

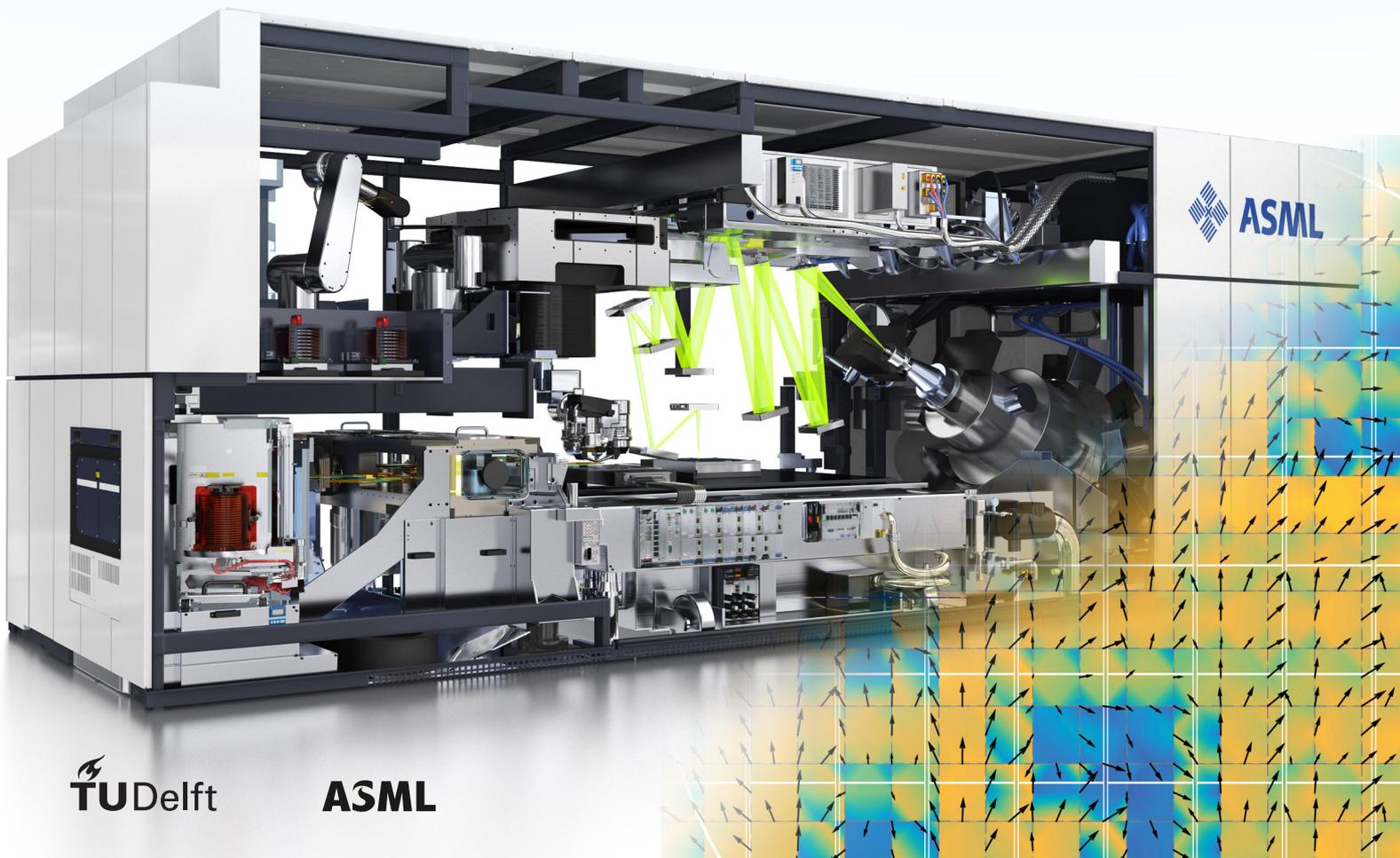
Department of Precision and Microsystems Engineering

Optimal Experiment Design for Improved Parameter Estimation in Thermo-Mechanical Feedforward Models

S.M. Hekner

Report no. : EM 2016.010  
Coach : Dr. ir. G. Van der Veen & Dr. ir. S. Van der Meulen  
Professor : Prof. dr. ir. F. Van Keulen  
Specialisation : Mechatronic System Design  
Type of report : MSc. Thesis  
Date : 8 March 2016

CONFIDENTIAL





# Optimal Experiment Design for Improved Parameter Estimation in Thermo-Mechanical Feedforward Models

by

S.M. Hekner

to obtain the degree of Master of Science  
at the Delft University of Technology,  
to be defended publicly on Tuesday March 8, 2016 at 02:00 PM

Student number: 4013735  
Project duration: May 1, 2015 – February 29, 2016  
Thesis committee: Prof. dr. ir. F. Van Keulen, TU Delft, chairman  
Dr. ir. G. Van der Veen, TU Delft & MI-Partners, supervisor  
Dr. ir. S. Van der Meulen, ASML, supervisor  
Dr. ir. R. A. J. Van Ostayen, TU Delft  
Dr. D. J. P. Lahaye, TU Delft

*This thesis is confidential and cannot be made public up to and including March 7, 2021.*





# Preface

For my graduation as Mechanical Engineer from the Technical University of Delft, I wanted to look out for a challenging topic. I succeeded my search by ASML, who was able to provide me an intriguing assignment. This thesis is the result which I will enjoy to present.

Firstly and foremost, I want to express my gratitude to my supervisors, Stan van der Meulen and Gijs van der Veen, for all the guidance, the constructive feedbacks, interesting discussions and the ability to bring out the best. Without their help and support, this thesis would not be the same as it is today.

Next, I would like to thank all the employees of ASML who helped me during this project. You all took your time to help me with my project, as well as the administrative problems, for which I am grateful.

In addition, I want to thank ASML for giving me the opportunity to work on such an interesting problem and to experience the business environment.

Finally, I want to thank my family for all the support during my graduation and my study in general. You have always been patient and encouraging.

*Sebastiaan Hekner  
Delft, March 2016*



# Abstract

ASML's wafer scanners are crucial machines in the production of integrated circuits (ICs). An important performance parameter in these photo-lithographic machines is the so-called *overlay error*, which is a measure for the accurate stacking of multiple layers during the photo-lithography process in the  $xy$ -plane. The overlay performance is, amongst others, hampered by thermal disturbances. Currently, a thermo-mechanical feedforward model is used to predict the heating effects and, subsequently, this information is used to compensate for overlay (substrate deformation) by active control.

Some of the model parameters, used in the thermo-mechanical feedforward, are not accurately known which result in inaccurate overlay compensation. The model parameter accuracy can be improved by using experiments subjected to model parameter calibration. Unfortunately, model parameter calibration may be hampered when choosing an insufficient experimental setup. For example, it can lead to severe model parameter correlation. One way to improve overlay performance is to calibrate the model parameters of the thermo-mechanical feedforward model by means of an optimal conducted experiment.

The information matrix is used to capture the information content of an experiment. This matrix is the key ingredient in order to investigate the performance of an experiment when subjected to model parameter calibration. By using competent objective functions, the information content of the information matrix can be captured by one single expression. Subsequently, this single expression can be used to perform optimization of the experiment.

Two experimental cases of the thermo-mechanical feedforward model are optimized using two different objective functions. Case 1 contains the heat capacitance  $c_p$  of the substrate and the IR/EUV ratio as model parameters, and case 2 contains the tangential burl stiffness  $G_{zy}$  and the IR/EUV ratio as model parameters. A D-optimality objective function, which is the determinant of the information matrix, and an ACE1-optimality objective function, which focusses on the correlation between specified model parameter and the eigenvalues of the information matrix associated to the same model parameters. The Simulated Annealing algorithm is used to perform the complex experiment optimization problem that originates by the characteristics of the feedforward model. Eventually, it is possible to acquire an optimal experiment which can increase model parameter accuracy and, therefore, give a reduction of worst-case overlay error by 54.4% and 9.4% when considering a 95% confidence bounce for model parameter case 1 and 2, respectively.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation and Problem Definition . . . . .	2
1.2	Structure of the Thesis . . . . .	8
<b>I</b>	<b>Literature and Analysis Tools</b>	<b>9</b>
<b>2</b>	<b>Thermo-Mechanical Model of ASML</b>	<b>11</b>
2.1	Exposure Sequence . . . . .	12
2.2	Physical Representation of Model . . . . .	13
2.3	Exposure Heat Load . . . . .	13
2.4	Governing Equations . . . . .	15
2.5	Discretization . . . . .	17
2.5.1	State Space . . . . .	17
2.6	Numerical Implementation . . . . .	17
2.6.1	Scan Move . . . . .	18
2.6.2	Numerical Solution Scheme . . . . .	19
2.7	Model Assumptions . . . . .	21
<b>3</b>	<b>Sensitivity Analysis Methods</b>	<b>23</b>
3.1	Static Problems in a General Case . . . . .	23
3.1.1	Direct Sensitivity Analysis . . . . .	24
3.1.2	Adjoint Sensitivity Analysis . . . . .	25
3.2	Transient Problems . . . . .	27
3.2.1	Direct Sensitivity Analysis in Discrete Time . . . . .	27
3.2.2	Adjoint Sensitivity Analysis in Discrete Time . . . . .	28
3.3	Conclusion . . . . .	29
<b>4</b>	<b>State of the Art Experiment Design</b>	<b>31</b>
4.1	Mathematical Model . . . . .	31
4.2	Sensitivity of Responses to Parameters . . . . .	32
4.3	Variance-Covariance of Parameters . . . . .	32
4.4	Response and Measurement Errors . . . . .	36
4.5	Experiment Design . . . . .	37
4.5.1	Standard Metrics to Measure the Information Content of an Experiment . . . . .	37
4.5.2	Alternative Formulations Based on Standard Experiment Design Criteria . . . . .	43
4.5.3	Association Between Eigenvalues and Parameters . . . . .	44
4.5.4	Newly Developed Metrics for Information Content Measurement . . . . .	45

<b>5</b>	<b>Nonlinear Regression Analysis</b>	<b>51</b>
5.1	Linear Regression . . . . .	51
5.1.1	Least Squares Estimates . . . . .	52
5.1.2	Lack of Fit . . . . .	58
5.2	Nonlinear Regression . . . . .	60
5.2.1	Least Squares Estimates . . . . .	61
5.2.2	Profiling . . . . .	66
<b>6</b>	<b>Preliminary Calibration Investigation</b>	<b>75</b>
6.1	Sensitivity Analysis . . . . .	75
6.1.1	Thermal Parameter Variation . . . . .	76
6.2	Thermal Parameter Distinguishability from Overlay Measurements . . . . .	78
6.3	Thermal Parameter Subset Selection . . . . .	79
6.4	Numerical Optimization . . . . .	80
6.4.1	Thermal Parameter Subset Optimization . . . . .	80
6.4.2	Optimization of Overlay Experiment . . . . .	83
6.5	Conclusion Preliminary Calibration . . . . .	83
<b>II</b>	<b>Application of Experiment Design</b>	<b>85</b>
<b>7</b>	<b>Simplified Thermo-Mechanical Model</b>	<b>87</b>
7.1	Finite Element Method . . . . .	87
7.1.1	Thermal Equation . . . . .	87
7.1.2	Mechanical Equations . . . . .	88
7.1.3	Thermo-Mechanical Equations . . . . .	89
7.1.4	Elements . . . . .	90
7.1.5	Galerkin Method . . . . .	91
7.2	Physical Model Representation . . . . .	92
7.2.1	Geometry and Mesh . . . . .	93
7.2.2	Burl layers . . . . .	94
7.2.3	Constraints . . . . .	95
7.2.4	Model Parameters . . . . .	96
7.3	Governing Equations . . . . .	96
7.3.1	Thermal Equations . . . . .	96
7.3.2	Mechanical Equations . . . . .	98
7.4	Heat Load . . . . .	100
7.5	Input Design Parameters . . . . .	102
7.6	Model Validation . . . . .	104
<b>8</b>	<b>Analysis of Simplified Model</b>	<b>107</b>
8.1	Sensitivity Analysis . . . . .	107
8.1.1	Validation . . . . .	109
8.1.2	Model Parameter Subset . . . . .	111
8.1.3	Sensitivities of Model Parameter Subset . . . . .	113
8.2	Model Characteristics and Behaviour . . . . .	116
8.2.1	Least Square . . . . .	117
8.3	Standard Experiments . . . . .	119
8.3.1	Routings . . . . .	119
8.3.2	Overlay Results . . . . .	120
8.4	Model Parameter Estimation . . . . .	121
8.5	Statistical Investigation . . . . .	123

8.5.1	T-profiling Standard Experiments . . . . .	123
8.5.2	Routing Performance . . . . .	125
8.6	Experiment Fingerprints . . . . .	126
8.6.1	Meander Routing . . . . .	127
8.6.2	Spiral Routing . . . . .	127
8.6.3	Jumpy Routing . . . . .	128
<b>9</b>	<b>Optimal Experiment Design</b>	<b>131</b>
9.1	Input Parameters for Experiment Design . . . . .	131
9.1.1	Simulated Annealing . . . . .	132
9.2	Experiment Optimization . . . . .	134
9.2.1	Objective Functions . . . . .	135
9.2.2	Dose Optimization Analysis . . . . .	135
9.2.3	Experiment Optimization . . . . .	138
9.3	Optimal Experiment Design Analysis . . . . .	144
9.3.1	Model Parameter Estimation . . . . .	144
9.3.2	Statistical Investigation . . . . .	145
9.3.3	Fingerprints of Optimal Experiments . . . . .	147
9.4	Efficiency Optimization Algorithm Investigation . . . . .	151
9.5	Overlay Accuracy . . . . .	152
<b>10</b>	<b>Conclusion and Recommendation</b>	<b>155</b>
10.1	Conclusion . . . . .	155
10.2	Recommendation . . . . .	156
	<b>Appendix</b>	<b>161</b>
<b>A</b>	<b>Figures In Report</b>	<b>163</b>
A.1	Least Squares Estimation Progress Standard Experiments . . . . .	163
A.2	Least Squares Estimation Progress Optimal Experiments . . . . .	164
A.3	T-profiling of Standard Experiments . . . . .	165
A.3.1	Spiral Experiment . . . . .	165
A.3.2	Jumpy Experiment . . . . .	166
A.4	T-profiling of Standard Experiments . . . . .	167
<b>B</b>	<b>Shape Functions</b>	<b>169</b>
<b>C</b>	<b>Interpolating Profile Pair Contours</b>	<b>171</b>



# List of Figures

1.1	Inside of the EUV wafer scanner ASML machine. . . . .	1
1.2	Overlay error due to substrate deformation caused by heating. . . . .	3
1.3	Position control loop with feedforward model. . . . .	4
1.4	Simplified block diagram of the system. . . . .	5
1.5	Full schematic block diagram to improve the overview of the problem. . . . .	6
2.1	Schematic overview of the thermo-mechanical feedforward model. . . . .	11
2.2	Multiple fields on a substrate. . . . .	12
2.3	Physical representation of the model. Different material layers can be seen. . . . .	13
2.4	Representation of the applied heat load. . . . .	14
2.5	Conservation of energy in an infinitesimal volume. . . . .	16
2.6	Scan move across a field . . . . .	18
2.7	Pre-defined $y$ -coordinates in a field . . . . .	19
4.1	System response with $\gamma = 0$ . . . . .	34
4.2	System response with $\gamma = \frac{1}{4}\pi$ . . . . .	35
4.3	System response with $\gamma = \frac{1}{2}\pi$ . . . . .	36
4.4	Geometrical interpretation of the standard criteria for the experiment design. . . . .	38
4.5	Confidence region for $\gamma = 0$ . . . . .	41
4.6	Confidence region for $\gamma = 1/4\pi$ . . . . .	41
4.7	Confidence region for $\gamma = 1/4\pi$ . . . . .	42
4.8	Interpretation of eigenvalues and eigenvectors. . . . .	45
5.1	Least square regression for linear model. . . . .	52
5.2	Least square geometrical interpretation. . . . .	53
5.3	Nominal simulation with and without noise. . . . .	54
5.4	Summed response represents measurement. . . . .	55
5.5	One hundred least square estimates for $\gamma = 0$ . . . . .	56
5.6	Nominal simulation with and without noise. . . . .	57
5.7	One hundred least square estimates for $\gamma = 1/4\pi$ . . . . .	58
5.8	Measurement and simulated response $\mathbf{f}(\hat{\mathbf{p}})$ . . . . .	59
5.9	One hundred least square estimates for $\gamma = 1/4\pi$ . . . . .	60
5.10	Geometrical interpretation Gauss-Newton method. . . . .	62
5.11	Geometrical interpretation of orthogonal convergence criterion. . . . .	63
5.12	Nominal simulation for deformation response vector $\mathbf{u}$ . . . . .	64
5.13	Nominal simulation with the addition of noise . . . . .	65
5.14	Least squares estimation for a non-linear model. . . . .	65
5.15	t-profile curve for one model parameter. . . . .	68
5.16	Profile trace plot for non-linear system. . . . .	69
5.17	t-profiling for $\mathbf{p}$ with $\gamma = 0$ . . . . .	71

5.18	True confidence region with profile trace vectors for $\gamma = 0$ .	72
5.19	t-profiling for $\mathbf{p}$ with $\gamma = 2/5\pi$ .	73
5.20	True confidence region with profile trace vectors for $\gamma = 0$ .	74
6.1	Deformation grid in $xy$ -plane for field 50.	75
6.2	Diagram of maximum absolute objective function values. Used for sensitivity.	77
6.3	An overlay measurement of the substrate.	78
6.4	Correlation coefficient matrix $\mathbf{C}$ for parameter variations $\Delta\hat{\mathbf{p}}$ .	79
7.1	Conservation of momentum in an infinitesimal volume.	88
7.2	Type of elements used in the FEM toolbox.	90
7.3	Schematic illustration of the thermo-mechanical model in Matlab.	92
7.4	Geometry of the simplified thermo-mechanical model in Matlab.	94
7.5	Burl layers are modelled as a shear layer	94
7.6	Mechanical constraints of simplified model.	95
7.7	Thermal constraints of simplified model.	96
7.8	Stiffness matrix with Cholesky decomposition without reordering.	99
7.9	Stiffness matrix with Cholesky decomposition with symrcm ordering.	100
7.10	Simplified heat load to substrate.	101
7.11	Heat load scan procedure.	102
7.12	Nominal experiment with meander routing.	104
7.13	Temperature response for overlay of meandering routing.	105
8.1	Overlay sensitivity response check.	109
8.2	Relative error between Finite Difference method and Direct method.	110
8.3	Different perturbation step-sizes for Finite Difference method and Direct method.	110
8.4	Relative impact on overlay per model parameter.	112
8.5	Overlay response for meander routing.	113
8.6	Fingerprint of $c_p$ for nominal experiment.	114
8.7	Fingerprint of $G_{zy}$ for nominal experiment.	115
8.8	Fingerprint of IR/EUV for nominal experiment.	115
8.9	Overlay response of the model parameters.	117
8.10	Least square regression for a non-linear model.	118
8.11	Standard meander, spiral, and jumpy routings.	119
8.12	Overlay meander routing.	120
8.13	Overlay spiral routing.	121
8.14	Overlay jumpy routing.	121
8.15	T-profiling meander routing, case 1.	124
8.16	T-profiling meander routing, case 2.	124
8.17	Confidence regions for meander routing for both cases.	125
8.18	Comparison of 95% confidence regions for both parameter cases.	126
8.19	Fingerprints of $c_p$ , $G_{zy}$ , and IR/EUV for spiral routing.	128
8.20	Fingerprints of $c_p$ , $G_{zy}$ , and IR/EUV for jumpy routing.	129
9.1	Geometrical interpretation of neighbourhood functions.	134
9.2	Basis functions for dose input parameter.	136
9.3	Dose input analysis for model parameter case 1.	137
9.4	Dose input analysis for model parameter case 2.	137
9.5	Simulated annealing optimization for case 1.	139
9.6	Established optimal experiment for parameters from case 1.	140
9.7	Overlay response of optimal experiment for parameters from case 1.	141
9.8	Simulated annealing optimization for case 2.	142

---

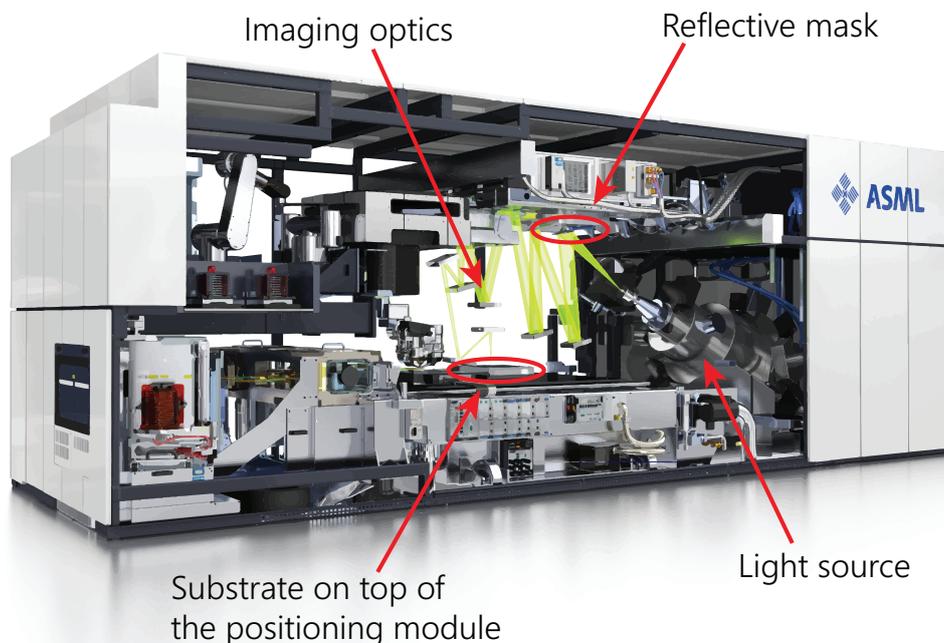
9.9	Established optimal experiment for parameters from case 2. . . . .	143
9.10	Overlay response of optimal experiment for parameters from case 2. . . . .	144
9.11	Confidence regions for both cases with optimal experiment. . . . .	145
9.12	Comparison of 95% confidence regions for both parameter cases. . . . .	146
9.13	Fingerprint of optimal experiment for $c_p$ . . . . .	147
9.14	Fingerprint of optimal experiment for IR/EUV ratio. . . . .	148
9.15	Temperature field for overlay of jumpy and optimal experiment. . . . .	149
9.16	Fingerprint of optimal experiment for $c_p$ . . . . .	149
9.17	Fingerprint of optimal experiment for IR/EUV ratio. . . . .	150
9.18	Brute force approach generating random experiments in order to compare. . . . .	151
9.19	Parameter bounds for optimal experiments. . . . .	152
A.1	Least squares estimation progress for meander routing. . . . .	163
A.2	Least squares estimation progress for spiral routing. . . . .	163
A.3	Least squares estimation progress for jumpy routing. . . . .	164
A.4	Least squares estimation progress for optimal experiments. . . . .	164
A.5	T-profiling spiral routing, case 1. . . . .	165
A.6	T-profiling spiral routing, case 2. . . . .	165
A.7	Confidence regions for spiral routing for both cases. . . . .	166
A.8	T-profiling jumpy routing, case 1. . . . .	166
A.9	T-profiling jumpy routing, case 2. . . . .	167
A.10	Confidence regions for jumpy routing for both cases. . . . .	167
A.11	T-profiling optimal routing, case 1. . . . .	168
A.12	T-profiling optimal routing, case 2. . . . .	168
C.1	Profile traces in the $\tau$ coordinates. . . . .	172



# 1

## Introduction

ASML's wafer scanners are crucial machines in the production of integrated circuits (ICs). An important performance parameter in these photo-lithographic machines is the so-called overlay error. It is desirably small (order of one nanometer, i.e.,  $1 \times 10^{-9}$  m ) for the accurate stacking of multiple electrical layers. An important part of a wafer scanner is a positioning module that is servo-controlled in six motion degrees of freedom (DOFs). This positioning module holds the substrate, i.e. the silicon disk on which the ICs are processed. The substrate is attached to the positioning module via a clamp. For good overlay, the substrate must be positioned very accurately with respect to the imaging optics in the photo-lithographic machine. In view of the required (sub-)nanometer performance, this accurate positioning and hence good overlay is hampered by structural deformations of the substrate that are inevitably induced by thermal disturbances.



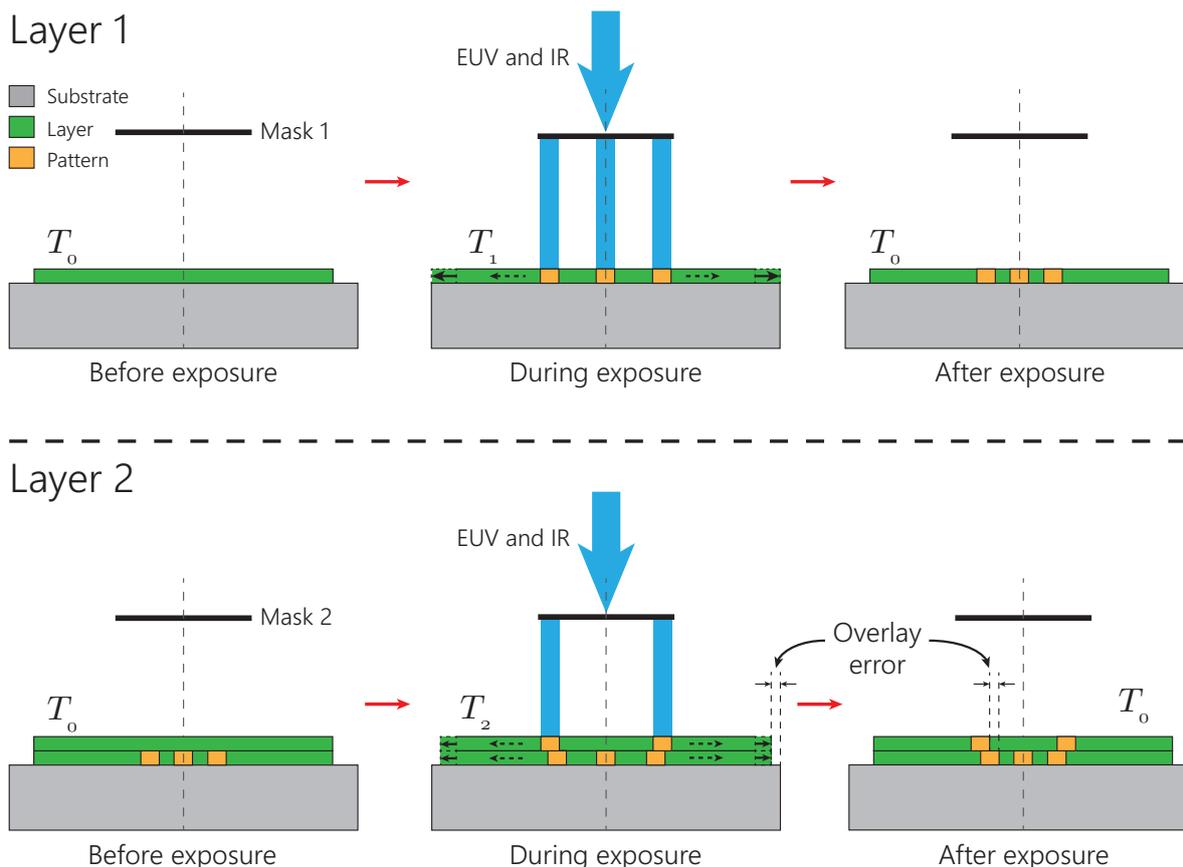
**Figure 1.1:** Inside of the EUV wafer scanner ASML machine. EUV light shines on the reflective mask in order to construct a light pattern. Subsequently, the EUV light continues through the imaging optics and will end up at the substrate.

## 1.1 Motivation and Problem Definition

In order to improve the understanding of the photo-lithographic machine and the previously mentioned overlay error, its inside is depicted in Figure 1.1. To put it simply, the photo-lithography process goes as follows: Extreme Ultra-Violet (EUV) light is generated by the *light source* and reflected against the *mask*. This mask is used to generate a light pattern, desired by the customer for the lithography process. The EUV light continues through the *imaging optics* where the pattern is reduced, and eventually the light pattern is printed on the *substrate*.

The *overlay* is a measure for the accurate stacking of multiple layers during the photo-lithography process in the  $xy$ -plane. A different measure for the accuracy of the photo-lithography process is *focus*. Like a photo camera, the focus plays an important role in the sharpness of an image. It is not different for the ASML machine, and it is desired to have the focal point of the lens at the surface of the substrate (in  $z$ -direction) for maximum sharpness of the image. The overlay and focus performance are, amongst others, hampered by thermal disturbances. In van de Wal (2014) an overview is given of the thermal problems that will arise during development of these photo-lithographic machines and what kind of measures should be taken to cope with them.

During the generation of EUV light in the light source, infra-red light (IR) is generated as well. This effect is not desirable. IR does not contribute to the lithography process and it is directly translated into a thermal load. Several components in the photo-lithographic machine will influence the overlay performance due to thermal loads. For instance, when the imaging optics are heated due to IR they will deform and, subsequently, a distorted image is the result. Another example would be a temperature difference of the substrate and substrate clamp. Studies have shown that this temperature difference gives rise to deformations of the substrate and its clamp, due to thermo-mechanical behaviour, which will lead to serious overlay errors. However, the photo-lithographic machine can – in addition of active cooling and heating – deal with these type of errors by adjusting its set point (e.g., adjust imaging optics, substrate positioning module, or mask positioning module). Therefore, it is useful to predict these deformations of the substrate and send this information back into the machine in order to improve the overlay performance.



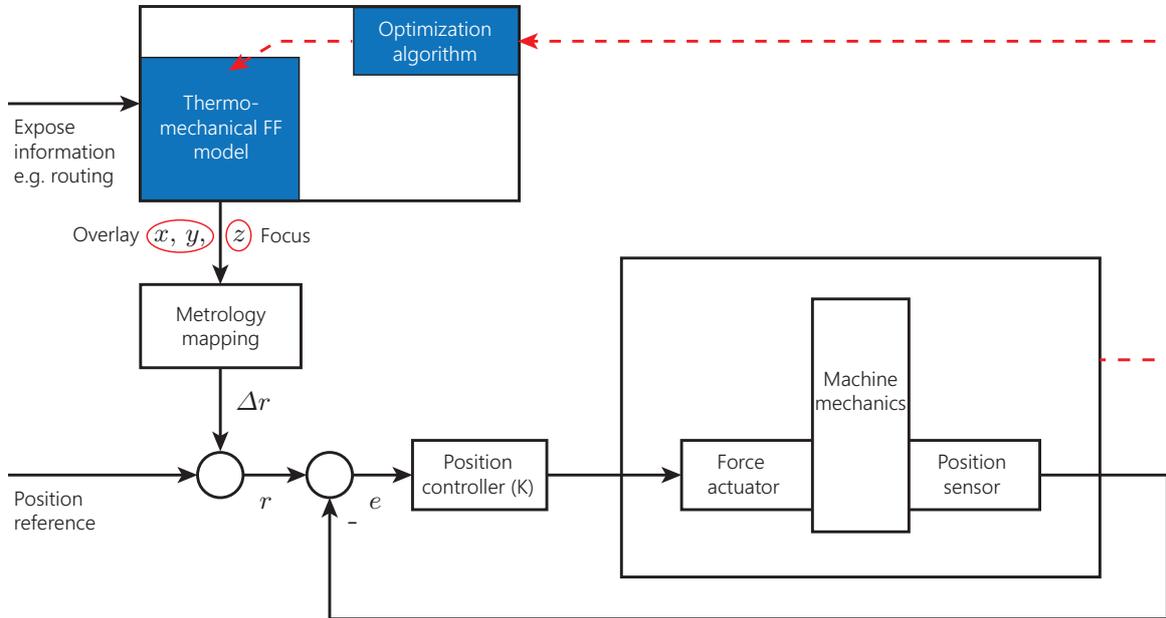
**Figure 1.2:** Overlay error due to substrate deformation caused by heating. Patterns are not properly stacked on top of each other due to different heat load (second layer).

A two layer exposure is schematically shown in Figure 1.2. Before exposure, the substrate and the layer are conditioned at temperature  $T_0$ . During exposure layer one is illuminated by three light beams, which is the result of the mask. For convenience, the mask is represented as if light passes through, but in reality light is reflected in the desired pattern. The first layer is subjected to thermal loads due to EUV and IR and the temperature rises to  $T_1$ . Hence, deformations will arise, and the imaging of the pattern is performed at the expanded layer. When the exposure is completed, the substrate and the layer are conditioned to initial temperature  $T_0$  over again and the layer and its pattern will shrink.

Next, a second layer is deposited on the substrate. Before the second exposure, the substrate and its layers are conditioned at temperature  $T_0$ . A new mask with a different pattern is used due to the lithography design specifications. This new mask generates a pattern which contains only two light beams. As a result, the thermal load is less compared with the exposure of layer one, which was illuminated by three light beams. During exposure, the temperature rises to  $T_2$ . Since the thermal load is less,  $T_2 < T_1$  and thus layer two will deform less. Subsequently, the imaging of the pattern is performed on a less expanded layer. When the exposure is complete and the system is at  $T_0$  an overlay error is the result.

In Figure 1.3 the position control loop of the positioning module is depicted. A feedforward is present in the form of a thermo-mechanical model. This model predicts the substrate deformation when it is processed by the machine, or in other words, when the substrate is exposed to thermal loads. The input of the thermo-mechanical model is expose information such as routing of the substrate or dose input (amount of energy per area that hits the surface of the substrate). The output of the model is the deformation of the substrate in  $x$ ,  $y$ , and  $z$  direc-

tion. Subsequently, this information is converted into information that is manageable for the machine in order to correct for the predicted substrate deformation. This step is performed in the *metrology mapping* block. Afterwards, the position reference (or set point) is corrected by  $\Delta r$  and the new set point  $r$  is fed into the feedback control loop.



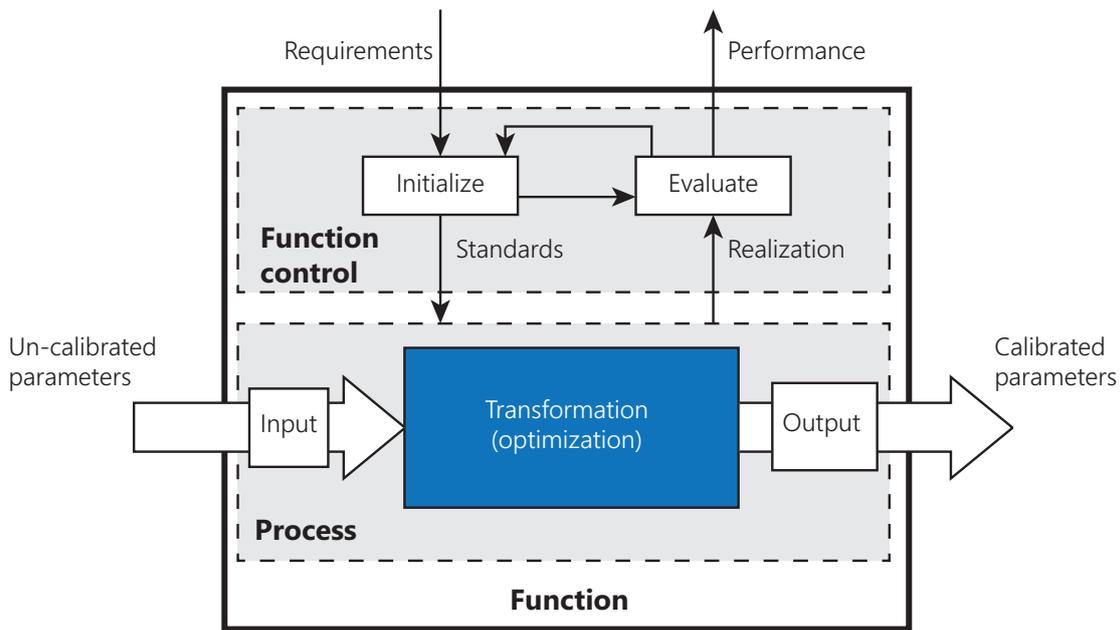
**Figure 1.3:** Position controller of the positioning module implemented in the machine (without optimizer). A thermo-mechanical model is present to provide feedforward. The optimization algorithm provides calibrated parameters for the feedforward model (not present currently).

In Figure 1.3, an optimization block can be seen as well. The thermo-mechanical model depends on several thermal and mechanical parameters  $[p_1 \dots p_k \dots p_K]^T$ . In theory, the values of these model parameters are known, but in reality they might differ slightly. In order to improve the accuracy of the model, the parameters need to be calibrated. The calibration of the parameters is the function of the optimization block. The input of this block is an overlay measurement obtained from the machine (red dotted line). The outputs of the block are the optimized parameters which are fed in the feedforward model. It was shown in van der Meulen (2015) that the optimization algorithm has difficulties in finding the correct values when the feedforward parameters  $[p_1 \dots p_k \dots p_K]^T$  are poorly distinguishable from the overlay measurement, or so to say, are non orthogonal to each other.

The feedforward model contains exposure information – also called recipe – as an input (e.g. routing and dose), see Figure 1.3. By changing the recipe, for instance the routing of the exposure, the overlay error differs. The following question comes to mind if it is possible to optimize the recipe such that the thermo-mechanical model parameters, belonging to overlay error, become orthogonal to each other. When such an optimized recipe exists, it is rather convenient to acquire the correct model parameters. Lets keep this in mind for now.

**Problem overview** In order to gain insight and to enlarge the overview of the problem, the Delft Systems Approach is applied (Veeke et al., 2008). It is stated that every problem can be expressed in a function block as shown in Figure 1.4. The main goal is to improve the overlay performance. This can be achieved in several ways, and one is *to calibrate the model parameters*, as was explained previously. This latter represents the function as shown in Figure 1.4. Uncalibrated model parameters enter the function and as a result, calibrated model parameters

come out. Within the function several blocks can be found.

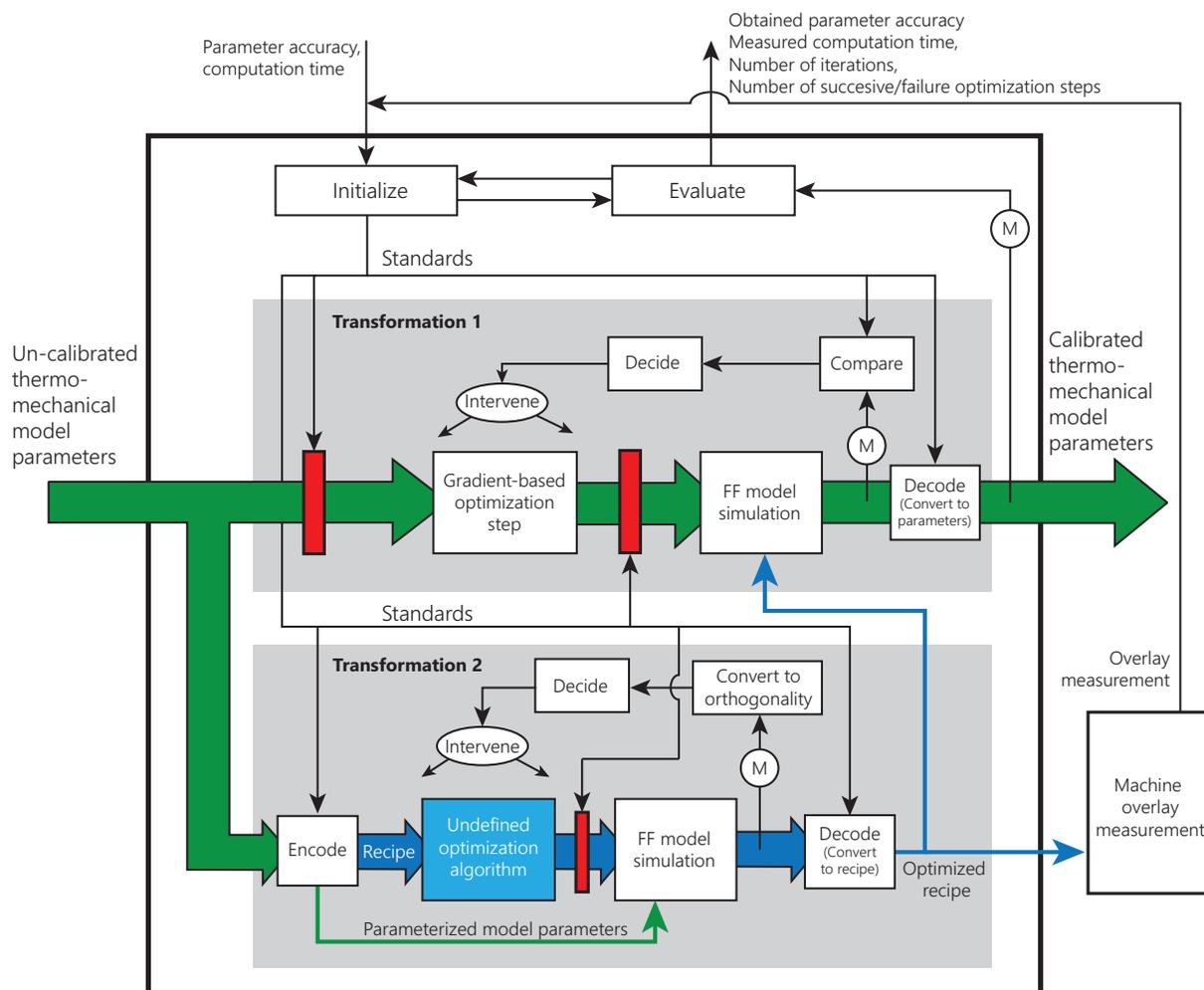


**Figure 1.4:** Simplified block diagram of the system.

Within *function control* the *initialize* and *evaluate* blocks are located. The initialize block translates the requirements of the function into standards which are measurable in the system (or process). The real results of the process are measured in terms of these standards. For example, if the requirement is to find an optimum in the lowest amount of time possible (vaguely defined) then the initialize block will translate this requirement into a measurable standard, say to find an optimum in one minute or less. This standard is measurable and is evaluated in the evaluate block. The evaluate block checks whether the requirements are achieved. When it happens to be that the requirements are often violated, it could indicate an infeasibility of the standards. The evaluate block will intervene by passing this information to the initialize block. Subsequently, the initialize block will adjust the standards such that the system remains feasible. Function control is responsible for guarding of the feasibility as well.

In *process* the *transformation* block can be seen. This block contains the task of the function, the actual transformation from the input to the output. In this simplified representation it cannot be seen how the transformation is done, a so-called *black box* approach. The *input* and *output* block are illustrated as well. The ingoing information is often provided in a form that cannot be handled by the transformation block. The input block will encode the input to manageable input for the transformation block. The same holds for the output block, only the encoding part is replaced by decoding.

In Figure 1.4 the transformation block is a "black box", but for now it is opened by just a little (see Figure 1.5). It can be divided into two separate transformation blocks. Transformation 1 represents the actual optimization of the thermo-mechanical model parameters. Transformation 2 represents the optimization of the recipe in which the thermo-mechanical parameters obtain the best possible orthogonal property to each other. Subsequently, the optimized recipe is used in transformation 1 (the blue arrow) to improve the optimization of the thermo-mechanical model parameters.



**Figure 1.5:** The simplified "black box" is opened a little. Two transformations take place. Transformation 1 ensures for the calibrated model parameters. Transformation 2 ensures an optimal recipe that is used by transformation 1 to accomplish excellent calibrated parameters. The blue block in transformation 2 is of most interest in this thesis.

In transformation 1, the input passes a *filter* (red rectangle). This filter checks whether the un-calibrated parameters lie within a predefined range. This could be necessary for the optimization algorithm to ensure convergence. Next, the un-calibrated parameters enter the *gradient-based optimization step* and new parameters are calculated. For now, this optimization step is out of scope, since it is successfully used in previous analysis. The new parameters are checked for feasibility by the filter and are, subsequently, fed into the *feedforward model simulation* – together with the optimized recipe – and an exposure is simulated. When the simulation is done, the result is compared with the machine overlay measurement with the same optimized recipe. Subsequently, the *decide* block decides to take action yes or no. This will depend on the objective function. When the objective function is not satisfied, intervention takes place and a new optimization step is performed. When the optimization is completed a *decoding* takes place. This block converts the result from the feedforward simulation into thermo-mechanical parameters, which are calibrated.

The input in transformation 2 passes an encoding block. The *encoding* block converts the un-calibrated parameters into parameterized parameters which are fed into the feedforward model simulation block. A feasible starting recipe is generated as well and goes into the *undefined optimization algorithm*. This optimization algorithm is supposed to generate a new recipe which is more convenient with the objective function. It is unknown how this optimization algorithm

is built. The new recipe is checked for feasibility by the filter. Next, a simulation is performed and the overlay result is measured and converted into a measurable orthogonality dependency. Subsequently, the decide block determines to take action yes or no. Intervention takes place when the objective function is not satisfied and the optimization algorithm continues. When the optimization is completed, the decoding function provides the optimized recipe. Thereafter, this recipe is fed into transformation 1 and is used in the machine overlay measurement. This machine overlay measurement will eventually end up in the compare block of transformation 1.

Transformation 2 is still an unknown process, especially the undefined optimization algorithm (blue block). Therefore, the main research question in this thesis is:

*How to make the appearance in measurements of the physical parameters subject to calibration more orthogonal via optimization of the experiment that provides these measurements?*

That is to say, numerical optimization of the overlay experiment via the recipe (for example, routing and dose input).

In order to be able to answer the main research question, several sub questions must be defined. By observing transformation block 2 in Figure 1.5 the following questions arise:

1. How to convert the information content of an measurement into a measurable criterion for the rate of orthogonality or parameter distinguishability?
2. What type of optimization algorithm is capable of efficiently optimizing an experiment which partly consists of non-continuous design parameters?
3. What is an adequate intervene process for the optimization algorithm?

