial in the presence of process faults. The second part expands on this result by taking into account the relationship between the proportion of samples going through VM and the quality of the prediction model. This is important since as more samples from off-line metrology are used for VM recalibration, the prediction quality of the VM model is improved. However, the cost of off-line metrology would also increase. This paper formulates and explores this tradeoff between re-calibration and off-line metrology to find the optimal number of samples that maximizes the profit. A sequence of metrology samples using a regression model with linearly drifting coefficients is simulated, a model realistically applying to a manufacturing process with linearly drifting hidden variables. Three different types of statistical models, ordinary least-squares (OLS), exponentially-weighted ordinary least-squares (WOLS), and the Kalman Filter are used as VM prediction tools. To simulate a blended metrology scheme, we alternate between training sets (VM calibration using off-line metrology), and testing sets (prediction through VM

model), and compare the resulting net profit, type I, and type II errors as a function of varying VM prediction sample sizes. In this work, the blended metrology scheme involved a periodic pattern starting with 20 actual metrology samples used as the re-calibration set, followed by a variable number of VM-predicted samples. Results show that each VM prediction model has a different tradeoff between the Type I and Type II errors that determine the optimal sampling scheme. The ultimate goal is to create a general framework that quickly leads to the optimal design of such schemes given the characteristics of the process in question.

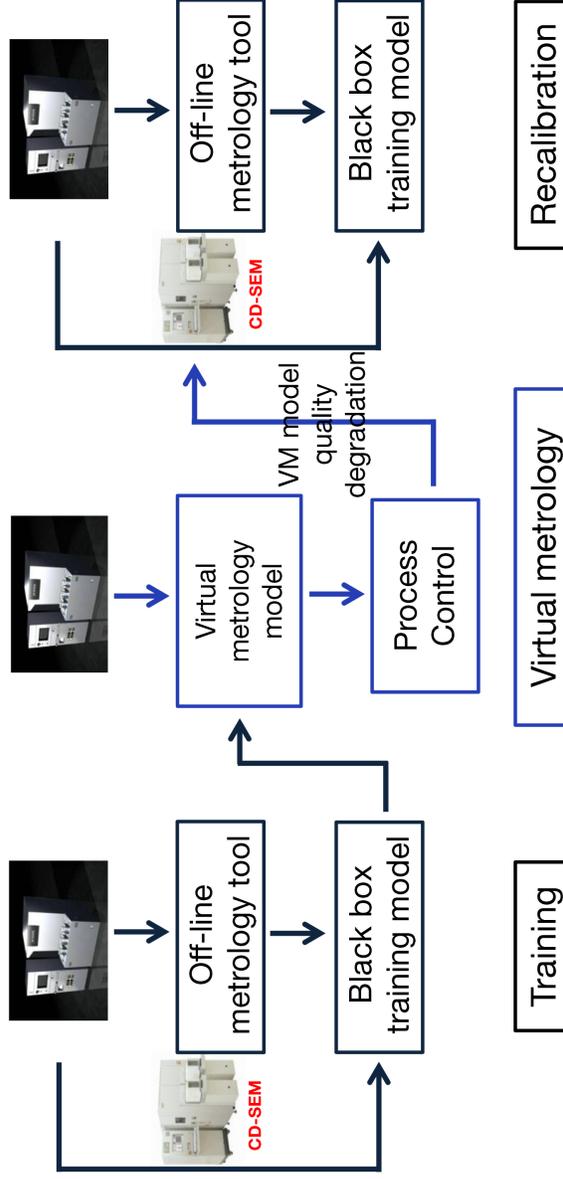
1 Introduction

1.1 Advanced Process Control and Virtual Metrology

An increase in complex manufacturing processes and reduced device dimensions has demanded even tighter quality control in semiconductor manufacturing. A factory-wide control strategy is required to pursue an optimal objective for factory-wide metrics such as yield, cycle time, fabrication cost, and electrical characteristics of product. [1] Recently, there has been a shift from statistically based techniques (SPC) [2] using sampled measurements to advanced process control (APC) techniques such as run-to-run control, fault detection and fault analysis, [3] R2R control already being widely implemented in fabs [4], [5]. A majority of these systems are lot-to-lot (L2L), where 2-3 wafers from a lot are sampled and examined as representations of the whole lot. Engineers have determined that this is even not enough and the need for monitoring every wafer-to-wafer (W2W) is imminent. [4] However a critical necessity for all these techniques and successful APC implementation is the supply of accurate wafer metrology data. [3]

Metrology in semiconductor processing allows us to measure critical process parameters and modify the process for tighter quality control. Moreover, manufacturing processes suffer from drift [6] and metrology is used to recalibrate the process recipe. Although all process parameters are important for effective product manufacturing, some critical dimensions (CDs) are determined to be the most important. Dimensions that are deemed to be critical typically encompass those that define the electrical performance of transistors or interconnect. Currently, the main workhorse for measuring the length of transistor gates has been the scanning electron microscope (SEM). SEM measurements are collected off-line using dedicated, throughput-limited equipment. Although this kind of metrology produces relatively accurate measurements, it is also costly and time-consuming. It is a L2L control tool as only a few samples from a lot can go through the SEM due to its cost and time of use. In addition, an increase in process cycle time occurs because of the time gap from sending the test wafer to the metrology equipment to getting the results. [7] The fab has to incur a significant cost due to these various disadvantages of conventional SEM metrology. [1]

Virtual metrology (VM) is an alternative scheme to conventional metrol-



Images courtesy of Hitachi

Figure 1: A blended metrology scenario: the VM model is constructed through the initial training step, then applied to the process, and recalibrated later to incorporate process drift.

ogy, which takes the processing data produced by the processing tool in real time, (plasma etching data during isolation trench formation, for example) and predicts an outcome of the wafer (e.g. critical dimension of the trench) utilizing an empirical model. Some alternative definitions are the prediction of metrology variables (either measurable or non-measurable) using information about the state of the process [8], and the correlation between process tool and final wafer results [4]. Implementation of a good virtual metrology model enables W2W real-time quality control and reduces the cycle time. [9] In addition, a decrease in the number of test wafers needed results in lower cost.

Various authors have developed VM models to predict wafer etch rate using high-dimensional sensor data. [10], [11], [12] Su *et al.* and Hung *et al.* developed VM models for CVD film thicknesses using neural networks. [8], [13] The performance of the VM model is usually measured against conventional metrology data by criteria such as the coefficient of determination R^2 or the Mean Squared Error (MSE). What really matters, however, is whether the introduction of VM actually improves the overall performance of the fabrication sequence. The goal of our work is to explore this question.

1.2 Blended Metrology

Although VM has its advantages, VM model prediction quality is not as good as that of conventional metrology and the models need frequent recalibration in order to maintain acceptable predictive capability. Cheng *et al.* have developed a reliance index as a measure of the reliability of VM. [14] In real life, we envision that practical metrology schemes will involve VM in combination with conventional or actual metrology, the latter being used for the needed periodic recalibration of the VM empirical model. Such a scenario is shown in Figure 1. The initial training step is used to construct the VM model; the model is then applied to incoming wafers and later recalibrated to incorporate process drift and other faults. We also envision an algorithm that would respond to a fault being predicted by the VM model by possibly requiring additional actual measurements.

1.3 Higher Level Perspective: Cost Analysis

Although generating more accurate empirical models is a critical aspect of implementing virtual metrology, not many have focused on the problem of analyzing the cost of using virtual metrology vs. actual metrology. Yung-Cheng *et al.* have proposed a model for calculating the fab's profitability of using VM [7]. However, this does not include a term for the quality of the VM model. As mentioned above, we think that virtual metrology alone will not suffice as a stand-alone metrology tool due to its limited predictive capability, but could be profitable for the fab with the use of conventional metrology as a recalibration tool.

