Statistical Methods for Semiconductor Manufacturing

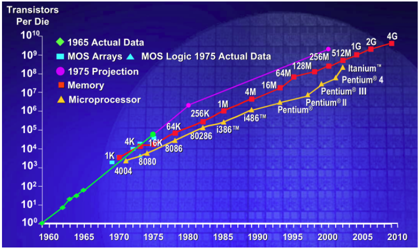
Gian Antonio Susto

PhD School on Information Engineering
XXV Series, ICT Section
Advisor: Prof. Alessandro Beghi

Padova - February 28th, 2013

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Motivation



- Semiconductor-based devices are pervasive (PCs, mobiles, cars, etc.)
- **Semiconductor Manufacturing**: one of the most technologically advanced with new products every months
- Constant search for improving process quality and control

- In the milestone paper [**Edgar 2000**] a huge margin of improvements for modeling and control in semiconductor manufacturing has been described
- 13 years have passed ...

T. Edgar, S. Butler, W. Campbell, C. Pfeiffer, C. Bode, S. Hwang, K. Balakrishnan, J. Hahn
Automatic Control in Microelectronics Manufacturing: Practices, Challenges, and Possibilities
Automatica, Vol. 36, Issue 11, Nov. 2000, Pag. 1567-1603

Outline

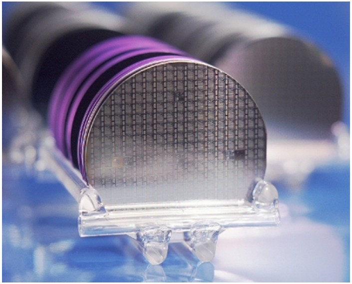
- Advanced Process Control (APC) systems have proliferated in the past years
- Two APC topics investigated
  - (a) **Virtual Metrology (VM)**
  - (b) **Predictive Maintenance (PdM)**
- Several challenging aspects tackled
- We will illustrate the following applications
  - (i) VM: Prediction of CVD thickness
  - (ii) VM: Multi-Step VM
  - (iii) VM with Time Series data
  - (iv) VM and Run-to-Run Control
  - (v) PdM for epitaxy
  - (vi) PdM for ion implantation

Summary of PhD activities

- Involvement in *IMPROVE* (*Implementing Manufacturing science solutions to increase equiPment pROductiVity and fab pErformance*): a 42 months, 37 million euro project
- Collaborations:
  - (a) Infineon Technologies AG, Austria (IFAT)
  - (b) STMicroelectronics
  - (c) University of Pavia
- Research periods abroad:
  - (a) IFAT in Villach, Austria
  - (b) National University of Ireland, Maynooth

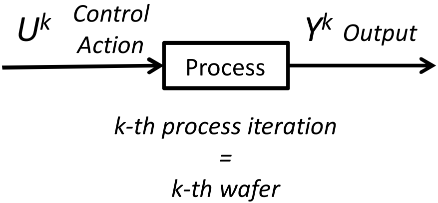
NUI MAYNOOTH
Ollscoil na hÉireann Má Nuad

Elements of Semiconductor Manufacturing



- The **Wafer**: main product of semiconductor industry
- Thin slice of silicon crystal used as substrate for microelectronic devices
- Production organized in **Lots**: groups of 25 wafers
- One process iteration, one wafer

- Hundreds of processes:
  (a) Chemical Vapor Deposition (CVD)
  (b) Lithography
  (c) Etching
  (d) ...

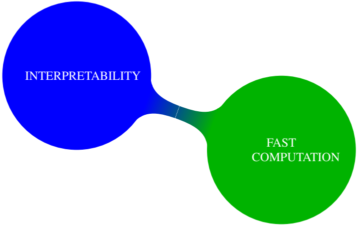


Virtual Metrology (VM)

- Several quantities (Y) on the wafer are *costly to be measured*, but they are necessary to monitor process quality
- Y only measured for few wafers in a lot
- **VM system**: model of a process for estimating Y based on the availability of tool and/or logistic data X

$\Rightarrow$ **Regression Problem**

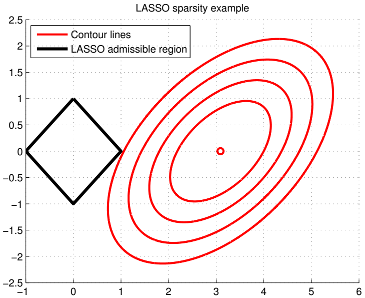
$\Rightarrow$ **Black Box Modeling**



- Besides accuracy a VM system must provide
  - (i) **Fast Computation**: new products every months, models needs to be recomputed frequently
  - (ii) **Interpretability**: to provide process insights and guidelines for control actions

VM: Challenge (1) - High Dimensionality

- **(1) High-Dimensionality**: hundreds/thousands of input variables
- Two approaches:
  (i) *Dimensionality Reduction* - (Correlation, PCA) two-step approach
  (ii) *Variable Selection* - (Stepwise Selection, LARS, Regularization Methods) interpretable results