Although transformation block 1 is known and currently used, still some questions arise:

4. Is the gradient-based optimization algorithm the most efficient and applicable method for parameter calibration?
5. What is the measure of the parameter calibration which decides converges and do there exist more suitable measures?

## 1.2 Structure of the Thesis

The thesis will be contained in two parts. The first part consists of a literature study, supported by several examples, which will be a useful "toolbox" to understand the principles of experiment design and everything that comes with it.

The second part consists of the application of experiment design to a simplified thermo-mechanical feedforward model, which is the main matter of the new piece of information.

### PART I - Literature and Analysis Tools

**Chapter 2** In this chapter, the current available thermo-mechanical model of ASML used for feedforward is described.

**Chapter 3** In this chapter, several sensitivity methods are described for their efficiency and application. Illustrations are made by using simple examples.

**Chapter 4** In this chapter, state of the art experiment design techniques are described and illustrated.

**Chapter 5** In this chapter, linear and non-linear regression techniques are explained in order to provide statistical information about model parameters.

**Chapter 6** In this chapter, a preliminary calibration investigation on the thermo-mechanical feedforward model of ASML is performed to obtain insight of the model behaviour. The current problems are highlighted as well.

### PART II - Application of Experiment Design

**Chapter 7** In this chapter, a simplified thermo-mechanical model is made and explained in detail.

**Chapter 8** In this chapter, the simplified thermo-mechanical model is analysed for its characteristics and behaviour.

**Chapter 9** In this chapter, the standard available experiments are investigated for their performance in the context of model parameter accuracy during estimation.

**Chapter 10** In this chapter, the experiment design techniques are used to find an optimal experiment by means of an optimization algorithm. Subsequent, the optimal experiments are compared with the standard experiments.

**Chapter 11** In this chapter, the conclusions and recommendations of this study are given.

## Part I

# Literature and Analysis Tools

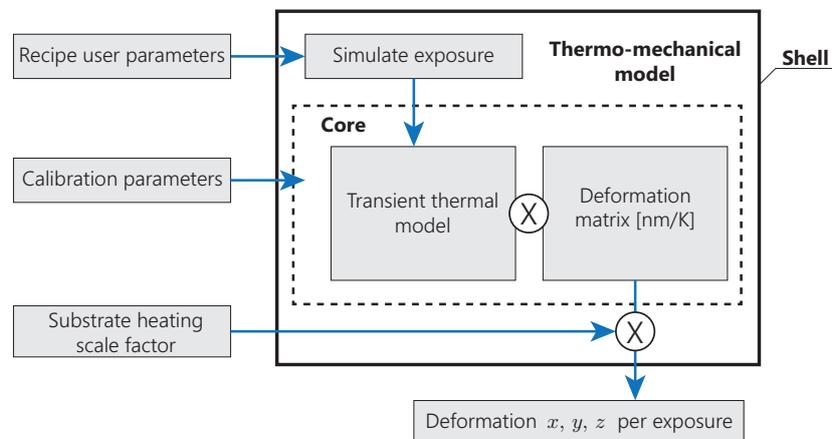


# 2

## Thermo-Mechanical Model of ASML

As is described in Castelijns (2014), the purpose of the thermo-mechanical feedforward model is to correct for the effects of substrate heating on overlay and focus. The model consists of a core and a shell layer. The latter is used to simulate an exposure sequence and to feed the core model with relevant parameters. The core model is a so-called 2.5D model. These type of models can represent surfaces in the third dimension ( $z$ ), but they are limited by one single elevation only. This implies that folded surfaces cannot be represented by 2.5D models.

A schematic overview of the implementation of the thermo-mechanical feedforward model is given in Figure 2.1.



**Figure 2.1:** Schematic overview of the thermo-mechanical feedforward model. It consists of a core and a shell layer. The latter is used to simulate an exposure.

In essence the core model is a transient dynamic thermal model which the user input acts upon. Thermal equations are solved numerically in order to acquire the temperature distribution in the substrate in space and time. Subsequently, the substrate deformation is obtained by multiplying the temperature distribution of the substrate by pre-calculated thermo-mechanical relations. These relations are, in turn, calculated offline by using a finite element method (FEM). They form a big data set, called *deformation matrix* or *C-matrix* Meulenbroeks (2014). Since the time constants of the thermal problem are relatively large compared to the mechanical time constants (eigenfrequencies) no dynamic coupling is modelled. Therefore, the deformation calculation can be separated from the thermal calculation.

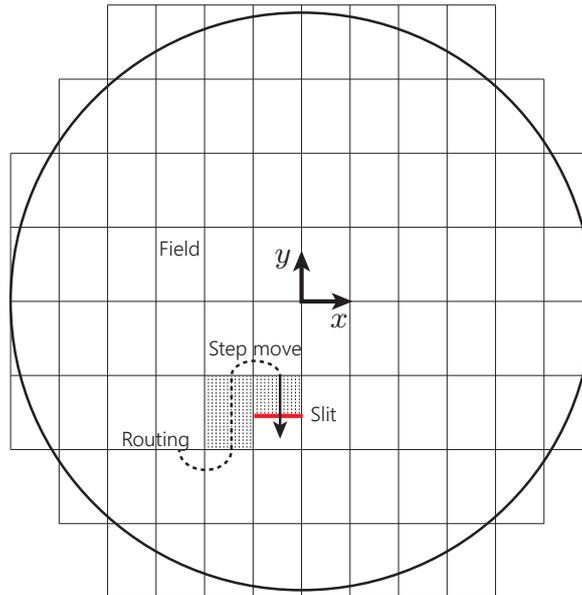
In Figure 2.1, a substrate heating scale factor is included. Currently, when performing an

overlay simulation, this scale factor is used to linearly scale the substrate deformation result in order to tune the model such that it (almost) overlaps a real overlay measurement. The substrate heating scale factor is, in fact, a tuning parameter to improve the accuracy of the model. However, this scaling factor is unnecessary when a successful experiment design is obtained. The designed experiment ensures an orthogonal (or close to orthogonal) relation between the model parameters in the overlay measurement, which in turn, can be used to obtain model parameters with high accuracy.

## 2.1 Exposure Sequence

As was mentioned previously the thermo-mechanical model consists of two parts, namely a core and a shell layer. The exposure sequence is located in the shell layer.

A substrate exposure consists of a number of field exposures. In each field the scanning motion is in  $y$ -direction when projecting the image onto the resist. The exposure beam projected on the substrate is called *slit*. In between the fields the slit moves from one field to the other, also known as *step move*. During the step move there is no exposure light present, but an extra unspecified heat load may still act upon the substrate. Each field has a dimension of  $26 \times 33$  mm<sup>2</sup> (width $\times$ height). When a smaller image is exposed the illumination and projection beam are partly blocked by blades, which effectively cuts off the image projected on the substrate. The slit size will be reduced, but the full slit may still project infra-red light coming from the exposure beam onto the substrate, since the blades do not block infra-red.



**Figure 2.2:** Multiple fields on a substrate. A scanning motion of the exposure beam in  $y$ -direction is represented. This routing is not fixed and can be chosen by the user.

The scan and step moves are subject to recipes. The recipe determines the exposure information like routing, field lay-out, and physical dose that is applied to the substrate. In order to achieve a successful lithography process, the substrate must receive a certain amount of energy. The scanning speed is relevant to determine the received amount of energy. It can be calculated by,

$$v_{\text{scan}} = \frac{\Phi}{\text{dose}} \quad (2.1)$$

where  $\Phi$  is the slit power per meter ( $\text{W m}^{-1}$ ) and dose the energy per area ( $\text{J m}^{-2}$ ) which is defined by the user. In addition,  $\Phi$  is calculated according:

$$\Phi = \frac{P_{\text{EUV}}}{b_0} \quad (2.2)$$

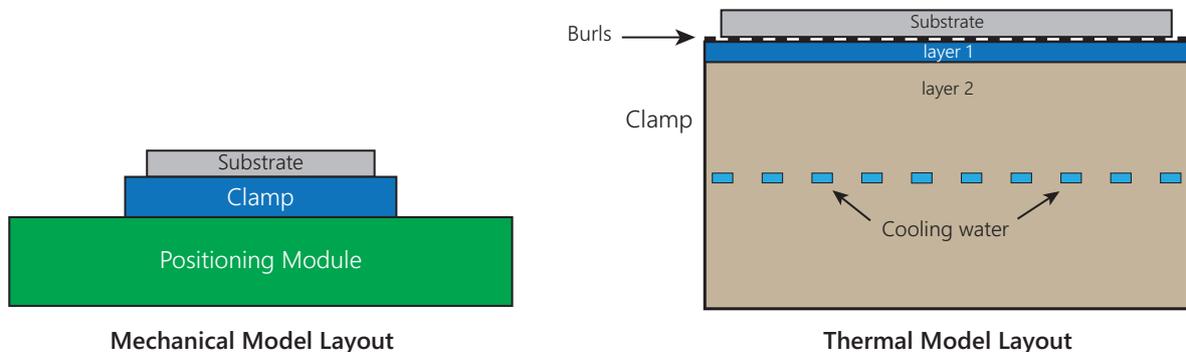
Here,  $P_{\text{EUV}}$  is the power coming from the light source – at substrate level – that is used for the lithography process (100% EUV transmission assumed) and  $b_0$  is the default slit width, i.e., 26 mm.

## 2.2 Physical Representation of Model

The substrate is clamped with an electrostatic clamp which, in turn, is clamped to the positioning module. A large number of burrs make up the surface at the interface between the substrate and the clamp. The same holds for the interface between the clamp and the positioning module. The components between substrate and cooling water form a thermal barrier. In Figure 2.3 the layout of the model is given.

The mechanical model consists of all components. This model is used to calculate the deformations at substrate level or, in other words, it is used to obtain the deformation C-matrix.

The thermal model consists of the substrate and the clamp, but the positioning module is omitted. In this area the heat transfer is limited and the clamp is conditioned by cooling water as well. The positioning module is not omitted in the mechanical model, since it has a strong influence on the deformation profile. The clamp is divided in several (material) layers. The burrs are not implemented in the geometry, but the thermal resistance is taken into account via the heat transfer coefficient from substrate to clamp. The clamp has cooling water channels located in the middle (see Figure 2.3).



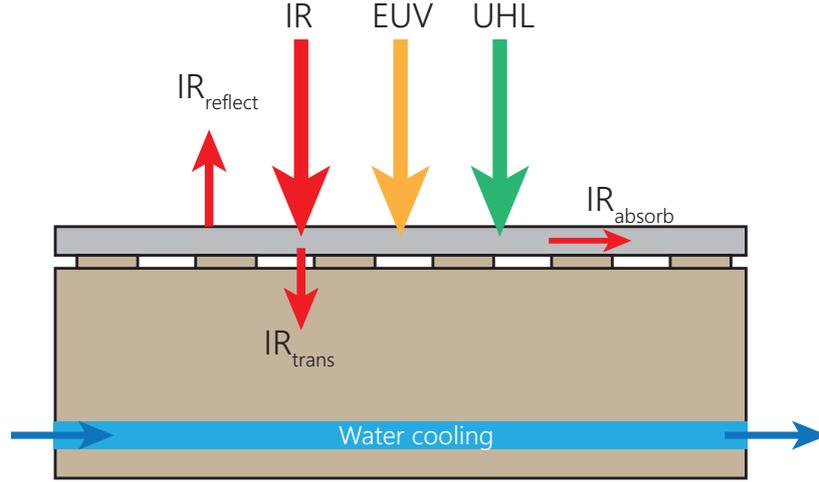
**Figure 2.3:** Layout of the model. The mechanical model layout contains the positioning module. The thermal model layout contains the clamp and substrate only. This latter model is used to calculate transient thermal behaviour.

## 2.3 Exposure Heat Load

The main heat load of the substrate and its clamp is due to EUV and infra-red (IR) light coming from the exposure beam or light source. An overview is shown in Figure 2.4. An extra unspecified heat load (UHL), not related to EUV or IR, is present as well. The UHL is not modelled in detail, but a simplified fixed heat load is added which represents the extra heat load. The following assumptions are made in the application of the heat loads:

- EUV is fully absorbed by the substrate due to the photo-resist. All photon energy is converted into heat.
- IR is partly reflected at the substrate surface.
- IR is partly absorbed in the substrate.
- IR is partly transmitted through the substrate.
- UHL is fully absorbed by the substrate.

Eventually, all heat will be removed from the system by the cooling water in absence of additional heat load.



**Figure 2.4:** Representation of the applied heat load. Some infra-red (IR) light is reflected or transmitted to the clamp. Extreme Ultra-Violet (EUV) light and an extra unspecified heat load (UHL) are both absorbed by the substrate.

The total heat load  $P_{\text{load}}$  is a summation of the loads depicted in Figure 2.4. It can be described by

$$P_{\text{load}} = P_{\text{IR}} + P_{\text{EUV}} + P_{\text{UHL}} \quad (2.3)$$

Note that  $P_{\text{UHL}}$  continues to be present during the step moves. The EUV heat load is determined according

$$P_{\text{EUV}} = \Phi \cdot b_0 \cdot f_{\text{dose}} \quad (2.4)$$

where  $f_{\text{dose}}$  is introduced, compared to equation (2.2). This is done to take into account the uncertainty of the absolute power measured by the energy sensor. The IR heat load at the substrate surface level is proportional to the EUV power and is determined according:

$$P_{\text{IR}} = R_{\text{IR/EUV}} \cdot P_{\text{EUV}} \quad (2.5)$$

The factor  $R_{\text{IR/EUV}}$  is the ratio of IR present in the EUV power at substrate level. Not all IR power will end up in the substrate and may be reflected from the substrate or transmitted through the substrate into the clamp. The following can be written in order to determine this quantity:

$$C_{\text{IR,substrate}} = R_{\text{IR/EUV}} (1 - r_{\text{sub}}) (1 - t_{\text{sub}}) \quad (2.6)$$

$$C_{\text{IR,clamp}} = R_{\text{IR/EUV}} (1 - r_{\text{sub}}) t_{\text{sub}} \quad (2.7)$$

$$C_{\text{IR,reflect}} = R_{\text{IR/EUV}} \cdot r_{\text{sub}} \quad (2.8)$$

Where  $r_{\text{sub}}$  is the reflective coefficient of IR on the substrate and  $t_{\text{sub}}$  is the transmission coefficient of IR through the substrate. Subsequently, the IR can be separated according:

$$P_{\text{IR}} = R_{\text{IR/EUV}} \cdot P_{\text{EUV}} \quad \rightarrow \quad \begin{cases} P_{\text{IR,substrate}} & = C_{\text{IR,substrate}} \cdot P_{\text{EUV}} \\ P_{\text{IR,clamp}} & = C_{\text{IR,clamp}} \cdot P_{\text{EUV}} \\ P_{\text{IR,reflect}} & = C_{\text{IR,reflect}} \cdot P_{\text{EUV}} \end{cases} \quad (2.9)$$

Currently, the heat load due to  $P_{\text{UHL}}$  is incorporated in the  $P_{\text{IR}}$  contribution. This is accomplished by increasing the IR/EUV ratio mentioned in equation (2.5). The heat load  $P_{\text{UHL}}$  is set to zero and as a consequence, the UHL is not applied during step moves. Alternatively, the whole substrate heating can be scaled to account for the larger total heat load.

The total load at the substrate and clamp can be written as:

$$P_{\text{substrate}} = (1 + C_{\text{IR,substrate}}) P_{\text{EUV}} \quad (2.10)$$

$$P_{\text{clamp}} = C_{\text{IR,clamp}} \cdot P_{\text{EUV}} \quad (2.11)$$

Which eventually becomes,

$$\begin{aligned} P_{\text{load}} &= P_{\text{substrate}} + P_{\text{clamp}} \\ &= P_{\text{EUV}} (1 + C_{\text{IR,substrate}} + C_{\text{IR,clamp}}) \end{aligned} \quad (2.12)$$

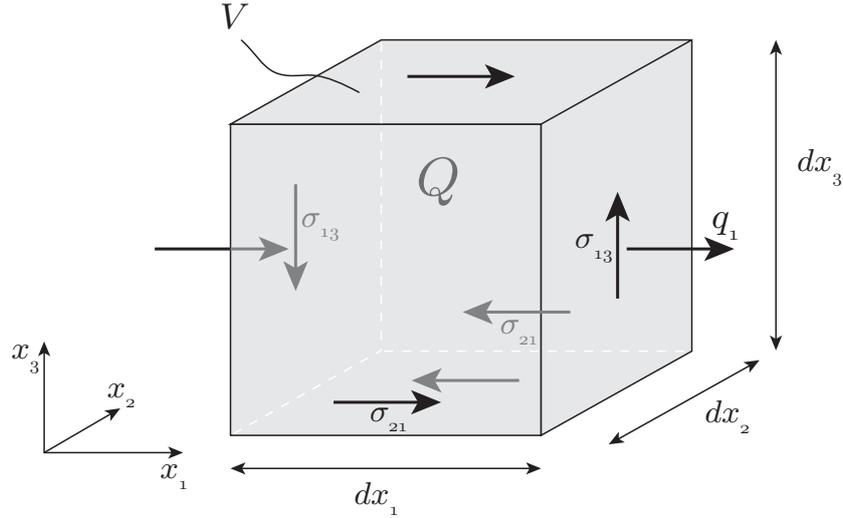
It should be noted that these equations only hold for a full slit exposure. When a smaller image is exposed the illumination and projection beam are partly blocked by blades. In order to cope with the blading, the equations must be adjusted by a little. This is done in Castelijns (2014, p. 29).

## 2.4 Governing Equations

In order to find the governing equations of basic thermal analysis in a solid, one can write the conservation of energy equation for an infinitesimal volume in the solid  $V$  (Rixen, 2011a), see Figure 2.5.

$$\frac{\partial e}{\partial t} + \frac{\partial q_i}{\partial x_i} - \sigma_{ij} \frac{\partial \varepsilon_{ij}}{\partial t} - Q = 0 \quad \in V(x_i) \quad (2.13)$$

Where  $e(\varepsilon_{ij}, T)$  is the internal energy density and is typically a function of the state of strain  $\varepsilon_{ij}$  and the temperature  $T$ . In addition,  $q_i$  represents the heat flux and  $\sigma_{ij}$  represents the mechanical stresses in the directions of  $\mathbb{R}^3$ . The heat source of the solid (in  $V$ ) is captured in  $Q$ . The first term in equation (2.13) represents the change of thermal energy which is created by the divergence of the heat flux (the second term). The third term represents the power produced by stresses related with the strain rates and the last term  $Q$  represents the internal heat source.



**Figure 2.5:** Conservation of energy in an infinitesimal volume. Not all heat fluxes  $q_i$ , stresses  $\sigma_{ij}$  and strains  $\varepsilon_{ij}$  are depicted.

The power produced by strain rates is assumed to be small, due to the slow thermal expansion behaviour, and is therefore disregarded. Furthermore, it is assumed that the internal energy  $e$  is not a function of the strains, due to small deformations. Therefore, the time derivative of the internal energy can be written as:

$$\frac{\partial e}{\partial t} = \rho c(T) \frac{\partial T}{\partial t} \quad (2.14)$$

Where  $\rho$  ( $\text{kg m}^{-3}$ ) is the mass density of the solid and  $c$  ( $\text{J kg}^{-1} \text{K}^{-1}$ ) the specific heat of the material. It could be possible that the specific heat is temperature dependent, but it is often assumed to be constant within a certain temperature range. Due to the small temperature range in which the feedforward model is subjected to, the heat capacity is:

$$c(T) = c = \text{constant} \quad (2.15)$$

By combining equation (2.13) and (2.14) and neglecting the power produced by strain rates, the basic thermal equation becomes:

$$\rho c \frac{\partial T}{\partial t} + \frac{\partial q_i}{\partial x_i} = Q \quad \in V(x_i) \quad (2.16)$$

Fourier's law is used to describe the heat flux within the solid. It is given by:

$$q_i = -k \frac{\partial T}{\partial x_i} \quad i = 1, 2, 3 \quad (2.17)$$

The parameter  $k$  is known as the thermal conductivity and, in this case, its representation is scalar (isotropic material is assumed). It can be seen that the heat flux is related to the gradient of the temperature. When combining Fourier's law with the basic thermal equation, the thermal problem writes:

$$\rho c \frac{\partial T}{\partial t} - k \frac{\partial^2 T}{\partial x_i^2} = Q \quad \in V(x_i), \quad i = 1, 2, 3 \quad (2.18)$$

## 2.5 Discretization

A finite difference scheme is used in order to perform spatial discretization of the model. Each layer in the thermal model consists of a 2D square grid made of rectangular cells. One cell represents the control volume of the corresponding thermal node. The default cell size is  $6.5 \times 6.5 \text{ mm}^2$  and the height ( $z$ -direction) may vary per layer. The total span area of the model is  $325 \times 325 \text{ mm}^2$ , which translates to  $50 \times 50$  cells with default cell size. Since the grid is square and the substrate and clamp are round, not all cells will contribute to the solution. The cell outside the system are isolated and their temperature will remain zero. The same holds for the heat transfer and diffusion.

In the feedforward model, the substrate grid layer is refined in order to improve the accuracy and to enhance the description of the model near the edges. The cell size at substrate level is  $6.5 \times 3.25 \text{ mm}^2$ , resulting in a grid of  $50 \times 100$  cells. The refinement is done in  $y$ -direction since it will improve the interpolation, which was mentioned in Section 2.6.1.

### 2.5.1 State Space

The continuous equation (2.18) can be rewritten in the following form:

$$\frac{\partial T}{\partial t} = \frac{k}{\rho c} \frac{\partial^2}{\partial x_i^2} T + \frac{Q}{\rho c} \quad \Rightarrow \quad \dot{\mathbf{T}} = \frac{1}{\rho c} (k \nabla^2 T + Q) \quad (2.19)$$

Where  $\nabla$  is the Laplace-operator which is a differential operator. In order to solve the thermo-mechanical problem, equation (2.19) is discretized and subjected to the spatial grid. The following well-known state space notation is obtained:

$$\begin{aligned} \dot{\mathbf{T}} &= \mathbf{A} \mathbf{T} + \mathbf{E}^{-1} \mathbf{Q} \\ \mathbf{y} &= \mathbf{C} \mathbf{x} + \mathbf{D} \mathbf{u} \end{aligned} \quad (2.20)$$

The state vector  $\mathbf{T}$  (K) contains the temperature in the substrate, the clamp and the cooling water subjected to the spatial grid points. The system is exposed to an external load  $\mathbf{Q}$  which is explained in Section 2.3. Matrix  $\mathbf{E}$  ( $\text{J m}^{-3} \text{K}^{-1}$ ) represents the thermal mass density, i.e.  $\rho c$ . The output vector  $\mathbf{y}$  (m) contains the substrate deformations and is calculated by the deformation matrix  $\mathbf{C}$  (as was mentioned on page 11). The model does not contain direct mechanical loads, which are represented by  $\mathbf{u}$ . Therefore, this last part is omitted.

In general the  $\mathbf{A}$  matrix describes the dynamics of the system. The dynamical part of the total feedforward model is the thermal model only (see Figure 2.3). Hence, the  $\mathbf{A}$  matrix contains the thermal diffusion and heat transfer and no mechanical relations. The matrix is built in the following manner:

$$\mathbf{A} = \mathbf{E}^{-1} (\mathbf{K} + \mathbf{H}) \quad (2.21)$$

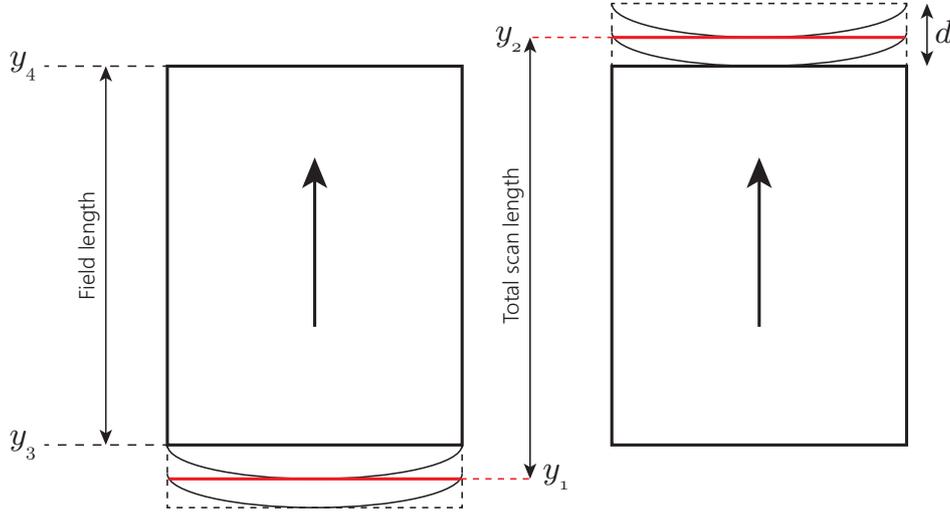
The thermal conductivity parameters for all layers and cooling channels are separated into two matrices. Matrix  $\mathbf{K}$  describes the diffusion and thermal convection in the channels in the horizontal plane and  $\mathbf{H}$  describes the heat transfer between the layers in the vertical plane. In Castelijns (2014) a description is given how these thermal diffusions and convections are calculated.

## 2.6 Numerical Implementation

In this section the implementation of the scan move, and the implementation of the time integration are described.

### 2.6.1 Scan Move

The scan motion is based on a constant scan velocity in  $y$ -direction solely. This is calculated according equation (2.1), which follows from the recipe. Due to the imaging optics the slit is curved. This will influence the scan distance as is shown in Figure 2.6. The scan length is given by  $y_2 - y_1 = (y_4 - y_3) + d$ .



**Figure 2.6:** Scan move across a field. The  $y$ -coordinates of the field are  $[y_3, y_4]$ . The scan start and end  $y$ -coordinates are  $[y_1, y_2]$ .  $d$  represents the bounding box slit thickness.

Per exposure (per field) a buffer is necessary in order to store the temperature states for a certain amount of slit positions. The deformation per field will be determined on a grid, which is defined by the coordinates  $(x_{\text{deform}}, y_{\text{deform}})$ . While the slit moves across the field, according the thermal time step  $\Delta t$ , the central  $y$ -coordinate  $y_{\text{slit}}$  passes across several pre-defined  $y$ -coordinates  $y_{n,\text{deform}}$ , with  $n$  the number of pre-defined grid coordinates per field. The central  $y_{\text{slit}}$  coordinate may not necessarily coincide with the pre-defined  $y_{\text{deform}}$  coordinates (see Figure 2.7). The  $y$ -grid is given by:

$$y_{n,\text{deform}} = y_3 + \left( \frac{y_4 - y_3}{N - 1} \right) n \quad n = 0, 1, \dots, N - 1 \quad (2.22)$$

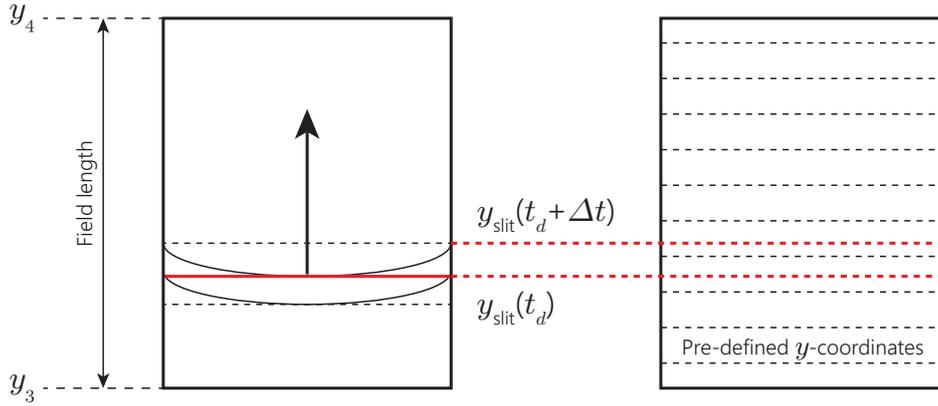
where  $N$  is the number of rows, which is 10 for a full field in this case. For each thermal time step it is verified whether one of the  $y_{n,\text{deform}}$  is found between  $y_{\text{slit}}(t_d)$  and  $y_{\text{slit}}(t_d + \Delta t)$ . When this happens to be, the temperature state for the corresponding position  $y_{n,\text{deform}}$  is calculated and stored. This calculation is a linear interpolation between  $T(t_d)$  and  $T(t_d + \Delta t)$ . The interpolation is given by:

$$T_{n,\text{deform}} = \phi T(t_d) + (1 - \phi) T(t_d + \Delta t) \quad (2.23)$$

with,

$$\phi = 1 + \left( \frac{y_{n,\text{deform}} - y_{\text{slit}}(t_d)}{y_{\text{slit}}(t_d) - y_{\text{slit}}(t_d + \Delta t)} \right) \quad (2.24)$$

This interpolated temperature state  $T_{n,\text{deform}}$  is stored in the thermal buffer, representing the position  $y_{n,\text{deform}}$  for the field that is being exposed. Once the scan of the field is completed and the thermal buffer is filled the deformations are calculated by multiplying the results with the deformation C-matrix. The deformation grid of the exposed field is obtained.



**Figure 2.7:** Pre-defined  $y$ -coordinates in a field. The central  $y$ -position of the slit may not necessarily coincide with the pre-defined  $y$ -coordinates.

### 2.6.2 Numerical Solution Scheme

In order to obtain the solution in time domain, the state-space representation (equation (2.20)) is integrated. The continuous time integration of the state-space system is described by

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \quad (2.25)$$

where  $\mathbf{x}$  represents the temperature state in the time domain. The discrete equivalent writes:

$$\mathbf{x}(t_d + \Delta t) = e^{\mathbf{A}\Delta t}\mathbf{x}(t_d) + \int_{t_d}^{t_d+\Delta t} e^{\mathbf{A}(t_d+\Delta t-\tau)}\mathbf{B}\mathbf{u}(\tau)d\tau \quad (2.26)$$

The matrix exponential is calculated according the Taylor expansion.

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \frac{(\mathbf{A}t)^3}{3!} + \dots \quad (2.27)$$

The *Euler forward* integration (explicit) scheme is used, which leads to a first order Taylor expansion of the matrix exponential given above.

**Stability** The thermal differential equations are solved explicitly, i.e., the next state of the system is determined via calculations of the current state. It is important to investigate the numerical stability, which is not the same as stability of the model itself. It is possible, even when the model is stable by its nature, that numerical integration creates instability due to an incautious selected time-step. Each numerical integration scheme has its own stability region.

A global stability analysis is performed on the feedforward model (Rook, 2014). In order to obtain the stability criterion for the *Partial Differential Equations* (PDEs), the Von Neumann stability analysis is performed. For an explicit Euler integration scheme the Fourier number ( $Fo$ ) for heat conduction in horizontal direction must satisfy the following criterion:

$$Fo_h = \frac{k}{\rho c(T)} \frac{\Delta t}{\min(\Delta x^2, \Delta y^2)} \leq \frac{1}{4} \quad (2.28)$$

A similar characteristic number can be defined for the vertical heat conduction between the layers:

$$Fo_v = \frac{h}{\rho c(T)} \frac{\Delta t}{\Delta z} \leq \frac{1}{2} \quad (2.29)$$

Here,  $h$  is the average heat transfer coefficient ( $\text{W m}^{-2} \text{K}^{-1}$ ) and  $\Delta z$  is the thickness of the layer. The feedforward model uses a time-step of 5 (ms) as a default. This is sufficient to guarantee stability for the Fourier numbers. What appears to be more critical for stability is the convection term. This can be characterized by the so-called Courant number,

$$Co = u \frac{\Delta t}{\Delta x} \leq 1 \quad (2.30)$$

where  $u$  represents the flow velocity in the cooling channels and  $\Delta x$  represents the 1D spacing of the channel grid. In other words, the numerical integration scheme remains stable when a flow particle does not skip a discretized cell – due to the flow velocity – with length  $\Delta x$  within a single time-step evaluation. For the default time-step, the Courant number is within the stability criterion.

**Time Step Sensitivity** The time-step has an important role for the numerical error that appears during integration. The error that occurs at every individual time-step is called the local truncation error  $\varepsilon$ . For the Euler integration scheme, the local truncation error  $\varepsilon$  is of order  $\mathcal{O}(\Delta t^2)$ . The global error  $E$ , equation (2.31), is more important. This global truncation error is the result after integration over the time-span  $t = 0 \dots T$ . For the Euler forward scheme  $E$  is of order  $\mathcal{O}(\Delta t)$ . It is necessary to investigate the accuracy of the model due to numerical integration. In Castelijns (2014) it was stated that the overlay error budget is 2.5 nm. It should be noted that the overlay budget is adjusted for the next generation machines to sub-nanometers.

$$E = \left( \frac{T}{\Delta t} \right) \varepsilon \quad (2.31)$$

The problem is solved in a two-way fashion. First the differential equations of the thermal part are solved by means of integration. Subsequently, the mechanical deformations are obtained by multiplying the thermal results with the C-matrix. Therefore, it is difficult to see the direct influence of the time-step in the overlay budget. In order to be able to give an impression of the accuracy of the feedforward model due to the selected time-step, the following must be done.

According Schwab and Delhaes (2009) the approximate solution of the problem is given by,

$$y = \hat{y} + C_1 \Delta t^p \quad (2.32)$$

with  $\hat{y}$  the true solution – which is unknown – and  $C_1 \Delta t^p$  the truncation error. For the Euler integration scheme  $p = 1$ . It is noticeable that the equation contains two unknowns, the true solution  $\hat{y}$  and the coefficient  $C_1$ . These unknowns can be determined by integrating the system twice. The first time integration is performed from  $t = 0$  to  $t = T$  with time-step  $\Delta t$ . Subsequently, the second time integration is performed from  $t = 0$  to  $t = T$  as well, but with half the time-step  $\Delta t/2$ . The solutions are referred to as  $y_{\Delta t}$  and  $y_{\Delta t/2}$  respectively. The following equations are obtained:

$$y_{\Delta t} = \hat{y} + C_1 \Delta t^p \quad (2.33)$$

$$y_{\Delta t/2} = \hat{y} + C_1 \left( \frac{\Delta t}{2} \right)^p \quad (2.34)$$

By subtracting the two equations the unknown coefficient  $C_1$  is determined:

$$y_{\Delta t} - y_{\Delta t/2} = C_1 \left( \Delta t^p - \left( \frac{\Delta t}{2} \right)^p \right) \Rightarrow C_1 = \frac{y_{\Delta t} - y_{\Delta t/2}}{\Delta t^p - \left( \frac{\Delta t}{2} \right)^p} \quad (2.35)$$

By filling in  $C_1$  into equation (2.34):

$$\hat{y} - y_{\Delta t/2} = -(y_{\Delta t} - y_{\Delta t/2}) \frac{\left(\frac{\Delta t}{2}\right)^p}{\Delta t^p - \left(\frac{\Delta t}{2}\right)^p}$$

Next, the global error  $E$  can be determined. However, the global error only gives its bounds (plus or minus) and the best result for the global truncation error is given by:

$$E = |\hat{y} - y_{\Delta t/2}| = \frac{1}{2^p - 1} |y_{\Delta t/2} - y_{\Delta t}| \quad (2.36)$$

It should be noted that for the Euler scheme a truncation error was assumed only. The condition error, due to finite computer precision, becomes important at a much smaller time-step.

This type of time-step analysis was not performed yet and may be necessary during development of the parameter optimization algorithm. In Castelijns (2014) a simple time-step sensitivity was performed. The integration was done with multiple time-steps, varying from  $\Delta t = 0.1$  ms to  $\Delta t = 5$  ms. Afterwards, the simulations were compared against the simulation with the smallest time-step. For the default time-step of 5 ms it was stated that the absolute error was 0.3 nm. In order to be able to perform simulations with sub-nanometer accuracy, the time-step might need to be adjusted.

## 2.7 Model Assumptions

The following assumption were made within the model.

- DGL heat load model is out of scope. No detailed model is available yet. The impact of DGL heat load is small compared to the EUV and IR heat load and therefore a fixed core load can be taken into account.
- No wafer slip assumed, it occurs above a certain thermo-mechanical stress. In other words, thermal stress must remain bounded for model to be valid.
- Continues EUV, light source is always on during the step move.
- Non-uniformity properties of the substrate are neglected, mechanical anisotropy as well. Non-uniform reticle patterns are not assumed (no realistic for customer use).
- Burls geometry simplified.
- DGL incorporated in the  $P_{IR}$  term. In this way the DGL is not present during step moves. When simulation is done, the model is scaled linearly to account for the extra total heat load.



# 3

## Sensitivity Analysis Methods

Sensitivity information of a system is useful for several reasons. It gives valuable insight into the behaviour of the system and it is useful for optimization algorithms. But besides that, sensitivity information is used to construct statistical information about the model parameters as well. The latter will be explained in Chapter 4.

In this chapter, two sensitivity methods are described and investigated for their accuracy and efficiency. These methods are the direct sensitivity method and the adjoint sensitivity method. In Section 3.1, the sensitivity methods are subjected to a static problem. In Section 3.2 the sensitivity methods are subjected to a transient problem. Finally, a conclusion is formed about which methods is best suited for the thermo-mechanical feedforward model.

### 3.1 Static Problems in a General Case

Consider the general formulation for a mechanical static problem:

$$\mathbf{K}\mathbf{u} = \mathbf{f} \tag{3.1}$$

Here,  $\mathbf{K}$  is the stiffness matrix,  $\mathbf{u}$  the displacement vector (state of the system), and  $\mathbf{f}$  the externally applied force vector. The model parameters of the system are stored in the vector  $\mathbf{p} \in \mathbb{R}^M$  where  $M$  is the number of model parameters. When the system is subjected to an optimization problem, an objective function  $\phi$  is required. Often, this objective function is a specified measure of the state of the system. Therefore, the objective function is dependent on the state  $\mathbf{u}$ . However, the state vector is dependent on the model parameters  $\mathbf{p}$  and thus the objective function is represented by:

$$\phi(\mathbf{u}(\mathbf{p})) \tag{3.2}$$

#### EXAMPLE 3.1

---

During this report, several examples will come along to illustrate some principles and definitions that are necessary to gain knowledge of the principle of optimal experiment design. In this example an academic model is constructed which represents an arbitrary mechanical system. The system is represented by,

$$\mathbf{K}\mathbf{u} = \mathbf{f} \tag{3.3}$$

with:

$$\mathbf{K} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} \sin(\omega t) \\ \cos(\omega t - \gamma) \end{bmatrix} \quad (3.4)$$

The model parameter are represented by  $\mathbf{p} = [A \ B]^T$ . The parameter  $\omega$  is a constant and is therefore not assigned as a model parameter. Furthermore,  $t$  represents time and  $\gamma$  is used as an experiment design variable which can influence the force vector  $\mathbf{f}$  and thus the state of the system. The state is determined according:

$$\begin{aligned} u_1 &= \frac{1}{A} \sin(\omega t) \\ u_2 &= \frac{1}{B} \cos(\omega t - \gamma) \end{aligned} \quad (3.5)$$

The objective function of interest is chosen to be:

$$\phi(\mathbf{u}(\mathbf{p})) = \frac{\mathbf{u}^T \mathbf{u}}{2} \quad (3.6)$$

For gradient-based optimization purposes, gradient information of the form,

$$\frac{d\phi(\mathbf{u})}{dp_m} \quad (3.7)$$

is necessary to find the minimum of the objective function in the design space  $\Omega$ , the space where the parameters  $\mathbf{p}$  are defined in. Gradient-based optimization algorithms are often more efficient in finding a minimum compared to direct-search optimization algorithms. Direct-search optimization algorithms uses straight objective function evaluations – in a clever way – in order to find a minimum of the objective function. A drawback of gradient based optimization algorithms is the requirement of differentiable functions.

The direct and adjoint sensitivity methods are described in Section 3.1.1 and Section 3.1.2 respectively. It will be seen that the adjoint method becomes more efficient when the number of model parameters is larger than the number of objective responses.

### 3.1.1 Direct Sensitivity Analysis

Direct sensitivity analysis (SA) for static problems of the form given in equation (3.1) can be calculated by:

$$\frac{d\phi(\mathbf{u}(\mathbf{p}))}{dp_m} = \underbrace{\frac{\partial \phi}{\partial p_m}}_{=0} + \frac{\partial \phi}{\partial \mathbf{u}^T} \frac{d\mathbf{u}}{dp_m} \quad (3.8)$$

The first term is often zero, since the objective function is not directly related to the design parameters  $\mathbf{p}$ , but a measure made from the state  $\mathbf{u}$ . In this section, this term will be neglected. In order to obtain  $\frac{d\mathbf{u}}{dp_m}$ , 3.1 is differentiated. This will give:

$$\frac{d\mathbf{K}}{dp_m} \mathbf{u} + \mathbf{K} \frac{d\mathbf{u}}{dp_m} = \frac{d\mathbf{f}}{dp_m} \quad \Rightarrow \quad \frac{d\mathbf{u}}{dp_m} = \mathbf{K}^{-1} \left( \frac{d\mathbf{f}}{dp_m} - \frac{d\mathbf{K}}{dp_m} \mathbf{u} \right) \quad (3.9)$$

Notice that for every design parameter in  $\mathbf{p}$ , a system of equations must be solved. For large problems with many design parameters this could be an inefficient method, even when the stiffness matrix is factorized. Finally, the sensitivity information is obtained by substitution of 3.9 into 3.8.

$$\boxed{\frac{d\phi(\mathbf{u}(\mathbf{p}))}{dp_m} = \frac{\partial \phi(\mathbf{u}(\mathbf{p}))}{\partial \mathbf{u}^T} \mathbf{K}^{-1} \left( \frac{d\mathbf{f}}{dp_m} - \frac{d\mathbf{K}}{dp_m} \mathbf{u} \right)} \quad (3.10)$$

### 3.1.2 Adjoint Sensitivity Analysis

The adjoint sensitivity method is used in order to improve the efficiency of the calculations by removing the state sensitivity  $\frac{d\mathbf{u}}{dp_m}$  from the objective sensitivity equation. The objective function is rewritten in the form:

$$\phi^*(\mathbf{u}(\mathbf{p})) = \phi(\mathbf{u}(\mathbf{p})) + \boldsymbol{\lambda}^T (\mathbf{K}\mathbf{u} - \mathbf{f}) \quad \forall \boldsymbol{\lambda} \neq \mathbf{0} \quad (3.11)$$

Where  $\lambda$  is a so-called LaGrange multiplier and is free to choose, except the zero vector since that would result in the direct method. Notice that the addition of the second term has no influence on the objective function, since its product is zero. Now follows,

$$\frac{d\phi^*(\mathbf{u}(\mathbf{p}))}{dp_m} = \frac{\partial\phi(\mathbf{u}(\mathbf{p}))}{\partial\mathbf{u}^T} \frac{d\mathbf{u}}{dp_m} + \boldsymbol{\lambda}^T \left( \frac{d\mathbf{K}}{dp_m} \mathbf{u} + \mathbf{K} \frac{d\mathbf{u}}{dp_m} - \frac{d\mathbf{f}}{dp_m} \right) \quad (3.12)$$

where  $\boldsymbol{\lambda}$  is chosen such that:

$$\begin{aligned} \left( \frac{\partial\phi(\mathbf{u}(\mathbf{p}))}{\partial\mathbf{u}^T} + \boldsymbol{\lambda}^T \mathbf{K} \right) \frac{d\mathbf{u}}{dp_m} = 0 &\Rightarrow \boldsymbol{\lambda}^T \mathbf{K} = -\frac{\partial\phi(\mathbf{u}(\mathbf{p}))}{\partial\mathbf{u}^T} \\ &\Rightarrow \boldsymbol{\lambda} = -\mathbf{K}^{-1} \frac{\partial\phi(\mathbf{u}(\mathbf{p}))}{\partial\mathbf{u}} \end{aligned} \quad (3.13)$$

In contrast to the direct sensitivity method, the system must be solved once, which is indicated by the latter equation. Solving the system is the most expensive computational step. By substituting this result into 3.12, a reduced expression for the sensitivity is obtained:

$$\boxed{\frac{d\phi^*(\mathbf{u}(\mathbf{p}))}{dp_m} = \boldsymbol{\lambda}^T \left( \frac{d\mathbf{K}}{dp_m} \mathbf{u} - \frac{d\mathbf{f}}{dp_m} \right)} \quad (3.14)$$

The state sensitivity is not present anymore and it can be observed that only one equation, which involves  $\mathbf{K}^{-1}$ , must be solved in order to obtain  $\boldsymbol{\lambda}$ . When  $\boldsymbol{\lambda}$  is known, the sensitivity for all the design parameters can be determined without intensive computational effort.

## EXAMPLE 3.2

In this example, the direct and adjoint sensitivity method is performed onto the academic model, given in Example 3.1, in order to give more insight about the difference in both methods.

**Direct Method** For the direct sensitivity method, the first step is finding the state derivative, which can be accomplished by using 3.9:

$$\begin{aligned}\frac{d\mathbf{u}}{dp_1} &= \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right) \\ &= \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} u_1 \\ &= \begin{bmatrix} -A^{-1} \\ 0 \end{bmatrix} u_1\end{aligned}\quad (3.15)$$

And,

$$\begin{aligned}\frac{d\mathbf{u}}{dp_2} &= \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \right) \\ &= \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} u_2 \\ &= \begin{bmatrix} 0 \\ -B^{-1} \end{bmatrix} u_2\end{aligned}\quad (3.16)$$

Finally, the direct sensitivity for the objective  $\phi(\mathbf{u}(\mathbf{p}))$  can be obtained by:

$$\begin{aligned}\frac{d\phi(\mathbf{u}(\mathbf{p}))}{dp_1} &= \frac{\partial\phi(\mathbf{u}(\mathbf{p}))}{\partial\mathbf{u}^T} \frac{d\mathbf{u}}{dp_1} = \mathbf{u}^T \begin{bmatrix} -A^{-1} \\ 0 \end{bmatrix} u_1 \\ \frac{d\phi(\mathbf{u}(\mathbf{p}))}{dp_2} &= \frac{\partial\phi(\mathbf{u}(\mathbf{p}))}{\partial\mathbf{u}^T} \frac{d\mathbf{u}}{dp_2} = \mathbf{u}^T \begin{bmatrix} 0 \\ -B^{-1} \end{bmatrix} u_2\end{aligned}\quad (3.17)$$

It can be seen that for each model parameter sensitivity, a full system must be solved as seen in 3.15 and 3.16. For this small example the computations are easy, but when  $\mathbf{K}$  becomes large, it is useful to factorize it according to the Cholesky decomposition, or other techniques, and perform backward- and forward substitutions to obtain the solution.

