Application of Virtual Metrology Techniques to combine Process Information from DOEs and individual Experiments performed during Equipment Assessment

Georg Roeder[1], Juergen Niess[2], Martin Schellenberger[1], Alexander Gschwandtner[2], Lothar Pfitzner[1], Rudolf Berger[3], Raimund Förg[3,4]

[1] Fraunhofer Institute for Integrated Systems and Device Technology IISB, Erlangen, Germany
[2] HQ-Dielectrics GmbH, Dornstadt, Germany
[3] Infineon Technologies AG, Wernerwerkstrasse 2, Regensburg, Germany
[4] Technische Hochschule Deggendorf, Edlmaierstrasse 6 and 8, Deggendorf, Germany

Application of Virtual Metrology Techniques to combine Process Information from DOEs and individual Experiments performed during Equipment Assessment

- Introduction
  - Overview of the SEAL project
  - Virtual Metrology for equipment assessment
- Assessment of low-temperature microwave plasma oxidation
  - Principle of the low-temperature microwave plasma oxidation
  - Overview of the plasma reactor
  - Summary of the investigations and results
- Approaches to combine process information and results
  - Overview of the DOEs and individual experiments
  - Combination of DOEs
  - Approaches in VM development
  - Combination of individual experiments applying VM techniques
- Summary

Introduction
Overview of the SEAL project

SEAL: Semiconductor Equipment Assessment Leveraging Innovation

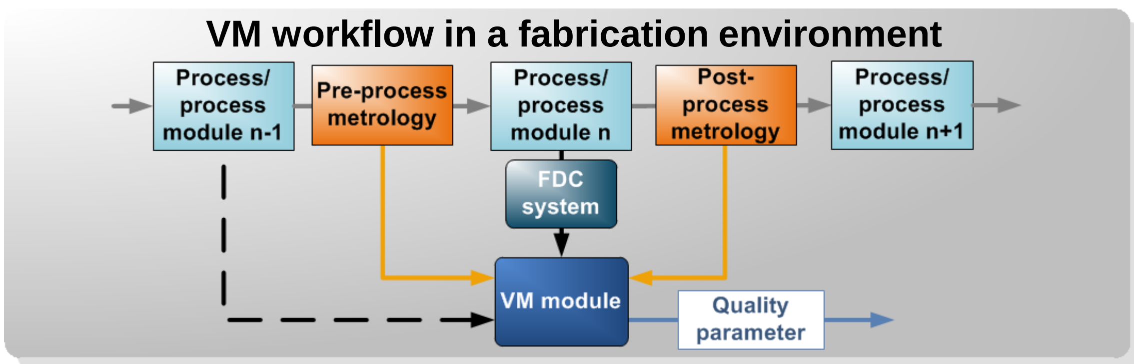
- Assessment of prototype equipment and novel enhancements to existing equipment, and their application to next generation semiconductor technologies and device architectures (http://www.seal-project.eu/)
- Funding, duration: EU FP7 Framework Programme, June 2010 – May 2013
- Partners: 8 end users, 19 equipment suppliers, 6 research institutes

Main process themes	• EUV masks, cleaning front end und back end, lithography (optics and multi e-beam), **low temperature oxidation**, bonding of thin wafers, plasma immersion ion implantation
Key metrology equipment	• Full wafer and multi column e-beam inspection, life time measurements, nano-topography, EDS/EDX, mass metrology, acoustic microscopy, overlay metrology
Cross-cut R&D	• **Equipment and process characterisation incl. virtual metrology** and predictive maintenance, equipment simulation, equipment automation, generic equipment topics and assessment, training

Introduction
Virtual Metrology for equipment assessment

Virtual Metrology (VM) objectives

- Prediction of post process quality parameters using process and wafer state information including upstream metrology data
- Support of metrology operations, FDC, RtR control, and predictive maintenance
- VM is currently a topic at fab level



Motivation for VM at the equipment level

- Gaining better understanding of unit processes is at the outset of VM development
- → Improved use of knowledge gained from equipment development and assessment
- → Provide basic process models for the benefit of the equipment supplier and IC manufacturer

Application of Virtual Metrology Techniques to combine Process Information from DOEs and individual Experiments performed during Equipment Assessment

- Introduction
  - Overview of the SEAL project
  - Virtual Metrology for equipment assessment
- **Assessment of low-temperature microwave plasma oxidation**
  - Principle of the low-temperature microwave plasma oxidation
  - Overview of the plasma reactor
  - Summary of the investigations and results
- Approaches to combine process information and results
  - Overview of the DOEs and individual experiments
  - Combination of DOEs
  - Approaches in VM development
  - Combination of individual experiments applying VM techniques
- Summary

Assessment of low-temperature microwave plasma oxidation
Principle of the low-temperature microwave plasma oxidation