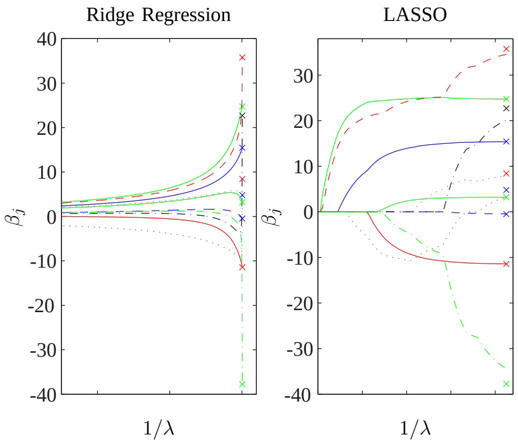


- *Regularization Methods*
  Penalty on the complexity of the model:

$$\mathcal{L}(\theta) = \|Y - X\theta\|^2 + \lambda\mathcal{R}(\theta)$$

- Different methods for different choice of $\mathcal{R}$
  (i) Ridge Regression $- R(\theta) = \sum_{j=1}^{p} \theta_j^2$
  (ii) LASSO $- R(\theta) = \sum_{j=1}^{p} |\theta_j|$
  (iii) Elastic Net
  $R(\theta) = \alpha \sum_{j=1}^{p} \theta_j^2 + (1-\alpha) \sum_{j=1}^{p} \theta_j^2$

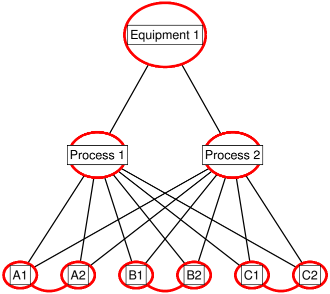
VM: Challenge (1) - High Dimensionality

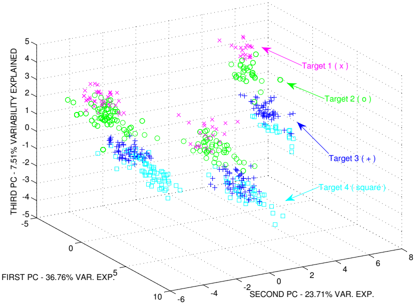


- RR suitable for highly correlated dataset
- LASSO provide sparse (more interpretable) models

VM: Challenge (2) - Data Fragmentation

- **(2) Data Fragmentation**: thousands of products run, different tool settings (*recipes*), equipment composed of 2/3 chambers
- No single modeling for each logistic path, not enough data. Two approaches:
  - (i) *smart data clustering* (PCA, information-theory elements)
  - (ii) *multi-task modeling*





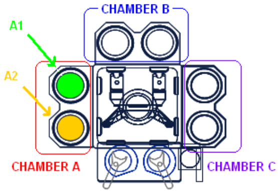
VM: Prediction of CVD thickness

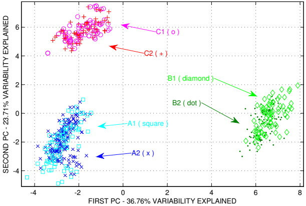
- Collaboration with Infineon Technologies AG, Austria
- Chemical Vapor Deposition (CVD): chemically growing a layer upon the wafer
- Prediction of the thickness of the deposited layer from tool variables (temperatures, pressures, flows, etc.)

(i) High data fragmentation:
- 10 products run, different recipes
- tool with 3 chambers (A, B, C) and 2 sub-chambers (1, 2)

(ii) High number of variables ($\approx$ 450)

- Modeling with Neural Networks and Least Angle Regression (LARS)

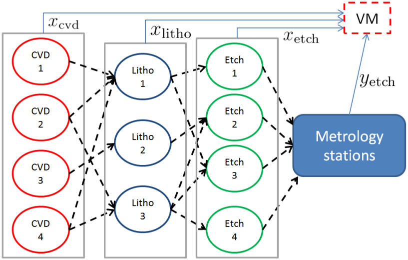


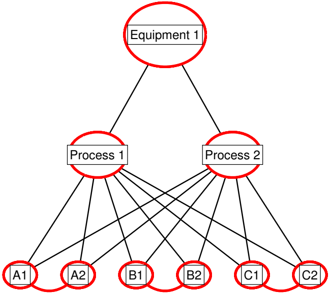


VM: Multi-Step Modeling

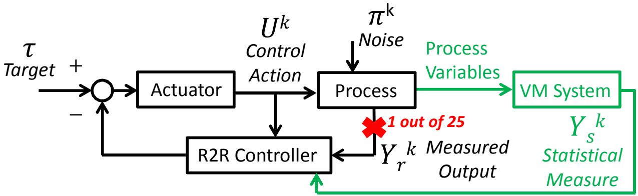
- Collaboration with University of Pavia and Infineon Technologies AG, Austria
- Quality features of a certain wafer depend on the whole processing
- Multi-level: Generalized Additive Model (sum η logistic effects)

$$f(X) = \sum_{k=1}^{\eta} f_k(X_k)$$

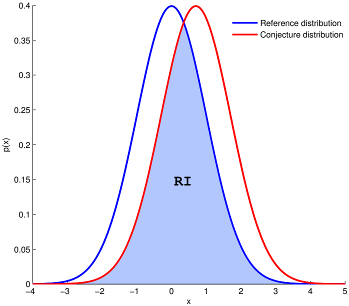




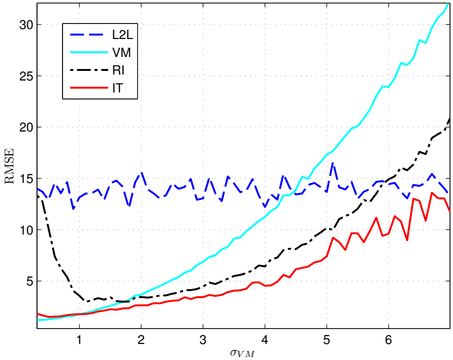
VM: Challenge (3) - Run-to-Run and VM 1/2