Any metrology scheme, whether conventional, virtual or blended, may produce false alarms (Type I error) or may fail to detect a fault (Type II error). Those two types of errors are associated with operational costs that depend on the metrology scheme and the algorithm used in response to an alarm. In this work, we formulate the costs associated with Type I and Type II errors that result from a blended metrology scheme, and we propose a general framework that can be used to quickly lead to the optimal design of such schemes given the characteristics of the process in question. This is done by exploring the effects of variables such as the frequency of samples that go through actual metrology, the prediction quality of the VM model, the cost of missed or false alarms, processing and actual metrology costs etc. To demonstrate this methodology, we calculate the total profit for three different but realistic VM models: Ordinary Least Squares (OLS), Exponentially-weighted Linear Regression (EWLR), and the Kalman Filter. The purpose of this work is to analyze the conditions under which such a blended metrology scheme would be advantageous, and to provide rules that will let us optimize such blended metrology schemes.

Chapter 2 explains the setup for blended metrology cost analysis. From hereon, our analysis assumes conventional metrology to be the SEM and the unknown quantities to be the critical dimension (CD).

2 Design of Cost Analysis

2.1 Defining Total Profit

In any metrology operation, the value obtained is an estimate of the true value that is being measured. For either conventional and virtual metrology, let us assume that the true value of the quantity in question (unknown to the process engineer), is y , and the value estimated by the metrology operation is \hat{y} . We assume there are three types of cost associated with each sample wafer going through a metrology scheme:

1. Revenue: If the true value of a wafer sample is in-spec and it goes through the processing line and ultimately becomes an integrated circuit product (e.g. microprocessor), then the fab profits from selling this product.
2. Processing cost: Cost required to further process one wafer until it becomes a final product.
3. SEM cost: Cost of putting one wafer sample through the SEM (conventional metrology). This includes the cost of physically operating the metrology tool and employment of the engineer.

Here we assume with no loss of generality that the entire wafer is accepted or rejected by the metrology operation. In addition, we assume for simplicity that there is no cost of operating a virtual metrology software tool. Thus, we can define a total profit for each wafer as:

$$\text{profit} = \text{revenue} - \text{process cost} - \text{SEM cost} \quad (1)$$

2.2 Costs Associated with Imperfect VM Predictions

We determine that a sample is "faulty" or "bad" if a metrology value is under or over the respective specification limits. Thus, if the estimated metrology value is bad, the process engineer will have to throw away (or re-work) the sample, and if the estimated value is good, the wafer continues onto subsequent manufacturing processes. Overall, there are four cases that can happen when a process engineer uses either of the metrology schemes. Depending on which category a wafer is in, it incurs a different set of costs.

1. Correctly classified as good: The metrology tool (either conventional or virtual) classifies the sample as good when the true value is good. The wafer goes through subsequent processes and is turned into final product, generating revenue for the company. Given some Upper Specification Limit (USL) and Lower Specification Limit (LSL), this directly translates to

$$P(\hat{y}_{vm} = \text{good} \mid y = \text{good}) = P(\text{USL} \geq \hat{y}_{vm} \geq \text{LSL} \mid \text{USL} \geq y \geq \text{LSL})$$

2. Correctly classified as bad: The metrology tool classifies the sample as bad when the true metrology value is bad, and the wafer is discarded. The correct metrology estimate saves the company cost that could have incurred if the wafer was processed subsequently, as the final product would have been defective due to the faulty true value.

$$P(\hat{y}_{vm} = \text{bad} \mid y = \text{bad}) = P(\hat{y}_{vm} > \text{USL} \parallel \hat{y}_{vm} < \text{LSL} \mid y > \text{USL} \parallel y < \text{LSL})$$

3. Type I error: Also known as a "false alarm", the metrology tool classifies the sample as bad when actually the true value is good (in spec). The proportion of samples with type I error is given by:

$$P(\hat{y}_{vm} = \text{bad} \mid y = \text{good}) = P(\hat{y}_{vm} > \text{USL} \parallel \hat{y}_{vm} < \text{LSL} \mid \text{USL} \geq y \geq \text{LSL})$$

In this case, the wafer is discarded even though it could have been processed to become a final product.

4. Type II error: Also known as a "missed alarm", the metrology tool classifies the sample as good when actually the true value is bad. Similarly, the proportion of samples with type II error is given by:

$$P(\hat{y}_{vm} = \text{good} \mid y = \text{bad}) = P(\text{USL} \geq \hat{y}_{vm} \geq \text{LSL} \mid y > \text{USL} \parallel y < \text{LSL})$$

Table 1: Different kinds of costs that occur for a metrology scheme

	Revenue	Process Cost	SEM Cost
Both Good	Yes	Yes	Yes
Both Bad	No	No	Yes
Type I Error	No	No	Yes
Type II Error	No	Yes	Yes

A missed alarm incurs the greatest cost because even though the wafer goes through the whole manufacturing process, it is not made into a final product at the end due to its defect. The different costs that occur for each classification category are summarized in Table 1. As a visual example, we plot

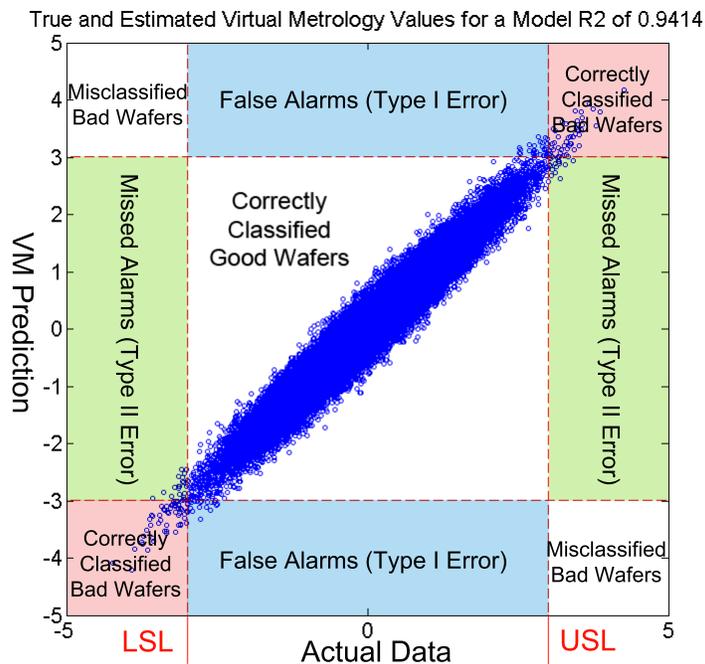


Figure 2: Plot of true and estimated metrology values for an R^2 of 0.94 for a VM model. The dotted lines are the $\pm 3\sigma$ limits. The Type I and Type II error regions are highlighted and labeled.

simulated pairs of true and estimated VM values with correlation ρ of 0.96, as seen in Figure 2. Notice that this is a very high correlation for a statistical model. The corresponding Type I and Type II error regions discussed above are highlighted and labeled.

3 Part I: Preliminary Simulations

We first go through a preliminary example in this section that will motivate our further analysis in section 4.