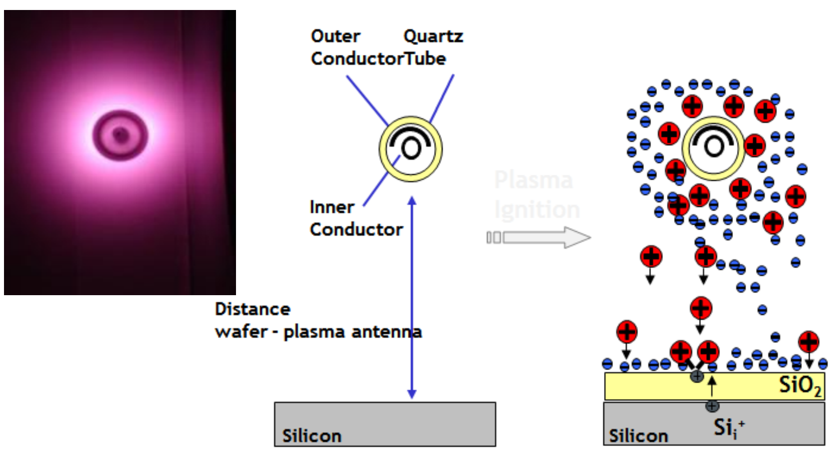
Motivation

- Thermal oxidation is generally not scalable and needs a high thermal budget
- Low thermal budget required for novel logic device structures (FINFET)
- Low thermal budget required for 450 mm wafer size (e.g. dislocation & distortion)

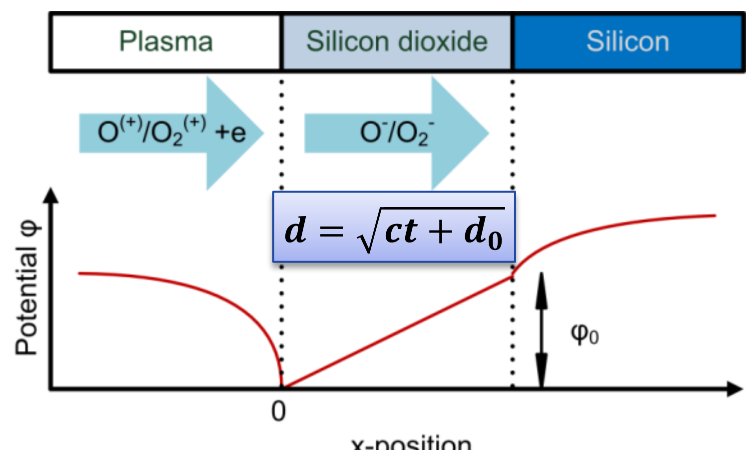
Reaction principle

- Low reaction temperature (T ≤ 400 °C) by microwave (MW) plasma-enhanced oxidation

Plasma generation at a planar MW antenna vs. a floating Si-substrate



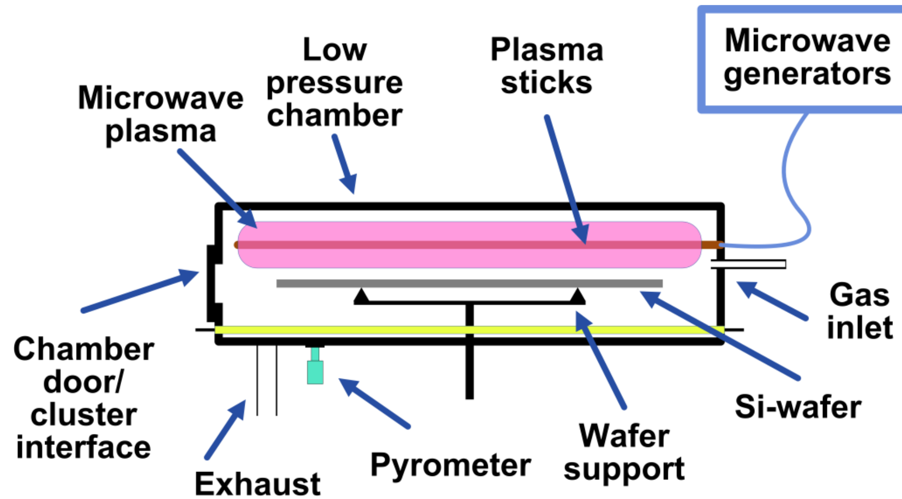
Field-enhanced diffusion of charged oxygen species through the silicon dioxide layer



Assessment of low-temperature microwave plasma oxidation
Overview of the plasma reactor

Microwave plasma oxidation module installed at the Fraunhofer IISB

Microwave plasma chamber details



- Reactor properties: High oxide growth rate with excellent uniformity and very low plasma damage compared to radicals oxidation and RF-based oxidation
- Selective oxidation on W vs. Si

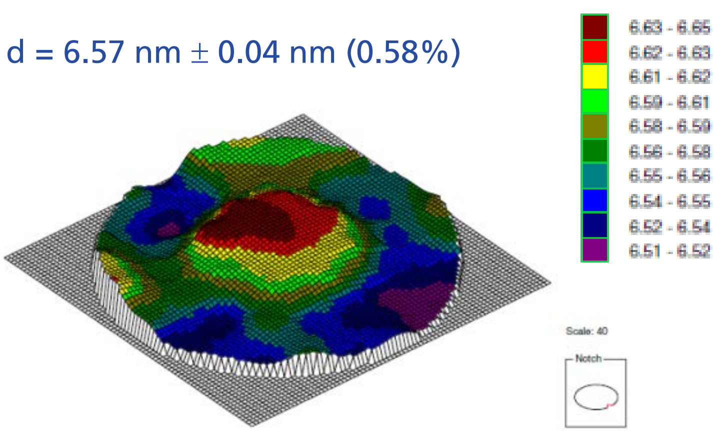
Assessment of low-temperature microwave plasma oxidation
Summary of the investigations and results