- Run-to-Run control
- 1 out of 25 measured: Lot-to-Lot control
- Mixture of measures in the control loop
  - (i) Physical Y_R, ≈ 0 error
  - (ii) Statistical Y_S, estimation i.e. Gaussian distributed error
- State of the art: Reliance Index (RI) based penalization



VM: Challenge (3) - Run-to-Run and VM 2/2



- Information Theory elements
- Penalization on VM measures based on
  - (i) statistical distance
  - (ii) iterations from last physical measure

VM: Challenge (4) - Time Series Learning (TSL)

Modeling

- Training dataset (n observations): p regressors x^j and a scalar target value y

$$\mathcal{S} = \left\{x_i \in \mathbb{R}^{1\times p},\ y_i \in \mathbb{R}\right\}_{i=1}^{n} \quad \Rightarrow \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \; X = \begin{bmatrix} x_1^1 & \dots & x_1^p \\ \vdots & & \vdots \\ x_n^1 & \dots & x_n^p \end{bmatrix}$$

- Goal: create a predictive model f to provide predictions for $\widetilde{x} \notin \mathcal{S}$

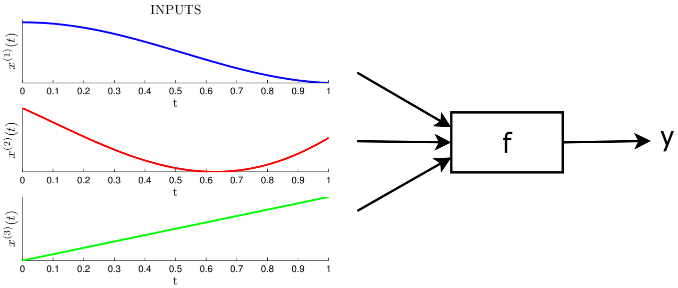
Input Space Representation

- Data not always organized in $n \times p$ matrix (x^j not always scalar)
- Representation with little information loss not always trivial

i.e. text data mining

Introduction: Problem Statement

- Typical settings for modeling of industrial processes:
  - (i) p time series input

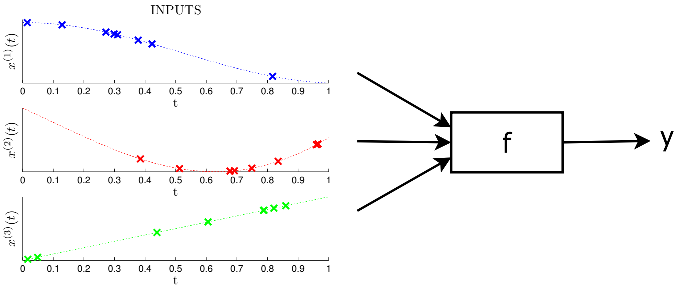


Goal

Build a regressor matrix and aggregate the input information in summary features with small loss of information

Introduction: Problem Statement

- Typical settings for modeling of industrial processes:
  - (i) *p* time series input
  - (ii) Irregularly sampled measures

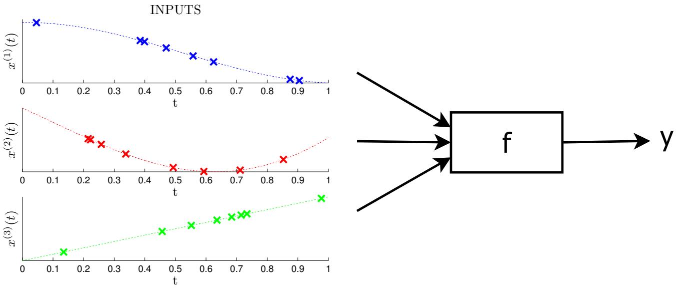


Goal

Build a regressor matrix and aggregate the input information in summary features with small loss of information

Introduction: Problem Statement

- Typical settings for modeling of industrial processes:
  - (i) p time series input
  - (ii) Irregularly sampled measures
  - (iii) Sampling can vary observation-wise

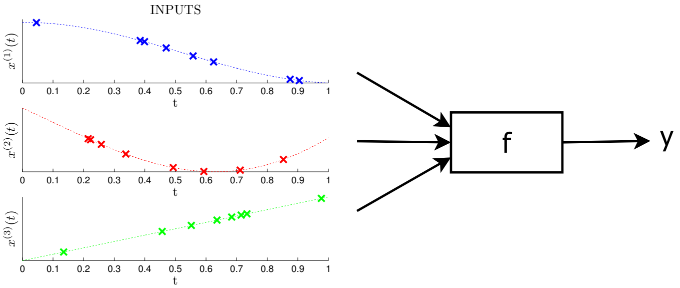


Goal

Build a regressor matrix and aggregate the input information in summary features with small loss of information

Introduction: Problem Statement

- Typical settings for modeling of industrial processes:
  - (i) *p* time series input
  - (ii) Irregularly sampled measures
  - (iii) Sampling can vary observation-wise



Goal

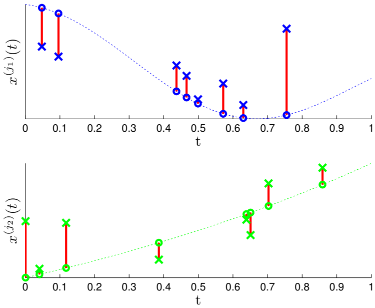
Build a regressor matrix and aggregate the input information in summary features with small loss of information