**Adjoint Method** For the adjoint method the objective function is extended by:

$$\phi^*(\mathbf{u}(\mathbf{p})) = \phi(\mathbf{u}(\mathbf{p})) + \boldsymbol{\lambda}^T (\mathbf{K}\mathbf{u} - \mathbf{f}) \quad \forall \boldsymbol{\lambda} \neq \mathbf{0} \quad (3.18)$$

As was seen in Section 3.1.2, the state sensitivity  $\frac{\partial\mathbf{u}}{\partial p_m}$  can be cancelled when the LaGrange multiplier is chosen as:

$$\boldsymbol{\lambda} = -\mathbf{K}^{-1} \frac{\partial\phi(\mathbf{u}(\mathbf{p}))}{\partial\mathbf{u}} = - \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} \mathbf{u} = - \begin{bmatrix} A^{-1}u_1 \\ B^{-1}u_2 \end{bmatrix} \quad (3.19)$$

By using 3.12, the adjoint sensitivity can be written into:

$$\frac{d\phi^*(\mathbf{u}(\mathbf{p}))}{dp_m} = \boldsymbol{\lambda}^T \left( \frac{d\mathbf{K}}{dp_m} \mathbf{u} - \frac{d\mathbf{f}}{dp_m} \right) \quad (3.20)$$

Which will result in the same sensitivity as the direct method.

$$\begin{aligned}
\frac{d\phi^*(\mathbf{u}(\mathbf{p}))}{dp_1} &= - [A^{-1}u_1 \quad B^{-1}u_2] \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \\
&= u_1 [-A^{-1} \quad 0] \mathbf{u} \\
\frac{d\phi^*(\mathbf{u}(\mathbf{p}))}{dp_2} &= - [A^{-1}u_1 \quad B^{-1}u_2] \left( \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \\
&= u_2 [0 \quad -B^{-1}] \mathbf{u}
\end{aligned} \tag{3.21}$$

The adjoint formulation allows more efficient computations of the sensitivities, but this only holds when the number of responses in  $\phi$  are less than the number of model parameters  $M$ . When calculating the LaGrange multiplier, the system in 3.19 is solved, which is the most expensive calculation in the adjoint method. However, this calculation only needs to be done once. When  $\boldsymbol{\lambda}$  is obtained, calculations of the sensitivities with respect to the model parameter are efficient. This is in contrast to the direct method, since a backward- and forward substitution is required for every sensitivity with respect to the model parameters. When the system is large, the difference in computational effort becomes more clear.

---

## 3.2 Transient Problems

In the following subsections, a description is given on how to obtain a sensitivity analysis for transient problems, e.g., the thermo-mechanical feedforward model described in Chapter 2. The transient problem of the feedforward model is described by the discrete equations:

$$\mathbf{T}_{k+1} = \mathbf{A}\mathbf{T}_k + \mathbf{E}^{-1}\mathbf{Q}_k \tag{3.22}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{T}_k \tag{3.23}$$

Here,  $T_k$  contains the temperature states at time instant  $k = 1, \dots, N-1$ ,  $\mathbf{Q}_k$  is the (heat) input vector, and  $\mathbf{y}_k$  are the substrate deformations. The matrix  $\mathbf{A}$  contains the system dynamics and  $\mathbf{C}$  is known as the deformation matrix, which translates changing temperatures to mechanical deformations. Matrix  $\mathbf{E}^{-1}$  contains the thermal mass density. In the static problem, the sensitivity was given by equation (3.7). For the transient case – together with the line of thought of experiment design – the sensitivity of the objective function is given by,

$$\frac{d\phi(\mathbf{y}_N(\mathbf{p}))}{dp_m} \tag{3.24}$$

where  $\mathbf{y}_N$  are the substrate deformations with  $N$  the last time instant. In Section 3.2.1 the direct sensitivity method for the transient model is described and in Section 3.2.2 the adjoint sensitivity method for the transient model is described.

### 3.2.1 Direct Sensitivity Analysis in Discrete Time

The direct sensitivity method for transient problems can be obtained by:

$$\frac{d\phi(\mathbf{y}_N(\mathbf{p}))}{dp_m} = \frac{\partial\phi(\mathbf{y}_N(\mathbf{p}))}{\partial\mathbf{y}_N^T} \frac{d\mathbf{y}_N}{dp_m} \tag{3.25}$$

Furthermore,

$$\frac{d\mathbf{y}_N}{dp_m} = \frac{d\mathbf{C}}{dp_m} \mathbf{T}_N + \mathbf{C} \frac{d\mathbf{T}_N}{dp_m} \quad (3.26)$$

according to 3.23. In order to obtain the end value  $\frac{d\mathbf{T}_N}{dp_m}$ , the following equation must be solved:

$$\frac{d\mathbf{T}_{k+1}}{dp_m} = \frac{d\mathbf{A}}{dp_m} \mathbf{T}_k + \mathbf{A} \frac{d\mathbf{T}_k}{dp_m} + \frac{d\mathbf{E}^{-1}}{dp_m} \mathbf{Q}_k + \mathbf{E}^{-1} \frac{d\mathbf{Q}_k}{dp_m} \quad (3.27)$$

The system matrix sensitivities can be obtained by a simple finite difference scheme, but the direct sensitivity for the thermal state must be solved by performing a time integration. It can be observed that  $M$  transient simulations are required in order to obtain the sensitivities for all model parameters, which is computationally expensive. This is similar to the direct sensitivity approach for the static case.

### 3.2.2 Adjoint Sensitivity Analysis in Discrete Time

The adjoint sensitivity method is used in order to improve the efficiency of the calculations. The discrete equations (3.22) and (3.23) are used. The objective function is rewritten in the form:

$$\begin{aligned} \phi^*(\mathbf{y}_N(\mathbf{p})) &= \phi(\mathbf{y}_N(\mathbf{p})) + \boldsymbol{\mu}^T (\mathbf{y}_N - \mathbf{C}\mathbf{T}_N) + \sum_{k=0}^{N-1} \boldsymbol{\lambda}_k^T (\mathbf{T}_{k+1} - \mathbf{A}\mathbf{T}_k - \mathbf{E}^{-1}\mathbf{Q}_k) \\ &\forall \boldsymbol{\mu}, \boldsymbol{\lambda} \neq \mathbf{0} \end{aligned} \quad (3.28)$$

In contrast to the static adjoint sensitivity, two types of LaGrange multipliers are present,  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$ , since the transient problem is described in state-space (two equations). As before, the additional parts do not contribute to the objective function, since their products are zero. Subsequently, the sensitivity problem writes:

$$\begin{aligned} \frac{d\phi^*(\mathbf{y}_N(\mathbf{p}))}{dp_m} &= \frac{\partial\phi(\mathbf{y}_N(\mathbf{p}))}{\partial\mathbf{y}_N^T} \frac{d\mathbf{y}_N}{dp_m} + \boldsymbol{\mu}^T \left( \frac{d\mathbf{y}_N}{dp_m} - \frac{d\mathbf{C}}{dp_m} \mathbf{T}_N - \mathbf{C} \frac{d\mathbf{T}_N}{dp_m} \right) + \\ &\sum_{k=0}^{N-1} \boldsymbol{\lambda}_k^T \left( \frac{d\mathbf{T}_{k+1}}{dp_m} - \frac{d\mathbf{A}}{dp_m} \mathbf{T}_k - \mathbf{A} \frac{d\mathbf{T}_k}{dp_m} - \frac{d\mathbf{E}^{-1}}{dp_m} \mathbf{Q}_k - \mathbf{E}^{-1} \frac{d\mathbf{Q}_k}{dp_m} \right) \end{aligned} \quad (3.29)$$

The LaGrange multipliers are free to choose, except zero, and when using smart selections for  $\boldsymbol{\mu}$  and  $\boldsymbol{\lambda}$ , the sensitivity problem reduces. The first term can be cancelled when:

$$\left( \frac{\partial\phi(\mathbf{y}_N(\mathbf{p}))}{\partial\mathbf{y}_N^T} + \boldsymbol{\mu}^T \right) \frac{d\mathbf{y}_N}{dp_m} = 0 \quad \Rightarrow \quad \boldsymbol{\mu} = -\frac{\partial\phi(\mathbf{y}_N(\mathbf{p}))}{\partial\mathbf{y}_N} \quad (3.30)$$

Subsequently,

$$(-\boldsymbol{\mu}^T \mathbf{C} + \boldsymbol{\lambda}_{N-1}^T) \frac{d\mathbf{T}_N}{dp_m} \quad \Rightarrow \quad \boldsymbol{\lambda}_{N-1} = \mathbf{C}^T \boldsymbol{\mu} \quad (3.31)$$

By using 3.30 and 3.31, the sensitivity problem can be rearranged to:

$$\begin{aligned} \frac{d\phi^*(\mathbf{y}_N(\mathbf{p}))}{dp_m} &= \sum_{k=0}^{N-2} \boldsymbol{\lambda}_k^T \left( \frac{d\mathbf{T}_{k+1}}{dp_m} - \frac{d\mathbf{A}}{dp_m} \mathbf{T}_k - \mathbf{A} \frac{d\mathbf{T}_k}{dp_m} - \frac{d\mathbf{E}^{-1}}{dp_m} \mathbf{Q}_k - \mathbf{E}^{-1} \frac{d\mathbf{Q}_k}{dp_m} \right) + \\ &\boldsymbol{\lambda}_{N-1}^T \left( -\frac{d\mathbf{A}}{dp_m} \mathbf{T}_{N-1} - \mathbf{A} \frac{d\mathbf{T}_{N-1}}{dp_m} - \frac{d\mathbf{E}^{-1}}{dp_m} \mathbf{Q}_{N-1} - \mathbf{E}^{-1} \frac{d\mathbf{Q}_{N-1}}{dp_m} \right) - \\ &\boldsymbol{\mu}^T \frac{d\mathbf{C}}{dp_m} \mathbf{T}_N \end{aligned} \quad (3.32)$$

The problem can be reduced further by choosing  $\lambda_k$ :

$$(-\lambda_k^T \mathbf{A} + \lambda_{k-1}^T) \frac{d\mathbf{T}_k}{dp_m} = 0 \quad \Rightarrow \quad \lambda_{k-1} = \mathbf{A}^T \lambda_k \quad (3.33)$$

And,

$$(-\lambda_{N-1}^T \mathbf{A} + \lambda_{N-2}^T) \frac{d\mathbf{T}_{N-1}}{dp_m} \quad \Rightarrow \quad \lambda_{N-2} = \mathbf{A}^T \lambda_{N-1} \quad (3.34)$$

When all intermediate solutions are combined, the final equation for the sensitivity problem writes:

$$\frac{d\phi^*(\mathbf{y}_N(\mathbf{p}))}{dp_m} = -\boldsymbol{\mu}^T \frac{d\mathbf{C}}{dp_m} \mathbf{T}_N + \sum_{k=0}^{N-1} \lambda_k^T \left( -\frac{d\mathbf{A}}{dp_m} \mathbf{T}_k - \frac{d\mathbf{E}^{-1}}{dp_m} \mathbf{Q}_k - \mathbf{E}^{-1} \frac{d\mathbf{Q}_k}{dp_m} \right)$$

(3.35)

It can be seen that the state sensitivities are successfully removed from the objective sensitivity equation. Furthermore, it can be noticed that the adjoint sensitivity problem for transient systems – in contrast to the direct method – reduces the number of simulations (solving the state-space problem) to only one. This is represented by the backward simulation to obtain the LaGrange multipliers  $\lambda_k$ .

### 3.3 Conclusion

The direct- and adjoint sensitivity methods are explained for a static problem and for a transient problem. It is useful to determine which method is best suitable to obtain sensitivities for the thermo-mechanical feedforward model.

For static problems the adjoint method is preferable when the number of responses in the objective is less than the number of model parameters. The direct method needs to solve a system for every model parameter. This can be done efficiently when factorizing the stiffness matrix and performing backward- and forward substitutions. However, when the system is large, the adjoint method is still more efficient.

For transient problems it was seen that the adjoint method requires a backward time integration in order to obtain the LaGrange multipliers. This is far more efficient than using the direct method, which requires a simulation for every model parameter. In theory the adjoint method seems a good alternative to calculate the sensitivities, but one practical problem arises. The storage of the LaGrange multipliers can become a problem when the system matrices are large and the transient simulation is long. In perspective, the existing thermo-mechanical feedforward model consists of approximately  $10^5$  DOFs. The time span of one simulation is around 30 seconds with a time step of 0.001 s, resulting in  $N = 30\,000$  integration steps. One would need a full three dimensional matrix of  $10^5 \times 10^5 \times 30 \cdot 10^3$  to store all the LaGrange multipliers. This requires  $\approx 2.1$  PetaByte of storage.

The thermo-mechanical feedforward model requires transient simulations to obtain overlay deformations. The overlay sensitivity with respect to the model parameters is of interest. In general, the overlay response is much larger than the number of model parameters, since the overlay is directly linked to the mesh size of the model. In addition, the overlay response is not determined by the last temperature state of the simulation (as was assumed in the derivation of the adjoint sensitivity for transient problems), but intermediate temperature states

are necessary. For the adjoint method, this requires extra sets of calculations of the LaGrange multipliers. This makes it almost impossible to store all the LaGrange multipliers in the available memory. In summary, it can be concluded that the direct sensitivity is the most suitable method to obtain sensitivity information for the thermo-mechanical feedforward model.

# 4

## State of the Art Experiment Design

In this chapter, a summary of the most important aspects for experiment design are given. Experiment design is useful in order to enhance the distinguishability or information content of model parameters from the overlay experiment. Multiple literature exists which enlightens the principles of experiment design. However, the information in Rojas et al. (2006) and Bernaerts et al. (2004) is mainly set to application of experiment design and the main working principles of experiment design are briefly described. In Garcia (1999), the basic objective function principles for optimization are briefly explained, but no thorough understanding of the main principles of experiment design are enlighten. The material in Atkinson and Hunter (2012) is very useful in understanding the core principles of experiment design by investigating the parameter regression field. However, the latter is limited by exploration of the type of objective functions used for experiment design. Finally, in Franceschini and Macchietto (2007) and Franceschini and Macchietto (2008) a complete description of the main principles of experiment design, together with a deeper understanding and new competitive objective functions are given.

Throughout this chapter, the standard form of the model is given at first, followed by the required sensitivities used for experiment design. Next, the variance-covariance between parameters is explained and illustrated by means of an example. The measurement errors are explained and in the last section, a thorough description of experiment design with the associated objective functions is given.

### 4.1 Mathematical Model

A mathematical model is considered in the form of a very general system of implicit differential algebraic equations (DAEs). The system is given by:

$$\begin{cases} \mathbf{f}(\dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t), \mathbf{w}, \mathbf{p}, t) = 0 \\ \mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t)) \end{cases} \quad (4.1)$$

where  $\mathbf{f}$  is assumed to have continuous first partial derivatives with respect to its arguments,  $t$  is the time between 0 and  $\tau$ . In most cases,  $\mathbf{h}(\mathbf{x}(t))$  is a “selector” function, selecting those state variables that are measured. Furthermore, the system implies:

- an  $M$ -dimensional set of parameters  $\mathbf{p}$  to be estimated. The vector  $\hat{\mathbf{p}}$  contains the best currently available estimate.
- an  $n_u$ -dimensional set of time-varying control inputs  $\mathbf{u}(t)$ .
- an  $n_w$ -dimensional set of constant controls  $\mathbf{w}$ .
- an  $n_s$ -dimensional set of state variables of the system  $\mathbf{x}(t)$ .

- an  $n_{\text{resp}}$ -dimensional set of measured response variables  $\mathbf{y}(t)$ .
- an  $n_{\text{eq}}$ -dimensional set of DAEs  $\mathbf{f}$  and an  $n_{\text{resp}}$ -dimensional set of relations  $\mathbf{h}$  between the measured response variables  $\text{maty}(t)$  and the state variables  $\mathbf{x}(t)$ .

## 4.2 Sensitivity of Responses to Parameters

The goal of experiment design is to construct experimental data such that fitting a model to this experimental data establishes the model parameter estimates with the lowest possible variance. Therefore, it is important to evaluate how much a variation of the model parameters influences the output of the system. This information is given in a first-order sensitivity matrix  $\mathbf{Q}$  and will be used later.

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial y_1}{\partial p_1} & \cdots & \frac{\partial y_1}{\partial p_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{n_{\text{resp}}}}{\partial p_1} & \cdots & \frac{\partial y_{n_{\text{resp}}}}{\partial p_M} \end{bmatrix} \quad (4.2)$$

The entries in the matrix are obtained by partial differentiation of 4.1. In dynamic models, the matrix  $\mathbf{Q}$  is time-varying and changes at every instant of  $t$ . This time-varying sensitivity matrix  $\mathbf{Q}_r$  is defined by:

$$\mathbf{Q}_r = \begin{bmatrix} \left. \frac{\partial y_r}{\partial p_1} \right|_{t_1} & \cdots & \left. \frac{\partial y_r}{\partial p_M} \right|_{t_1} \\ \vdots & \ddots & \vdots \\ \left. \frac{\partial y_r}{\partial p_1} \right|_{t_{n_{\text{sp}}}} & \cdots & \left. \frac{\partial y_r}{\partial p_M} \right|_{t_{n_{\text{sp}}}} \end{bmatrix} \quad (4.3)$$

Here,  $\mathbf{Q}_r$  represents the sensitivity matrix of the  $r$ th response variable and  $n_{\text{sp}}$  represents the sampling time instant.

## 4.3 Variance-Covariance of Parameters

The variance-covariance matrix indicates the correlation of parameters in terms of their associated responses. In probability and statistics theory, the covariance is a measure of how much two random variables – or parameters – change together in relation. This information is valuable when an experiment measurement is used for parameter estimation. The variance-covariance matrix reveals the underlying connection for parameters in a certain model (e.g., the feedforward model). The variance-covariance matrix  $\mathbf{V}$  in the parameter space is known by:

$$\mathbf{V} = (\mathbf{Q}^T \mathbf{Q})^{-1} \quad (4.4)$$

Here,  $\mathbf{Q}$  is the sensitivity matrix defined in 4.2. The correlation coefficient matrix  $\mathbf{C}$  is a scaled version of the covariance matrix  $\mathbf{V}$  and is used in practice. The correlation coefficient matrix  $\mathbf{C}$  is defined by:

$$\mathbf{C} = \frac{V_{i,j}}{\sqrt{V_{i,i}} \sqrt{V_{j,j}}} \quad i, j = 1, \dots, M \quad (4.5)$$

Each entry in  $\mathbf{C}$  ranges from -1 to 1 and indicates the correlation between the different parameters. When the absolute value is 1, it indicates a 100% correlation between two parameters and, hence, they are not distinguishable from the experiment measurement. When the value of the  $\mathbf{C}$  matrix is zero, it indicates an orthogonal relation between the concerned parameters and, consequently, the parameters are uniquely distinguishable from the experiment measurement.

Note that the diagonal terms are always equal to 1. It can be concluded that it is desired to obtain a square identity matrix of dimension  $\mathbb{R}^M$ , since it will indicate orthogonality between all parameters and thus unique distinguishability from the experiment measurement.

According to Dekking et al. (2005, p. 140) a covariance of zero between two parameters (off-diagonal entries in  $\mathbf{C}$ ) does not imply two parameters being independent, only uncorrelated. This implies that one parameter could influence the remaining parameter. However, the opposite is true. When two parameters are independent from each other, they have a covariance of zero. It is not known how this will translate in a successful experiment design so far, but it is useful to keep in mind. Furthermore, it should be noted that the correlation matrix  $\mathbf{C}$  is always symmetric.

#### EXAMPLE 4.1

---

In this example, the correlation matrix  $\mathbf{C}$  will be clarified by means of some illustrations. To illustrate the correlation matrix, the model explained in Example 3.1 is used, but with a slight modification. The system is still represented by,

$$\mathbf{K}\mathbf{u} = \mathbf{f}$$

where the stiffness matrix remains:

$$\mathbf{K} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \quad (4.6)$$

However, the displacement vector  $\mathbf{u}$  is represented by:

$$\mathbf{u} = \begin{bmatrix} \sin(\omega t) \\ \cos(\omega t - \gamma) \end{bmatrix} \quad (4.7)$$

This can be interpreted as a prespecified deformation profile for the mechanical system. As a consequence, the force response  $\mathbf{f} = [f_1 \ f_2]^T$  is given by:

$$\begin{aligned} f_1 &= A \sin(\omega t) \\ f_2 &= B \cos(\omega t - \gamma) \end{aligned} \quad (4.8)$$

In this example,  $\omega$  is a constant and fixed at  $1 \text{ rad s}^{-1}$ . The variance-covariance matrix  $\mathbf{V}$  is necessary to obtain the correlation matrix  $\mathbf{C}$ . The time-varying sensitivity matrix  $\mathbf{Q}_r$ , given in 4.3, is necessary to construct  $\mathbf{V}$ . The response  $\mathbf{f}$  is measured and consists of two individual responses  $f_1$  and  $f_2$ , hence the index  $r$  goes up to two. Since the response  $\mathbf{f}$  is measured, the sensitivity is linear in the model parameters  $\mathbf{p}$  and given by:

$$\begin{aligned} \mathbf{Q}_1 &= \begin{bmatrix} \frac{\partial f_1}{\partial p_1} \Big|_{t_1} & \frac{\partial f_1}{\partial p_2} \Big|_{t_1} \\ \vdots & \vdots \\ \frac{\partial f_1}{\partial p_1} \Big|_{t_{\text{end}}} & \frac{\partial f_1}{\partial p_2} \Big|_{t_{\text{end}}} \end{bmatrix} = \begin{bmatrix} \sin(\omega t_1) & 0 \\ \vdots & \vdots \\ \sin(\omega t_{\text{end}}) & 0 \end{bmatrix} \\ \mathbf{Q}_2 &= \begin{bmatrix} \frac{\partial f_2}{\partial p_1} \Big|_{t_1} & \frac{\partial f_2}{\partial p_2} \Big|_{t_1} \\ \vdots & \vdots \\ \frac{\partial f_2}{\partial p_1} \Big|_{t_{\text{end}}} & \frac{\partial f_2}{\partial p_2} \Big|_{t_{\text{end}}} \end{bmatrix} = \begin{bmatrix} 0 & \cos(\omega t_1 - \gamma) \\ \vdots & \vdots \\ 0 & \cos(\omega t_{\text{end}} - \gamma) \end{bmatrix} \end{aligned} \quad (4.9)$$

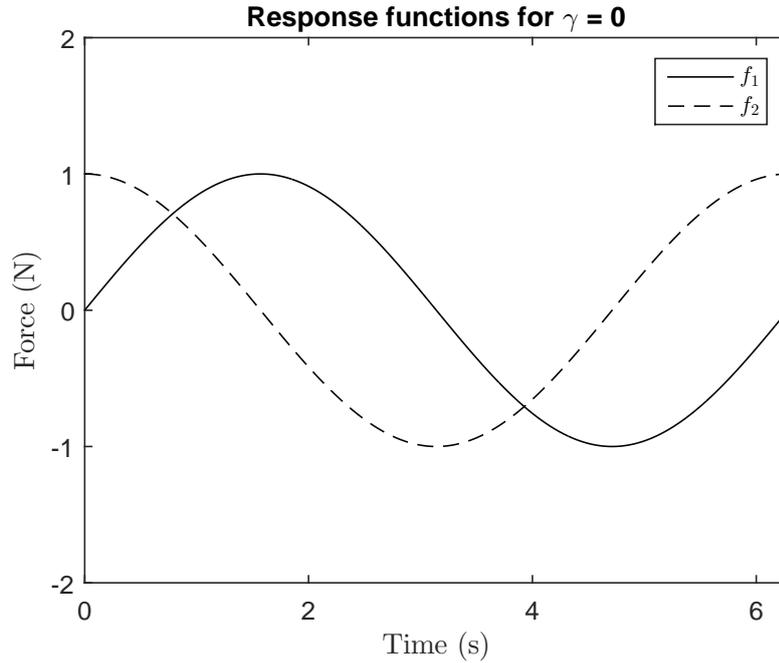
The time  $t$  varies from  $t_1$  to  $t_{\text{end}}$ . In order to construct the variance-covariance matrix, a slight modification of 4.4 is necessary due to the time-varying sensitivity matrices (this will be seen in Section 4.5.1).

$$\mathbf{V} = \left[ \sum_{r=1}^2 \sum_{s=1}^2 \mathbf{Q}_r^T \mathbf{Q}_s \right]^{-1} \quad (4.10)$$

The model parameters  $A$  and  $B$  are both set to the value 1. Three different cases will be investigated, which consist of three different values for the experiment design parameter, i.e.,  $\gamma = 0$ ,  $\gamma = 1/4\pi$  and  $\gamma = 1/2\pi$ .

**Case 1** The experiment design parameter  $\gamma$  is evaluated at a value of 0. This implies a pure sine and cosine function in the displacement vector. An artificial experiment is performed by calculating the force response when the system is constrained in  $\mathbf{u}$ . The displacement vector varies during time and the obtained response  $\mathbf{f}$  is depicted in Figure 4.1.

$$\mathbf{Q}_1 = \begin{bmatrix} \sin(\omega t_1) & 0 \\ \vdots & \vdots \\ \sin(\omega t_{\text{end}}) & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0 & \cos(\omega t_1) \\ \vdots & \vdots \\ 0 & \cos(\omega t_{\text{end}}) \end{bmatrix} \quad (4.11)$$



**Figure 4.1:** System response with  $\gamma = 0$ . A pure sine and cosine function can be seen, which are orthogonal to each other.

A sine and cosine response is visible, as is expected. The correlation between the model parameters  $A$  and  $B$  in  $\mathbf{f}$  can be calculated according to 4.5 with the time-varying sensitivity matrix in 4.11. For  $\gamma = 0$ , the correlation matrix  $\mathbf{C}$  for the model parameters is expressed by:

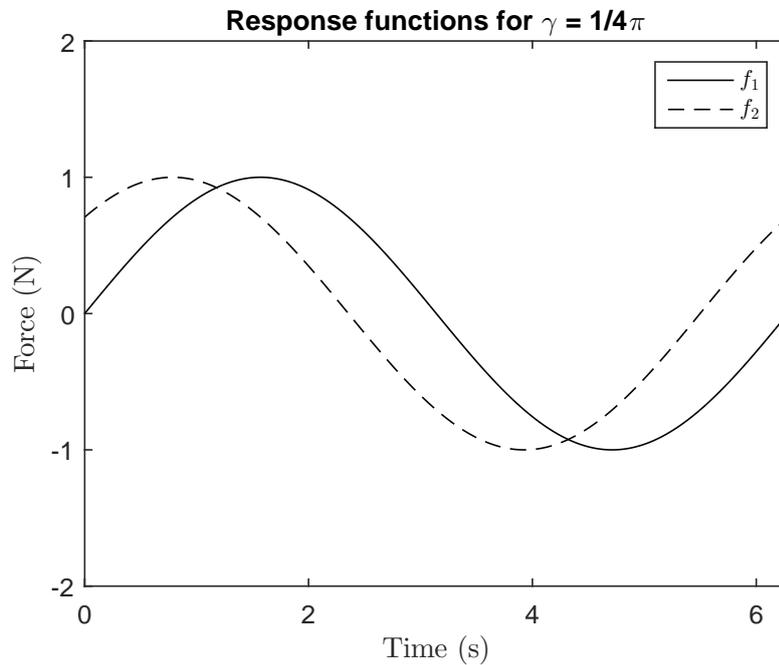
$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

It can be seen that  $\mathbf{C}$  represents an identity matrix. This indicates an orthogonal relationship of  $A$  and  $B$  in the measured system response  $\mathbf{f}$  and no correlation exists. The model parameters are perfectly distinguishable from the measurement.

**Case 2** In the following case, the experiment design parameter is set to  $\gamma = 1/4\pi$ . The cosine function in the displacement vector is shifted in phase. This has an influence in the time-varying sensitivity matrix.

$$\mathbf{Q}_1 = \begin{bmatrix} \sin(\omega t_1) & 0 \\ \vdots & \vdots \\ \sin(\omega t_{\text{end}}) & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0 & \cos(\omega t_1 - \frac{1}{4}\pi) \\ \vdots & \vdots \\ 0 & \cos(\omega t_{\text{end}} - \frac{1}{4}\pi) \end{bmatrix} \quad (4.12)$$

A new experiment is performed with a different experiment parameter  $\gamma$ . The obtained response  $\mathbf{f}$  is depicted in Figure 4.2.



**Figure 4.2:** System response with  $\gamma = \frac{1}{4}\pi$ . A phase shift in  $f_2$  can be observed.

As before, a sine and cosine function are illustrated, but the cosine function is shifted in phase due to the experiment design parameter. The correlation between the model parameters  $A$  and  $B$  in  $\mathbf{f}$  is calculated and given by:

$$\mathbf{C} = \begin{bmatrix} 1 & -0.706 \\ -0.706 & 1 \end{bmatrix}$$

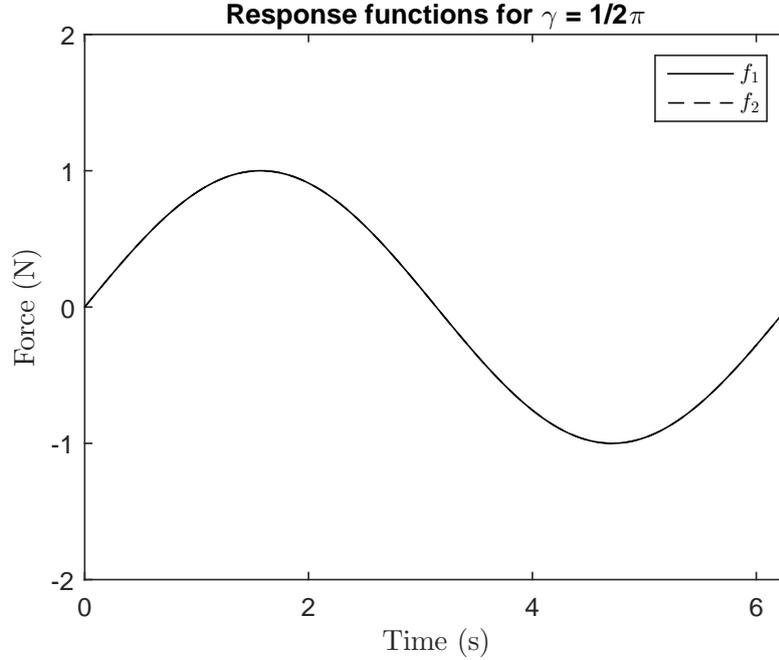
The correlation matrix  $\mathbf{C}$  is not an identity matrix, since the off-diagonal terms are negative non-zero. This indicates a negative correlation between parameter  $A$  and  $B$ , or rather, the parameters  $A$  and  $B$  are less distinguishable from the response  $\mathbf{f}$ . When performing least squares estimation, it might take more iterations to estimate the model parameters  $A$  and  $B$  from the measured response. The found estimates are most likely uncertain, since the combination of  $A$  and  $B$  is estimated as well.

**Case 3** In this last case, the experiment design parameter is set to  $\gamma = 1/2\pi$ . With this amount of phase shift, the cosine function can be rewritten in terms of a sine

function without any phase. Therefore, the time-varying sensitivities are given by:

$$\mathbf{Q}_1 = \begin{bmatrix} \sin(\omega t_1) & 0 \\ \vdots & \vdots \\ \sin(\omega t_{\text{end}}) & 0 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0 & \sin(\omega t_1) \\ \vdots & \vdots \\ 0 & \sin(\omega t_{\text{end}}) \end{bmatrix} \quad (4.13)$$

A new experiment is performed with a different experiment parameter  $\gamma$ . The obtained response  $\mathbf{f}$  is depicted in Figure 4.3.



**Figure 4.3:** System response with  $\gamma = \frac{1}{2}\pi$ . The phase shift in  $f_2$  is such that both responses are overlapping.

From the figure, it can be seen that both responses overlap each other and cannot be distinguished any more. This can be seen by the time-varying sensitivities as well. Both  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  contain the same information. The correlation matrix is given below.

$$\mathbf{C} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

It can be seen that  $\mathbf{C}$  is degenerated and has rank 1. This implies that the model parameters are no longer distinguishable, since the parameter space has been reduced to 1 dimension. This concludes that the model parameters can not be individually estimated, only their combination. This is the worst-case scenario for experiment design and is far from optimal.

## 4.4 Response and Measurement Errors

The measured response of the  $i$ th experiment can be expressed as:

$$\hat{\mathbf{y}}_i(t) = \mathbf{y}_i(t) + \boldsymbol{\eta}_i(t) \quad i = 1, \dots, n_{\text{exp}} \quad (4.14)$$

with  $\hat{\mathbf{y}}_i$  as the true response of all response variables and  $\boldsymbol{\eta}_i(t)$  the experimental measurement error in the  $i$ th experiment. This stochastic component is assumed to be multivariate normally distributed with,

$$E[\boldsymbol{\eta}_i] = 0 \quad \text{and} \quad E[\boldsymbol{\eta}_i, \boldsymbol{\eta}_j] = \begin{cases} \Sigma_y & i = j \\ 0 & i \neq j \end{cases} \quad (4.15)$$

where  $\Sigma_y$  is the variance-covariance matrix of the experimental errors:

$$\Sigma_y = \begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1 y_2}^2 & \cdots & \sigma_{y_1 y_{n_{\text{resp}}}}^2 \\ \sigma_{y_2 y_1}^2 & \sigma_{y_2}^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{y_{n_{\text{resp}}} y_1}^2 & \cdots & \cdots & \sigma_{y_{n_{\text{resp}}}}^2 \end{bmatrix} \quad (4.16)$$

Equation (4.15) implies that different measurements are independent, but measurements of multiple responses ( $y_1, y_2, \dots, y_{n_{\text{resp}}}$ ) taken under the same experimental conditions are subjected to a certain variance. The off-diagonal terms of matrix  $\Sigma_y$  imply a correlation between the output responses. It is assumed that noise is not related between the individual response variables  $\hat{\mathbf{y}}_i$ . Hence, the off-diagonal terms in 4.16 are zero. In Franceschini and Macchietto (2007), a description is given on how to find the elements in 4.16.

## 4.5 Experiment Design

When designing an optimal experiment, the variables that will control the experiment are determined. All the control variables which can be manipulated and optimized during the design are collected in the so-called design vector  $\boldsymbol{\varphi}$ .

$$\boldsymbol{\varphi} = \boldsymbol{\varphi}(\mathbf{u}(t), \mathbf{y}_0, \tau, \mathbf{w}) \quad (4.17)$$

Here,  $\mathbf{u}(t)$  is the vector of time-varying input controls which will influence the experiment,  $\mathbf{y}_0$  contains the initial conditions,  $\mathbf{w}$  is the time-invariant input controls vector, and  $\tau$  represents the experiment duration. The design vector  $\boldsymbol{\varphi}$  must be constrained to lie within the experiment design space  $\Phi$ . The space  $\Phi$  is the space where the experiment can exist, or in other words, the allowable range of all the design variables included in the vector  $\boldsymbol{\varphi}$ . For example, illumination of the substrate cannot be performed outside the surface area of the substrate and is therefore excluded from the experiment design space.

### 4.5.1 Standard Metrics to Measure the Information Content of an Experiment

Improving parameter precision in mathematical terms is equivalent to decreasing the size of the confidence regions of the model parameters. In other words, reducing the values in the parameter variance-covariance matrix. By using the Bayesian formulation of the high posterior density region, the following expression for the marginal posterior covariance can be obtained:

$$\mathbf{V}(\hat{\mathbf{p}}, \boldsymbol{\varphi}) = \left[ \sum_r^{n_{\text{resp}}} \sum_s^{n_{\text{resp}}} \tilde{\sigma}_{rs} \mathbf{Q}_r^T \mathbf{Q}_s + \Sigma_p(\hat{\mathbf{p}})^{-1} \right]^{-1} \quad (4.18)$$

where  $n_{\text{resp}}$  is the number of responses. The parameter variance-covariance matrix  $\mathbf{V}_{M \times M}$  is symmetric and depends on:

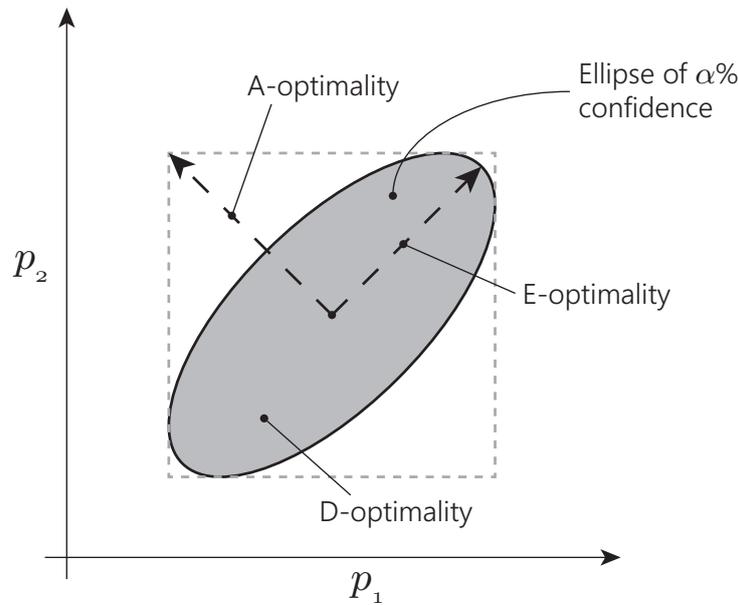
- The dynamic sensitivity coefficients of the  $r$ th response,  $\mathbf{Q}_r$  in equation (4.3).

- The variance-covariance matrix of the experimental measurements  $\Sigma_y$  ( $\tilde{\sigma}_{rs}$  is the (r,s) element of the inverse of  $\Sigma_y$ , see equation (4.16)).
- An initial approximation of the variance-covariance matrix ( $\Sigma_p$ ) of the model parameters, which contains preliminary information on the parameter uncertainty (same structure as equation (4.16)). This initial approximation can often be neglected, according to Box and Lucas (1959).

In Franceschini and Macchietto (2007) an adapted version of equation (4.18) is given, which is capable of defining an information matrix  $\mathbf{M}$  which includes multiple experiments. The information matrix can be rewritten into the variance-covariance matrix by  $\mathbf{V}(\hat{\mathbf{p}}, \varphi) = [\mathbf{M}(\hat{\mathbf{p}}, \varphi)]^{-1}$  and is given by:

$$\mathbf{M}(\hat{\mathbf{p}}, \varphi) = \sum_{i=1}^{n_{\text{exp}}} \sum_r^{n_{\text{resp}}} \sum_s^{n_{\text{resp}}} \tilde{\sigma}_{rs,i} \mathbf{Q}_{r,i}^T \mathbf{Q}_{s,i} + \mathbf{M}_0 \quad (4.19)$$

Here,  $\mathbf{M}_0$  is optional and represents the initial approximation of the parameter uncertainty (similar to equation (4.18)), and  $n_{\text{exp}}$  is the number of experiments. As can be seen, the time-varying sensitivity matrix might change when different experiments with different experimental conditions are performed.



**Figure 4.4:** Geometrical interpretation of the standard criteria for the experiment design. The grey area represents the confidence region of the parameters (typically, 90% or 95%).

The experiment design algorithm can either minimise a suitable metric of  $\mathbf{V}$ , or maximise the equivalent metric of  $\mathbf{M}$  by varying the elements of the design vector  $\varphi$ . Various metrics have been suggested that all aim at the goal of mapping the information in the variance-covariance matrix into a scalar number. The most common criteria are:

1. D-optimality, which minimises the determinant of the variance-covariance matrix, or maximises the determinant of the information matrix.
2. E-optimality, which minimises the largest eigenvalue of the variance-covariance matrix, or maximises the smallest eigenvalue of the information matrix.
3. A-optimality, which minimises the trace of the variance-covariance matrix, or maximises the trace of the information matrix.

In Figure 4.4, a geometrical interpretation is illustrated for a two parameter case. The uncertainty associated with the estimates of the parameters can be represented by joint parameter confidence regions (typically, at  $\alpha = 90\%$  or  $95\%$  confidence region). From Figure 4.4, the D-optimal criterion aims at minimising the area of this joint confidence region, the E-optimal criterion aims at reducing the size of its major axis, and the A-optimal criterion aims at reducing the dimensions of the enclosing box around the joint confidence region.

The D-optimal criterion is the most used of the three standard criteria cited above, due to some appealing properties:

- Easy geometrical interpretation, see Figure 4.4.
- Theoretical invariance with respect to any non-degenerated transformation applied on the model parameters (such as scaling).
- Yielding of optimal experiments that correspond to replications of a small number of different experimental conditions.

However, the D-optimal criterion is not free from drawbacks. It tends to give excessive importance to the most sensitive model parameter. Although the confidence region volume is reduced, due to the decrease in the variance of this sensitive model parameter, the uncertainties of all the other parameters may remain large.

By some authors (Goodwin, 1987; Munack and Posten, 1989; Vanrolleghem et al., 1995; Zullo, 1991; Pauw, 2005), the A-optimal criterion is considered unreliable. This criterion does not include the off-diagonal terms of the variance-covariance matrix in the objective function, which is an appreciable loss of information. As a consequence, this criterion is not desired for cases of high correlations between the model parameters  $\mathbf{p}$ .

The drawbacks of the standard criteria have led to the search for alternatives. As a result, the modified E-optimality was created. This criterion aims at minimizing the condition number of the information matrix. The condition number is the ratio between the largest and smallest eigenvalue. This enforces a joint confidence region of the parameters that is as spherical as possible. The modified E-optimality is often used, but it is valid only for two-parameter models according to Franceschini and Macchietto (2007). Also, it has a significant drawback: its mathematical property (ratio of eigenvalues) implies that the criterion is discontinuous and, therefore, can cause convergence problems when used with a gradient-based optimisation algorithm. Another limitation of the modified E-optimality criterion is that due to the pure shape criterion, ellipsoids with large volumes may be obtained since eigenvalues could be increased. One could think of combining several criteria, i.e., a combination of the modified E-optimality criterion and the D-optimality criterion. However, this could be difficult to implement successfully, since the D-optimality criterion tends to lose the spherical shape and the modified E-optimality criterion does not guarantee an equal or less confidence area (determinant) during optimization. It could be possible that one criterion reduces the confidence area, while the other criterion increases the confidence area.

The described standard criteria seem to be difficult to implement successfully for the feedforward model. The feedforward model contains many parameters and is expected to be correlated. The A-optimal criterion is, according to literature, not reliable and the modified E-optimal criterion only holds for two parameters. Therefore, these criteria are not suitable. The D-optimal and E-optimal criteria might be useful when a subset of parameters in the feedforward model are used, although a reduction of the correlation between parameters is preferable, which is neglected in these criteria.

---

**EXAMPLE 4.2**

In this example, the confidence ellipsoids together with the optimality criteria are illustrated. The construction of the confidence ellipsoids is explained in Chapter 5. The slightly modified academic model, used in Example 4.1, is considered again.

The confidence ellipsoids, or confidence regions, graphically illustrate the region where there is a  $1 - \alpha$  confidence that the true model parameters lie within that region after parameter estimation (e.g., least squares). A small confidence region implies low variation of the model parameters and thus implies a high accuracy of the associated model parameters. Experiment design focusses on reducing the confidence regions. Note, for two model parameters, the concept can be made visible with ellipsoids. This cannot be graphically illustrated for more than 2 model parameters, but the concept remains the same.

In order to construct the confidence regions, the information matrix must be built according to equation (4.19). The variance  $\sigma^2$  is set to 1 N. For now, this value is arbitrarily chosen and represents the variance of the measurement noise while measuring the response  $\mathbf{f}$ . It is assumed that there is no difference of the variance between  $f_1$  or  $f_2$ . The initial approximation  $\mathbf{M}_0$  is not required, and thus the information matrix for one experiment  $n_{\text{exp}} = 1$  is given by:

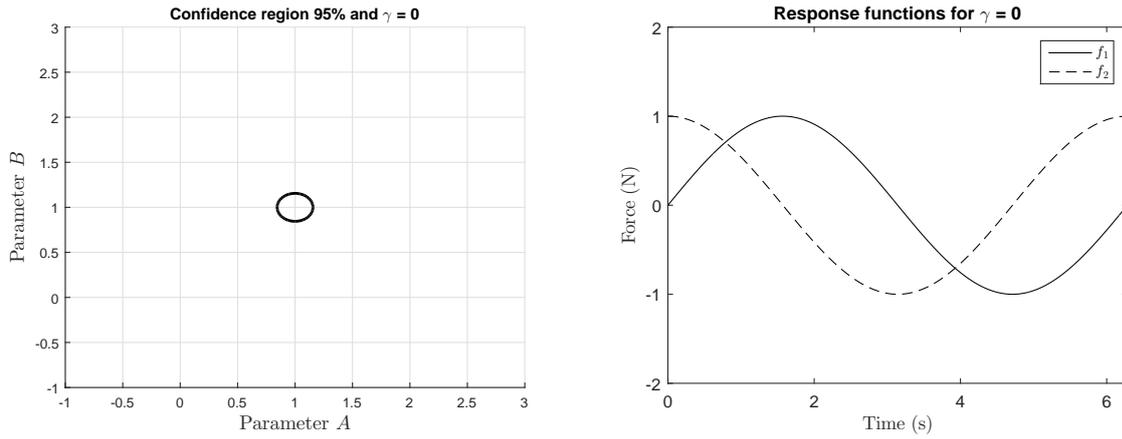
$$\mathbf{M} = \frac{1}{\sigma^2} \sum_{r=1}^2 \sum_{s=1}^2 \mathbf{Q}_r^T \mathbf{Q}_s \quad (4.20)$$

As in Example 4.1, three different experiments are performed by changing the experiment design parameter to  $\gamma = 0$ ,  $\gamma = 1/4\pi$  and  $\gamma \approx 1/2\pi$ . For these cases, the confidence regions are constructed. As before, the model parameters  $A$  and  $B$  are equal to 1.