Investigations
- Optimization of low-temperature MW oxidation
- Extensive characterization of MOS cap test devices
- Comparison of plasma oxide vs. thermal oxide on regular product at Infineon

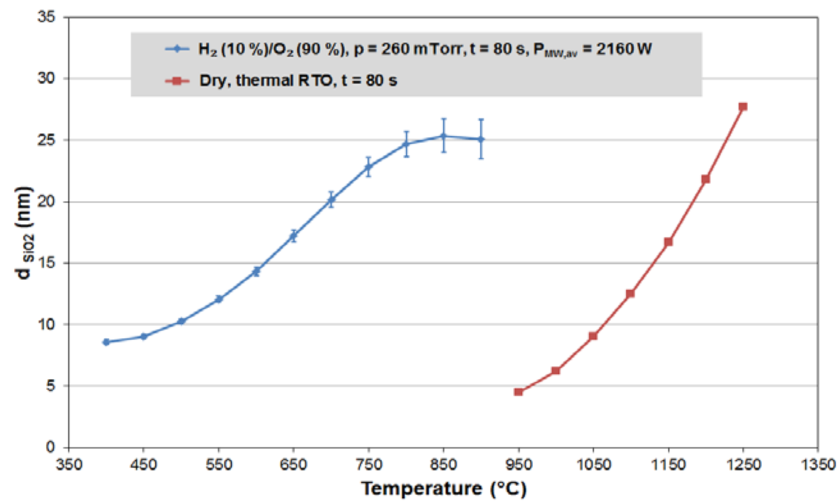
Results
- Excellent SiO_2 thickness uniformity and quality
- Electrical parameters comparable to furnace oxides

Excellent SiO_2 layer thickness uniformity on 300 mm wafer

d = 6.57 nm ± 0.04 nm (0.58%)



Oxidation Temperature ~ 500 °C lower than RTO



HQ-Dielectrics

Application of Virtual Metrology Techniques to combine Process Information from DOEs and individual Experiments performed during Equipment Assessment

- Introduction
  - Overview of the SEAL project
  - Virtual Metrology for equipment assessment
- Assessment of low-temperature microwave plasma oxidation
  - Principle of the low-temperature microwave plasma oxidation
  - Overview of the plasma reactor
  - Summary of the investigations and results
- **Approaches to combine process information**
  - Overview of the DOEs and individual experiments
  - Combination of DOEs
  - Approaches in VM development
  - Combination of individual experiments applying VM techniques
- Summary

Approaches to combine process information and results
Overview of the DOEs and individual experiments

Overview of the DOEs

- Investigation and adjustment of thickness (d_{av}) and homogeneity σ_d for MOS cap oxides (4 nm, 7 nm, 12 nm, 20 nm) by DOEs
- Completion of 3 individual designs for different gas mixtures ($P_{MW, av}$ = 2190 W)

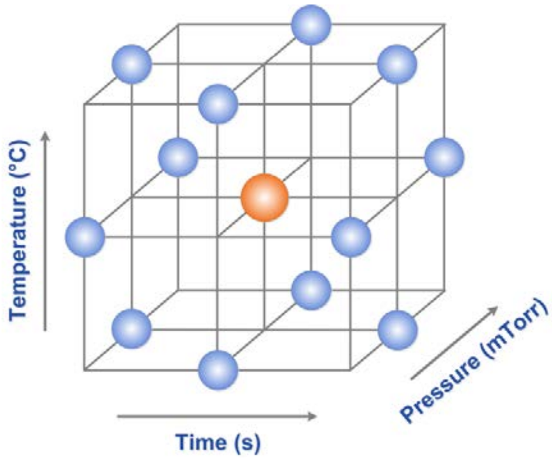
Individual experiments

- Exploit limits of the parameter space, detailed investigation of temperature dependence (experiments at $P_{MW, av}$ = 1095 W, T: RT to 900°C, t: 10 s to 560 s)

Overview of the factor levels in the DOEs

Factor levels	t (s)	T (°C)	p (mTorr)
High	160	800	200
Center	80	600	260
Low	20	400	320
Design applied for	$H_2/He/O_2$ (10 %/ 50 % / 40 %), H_2/O_2 (10 % / 90 %), O_2 (100 %)		