TSL: Mathematical Formulation

- n observations, p time series: i-th observation $\mathcal{X}_i$

$$\mathcal{X}_i = [x_i^{(1)}(t) \; \dots \; x_i^{(j)}(t) \; \dots \; x_i^{(p)}(t)], \; t \in [0,1], \forall \, j$$

- A scalar output y: training dataset $\mathcal{S} = \{\mathcal{X}_i, \; y_i\}_{i=1}^{n}$ for learning f



- $x_i^{(j)}(t)$ samples

$$\left\{ t_{i,s}^{(j)}, \; z_{i,s}^{(j)} \right\}_{s=1}^{\mathcal{N}_{i,j}}$$

- Differm length and sampling timestamps

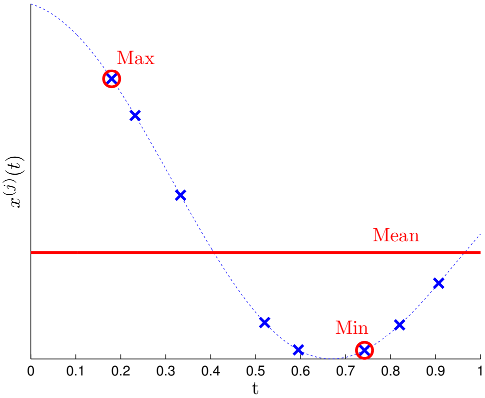
$$\mathcal{N}_{i,j} \neq \mathcal{N}_{i,m}, \; \mathcal{N}_{i,j} \neq \mathcal{N}_{k,j}$$

$$t_{i,s}^{(j)} \neq t_{i,s}^{(m)}, t_{i,s}^{(j)} \neq t_{k,s}^{(j)}$$

- Noisy observations

$$z_{i,s}^{(j)} = x_i^{(j)}(t_{i,s}^{(j)}) + v_{i,s}^{(j)} \quad v_{i,s}^{(j)} \sim N(0, \rho_j^2)$$

TSL: Classical Approaches - Statistical Moments



- Regressor matrix

$$\Phi = [\Phi_1 \ \ldots \ \Phi_j \ \ldots \Phi_p]$$

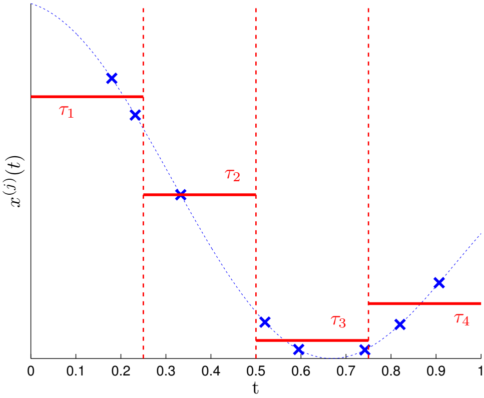
where $\Phi_j \in \mathbb{R}^{n \times k_{\max}}$ is

$$\Phi_j[i,k] = m^{(k)}\left(\left\{z_{i,s}^{(j)}\right\}_{s=1}^{\mathcal{N}_{i,j}}\right)$$

$k_{\max}$ highest considered moment order, $m^{(k)}(\cdot)$ is the k-th sample moment

- Drawbacks:
  - (i) Loss of dependency between information and time
  - (ii) Independent data points: little statistical meaning for autocorrelated time series

TSL: Classical Approaches - Systematic Sampling



- Interval $[0, 1]$ divided into $\mathcal{N}$ segments $[\tau_1 \ \ldots \ \tau_{\mathcal{N}}]$
- Regressor matrix

$$\Phi = [\Phi_1 \ \ldots \ \Phi_j \ \ldots \Phi_p]$$

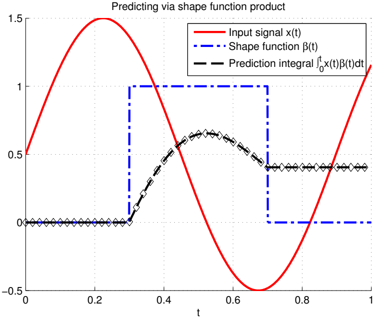
populated with

$$\Phi_j[i, k] = \mathrm{Avg}[z_{i,s}^{(j)} : t_{i,s}^{(j)} \in \tau_k].$$

- Drawbacks:
  - (i) Number of segments to be selected as trade-off: temporal resolution vs robustness to noise
  - (ii) Different features computed with different numbers of values: data reliability issues (*PCA a possible solution*)

SAFE: Supervised Aggregative Feature Extraction

- **SAFE**: A methodology to include temporal information in the regression matrix Φ.
- **Assumption**: the input $x_i^{(j)}$ influences y through a weighted integration



- Model structure:

$$f(\mathcal{X}_i) := \sum_{j=1}^{p} \left\langle x_i^{(j)}(t), \beta^{(j)}(t) \right\rangle_{L^2}$$

$$\langle f, g \rangle_{L^2} = \int_{-\infty}^{\infty} f(t)g(t)dt$$

- $\beta = \left[\beta^{(1)}(t) \dots \beta^{(j)}(t) \dots \beta^{(p)}(t)\right]$ to be estimated

SAFE: Regularization

- **Objective Function**:

$$\mathcal{L}(\beta) = \mathcal{F}(\beta) + \lambda\mathcal{R}(\beta)$$

where

$$\mathcal{F}(\beta) = \sum_{i=1}^{n}\left(\sum_{j=1}^{p}\int_{-\infty}^{\infty}\beta^{(j)}(t)x_i^{(j)}(t)dt - y_i\right)^2$$

$$\mathcal{R}(\beta) = \sum_{j=1}^{p}\left\langle\beta^{(j)}, \beta^{(j)}\right\rangle_{L^2} \qquad \text{Ridge Regression}$$