3.1 Sampling Scenario

One question that remains from Chapter 2 is how to devise a metrology sampling scenario. Traditionally, 2-3 wafers are sampled out of a lot and are observed through the SEM. For the preliminary analysis, we start with a sampling scheme where a fixed proportion of the samples are sent to the SEM (hereon called SEM Proportion) and the rest are monitored on-line via virtual metrology. As seen in Figure 3, we call the traditional metrology sampling method the default scenario, where a certain proportion of the wafers go through SEM metrology, are discarded or kept, and the rest of the wafers continue on through subsequent processing regardless. That is, the remaining wafers go through the manufacturing line unconditionally.¹ On the other hand, for the blended metrology scheme, the virtual metrology model adds an additional monitoring component for each wafer.

3.2 Simulation Setup

To make things simple, we disregard any kind of VM model for now and simulate true CD and VM estimated CD values by generating pairs of correlated points with noise. Assume that the true CD values come from standard normal distribution given by

$$y \sim N(0, 1).$$

For SEM metrology, we assume that the estimated metrology value is centered around the true value due to some random measurement error. That is,

$$\hat{y}_{sem} = y + \epsilon_{sem}$$

where ϵ_{sem} is very small. For simplicity, we assume that ϵ_{sem} is 0, i.e., the SEM accurately predicts the true CD. For virtual metrology, the estimated

¹A similar scheme, where the decision about further wafer processing is done in groups of 24 wafers at a time (also known as wafer lots) is also possible and can be similarly simulated by our method.

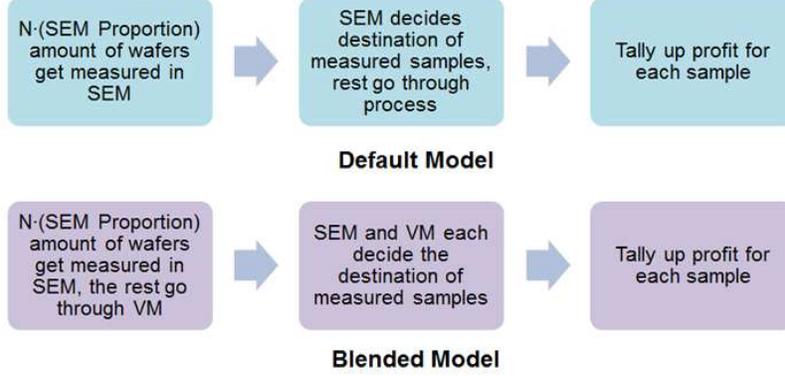


Figure 3: Two different metrology sampling scenarios. The default scenario is when a certain proportion of samples go through SEM metrology and the rest are sent through the subsequent processing line. The blended metrology scenario is when a certain proportion of samples go through SEM metrology, but the destiny of the remaining wafers are determined by the virtual metrology estimate.

values come from an empirical model. We will simulate that relationship by generating \hat{y}_{vm} values so that they have a certain correlation ρ with the true metrology values, as mentioned above. This is given by

$$\hat{y}_{vm} = \rho \cdot y + \sqrt{1 - \rho} \cdot x$$

where x is sampled from a standard normal distribution and is uncorrelated to y . For each metrology scheme, we can now calculate some statistic of accuracy, such as the coefficient of determination (R^2) or the mean-squared error. The formulas for these well-known statistics are given below. [15]

$$R_{vm}^2 = 1 - \frac{\sum_{i=1}^N (y - \hat{y}_{vm})^2}{\sum_{i=1}^N (y - \bar{y})^2}, \quad \text{MSE} = \frac{\sum_{i=1}^N (y - \hat{y}_{vm})^2}{N}$$

where \bar{y} is the average of the true values.

Using this setup, 5,000 true and estimated metrology pairs were simulated for preliminary analysis. The specific quantities used for the costs and revenue are listed in 2. These were extracted from a metrology-related cost analysis

done by the National Institute of Standard and Technology [16] and pricings for Intel microprocessor chips [17]. Note that the estimated metrology cost is for only one metrology step, and in reality, there will be many more. The effects of the VM R^2 and SEM proportion when the SEM metrology cost was fixed at 0.17 (Figure 4) was explored.

Table 2: Cost and profit values used for simulations. These numbers were calculated using reference data from NIST [16] and pricings for Intel microprocessor chips [17].

	Value
Processing Cost	30
Metrology Cost	0.12-0.24
Revenue	500

3.3 Results

In Figure 4(a), the curved surface plot is that of the blended metrology scenario, and the top flat-shaped plot is that of the default case. Unfortunately, we see that for most cases, the default case is more profitable than the blended metrology case. The only time blended metrology generates more profit is when the R^2 of the model is very high ($R^2 \sim 0.98$). Although for higher SEM metrology costs, this portion might be larger, the general trend is that a very high R^2 or very low MSE is needed in order for the blended scheme to be profitable. However, we also notice that most of the samples generated for Figure 4(a) have both 'good' estimated and true values. This is in a sense favorable for the default scenario because most samples will generate revenue after processing, and the magnitude of revenue is significantly larger than the SEM metrology cost. Keeping this in mind, the natural subsequent problem is to explore what would happen if the process went into a fault. For example, we assume that the mean value of the true values have shifted due to some error in the manufacturing process. As seen in Figure 4(b), the smallest VM R^2 needed is now around 0.90 instead of 0.98. This is because many more wafers are defective and do not generate revenue, and the cost

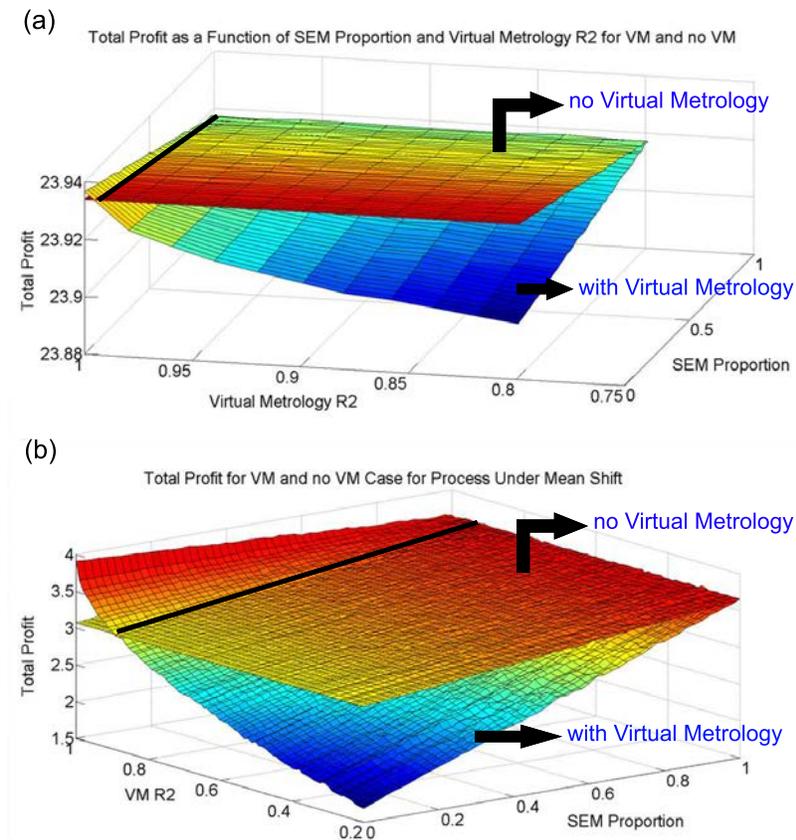


Figure 4: Surface plots for total profit as a function of SEM proportion and VM R² for a fixed SEM metrology cost of 0.17. (a) Case of no mean shift, most samples are good and are estimated to be good. (b) Case of mean shift of +4. Most samples are bad and are estimated to be bad.

of conventional metrology catches up. Moreover, in the default case, the rest of the wafers that do not go through SEM metrology generate unnecessary processing cost because a majority of the wafers are defective. Thus, in this case, although the virtual metrology R^2 may be poor, it is still useful in classifying bad wafers.