**Case 1** The experiment design parameter is set to  $\gamma = 0$ . An experiment is performed and the information matrix  $\mathbf{M}$  is calculated:

$$\mathbf{M} = \begin{bmatrix} 250 & 0 \\ 0 & 250 \end{bmatrix} \quad (4.21)$$

In Figure 4.5, the 95% confidence region of the parameters  $A$  and  $B$  is illustrated. As was seen before, this situation creates an orthogonal relation of the geometric functions with respect to the parameters  $A$  and  $B$ . The correlation matrix turned out to be an identity matrix, hence the information matrix is a diagonal matrix with the eigenvalues located at the diagonal. The confidence region of the parameters is constructed by the variance-covariance matrix  $\mathbf{V}$ , which is the inverse of  $\mathbf{M}$ . As can be seen in the figure, the confidence region is circular, since the eigenvalues of the information matrix are equal.



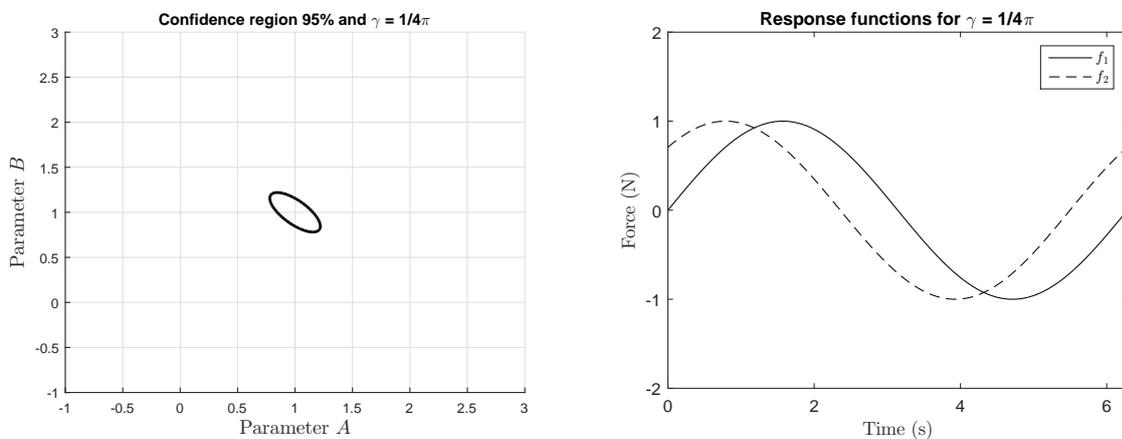
**Figure 4.5:** 95% confidence region for  $\gamma = 0$ . Confidence region is small which translates to high accuracy of model parameters. The confidence shape is circular due to zero correlation and equal eigenvalues. As was seen before, the responses are orthogonal.

Next, the D-, E-, and A-optimality criteria are calculated to gain insight of their working principle. The D-optimality criterion determines the surface area of the confidence region, the E-optimality criterion determines the smallest eigenvalue of the information matrix and the A-optimality criterion calculates the trace of the information matrix. The criteria are listed in Table 4.1 under case 1.

**Case 2** In this case, the experiment design parameter is set to  $\gamma = 1/4\pi$ . An experiment is performed and the information matrix  $\mathbf{M}$  is calculated:

$$\mathbf{M} = \begin{bmatrix} 250 & 176.4 \\ 176.4 & 250 \end{bmatrix} \quad (4.22)$$

The confidence region is illustrated in Figure 4.6, which is an ellipsoidal shape due to the correlation and thus the difference in magnitude of the eigenvalues. This correlation was found in Example 4.1 as well. The set-up of the performed experiment is less desirable for parameter estimation due to increasing correlation (or reduction of distinguishability) and an increase of the uncertainty, which is represented by the increasing confidence region.



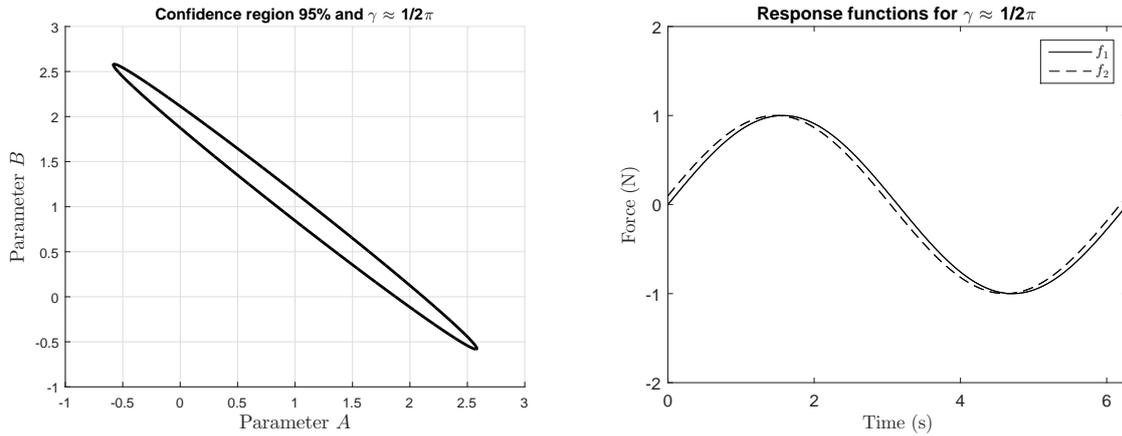
**Figure 4.6:** 95% confidence region for  $\gamma = 1/4\pi$ . Confidence region is slightly larger than for  $\gamma = 0$ , which yields in less accuracy of model parameters.

The D-, E-, and A-optimality criteria are calculated and are listed in Table 4.1 under case 2.

**Case 3** In this case, the experiment design parameter is set to  $\gamma \approx 1/2\pi$ , since  $\gamma = 1/2\pi$  results in a singular information matrix and the confidence region cannot be constructed. In this way, it can be seen what happens when the model parameters are close to non-distinguishability. As before, an experiment is performed and the information matrix  $\mathbf{M}$  is calculated:

$$\mathbf{M} = \begin{bmatrix} 249.5 & 248.3 \\ 248.3 & 249.5 \end{bmatrix} \quad (4.23)$$

The information matrix is almost singular. The confidence region is obtained and illustrated in Figure 4.7. The parameter space in  $\mathbf{M}$  is almost reduced to rank 1, which is indicated by the small eigenvalue that is approaching zero (see Table 4.1 the E-optimality criterion for case 3). The bounds of the confidence region are significantly larger than previous cases. This results in very poor accuracy of the model parameters and, therefore, this set-up of the experiment is the worst-case scenario for parameter estimation.



**Figure 4.7:** 95% confidence region for  $\gamma \approx 1/2\pi$ . The bounds of the confidence region are significantly larger than for previous cases. Accuracy of the model parameters is very poor and this set-up of the experiment is the worst-case scenario.

The D-, E-, and A-optimality criteria are calculated to see how the criteria are effected by the poor experiment design. The result is listed in the Table 4.1 under case 3.

**Table 4.1:** Optimality criteria for case 1 with experiment design parameter  $\gamma = 0$ , case 2 with  $\gamma = 1/4\pi$  and case 3 with  $\gamma \approx 1/2\pi$ .

	Case 1	Case 2	Case 3
CRITERION	OBJECTIVE VALUE	OBJECTIVE VALUE	OBJECTIVE VALUE
D	$1.60 \cdot 10^{-5}$	$3.20 \cdot 10^{-5}$	$1.70 \cdot 10^{-3}$
E	250	73.3	1.2
A	500	500	499

**Conclusion** By illustrating the confidence regions graphically, it is evident that non-correlated orthogonal response sensitivities result in the highest accuracy of the model parameters during parameter estimation. When observing the corresponding optimality criteria, listed in Table 4.1, it is noticeable that the A-optimality criterion seems invariant for changes of the experiment. This can be explained by the fact that the experiment design parameter  $\gamma$ , in this model, affects the correlation only. In Section 4.5.1, it was explained that this criterion was not classified as reliable by several authors due to the fact that the correlation is neglected. From this example, it is clear that the trace of  $\mathbf{M}$  never changes.

The  $D$ -optimality seems an appropriate criterion when looking at the tables. The values correspond to the surface area of the confidence regions, which matches with the illustrations. It should be noted that the  $D$ -optimality criterion may not perform well when the number of model parameters is increased.

The  $E$ -optimality seems to be an appropriate criterion too. The decreasing values in the tables are corresponding to the increasing confidence regions. However, other eigenvalues are not included and might be affected in a negative sense.

## 4.5.2 Alternative Formulations Based on Standard Experiment Design Criteria

It is difficult to find a criterion that is able to reduce the complex information in the variance-covariance (or information) matrix to a scalar value. The standard based criteria have their strong points, but they are not free from drawbacks. Therefore, several alternative formulations have been proposed for objective functions  $\phi$  which are based on (minor) modifications of the standard criteria.

One of the proposed modifications is the following objective function:

$$\phi = (1 - \beta) \log \bar{\Lambda} - \beta \log |\mathbf{M}| \quad (4.24)$$

Here,  $\beta$  is a weighting factor and  $\bar{\Lambda}$  and  $|\mathbf{M}|$  are, respectively, the condition number and the determinant of the information matrix. It can be noticed that this objective function is a generalization of the  $D$ -,  $E$ -, and modified  $E$ -optimality criteria. By choosing  $\beta$  equal to 0, 0.5 and 1, the modified  $E$ -optimal, the  $E$ -optimal and the  $D$ -optimal criterion are obtained, respectively. Unfortunately, no recommendations are given on how to choose the weighting factor  $\beta$ . Also, these criteria only work well for two-parameter models. The feedforward model contains more than two parameters and, therefore, this criterion is not suitable for experiment design for the feedforward model.

A new alternative criterion to the traditional  $D$ -optimal criterion was proposed by Pritchard and Bacon (1978) Franceschini and Macchietto (2007). This criterion measures the overall correlation of parameters and uses this information directly as an objective function to be minimized:

$$\phi = \left( \sum_{i,j} \frac{C_{ij}^2}{(M^2 - M)} \right)^{1/2} \quad (4.25)$$

Where  $C_{ij}$  is a correlation entry of the correlation matrix and  $M$  is the number of model parameters. It has been shown that these kind of design criteria are limited to two-parameter models. In addition, it was shown that design criteria that aim only at reducing parameter correlation, are likely to produce a large confidence region for the parameters. This is the result when no further information content of  $\mathbf{M}$  is included in the design.

### 4.5.3 Association Between Eigenvalues and Parameters

In the next subsection, new metrics are developed to measure the information content. In order to understand the underlying principle of these new metrics, the association between eigenvalues-eigenvectors with their corresponding model parameter will be clarified.

For one experiment, the information matrix can be written as:

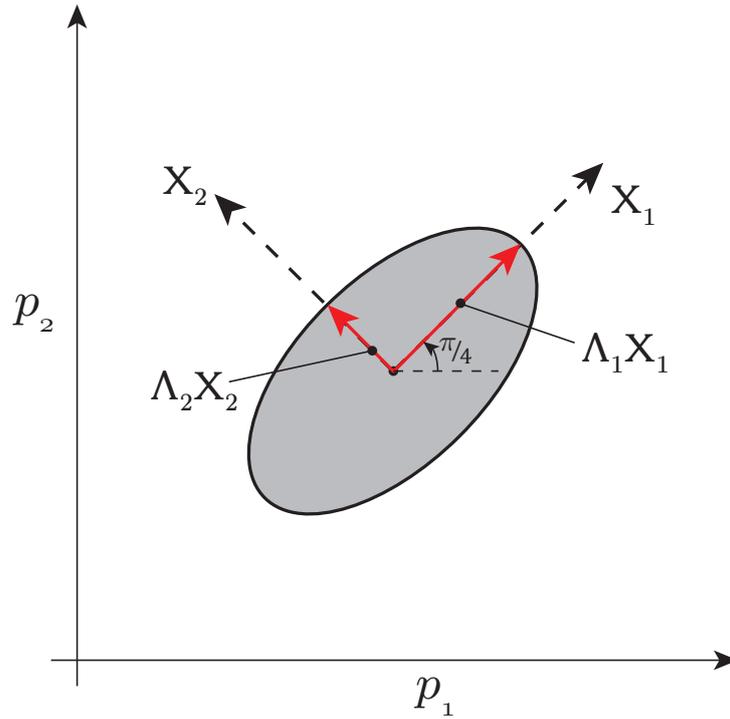
$$\mathbf{M}(\hat{\mathbf{p}}, \varphi) = \sum_{r=1}^{n_{\text{resp}}} \sum_{s=1}^{n_{\text{resp}}} \tilde{\sigma}_{rs} \mathbf{Q}_r^T \mathbf{Q}_s \quad (4.26)$$

This result is achieved by using equation (4.19). The information matrix can be diagonalized into eigenvalues with their corresponding eigenvectors by:

$$\mathbf{M} = \mathbf{X} \mathbf{\Lambda} \mathbf{X}^T \quad (4.27)$$

where  $\mathbf{\Lambda}$  is a diagonal matrix containing the eigenvalues of the information matrix and  $\mathbf{X}$  is a matrix containing the normalized eigenvectors corresponding to its eigenvalue. From the inverse of the information matrix, a hyper-ellipsoid can be obtained in the parameter space, which represents the true joint confidence region of the parameters. The eigenvectors represent the orientation of the ellipsoidal confidence region with respect to the parameter axes, whereas the eigenvalues determine the relative lengths of the axes of the ellipsoid. See Figure 4.8 for a geometrical interpretation.

When the eigenvectors are parallel to the parameter axes, the correlation between the parameters is zero. A rotation of  $\pi/4$  rad of the principle axis of the ellipsoid with respect to the parameter axis indicates a full correlation of the parameters. Furthermore, the eigenvalue of the information matrix indicates the experiment information content of the corresponding parameter. A large eigenvalue in  $\mathbf{M}$  will result in a confidence region with a small uncertainty for that particular parameter, which is desired.



**Figure 4.8:** Interpretation of eigenvalues and eigenvectors. The eigenvectors represent the orientation of the ellipsoidal confidence region with respect to the parameter axes. The eigenvalues determine the relative lengths of the axes of the ellipsoid. In the figure, the parameters are fully correlated.

#### 4.5.4 Newly Developed Metrics for Information Content Measurement

A new approach was developed by Franceschini and Macchietto (2008). Four novel criteria were proposed, all of which work directly on the parameter correlation matrix in order to reduce the interaction between (selected) parameters by means of suitable objective functions and/or constraints. The new criteria can cope with highly correlated parameters and are capable of reducing the confidence region. They are referred to as: PAC design, ACE design, E-AC design, and AC-V design.

The PAC design formulation only aims to decrease parameter correlations, whereas the other three (ACE, E-AC and AC-V designs) combine a correlation reduction with the traditional aim of standard experiment design (by improvement of the information matrix). For example, the E-AC design maximises the information content of the experiment in the objective function, while bounding all the correlation coefficients in constraints. Basically, these criteria improve the information matrix via traditional experiment design, but are constrained by a certain allowable parameter correlation bound. Mathematically written as,

$$\max_{\varphi \in \Phi} f(\mathbf{M}(\hat{\mathbf{p}}, \varphi)), \quad g = C_{ij}^2 |_{i \neq j} < \varepsilon_{ij}^C \quad (4.28)$$

where  $g$  represents the constraint,  $C_{ij}$  is the correlation coefficient between parameters  $i$  and  $j$ , and  $\varepsilon_{ij}^C$  represents the bound on this correlation coefficient. The ACE, E-AC and AC-V criteria aim at finding the best possible trade-off between correlation reduction and information content enhancement.

**PAC Design** The aim of this design criterion is to eliminate, or when not possible, reduce the correlation between the parameters without using additional experiment information. This criterion involves the correlation coefficients only and two different formulations can be defined.

**CRITERION PAC 1:**

The objective is to minimize the sum of the square of the upper triangular of the correlation coefficients.

$$\phi = \min_{\varphi \in \Phi} \sum_{i=1}^{M-1} \sum_{j=2}^M C_{ij}^2(\hat{\mathbf{p}}, \varphi) \quad i \neq j \text{ and } j > i \quad (4.29)$$

**CRITERION PAC 2:**

The objective is to minimize the largest correlation coefficient (selected based on the values of the correlation matrix  $\mathbf{C}$  at the base point, the starting point of the values), subject to upper bounds on the acceptable magnitude of each of the other correlation coefficients. A different bound  $\varepsilon_{kl}^C$  can be set for each of the correlation coefficients involved in the constraint  $g$ :

$$\phi = \min_{\varphi \in \Phi} C_{ij}^2(\hat{\mathbf{p}}, \varphi) \quad (4.30)$$

with  $i, j$  such that:

$$C_{ij} = \max \mathbf{C} \Big|_{\text{basepoint}} \quad i, j \in \{1, M\} \quad (4.31)$$

$$g = C_{kl}^2(\hat{\mathbf{p}}, \varphi) \Big|_{\substack{k \neq l \\ k, l \neq i, j}} < \varepsilon_{kl}^C \quad k, l \in \{1, M\} \quad (4.32)$$

The PAC criterion 2 can be more advantageous when the model contains a large number of parameters, since a sum over all the entries may be a difficult function to minimize in terms of obtaining a flat response or obtaining several minima.

**ACE Design** The PAC design aims at reducing the correlation between the model parameters. No experimental information is included which can result in large confidence regions. Reduction in correlation surely enhances the estimation of the parameters, but the optimal designed experiments are not expected to be necessarily informative.

The ACE design criterion combines the PAC design formulation and the enhancement of the information content. The main aim of the ACE design remains reducing the correlation between the model parameters, the PAC objective function with its constraint is used. As an addition, an extra constraint is introduced to ensure at least a minimum level for the information content of the experiments. Like the PAC design, two variants of the ACE design can be adopted:

1. Criterion 1: the information content is expressed by means of a single eigenvalue.
2. Criterion 2: the information content is expressed by means of two or more eigenvalues.

**CRITERION ACE 1:**

The first ACE criterion objective can be expressed as:

$$\phi = \min_{\varphi \in \Phi} C_{ij}^2(\hat{\mathbf{p}}, \varphi) \quad (4.33)$$

And containing:

$$C_{ij} = \max \mathbf{C} \Big|_{\text{basepoint}} \quad i, j \in \{1, M\} \quad (4.34)$$

$$g_{1,2} = \begin{cases} C_{kl}^2(\hat{\mathbf{p}}, \varphi) \Big|_{\substack{k \neq l \\ k, l \neq i, j}} < \varepsilon_{kl}^C & k, l \in \{1, M\} \\ \lambda_i - \lambda_i^0 > \varepsilon_i^\lambda & i \in \{1, M\} \end{cases} \quad (4.35)$$

A bound is introduced for one eigenvalue of the information matrix  $\mathbf{M}$ , where  $\lambda_i$  represents the new desired eigenvalue and  $\lambda_i^0$  its initial value. Depending on the desired accuracy, the bound  $\varepsilon_i^\lambda$  can be chosen. Unfortunately, there are no guidelines how to determine  $\varepsilon_i^\lambda$ , since it is model dependent and problem. As was mentioned previously, when the correlation between parameters is eliminated, the eigenvalues represent the information content of the experiment for their corresponding parameter. Hence, the information content for a specific parameter can be enhanced by choosing  $\varepsilon_i^\lambda$  positive. The lowest eigenvalue in the information matrix is a logical choice for  $\lambda_i$ . In this way, the design will enhance the information content of the most uncertain parameter.

#### CRITERION ACE 2:

The second ACE criterion objective is similar and can be expressed as:

$$\phi = \min_{\varphi \in \Phi} C_{ij}^2(\hat{\mathbf{p}}, \varphi)$$

But containing more constraints:

$$C_{ij} = \max \mathbf{C} \Big|_{\text{basepoint}} \quad i, j \in \{1, M\} \quad (4.36)$$

$$g_1 = C_{kl}^2(\hat{\mathbf{p}}, \varphi) \Big|_{\substack{k \neq l \\ k, l \neq i, j}} < \varepsilon_{kl}^C \quad k, l \in \{1, M\} \quad (4.37)$$

$$g_{2, \dots, n_c^\lambda} = \lambda_i - \lambda_i^0 > \varepsilon_i^\lambda \quad i \in \{1, M\} \quad (4.38)$$

Here,  $n_c^\lambda$  is the number of constraints for the chosen eigenvalues to bound. The objective of ACE 2 is to minimize the correlations and simultaneously increase the information content of the experiment for more than one parameter. It should be noted that the higher  $n_c$  is, the more difficult it is to find bounds that result in a feasible optimization problem. One way to overcome this issue is to loosen  $\varepsilon_i^\lambda$  per constraint (descending order) by choosing this bound slightly above zero or, in the end, zero itself. That is to say, the first constraint could be tight, the second constraint a little bit less, etcetera.

**E-AC Design** The E-AC design criterion is a modification of the ACE design criterion. The objective function aims to maximize one of the eigenvalues, and thus the information content of the experiment. The eigenvalue to maximize is free to choose and, therefore, the objective function is given by:

$$\phi = \max_{\varphi \in \Phi} \lambda_i(\hat{\mathbf{p}}, \varphi) \quad i \in \{1, M\} \quad (4.39)$$

A constraint is applied and deals with the reduction of correlation coefficients by keeping the values below a certain bound  $\varepsilon_{ij}^C$ .

$$g = C_{ij}^2(\hat{\mathbf{p}}, \varphi) \Big|_{i \neq j} < \varepsilon_{ij}^C \quad \begin{cases} i = 1, \dots, M-1 \\ j = 2, \dots, M \end{cases} \quad (4.40)$$

The less strict the constraints are, the larger the eigenvalue will be. An upper bound for  $\lambda_i$  to obtain is the value achieved while using the standard E-optimality criterion (optimization of eigenvalues, correlation constraint is neglected). A lower bound for  $\lambda_i$  is the value found with PAC design, which does not include information content. Therefore, one can state:

$$\lambda_i \Big|_{\text{PAC}} < \lambda_i < \lambda_i \Big|_{\text{E-optimal}} \quad i \in \{1, M\} \quad (4.41)$$

This is the range in which the solution is found.

**AC-V Design** The AC-V design criterion uses the same structure as the ACE design criterion. The largest correlation coefficients are partly involved in the objective function and are partly bounded by constraints. In the ACE design, extra constraint were added in order to enhance the information content of the experiment. In the AC-V design, extra constraints are introduced as well, however, these new constraints are not applied at the eigenvalues but at the variance of one or more parameters. The objective function is given by:

$$\phi = \min_{\varphi \in \Phi} C_{ij}^2(\hat{\mathbf{p}}, \varphi) \quad (4.42)$$

With  $i, j$  chosen such that:

$$C_{ij} = \max \mathbf{C} \Big|_{\text{basepoint}} \quad i, j \in \{1, M\} \quad (4.43)$$

$$g_1 = C_{kl}^2(\hat{\mathbf{p}}, \varphi) \Big|_{\substack{k \neq l \\ k, l \neq i, j}} < \varepsilon_{kl}^C \quad k, l \in \{1, M\} \quad (4.44)$$

$$g_{2, \dots, n_C^V} = V_{ii} < \varepsilon_i^V \quad i \in \{1, M\} \quad (4.45)$$

where  $V_{ii}$  is the variance of the  $i$ th parameter,  $n_C^V$  is the number of constraints for the variance, and  $\varepsilon_i^V$  is a specified upper bound for the variance in entry  $V_{ii}$ . By constructing the design criterion in this way, no association between parameters and eigenvalues is required. According to Franceschini and Macchietto (2008), this design is very suitable for cases where residual parameter correlations are expected due to model structure. Residual parameter correlation is expected for the feedforward model. Therefore, this criterion might be interesting to use for the experiment design optimization.

To identify a suitable value for the bound  $\varepsilon_i^V$ , the relationship between the standard deviation of a parameter and its  $t$  value can be adopted (Bard, 1974):

$$t_i = \frac{\Delta p_i}{\sqrt{V_{ii}}} \quad (4.46)$$

Where  $\Delta p_i$  is the difference in the parameter value and represents the desired accuracy, or range, of the parameter. The standard error of the parameter is represented by  $\sqrt{V_{ii}}$ . More detail about the t-distribution will be given in Chapter 5. A value for  $\varepsilon_i^V$  can easily be calculated after choosing the desired confidence level  $1 - \alpha$  and the desired parameter accuracy  $\Delta p_i$ .

$$\varepsilon_i^V = \left( \frac{\Delta p_i}{t \left( \frac{1+\alpha}{2}, N - M \right)} \right)^2 \quad (4.47)$$

Here  $\alpha$  is the right tail probability level (typically  $\alpha = 0.05$  which results in 95% confidence level) and  $N$  is the number of measurement data responses.

It should be noted that variations between objective functions and constraints for the above mentioned criteria are possible to obtain new design criteria.

In Table 4.2 all the mentioned optimality criteria are listed enriched with a description. This table allows for choosing a proper objective function for the right optimization problem .

**Table 4.2:** Optimality criteria with description which will help to make a proper selection of objective function for right problem.

CRITERION	DESCRIPTION
D	Standard criterion which targets at the volume of confidence region. Usually good performance for 2 model parameters. This criterion is often used for first try.
E	Standard criterion which targets at the smallest eigenvalue of the information matrix. Can be used for multiple model parameters, but affects only one parameter.
A	Standard criterion which targets at the trace of the information matrix. Considered unreliable by most authors, but can be used when correlation is not present
mod. E	Modification of E-optimality. Tries to obtain a spherical joint confidence region by optimizing the condition number of the information matrix. Often used, but only valid for 2 model parameters.
PAC 1	Criterion aims at reducing the largest correlation between a set of parameters. No further constraints.
PAC 2	Criterion aims at reducing all correlation entries. May be advantageous when number of model parameters is high. Local minima may appear, or flat responses.
ACE 1	Criterion aims at reducing the largest correlation between a set of parameters, but constraining one eigenvalue in order to keep an required information level.
ACE 2	Criterion aims at reducing the largest correlation between a set of parameters, but constraining multiple eigenvalue in order to keep an required information level.
E-AC	Criterion aims at improving selected eigenvalue, implying information content, while constraining the correlations between model parameters.
AC-V	Criterion aims at reducing the largest correlation between a set of parameters, but instead of constraining the eigenvalues the variance is constrained. Advantage: no association between parameters and eigenvalues needs to be established.



# 5

## Nonlinear Regression Analysis

In this chapter a brief review of linear regression and an introduction to nonlinear regression is given. These analyses are useful to investigate the rate of (non)linearity of the analytical model. The obtained information gives insight for model parameter estimation.

At first, a brief review of linear regression is given to provide basic knowledge. Subsequently, the linear case is extended to the non-linear case where new techniques, like profiling, are explained.

### 5.1 Linear Regression

A thorough grounding in linear regression is fundamental to understanding non-linear regression. Therefore, a brief review of linear regression is provided.

Linear regression provides estimates and other inferential results for the model parameters  $\mathbf{p} = [p_1, p_2, \dots, p_M]^T$  in the model:

$$\begin{aligned} Y_n &= p_1 Q_{n1} + p_2 Q_{n2} + \dots + p_M Q_{nM} + Z_n \\ &= (Q_{n1}, \dots, Q_{nM}) \mathbf{p} + Z_n \end{aligned} \quad (5.1)$$

In the above given model description, the random variable  $Y_n$  represents the response for  $n = 1, 2, \dots, N$  with  $N = n_{\text{resp}}$  and it consists of a *deterministic* part and a *stochastic* part. The deterministic part,  $(Q_{n1}, \dots, Q_{nM}) \mathbf{p}$ , depends upon the model parameters  $\mathbf{p}$  and the so-called regressor variables  $Q_{nm}$  with  $m = 1, 2, \dots, M$  with  $M$  the number of model parameters. The stochastic part, which is represented by  $Z_n$ , is a disturbance which perturbs the response for case  $n$ . In matrix notation, the model can be represented by:

$$\mathbf{Y} = \mathbf{Q}\mathbf{p} + \mathbf{Z} \quad (5.2)$$

Where  $\mathbf{Y}$  is the response vector,  $\mathbf{Z}$  the vector of random disturbances and  $\mathbf{Q}$  the sensitivity matrix defined in equation (4.2). The disturbance mean is assumed to be,

$$\text{E}[\mathbf{Z}] = 0 \quad (5.3)$$

and thus the expected mean of the model:

$$\text{E}[\mathbf{Y}] = \mathbf{Q}\mathbf{p} \quad (5.4)$$

Therefore, in Bates and Watts (1988)  $\mathbf{Q}\mathbf{p}$  is called the *expectation function*. Note that for linear models derivatives with any of the model parameters are independent of all the model parameters. Furthermore, it is assumed that  $\mathbf{Z}$  is normally distributed with

$$\text{Var}[\mathbf{Z}] = \sigma^2 \mathbf{I} \quad (5.5)$$

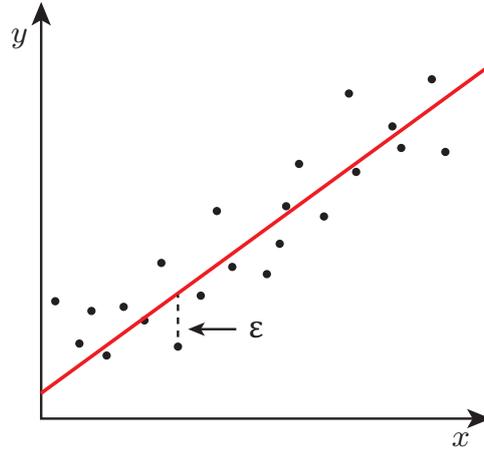
where  $\mathbf{I}$  is the identity matrix and  $\sigma^2$  the variance defined in Section 4.4.

### 5.1.1 Least Squares Estimates

The least square estimation technique allows one to estimate model parameters with experimental data. In order to do so, the *residual sum of squares* is minimized which is described by:

$$S(\mathbf{p}) = \|\mathbf{y} - \mathbf{Q}\mathbf{p}\|^2 \quad (5.6)$$

This equation calculates the summation of the error of all the experimental data with respect to the expectation plane. In Figure 5.1 experimental data of an arbitrary linear model is represented (dots) and the estimated linear model to describe the experimental data (red line) found by minimization of  $S(\mathbf{p})$ .



**Figure 5.1:** Least square regression for a linear model. Experimental data error  $\varepsilon$  is minimized with respect to expectation plane.

In terms of linear algebra, one can say that  $\mathbf{y}$  does not lie within the column space of  $\mathbf{Q}$ . The closest solution, or the minimum of the residual vector  $S(\mathbf{p})$ , is found by an orthogonal projection of  $\mathbf{y}$  onto the column space of  $\mathbf{Q}$ . In Figure 5.2 a geometrical interpretation for  $M = 2$  case is illustrated. The projection of  $\mathbf{y}$  onto the column space of  $\mathbf{Q}$  can be written as:

$$\mathbf{Q}\hat{\mathbf{p}} = \text{proj}(\mathbf{y}) \quad (5.7)$$

The residual vector is found when the projection is subtracted by  $\mathbf{y}$ .

$$\mathbf{Q}\hat{\mathbf{p}} - \mathbf{y} = \text{proj}(\mathbf{y}) - \mathbf{y} \quad (5.8)$$

From the figure, it can be seen that the residual vector is an orthogonal vector of the column space of  $\mathbf{Q}$ . However, this orthogonal vector is the same as the null space of  $\mathbf{Q}^T$ .

$$\mathbf{Q}\hat{\mathbf{p}} - \mathbf{y} \in C(\mathbf{Q})^\perp, \quad C(\mathbf{Q})^\perp = N(\mathbf{Q}^T) \quad (5.9)$$

Subsequently,

$$\mathbf{Q}\hat{\mathbf{p}} - \mathbf{y} \in N(\mathbf{Q}^T) \quad (5.10)$$

Since the residual vector is a member of the null space of  $\mathbf{Q}^T$  the inner product equals:

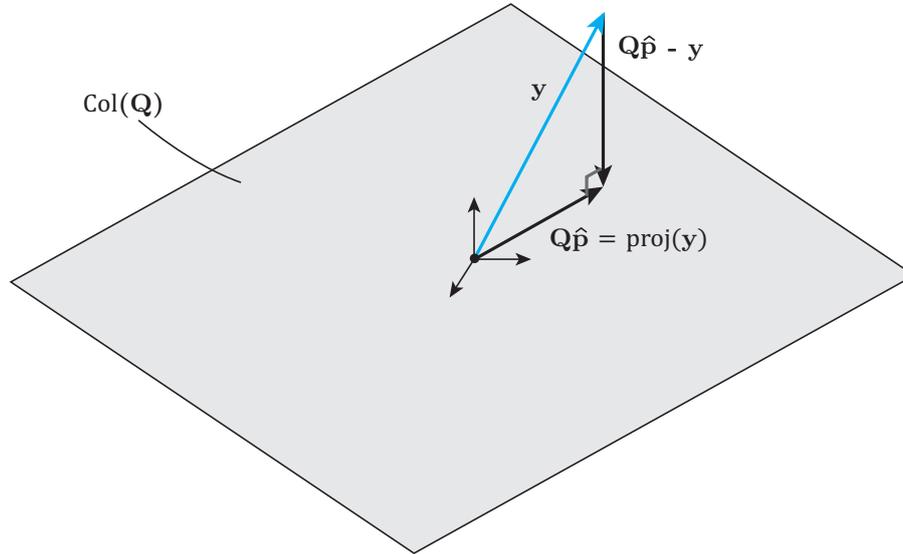
$$\mathbf{Q}^T (\mathbf{Q}\hat{\mathbf{p}} - \mathbf{y}) = 0 \quad (5.11)$$

And thus:

$$\mathbf{Q}^T \mathbf{Q}\hat{\mathbf{p}} = \mathbf{Q}^T \mathbf{y} \quad (5.12)$$

And finally the estimated model parameters  $\hat{\mathbf{p}}$ , which corresponds to the minimum of  $S(\mathbf{p})$ , are determined by:

$$\hat{\mathbf{p}} = (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{y} \quad (5.13)$$



**Figure 5.2:** Geometrical interpretation of the least squares method. Minimized residual vector  $S(\mathbf{p})$  is the orthogonal vector  $\mathbf{Q}\hat{\mathbf{p}} - \mathbf{y}$  onto the column space of  $\mathbf{Q}$ .

The least square estimator has a number of desirable properties as shown in Bates and Watts (1988):

1. The least squares estimator  $\hat{\mathbf{p}}$  is normally distributed. This follows because the estimator is a linear function of  $\mathbf{Y}$ , which in turn is a linear function of  $\mathbf{Z}$ . Since  $\mathbf{Z}$  is assumed to be normally distributed,  $\hat{\mathbf{p}}$  is normally distributed.
2.  $E[\hat{\mathbf{p}}] = \mathbf{p}$ : the least squares estimator is unbiased.
3.  $\text{Var}[\hat{\mathbf{p}}] = \sigma^2 (\mathbf{Q}^T \mathbf{Q})^{-1}$ : the covariance matrix of the least squares estimator depends on the variance of the disturbances and on the derivative matrix  $\mathbf{Q}$ . The variance of  $\hat{\mathbf{p}}$  can be coupled to the variance-covariance matrix found in Section 4.5.1.

Furthermore, the  $1 - \alpha$  ellipsoid joint confidence region, as seen in Section 4.5.1, can be constructed according

$$(\mathbf{p} - \hat{\mathbf{p}})^T \mathbf{Q}^T \mathbf{Q} (\mathbf{p} - \hat{\mathbf{p}}) \leq M s^2 F(M, N - M; \alpha) \quad (5.14)$$

where

$$s^2 = \frac{S(\hat{\mathbf{p}})}{N - M} \quad (5.15)$$

is the *variance estimate* based on  $N - M$  degrees of freedom and  $F(M, N - M; \alpha)$  the upper  $\alpha$  quantile for Fisher's F distribution with  $M$  and  $N - M$  degrees of freedom. The marginal confidence region for a single model parameter  $\mathbf{p}$  can be found according the t-value and is described by:

$$\hat{p}_m \pm \text{se}(\hat{p}_m) t(N - M, \alpha/2) \quad (5.16)$$

Here, the  $t$ -value  $t(N - M, \alpha/2)$  is the upper  $\alpha/2$  quantile for Student's T distribution with  $N - M$  degrees of freedom. The standard error  $se$  of the parameter estimator is given by:

$$se(p_m) = s\sqrt{\{\mathbf{Q}^T\mathbf{Q}\}_{mm}^{-1}} \quad (5.17)$$

Or equivalent to,

$$se(p_m) = \sqrt{\mathbf{V}_{mm}} \quad (5.18)$$

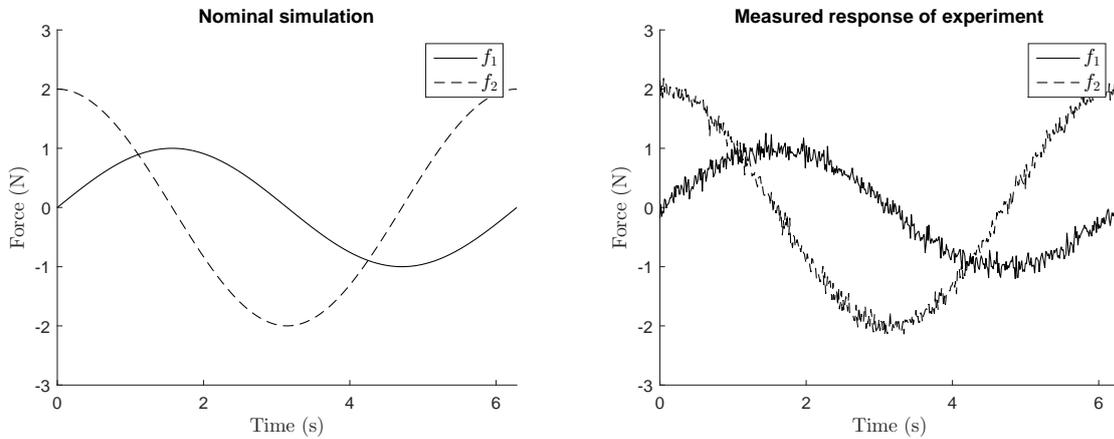
where  $\mathbf{V}_{mm}$  is the  $m$ th diagonal entry of the variance-covariance matrix given in equation (4.18). The  $t$ -test is used to establish the statistical significance of the model parameter estimates.

### EXAMPLE 5.1

In this example, linear regression techniques are applied at the system used in Example 4.1. This model is linear in its model parameters when the system is constraint in the deformation  $\mathbf{u}$  and the force vector  $\mathbf{f}$  is measured as the response.

$$\mathbf{K}\mathbf{u} = \mathbf{f}, \quad \text{with} \quad \mathbf{K} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \sin(\omega t) \\ \cos(\omega t - \gamma) \end{bmatrix}$$

The model parameters are set to  $A = 1$  and  $B = 2$ ,  $\omega = 1$  rad,  $t$  is the time vector from  $t_1$  up to  $t_{\text{end}}$  and for now the experiment design parameter  $\gamma = 0$ . A nominal simulation is performed and subsequent, noise is added to represent an experiment with measurement noise. Both responses are illustrated in Figure 5.3.



**Figure 5.3:** Nominal simulation of the system on the left. On the right, response of the system with noise, which acts as a measured experiment.

The measured response of the experiment is constructed according,

$$\mathbf{f}_\eta = \mathbf{f} + \boldsymbol{\eta} \quad (5.19)$$

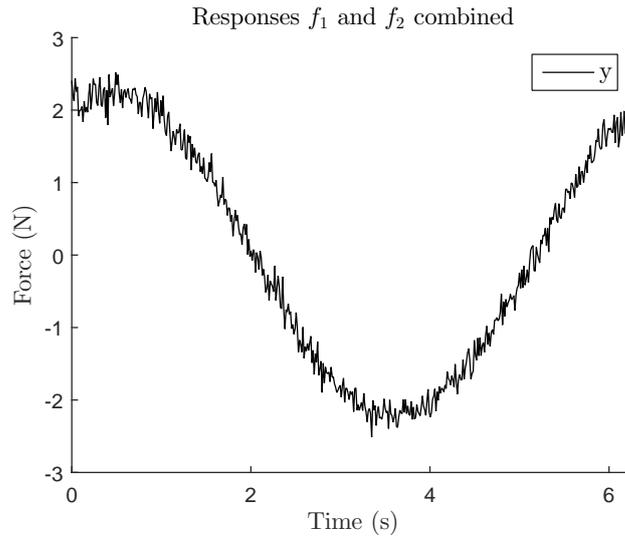
where  $\boldsymbol{\eta}$  represents normally distributed noise with mean zero and:

$$\text{var}[\boldsymbol{\eta}] = \sigma^2\mathbf{I} \quad (5.20)$$

The variance is assumed to be  $\sigma^2 = 0.01 \text{ N}^2$ . Since the model has two responses in time, the response  $\mathbf{y}$  is a matrix. In order to use the information of both responses during least square, the responses must be summed:

$$\mathbf{y} = \sum_{j=1}^2 \mathbf{f}_{\eta,j} \quad (5.21)$$

The measured responses used for linear least square estimation is illustrated in Figure 5.4. From  $\mathbf{y}$  the model parameter  $\mathbf{p}$  could be estimated. Since the model is linear in the model parameters, the least square estimator is described by equation (5.13).



**Figure 5.4:** Individual responses  $f_1$  and  $f_2$  are summed to use the information of both during least square.

Note that only one response is present, the time-varying sensitivity matrix is reduced to  $r = 1$  and is given by:

$$\mathbf{Q} = \begin{bmatrix} \left. \frac{\partial \mathbf{y}}{\partial p_1} \right|_{t_1} & \left. \frac{\partial \mathbf{y}}{\partial p_2} \right|_{t_1} \\ \vdots & \vdots \\ \left. \frac{\partial \mathbf{y}}{\partial p_1} \right|_{t_{\text{end}}} & \left. \frac{\partial \mathbf{y}}{\partial p_2} \right|_{t_{\text{end}}} \end{bmatrix} = \begin{bmatrix} \sin(\omega t_1) & \cos(\omega t_1) \\ \vdots & \vdots \\ \sin(\omega t_{\text{end}}) & \cos(\omega t_{\text{end}}) \end{bmatrix} \quad (5.22)$$

The least square estimates are obtained using  $\mathbf{Q}$  and  $\mathbf{y}$ . The result is listed in the table below.

**Table 5.1:** Least square estimates for design parameter  $\gamma = 0$ .

$\mathbf{p}$	TRUE VALUE	$\hat{\mathbf{p}}$ VALUE	ERROR (%)
$A$	1	1.0019	0.19
$B$	2	1.9960	0.20

From Table 5.1 it can be seen that the true values are not found, but no further statistical information of the estimates is available as well. Normally, one would perform multiple experiments and construct statistical data. This is not necessary and statistical information about the model parameters can be obtained using the

regression techniques described in Section 5.1.1. The Student t-test can be used to determine the marginal confidence bounds of the estimates  $\hat{\mathbf{p}}$ . The variance is known, thus the variance estimate  $s$  can be substituted by  $\sigma$ . Therefore, the standard error  $\mathbf{se}$  is given by:

$$\mathbf{se} = \sqrt{(2\sigma^2)} \sqrt{\{\mathbf{Q}^T \mathbf{Q}\}^{-1}} = 0.14 \sqrt{\begin{bmatrix} 250 & 0 \\ 0 & 250 \end{bmatrix}^{-1}} \quad (5.23)$$

The variance  $\sigma^2$  is multiplied by a factor 2 due to the summation in  $\mathbf{y}$ . Using equation (5.16) the 95% marginal confidence bounds for the estimators are:

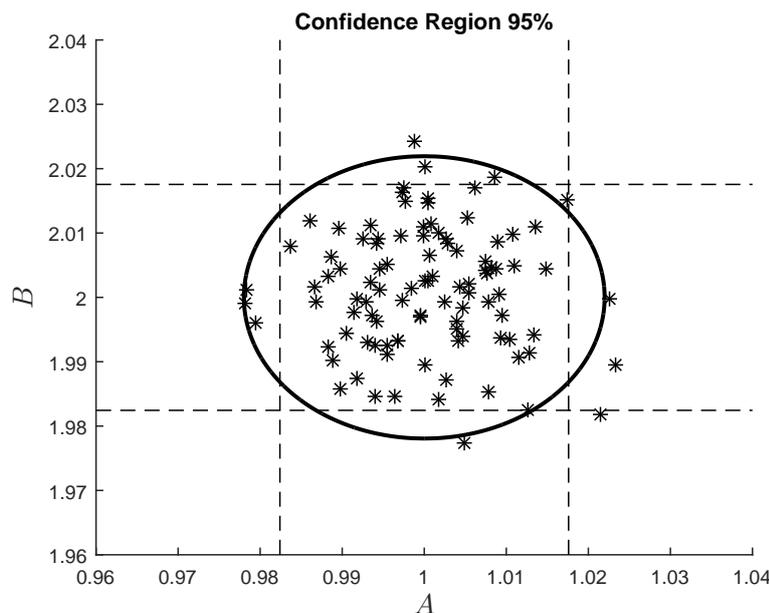
$$\begin{aligned} \hat{A} &= 1.0019 \pm 0.0089 \cdot 1.9647 = 1.0019 \pm 0.0175 \\ \hat{B} &= 1.9960 \pm 0.0089 \cdot 1.9647 = 1.9960 \pm 0.0175 \end{aligned} \quad (5.24)$$

Since the parameters are uncorrelated, the Student t-test is reliable. However, the parameters  $A$  and  $B$  are multivariate distributed and the joint confidence region represents the true confidence bounce. These bounds can be calculated according equation (5.14) and are given by:

$$\begin{aligned} \hat{A} &= 1.0019 \pm 0.022 \\ \hat{B} &= 1.9960 \pm 0.022 \end{aligned} \quad (5.25)$$

The confidence bounds are slightly higher, as is expected. It can be stated that the true parameters  $\mathbf{p}$  lie within these bounds with a confidence of 95%.