- Looking for

$$\beta^* = \text{arg min}_\beta \mathcal{F}(\beta) + \lambda\mathcal{R}(\beta)$$

- The previous cannot be handled directly:
  (i) we just have samples of $x_i^{(j)}(t)$
  (ii) $\beta^{(j)}(t)$ have infinite degrees of freedom

SAFE: Regularization

- **Objective Function**:

$$\mathcal{L}(\beta) = \mathcal{F}(\beta) + \lambda \mathcal{R}(\beta)$$

where

$$\mathcal{F}(\beta) = \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \int_{-\infty}^{\infty} \beta^{(j)}(t) x_i^{(j)}(t) dt - y_i \right)^2$$

$$\mathcal{R}(\beta) = \sum_{j=1}^{p} \left\langle \beta^{(j)}, \beta^{(j)} \right\rangle_{L^2} \qquad \text{Ridge Regression}$$

- Looking for

$$\beta^* = \text{arg min}_\beta \mathcal{F}(\beta) + \lambda \mathcal{R}(\beta)$$

- The previous cannot be handled directly:
  - (i) we just have samples of $x_i^{(j)}(t)$ $\Rightarrow$ **approximate $x_i^{(j)}(t)$**
  - (ii) $\beta^{(j)}(t)$ have infinite degrees of freedom $\Rightarrow$ **parametrize $\beta^{(j)}(t)$**

SAFE: Time Series Approximation

- We consider an approximation of the fitness function $\mathcal{L}$

$$\hat{\mathcal{L}} = \hat{\mathcal{F}} + \lambda \mathcal{R}$$

$$\hat{\mathcal{F}} = \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \int_{-\infty}^{\infty} \beta^{(j)}(t) \hat{x}_i^{(j)}(t) dt - y_i \right)^2$$

- Radial Basis function Kernel

$$\begin{aligned} \mathcal{K}(t_1, t_2) &:= e^{-\frac{(t_1 - t_2)^2}{2\omega^2}} \\ &= \sqrt{2\pi}\omega G(t_1, \omega^2; t_2) \end{aligned}$$

it follows that

$$\hat{x}_i^{(j)}(t) = \sqrt{2\pi}\omega_{(j)} \sum_{s=1}^{\mathcal{N}_{i,j}} c_{i,s}^{(j)} G(t_{i,s}^{(j)}, \omega_{(j)}^2; t)$$

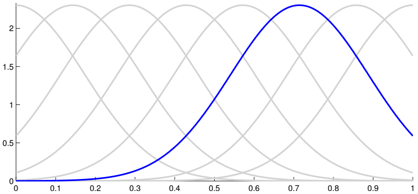
- Approximation of $x_i^{(j)}(t)$ obtained as a weighted sum of Gaussian densities

SAFE: Shape Function Parametrization

- Parametrization of $\beta^{(j)}$ as linear combination of γ Gaussian densities

$$\beta^{(j)}(t) = \sum_{k=1}^{\gamma} \alpha_k^{(j)} G(\mu(k), \sigma^2; t)$$

$$\mu(k) = \frac{k-1}{\gamma-1}$$



- The loss function then becomes

$$\hat{\mathcal{F}} = \sum_{i=1}^{N} \left[\sqrt{2\pi} \sum_{j=1}^{p} \omega_{(j)} \sum_{k=1}^{\gamma} \alpha_k^{(j)} \sum_{s=1}^{\mathcal{N}_{i,j}} c_{i,s}^{(j)} \int_{-\infty}^{\infty} \left(G(\mu(k), \sigma^2; t) G(t_{i,s}^{(j)}, \omega_{(j)}^2; t) \right) dt - y_i \right]^2$$

Theorem

Let $a, b, x \in \mathbb{R}^p$ and $A, B \in \mathbb{R}^{p \times p}$, it holds that

$$\int_{-\infty}^{\infty} G(a, A; x) G(b, B; x) dx = G(a, A + B; b)$$

SAFE: Shape Function Parametrization

$$\hat{\mathcal{F}} = \sum_{i=1}^{n} \left(\sqrt{2\pi} \sum_{j=1}^{p} \omega_{(j)} \sum_{k=1}^{\gamma} \alpha_k^{(j)} \sum_{s=1}^{\mathcal{N}_{i,j}} c_{i,s}^{(j)} G(\mu(k), \sigma^2 + \omega_{(j)}^2; t_{i,s}^{(j)}) - y_i \right)^2$$

- Defining

$$\delta_{i,s}^{(j)}(k) = \sqrt{2\pi} c_{i,s}^{(j)} \omega_j G(\mu(k), \sigma^2 + \omega_{(j)}^2; t_{i,s}^{(j)}), \qquad \overline{\delta}_i^{(j)}(k) = \sum_{s=1}^{\mathcal{N}_{i,j}} \delta_{i,s}^{(j)}(k),$$

we obtain the compact formulation

$$\hat{\mathcal{F}} = \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \sum_{k=1}^{\gamma} \alpha_k^{(j)} \overline{\delta}_i^{(j)}(k) - y_i \right)^2 = ||\Phi\theta - Y||^2$$

$$\Phi = \begin{bmatrix} \overline{\delta}_1^{(1)}(1) & \dots & \overline{\delta}_1^{(1)}(\gamma) & \overline{\delta}_1^{(2)}(1) & \dots & \overline{\delta}_1^{(p)}(\gamma) \\ \vdots & & \vdots & \vdots & & \vdots \\ \overline{\delta}_n^{(1)}(1) & \dots & \overline{\delta}_n^{(1)}(\gamma) & \overline{\delta}_n^{(2)}(1) & \dots & \overline{\delta}_n^{(p)}(\gamma) \end{bmatrix} \qquad \theta = [\alpha_1^{(1)} \; \alpha_2^{(1)} \; \cdots \; \alpha_k^{(j)} \; \cdots \; \alpha_\gamma^{(p)}]'$$