3.4 Use of VM in a Faulty Process and the Dependence of VM R^2 on SEM Proportion

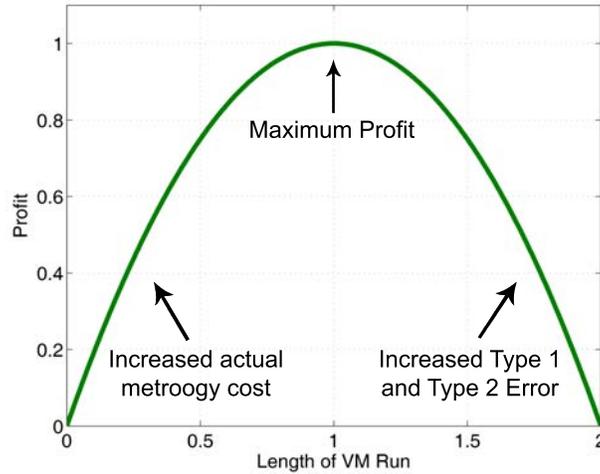


Figure 5: Figure illustrating the tradeoff between actual and virtual metrology in total profit. As the Length of VM Run increases, the actual metrology cost decreases due to less frequent measurements; however, the VM model accuracy decreases due to less training data. The opposite occurs for a decrease in Length of VM Run.

Although the results in Figure 4 give us insight on how the SEM Proportion and accuracy of the VM model effect the overall profit, it is a very simple model. No specific VM model has been used to generate the estimates \hat{y} ; moreover, a critical point not addressed is the coupled relationship between SEM proportion and VM R^2 . In reality, VM models will be trained based on significantly more accurate SEM data, and two effects will counteract each other as the SEM Proportion is increased. One is the increase in the cost of

actual metrology as the engineer makes more frequent measurements. The second is the increase in robustness and accuracy of the VM model due to a larger number of accurate training samples. This means that there are fewer costly Type I and Type II errors mentioned in Chapter 2.2. The opposite occurs when the SEM Proportion is decreased. One can imagine that the total metrology profit is the sum of these two effects, ideally with a maximum point. This is visualized in Figure 5, where the x-axis is now taken to be the Length of VM Run, a quantity opposite to SEM Proportion that we define next.

In the next chapter, we provide a more detailed simulation which incorporates this dependence between the frequency of SEM samples and model accuracy for three different realistic VM models. This is done for a faulty process when VM is useful for the manufacturing process.

4 Part II: Blended Metrology Optimization Using Virtual Metrology Recalibration

Now that we have three different VM models, we present our updated sampling scenario and define some terms before we go into details of the next simulation. As seen in Figure 6, we directly implement the blended metrology scheme introduced in Figure 1, where X_i represents a $n \times p$ matrix of process conditions, and y_i is a $n \times 1$ vector that represents actual metrology values or VM predictions depending on the step. Some terms to note are:

1. Initial Training Length: Number of samples used to initially train the VM model.
2. Length of VM Run: Number of samples for each VM step.
3. Recalibration Length: Number of samples for each recalibration step.

We note that the sum of Initial Training Length and Recalibration Length are analogous to the SEM Proportion in Section 3.1. For convenience, we now work with the Length of VM Run instead of SEM Proportion.

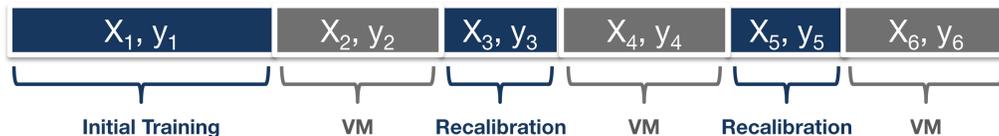


Figure 6: Sampling scenario showing process condition matrices and CD values/estimates.

4.1 Drifting Regression Model

As mentioned in Chapter 3, the case where the process is well-controlled is not a compelling application for virtual metrology, as the default scenario (no VM) is very likely to be the favorable one. In addition, we want a dynamic yet realistic process that will allow us to explore the dependence between SEM Proportion and VM model accuracy (or Length of VM Run

and VM model accuracy). Such a process is a regression model with drifting coefficients. Consider a general linear process

$$y_t = \beta_0^t + \beta_1^t x_1^t + \beta_2^t x_2^t + \dots + \beta_{hid}^t x_{hid}^t + \epsilon$$

where ϵ is the observation error, and t is the superscript for time. It is possible that the VM model misses estimating an independent variable x_{hid}^t and its coefficient β_{hid}^t , and suppose this hidden variable drifts over time by the amount Δx_{hid} . At the next time step, the process is now explained by

$$\begin{aligned} y_{t+1} &= \beta_0^t + \beta_1^t x_1^t + \beta_2^t x_2^t + \dots + \beta_{hid}^t (x_{hid}^t + \Delta x_{hid}) + \epsilon \\ &= \beta_0^t + (\beta_1^t + \Delta\beta)x_1^t + \beta_2^t x_2^t + \dots + \beta_{hid}^t x_{hid}^t + \epsilon. \end{aligned}$$

Now the drift Δx_{hid} is captured in β_1^t with $\Delta\beta x_1^t = \beta_{hid}^t \Delta x_{hid}$ as a linearly drifting coefficient. Thus, the final model of the true process would be

$$y_t = \vec{\beta}_t \mathbf{X}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (2)$$

$$\vec{\beta}_{t+1} = \vec{\beta}_t + \vec{u} + \vec{\eta}_t, \quad \vec{\eta}_t \sim N(0, \mathbf{Q}) \quad (3)$$

where y_t is the true CD, \mathbf{X}_t are the regressor variables, and $\epsilon_t, \vec{\eta}_t$ are white noise variables. Finally, \vec{u} controls how much the coefficient drifts over time.

4.2 Virtual Metrology Models

We provide a brief theoretical background on the three VM models that are used in the analysis: Ordinary Least Squares (OLS), Exponentially-weighted Least Squares (EWLS), and the Kalman Filter (KF) in the context of the sampling scenario in Figure 6.

4.2.1 Ordinary Least Squares

Assuming the process is of the form

$$y = \vec{\beta} \mathbf{X} + \epsilon, \quad \epsilon \sim N(0, \sigma_\epsilon^2)$$

the well-known estimate of the coefficients is

$$\vec{\hat{\beta}}_{SEM} = (\mathbf{X}_{SEM}^T \mathbf{X}_{SEM})^{-1} \mathbf{X}_{SEM}^T \vec{y}_{SEM}$$

In the context of this work,

$$\mathbf{X}_{SEM} = \begin{pmatrix} \cdots \mathbf{X}_1 \cdots \\ \cdots \mathbf{X}_3 \cdots \\ \vdots \end{pmatrix}, \quad \vec{y}_{SEM} = \begin{pmatrix} \vec{y}_1 \\ \vec{y}_3 \\ \vdots \end{pmatrix}$$

where \mathbf{X}_i and \vec{y}_i , $i = 1, 3, 5 \dots$ refer to Figure 6 and denote the training data from the SEM. That is, we accumulate all data from the SEM not knowing the true process is faulty. Given this is our VM model, the CD prediction for subsequent VM samples is:

$$\begin{aligned} \vec{y}_{VM} &= \mathbf{X}_{VM} \vec{\beta}_{SEM} \\ &= \mathbf{X}_{VM} (\mathbf{X}_{SEM}^T \mathbf{X}_{SEM})^{-1} \mathbf{X}_{SEM}^T \vec{y}_{SEM} \end{aligned}$$