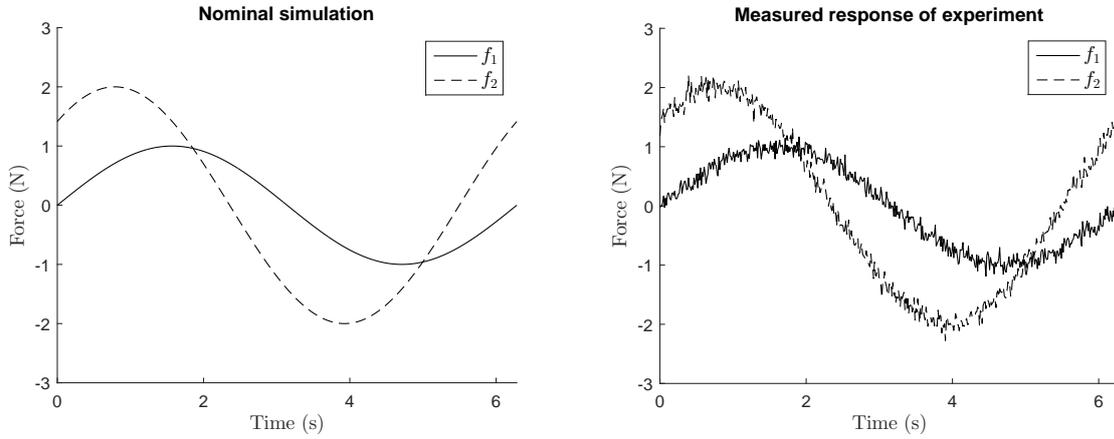
Multiple experiments can be performed with the associated least square estimation. This allows one to graphically verify the estimates and their confidence. Each experiment is "measured" with a different set of noise. The 95% confidence region of the true parameters is constructed for one hundred realization, which implies that once per 20 estimates an estimate will fall outside the ellipsoid. The result can be seen Figure 5.5.



**Figure 5.5:** One hundred least square estimates for  $\gamma = 0$ . The dotted lines represents the marginal confidence bounds.

Some of the estimates fall outside of the confidence region, but this is expected due to the chosen level of confidence. When increasing the confidence level, the ellipsoid will become larger and more estimates will fall within. The dotted lines represents the marginal confidence interval of the parameters. When the system is uncorrelated, these marginal bounds can be used to check whether the parameters fall within the confidence region.

Next, the experiment design parameter is set to  $\gamma = 1/4\pi$ . A correlation of  $A$  and  $B$  is introduced and it is expected that this correlation will influence the accuracy of the estimates. As before, a nominal simulation is performed. Subsequent, multiple noise responses are made and added to the nominal simulation. This represents multiple experiments, which are subjected to least squares estimation.



**Figure 5.6:** Nominal simulation of the system on the left. On the right, response of the system with noise, which acts as a measured experiment.

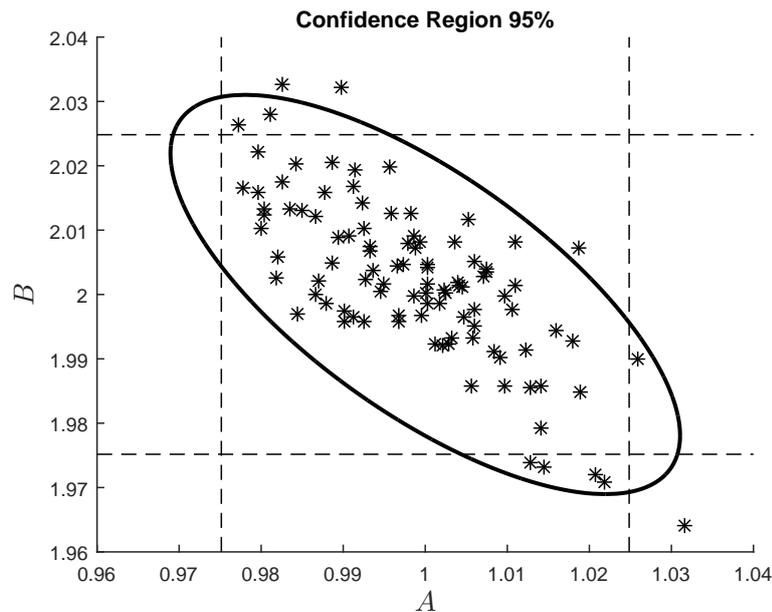
The time-varying sensitivity matrix is not the same as before, since the experiment design parameter has changed. The new sensitivity matrix – used for least square estimation – is described by:

$$\mathbf{Q} = \begin{bmatrix} \sin(\omega t_1) & \cos(\omega t_1 - 1/4\pi) \\ \vdots & \vdots \\ \sin(\omega t_{\text{end}}) & \cos(\omega t_{\text{end}} - 1/4\pi) \end{bmatrix} \quad (5.26)$$

The result of the obtained estimates  $\hat{\mathbf{p}}$ , together with the confidence region of the true parameters  $\mathbf{p}$ , is illustrated in Figure 5.7. It can be seen that most of the estimates lie within the confidence region, as was seen in the previous case. Due to the correlation, the confidence bounds has increased and accuracy of the model parameters are reduced. The value of the bounds are:

$$\begin{aligned} \hat{A} &= 1.0 \pm 0.0249 \\ \hat{B} &= 2.0 \pm 0.0249 \end{aligned} \quad (5.27)$$

Also, due to correlation the t-values cannot be used to check whether the estimates  $\hat{\mathbf{p}}$  lie within the confidence region. The lower-left and upper-right corner of the square that is made by the marginal confidence bounds, represents regions wherein the estimates pass the t-test. However, when looking at the true joint confidence ellipsoid, these regions cannot be identified as valid.



**Figure 5.7:** One hundred least square estimates for  $\gamma = 1/4\pi$ . The dotted lines represents the marginal confidence bounds.

### 5.1.2 Lack of Fit

The *Lack of Fit* (LOF) test is used to check whether or not the model response explains the observed data in a satisfactory way. The LOF can be determined using the Pearson's chi-square test and is calculated according:

$$\chi^2 = \frac{S(\hat{\mathbf{p}})}{\sigma^2} \quad (5.28)$$

The Pearson's chi-square test is a statistical test to evaluate if the difference between observed categorical data can be explained by chance. The  $\chi^2$  value is compared with a reference value calculated from the chi-square distribution. When the  $\chi^2$  value is higher than the reference value, there is a significant lack of fit at level  $\alpha$ . Otherwise, when the  $\chi^2$  value is lower than the reference value, the difference between the observed data can be assigned due to chance.

It should be noted that the chi-square may lead to erroneous results when the sample size is small. Type II errors may occur, which is the failure of rejecting a false hypothesis. When the sample sizes are small, the Fisher's exact test must be used instead.

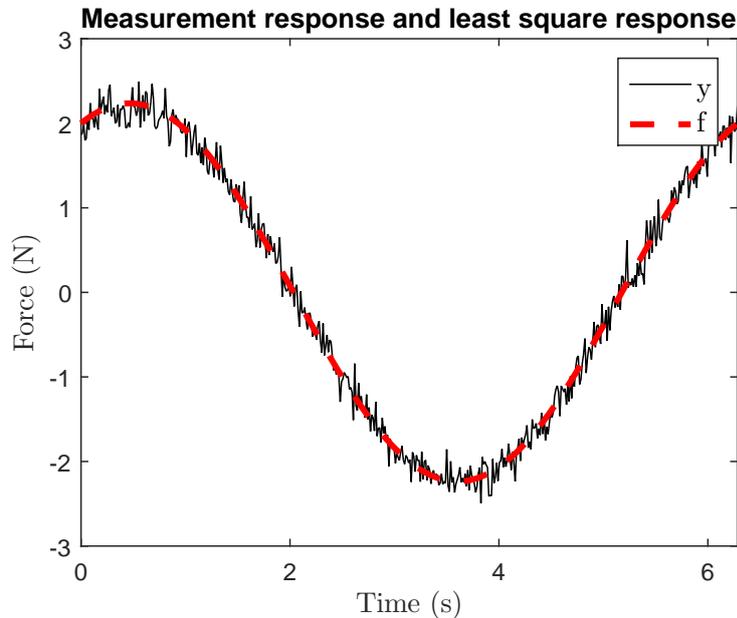
## EXAMPLE 5.2

In this example two responses of the model described in Example 5.1 are subjected to the Pearson's chi-square test.

At first, a nominal experiment is performed and extended with normally distributed mean zero noise in order to represent an experimental measurement. This measurement is constructed as is described in equation (5.21). Subsequent, the model parameters are estimated by means of the least squares method. The found estimates are:

$$\hat{\mathbf{p}} = \begin{bmatrix} 1.01 \\ 1.97 \end{bmatrix} \quad (5.29)$$

A simulation with  $\hat{\mathbf{p}}$  as the model parameters is performed and the response  $\mathbf{f}(\hat{\mathbf{p}})$  is obtained. In order to compare the measurement and the simulated response,  $\mathbf{f}(\hat{\mathbf{p}})$  is summed in the same manner. In Figure 5.8 the measurement response and the simulated response  $\mathbf{f}(\hat{\mathbf{p}})$  are illustrated.



**Figure 5.8:** Measurement and simulated response  $\mathbf{f}(\hat{\mathbf{p}})$  are illustrated. Simulated responses overlaps measurement in satisfactory way. 492.7235

From the figure, it seems that the simulated response overlaps the measurement in a satisfactory way. Using the Pearson's chi-square test this can be verified. The LOF value is calculated according equation (5.28) and its value is:

$$\chi^2 = 492.7 \quad (5.30)$$

The reference  $\chi_{\text{ref}}^2$  value for a  $1 - \alpha$  confident can be found in associated chi-square distribution tables. The  $\chi_{\text{ref}}^2$  value for a system of  $N - M$  degrees of freedom, with  $N$  the number of responses (time samples), and a confidence of 95% is:

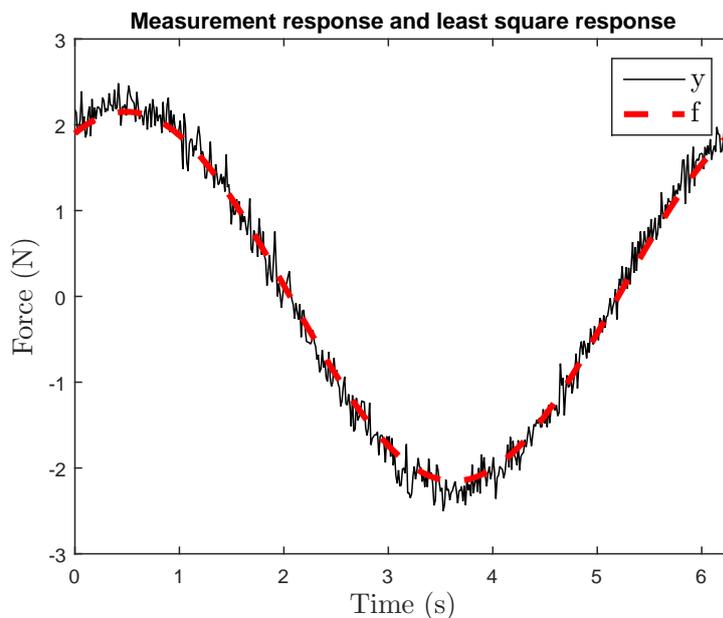
$$\chi_{\text{ref}}^2 = 551.0 \quad (5.31)$$

It can be seen that  $\chi^2 < \chi_{\text{ref}}^2$ , which implies that the differences in responses can totally be assigned by the measured noise.

Next, a simulation is performed with found estimates which are significantly deviated from the nominal parameters. The estimates are given by:

$$\hat{\mathbf{p}} = \begin{bmatrix} 1.1 \\ 1.85 \end{bmatrix} \quad (5.32)$$

The simulated response  $\mathbf{f}(\hat{\mathbf{p}})$  is obtained and, as before, the measured and simulated response are illustrated in Figure 5.9.



**Figure 5.9:** One hundred least square estimates for  $\gamma = 1/4\pi$ . 590.6923

It can be seen that the simulated response shows similar behaviour as the measurement, but it does not overlap as satisfactory as before. The corresponding  $\chi^2$  value is:

$$\chi^2 = 796.1 \quad (5.33)$$

This value is larger than the reference  $\chi_{\text{ref}}^2$  value, which implies that the simulated response falls in the right probability tail of the chi-square distribution. There is only 5% chance, or less, that the difference between the measured and simulated response can be allocated to noise. Hence, the found estimates are not reliable and the least square estimation has failed.

## 5.2 Nonlinear Regression

Linear regression is a powerful method to analyse data described by models which are linear in the model parameters. Unfortunately, most models do not behave linear in their parameters and thus linear regression techniques must be extended.

A nonlinear regression model can be written in the form,

$$Y_n = f(\mathbf{Q}_n, \mathbf{p}) + Z_n \quad (5.34)$$

where  $f$  is the expectation function. The nonlinear model is of exactly the same form as equation (5.1), except that the responses of the expectation function are nonlinear related in the model parameters. That is, according Bates and Watts (1988), at least one of the derivatives of the expectation function with respect to the model parameters depends on at least one of the model parameters. Since the expected responses of  $f$  depends on the model parameters  $\mathbf{p}$ , the following can be written:

$$\zeta_n(\mathbf{p}) = f(\mathbf{Q}_n, \mathbf{p}) \quad n = 1, 2, \dots, N \quad (5.35)$$

The nonlinear regression model becomes:

$$\mathbf{Y} = \zeta(\mathbf{p}) + \mathbf{Z} \quad (5.36)$$

With the same assumption of  $\mathbf{Z}$  as in the linear case. That is, the disturbance is normally distributed with:

$$E[\mathbf{Z}] = 0 \quad (5.37)$$

$$\text{Var}[\mathbf{Z}] = \sigma^2 \mathbf{I} \quad (5.38)$$

### 5.2.1 Least Squares Estimates

For the nonlinear model, estimation of the model parameters is quite different. There exists no solution of the form seen in equation (5.13). The  $n_{\text{resp}}$ -vectors  $\zeta(\mathbf{p})$  span a  $M$ -dimensional surface called the *expectation surface* in the response space. The least squares estimates correspond to the point on the expectation surface,

$$\hat{\zeta} = \zeta(\hat{\mathbf{p}}) \quad (5.39)$$

which is closest to response data  $\mathbf{y}$ . Therefore,  $\hat{\mathbf{p}}$  is found by minimizing the sum of residuals defined by,

$$S(\mathbf{p}) = \|\mathbf{y} - \zeta(\mathbf{p})\|^2 \quad (5.40)$$

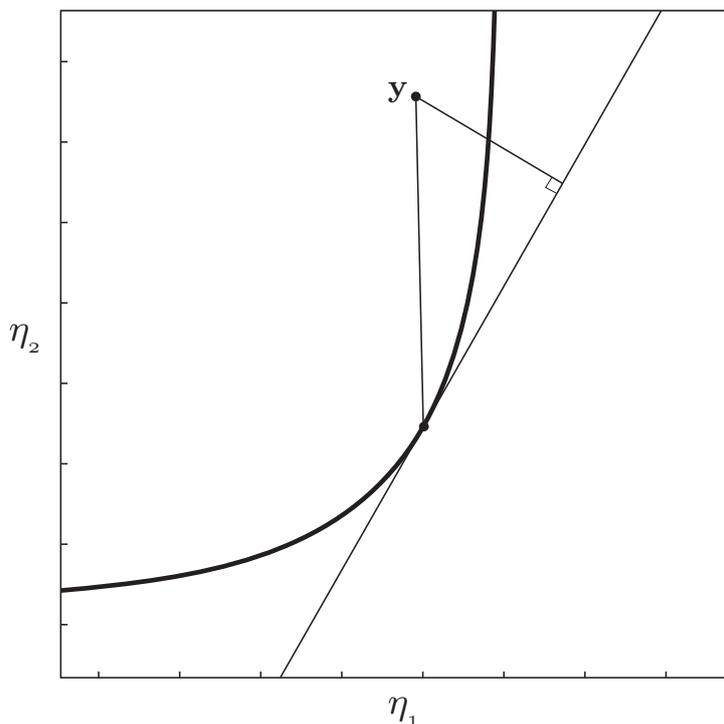
In contrast to the linear case, there is no analytical solution for the minimization of equation (5.40). However, an approach suggested by Gauss is to use a linear approximation of the expectation surface and iteratively improve the least squares estimate until there is no change. This method is modification of the Newton method and is known as the Gauss-Newton method. It is used in least squares problems only, unlike the Newton method.

Different types of algorithms do exists and might be more advantageous depending on the type of problem. The Levenberg-Marquardt algorithm is a popular algorithm when stability issues arise. In short, this algorithm begins with the Steepest-Descent search method (due to stability) and gradually converts to the Gauss-Newton method.

In Figure 5.10 a geometrical interpretation of the Gauss-Newton method is illustrated. The heavy solid line is a slice of the expectation surface  $\zeta$ , which is non-linear. The linear approximation of the expectation surface  $\mathbf{W}$  is indicated by the straight line which is the tangent of the expectation surface at initial point  $\hat{\mathbf{p}}_0$ . The linearised expectation surface is calculated according:

$$\mathbf{W} = \frac{\partial f(\mathbf{Q}, \mathbf{p})}{\partial \mathbf{p}} \quad (5.41)$$

Next, the response  $\mathbf{y}$  is projected onto the tangent plane and, hence, new approximated parameters  $\hat{\mathbf{p}}_1$  are obtained. These parameters are used to create a new approximation at  $\zeta(\hat{\mathbf{p}}_1)$  and the process is repeated until convergences is obtained.



**Figure 5.10:** A geometrical interpretation of the Gauss-Newton iteration increment calculation. The heavy solid line is a slice of the expectation surface  $\zeta$ , which is non-linear. The linear approximation of the expectation surface is indicated by the straight line which is the tangent of the expectation surface at initial point  $\hat{\mathbf{p}}_0$ . Next, the response  $\mathbf{y}$  is projected onto the tangent plane and, hence, new approximated parameters  $\hat{\mathbf{p}}_1$  are obtained. These parameters are used to create a new approximation at  $\zeta(\hat{\mathbf{p}}_1)$  and the process is repeated until convergence is obtained.

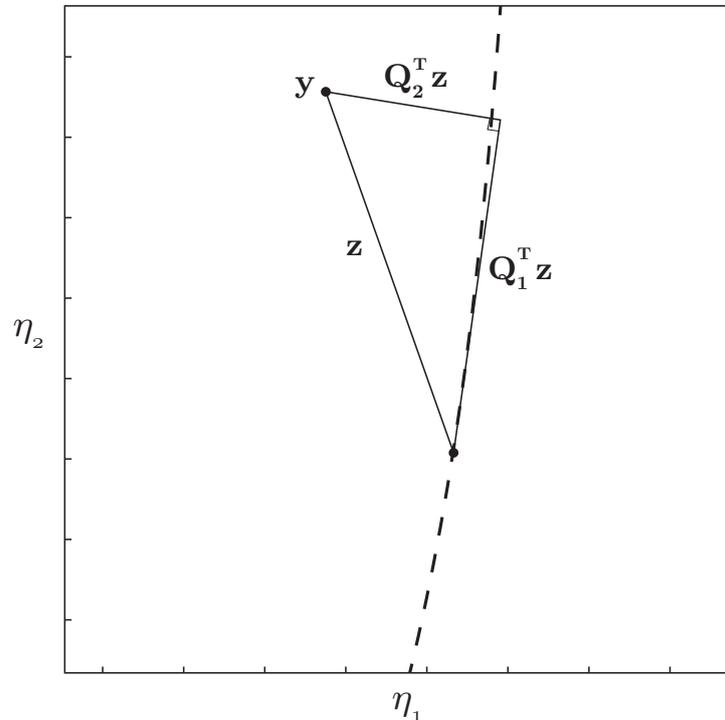
**Convergence** The Gauss-Newton iterative method can be used to solve a non-linear least squares problem. Several measures exist to determine the stability of the iterations or, they tell when convergence has reached. One measure could be the size of each parameter increment relative to the previous parameter value. Another convergence criterion used is the relative change in the sum of squares and convergence is achieved when this change is small. However, as discussed in Bates and Watts (1981), compliance even with both relative change criteria does not guarantee convergence.

The main criticism of these criteria is that they indicate lack of progress rather than convergence. Bates and Watts (1988) suggest a new convergence criterion based on the geometry of the non-linear least squares problem. From previous, it is known that a critical point is reached when the residual vector is orthogonal to the expectation surface and thus the tangent plane  $\mathbf{W}$ . This orthogonality of the residual vector to the tangent plane can be used as a convergence criterion. A tolerance level needs to be established since perfect orthogonality may not be reached due to numerical errors. One way to do this is to consider the statistical variability in the least squares estimates, but unfortunately, this involves the unknown least squares vector  $\hat{\mathbf{p}}$ , see Bates and Watts (1988). A so-called *relative offset convergence criterion* is eventually made which is described by,

$$\frac{\|\mathbf{Q}_1^T(\mathbf{y} - \zeta(\mathbf{p}_i))\|/\sqrt{M}}{\|\mathbf{Q}_2^T(\mathbf{y} - \zeta(\mathbf{p}_i))\|/\sqrt{N-M}} \quad (5.42)$$

where  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  are the first  $M$  and  $N-M$  columns of the *QR-decomposition* of  $\mathbf{W}$  respectively. Note,  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  must not be confused with the sensitivity matrix  $\mathbf{Q}$ . The index  $i$  indicates the current estimate of the model parameter  $\hat{\mathbf{p}}$ . Equation 5.42 measure the scaled length of the

tangent plane component of the residual vector relative to the scaled orthogonal component of the residual vector. Therefore, this criterion is related to angle of the residual vector that it makes with the tangent plane. Small relative offset of convergence corresponds to an angle near  $90^\circ$ . According Bates and Watts (1988), to declare converges a relative offset equal or less than 0.001 is required. See Figure 5.11 for a geometrical interpretation.



**Figure 5.11:** Geometrical interpretation of orthogonal convergence criterion. The dotted line represents a slice of the expectation surface and  $\mathbf{z} = \mathbf{y} - \zeta(\mathbf{p}_i)$  represents the residual vector at estimate  $\mathbf{p}_i$ . QR-decomposition is used to determine the project of  $\mathbf{z}$  onto the tangent plane  $\mathbf{W}$  and to determine the orthogonal part of  $\mathbf{z}$ .

### EXAMPLE 5.3

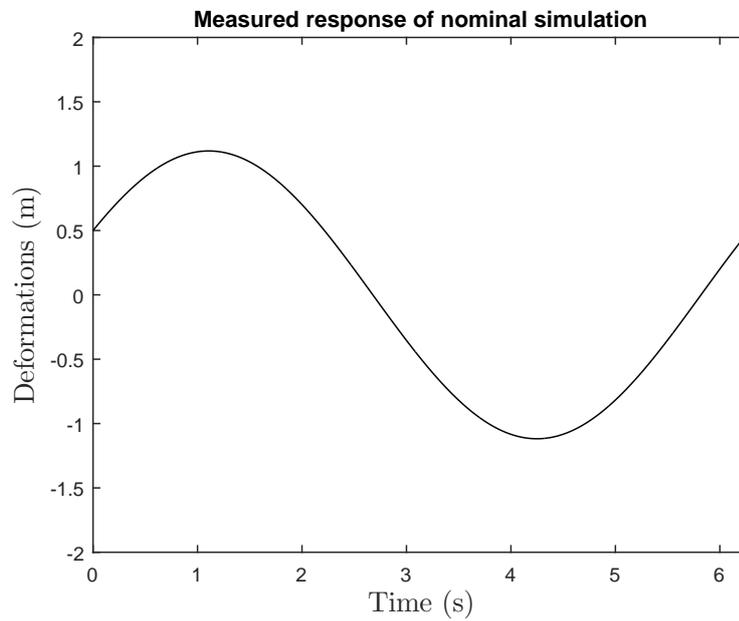
In this example a non-linear least square estimation is performed. The academic model explained in Example 3.1 will be used. For the non-linear least squares, the response must behave non-linear in the model parameters. This is the case for the deformation vector  $\mathbf{u}$ . The model is written as:

$$\mathbf{u} = \mathbf{K}^{-1}\mathbf{f} \quad (5.43)$$

It can be seen that the response of  $\mathbf{u}$  behaves non-linear in the model parameters due to the inverse of  $\mathbf{K}$ . The model represents a mechanical system with stiffness  $A$  and  $B$  which is subjected to a known force function over time. Hence, deformations  $\mathbf{u}$  will arise. As before,  $\omega$  and  $t$  are a constant and the time vector respectively and  $\gamma$  is the experiment design parameter.

A nominal experiment is performed with  $\mathbf{p} = [1 \ 2]^T$  (model parameters  $A$  and  $B$  respectively) and the experiment design parameter  $\gamma = 0$ . The deformation response

vector  $\mathbf{u}$  is obtained and, like the linear least squares case, the deformation responses are summed in order to obtain one response in time. This response is illustrated in Figure 5.12.



**Figure 5.12:** Nominal simulation for deformation response vector  $\mathbf{u}$ .

In order to simulate an experiment measurement, noise is added according,

$$\mathbf{u}_\eta = \mathbf{u} + \boldsymbol{\eta} \quad (5.44)$$

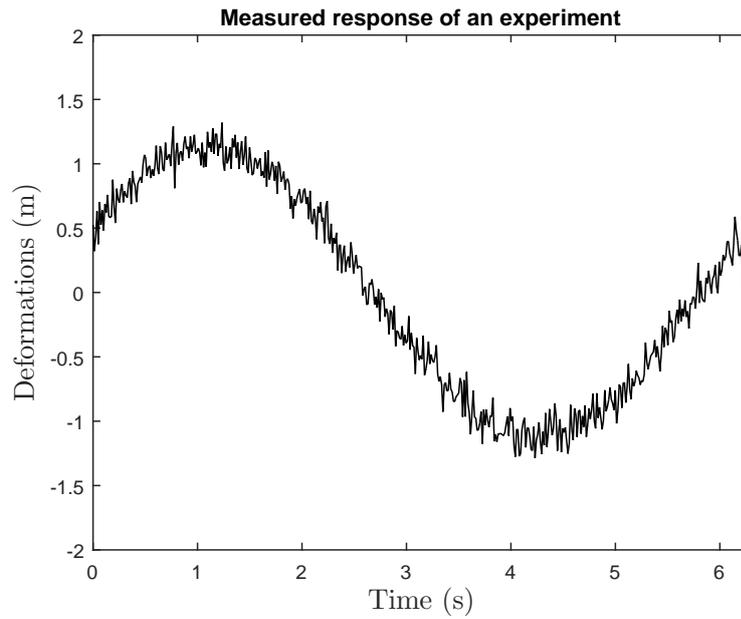
where  $\boldsymbol{\eta}$  represents normally distributed noise with mean zero and:

$$\text{var}[\boldsymbol{\eta}] = \sigma^2 \quad (5.45)$$

The variance is assumed to be  $\sigma^2 = 0.01 \text{ m}^2$ . The nominal simulation with the addition of noise is used as a measured experiment subjected to the non-linear least square estimation, see Figure 5.13. The least square estimation is performed with the Gauss-Newton iterative method, which minimizes the residual vector  $S(\mathbf{p})$ .

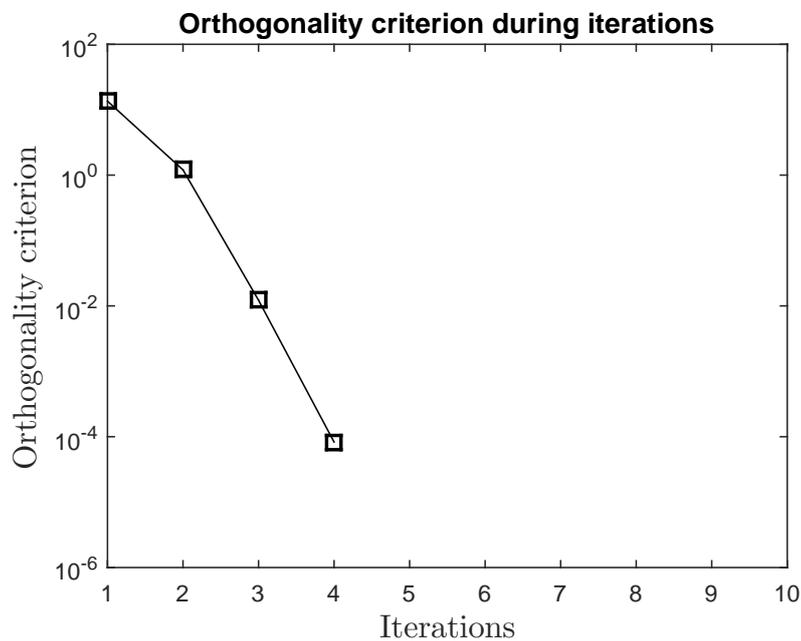
$$S(\mathbf{p}) = \|\mathbf{u}_\eta - \mathbf{u}(\mathbf{p})\|^2 \quad (5.46)$$

The orthogonality criterion of the residual vector is used to determine when converges has been reached.



**Figure 5.13:** Nominal simulation with the addition of noise in order to represent an experimental measurement  $\mathbf{u}_\eta$ .

In Figure 5.14 it can be seen that convergence has been reached after four iterations. When the algorithm stops, the residual vector is (approximately) orthogonal to the expectation plane, indicating the least squares solution which is listed in Table 5.2.



**Figure 5.14:** Least squares estimation for a non-linear model. Orthogonality criterion reduces per iteration, indicating convergence.

From the table it can be seen that parameter  $B$  has a 10 times larger error. Since the model is non-linear, confidence ellipsoids cannot be constructed. However, the Pearson's chi-square test can be used to check whether the response of the estimates

represents the measured experimental response. The chi-square value is:

$$\chi^2 = 244.6 \quad (5.47)$$

This value is far below the reference value given in Example 5.2. Therefore, the found estimates can be marked as reliable.

**Table 5.2:** Least square estimates for design parameter  $\gamma = 0$ .

$\mathbf{p}$	TRUE VALUE	$\hat{\mathbf{p}}$ VALUE	ERROR (%)
<i>A</i>	1	0.9979	0.21
<i>B</i>	2	2.0600	2.90

When the same procedure is repeated, but with design parameter  $\gamma = 3/8\pi$ , it is expected that due to correlation the estimates error is worse. When observing Table 5.3, it can be seen that this is indeed the case.

**Table 5.3:** Least square estimates for design parameter  $\gamma = 3/8\pi$ .

$\mathbf{p}$	TRUE VALUE	$\hat{\mathbf{p}}$ VALUE	ERROR (%)
<i>A</i>	1	0.9824	1.79
<i>B</i>	2	2.0851	4.08

## 5.2.2 Profiling

Joint and marginal confidence regions as seen before are not directly applicable to non-linear models. One could linearise the model around  $\hat{\mathbf{p}}$  and obtain the tangent plane  $\mathbf{W}$ . The joint confidence region would be described by,

$$(\mathbf{p} - \hat{\mathbf{p}})^T \mathbf{W}^T \mathbf{W} (\mathbf{p} - \hat{\mathbf{p}}) \leq Ms^2 F(M, N - M; \alpha) \quad (5.48)$$

and the marginal confidence region by equation (5.16) with,

$$se(p_m) = s \sqrt{\{\mathbf{W}^T \mathbf{W}\}_{mm}} \quad (5.49)$$

Statistical information is obtained, however, it is not known if this data is reliable due to non-linearities. In order to investigate the non-linearity of the model profile techniques can be applied.

### 5.2.2.1 t-profile

The t-profiling technique is useful to determine the marginal confidence interval for a non-linear model parameter  $p_m$ . It must be noticed that all the important information about the parameters is embodied in the sum of squares function given in equation (5.40). The marginal confidence can be represented by its sum of squares according:

$$S_t = S(\hat{\mathbf{p}}) \left[ 1 + \frac{t^2(N - M; \alpha/2)}{N - M} \right] \quad (5.50)$$

Here  $S_t$  represent the value of the sum of squares for the required marginal confidence interval. When the system is linear in its model parameters, equation (5.50) can be written as equation (5.16), since the standard error remains constant and the mapping of the points on the expectation surface back to the parameters space is easy.

In order to construct the non-linear marginal confidence interval, the surface of the sum of squares must be profiled at first which is described in Ruckstuhl (2010); Asprey and Naka (1998); Watts (2010). In the following steps, the calculations that are required to obtain the desired information is given.

1. Begin with a least squares estimation of the model for all parameters. The estimates  $\hat{\mathbf{p}}$  are found.
2. Select the profile parameter  $p_m$ , which is the parameter that is investigated for its confidence interval.
3. Specify increment  $\Delta$  for  $p_m$ . This increment can be determined according,

$$\delta(p_m) = \frac{p_m - \hat{p}_m}{\text{se}(\hat{p}_m)} \quad (5.51)$$

which is equivalent to the t-value. The standard error is calculated from linearisation at the point of interest.

4. Initialize  $p_m = \hat{p}_m$  and  $\tilde{\mathbf{p}}(p_m) = \hat{\mathbf{p}}$ . Where  $\tilde{\mathbf{p}}(p_m)$  is the profile trace vector and contains all parameters in  $\hat{\mathbf{p}}$  except the profile parameter  $p_m$ . Therefore, the profile trace vector is dependent on the profile parameter.
5. Increment  $p_m = p_m + \Delta$ .
6. Perform least squares estimation, e.g. Gauss-Newton, to converge and obtain  $\tilde{\mathbf{p}}(p_m)$  and  $\tilde{S}(p_m)$ . It should be noted that during the least squares estimation  $p_m$  is fixed and constant, and the remaining parameters in  $\tilde{\mathbf{p}}(p_m)$  are free.
7. Store  $p_m$ ,  $\tilde{\mathbf{p}}(p_m)$  and  $\tilde{S}(p_m)$ , the profile sum of squares.
8. Repeat step 3 to 7 as necessary in order to make sure that the increments are reaching the t-value that is required to calculate the desired  $-\alpha$  confidence.
9. When finished, repeat the process for a different profile parameter  $p_m$  until all parameters are profiled.

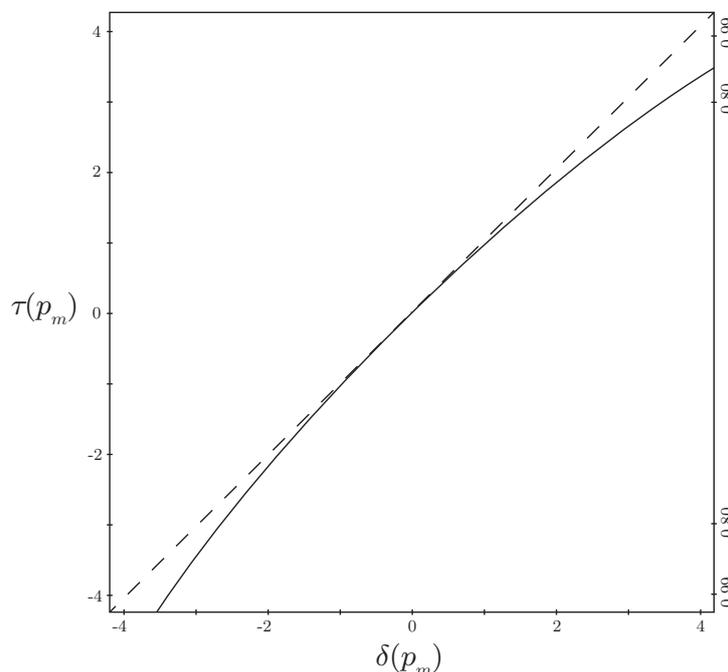
Note, although several least squares problems must be solved, this is not computational intensive due to good initial estimates.

The next step, in order to obtain the non-linear marginal confidence intervals, is to convert the profile sum of squares in terms of likelihood regions or confidence regions. First for convenience, the profile parameter values  $p_m$  needs to be studentized according equation (5.51). Subsequently, convert the profile sum of squares  $\tilde{S}(p_m)$  to t-values according,

$$\tau(p_m) = \text{sign}(p_m - \hat{p}_m) \frac{\sqrt{\tilde{S}(p_m) - S(\hat{p}_m)}}{s} \quad (5.52)$$

which is derived from equation (5.50). Here,  $\tau(p_m)$  is called the *profile t function* with  $\tilde{S}(p_m) = S_t$ . When the profile parameters and profile sum of squares are converted into t-values, the non-linear marginal confidence can be plotted, see Figure 5.15. When the system is linear, the t-profile is a straight line with slope 1. This reference line is illustrated as the dashed line. When

the t-profile deviates from the reference line, non-linearity of the associated model parameter is revealed. However, if the deviation is small and the t-profile is nearly linear, one can safely assume that the situation can be approximated by the tangent plane  $\mathbf{W}$ .



**Figure 5.15:** t-profile curve for one model parameter. The dashed line has slope 1 and indicates full linearity. The t-profile line is curved, which reveals the non-linearity of model parameter  $p_m$ . The t-values  $\tau(p_m)$  can be used to determine the marginal confidence intervals and, subsequently, the corresponding parameter value can be determined by  $\delta(p_m)$ .

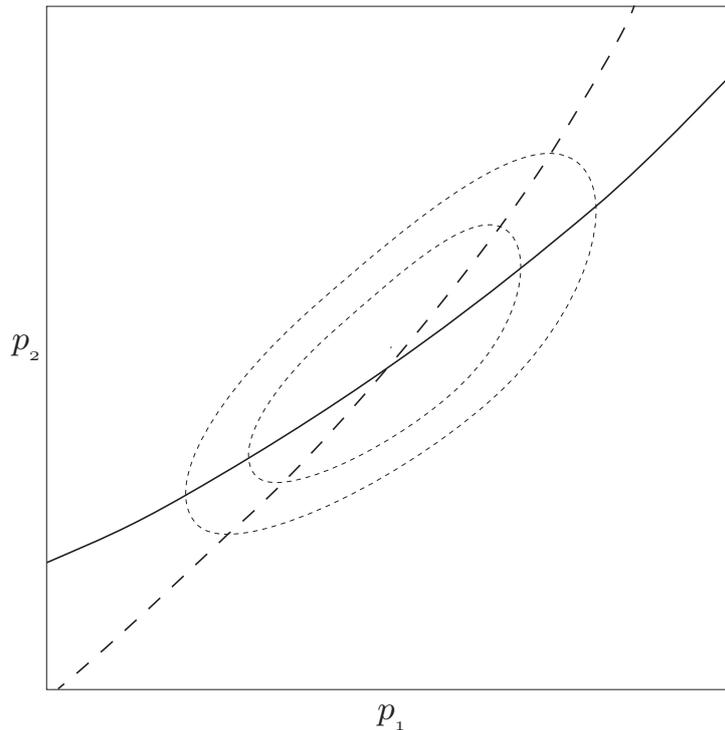
The marginal confidence region can be determined using the t-value  $\tau(p_m)$  on the vertical axis. The parameter value associated to the confidence level can be found through the profile curve onto the horizontal axis.

### 5.2.2.2 Profile Traces

Another useful plot is the *profile trace* obtained during t-profiling. For example, after evaluating the t-profile likelihood for  $p_1$ ,  $\tilde{\mathbf{p}}_m$  can be plotted versus  $p_1$  (e.g.,  $\tilde{p}_2$  versus  $p_1$ ,  $\tilde{p}_3$  versus  $p_1$ , ... etc.). This procedure can be continued for all profile parameters with the associated profile trace parameters. Plots of the profile traces provide useful information on how the parameters interact. For a linear model, the profile traces on a plot  $p_m$  versus  $p_n$  consists of straight lines. The angle of intersection depends upon the correlation between the parameters. When the parameters are uncorrelated the profile traces intersect at an angle of  $90^\circ$ . Perturbations of one parameter does not influence the value of the other parameter. When full correlation exists, the profile traces coincide and are not distinguishable.

When the model is non-linear, the profile traces will be curved. This curving provides information on how the parameter estimates affect one another and provides information on the shape of the confidence regions. If the non-linear confidence contours will be long and thin, the profile traces will be close together. When the contours are nearly elliptical, the profile traces will be straight (linear case). In Figure 5.16 the profile traces for  $p_1$  and  $p_2$  are illustrated. The solid line is obtained by plotting the profile parameter  $p_1$  versus the profile trace parameter  $\tilde{p}_2$  and the dashed line is obtained by plotting  $p_2$  versus  $\tilde{p}_1$ . The curving profile traces reveals the non-linearity between the parameter estimates. The intersection of the profile traces is not

perpendicular, hence correlation between the parameters is present. For convenience, two levels of (non-linear) confidence regions are plotted.



**Figure 5.16:** Profile trace plot for non-linear system. The curving lines reveal non-linearity between parameter estimates. The angle of intersection indicates correlation between the parameters. Two level of non-linear confidence regions are plotted.

### 5.2.2.3 Profile Pair Sketches

For non-linear problems, confidence regions are hard to obtain since there is no analytical solution. However, using the profile sum of squares and the profile traces, very accurate approximations to the 2-dimensional confidence regions can be created. This gives a visual indication of the region and the non-linear dependence of parameter estimates upon each other. To determine, for instance, a 95% confidence contour, the intersection of the trace of  $\tilde{p}_2$  on  $p_1$  needs to be found. This intersection for  $p_1$  occurs at the point of desired confidence found from the previous t-profiling (see Figure 5.15). The intersections will give two points on the contour. In addition, at these points it is known that the tangent to the contour must be vertical, since they represent the confidence bounds in  $p_1$  direction. This is represented in Figure 5.16. Similarly, two more points can be determined from the trace of  $\tilde{p}_1$  on  $p_2$  and it is known that the tangent to the contour of these points must be horizontal.

By using the available information of the t-profiling, the points on the contour, the direction of their tangents and the fact that the contour is bounded by the parameter values at these points, very accurate interpolations of the confidence contour for non-linear models can be created (such as Figure 5.16). In Bates and Watts (1988), these interpolations are called *profile pair sketches*. The method of the interpolation is described in appendix C.

---

**EXAMPLE 5.4**


---

In this example, the academic model is subjected to the profiling technique. The non-linear form used in Example 3.1 is examined, since profiling techniques are applicable to non-linear models only (where non-linearity occurs in the model parameters).

T-profiling of the parameters  $\mathbf{p}$  is performed. The necessary steps are listed in Section 5.2.2.1 and the first step is performing a least-squares estimation from an experiment measurement. One would use the profiling techniques in order to obtain more information about the least-squares estimates. A nominal simulation with  $\mathbf{p} = [1 \ 2]^T$  is performed with experiment design parameter  $\gamma = 0$ . Normally distributed noise is added to represent a real experimental measurement with associated measurement noise. The variance  $\sigma^2 = 0.01$  m, as was seen in the previous example. A non-linear least square estimation is performed which resulted in:

$$\hat{\mathbf{p}} = \begin{bmatrix} 1.0046 \\ 2.0351 \end{bmatrix} \quad (5.53)$$

Both parameters are subjected to the profiling and in order to determine the increment  $\Delta$ , the standard error is required. The system is linearised around  $\hat{\mathbf{p}}$  and the tangent plane  $\mathbf{W}$  is obtained, see equation (5.41). From this linearised expectation plane, the standard error can be determined.

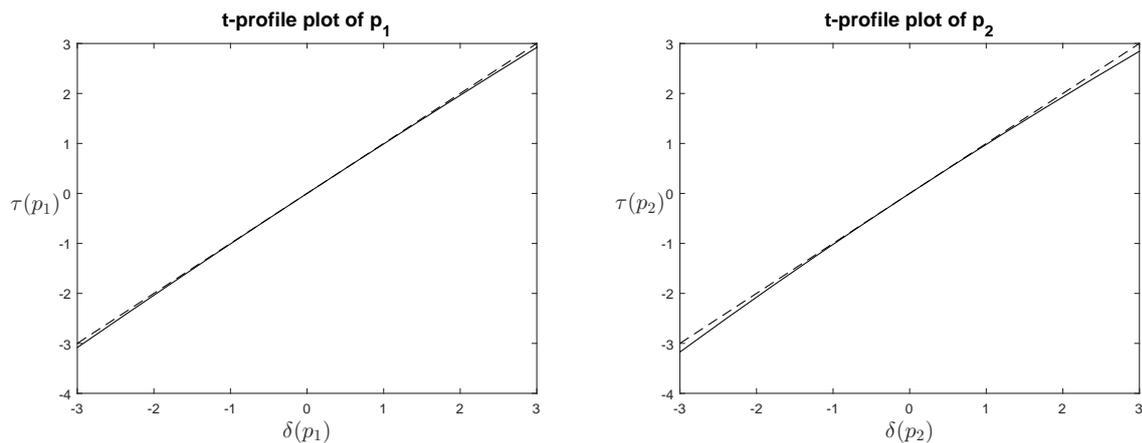
$$\begin{aligned} se(p_1) &= \sigma \sqrt{\{\mathbf{W}^T \mathbf{W}\}_{11}^{-1}} = 0.009 \\ se(p_2) &= \sigma \sqrt{\{\mathbf{W}^T \mathbf{W}\}_{22}^{-1}} = 0.037 \end{aligned} \quad (5.54)$$

The increments are chosen between a t-value of  $[-3 \ 3]$ , this will make sure that a confidence of 99% is included. When using equation (5.51) the profiling region per parameter can be determined and, subsequent a suitable increment can be chosen. The bounds are calculated and given by:

$$\begin{aligned} 0.9775 &\leq p_1 \leq 1.0317 \\ 1.9241 &\leq p_2 \leq 2.1461 \end{aligned} \quad (5.55)$$

First, parameter  $p_1$  is subjected to profiling, which makes  $p_2$  the trace parameter stored in the trace vector  $\tilde{\mathbf{p}}(p_1)$ . Note, for convenience the vector notation is used for the profile trace vector, but in this example the profile trace vector is actually a scalar value. The profile parameter is fixed at  $p_1 = \hat{p}_1$  and a least squares estimation is performed on the trace parameter  $p_2$ . When the least squares algorithm has converged, the profile trace vector  $\tilde{\mathbf{p}}(p_1)$  and the residual  $\tilde{\mathbf{S}}(p_1)$  vector are stored. Next, the profile parameter  $p_1$  is incremented with  $\Delta$  and a new least square estimation is performed. This process is repeated until the last increment.