Box-Behnken design factor level arrangement



Approaches to combine process information and results

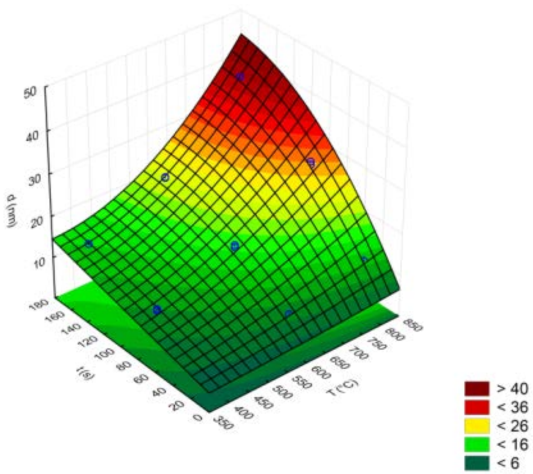
Main results of the DOEs and individual experiments

Results summary

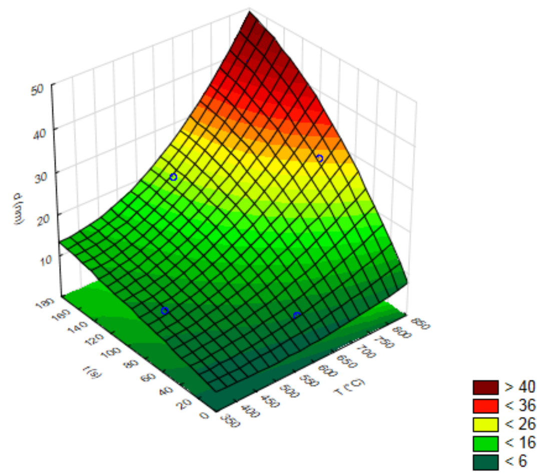
- d_{av} > 4 nm to 50 nm achievable in process window; σ_d < 1% achievable
- Similar dependence of d_{av} for different gas mixtures; individual dependence of σ_d
- Typically, additional MW power tuning required to further minimize σ_d

Investigation of thickness for different gas mixtures

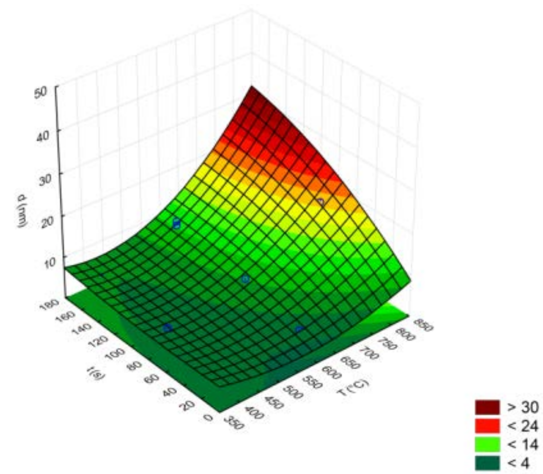
$H_2/He/O_2$, p = 200 mTorr



H_2/O_2, p = 200 mTorr



O_2, p = 200 mTorr



- Combination of the individual DOEs: Joint assessment of differences in d_{av} for the gas mixtures

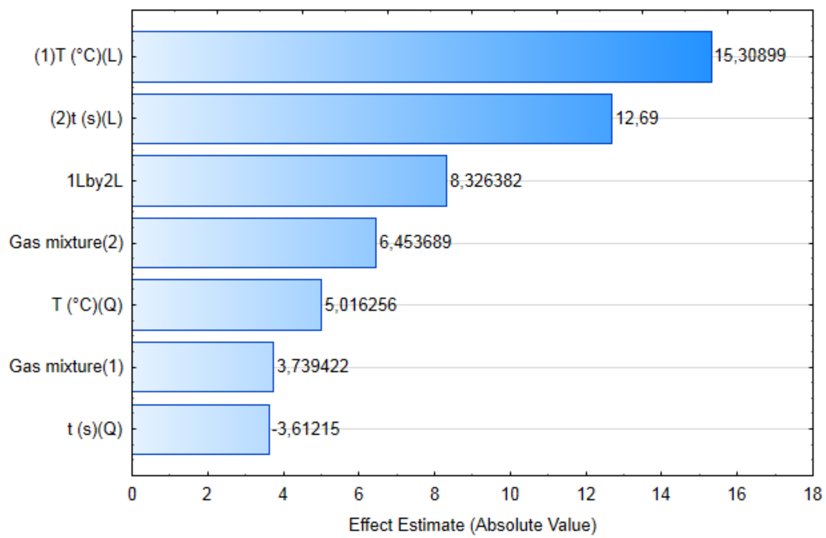
Approaches to combine process information and results
Combination of individual DOEs