Experimental Settings

- Testing against classical feature extraction approaches
  (i) Statistical Moments
  (ii) Systematic Sampling
  (iii) PCA on Systematic Sampling
- Results based on Ridge Regression
- $n = 500$:
  70% training
  30% validation
- $p = 1$ with $[35, 45]$ samples for observation
- Noise:
  $v_{i,s}^{(j)} \sim N(0, 0.1)$
- Comparison in terms of RMSE

Experiment 1: Sinusoid Dataset

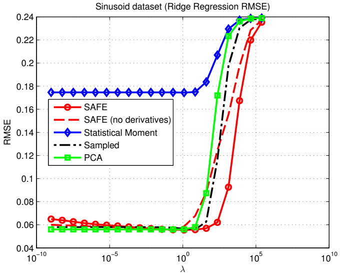
- Only an unknown part of the input signal affects the output
- Input:

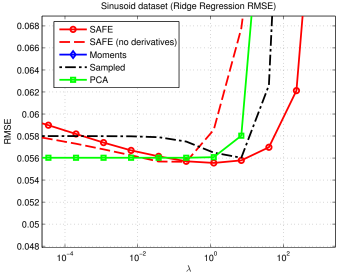
$$\begin{aligned} x(t) &= \sin(t\omega + \delta) \\ \omega &\sim \mathcal{U}(0.01, 10) \\ \delta &\sim \mathcal{U}(0, 2\pi) \end{aligned}$$

- Output:

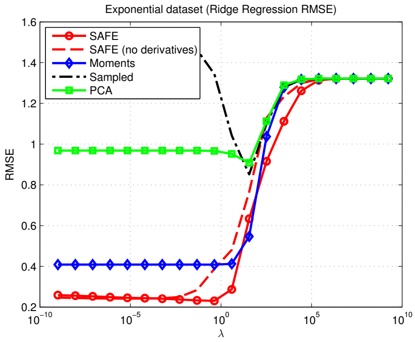
$$\begin{aligned} y &= \int_{0.3}^{0.7} x(t)dt \\ &= \frac{\cos(0.3\omega + \delta) - \cos(0.7\omega + \delta)}{\omega} \end{aligned}$$

- Statistical Moments based approach cannot filter important portions of the signal





Experiment 2: Exponential Dataset



- Target entirely explained by global features of the inputs
- Input:

$$x(t) = ae^{-bt}$$
$$a \sim \mathcal{U}(8, 12) \quad b \sim \mathcal{U}(0.5, 1.5)$$

- Output:

$$y = \int_0^1 \left(x(t) - \int_0^1 x(t)dt \right)^2 dt$$

$$\Rightarrow y = \frac{a^2\,\left(1 - e^{-2\,b}\right)}{2\,b} - \frac{a^2\,\left(e^{-2\,b} - 2e^{b} + 1\right)}{b^2}$$

- SAFE outperforms statistical moments even when the target is a global feature of the signal

VM: Publications 1/2

G.A. Susto, A. Beghi, C. De Luca
Estimating CVD Thickness through Statistical Inference Methods
Intel European Research and Innovation Conference 2010, Leixlip (Ireland), October 12-14th, 2010

G.A. Susto, A. Beghi, C. De Luca
A Virtual Metrology System for Predicting CVD Thickness with Equipment Variables and Qualitative Clustering
16th IEEE Conference on Emerging Technologies and Factory Automation, Toulouse (France), September 5-9th, 2011

G.A. Susto, A. Beghi
Least Angle Regression for Semiconductor Manufacturing Modeling
IEEE Multi-Conference on Systems and Control, Dubrovnik (Croatia), October 3-5th, 2012, pp. 658-663
Best Student Paper Winner awarded by IEEE Control Systems Society

G.A. Susto, A. Beghi
An Information Theory-based Approach to Data Clustering for Virtual Metrology and Soft Sensors
3rd International Conference on Circuits, Systems, Control, Signals, Barcelona (Spain), October 17-19th, 2012, pp. 198-203

G.A. Susto, A. Beghi
A Virtual Metrology System Based on Least Angle Regression and Statistical Clustering
Applied Stochastic Models in Business and Industry, On-line access stage, 2012

VM: Publications 2/2

G.A. Susto, A. Schirru, S. Pampuri, G. De Nicolao, A. Beghi
An Information-Theory and Virtual Metrology-based approach to Run-to-Run Semiconductor Manufacturing Control
8th IEEE International Conference on Automation Science and Engineering, Seoul (South Korea), August 20-24th, 2012, pp. 91-96
Best Student Paper Finalist awarded by IEEE Robotics and Automation Society