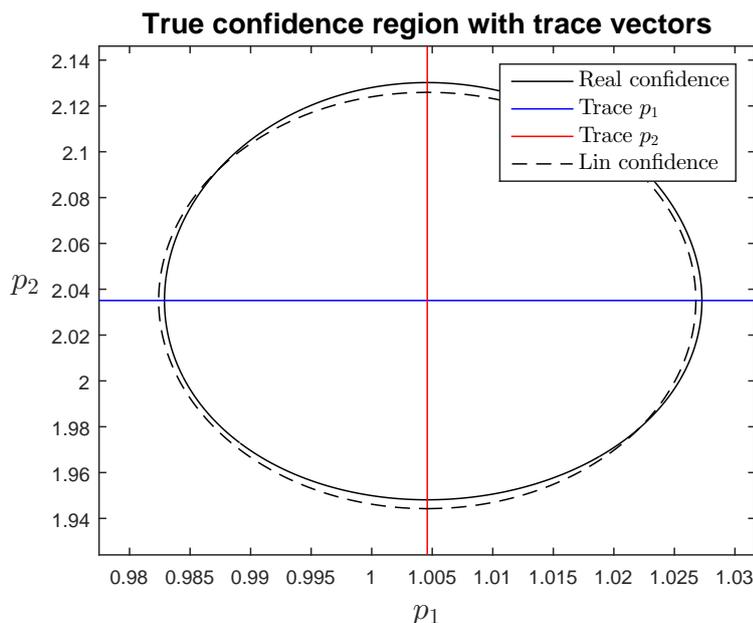
The same procedure is executed for profile parameter  $p_2$  and trace vector  $\tilde{\mathbf{p}}(p_2)$ . When all the profile information is obtained, the true t-values can be calculated according equation (5.52). All the required information is available in order to plot the t-profiling plots, which are illustrated in Figure 5.17.



**Figure 5.17:** t-profiling for  $\mathbf{p}$  with  $\gamma = 0$ . The dotted line represent the response for a linear system, the solid line represent the real response.

In the figure, the dotted line with slope 1 represents the response for a linear system and the solid line represents the t-value response for the system investigated. At the point  $[0, 0]$  the lines coincide and the associated model parameter values are the least squares solution. Any deviation unveils non-linear behaviour of the model parameters. From the left figure, it can be seen that the solid line is almost linear and approximated by the dashed line. For this experiment, the model parameter  $A$  behaves approximately linear. The same can be concluded for  $B$  (right figure), although the deviation is slightly larger. This comes from the fact that this parameter is more uncertain when observing the standard error in equation (5.54).

With the profiling information, the profile trace vectors and the true confidence regions (profile pair sketches) can be obtained. The true confidence region, the profile trace vectors and the confidence region made by the linear approximation are illustrated in Figure 5.18.



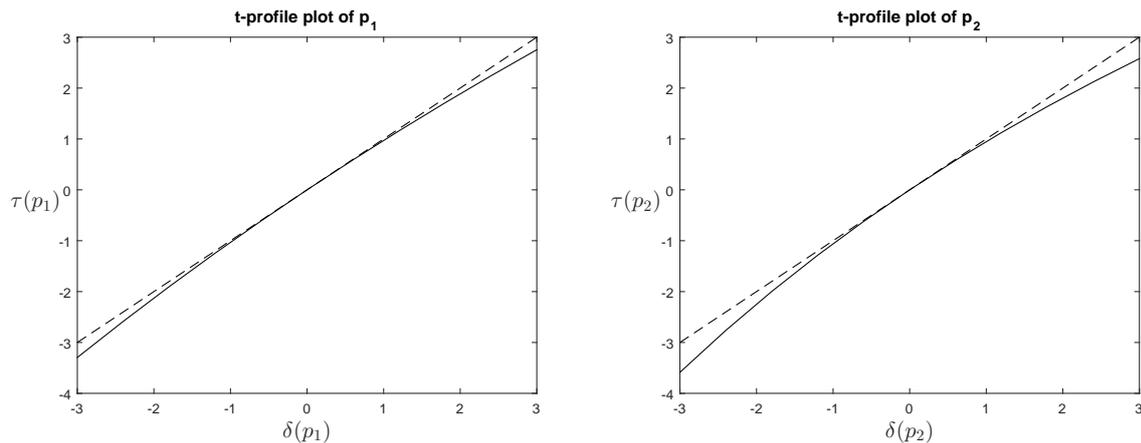
**Figure 5.18:** True confidence region with profile trace vectors for  $\gamma = 0$ . The confidence region of the linear approximation represent the true confidence region quite accurate. The trace plots intersect perpendicular, indicating no correlation between the model parameters.

The design parameter  $\gamma = 0$ , and as a consequence the model parameters are un-correlated as was seen in previous examples. This un-correlation can be seen in the figure by the perpendicular intersection of the profile trace vectors, but it is also evident from the fact that, for instance for the trace plot  $p_1$ , a perturbation of  $p_1$  does not result in a variation of  $p_2$ . The trace plot remains horizontal for the whole domain. As a consequence of the un-correlated relation, the shape of the confidence ellipsoid is circular. The confidence region of the linear approximation is very similar to non-linear confidence region. This is expected, since the t-profiling plot showed little deviation from the linear reference line. The approximated confidence region is least accurate in the direction of  $p_2$ , which again can be explained by the information obtained from the t-profiling plot. It can be concluded that a linear approximation of the least squares solution can be used to obtain statistical information about the estimates  $\hat{\mathbf{p}}$ .

The same procedure is performed once more, but the experiment design parameter is set to  $\gamma = 2/5\pi$ . It is known that this will introduce significant correlation between the model parameters. A least squares estimation is performed on the new measured experimental response, constructed in the same manner as before. The least squares estimates are:

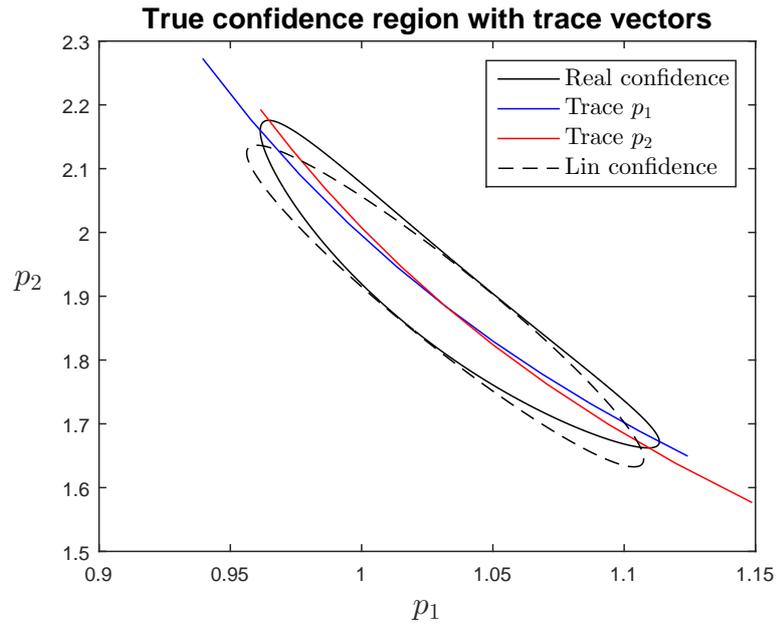
$$\hat{\mathbf{p}} = \begin{bmatrix} 1.0318 \\ 1.8847 \end{bmatrix} \quad (5.56)$$

The t-profiling technique is performed and the t-profile plots are illustrated in Figure 5.19.



**Figure 5.19:** t-profiling for  $\mathbf{p}$  with  $\gamma = 2/5\pi$ . The dotted line represent the response for a linear system, the solid line represent the real response.

In contrast to Figure 5.17, the solid lines deviate more from the dashed line. This indicates that the system behaves more non-linear in the model parameters when the experiment is performed with  $\gamma = 2/5\pi$ . The true confidence region together with the trace vector can be constructed and are illustrated in Figure 5.20. The profile trace plots are almost coinciding, indicating severe correlation between the model parameters. This is also evident when looking at the individual trace plots. For the trace plot of  $p_1$ , a perturbation of  $p_1$  results in a variation of  $p_2$  when subjected to least squares estimation. This is not the case when  $\gamma = 0$ . The t-profiling plots showed considerable non-linear behaviour of the model parameters, which affects the ellipsoid shape of the true confidence region. It follows the direction of the trace plots and becomes banana shaped. From the figure, it is obvious that the linearised confidence region cannot be used as an approximation of the true confidence region. Therefore, every applied linear regression technique – on this experiment – cannot be considered reliable.



**Figure 5.20:** True confidence region with profile trace vectors for  $\gamma = 2/5\pi$ . The confidence region of the linear approximation represent the true confidence region not accurate. The trace plots almost coincide, indicating severe correlation between the model parameters.

---

# 6

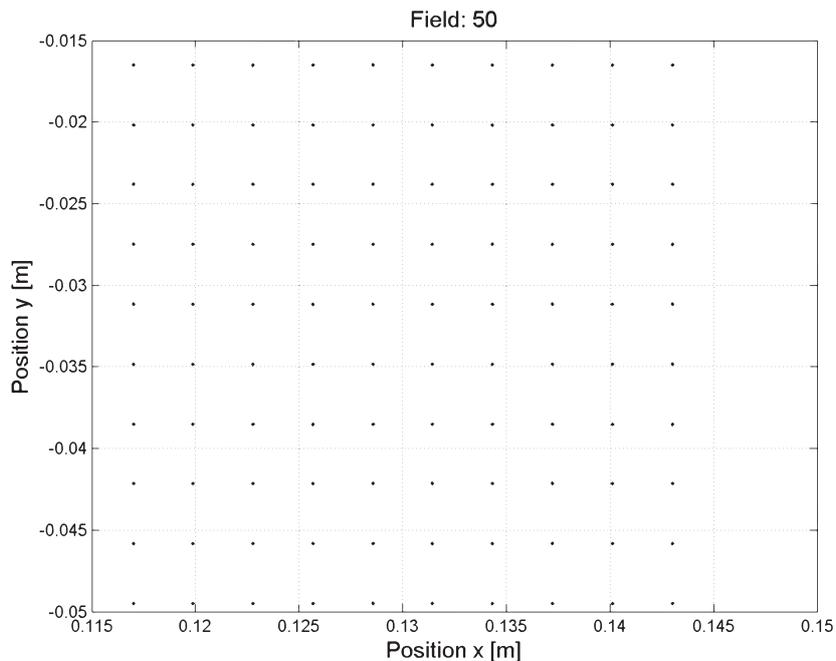
## Preliminary Calibration Investigation

A preliminary investigation of the calibration of the model parameters on the thermo-mechanical model of ASML was performed in van der Meulen (2015). In this investigation, several analyses of the thermal model parameters were done, e.g., sensitivity analysis and correlation analysis. This is necessary to select a well chosen subset of thermal parameters in order to perform the calibration successfully.

Next, a gradient-based optimization algorithm (mentioned in Section 1.1, Figure 1.5) was used in order to estimate the real model parameters.

### 6.1 Sensitivity Analysis

In this section, the sensitivity of overlay for each thermal model parameter when it is subjected to a certain variation is investigated. Furthermore, the distinguishability of the thermal model parameter from the overlay measurement is analysed.



**Figure 6.1:** Deformation grid in  $xy$ -plane for field 50.

The thermo-mechanical FF model incorporates 18 thermal parameters. On the basis of a recipe,

a load case is constructed. The load case that is used in this preliminary calibration investigation consists of 87 fields that are exposed on one substrate. Each field contains a deformation grid, see Section 2.5. In Figure 6.1 the deformation grid in the  $xy$ -plane, for field 50, is illustrated.

The response on terms of overlay can be expressed by the following vector,

$$\mathbf{y} = [\Delta x_1 \quad \dots \quad \Delta x_g \quad \dots \quad \Delta x_G \quad \Delta y_1 \quad \dots \quad \Delta y_g \quad \dots \quad \Delta y_G]^T \quad (6.1)$$

in which  $g = 1, \dots, G$  and  $G$  the number of grid points for all the fields. It is difficult to use a real overlay measurement that is obtained from a machine. Therefore, in order to investigate the sensitivity of the thermal parameters, a reference simulation is performed with the true (nominal) parameter values  $\mathbf{p}$ . For the parameter values  $\hat{\mathbf{p}}$  – the thermal parameters to be calibrated – the response  $\mathbf{y}_{\text{model}}(\hat{\mathbf{p}})$  is obtained. Now, the residual function is defined by:

$$\mathbf{r}(\hat{\mathbf{p}}) = \mathbf{y}_{\text{model}}(\hat{\mathbf{p}}) - \mathbf{y}_{\text{ref}}(\mathbf{p}) \quad (6.2)$$

The residual function is a measure for the error between the model and the simulated overlay measurement in general. For analysis and optimization purposes, it is convenient to express the residual function (column) as a scalar. This is realized by the objective function  $\phi$ . The objective function simplifies the interpretation in terms of (worst-case) absolute substrate deformation and is defined by:

$$\phi(\hat{\mathbf{p}}) = \rho(\mathbf{V}\mathbf{r}(\hat{\mathbf{p}})) \quad (6.3)$$

Here,  $\rho$  is known as Huber's function. It shall be a symmetric, positive-definite function with a unique minimum at zero. Huber's function is defined by:

$$\rho(\mathbf{x}) = \max(|\mathbf{x}|) \quad (6.4)$$

The matrix  $\mathbf{V}$  is a weighting matrix. This matrix scales the residual vector  $\mathbf{r}(\hat{\mathbf{p}})$  to be independent and identically distributed with a variance that equals one. Here, the weighting matrix  $\mathbf{V}$  can be interpreted as the conversion from the unit m to the unit nm and is defined by:

$$\mathbf{V} = 1 \cdot 10^9 \mathbf{I} \quad (6.5)$$

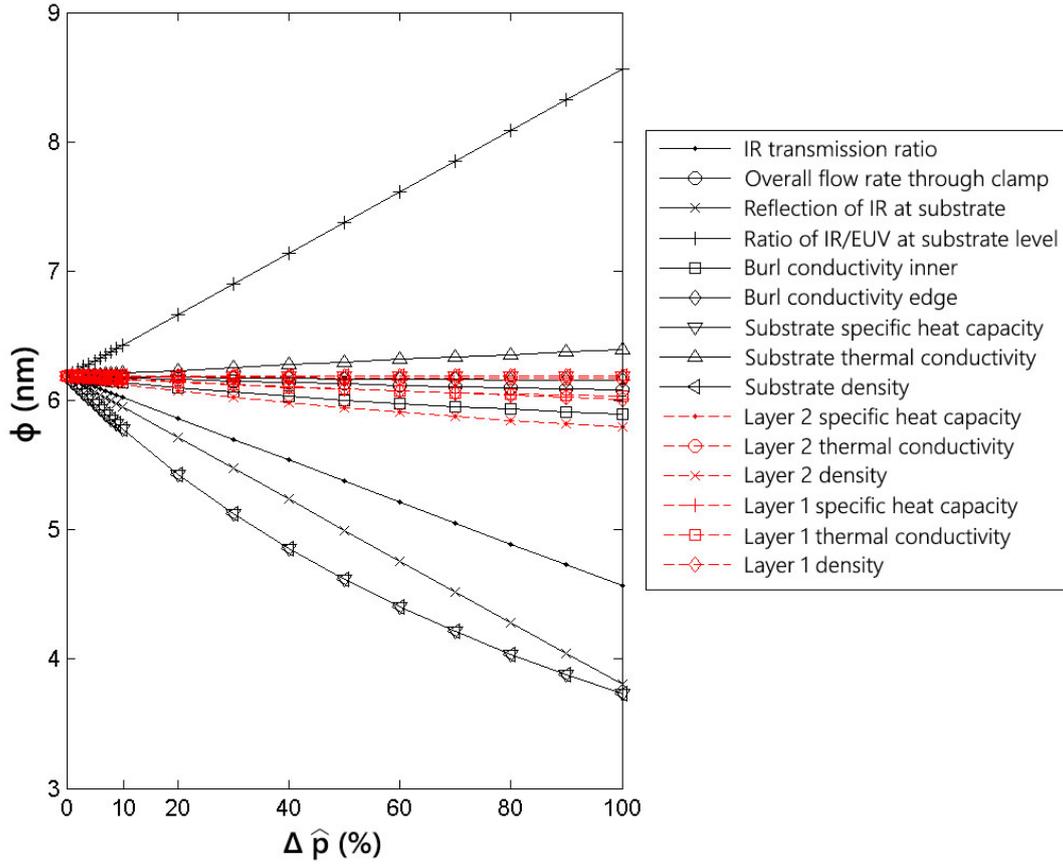
The matrix  $\mathbf{I}$  denotes the square identity matrix of appropriate dimensions.

### 6.1.1 Thermal Parameter Variation

In this section, the residual function is adjusted. The reference response  $\mathbf{y}_{\text{ref}}(\mathbf{p})$  is not subtracted any more. This enables to distinguish whether a positive perturbation of each thermal parameter  $\hat{\mathbf{p}}$  with respect to its true (nominal) parameter value  $\mathbf{p}$  increases or decreases the objective function.

$$\mathbf{r}(\hat{\mathbf{p}}) = \mathbf{y}_{\text{model}}(\hat{\mathbf{p}}) \quad (6.6)$$

The objective function is evaluated, where each parameter  $\hat{\mathbf{p}}$  is perturbed with respect to its true (nominal) parameter  $\mathbf{p}$ . The perturbation is chosen as an increase of  $c\%$  (from 1% to 100%) with respect to the nominal parameter value or as  $c\%$  when the parameter is a ratio for which the nominal parameter value is equal to zero. The results are shown in Figure 6.2.



**Figure 6.2:** Diagram of maximum absolute objective function values  $\phi(\hat{\mathbf{p}})$  for perturbation  $\Delta\hat{\mathbf{p}}$  of each thermal parameter  $\hat{\mathbf{p}}$  with respect to its true (nominal) parameter value  $\mathbf{p}$ .

From this analysis, it can be seen that the objective function behaves linear for most of the thermal parameters. An exception is given by the specific heat capacity and density of the substrate and of layer 2. For those parameters, the objective function behaves almost linear. This behaviour can be explained by equation (2.21), and for convenience, was given by:

$$\mathbf{A} = \mathbf{E}^{-1} (\mathbf{K} + \mathbf{H})$$

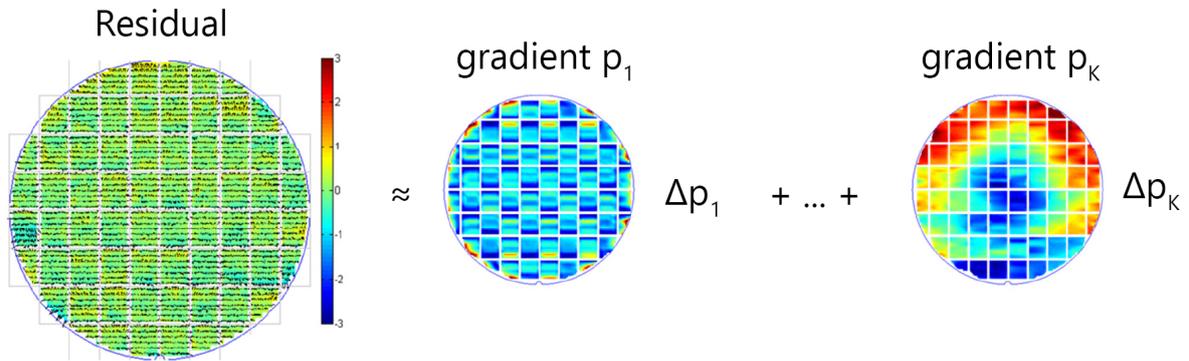
The matrix  $\mathbf{E}^{-1}$  contains the information of the density  $\rho$  and the heat capacity  $c$  for the different type of materials. They appear as the inverse, which explains the non-linear behaviour in Figure 6.2. In equation (2.18), the product of the density and heat capacity are used. Hence, they are not distinguishable and a variation of these parameters results in a similar sensitivity.

From Figure 6.2 it can be seen that five parameters show the largest sensitivity, substrate specific heat capacity, substrate density, reflection of IR at substrate, ratio of IR/EUV at substrate level, and IR transmission ratio. A positive perturbation of the IR/EUV ratio result in an increasing objective function, whereas a positive perturbation of the IR transmission ratio, the reflection of IR at substrate, the specific heat capacity and density of the substrate, result in a decreasing objective function. An increasing objective function implies an increasing impact on substrate deformation and a decreasing objective function implies a decreasing impact on substrate deformation. This kind of sensitivity information is useful to determine a subset of thermal parameters that is used for calibration.

It should be noted that the same kind of analysis is performed for the focus error. Similar results are achieved with the same thermal parameters. These results can be found in van der Meulen (2015).

## 6.2 Thermal Parameter Distinguishability from Overlay Measurements

The parameter variation for each thermal parameter  $\hat{\mathbf{p}}$  manifests itself as a certain shape via the residual function  $\mathbf{r}(\hat{\mathbf{p}})$  (column vector), see Figure 6.3. The mismatch of the corresponding parameters is not distinguishable from the overlay measurement when two shapes of the residual function are identical. The parameters are uniquely distinguishable from the overlay measurement when the residual shapes are not identical to each other.



**Figure 6.3:** An overlay measurement of the substrate. Each thermal parameter  $\hat{\mathbf{p}}$  manifests itself as a certain shape in the residual function  $\mathbf{r}(\hat{\mathbf{p}})$  via its gradient.

A measure for the correlation between the shapes of the thermal parameters is given by the correlation coefficient matrix  $\mathbf{C}$ . This correlation matrix is explained in Section 4.3 in “State of the Art Experiment Design”.

Several simulations with constant input parameters  $\varphi$  (e.g., routing, dose, ...) are performed in order to obtain the necessary residual sensitivity  $\frac{\mathbf{r}_i}{\Delta \hat{\mathbf{p}}_i}$  with  $i = 1, \dots, m$ . Subsequently, the correlation matrix  $\mathbf{C}$  is constructed which is illustrated in Figure 6.4.

IR transmission ratio	Overall flow rate through clamp	Reflection of IR at substrate	Ratio of IR/EUV at substrate level	Burl conductivity inner	Burl conductivity edge	Substrate specific heat capacity	Substrate thermal conductivity	Substrate density	Layer 2 specific heat capacity	Layer 2 thermal conductivity	Layer 2 density	Layer 1 specific heat capacity	Layer 1 thermal conductivity	Layer 1 density
1.0	0.4	0.9	-0.9	0.6	0.4	1.0	-0.2	1.0	0.5	0.3	0.5	0.5	0.5	0.5
0.4	1.0	0.7	-0.7	0.9	0.4	0.4	0	0.4	1.0	0.9	1.0	0.8	0.9	0.8
0.9	0.7	1.0	-1.0	0.9	0.5	0.9	-0.1	0.9	0.8	0.6	0.8	0.8	0.7	0.8
-0.9	-0.7	-1.0	1.0	-0.9	-0.5	-0.9	0.1	-0.9	-0.8	-0.6	-0.8	-0.8	-0.7	-0.8
0.6	0.9	0.9	-0.9	1.0	0.5	0.6	0.1	0.6	0.9	0.8	0.9	0.9	0.9	0.9
0.4	0.4	0.5	-0.5	0.5	1.0	0.4	0	0.4	0.4	0.4	0.4	0.4	0.6	0.4
1.0	0.4	0.9	-0.9	0.6	0.4	1.0	-0.3	1.0	0.5	0.2	0.5	0.6	0.4	0.6
-0.2	0	-0.1	0.1	0.1	0	-0.3	1.0	-0.3	0	0.2	0	-0.1	0.2	-0.1
1.0	0.4	0.9	-0.9	0.6	0.4	1.0	-0.3	1.0	0.5	0.2	0.5	0.6	0.4	0.6
0.5	1.0	0.8	-0.8	0.9	0.4	0.5	0	0.5	1.0	0.9	1.0	0.9	0.8	0.9
0.3	0.9	0.6	-0.6	0.8	0.4	0.2	0.2	0.2	0.9	1.0	0.9	0.7	0.9	0.7
0.5	1.0	0.8	-0.8	0.9	0.4	0.5	0	0.5	1.0	0.9	1.0	0.9	0.8	0.9
0.5	0.8	0.8	-0.8	0.9	0.4	0.6	-0.1	0.6	0.9	0.7	0.9	1.0	0.7	1.0
0.5	0.9	0.7	-0.7	0.9	0.6	0.4	0.2	0.4	0.8	0.9	0.8	0.7	1.0	0.7
0.5	0.8	0.8	-0.8	0.9	0.4	0.6	-0.1	0.6	0.9	0.7	0.9	1.0	0.7	1.0

**Figure 6.4:** Correlation coefficient matrix  $\mathbf{C}$  for parameter variations  $\Delta\hat{\mathbf{p}}$ . The grey fields are discussed in Section 6.4.1.

From the figure it is clear that parameter  $i = 3$  (Reflection of IR) and  $j = 4$  (Ratio of IR/EUV) are not distinguishable from the overlay measurement, since their value is equal to -1 (blue numbers). This is understood from the fact that the product  $(1 - p_3) \cdot p_4$  determines the amount of IR light that is supplied to the substrate. Hence, it is useless to optimize or calibrate for both parameters. The same holds for several other parameters. They are colored in Figure 6.4 as well.

Some parameters do have a high correlation value, but are still beneath absolute value 1. This indicates that it is possible to distinguish the thermal parameters, but it will often hamper the optimization algorithm due to poor distinguishability.

### 6.3 Thermal Parameter Subset Selection

The previously performed sensitivity- and correlation analysis was done in order to select a well-chosen subset of thermal parameters. This subset of parameters is used for calibration. In van der Meulen (2015) a ranking based on the previous analyses is given. The ranking is based on the expected maximum absolute objective function  $\phi(\hat{\mathbf{p}})$  values. The results of the ranking and thus the final subset of parameters is given in Table 6.1.

**Table 6.1:** Overview of thermal parameter subset selection with ranking for overlay, focus, and choice of the final subset.

NO.	RANKING FOR OVERLAY	RANKING FOR FOCUS	FINAL SUBSET
1	Ratio of IR/EUV at substrate	Ratio of IR/EUV at substrate level	Combined IR thermal load parameter
2	Reflection of IR at substrate	Reflection of IR at substrate	
3	Substrate specific heat capacity	IR transmission ratio	IR transmission ratio
4	Substrate density	Burl conductivity inner	Burl conductivity inner
5	IR transmission ratio	Substrate specific heat capacity	
6		Substrate density	

Since the ratio of IR/EUV at substrate level and the reflection of IR at substrate level are not distinguishable from the overlay measurement, these thermal parameters are combined into one thermal parameter "*Combined IR thermal load parameter*". This thermal parameter determines the IR heat load that is supplied to the substrate. In Section 6.2 it can be seen that the product of heat capacity and density of the substrate and the IR transmission ratio are not distinguishable from the overlay measurement. The product of specific heat capacity and density implies the calibration of substrate-to-substrate variation. However, the IR transmission ratio implies the calibration of product-to-product variation. Hence, the latter is included in the subset.

## 6.4 Numerical Optimization

In the following section several optimization are performed in order to find an initial overlay experiment optimization and to calibrate the thermal parameters.

**Optimization Algorithm** The optimization algorithm is a gradient-based optimization algorithm for non-linear least squares alike problems. This algorithm is extensively described in Tinnemans (2010). The parameter optimization problem is posed as:

$$\mathbf{p} = \arg \min \phi(\hat{\mathbf{p}}) \quad \hat{\mathbf{p}} \in \Omega \quad (6.7)$$

Here,  $\Omega \in \mathbb{R}^m$  is the feasible set of parameters and  $\mathbf{p} \in \Omega$  is a (local) minimum of the objective function  $\phi(\hat{\mathbf{p}})$ .

Lower and upper bounds (constraints) are applied to the optimization problem of the thermal parameters which will influence  $\Omega$ . The parameters which are ratios, are bounded with a lower bound of 0 and an upper bound of 1. The thermal parameters that represent a physical quantity are bounded by a lower bound of 0.7 times the nominal parameter value  $\mathbf{p}$  and an upper bound of 1.3 times the nominal parameter value  $\mathbf{p}$ . For now, these bounds for the latter parameters are arbitrarily chosen.

### 6.4.1 Thermal Parameter Subset Optimization

In order to succeed the calibration of the thermal model parameters, a subset of parameters was selected (see Section 6.3). The gradient-based optimization algorithm is applied in order to calibrate the subset of parameters with an overlay measurement. The parameters are given by:

1. Combined IR thermal load parameter
2. IR transmission ratio

### 3. Burl conductivity inner

In the following paragraphs the overlay measurement is obtained from simulation and two cases are investigated. An overlay measurement without the addition of measurement noise is considered and an overlay measurement with the addition of measurement noise is considered. Furthermore,  $LB$  and  $UB$  represent the lower-, and upper bound respectively.  $Step$  represents the absolute step size that is used to calculate the first order derivatives or the residual function for each parameter.

**No Measurement Noise** The subset of parameters is optimized for calibration. The results are shown in Table 6.2. A comparison of the columns *final* and *true* reveals that the true (nominal) parameter values are returned by the optimization algorithm. The result is achieved after 6 iterations. This implies that the distinguishability of the corresponding parameters is sufficient. In Figure 6.4 it can be seen that the corresponding entries are equal to -0.9, 0.6, and -0.9 (see the grey boxes in the figure), which relates to a high correlation. It can be concluded that a high correlation without any measurement noise is not sufficient to let the optimization algorithm fail. All the thermal parameters are correctly retrieved, since their appearance was not identical in the overlay measurement. However, the number of iterations of the optimization algorithm could be high for highly correlated parameters.

**Table 6.2:** Results of the thermal parameter subset optimization, without the presence of measurement noise.

NO.	PARAMETER	LB	UB	STEP	INITIAL	FINAL	TRUE
1	IR transmission ratio	0	1	0.01	0.1	0	0
2	Combined IR thermal load parameter	0.2188	0.4063	0.0031	0.3438	0.3125	0.3125
3	Burl conductivity inner	1050.0	1950.0	15.0	1650.0	1500.0	1500.0

**Measurement Noise** In reality, the overlay measurement suffers from measurement noise. This noise is taken into account via a modification of the simulated response  $\mathbf{y}_{\text{ref}}(\mathbf{p})$ , which was explained in Section 6.1. A disturbance  $\boldsymbol{\eta}$  is added which leads to the following response:

$$\mathbf{y}_{\text{ref},\boldsymbol{\eta}}(\mathbf{p}) = \mathbf{y}_{\text{ref}}(\mathbf{p}) + \boldsymbol{\eta} \quad (6.8)$$

Here,  $\boldsymbol{\eta}$  denotes a vector of appropriate dimension and its entries are normally distributed random variables with mean = 0 nm and  $3^*$  standard deviation = 0.5 nm<sup>1</sup>. The residual function  $\mathbf{r}(\hat{\mathbf{p}})$  is now defined by:

$$\mathbf{r}(\hat{\mathbf{p}}) = \mathbf{y}_{\text{model}}(\mathbf{p}) - \mathbf{y}_{\text{ref},\boldsymbol{\eta}}(\mathbf{p}) \quad (6.9)$$

The new residual function is used for this optimization and, for completeness, the total objective function is given below together with the required definitions.

$$\phi(\hat{\mathbf{p}}) = \rho(\mathbf{V}\mathbf{r}(\hat{\mathbf{p}})) \quad (6.10)$$

$$\rho(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{x}}{2} \quad (6.11)$$

$$\mathbf{V} = 1 \cdot 10^9 \mathbf{I} \quad (6.12)$$

<sup>1</sup>Information obtained from Hein Castelijns (ASML) d.d. April 09, 2014.

An initial investigation for 1 run is performed. The gradient-based optimization algorithm is executed for 1 arbitrary realization of the disturbed response  $\mathbf{y}_{\text{ref},\eta}(\mathbf{p})$ . The results for the case with the addition of measurement noise are shown in Table 6.3 and are achieved after 12 iterations. Although it is not shown, the objective function decreases in the first two iterations, whereas the objective function hardly decreases in the remaining iterations. This suggests a change in the stopping criterion of the optimization algorithm to limit the duration of the optimization.

It can be seen that the true (nominal) parameters are not returned by the optimization algorithm. Nevertheless, the parameter values after optimization are close to the (nominal) parameters. This is easily observed when the deviations of the optimized parameters are expressed in terms of percentages (see the final column). These percentages are +0.75%, -0.18%, and -1.5% for the IR transmission ratio, the combined IR thermal load parameter, and the burl conductivity inner, respectively. Notice that these percentages are equal to +10% before optimization. Hence, this implies that convergence is achieved.

**Table 6.3:** Results of the thermal parameter subset optimization with the addition of measurement noise for 1 run.

NO.	PARAMETER	LB	UB	STEP	INITIAL	FINAL	TRUE
1	IR transmission ratio	0	1	0.01	0.1	0.0075 (+0.75 %)	0
2	Combined IR thermal load parameter	0.2188	0.4063	0.0031	0.3438	0.3119 (-0.18 %)	0.3125
3	Burl conductivity inner	1050.0	1950.0	15.0	1650.0	1477.0 (-1.5 %)	1500.0

Next, a statistical analysis for 100 runs is performed. The gradient-based optimization algorithm is executed for 100 different realizations of the disturbed response  $\mathbf{y}_{\text{ref},\eta}(\mathbf{p})$ . The results of the statistical analysis are shown in Table 6.4. Here, the columns Mean and Std contain the mean and the standard deviation of the corresponding parameters. Both the mean and the standard deviation are expressed in terms of a percentage with respect to the true (nominal) parameter value  $T$  as well. These percentages are found in the columns  $(\cdot-T)/T$  and  $(\cdot/T)$ , respectively. The dot denotes the value that is expressed in terms of a percentage. It can be noticed that the mean and standard deviation of the IR transmission ratio are directly converted into a percentage, since the true parameter value is equal to 0.

**Table 6.4:** Results of the thermal parameter subset optimization with the addition of measurement noise for 100 (different) runs of  $\mathbf{y}_{\text{ref},\eta}(\mathbf{p})$ .

NO.	PARAMETER	TRUE	INITIAL	MEAN	$(\cdot-T)/T$	STD	$(\cdot/T)$
1	IR transmission ratio	0	0.1	0.00495	+0.50 %	0.0071	0.71 %
2	Combined IR thermal load parameter	0.3125	0.3438	0.3140	+0.48 %	0.0026	0.83 %
3	Burl conductivity inner	1500.0	1650.0	1514	+0.94 %	36	2.4 %

### 6.4.2 Optimization of Overlay Experiment

The motivation for optimization of the overlay experiment is to make the appearance of two physical parameters in overlay more orthogonal. For now, two physical parameters are used to illustrate the concept. The goal is to improve the optimization problem described in Section 6.4.1, such that it is efficiently solved by the optimization algorithm.

Parameter  $i = 1$  (IR transmission ratio) and parameter  $j = 2$  (overall flow rate through clamp) are considered in order to calibrate. The experimental input parameter to be optimized, is selected as the effective source power at substrate level. Recalling Figure 1.5, the optimization that is going to be performed is illustrated in the transformation 2 block. Thus, the thermal model parameters are **not** calibrated by this optimization.

The initial experimental input parameter value is equal to  $\hat{u} = 44.2 \text{ W m}^{-1}$ . The optimal input parameter  $u^*$  is unknown. The objective function is defined as:

$$\phi(\hat{u}) = \mathbf{C}(i, j) \quad (6.13)$$

The correlation coefficient matrix  $\mathbf{C}$  is calculated according equation (4.5). The results of the optimization of the experimental input parameter are listed in Table 6.5. The optimal input parameter  $u$  is equal to  $u = 39.2512 \text{ W m}^{-1}$ . In Table 6.5 it can be seen that the correlation coefficient matrix  $\mathbf{C}$  shows a decrease in the off-diagonal term from 0.4039 to 0.3962. Although this is an improvement of orthogonality between the thermal parameter  $i = 1$  and  $j = 2$ , it can be considered as a small improvement. Apparently, this experimental input parameter is not very well able to improve the distinguishability of these two thermal model parameters. Nevertheless, the concept works.

**Table 6.5:** Results of overlay experiment optimization for the input parameter *source power at substrate level*.

PARAMETER	UNIT	ORIGINAL	OPTIMIZED
Effective source power at substrate level	$\text{W m}^{-1}$	44.2	39.2512
Correlation coefficient matrix	-	$\begin{bmatrix} 1 & 0.4039 \\ 0.4039 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.3962 \\ 0.3962 & 1 \end{bmatrix}$

## 6.5 Conclusion Preliminary Calibration

In this preliminary calibration investigation, it is shown that the subset of thermal parameters can be estimated from an experiment to their true values, without the presence of any measurement noise. When measurement noise is included, the subset of thermal parameters can be estimated as well. However, their true values are not found. When performing multiple experiments together with the associated parameter optimization, a statistical expression can be made for each thermal parameter (see Table 6.4). Unfortunately, this method requires multiple experiments which can become quite expensive. An alternative method, as is described in Chapter 5, might be useful, since only one experiment is sufficient to present statistical data of the thermal parameters.

Optimization of the correlation coefficient between parameters is useful to enhance parameter estimation, and therefore improves overlay performance. The concept of optimization for overlay experiments is described in Section 6.4.2, where the correlation between IR transmission ratio and the overall flow rate through clamp was optimized by adjusting the effective source power at substrate level. A small improvement was achieved, as can be seen in Table 6.5. The

concept works, but even more improvement of the overlay experiment might be achieved by using more sophisticated experiment design methods as described in Chapter 4.

By extending the preliminary calibration investigation, useful information can be achieved when using more advanced methods. Overlay experiments could be optimized, for enhancement of parameter estimation, using more sophisticated objective functions and valuable statical information about the model parameters can be achieved within a single experiment.

## Part II

# Application of Experiment Design



# 7

## Simplified Thermo-Mechanical Model

In Chapter 2 the thermo-mechanical feedforward model is explained. The purpose of this model is to correct for the effects of substrate heating on overlay and focus. Although the performance of the feedforward model is satisfactory, there are some beneficial reasons to build a simplified version in Matlab of this model. These reasons are listed below:

- Thermal and mechanical parts can be combined in one model.
- Fast evaluation for different mechanical parameters is possible. Currently, the mechanical part is solved via a pre-calculated (Ansys) deformation matrix  $\mathbf{C}$ .
- It is easy to modify the geometry.
- A good understanding of the model and working principle is obtained.

It is of great importance that the simplified Matlab model can capture all relevant aspects of substrate heating and overlay performance. The purpose of this model is to apply experiment design techniques and for proof of concept.

At first, the Finite Element Method (FEM) used for the simplified Matlab model is explained. Next, the model layout along with its assumptions is given, followed by the governing equations that are required to perform simulations. Thereupon, the modelling of the applied heat load is described and finally, the input design parameters are given together with a validation of the simplified thermo-mechanical feedforward model.

### 7.1 Finite Element Method

The simplified version of the thermo-mechanical feedforward model is built in Matlab using the Finite Element Method toolbox made by Jeroen de Best. This toolbox is able to describe thermo-mechanical problems and is extensively described in de Best (2015). In this section, relevant aspects of the FEM toolbox are given to provide background information of the working principles on the simplified model.

#### 7.1.1 Thermal Equation

The thermal differential equation and its derivation used to describe thermal interaction with an infinitesimal small volume is already given in Section 2.4. For convenience, the thermal differential equation will be repeated.

$$\rho c \frac{\partial T}{\partial t} - \kappa \frac{\partial^2 T}{\partial x_i^2} = Q \quad \in V(x_i), \quad i = 1, 2, 3 \quad (7.1)$$

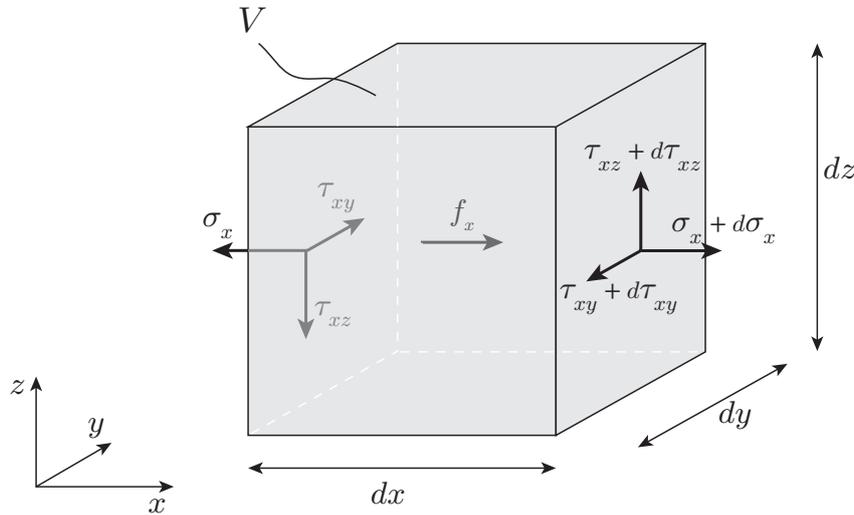
Here  $\rho$ ,  $c$  and  $\kappa$  are the mass density ( $\text{kg m}^{-3}$ ), heat capacitance ( $\text{J kg}^{-1} \text{K}^{-1}$ ), and heat conductivity ( $\text{W m}^{-1} \text{K}^{-1}$ ), respectively, in the volume  $V$ . The heat load is expressed by  $Q$  in W.

### 7.1.2 Mechanical Equations

The mechanical equations are derived from the conservation of momentum principle. A force balance in  $x$  direction can be described as:

$$-\sigma_x dydz + (\sigma_x + d\sigma_x) dydz - \tau_{xy} dx dz + (\tau_{xy} + d\tau_{xy}) dx dz - \tau_{xz} dx dy + (\tau_{xz} + d\tau_{xz}) dx dy + f_x dx dy dz = \rho dx dy dz \frac{\partial^2 u}{\partial t^2} \quad (7.2)$$

Where  $\sigma$  are the normal stresses in  $\text{N m}^{-2}$ ,  $\tau$  the shear stresses in  $\text{N m}^{-2}$ ,  $f_x$  the applied body force in  $\text{N m}^{-3}$ ,  $\rho$  the density in  $\text{kg m}^{-3}$ ,  $u$  the displacement in  $x$  direction in m and  $t$  is the time in s.



**Figure 7.1:** Conservation of momentum in an infinitesimal volume. Not all normal stresses  $\sigma$  and shear stresses  $\tau$  are depicted.

In order to derive the normal- and shear stresses at a small deviation of the initial positions, a first order Taylor expansion is used (see de Best (2015)):

$$\sigma_x + d\sigma_x = \sigma_x + \frac{\partial \sigma_x}{\partial x} dx \quad (7.3)$$

$$\sigma_y + d\sigma_y = \sigma_y + \frac{\partial \sigma_y}{\partial y} dy \quad (7.4)$$

$$\sigma_z + d\sigma_z = \sigma_z + \frac{\partial \sigma_z}{\partial z} dz \quad (7.5)$$

After substitution of these equations into equation (7.2) and dividing by  $dx dy dz$ , the following result is achieved:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} + f_x = \rho \frac{\partial^2 u}{\partial t^2} \quad (7.6)$$

Similar equations are achieved for the directions  $y$  and  $z$ .

$$\frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{zy}}{\partial z} + f_y = \frac{\rho \partial^2 v}{\partial t^2} \quad (7.7)$$

$$\frac{\partial \sigma_z}{\partial z} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \tau_{xz}}{\partial x} + f_z = \frac{\rho \partial^2 w}{\partial t^2} \quad (7.8)$$

Here,  $v$  and  $w$  are the displacements in  $y$  and  $z$  direction respectively. Hooke's law is used to convert stresses into strain:

$$\boldsymbol{\sigma} = \mathbf{H}\boldsymbol{\varepsilon} \quad (7.9)$$

With  $\boldsymbol{\varepsilon}$  the strain vector and  $\mathbf{H}$  the Hooke's matrix

$$\mathbf{H} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix} \quad (7.10)$$

where  $E$  is the Young's modulus (Pa) and  $\nu$  the Poisson ratio (-). When combining Hooke's equation together with the conservation of momentum equations, the mechanical solution can be written as:

$$\nabla^T \cdot \boldsymbol{\sigma} + \mathbf{f} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad \rightarrow \quad \nabla^T \cdot \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{f} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (7.11)$$

Where the nabla operator  $\nabla$ , stress vector  $\boldsymbol{\sigma}$ , force vector  $\mathbf{f}$ , and displacement vector  $\mathbf{u}$  are defined as:

$$\nabla = \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial y} & 0 \\ 0 & 0 & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} \end{bmatrix}, \quad \boldsymbol{\sigma} = \begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{xz} \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \quad (7.12)$$

### 7.1.3 Thermo-Mechanical Equations

In the mechanical part, strains arise from mechanical stresses via Hooke's law. However, in the thermo-mechanical coupling, strains also originate due to thermal expansion. This can be written as:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_m - \boldsymbol{\varepsilon}_t \quad (7.13)$$

The strains that arise due to mechanical stresses are represented by  $\boldsymbol{\varepsilon}_m$ , whereas strains that arise due to thermal expansion are represented by  $\boldsymbol{\varepsilon}_t$ . It should be noted that in contrast to the mechanical strain vector, the thermal strain vector is written as,

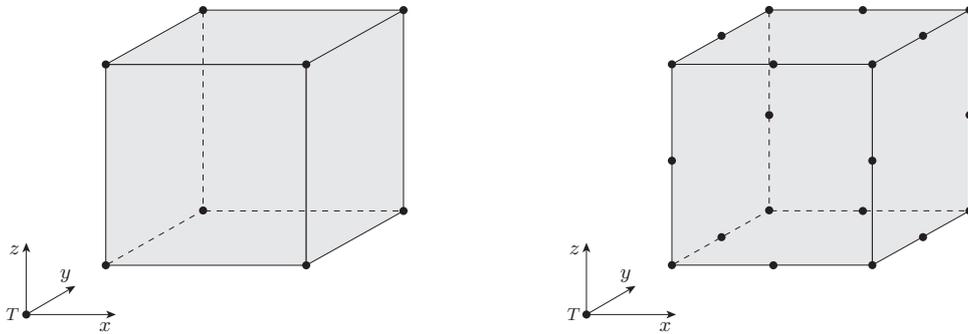
$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \boldsymbol{\varepsilon}_t = \mathbf{s}\boldsymbol{\varepsilon}_t \quad (7.14)$$

since a thermal expansion will only lead to an elongation and, therefore, introduces no shear strains. When substituting 7.13 into 7.11, the thermo-mechanical equation is defined by:

$$\nabla^T \cdot \mathbf{H}(\boldsymbol{\varepsilon}_m - \boldsymbol{\varepsilon}_t) + \mathbf{f} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad (7.15)$$

#### 7.1.4 Elements

In order to discretize the model, elements are being used. It is important to have knowledge about the type of elements that are used. In the FEM toolbox, three dimensional linear and quadratic hexahedrons finite elements exist. These type of finite elements are given in Figure 7.2. The linear hexahedron is often referred to as the 8-node brick, whereas the quadratic hexahedron is referred to as the 20-node brick. The 8-node brick has 8 nodes and, obviously, the 20-node brick has 20 nodes. Each node consists of three mechanical DOFs  $u$ ,  $v$  and  $w$  which represent displacements in  $x$ ,  $y$ , and  $z$  direction, respectively. A thermal degree of freedom  $T$  for each node is included as well.