Approach and results

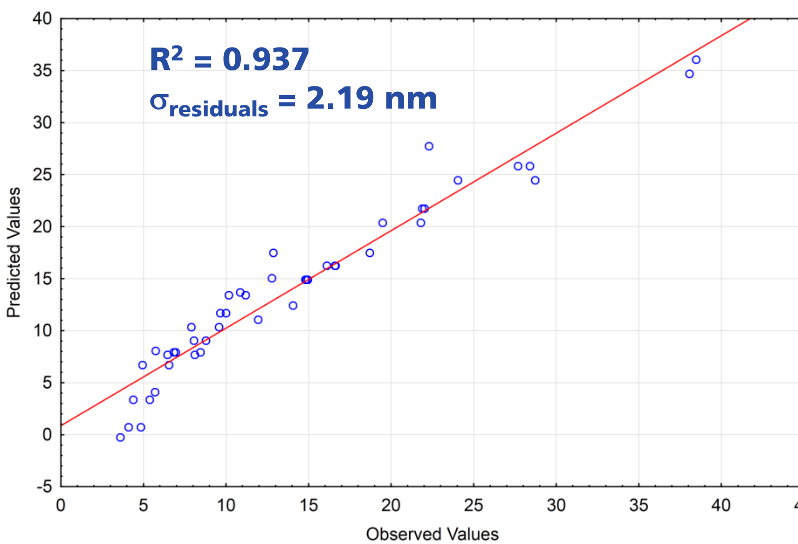
- Combination of the individual DOEs modeling the gas mixtures as blocks
- Analysis using a quadratic model with linear interactions

Investigation of parameter effects and regression modeling

Estimation of relevant parameter effects



Regression modeling with selected predictors



→ Assessment of joint effects and regression modeling using all parameters possible

HQ-Dielectrics Infineon technologies Fraunhofer IISB

Approaches to combine process information and results
Approaches for VM development (1)

VM model objectives

- Development of a statistical model from a sample of training data
- Application of the model to predict a process quality variable for a new data vector in the same process

VM model development regression

- Find function $g(X)$ to approximate the true function $f(X)$ to describe the process output Y

Metrology data

Equipment data

Error

No. of processes

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = g\left(\begin{bmatrix} x_{11} & \cdots & x_{1D} \\ x_{21} & & x_{2D} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{ND} \end{bmatrix}\right) + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_D \end{pmatrix}$$

Experiment	T (°C)	p (mTorr)	Mittelwert P(W)	t (s), Plasmabrenn dauer	O2 (%)	H2 (%)	He (%)	d_av (nm)
DOE HeH2O2	400	260	2190	20	50	10	40	5.735
DOE HeH2O2	400	260	2190	160	50	10	40	14.06
DOE HeH2O2	400	200	2190	80	50	10	40	10
DOE H2O2	400	260	2190	20	90	10	0	4.946

Approaches to combine process information and results
Approaches for VM development (2)

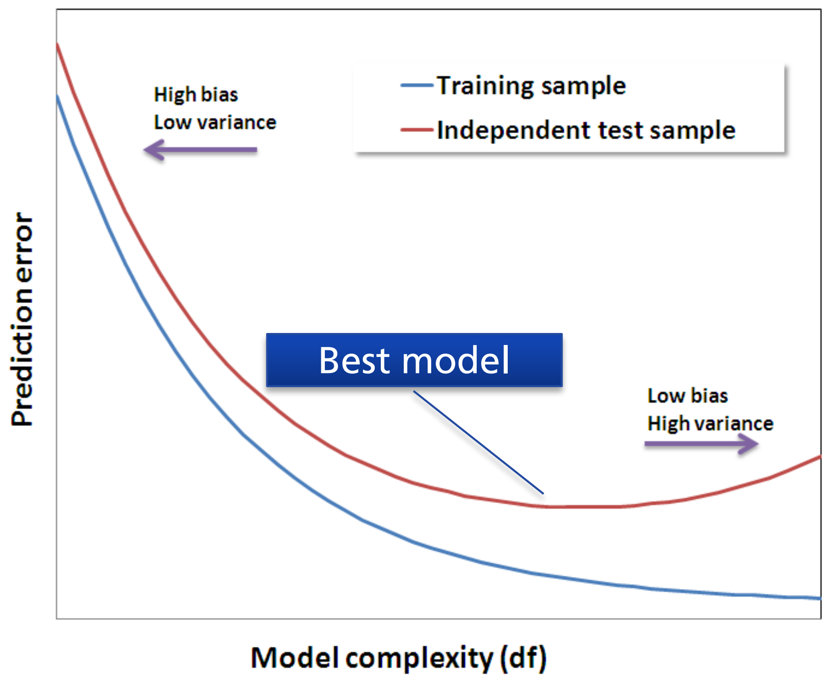
Approaches to obtain stable models

- Selection of a variable subsets (orthogonal variable vectors)
- Regularization by variable ranking or elimination
- Regularization by penalization of the model