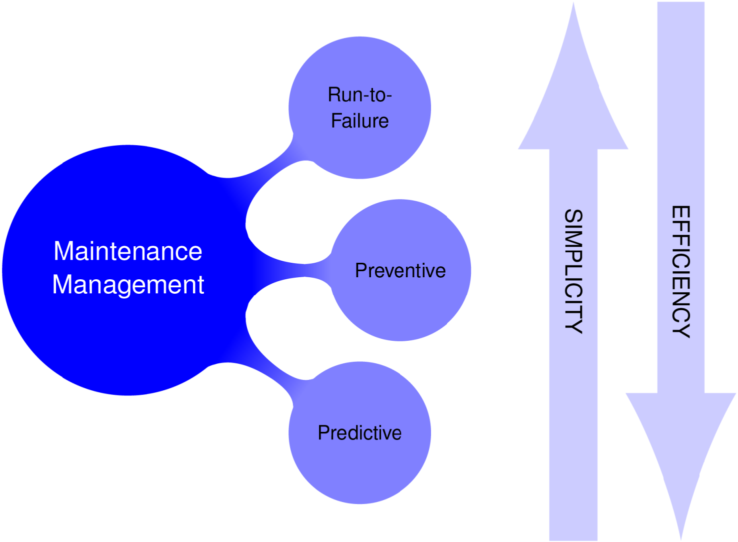
S. Pampuri, A. Schirru, G.A. Susto, G. De Nicolao, A. Beghi, C. De Luca
Multistep Virtual Metrology Approaches for Semiconductor Manufacturing Processes
8th IEEE International Conference on Automation Science and Engineering, Seoul (South Korea), August 20-24th, 2012, pp. 358-363

A. Schirru, G.A. Susto, S. Pampuri, S. McLoone
Learning from Time Series: Supervised Aggregative Feature Extraction
51st IEEE Conference on Decision and Control, Maui (US), December 10-14th, 2012, pp. 5254-5259

A. Schirru, G.A. Susto, S. Pampuri, S. McLoone
Supervised Aggregated Feature Extraction for Functional Regression in Time Series Space
IEEE Transactions on Neural Networks and Learning Systems, Under Review

G.A. Susto, A. Schirru, S. Pampuri, S. McLoone
Enhanced Virtual Metrology with Time Series Data
Intel European Research and Innovation Conference 2012, Leixlip (Ireland), October 3-4th, 2012

PdM Intro: Maintenance Policies



PdM Intro: Maintenance Policies

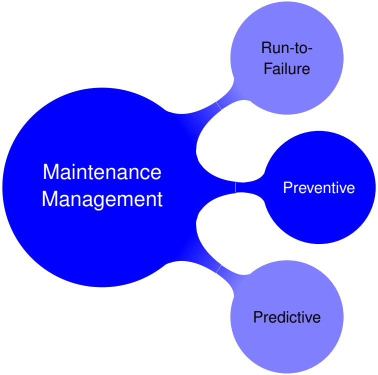
1) Run-to-Failure Maintenance - R2F

When repairs or restore actions are performed only after the occurrence of a failure

“If it’s not broken don’t fix it”

Common Policy in the fabs

PdM Intro: Maintenance Policies

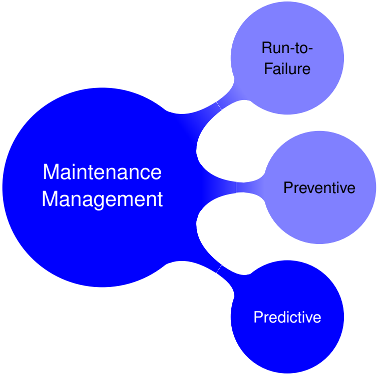


2) Preventive Maintenance

Maintenances carried out on a planned schedule with the aim of anticipating the process failures

Failures are usually warded off but unnecessary maintenances are performed

PdM Intro: Maintenance Policies



3) Predictive Maintenance - PdM

Maintenance actions are taken after the verification of conditions indicating the degradation of the process/equipment. A *PdM* system predicts when such actions have to be taken

Proposed Policy

Predictive Maintenance (PdM): Techniques

- While all VM problems are regression ones, for PdM, depending on the problem, several techniques may be suitable
  - **[Wu 2007]**: *regression methods* (Neural Networks, Elastic Nets)
  - **[Baly 2012]**: *classification methods* (Support Vector Machines)
  - **[Pampuri 2011]**: *survival models*
  - **[Butler 2010]**: *filtering and prediction* (Particle Filters)

S. Wu, N. Gebraeel, M. Lawley, Y. Yih
A Neural Network Integrated Decision Support System for Condition-Based Optimal Predictive Maintenance Policy
IEEE Transactions on Systems, Man and Cybernetics, (37), pp. 226-236, 2007

R. Baly, H. Hajj
Wafer Classification Using Support Vector Machines
IEEE Transactions on Semiconductor Manufacturing, (25), pp. 373-383, 2012

S. Pampuri, A. Schirru, C. De Luca, G. De Nicolao
Proportional Hazard Model with L1 Penalization Applied to Predictive Maintenance in Semiconductor Manufacturing
IEEE Conference on Automation Science and Engineering, Trieste, 24-27 Aug, 2011

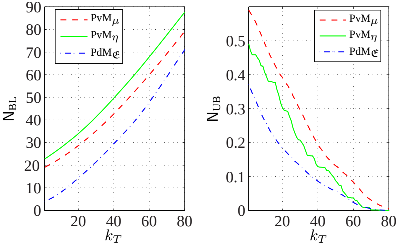
S. Butler, J. Ringwood
Particle Filters for Remaining Useful Life Estimation of Abatement Equipment Used in Semiconductor Manufacturing
IEEE Conference on Fault-Tolerant Systems, Nice, 6-8 Oct, 2010