4.2.2 Exponentially-weighted Linear Regression

For this model, we assume that the engineer knows there is some sort of drift going on and wants to weight the recent observations more heavily than the earlier ones. The weights are given by

$$w_t = \alpha(1 - \alpha)^{(N_{train}-t)}$$

where α is a tunable parameter. The best value for α was found by training the model on different α values and finding the one that minimized the MSE. With $\mathbf{W} = \text{diag}(\vec{w})$, the EWLS estimates of $\vec{\beta}$ are given by

$$\vec{\hat{\beta}}_{SEM} = (\mathbf{X}_{SEM}^T \mathbf{W} \mathbf{X}_{SEM})^{-1} \mathbf{X}_{SEM}^T \mathbf{W} \vec{y}_{SEM}$$

and similarly, the estimated y are given by

$$\begin{aligned} \vec{\hat{y}}_{VM} &= \mathbf{X}_{VM} \vec{\hat{\beta}}_{SEM} \\ &= \mathbf{X}_{VM} (\mathbf{X}_{SEM}^T \mathbf{W} \mathbf{X}_{SEM})^{-1} \mathbf{X}_{SEM}^T \mathbf{W} \vec{y}_{SEM} \end{aligned}$$

4.2.3 Kalman Filter

For this section, we follow the notation in [18]. Consider the linear Gaussian state space model

$$\begin{aligned} y_t &= Z_t \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, H_t) \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \quad t = 1, \dots, n, \\ & & \alpha_1 &\sim N(a_1, P_1) \end{aligned}$$

where y_t are the observations, α_t are hidden state variables controlling the process, and ϵ_t, η_t are independent Gaussian noise sequences with corresponding covariance matrices H_t and Q_t .

Our objective is to obtain the conditional distribution of α_{t+1} given Y_t for $t = 1, \dots, n$ where $Y_t = y_1, \dots, y_t$. Since all distributions are normal, conditional distributions of subsets of variables given other subsets of variables are also normal (Gaussian Lemma). Now all we need is $\alpha_t|Y_{t-1} \sim N(a_t, P_t)$ and then

$$\begin{aligned} a_{t+1} &= E(\alpha_{t+1}|Y_t) \\ P_{t+1} &= \text{Cov}(\alpha_{t+1}|Y_t). \end{aligned}$$

Since $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$, we have

$$\begin{aligned} a_{t+1} &= E(T_t\alpha_t + R_t\eta_t|Y_t) \\ &= T_t E(\alpha_t|Y_t), \\ P_{t+1} &= \text{Cov}(T_t\alpha_t + R_t\eta_t|Y_t) \\ &= T_t \text{Cov}(\alpha_t|Y_t)T_t^T + R_tQ_tR_t^T \end{aligned}$$

for $t = 1, \dots, n$. We now define v_t , the one-step forecast error of y_t given T_{t-1} .

$$\begin{aligned} v_t &= y_t - E(y_t|Y_{t-1}) \\ &= y_t - E(Z_t\alpha_t + \epsilon_t|Y_{t-1}) \\ &= y_t - Z_t a_t \end{aligned}$$

Notice that when Y_{t-1} and v_t are fixed, Y_t is also fixed. Since observing Y_t is the same as observing Y_{t-1}, v_t , we see that $E(\alpha_t|Y_t) = E(\alpha_t|Y_{t-1}) + E(\alpha_t|v_t)$. It is easy to see that

$$\begin{aligned} E(v_t) &= 0 \\ \text{Cov}(v_t, Y_{t-1}) &= 0. \end{aligned}$$

Through the lemma in multivariate normal regression, we know that

$$\begin{aligned} E(\alpha_t|Y_t) &= E(\alpha_t|Y_{t-1}, v_t) \\ &= E(\alpha_t|Y_{t-1}) + E(\alpha_t|v_t) \\ &= E(\alpha_t|Y_{t-1}) + \text{Cov}(\alpha_t, v_t)[\text{Var}(v_t)]^{-1}v_t \\ &= a_t + M_t F_t^{-1}v_t, \end{aligned}$$

where $M_t = \text{Cov}(\alpha_t, v_t) = P_t Z_t^T$, and $F_t = \text{Var}(v_t) = Z_t P_t Z_t^T + H_t$.

Substituting everything into expression for a_{t+1} and P_{t+1} , we get

$$\begin{aligned} a_{t+1} &= T_t a_t + T_t M_t F_t^{-1} v_t \\ &= T_t a_t + K_t v_t, \quad t = 1, \dots, n, \\ \text{where } K_t &= T_t M_t F_t^{-1} = T_t P_t Z_t^T F_t^{-1}. \end{aligned}$$

Through similar analysis,

$$\begin{aligned} P_{t+1} &= T_t P_t L_t^T + R_t Q_t R_t^T, \quad t = 1, \dots, n, \\ \text{where } L_t &= T_t - K_t Z_t. \end{aligned}$$

We collect all the filtering recursion equations given by

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t^T + H_t, \\ K_t &= T_t P_t Z_t^T F_t^{-1}, & L_t &= T_t - K_t Z_t, \quad t = 1, \dots, n, \\ a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t L_t^T + R_t Q_t R_t^T \end{aligned}$$

For missing observations and observations to be forecasted, v_t and K_t of the filter are set to zero, and the updates just become

$$a_{t+1} = T_t a_t, \quad P_{t+1} = T_t P_t T_t^T + R_t Q_t R_t^T.$$

Given these update equations, the following Z_t , T_t , and α_t matrices gives us the model in equations 2 and 3.

$$\alpha_t = \begin{pmatrix} u \\ \beta_t \end{pmatrix}, \quad Z_t = \begin{pmatrix} 0 & \dots & 0 & X_t \end{pmatrix}, \quad T_t = \begin{pmatrix} I & 0 \\ I & I \end{pmatrix}, \quad R_t = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

For this analysis, we assume that this state space form is known and therefore T_t and R_t are known. In addition, we assume that the variances of the error terms are also known. Realistically when implementing the Kalman Filter, these parameters should be estimated via some parameter estimation algorithm (e.g., Expectation-Maximization algorithm) since the engineer does not know the true form of the state space model.

4.3 Results

We generate data based on equations 2 and 3 with the following parameters

$$\beta_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad u = \begin{pmatrix} 0.001 \\ 0 \\ 0.001 \end{pmatrix}, \quad Q_t = \begin{pmatrix} \sigma_\eta^2 & 0 & \cdots & 0 \\ 0 & \sigma_\eta^2 & & \vdots \\ \vdots & & \sigma_\eta^2 & 0 \\ 0 & \cdots & 0 & \sigma_\eta^2 \end{pmatrix}, \quad \sigma_\eta^2 = 0.00001, \quad H_t = 0.2$$

and an Initial Training Length of 200 and Recalibration Length of 20 samples. Values of Revenue, Processing Cost, and SEM Cost would be determined by the manufacturing process, but to clearly show the existence of maximum Figure 5, we use values of 300, 80, and 20 (we now assume the process has many more metrology steps than that of Chapter 3), respectively. 4000 samples are simulated and the corresponding total profit, Type I / Type II error rate, and the MSE are calculated for each VM model according to the scenario in Figure 6. This is done 100 times. For the Kalman Filter, we start with a randomized hidden state vector as our initial guess.