Investigated models

- Stepwise regression
- Lasso regression
- Stochastic gradient boosting trees

Bias-Variance tradeoff vs. model complexity

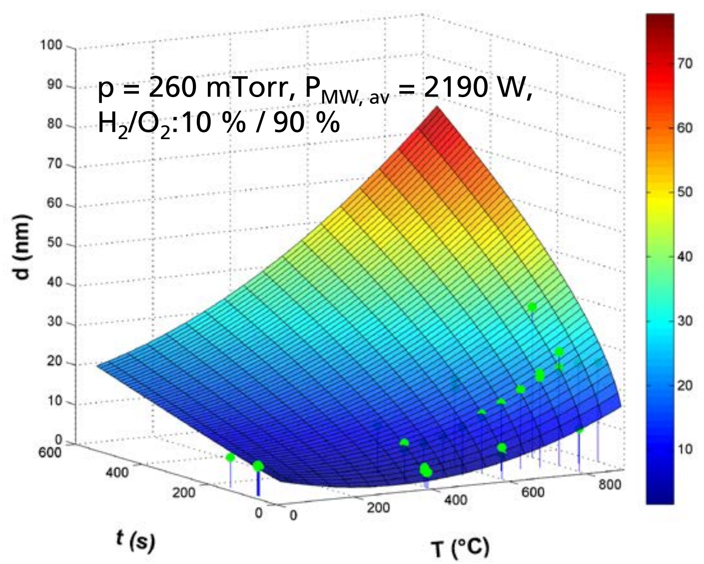


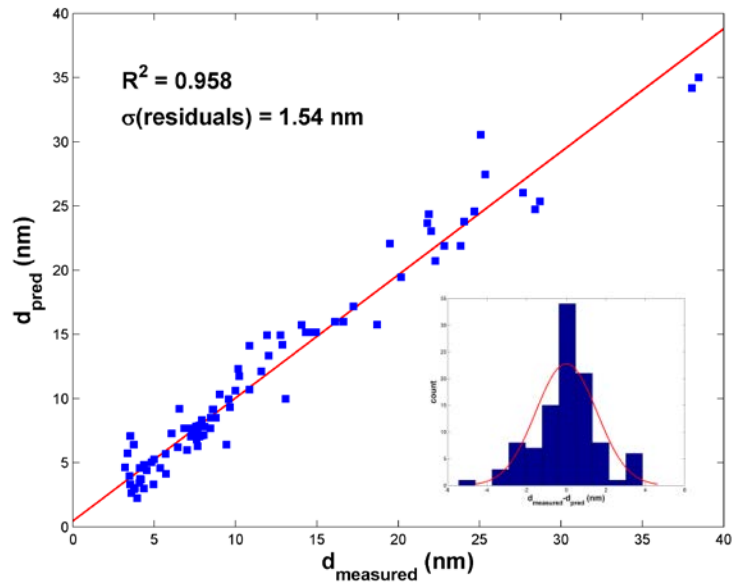
Combination of individual experiments applying VM techniques
Stepwise regression – overview

Stepwise regression properties

- Parameter selection method (parametric) adding and removing terms from a multilinear model based on their statistical significance in a regression
- Start with initial model and comparison of explanatory power of incrementally larger and smaller models

First overview on complete data set (transformation of $\tau = t^{1/2}$)





HQ-Dielectrics Infineon technologies Fraunhofer IISB

Combination of individual experiments applying VM techniques

Stepwise regression – application for prediction

Stepwise regression application

- Selection of first important parameters possible from complete data set
- Parameter selection depends on selected sample subset
- Selection of parameters applying data subsets and 10fold cross-validation

Stepwise regression results applying the full data set and 10fold cross-validation

Data	Relevant parameters	R^2	$\sigma_{residuals}$ training data	$\sigma_{residuals}$ test data
Full data set	T, τ, $[H_2]$, T*P, T*τ, T*$[H_2]$, p*[He], τ*$[H_2]$, $[O_2]$*$[H_2]$, T^2, τ^2	0.958	1.54 nm	-
10fold cross-validation	T, τ, $[H_2]$, T*P , T*τ , T*$[H_2]$, p*[He], τ*$[H_2]$, $[O_2]$*$[H_2]$, T^2, τ^2, p, T*$[O_2]$, p^2 , P*τ, P*$[O_2]$, $[O_2]^2$	0.959 $\pm$ 0.004	1.54 nm $\pm$ 0.06 nm	2.05 nm $\pm$ 0.40 nm

→ Application of regression model for prediction possible

→ Parameters selection from different data subsets difficult

→ More formal method preferred

Combination of individual experiments applying VM techniques

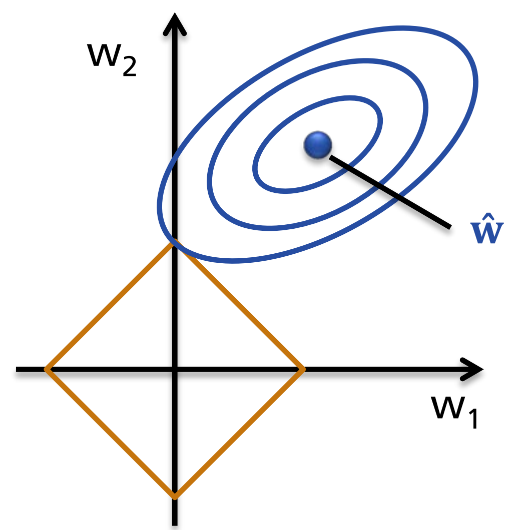
Lasso regression – overview

Lasso regression properties

- Parameter shrinkage (parametric) method achieved by adding a penalty term to the regression coefficients
- Coefficient estimates are shrunk towards zero forcing some coefficient estimates to be exactly equal to zero if the tuning parameter λ is sufficiently large
- Lasso regression provides parameter selection