PdM: Challenges

(1) **Problem Definition** - not standard approaches, solutions tailored for each problem

(2) **Small Amount of Observations** - maintenances are usually not many for statistical modeling

(3) **PdM Evaluation** - non-trivial evaluation of the impact of a PdM module and its advantages w.r.t. R2F and PvM approaches

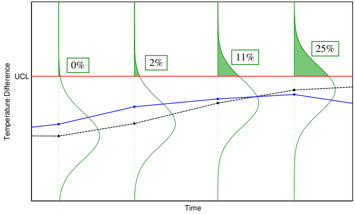


Performances must be evaluated in terms of

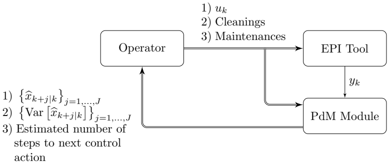
- *Type I error* - not prevented maintenances (N_{UB})
- *Type II error* - process iterations (N_{BL}) that may have been performed if the maintenance interventions suggested by the PdM would not have been performed

PdM: Filtering and Prediction for Epitaxy 1/2

- Temperature difference measured by pyrometers is a key value in Epitaxy control
- With Filtering and Prediction techniques (Kalman Predictor, Particle Filter) an estimation of temperature difference can be provided
- The PdM module provide a confidence level of control action need at next process step

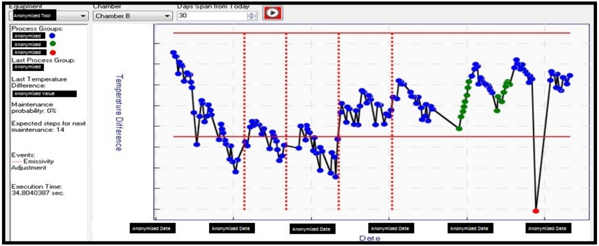


Wafer temperature (solid, blue) and prediction (dash, black).



- The end-user is provided with:
  - (i) an estimation of next values of temperature difference with confidence levels;
  - (ii) an estimation of runs to be processed before a maintenance action is needed.

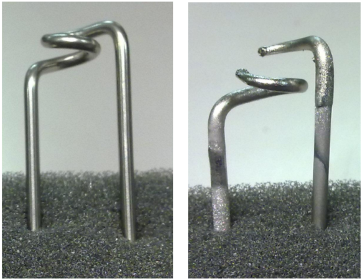
PdM: Filtering and Prediction for Epitaxy 2/2



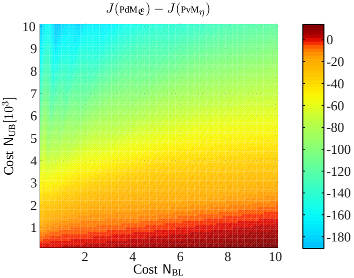
- The previous work has been developed with Infineon Technologies AG, Austria sited in Villach and with the collaboration of University of Pavia
- The PdM module, ROOME (pRedictive tOOl for Monitoring Epitaxy) has been implemented in C# in a GUI that it is currently used in production.

PdM: Regression for Ion-Implantation

- Collaboration with STMicroelectronics Italia and University of Pavia
- Breakings of Tungsten Filament is the biggest problem for maintenance engineers
- We propose a PdM module to predict filament breakings based on tool variables
- Module based on Elastic Nets to deal with high dimensionality



(a) New Filament (b) Broken Filament



PdM: Publications

- G.A. Susto, A. Schirru, S. Pampuri, A. Beghi
  A Predictive Maintenance System based on Regularization Methods for Ion-Implantation
  23rd IEEE/SEMI Advanced Semiconductor Manufacturing Conference, pp.175-180, Saratoga Springs (US), May 15-17th, 2012
  Best Student Paper Winner awarded by the Semiconductor Equipment and Materials International
- G.A. Susto, A. Beghi, C. De Luca, M.Holzinger, M.Huber
  A Predictive Maintenance System for Epitaxy Process
  Intel European Research and Innovation Conference 2010, Leixlip (Ireland), October 12-14th, 2010
- G.A. Susto, A. Beghi, C. De Luca
  A Statistical Approach for Maintenance Management in Semiconductor Manufacturing Processes
  Workshop on Statistical Methods Applied in Microelectronics, Milan (Italy), June 13th, 2011
- G.A. Susto, A. Beghi, C. De Luca
  A Predictive Maintenance System for Silicon Epitaxial Deposition
  IEEE Conference on Automation Science and Engineering, 262-267, Trieste (Italy), Aug. 24-27th, 2011
  Best Student Paper Winner awarded by the Robotics and Automation Society
- G.A. Susto, S. Pampuri, A. Schirru, A. Beghi
  Optimal Tuning of Epitaxy Pyrometers
  23rd IEEE/SEMI Advanced Semiconductor Manufacturing Conference, pp. 294-299, Saratoga Springs (US), May 15-17th, 2012
- G.A. Susto, A. Beghi, C. De Luca
  A Predictive Maintenance System for Epitaxy Processes
  IEEE Transactions on Semiconductor Manufacturing, vol. 25(4), pp. 638–649, 2012

Thank you for your attention !

Statistical Methods for Semiconductor Manufacturing

Gian Antonio Susto

PhD School on Information Engineering
XXV Series, ICT Section

Advisor: Prof. Alessandro Beghi