The main results of the simulation are shown in Figures 7, 8, and 9. The first column of each plot is the total profit calculated via equation 1, and the second column is the total MSE for all VM samples. Finally, the third column shows the Type I and Type II error probabilities. One subtle difference is that in Section 2.2, we defined the errors in terms of conditional probabilities. As it is hard to calculate $P(y = \text{good or bad})$ explicitly, we plot the joint probabilities instead. Results for the first 20 simulations are plotted in a faded color for each column, as well as the mean value of the 100 simulations in solid color. Figure 10 shows the coefficient estimates for each VM model as time passes, where the faded color plots are the true coefficient values given by the model in equation 3.

4.3.1 Ordinary Least Squares

One noticeable feature about Figure 7 is the non-existent Type I errors and the very large amount of Type II errors. This is due to the under-estimation of the regression coefficients as seen in Figure 10(a). Since all past and present observations are weighed equally in the recalibration phase, the coefficient estimates have significantly lower values than the true ones. This constraint

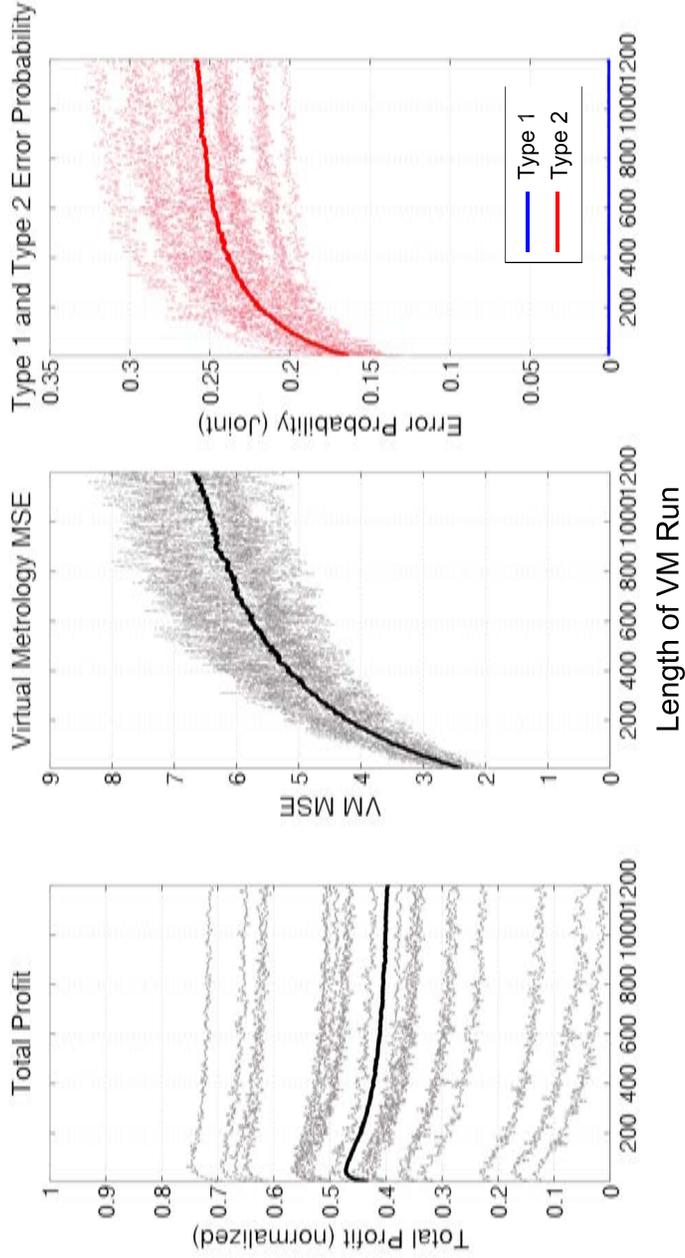


Figure 7: Total profit, MSE, and Type I / II error probabilities for Ordinary Least Squares VM model. The maximum profit occurs at 0.4717 when the Length of VM Run is 50. The increase in estimation variance is significant as the Length of VM Run increases.

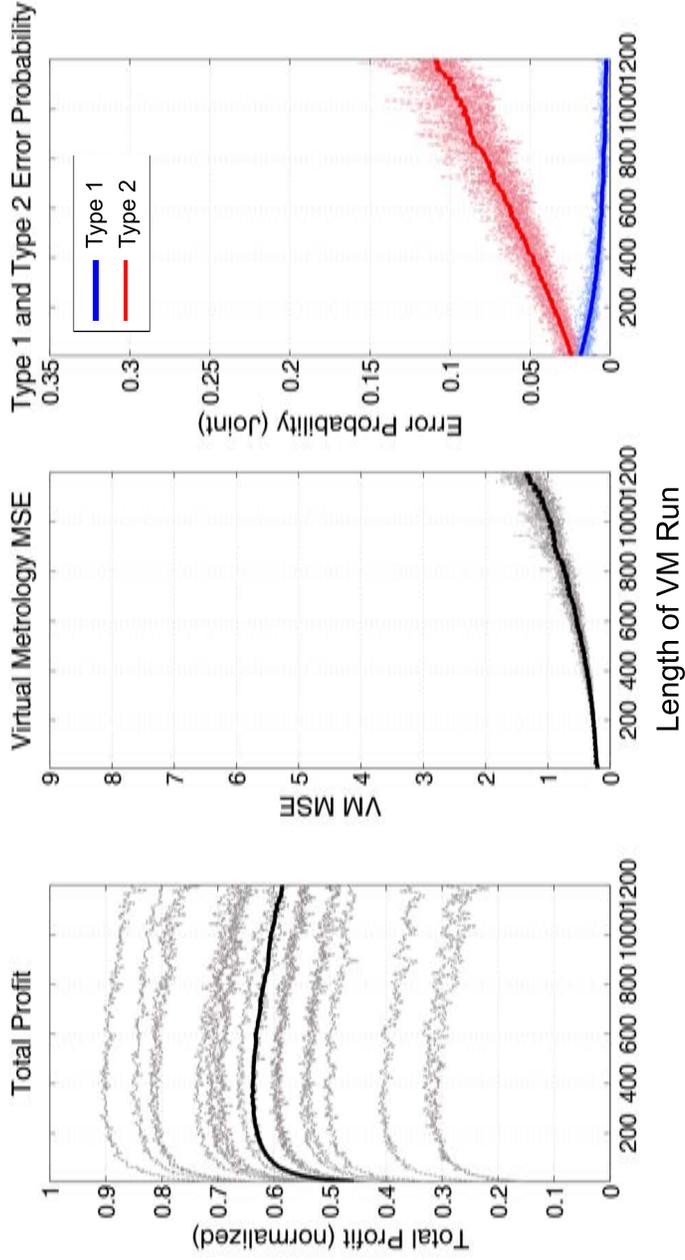


Figure 8: Total profit, MSE, and Type I / II error probabilities for Exponentially-weighted Linear Regression VM model. MSE and error probabilities have significantly improved compared to OLS. The maximum profit occurs at 0.6388 when the Length of VM Run is 350.