Lasso regression and parameter shrinkage

$$\hat{w}_{lasso} = \min_{w} \{ \underbrace{\frac{1}{2} \sum_{i=1}^{N} (y_i - w_0 - \sum_{j=1}^{P} x_{ij} w_i)^2}_{\text{Coefficient estimation}} + \underbrace{\lambda \sum_{j=1}^{P} |w_i|}_{\text{Penalization}} \}$$



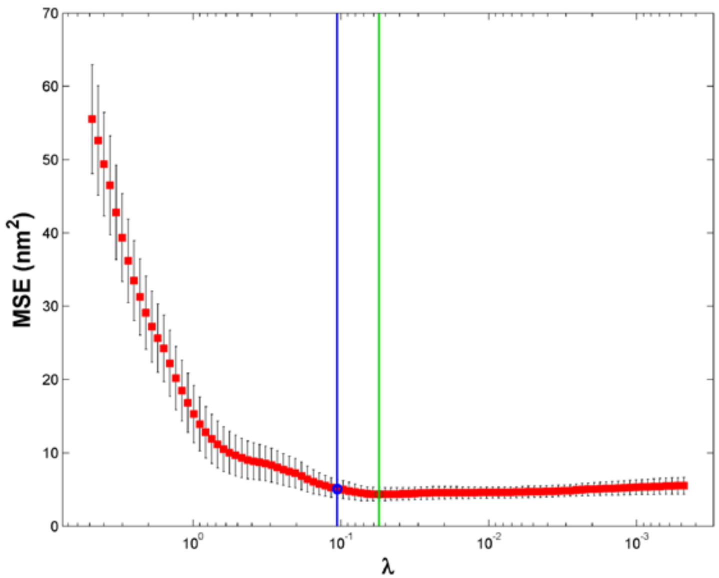
Combination of individual experiments applying VM techniques

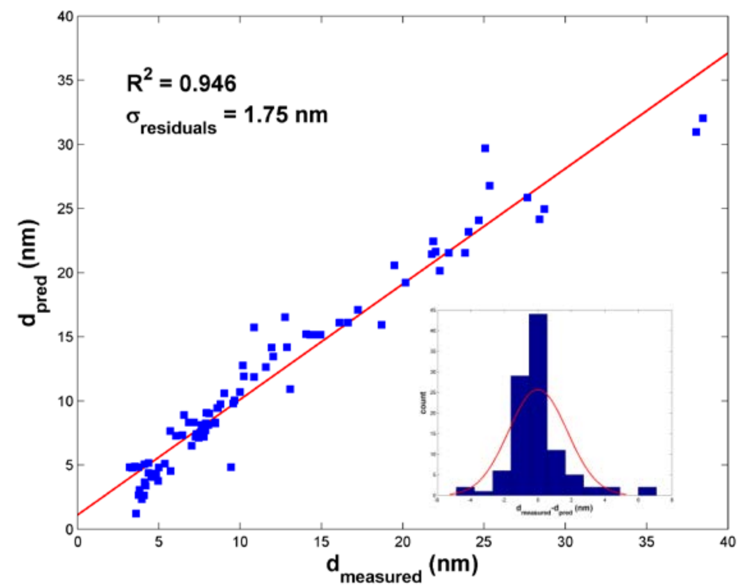
Lasso regression – application for prediction

Determination of penalty term and regression coefficients

- Lasso regression fit with 10-fold cross validation for different λ
- Determine optimum lambda at minimum average ($MSE + \sigma_{MSE}$) of test sample

Parameter estimation by Lasso regression





→ Lasso fit reveals slightly less R^2 than stepwise regression; overall 18 model coefficients are determined (4 different from stepwise regression)

Combination of individual experiments applying VM techniques
Boosted regression trees – overview

Principle of boosted trees (BT)

- Ensemble learning: improve performance of a single model by fitting and combining many simple models
- BT applies two algorithm types:
  - algorithms for classification and regression trees
  - boosting algorithm to build a collection of models

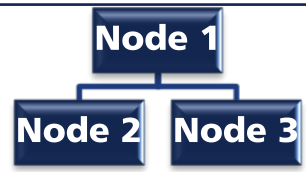
Advantages

- Applicability to classification and regression problems
- Accommodation of continuous and categorical predictors
- Tree methods are nonparametric and nonlinear
- Inherent variable ranking
- Improved predictive performance vs. C&RT

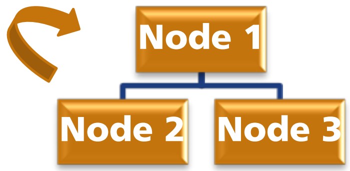
Limitations

- Model is complex and cannot be visualized like a single tree

Principle of BT



Initialization:
- Draw random sample
- Build tree on actual data and predict values



Iteration:
- Draw random sample
- Build tree on pseudo residuals introducing gradient, adjust prediction, minimize on pseudo error function
- Update model adjusted by learning rate
- Repeat step on new random sample