**Figure 7.2:** Type of elements used in the FEM toolbox. On the left the linear 8-node brick is illustrated. On the right the quadratic 20-node brick is illustrated.

Between the nodes in an element, interpolation using shape functions is used to construct continuous functions of  $T$ ,  $u$ ,  $v$  and  $w$ . The interpolation is written as:

$$\begin{aligned} T(x, y, z) &= [N_1(x, y, z) \quad N_2(x, y, z) \quad \dots \quad N_n(x, y, z)] \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{bmatrix} = \mathbf{N}(x, y, z) \mathbf{T} \\ u(x, y, z) &= [N_1(x, y, z) \quad N_2(x, y, z) \quad \dots \quad N_n(x, y, z)] \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \mathbf{N}(x, y, z) \mathbf{u} \\ v(x, y, z) &= [N_1(x, y, z) \quad N_2(x, y, z) \quad \dots \quad N_n(x, y, z)] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \mathbf{N}(x, y, z) \mathbf{v} \\ w(x, y, z) &= [N_1(x, y, z) \quad N_2(x, y, z) \quad \dots \quad N_n(x, y, z)] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \mathbf{N}(x, y, z) \mathbf{w} \quad (7.16) \end{aligned}$$

For the 8-node brick  $n$  is equal to 8 and for the 20-node brick  $n$  is equal to 20. The shape functions are represented by  $\mathbf{N}$ , the nodal temperature vector by  $\mathbf{T}$  and the nodal displacement

vectors by  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  in the directions  $x$ ,  $y$ , and  $z$ , respectively. The shape functions can be found in appendix B.

### 7.1.5 Galerkin Method

The Galerkin method is a method for converting a continuous differential equation into a discrete problem. It provides a residual minimization by:

1. Multiplying the differential equation by weighting functions,
2. Integrating over the element and,
3. Equation it to zero.

In de Best (2015) the Galerkin method will be explained in more detail for the thermal, mechanical, and thermo-mechanical case. In this section, the end result of the thermo-mechanical case will be given.

The thermo-mechanical equation (7.15) is given in the Galerkin statement,

$$\iiint_V \mathbf{N}^T \left( \nabla^T \cdot \boldsymbol{\sigma} + \mathbf{f} - \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \right) dV = 0 \quad (7.17)$$

with  $\boldsymbol{\sigma} = \mathbf{H}(\boldsymbol{\varepsilon}_m - \boldsymbol{\varepsilon}_t)$ . Although the derivation is not given, by writing out equation (7.17) and using the divergence theorem, the elementary system matrices for an element are described by:

$$\begin{aligned} - \iiint_V \mathbf{B}^T \mathbf{H} \mathbf{B} dV \mathbf{q} + \iiint_V \mathbf{B}^T \mathbf{H} \boldsymbol{\varepsilon}_t(T) dV + \iint_S \mathbf{N}^T \mathbf{t} dS + \iiint_V \mathbf{N}^T \mathbf{f} dV - \\ \iiint_V \rho \mathbf{N}^T \mathbf{N} dV \frac{\partial^2 \mathbf{q}}{\partial t^2} = 0 \end{aligned} \quad (7.18)$$

Or can be written as:

$$\mathbf{M} \ddot{\mathbf{q}} + \mathbf{K} \mathbf{q} = \mathbf{f}_b + \mathbf{f}_s + \mathbf{f}_t \quad (7.19)$$

With,

$$\begin{aligned} \mathbf{M} &= \iiint_V \rho \mathbf{N}^T \mathbf{N} dV, & \mathbf{K} &= \iiint_V \mathbf{B}^T \mathbf{H} \mathbf{B} dV, \\ \mathbf{f}_b &= \iiint_V \mathbf{N}^T \mathbf{f} dV, & \mathbf{f}_s &= \iint_S \mathbf{N}^T \mathbf{t} dS \\ \mathbf{f}_t &= \iiint_V \mathbf{B}^T \mathbf{H} \boldsymbol{\varepsilon}_t(T) dV \end{aligned} \quad (7.20)$$

where  $\mathbf{M}$  and  $\mathbf{K}$  are the mass- and stiffness matrices of the element, respectively,  $\mathbf{B}$  is the spatial derivative of the shape functions  $\mathbf{N}$ ,  $\mathbf{f}_b$  the applied body force,  $\mathbf{f}_s$  the applied surface force vector and  $\mathbf{f}_t$  the thermal force vector. The implementation of the thermal force is briefly described in more detail. The scalar value of the thermal strain can be expressed as:

$$\varepsilon_t(T) = \alpha_0 + \alpha_1 T \quad (7.21)$$

Where  $\alpha_1$  is the thermal expansion coefficient of the associated material and  $\alpha_0$  is the reference coefficient determined by:

$$\alpha_0 = -\alpha_1 T_{\text{ref}} \quad (7.22)$$

Here,  $T_{\text{ref}}$  is the initial or reference temperature of the system. In equation (7.21) the temperature  $T$  can be substituted by its shape functions.

$$T = \mathbf{N}\mathbf{T} \quad (7.23)$$

Using equation (7.14), the thermal strain vector can be written as,

$$\boldsymbol{\varepsilon}_t(T) = \alpha_0 \mathbf{s} + \alpha_1 \mathbf{s}\mathbf{N}\mathbf{T} \quad (7.24)$$

and substitution of this result into the thermal force vector  $\mathbf{f}_t$  leads to:

$$\mathbf{f}_t = \iiint_V \alpha_0 \mathbf{B}^T \mathbf{H} \mathbf{s} \, dV + \iiint_V \alpha_1 \mathbf{B}^T \mathbf{H} \mathbf{s} \mathbf{N} \, dV \mathbf{T} \quad (7.25)$$

Or can be written as:

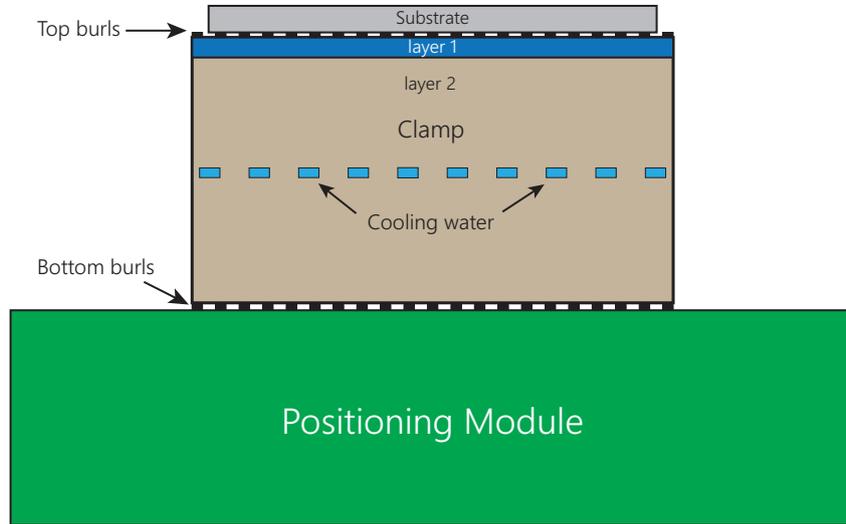
$$\mathbf{f}_t = \mathbf{L}_0 + \mathbf{L}_1 \mathbf{T} \quad (7.26)$$

With,

$$\mathbf{L}_0 = \iiint_V \alpha_0 \mathbf{B}^T \mathbf{H} \mathbf{s} \, dV, \quad \mathbf{L}_1 = \iiint_V \alpha_1 \mathbf{B}^T \mathbf{H} \mathbf{s} \mathbf{N} \, dV \quad (7.27)$$

## 7.2 Physical Model Representation

The thermo-mechanical model is rebuilt in Matlab. In Figure 7.3, the structure of the Matlab model is illustrated.



**Figure 7.3:** Schematic illustration of the thermo-mechanical model in Matlab. The clamp with substrate are identical as in Section 2.2, except burls located at the bottom. The positioning module, used for mechanical calculations, is included as well.

From the figure, it can be seen that the structure of the clamp is unchanged when compared to the already available thermo-mechanical model. The main difference is that the clamp is positioned on top of the bottom burls and the positioning module, and are combined in one model.

### 7.2.1 Geometry and Mesh

In the Matlab model, the geometry of the physical system is simplified. In reality the substrate and clamp are both circular, but they are modelled as square plates with different type of layers. This simplification will have effect on substrate heating (more mass is present) and stiffness, but it is assumed that these effects are small and certainly will not influence the concept of experiment design.

The positioning module contains all kinds of structures inside it self to reduce mass, but maintain stiffness. In this simplified Matlab model, the position module is modelled as one solid block. This simplification has no effect on heating (will be seen in Section 7.2.3), but it will have effect on the stiffness. As before, it is assumed that this will influence overlay accuracy by a small amount, but will not influence the concept of experiment design.

In the table below, the geometrical specifications of the model are listed.

**Table 7.1:** Geometry of the simplified Matlab model.

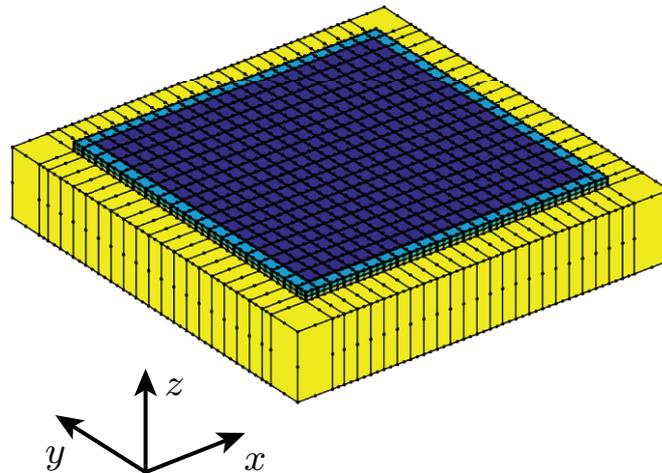
PART	LENGTH (mm)	WIDTH (mm)	HEIGHT (mm)
Substrate	300	300	0.775
Top burls	300	300	0.01
Layer 1	325	325	0.8
Layer 2	325	325	8.3
Bottom burls	325	325	0.01
Positioning module	400	400	70

**Mesh Grid** A suitable mesh size must be chosen in order to solve the substrate heating problem within the specified requirements. There are a few consideration to take into account. These are:

- Desired overlay accuracy
- Computation performance (CPU time, memory,...)
- Numerical stability during transient simulations

There are more considerations possible for choosing a suitable mesh size, for instance when high non-linearity occurs, but the criteria mentioned above have the highest priority for this problem.

In order to keep a balance between computation performance and desired overlay accuracy, the mesh grid is chosen such that each field on the substrate consists in total of 4 elements, two in each direction  $x$  and  $y$ . On the substrate, 11 fields fit in  $x$  direction and 9 fields fit in  $y$  direction and this results in a mesh grid of 22x18 elements in the  $xy$  plane at substrate level. However, this choice is inconvenient, since some of the nodes at the edge of the elements would be located in the space between the fields. Therefore their information is lost when calculating overlay (overlay is calculated in the space of the fields). A simple fix would be a shift in position of all the fields, but a slightly smaller grid size of 21x17 elements in the  $xy$  plane is chosen instead. This improves the computation performance as well. Since an implicit time integration scheme with unconditional stability criterion is used, the numerical stability during transient simulations is guaranteed (will be seen in Section 7.3.1).



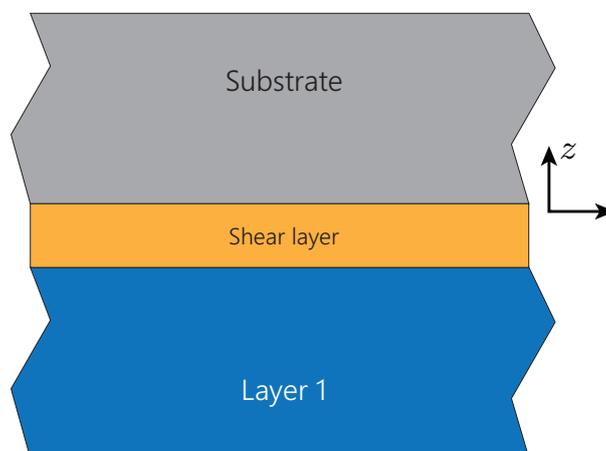
**Figure 7.4:** Geometry of the simplified thermo-mechanical model in Matlab. Unlike the real system, the substrate and clamp are square instead of a circle. The substrate consists of  $21 \times 17$  elements, and the mesh and elements are illustrated by the black lines.

In  $z$  direction, it is chosen to use one element per layer, except layer 2 of the clamp. This layer is divided into an upper and lower part. The final mesh grid is  $21 \times 17 \times 1$  seen from the substrate layer and is propagated throughout the rest of the layers. The total amount of elements is 2987. In Figure 7.4 the Matlab model with mesh is shown.

### 7.2.2 Burl layers

The implementation of the top and bottom burl layers is simplified compared to the existing thermo-mechanical feedforward model. The burl layers are modelled as a shear layer and, therefore, the burl geometry is omitted. In Figure 7.5 the implementation of the burl layer is shown.

Since the mass of the burls is negligible, the thermal heat capacity is neglected, but thermal conduction is included to describe heat transfer from substrate to clamp.



**Figure 7.5:** Burl layers are modelled as a shear layer that conducts heat but has no thermal capacity. The layers have a stiffness in  $z$  direction and in the tangential  $xy$  plane.

From de Best (2015), it is assumed that the burls have stiffness in  $z$  direction and a tangential stiffness in the  $xy$  plane to capture shear stresses. Furthermore, it is assumed that no Poisson effect is present. In Table 7.2 the properties of the burls are listed.

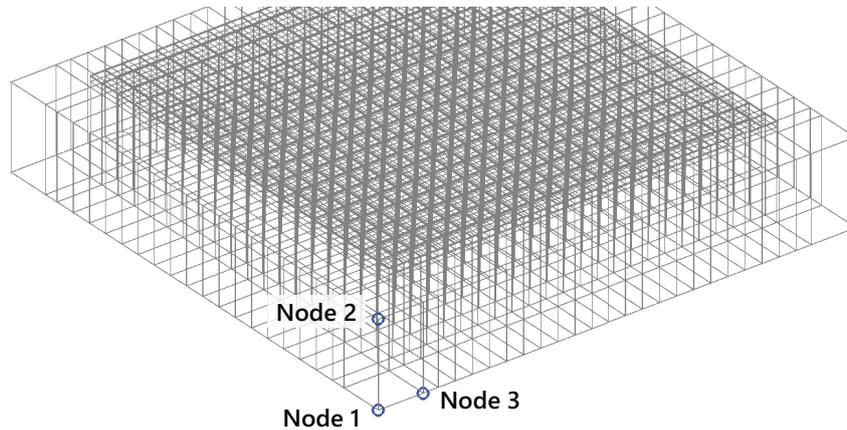
**Table 7.2:** Thermal and mechanical properties of the top- and bottom burls.

PROPERTY	TOP BURLS	BOTTOM BURLS	UNIT
$\rho$	0	0	$\text{kg m}^{-3}$
$E$	12.8	80.7	MPa
$G$	12.8	48.6	MPa
$\nu$	0	0	-
$c_p$	0	0	$\text{J kg}^{-1} \text{K}^{-1}$
$\kappa$	0.015	0.015	$\text{W m}^{-1} \text{K}^{-1}$
$\alpha$	0	0	$\text{K}^{-1}$

### 7.2.3 Constraints

The model has two types of constraints, namely thermal and mechanical constraints.

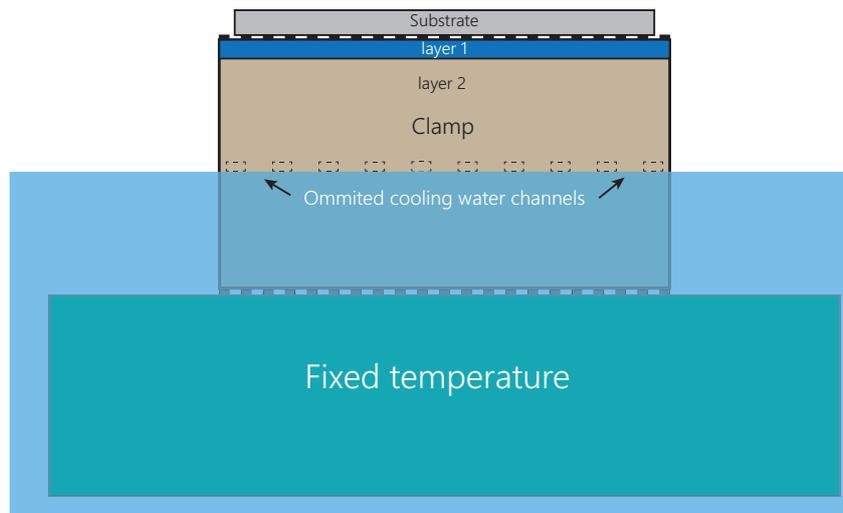
**Mechanical Constraints** In order to prevent singularity (rigid body motions) while calculating mechanical deformations, the model is mechanically attached to the fixed world. The model must be constrained carefully to prevent an over constrained situation. As mentioned before, each node has 3 DOFs. Unfortunately, it is not possible to constrain rotational DOFs of a node. This is a limitation of the used FEM package. According to Figure 7.6, node 1 is constrained in  $x$ ,  $y$  and  $z$ , direction (all DOFs). Node 2 is constrained in  $x$  and  $y$  direction, and node 3 is constrained in  $y$ . The system is properly constrained in all directions, including rotations, but is not over constrained. The clamp and substrate are free to deform.



**Figure 7.6:** Mechanical constraints for simplified thermo-mechanical model. The system is properly constrained in all directions, including rotations.

**Thermal Constraints** In the feedforward model, described in Chapter 2, cooling water is present. In Castelijns (2014) it is stated that the heat transfer below the cooling water channels is little and that heat transfer in the positioning module does not take place. In order to improve computational efficiency for the simplified model, it is assumed that the positioning module and the lower part of the clamp remain constant at initial temperature of  $22\text{ }^\circ\text{C}$ . Furthermore, the

cooling water channels with their controllers are omitted. In Figure 7.7 the fixed thermal constraint is illustrated.



**Figure 7.7:** Thermal constraints for simplified thermo-mechanical model. The lower part of the clamp and the positioning module are assumed to operate at fixed temperature.

### 7.2.4 Model Parameters

In Table 7.3 the mechanical and thermal parameters of the model parts are listed. Properties of the burls can be seen in Table 7.2.

**Table 7.3:** Thermo and mechanical parameters for the different types of layers.

PROPERTY	SUBSTRATE	LAYER 1	LAYER 2	POS MODULE	UNIT
$\rho$	2329	2230	3070	2530	$\text{kg m}^{-3}$
$E$	167	64	373	90.3	GPa
$\nu$	0.30	0.20	0.18	0.30	-
$c_p$	705	830	700	800	$\text{J kg}^{-1} \text{K}^{-1}$
$\kappa$	149	1.2	120	1.5	$\text{W m}^{-1} \text{K}^{-1}$
$\alpha$	$2.4 \cdot 10^{-6}$	$3.25 \cdot 10^{-6}$	$2.5 \cdot 10^{-6}$	$2.0 \cdot 10^{-8}$	$\text{K}^{-1}$

## 7.3 Governing Equations

In this section, the underlying calculations are given in order to solve the transient thermal problem and the mechanical problem. It is assumed that these calculations can be separated since the transient thermal fluctuations occur at a (far) lower rate than the resonances of the mechanical system. Another way to say, the calculations can be separated since the thermal eigenvalues are far below the eigenvalues of the mechanical system.

### 7.3.1 Thermal Equations

The general equation for conservation of energy in an infinitesimal volume is given in equation (2.18). This equation is rewritten in the following thermal problem,

$$\mathbf{C}_p \dot{\mathbf{T}} + \mathbf{K}_c \mathbf{T} = \mathbf{Q} \quad (7.28)$$

where  $\mathbf{C}_p$  is the thermal capacitance matrix containing the thermal masses  $\rho c$ ,  $\mathbf{K}_c$  is the thermal conductivity matrix, and  $\mathbf{Q}$  represents the heat source. Note that, in contrast to equation (2.20), the system is not written in state space notation. This has several advantages such as the avoidance of preliminary inverse calculations – which could introduce numerical noise – and the implementation of the burls is not hampered. In Section 7.2.2, it was assumed that the burls have no mass and thus no heat capacitance. Therefore,  $\mathbf{C}_p$  becomes singular and when the state space notation is used, the inverse does not exist. This implies that a repartition and division of the system matrices is necessary.

In order to solve this first order differential equation, the system is discretized with an Euler backwards integration scheme. This integration scheme in its general form is given by:

$$\frac{dy}{dt} = f(t, y) \quad (7.29)$$

$$y_{k+1} = y_k + \Delta t f(t_{k+1}, y_{k+1}) \quad (7.30)$$

This integration scheme is an implicit method, since the new response depends on the gradient at the next time step. This implicit property is necessary to prevent very small step sizes  $\Delta t$ , since equation (7.28) is a stiff differential equation. This stiff behaviour comes from the fact that the height of the layers of the model vary several orders (see Table 7.1). This leads to fast and slow heating/cooling phenomena and, therefore, small and large time constants are present. The Euler backward scheme is an unconditionally stable integration scheme of order  $\mathcal{O}(\Delta t)$ . The trapezoidal rule is an implicit integration scheme as well, is of order  $\mathcal{O}(\Delta t^2)$  and, is unconditionally stable as well but not so-called *super-stable*, see Vuik et al. (2006). The latter leads to undesired oscillatory effects and is therefore not used.

The Euler backward scheme is applied to equation (7.28):

$$\mathbf{C}_p \left( \frac{\mathbf{T}_{k+1} - \mathbf{T}_k}{\Delta t} \right) + \mathbf{K}_c \mathbf{T}_{k+1} = \mathbf{Q}_{k+1} \quad (7.31)$$

And is rewritten into:

$$\boxed{(\mathbf{C}_p + \mathbf{K}_c \Delta t) \mathbf{T}_{k+1} = \mathbf{C}_p \mathbf{T}_k + \mathbf{Q}_{k+1} \Delta t} \quad (7.32)$$

The system can be partitioned into free and prescribed DOFs (see Figure 7.7).

$$\left( \begin{bmatrix} \mathbf{C}_{p,cc} & \mathbf{C}_{p,cf} \\ \mathbf{C}_{p,fc} & \mathbf{C}_{p,ff} \end{bmatrix} + \begin{bmatrix} \mathbf{K}_{c,cc} & \mathbf{K}_{c,cf} \\ \mathbf{K}_{c,fc} & \mathbf{K}_{c,ff} \end{bmatrix} \Delta t \right) \begin{bmatrix} \mathbf{T}_c^{k+1} \\ \mathbf{T}_f^{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{p,cc} & \mathbf{C}_{p,cf} \\ \mathbf{C}_{p,fc} & \mathbf{C}_{p,ff} \end{bmatrix} \begin{bmatrix} \mathbf{T}_c^k \\ \mathbf{T}_f^k \end{bmatrix} + \begin{bmatrix} \mathbf{Q}_c^{k+1} \\ \mathbf{Q}_f^{k+1} \end{bmatrix} \Delta t \quad (7.33)$$

The first equation can be simplified into  $\mathbf{T}_c^{k+1} = \mathbf{T}_c^k$ , since the temperature of the constrained DOFs are fixed during the whole simulation. In the second equation, the thermal capacitance coupling term  $\mathbf{C}_{p,fc}$  can be omitted. However, thermal conductivity between the constrained and free DOFs must be included. This conduction coupling describes the heat flux from the upper part of the clamp to the simplified cooling water channels. The equations can be rewritten as:

$$\left( \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{p,ff} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \Delta t^{-1} & \mathbf{0} \\ \mathbf{K}_{c,fc} & \mathbf{K}_{c,ff} \end{bmatrix} \Delta t \right) \begin{bmatrix} \mathbf{T}_c^{k+1} \\ \mathbf{T}_f^{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{p,ff} \end{bmatrix} \begin{bmatrix} \mathbf{T}_c^k \\ \mathbf{T}_f^k \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{Q}_f^{k+1} \end{bmatrix} \Delta t \quad (7.34)$$

The second equation is of interest,

$$(\mathbf{C}_{p,ff} + \mathbf{K}_{c,ff} \Delta t) \mathbf{T}_f^{k+1} + (\mathbf{K}_{c,fc} \Delta t) \mathbf{T}_c^{k+1} = \mathbf{C}_{p,ff} \mathbf{T}_f^k + \mathbf{Q}_f^{k+1} \Delta t \quad (7.35)$$

which is simplified as:

$$\mathbf{M}_1 \mathbf{T}_f^{k+1} = \mathbf{M}_2 \mathbf{T}_f^k + \mathbf{M}_3 \mathbf{Q}_f^{k+1} + \mathbf{M}_4 \mathbf{T}_c^{k+1} \quad (7.36)$$

Where,

$$\mathbf{M}_1 = \mathbf{C}_{p,ff} + \mathbf{K}_{c,ff} \Delta t \quad (7.37)$$

$$\mathbf{M}_2 = \mathbf{C}_{p,ff} \quad (7.38)$$

$$\mathbf{M}_3 = \mathbf{I} \cdot \Delta t \quad (7.39)$$

$$\mathbf{M}_4 = -\mathbf{K}_{c,fc} \Delta t \quad (7.40)$$

Equation (7.36) is used to solve the transient thermal problem. It can be seen that for every time step, a system must be solved. This is a result of the choice of the implicit time integration scheme. Since  $\mathbf{M}_1$  is a symmetric matrix, an *Approximate Minimum Degree* (AMD) reordering permutation is applied followed by a Cholesky decomposition. Reordering of  $\mathbf{M}_1$  prevents loss of sparsity when the Cholesky decomposition is performed. Because sparsity is maintained and the Cholesky decomposition is an upper triangular matrix, fast forward/backward substitutions are possible which enhances computation performance. In the following section, more details about Cholesky decomposition and preconditioning are given.

### 7.3.2 Mechanical Equations

Heating of a mechanical structure introduces internal stresses due to thermal expansion. Subsequently, these stresses lead to deformation of the mechanical structure. As mentioned before, the time constants of the structural heating are (far) below the eigenvalues of the mechanical system and the structural dynamics can be considered to have an instantaneous response. Therefore, the mechanical solution can be solved separately according:

$$\mathbf{K} \mathbf{u} = \mathbf{f}, \quad (7.41)$$

where  $\mathbf{K}$  is the stiffness matrix,  $\mathbf{u}$  the deformations of the system and  $\mathbf{f}$  the force vector due to thermal expansion. The thermal force vector  $\mathbf{f}$  is calculated according,

$$\mathbf{f} = \mathbf{L}_0 + \mathbf{L}_1 \mathbf{T} \quad (7.42)$$

with  $\mathbf{T}$  the thermal state of the system and  $\mathbf{L}_0$  and  $\mathbf{L}_1$  defined in equation (7.27).

When observing the stiffness matrix  $\mathbf{K}$ , it can be seen that its size is 9 times larger than the thermal system matrices from equation (7.36). Each node has three mechanical DOFs ( $x$ ,  $y$ , and  $z$ ), while the same node only has one thermal DOF  $T$ . When the size of the system increases, computation time could become a serious problem when no measures are taken.

*Direct solvers*, to solve mechanical systems of the form  $\mathbf{K} \mathbf{u} = \mathbf{f}$ , are popular for their robustness and accuracy. Iterative solvers do exist, but are only favoured when the size of the system is very large, e.g., billions of DOFs (Vuik and Lahaye, 2014). Their application is mostly found in fluidic systems.

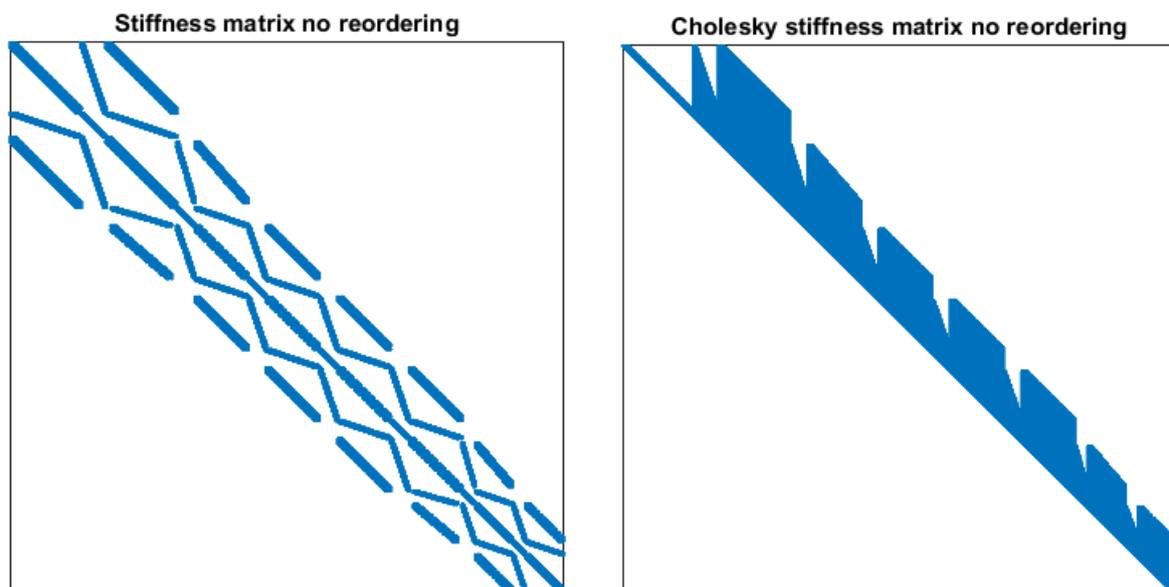
Equation (7.41) is solved via a direct solver by performing a Cholesky decomposition of the stiffness matrix  $\mathbf{K}$  and, subsequently, execute forward and backward substitutions. The Cholesky decomposition can only be used on symmetric positive definite matrices. When this is not the case, LU or LDLT decomposition can be considered, followed by the same forward/backward substitutions.

In order to improve the performance of the direct solver with Cholesky decomposition, the stiffness matrix is reordered via the *Sparse Reverse Cuthill-McKee* ordering algorithm. This

algorithm tries to squeeze the *bandwidth* of the stiffness matrix (Rixen, 2011b). The bandwidth of a matrix is a measure of the distance between the diagonal and the most off-diagonal entries. When the bandwidth is small, the Cholesky decomposition will result in a more sparse matrix, which will improve memory storage and computation time. There exist more preconditioning algorithms, like Approximate Minimum Degree. This pre-conditioner constructs a right-down arrow shape of the matrix, which prevents filled factorized upper and lower triangular matrices. Although these pre-conditioners have different working principles, the final goal is the same: make sure that the Cholesky matrix remains sparse. The Cholesky matrix  $\mathbf{C}$  is an upper triangular matrix and is given by:

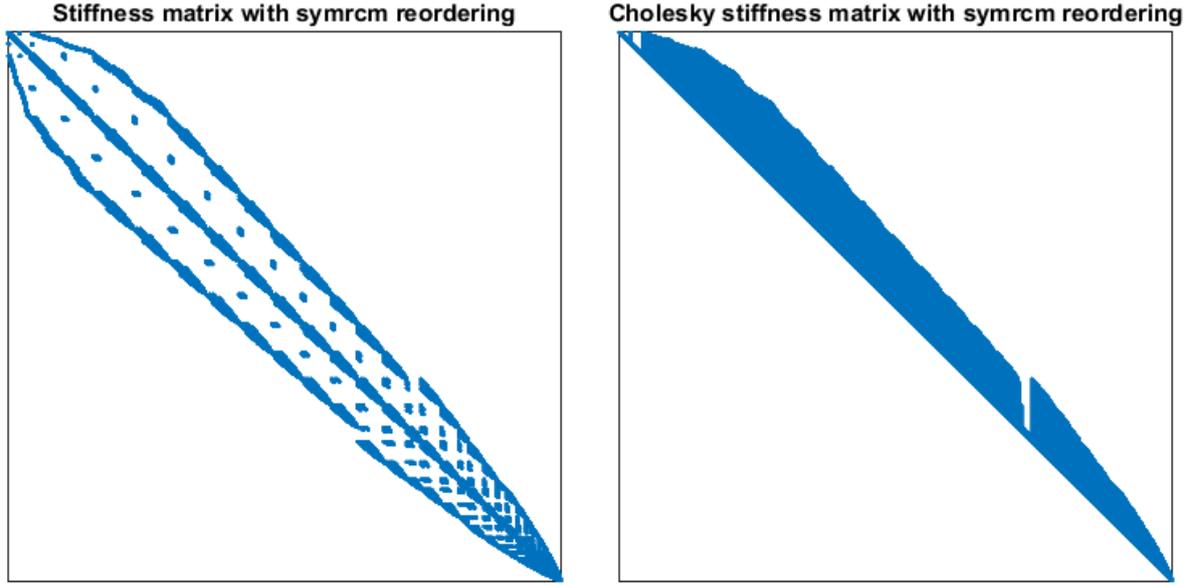
$$\mathbf{K} = \mathbf{C}\mathbf{C}^T \quad (7.43)$$

In Figure 7.8 the structure of the stiffness matrix, together with its Cholesky decomposition, is illustrated. It can be seen that the stiffness matrix is symmetric and that the Cholesky matrix is an upper triangular matrix. The stiffness matrix has a diagonal form, this is caused by the stacking structure of the layers in the model.



**Figure 7.8:** Stiffness matrix with Cholesky decomposition without the use of a reordering scheme. Sparsity of the Cholesky decomposition is 10.6%.

In Figure 7.9 the structure of the reordered stiffness matrix, together with its Cholesky decomposition, is illustrated. The structure of the stiffness matrix has changed due to the preconditioning. The entries in the matrix are more squeezed towards the diagonal. Although it is difficult to see, a serious improvement is obtained in sparsity of the associated Cholesky matrix. The Cholesky decomposition without the use of a reordering scheme has a sparsity of 10.6%, whereas the Cholesky decomposition with the use of the Sparse Reverse Cuthill-McKee ordering scheme has a sparsity of 3.8%.



**Figure 7.9:** Stiffness matrix with Cholesky decomposition with the use of the Sparse Reverse Cuthill-McKee ordering scheme. Sparsity of the Cholesky decomposition is 3.8%.

**Overlay** The measured response of the model is overlay, see Section 1.1. Overlay arises at the moment when the EUV light alters the characteristics of the photo-resist. Bookkeeping is done to keep track at which time instant, during the simulation, the photo-resist is being processed and overlay occurs.

The total overlay response is determined when the transient simulation is completed. Mechanical deformations of the model are calculated from the temperature states at the moments when overlay occurred. It should be noted that only a small part of these mechanical deformations are of interest, namely of the substrate only.

Suppose that during a simulation  $N$  time instances are stored when overlay occurred due to exposure. In order to calculate the overlay, equation (7.41) must be solved  $N$  times with  $N$  different force vectors. The force vectors are calculated via equation (7.42) where  $\mathbf{T}$  is the temperature state at the stored time instance. The deformations  $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_N]$  are obtained, but only a small part of  $\mathbf{U}_i$  is of interest. In order to select the required parts of  $\mathbf{U}_i$  to construct the total overlay, a selection matrix  $\mathbf{S}$  is introduced.  $\mathbf{S}$  is a 3 dimensional sparse matrix which contains ones at the positions where the DOFs of the associated overlay are located. The construction of the overlay  $\mathbf{O}$  is given by:

$$\mathbf{O} = \sum_{i=1}^N \mathbf{S}_i \mathbf{U}_i \quad (7.44)$$

The overlay  $\mathbf{O}$  is a vector which corresponds to substrate deformations in  $x$  and  $y$  directions.

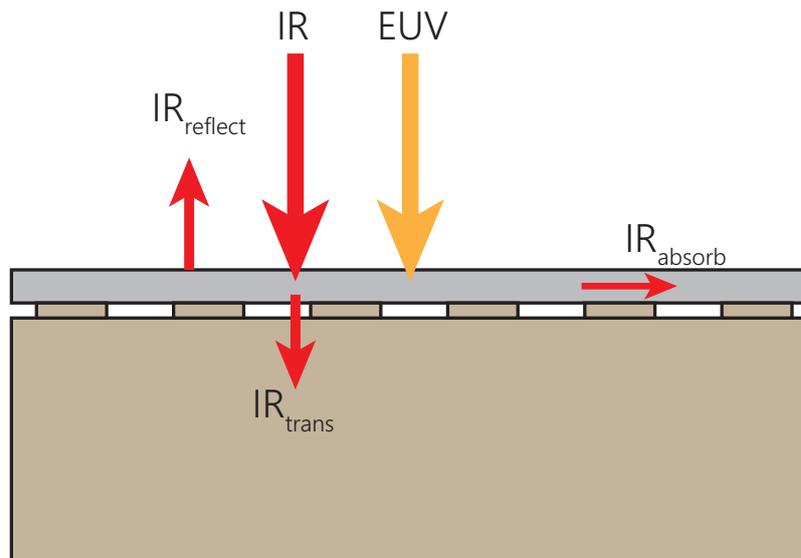
## 7.4 Heat Load

During the exposure of the substrate an external heat load is applied. The heat load is modelled as is described in Section 2.3 except one simplification. The external *unspecified heat load* (UHL) is neglected and can be compensated by additional IR light. For convenience, the assumptions are repeated.

- EUV is fully absorbed by the substrate due to the photo-resist. All photon energy is converted into heat.

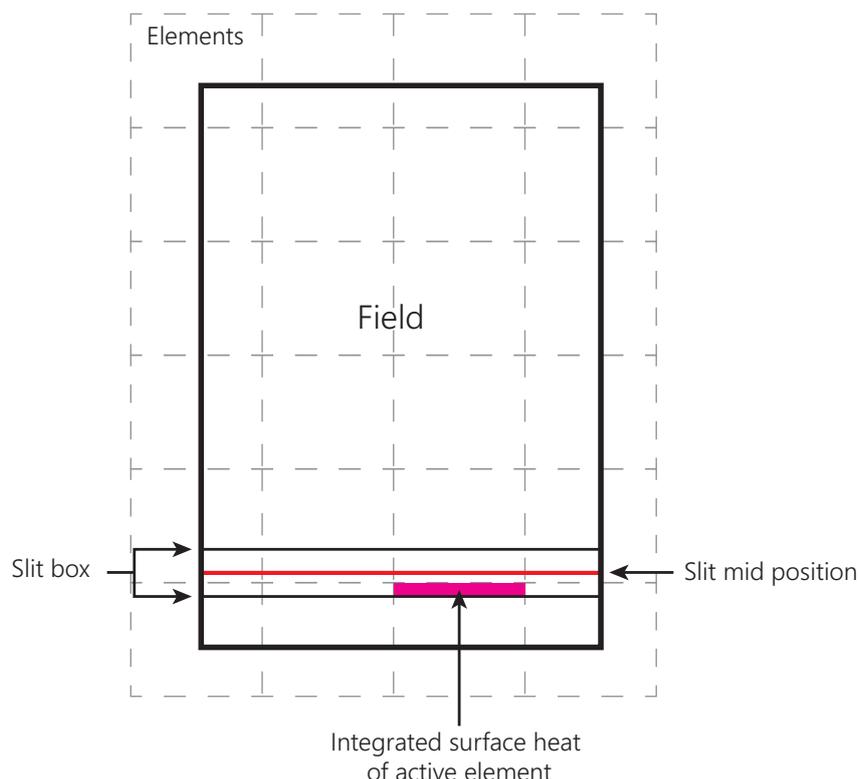
- IR is partly reflected at the substrate surface.
- IR is partly absorbed in the substrate.
- IR is partly transmitted through the substrate.

Figure 7.10 illustrates the applied heat load to the substrate. The cooling water channel is removed, since the cooling is captured in the thermal constraints. Another difference is the UHL load that is neglected in the simplified model. Additional IR could compensate for the missing UHL load, but is omitted in the simplified model.



**Figure 7.10:** Representation of the simplified heat load. Some IR light is reflected or transmitted to the clamp, whereas EUV light is entirely absorbed.

Fields are defined on top of the substrate, but do not necessarily coincide with the elements that represent the substrate (see Figure 7.11). To apply a correct heat load to each element, the overlapping surface of the slit box and its associated elements is calculated. This principle is indicated by the pink box in the figure. Subsequently, each integrated surface is used to apply a correct heat load per active element. All these integrated surface heat loads together represent the dose, which is defined in equation (2.1).



**Figure 7.11:** Slit scanning over elements. Heat load is applied at an overlapping surface of the slit box and the associated element(s).

When the slit box passes associated nodes of the exposure, or in other words when the exposure for the corresponding nodes is completed, the temperature state of the system is stored in order to calculate overlay at the end of the simulation.

In Table 7.4, the heat load parameters are listed which are necessary to determine the heat load from the source up to substrate level. The heat load at substrate level can be calculated according:

$$P_{\text{EUUV}} = P_{\text{source}} \cdot \alpha_{\text{trans}} \cdot \alpha_{\text{mask}} \quad (7.45)$$

Here,  $\alpha_{\text{trans}}$  is the transmission efficiency through the optics and  $\alpha_{\text{mask}}$  is the mask transmission coefficient.

**Table 7.4:** Parameters of applied heat load.

PARAMETER	VALUE
Source power	250 W
transmission efficiency	0.0046 (-)
Mask transmission	0.5 (-)
Substrate reflection	0.5 (-)
Substrate transmission	0 (-)

## 7.5 Input Design Parameters

The simplified model contains several parameters, which can be separated into *model parameters* and *model input parameters*. Model parameters describe how the system behaves when it is

isolated from the environment (e.g., homogeneous solution). For instance, material properties or the geometry of parts.

Model input parameters are used to provide the system required information before it can execute a simulation. One can think of model input parameters as the set-up of the environment of the system (e.g., external forces or external temperatures).

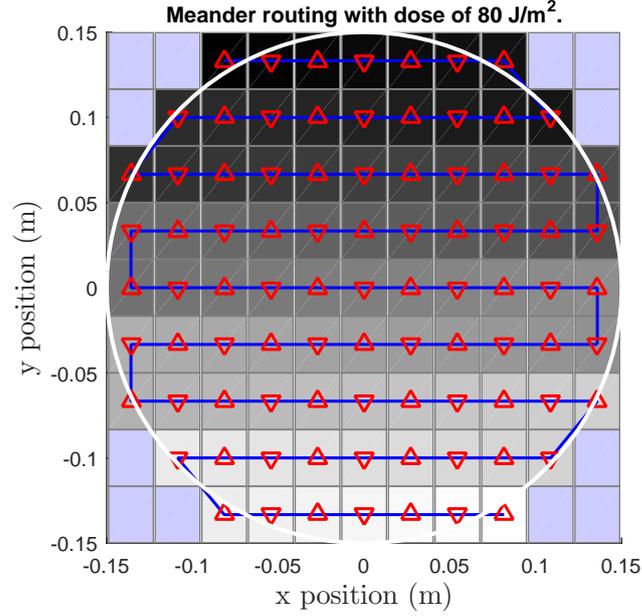
The purpose of experiment design is the enhancement of estimating model parameters by means of optimal conducted experiments. Experiments are defined by the environment and, thus, by the model input parameters. In order to achieve an optimal experiment, the input parameters must be optimized. As was stated in Section 4.5, the input parameters are stored in the design vector  $\varphi$ .

The simplified model consists of several input parameters. These input parameters are listed and briefly explained below.

- **Dose**, the dose determines the amount of energy received per field and is determined according to equation (2.1). Dose will influence heat input and it is noticeable that a change in dose will influence the exposure scan speed for a given source power.
- **Routing**, the routing is responsible for the order of the fields that are exposed. The routing influences the temperature field during simulation.
- **Scan direction**, the scan direction determines the direction of the slit during exposure of a field. Per field, the scan direction can be upwards or downwards. This effect will influence the direction of overlay.

These input parameters are stored in  $\varphi$ , where the dose is a continuous parameter in time in contrast to the routing and the scan direction. The latter input parameters consist of integers only with predefined values.