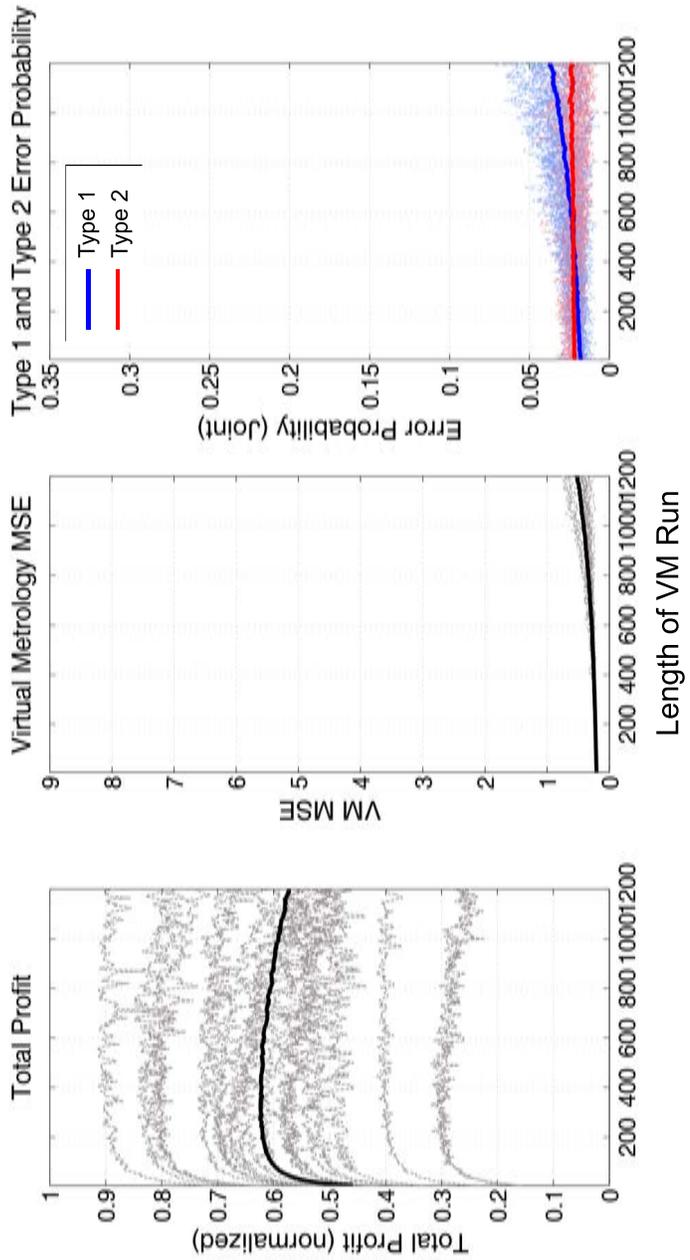


Figure 9: Total profit, MSE, and Type I / II error probabilities for the Kalman Filter VM model. This model shows the highest accuracy and lowest error probabilities as it is a filter constructed for a state-space model, and we have assumed we know the exact values of some parameters. The maximum profit occurs at 0.6241 when the Length of VM Run is 275.

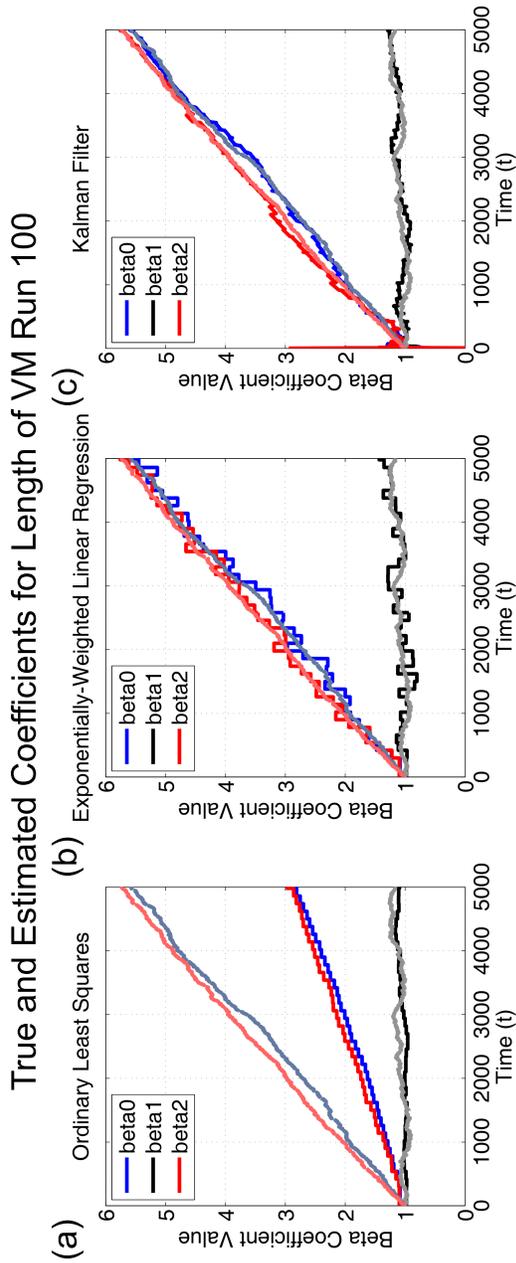


Figure 10: Regression coefficient estimates for each VM model for a Length of VM Run of 100. (a) Estimates for OLS. Notice that the coefficients are underestimated due to retained memory from start of the process, leading to an increase in Type II error probabilities. (b) Estimates for EWLR. Variance of estimations have increased due to smaller training size, but tracks the drifting coefficients. (c) Estimates for KF. Most accurate performance out of the three models.

in the gain leads to VM estimates that are mostly inside the LSL/USL limits and as a result, the OLS fails to catch most of the alarms that happen, leading to the significant Type II error probabilities. In addition, variance of the MSE and Type II errors drastically increase as the Length of VM Run is increased due to the decrease in training sample size and failure to incorporate the process drift in the model. The maximum normalized mean profit occurs at 0.4717 when the Length of VM Run is 50.

4.3.2 Exponentially-weighted Linear Regression

1EWLR shows a significantly reduced VM MSE than OLS (Figure 8), as only the recent observations are used in the recalibration stage. This is also seen in Figure 10(b), where the EWLR estimates accurately track the drift in the true coefficients. However, in contrast to Figure 10(a), the estimates have more variance. This translates to an increased probability of Type I error, and as seen in the third column of Figure 8, as the Type I probability is not zero anymore. We see an increase in Type II errors as the Length of VM Run increases because the coefficient estimates for a certain VM stage stay constant while the process keeps on drifting. In addition, the variance of the MSE and Type I / Type II error probabilities also increase as the Length of VM Run increases. The maximum normalized mean profit occurs at 0.6388 when the Length of VM Run is 350.

4.3.3 Kalman Filter

The Kalman Filter gives us the least MSE out of all three VM models. Note that this is because the filter is constructed on such a state-space model like equations 2 and 3. Moreover, we have assumed that the exact value of some of the parameters are known, including the noise covariance matrices. As seen in Figure 10(c), the Kalman Filter estimates accurately track the true values. Although the Type I and Type II error probabilities are quite low, we again see that the variances of these quantities increase as Length of VM Run increases. The maximum normalized mean profit occurs at 0.6241 when the Length of VM Run is 275.

Table 3: Maximum mean profit and corresponding Length of VM Run for each VM model.

	Max. Profit	Length of VM Run
Ordinary Least Squares	0.4717	50
Exponentially-weighted LR	0.6388	350
Kalman Filter	0.6241	275

4.4 Discussion

The total profit, MSE, and Type I / Type II error probabilities have been plotted for each VM model. In addition, the maximum mean profits were identified with the corresponding Length of VM Run. Given a LSL/USL pair, each model generates a different Type I / Type II error pattern that directly translates to total profit. It is the combination of model accuracy and missed / false alarm patterns that determine the optimal total metrology profit. Moreover, one must also take into account the increase in VM estimation variance as the Length of VM Run increases, as a robust estimator is as important as an accurate estimator.