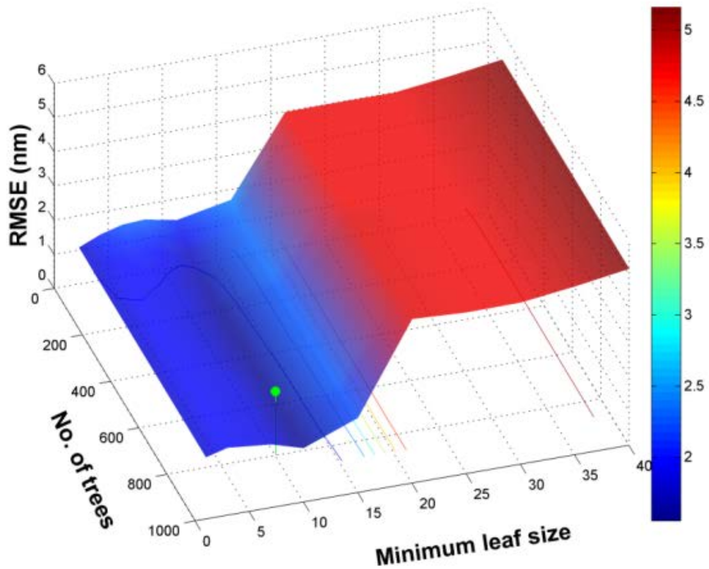
Output of final model

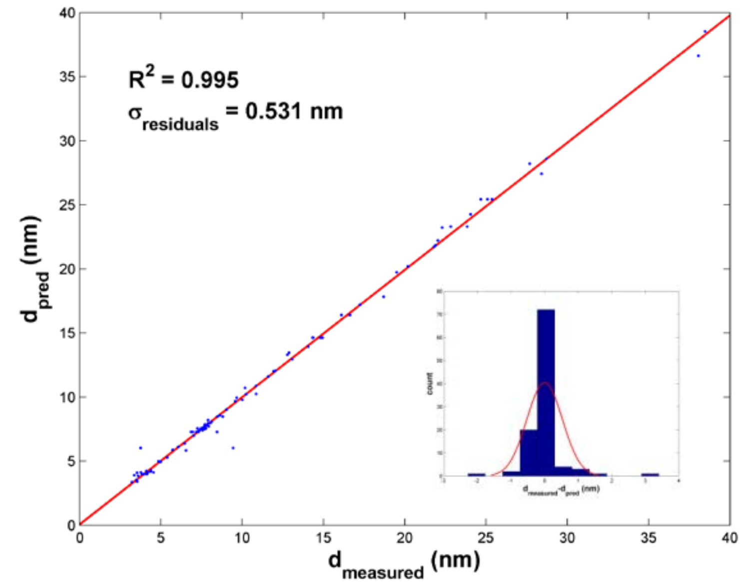
Combination of individual experiments applying VM techniques
BT regression – application for prediction

BT regression application

- Regression is performed on matrix with initial parameters (not design matrix)
- BT parameters (number of trees and tree size) are determined from training samples vs. test samples applying 10-fold cross validation

Optimization of tree parameters and regression by stochastic gradient boosting





➔ Stochastic BT regression provides accurate modeling

HQ-Dielectrics Infineon technologies Fraunhofer IISB

Combination of individual experiments applying VM techniques
Comparison of VM methods

Result summary

- All methods identify relevant parameters and provide good regression and prediction results on the training and test sample, respectively
- Variable selection in stepwise regression and lasso identify different parameters
- BT provides best regression results using original data as non parametric method

Model comparison applying the full data set and 10fold cross-validation

Data	Relevant parameters	R^2	$\sigma_{residuals}$ training data	$\sigma_{residuals}$ test data
Stepwise regression	T, τ , $[H_2]$, T*τ , T*$[O_2]$, T*$[H_2]$, p^2, p*[He], P*τ, P*$[O_2]$, τ*$[H_2]$, $[O_2]$*$[H_2]$, T^2, τ^2, p, $[O_2]^2$, T*P	0.959 ± 0.004	1.54 nm ± 0.06 nm	2.05 nm ± 0.40 nm
Lasso regression	T, τ, $[H_2]$, T*τ, T*$[O_2]$, T*$[H_2]$, p^2, p*[He], P*τ, P*$[O_2]$, τ*$[H_2]$, $[O_2]$*$[H_2]$, T^2, T*p, P*$[H_2]$, τ*$[O_2]$, τ*[He], $[He]^2$	0.946 ± 0.002	1.75 nm ± 0.07 nm	1.68 nm ± 0.52 nm
BT regression	T, p, P, τ, $[O_2]$, $[H_2]$, [He]	0.995 ± 0.001	0.54 nm ± 0.07 nm	1.43 nm ± 0.57 nm

Summary

- Virtual Metrology provides techniques to make available joint information from experiments at the early stage of equipment assessment
- Virtual Metrology techniques can build on information obtained from designs of experiments and from individual experiments
- VM provides the path from regression on the complete data set towards predictive modeling
- Stepwise regression, lasso regression and stochastic gradient boosting discussed as parametric and non parametric techniques

→ Improved use of knowledge gained from equipment development and assessment demonstrated

→ Basis for process models for the benefit of the equipment supplier and IC manufacturer

Thank you for your attention!

www.iisb.fraunhofer.de

Acknowledgment

The presented work was supported within the SEAL project, which was funded by the European Commission under contract number IST-257379. The partners are grateful for the support of the EC.