Although Figure 11 shows that the Kalman Filter has the highest average total cost out of all three VM models, we want to emphasize that the example given in this paper is one of many cases and that the results depend on different cost and revenue values. Consider an extreme example of a process with a very high Type I cost; in this case, an OLS-like model would be the most beneficial to the fab, even though the OLS has the worst MSE out of the three models. However, going back to the results of Chapter 3, VM was found to be most useful for deviating processes. Taking into account that most of these faulty processes have some kind of drift incorporated in them, we think that using an adaptive model like the Kalman Filter is crucial to implementing a successful VM system.

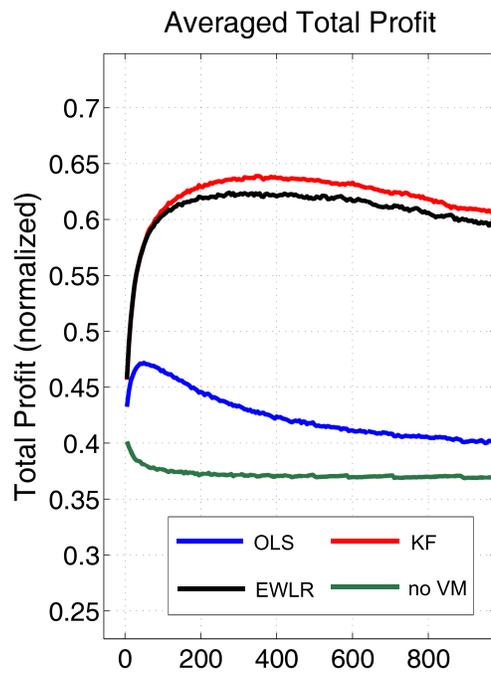


Figure 11: Averaged total profit for the three VM models and default scenario, where no VM is introduced. We see that specifically for the example presented in this paper, all scenarios with VM perform better than the default case.

5 Conclusion and Outlook

This work takes a step back from empirical VM model construction and analyzes when VM is actually useful for the fab and how it can be optimized as a function of VM recalibration frequency. This is done by identifying 4 categories of VM estimate classification and defining critical parameters such as the revenue, processing cost, and actual metrology cost. These parameters were combined into a metric we called the total profit. Preliminary simulations showed that VM models require a very high accuracy in order for it to be beneficial for well-controlled processes. Thus, we concentrated our analysis on a largely deviating process, the regression model with drifting coefficients, to link the relationship between VM recalibration frequency, VM model accuracy, and total profit.

In practical reality this means that a high VM proportion may be beneficial early into the process lifecycle, while this proportion maybe be gradually reduced as the process becomes more mature, and therefore more stable. This finding is similar to the intuition that the overall role of metrology changes with process maturity, starting with detecting and diagnosing significant deviations, and continuing to using metrology to drive more subtle run-to-run control adjustments.

A blended metrology scenario was carried out on simulated data for 3 VM models, the OLS, EWLR, and the Kalman Filter. Results indicated each VM model had different false (Type I) and missed (Type II) alarm patterns that translated to different total profit patterns as a function of Length of VM Run. Moreover, an optimal value of Length of VM Run that gave the maximal total profit was identified for each model. These all indicated the need for missed and false alarm pattern analysis for a given VM model, rather than solely focusing on increasing a certain accuracy metric.

Although this work gives us a view into how the cost, error probabilities, and recalibration sizes are related, a lot of assumptions were incorporated into the process model. For instance, we assumed the observation errors were independent from one time to the next, whereas in reality some auto-correlation would need to be accounted for. In addition, we assumed exact parameter values for the Kalman Filter. It would be interesting to see how the results change for a real processing dataset where the parameters of the

Kalman Filter would have to be estimated via some parameter estimation algorithm. Moreover, we believe the next step is to provide a theoretical framework past Monte Carlo simulations that would provide robust and accurate decision rules for the metrology industry.

References

- [1] Aftab A. Khan and James R. Moyne. An approach for factory-wide control utilizing virtual metrology. *IEEE Trans. on Semiconductor Manufacturing*, 20(4):364–375, November 2007.
- [2] Costas J. Spanos and Gary J. May. *Fundamentals of Semiconductor Manufacturing and Process Control*. Wiley-IEEE Press, first edition, 2006.
- [3] Claus Schneider, Lothar Pfitzer, and Heiner Ryssel. Integrated metrology: An enabler for advanced process control (apc). volume 4406 of *Proceedings of SPIE*, pages 118–123, 2001.
- [4] PingHsu Chen, Sunny Wu, Junshien Lin, Francis Ko, Henry Lo, Jean Wang, C.H. Yu, and M.S. Liang.
- [5] John Musacchio. Run to run control in semiconductor manufacturing. Master’s thesis, U.C. Berkeley, Berkeley.
- [6] Yaw-Jen Chang, Yuan Kang, Chih-Liang Hsu, Chi-Tim Chang, and Tat Yan Chan. Virtual metrology technique for semiconductor manufacturing. Intl. Joint Conference on Neural Networks, pages 5289–5293, July 2006.
- [7] Jonathan Chang Yung-Cheng and Fan-Tien Cheng. Application development of virtual metrology in semiconductor industry. IECON, pages 124–129, 2005.
- [8] Min-Hsiung Hung, Tung-Ho Lin, Fan-Tien Cheng, and Rung-Chuan Lin. A novel virtual metrology scheme for predicting cvd thickness in semiconductor manufacturing. *IEEE/ASME Trans. on Mechatronics*, 12(3):308–312, 2007.
- [9] An-Jhih Su, Jyh-Cheng Jeng, Hsiao-Ping Huang, Cheng-Ching Yu, Shih-Yu Hung, and Ching-Kong Chao. Control relevant issues in semiconductor manufacturing: Overview with some new results. *Control Engineering Practice*, 15:1268–1279, 2007.
- [10] Dekong Zeng and Costas J. Spanos. Virtual metrology modeling for plasma etch operations. *IEEE Trans. on Semiconductor Manufacturing*, 22(4):419–431, 2009.

- [11] Pilsung Kang, Hyoung joo Lee, Sungzoon Cho, Dongil Kim, Jinwoo Park, and Chan-Kyoo Park. A virtual metrology system for semiconductor manufacturing. *Expert Systems with Applications*, 36:12554–12561, 2009.
- [12] Shane Lynn, John Ringwood, Emanuele Ragnoli, Sean McLoone, and Niall MacGearailt. Virtual metrology for plasma etch using tool variables. Advanced Semiconductor Manufacturing Conference, pages 143–148, 2009.
- [13] Yu-Chuan Su, Tung-Ho Lin, Fan-Tien Cheng, and Wei-Ming Wu. Implementation considerations of various virtual metrology algorithms. IEEE Conf. on Automation Science and Engineering, pages 276–281, September 2007.
- [14] Fan-Tien Chang, Yeh-Tung Chen, Yu-Chuan Su, and Deng-Lin Zeng. *IEEE Trans. on Semiconductor Manufacturing*, 21(1), February 2008.
- [15] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, fourth edition, 2001.
- [16] William F. Finan. Metrology-related costs in the u.s. semiconductor industry, 1990, 1996, and 2001. Technical report, National Institute of Standards and Technology, May 1998.
- [17] Michael Kanellos. Intel’s manufacturing cost: \$40 per chip. *CNET News*, September 2005.
- [18] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford Univ., first edition, 2001.