In order to perform simulations, a nominal set of input parameters is defined. The substrate is exposed to a dose of  $80 \text{ J m}^{-2}$  and a meandering routing with alternating scan directions per field is used. This nominal experiment is illustrated in Figure 7.12. The colours indicate the field order during exposure, with black indicating the start position and white indicating the end position. The red markers indicate the scan direction per field, where an upper triangle indicates a scan motion in positive  $y$  direction and a lower triangle indicates a scan motion in negative  $y$  direction.



**Figure 7.12:** Nominal experiment with meander routing. The colours indicate the field order during exposure, with black indicating the start position and white indicating the end position. The red markers indicate the scan direction per field and the fields are connected with the blue line.

## 7.6 Model Validation

The model is validated by means of a simple hand calculation. First, a nominal simulation is performed with the routing illustrated in Figure 7.12. The corresponding temperature response that makes up the overlay is illustrated in Figure 7.13.

A simple hand calculation is performed in order to validate the model. First, the power at substrate level according equation (7.45) is:

$$P_{EUV} = 250 \cdot 0.0046 \cdot 0.5 = 0.575 \text{ W} \quad (7.46)$$

The scan slit has a width of 26 mm and according equation (2.1) scan speed becomes:

$$v_{scan} = \frac{0.575/0.026}{80} = 0.276 \text{ m s}^{-1} \quad (7.47)$$

One field consists four nodes in  $y$  direction. Therefore, the exposure duration per node is:

$$t_{exp} = \frac{f_h/4}{v_{scan}} = \frac{33 \cdot 10^{-3}/4}{0.276} = 0.0298 \text{ s} \quad (7.48)$$

where  $f_h$  is the height of one field. The total heat energy received by the exposed slice of the field is:

$$q_{in} = P_{EUV} \cdot t_{exp} = 1.7 \cdot 10^{-2} \text{ J} \quad (7.49)$$

The outgoing heat is:

$$q_{out} = \kappa_{sub} \cdot h_{sub} \cdot t_{exp} = 149 \cdot 0.775 \cdot 10^{-3} \cdot 0.0298 = 3.4 \cdot 10^{-3} \text{ J} \quad (7.50)$$

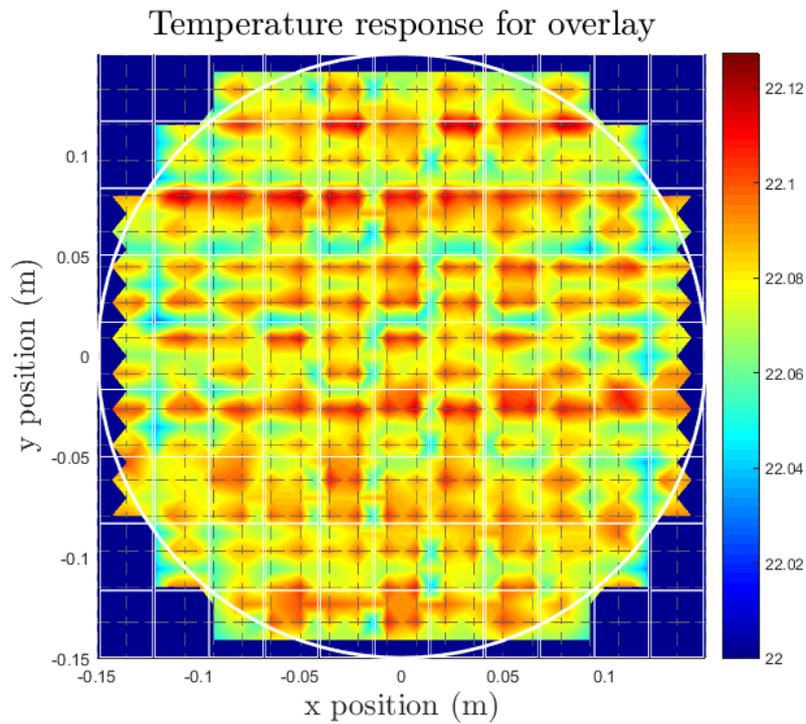
Here,  $\kappa_{\text{sub}}$  is the heat conductivity and  $h_{\text{sub}}$  the height of the substrate respectively. These values can be found in Table 7.3 and Table 7.1. By using the thermal heat capacitance, the increase in temperature can be calculated. First, the mass of exposed substrate is calculated:

$$m_{\text{sub}} = \rho_{\text{sub}} \cdot d \cdot b_0 \cdot h_{\text{sub}} = 2329 \cdot 4.7 \cdot 10^{-3} \cdot 26 \cdot 10^{-3} \cdot 0.775 \cdot 10^{-3} = 2.21 \cdot 10^{-4} \text{ kg} \quad (7.51)$$

Here,  $d$  and  $b_0$  are the depth and width of the scan slit respectively and  $\rho_{\text{sub}}$  is the density of the substrate. The temperature rise calculated according:

$$\Delta T = \frac{q_{\text{in}} - q_{\text{out}}}{c_p m_{\text{sub}}} = \frac{1.7 \cdot 10^{-2} - 3.4 \cdot 10^{-3}}{705 \cdot 2.21 \cdot 10^{-4}} = 0.09 \text{ K} \quad (7.52)$$

A nominal simulation with meander routing and with a time step of  $\Delta t = 0.01$  is performed. The temperature field used for overlay deformations is illustrated in Figure 7.13. It can be seen that the hand calculation is of the same order as the result of the simulation.



**Figure 7.13:** Temperature response for overlay of meandering routing. Colours indicate magnitude of temperature.



# 8

## Analysis of Simplified Model

This chapter is devoted to the analysis of the simplified thermo-mechanical model and the performance of existing experiment designs. In Section 8.1, a thorough sensitivity analysis is performed on the simplified thermo-mechanical model. Sensitivity analyses give useful insight into the behaviour of the model, but are also necessary in order to build statistical information about model parameters and, therefore, are required for experiment design. Next, in Section 8.2 the non-linear behaviour of the simplified model is investigated. In Section 8.3 the three standard routings that are currently used for experiments are described. Thereafter in Section 8.4, a least squares estimation is performed in order to obtain the model parameter subset estimates. In Section 8.5, the routings are evaluated on their performance by determining statistical information about the model parameters. Finally in Section 8.6, sensitivity information of the experiments – or so-called fingerprints – is interpreted in order to gain a deeper understanding where the performance difference of the experiments come from.

### 8.1 Sensitivity Analysis

In Chapter 3 it was shown that the direct sensitivity was the most applicable method for the simplified model in terms of accuracy and computational power. The direct sensitivity is developed by expressing the analytical derivative of the system and solving that expression. The sensitivity of interest is the overlay sensitivity with respect to the model parameters. Overlay can be calculated according equation (7.41) with  $\mathbf{f}$  the thermal force vector defined in equation (7.42). The simplified model consists of two parts, a thermal part and a mechanical part, where the thermal part, or thermal solution, is captured in this force vector  $\mathbf{f}$ .

**Thermal Part** In order to construct the direct sensitivity of overlay to model parameters, it is necessary to derive the thermal sensitivity first. From Section 7.3.1 the thermal equation of the model was described by:

$$\mathbf{M}_1 \mathbf{T}_f^{k+1} = \mathbf{M}_2 \mathbf{T}_f^k + \mathbf{M}_3 \mathbf{Q}_f^{k+1} + \mathbf{M}_4 \mathbf{T}_c^{k+1}$$

The direct sensitivity with respect to the model parameters  $\mathbf{p}$  gives the following expression,

$$\frac{d\mathbf{M}_1}{dp_m} \mathbf{T}_f^{k+1} + \mathbf{M}_1 \frac{d\mathbf{T}_f^{k+1}}{dp_m} = \frac{d\mathbf{M}_2}{dp_m} \mathbf{T}_f^k + \mathbf{M}_2 \frac{d\mathbf{T}_f^k}{dp_m} + \frac{d\mathbf{M}_3}{dp_m} \mathbf{Q}_f^{k+1} + \mathbf{M}_3 \frac{d\mathbf{Q}_f^{k+1}}{dp_m} + \frac{d\mathbf{M}_4}{dp_m} \mathbf{T}_c^{k+1} + \mathbf{M}_4 \frac{d\mathbf{T}_c^{k+1}}{dp_m} \quad (8.1)$$

where the sensitivity of  $\frac{d\mathbf{M}_3}{dp_m} = 0$  and the derivative of the constraint temperature vector  $\frac{d\mathbf{T}_c^{k+1}}{dp_m} = 0$ . This will lead to the final expression of the direct sensitivity for the thermal part as,

$$\boxed{\mathbf{M}_1 \frac{d\mathbf{T}_f^{k+1}}{dp_m} = -\frac{d\mathbf{M}_1}{dp_m} \mathbf{T}_f^{k+1} + \frac{d\mathbf{M}_2}{dp_m} \mathbf{T}_f^k + \mathbf{M}_2 \frac{d\mathbf{T}_f^k}{dp_m} + \mathbf{M}_3 \frac{d\mathbf{Q}_f^{k+1}}{dp_m} + \frac{d\mathbf{M}_4}{dp_m} \mathbf{T}_c^{k+1}}$$
(8.2)

with,

$$\begin{aligned} \frac{d\mathbf{M}_1}{dp_m} &= \frac{d\mathbf{C}_{p,ff}}{dp_m} + \frac{d\mathbf{K}_{c,ff}}{dp_m} \Delta t \\ \frac{d\mathbf{M}_2}{dp_m} &= \frac{d\mathbf{C}_{p,ff}}{dp_m} \\ \frac{d\mathbf{M}_4}{dp_m} &= -\frac{d\mathbf{K}_{c,fc}}{dp_m} \Delta t \end{aligned} \quad (8.3)$$

The analytical expressions of the system matrix sensitivities in equation (8.3) are not known, but can be computed using a *Finite Difference* (FD) scheme. In contrast to normal FD sensitivities, these are computationally efficient. The direct method is no longer purely analytical, but it is so-called *Semi-Analytical* (SA). It should be noted that a forward finite difference scheme is sufficient (see equation (8.4)), since the system matrices  $\mathbf{C}_p$  and  $\mathbf{K}_c$  are linearly dependent on the model parameters. This also implies that the semi-analytical method leads to exact sensitivities just as the purely analytical method, since no truncations are imposed.

$$\frac{d\mathbf{C}_{p,ff}}{dp_m} = \frac{\mathbf{C}_{p,ff}(p_m + \Delta p_m) - \mathbf{C}_{p,ff}(p_m)}{\Delta p_m} \quad (8.4)$$

**Mechanical Part** The mechanical solution is used to obtain the sensitivity of the overlay with respect to the model parameters. The thermal force vector is used to combine the thermal solution with the mechanical solution. For convenience, the thermal force vector  $\mathbf{f}$  is repeated.

$$\mathbf{f} = \mathbf{L}_0 + \mathbf{L}_1 \mathbf{T}$$

The direct sensitivity method for the mechanical equation of the system is expressed by:

$$\frac{d\mathbf{K}}{dp_m} \mathbf{u} + \mathbf{K} \frac{d\mathbf{u}}{dp_m} = \frac{d\mathbf{f}}{dp_m} \quad (8.5)$$

Which can be rewritten in the form:

$$\frac{d\mathbf{u}}{dp_m} = \mathbf{K}^{-1} \left( \frac{d\mathbf{f}}{dp_m} - \frac{d\mathbf{K}}{dp_m} \mathbf{u} \right) \quad (8.6)$$

The thermal force vector sensitivity is given by,

$$\frac{d\mathbf{f}}{dp_m} = \frac{d\mathbf{L}_0}{dp_m} + \frac{d\mathbf{L}_1}{dp_m} \mathbf{T} + \mathbf{L}_1 \frac{d\mathbf{T}}{dp_m} \quad (8.7)$$

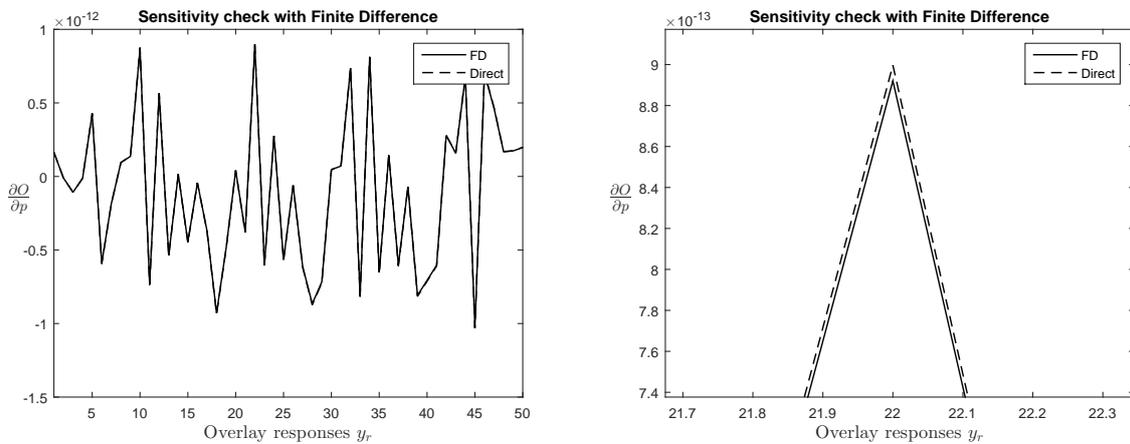
and therefore the overlay sensitivity is expressed by:

$$\boxed{\frac{d\mathbf{u}}{dp_m} = \mathbf{K}^{-1} \left( \frac{d\mathbf{L}_0}{dp_m} + \frac{d\mathbf{L}_1}{dp_m} \mathbf{T} + \mathbf{L}_1 \frac{d\mathbf{T}}{dp_m} - \frac{d\mathbf{K}}{dp_m} \mathbf{u} \right)} \quad (8.8)$$

### 8.1.1 Validation

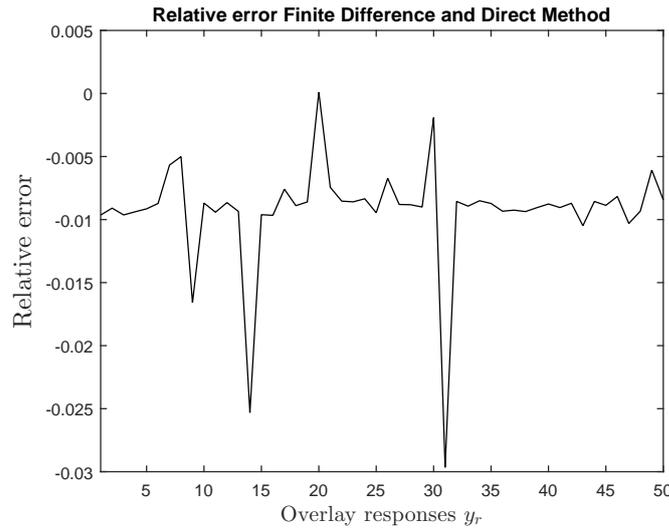
The derived sensitivity equations must be validated and this can be achieved by using the finite difference (FD) sensitivity as a reference. The finite difference method perturbs the model parameters by a certain amount and performs a complete new simulation. Subsequently, the FD sensitivity is calculated in the same manner as equation (8.4), but overlay  $\mathbf{O}$  is used instead.

A nominal simulation is performed as is illustrated in Figure 7.12. In Figure 8.1 the overlay sensitivity response with respect to model parameter  $p_1$  is illustrated for the finite difference method and the direct method. A 1% perturbation of the model parameter is used. It can be seen that the shape of the direct method is equivalent to the shape of the finite difference method. It is tempting to say that the implementation of the direct sensitivity method is correct, however, the remaining error must still be evaluated. Note, although only one model parameter is illustrated in the figure, all the remaining model parameter sensitivities show similar results.



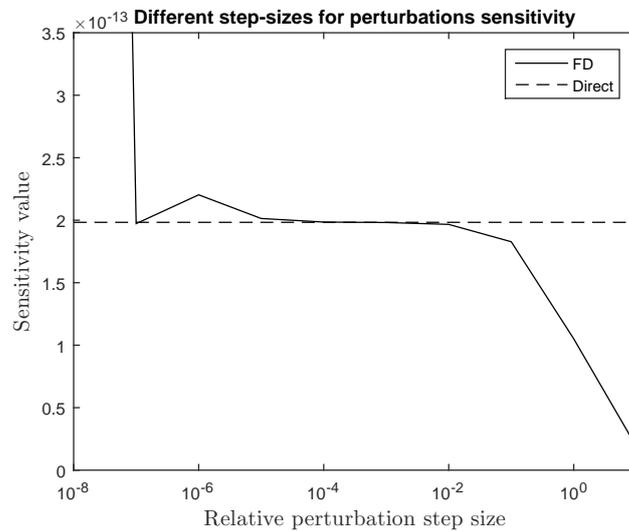
**Figure 8.1:** Overlay sensitivity response for  $p_1$  calculated with finite difference method and direct method. On the left, a slice of the overlay sensitivity response is illustrated. On the right a close-up is made to show the difference.

As was mentioned, an error still remains. This error can be explained by the truncation error that is made when using the FD method. The truncation error of the FD method is given by  $\mathcal{O}(\Delta p)$ , and since a perturbation of 1% is used, a truncation error of order  $10^{-2}$  is expected. It is known that the direct sensitivity leads to the exact solution and from Figure 8.2 it can be noticed that the relative error between both methods is of order  $10^{-2}$ . Hence, it can be concluded that the implementation of the direct method is reliable.



**Figure 8.2:** Relative error between Finite Difference method and Direct method.

In order to clarify the difference between the finite difference method and the direct method and their accuracy, multiple step-sizes are used as model perturbations. For one overlay response  $y_r$  the result is shown in Figure 8.3. From the figure, it can be seen that the ideal (relative) perturbation step-size for the FD method lies between  $10^{-4}$  and  $10^{-2}$ . In this region, the response remains almost constant. When a step-size smaller than the ideal region is chosen, the response becomes more noisy and will end up in an unstable response. This can be appointed as numerical noise and is known as the condition error. The condition error is indicated by the growing response on the left in the figure. When a step-size larger than the ideal region is chosen, the derivative becomes unreliable due to truncation. This is indicated by the decreasing response on the right.



**Figure 8.3:** Different perturbation step-sizes for Finite Difference method and Direct method.

It is noticeable that the response of the direct method remains constant and is not influenced by the different relative step-sizes. Although it cannot be seen, the direct method can be influenced by condition errors, since condition errors occur due to computational limitations. The advantage of the direct method is that it is linear in its response. Hence, accuracy remains, even when perturbations are off proportional. Note, this advantage only holds since  $\mathbf{C}_p$  and

$\mathbf{K}_c$  are linear in the model parameters.

### 8.1.2 Model Parameter Subset

When the simplified model is subjected to experiment design, a well chosen subset of the model parameters  $\mathbf{p}^*$  is selected to improve their statistical accuracy. A subset of model parameters is useful to reduce the complexity of finding an optimal experiment. It will reduce the complexity of finding least square estimates as well. However, the parameter subset cannot be chosen arbitrary, it must consist of parameters that have a significant impact on overlay performance but also in balance with the currently available uncertainty information. In other words, a model parameter with moderate impact on overlay, but with a currently uncertain parameter value, might be chosen above a parameter that has a significant impact on overlay, but with a currently high accuracy of its value.

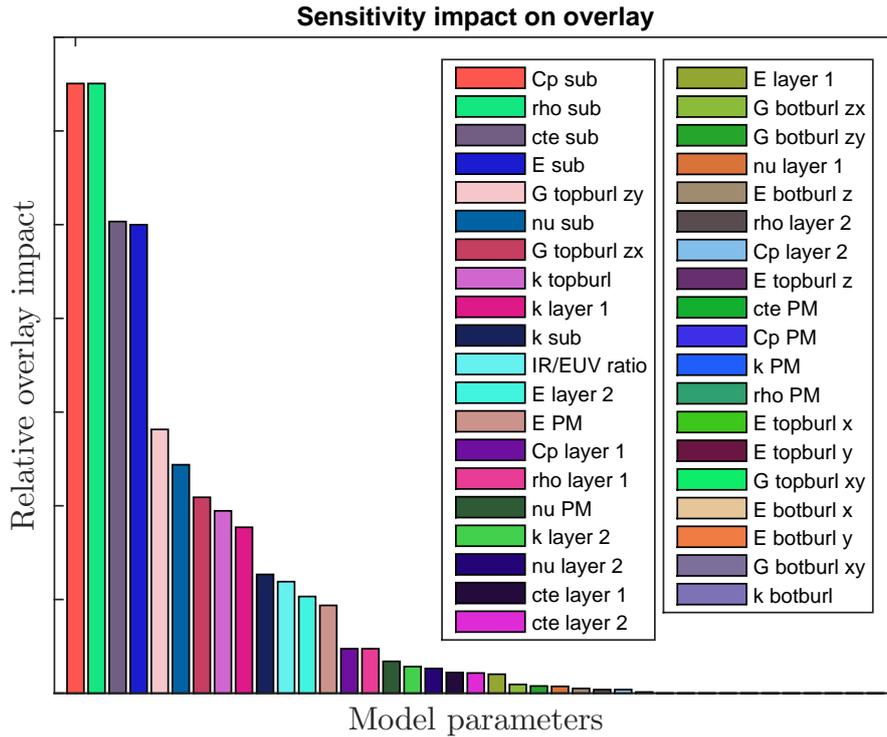
A sensitivity analysis is useful to investigate the impact on overlay per model parameter. Since the model parameters do not have the same unit, it is not possible to make a fair comparison on the overlay impact per parameter. Fortunately, a small adjustment can be made in order to obtain dimensionless overlay sensitivities. These kind of sensitivities are known as *logarithmic sensitivities* and are calculated according:

$$\frac{d(\log \mathbf{O})}{d(\log p_m)} = \frac{\frac{1}{\mathbf{O}} d\mathbf{O}}{\frac{1}{p_m} dp_m} = \frac{p_m}{\mathbf{O}} \frac{d\mathbf{O}}{dp_m} \quad (8.9)$$

The logarithmic sensitivities indicate the relative "strength" of the influence between the model parameters on overlay. The logarithmic sensitivities are calculated for all model parameters and the absolute mean value per parameter sensitivity response is calculated. This mean value allows for relative comparison on impact on overlay  $\text{IO}(p_m)$  between the different model parameters.

$$\text{IO}(p_m) = \text{mean} \left( \left| \frac{d(\log \mathbf{O})}{d(\log p_m)} \right| \right) \quad (8.10)$$

The result of the relative impact on overlay per model parameter is graphically represented in Figure 8.4.



**Figure 8.4:** Relative impact on overlay per model parameter, obtained by calculating the logarithmic sensitivities.

From the figure, it can be seen that the first four highest mean impact values are associated with model parameters of the substrate. This result is not surprising, since the heat load is directly applied to the substrate. The heat capacitance and density of the substrate show the same impact on overlay. This can be appointed to the relationship of both model parameters in equation (7.1). As a consequence, these model parameters are completely correlated. The same behaviour can be seen for the heat capacitance and density of layer 1 and layer 2, but their impact on overlay is minor. In the preliminary calibration investigation, mechanical model parameters were neglected and, as a result, not included in the model parameter subset. However, from Figure 8.4 it becomes clear that mechanical model parameters have significant impact on overlay. Note that the impact on overlay for, for instance, the stiffness in  $x$  direction of the top burls is zero. The stiffness in  $x$  direction is neglected and captured in the shear stresses, as was explained in Section 7.2.2. It can be observed that the tangential stiffness of the top burls in  $yz$  has more impact on overlay than the tangential stiffness of the top burls in  $xz$ . This may be surprising, since the model is symmetric, but this effect can be explained by the fact that the fields are exposed by a scanning motion in  $y$  direction. Therefore, perturbation of  $G_{zy}$  will have more influence on overlay than perturbation of  $G_{zx}$ .

The tangential stiffness of the top burls is a mechanical model parameter with moderate to high impact on overlay and is currently rather uncertain. Therefore, this model parameter is useful to investigate during experiment design.

The IR/EUV ratio is a thermal model parameter and has moderate impact on overlay, but is currently quite uncertain. Therefore, the impact on overlay performance is seriously affected and its statistical variance must be improved during experiment design.

At last the heat capacitance of the substrate is included. This model parameter is a thermal parameter as well, has a very high impact on overlay performance, but its value is already

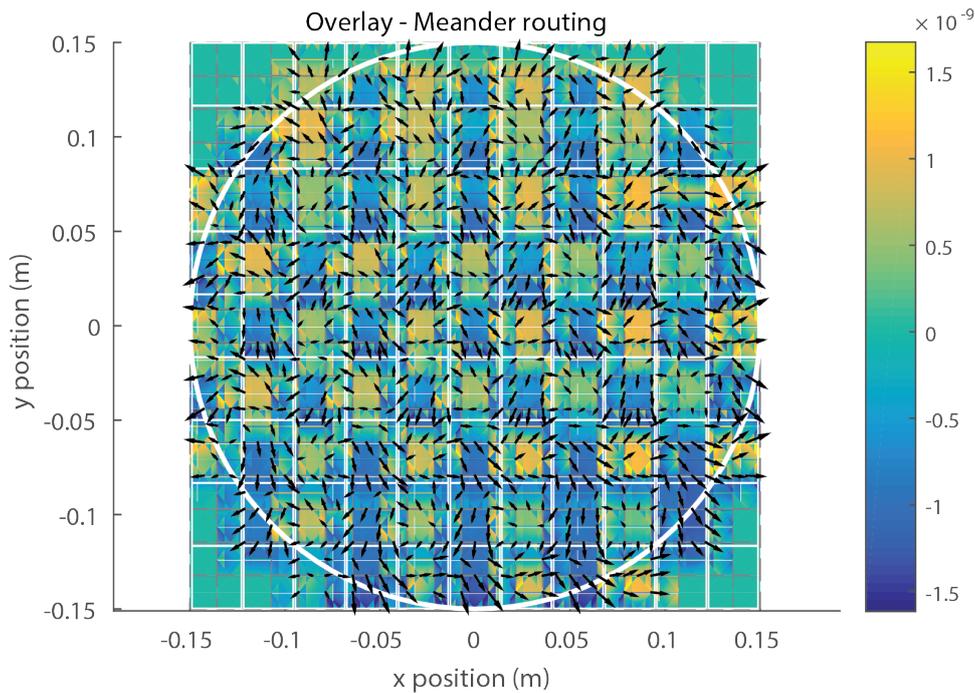
known with high accuracy. Still this value is included in the parameter subset to investigate experiment design with two model parameters that may look alike in their response (this will be seen later on). A list of the model parameter subset  $\mathbf{p}^*$  can be seen in Table 8.1.

**Table 8.1:** Overview of model parameter subset  $\mathbf{p}^*$  subjected to experiment design.

NO.	MODEL PARAMETER SUBSET
1	Heat capacitance $c_p$ of the substrate.
2	Tangential stiffness $G_{zy}$ of the top burls.
3	The IR/EUV ratio.

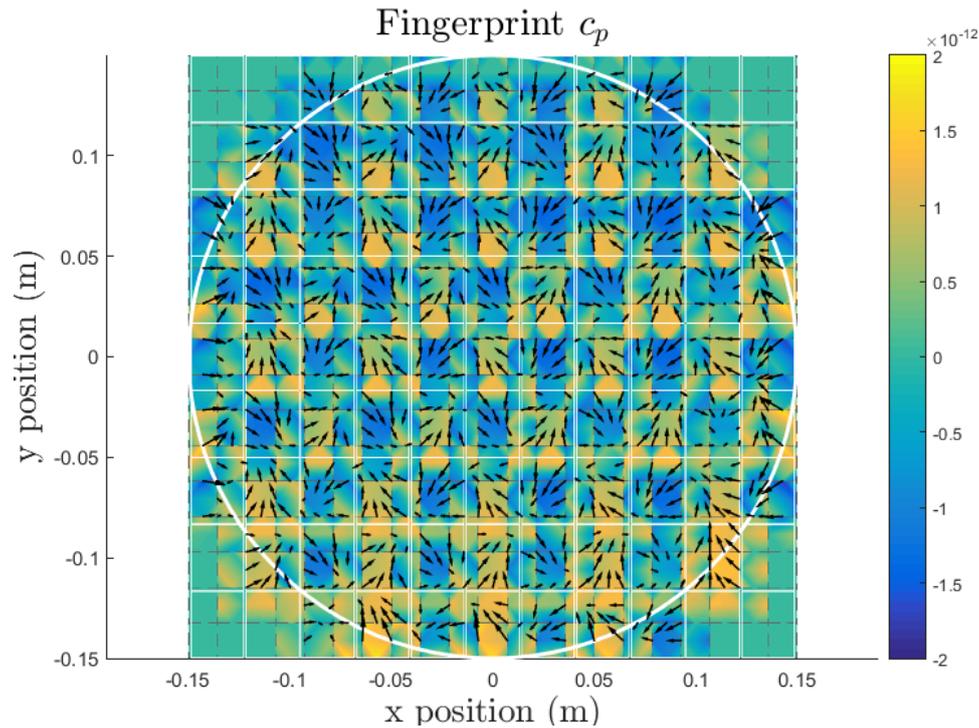
### 8.1.3 Sensitivities of Model Parameter Subset

The overlay sensitivity of the model parameter subset can be made visible. These kind of images will improve the understanding of the parameter sensitivity information and how each parameter affects the overlay. For the nominal simulation, the overlay image is depicted in Figure 8.5. The arrows indicate the magnitude and direction of deformation. The colours indicate the direction of deformation in  $y$  direction.



**Figure 8.5:** Overlay response for meander routing.

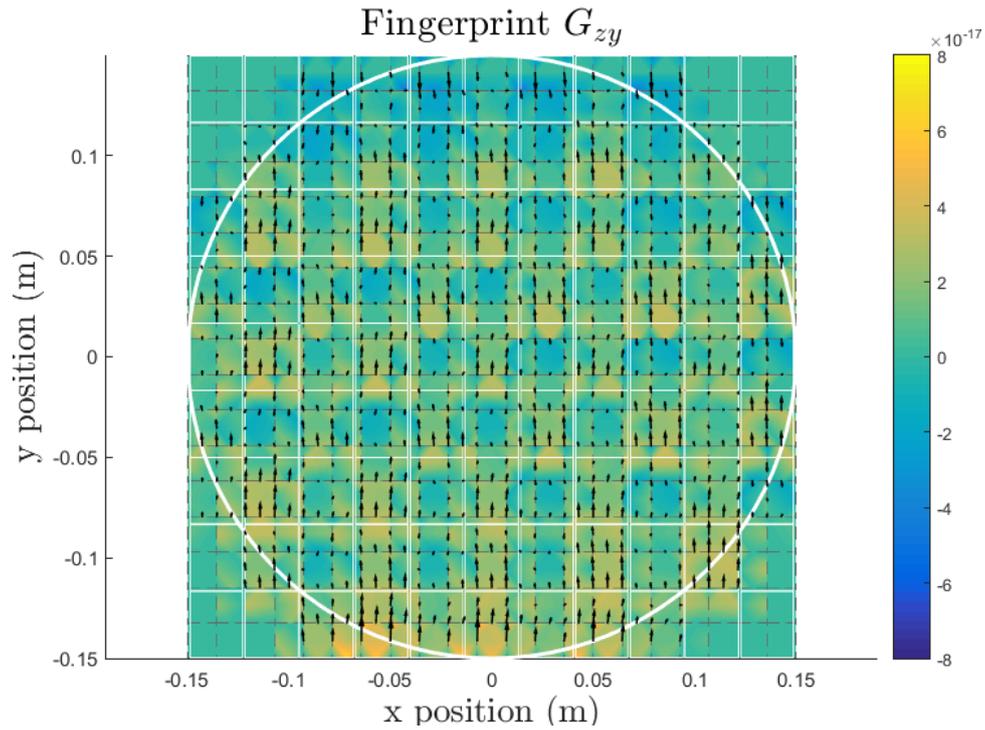
When the sensitivity for  $c_p$  is calculated, its direction can be plotted for all nodes. As before, the arrows indicate the direction and magnitude at the associated node and the colours indicate the sensitivity in  $y$  direction. In Figure 8.6 the sensitivity information, or so-called *fingerprint*, for  $c_p$  is illustrated.



**Figure 8.6:** Fingerprint of  $c_p$  for nominal experiment.

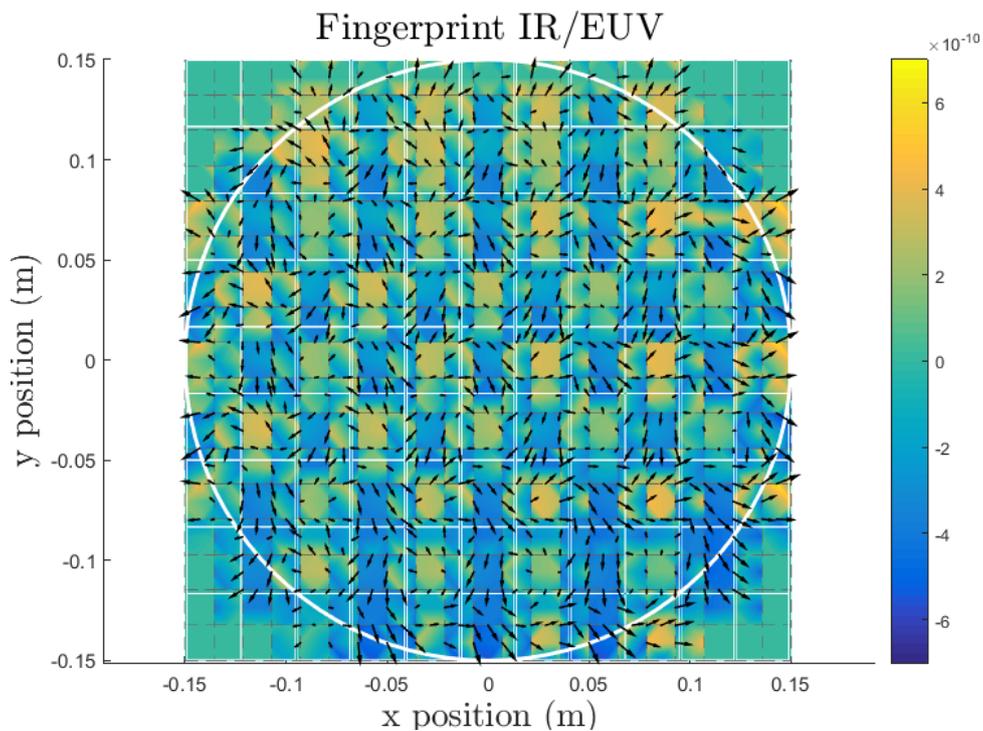
The direction of the fingerprint of  $c_p$  is (almost) a reversed copy of the overlay response. This is a logical consequence when understanding the characteristic of heat capacitance. When the heat capacitance is increased, more energy is required to rise the temperature. Subsequently, the overlay deformations will be reduced.

In Figure 8.7 the fingerprint of the tangential stiffness of the top burls in  $y$  direction is illustrated. It is noticeable that the sensitivity direction is mostly in  $y$  direction. This result is expected, since the tangential stiffness  $G_{zy}$  interacts at the deformations in  $y$  direction.



**Figure 8.7:** Fingerprint of  $G_{zy}$  for nominal experiment.

In Figure 8.8 the fingerprint of IR/EUV is illustrated. The IR/EUV ratio is a linear model parameter with the same amplification factor at each node. Therefore, the shape and directions are identical to the overlay response illustrated in Figure 8.5.



**Figure 8.8:** Fingerprint of IR/EUV for nominal experiment.

Additional information about the model parameters can be obtained when comparing the fin-

gerprints. The  $c_p$  and IR/EUV ratio are both thermal parameters and vary in the (reversed) direction of overlay. Therefore, it is hard to distinguish these parameters from each other, since one parameter could counterbalance the other parameter. The shape of the fingerprint of  $G_{zy}$  is easy to distinguish and unique. Therefore, it is expected that this parameter is less correlated with the  $c_p$  and IR/EUV.

**Correlation** The correlation between the model parameters  $\mathbf{p}^*$  can be determined according the scaled version of the variance-covariance matrix that was given in equation (4.5). The following correlation matrix is obtained:

$$\mathbf{C}(\mathbf{p}^*) = \begin{bmatrix} 1 & 0.47 & -0.95 \\ & 1 & -0.56 \\ & & 1 \end{bmatrix} \quad (8.11)$$

Here, the rows and columns indicate the model parameters  $c_p$ ,  $G_{zy}$  and IR/EUV ratio, respectively. It can be seen that the correlation between heat capacitance and the IR/EUV ratio is negative and has a high value of -0.95. The negative sign comes from the fact that an increase of  $c_p$  yields in less overlay deformation, whereas an increase of the IR/EUV ratio yields in more overlay deformation. The high correlation value is due to the characteristic behaviour of both thermal parameters. As was mentioned, the IR/EUV fingerprint has the exact shape of the overlay pattern and the fingerprint of  $c_p$  seems very similar as well (but in reversed direction). When calculating the inner product of both shapes (that is the calculation of  $\mathbf{C}$ ) the shape "vectors" are not orthogonal due to the similarity, hence the inner product results in a large projection.

The correlation of  $G_{zy}$  between  $c_p$  and IR/EUV ratio is less when compared between the thermal parameters. This result was expected when looking at the fingerprints. The fingerprint of the tangential stiffness is more unique and is more easily to distinguish and this concept is illustrated in Figure 6.3 as well. As with  $c_p$  and IR/EUV ratio, the correlation between  $G_{zy}$  and IR/EUV is negative also due to the same principle.

## 8.2 Model Characteristics and Behaviour

In this section the behaviour of the model, when perturbing the model parameter subset, is investigated. The parameter perturbation is chosen from 10% to 200% with respect to its nominal value. Subsequently, simulations are performed and the mean absolute overlay response  $\text{MA}(\mathbf{O})$  is stored. The nominal parameter values listed in Table 7.3 are used, together with the experimental set-up as was mentioned in Section 8.1.1, in order to perform the simulations.

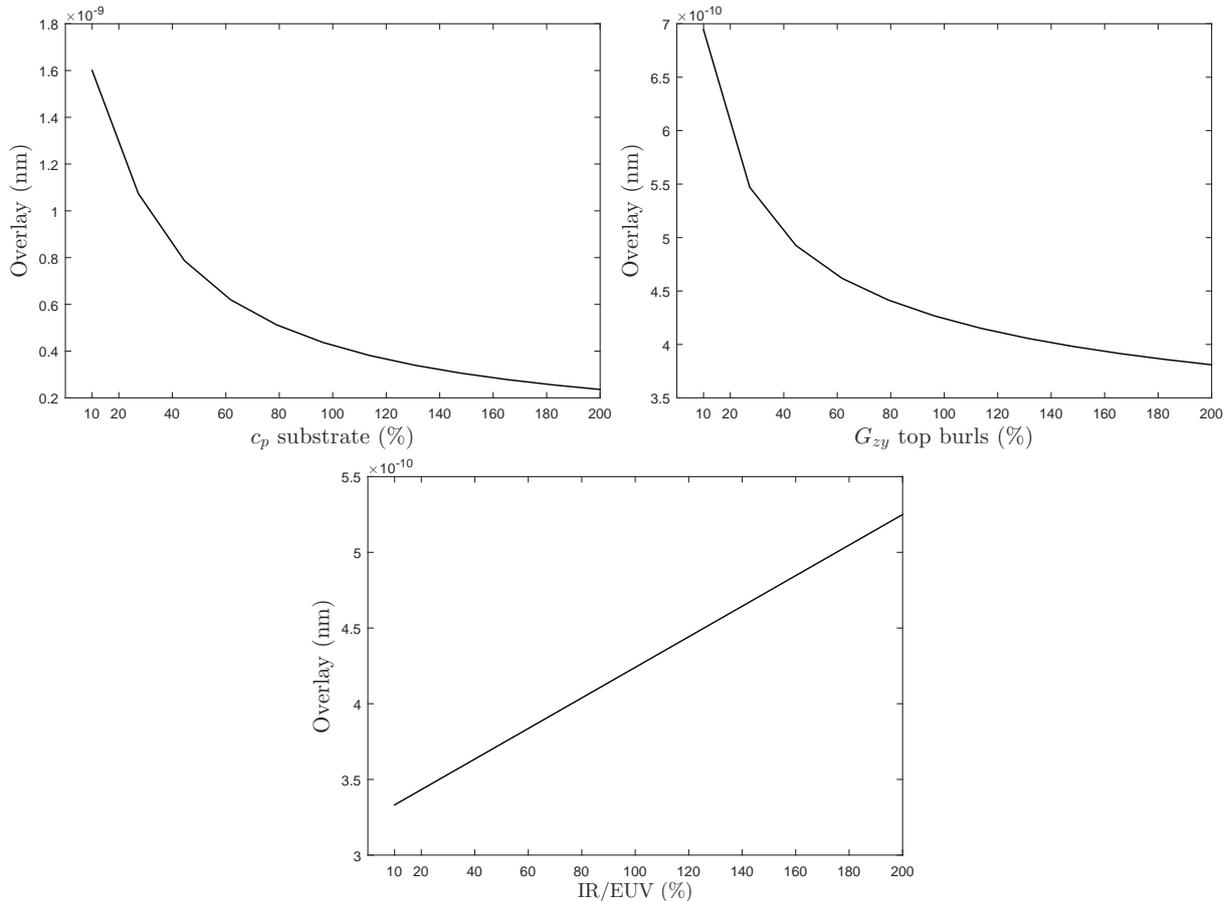
$$\text{MA}(\mathbf{O}) = \text{mean}(|\mathbf{O}|) \quad (8.12)$$

In Figure 8.9 the overlay response behaviour when perturbing the model parameters is shown. In the upper left figure the overlay response, associated with the perturbation of the heat capacitance  $c_p$  of the substrate, is illustrated. A monotonic decreasing response is visible and from the figure it is evident that  $c_p$  does not behave linear. This can be explained by observing the relation of  $c_p$  in equation (7.36). The heat capacitance is captured in the right hand side of the equation, but more important, it is captured in the left hand side as well. This implies an inverse of  $c_p$  when obtaining the thermal solution, hence the  $1/x$  behaviour. When the heat capacitance is increased, more energy is required to let the temperature rise. However, the energy remains constant and therefore the temperature decreases. This results in less overlay, which can be seen in the figure.

In the upper right figure, the tangential burl stiffness  $G_{zy}$  in  $y$  direction is shown. The same monotonic non-linear response can be seen. The tangential stiffness is captured in the stiffness

matrix  $\mathbf{K}$  and according to equation (7.41) an inverse of  $\mathbf{K}$  – and thus burl stiffness – is required to obtain the overlay deformations. The decreasing behaviour is logical as well, since increasing stiffness results in decreasing deformations.

The lower figure shows the characteristic behaviour of the IR/EUV ratio. The overlay response is an increasing straight line, which indicates linear behaviour. This linear behaviour can be explained by equation (7.28) as well. The IR/EUV ratio is a thermal model parameter linearly captured in  $\mathbf{Q}$  and appears on the right hand side of the equation. Therefore, the overlay response remains linear. The positive slope can be appointed to the fact that more energy results in increasing temperatures and, hence, results in more overlay.



**Figure 8.9:** Overlay response when perturbing the model parameter subset. The top left figure illustrates the overlay response of  $c_p$  substrate, the top right figure illustrates the overlay response for the tangential stiffness of the top burls in  $y$  direction (equivalent to tangential stiffness in  $x$  direction) and the lower figure represents the overlay response when perturbing the IR/EUV ratio.

### 8.2.1 Least Square

The selected model parameters do not behave linear in the system, except for IR/EUV. Therefore, iterative methods must be applied when solving the least square problem. It is expected that the non-linear least square problem can be solved using the Gauss-Newton algorithm, since the curved responses in Figure 8.9 are smooth with low order of non-linearity. Stability and convergence problems will mostly not arise.

A nominal overlay response  $\mathbf{O}(\mathbf{p})_{\text{ref}}$  is performed which acts as reference overlay measurement. A disturbance  $\boldsymbol{\eta}$  is added which represents measurement noise. The measurement noise is

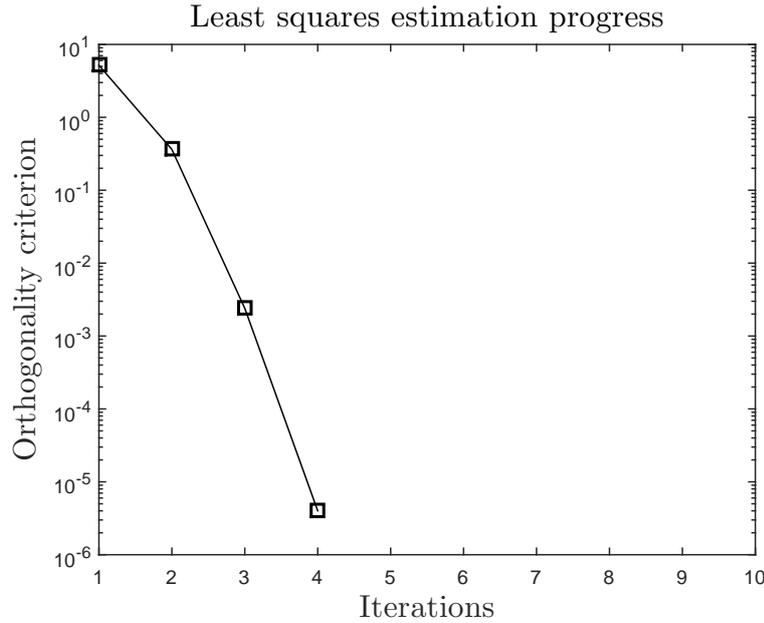
assumed normally distributed with mean zero and standard deviation of  $3\sigma = 0.5$  nm.

$$\mathbf{O}(\mathbf{p})_{\text{ref},\eta} = \mathbf{O}(\mathbf{p})_{\text{ref}} + \boldsymbol{\eta} \quad (8.13)$$

The least square estimator is found by minimizing the residual vector. The residual vector is given by:

$$S(\mathbf{p}) = \|\mathbf{O}_{\text{ref},\eta} - \mathbf{O}(\mathbf{p})\|^2 \quad (8.14)$$

The Gauss-Newton algorithm, together with the orthogonality criterion described in Section 5.2.1, is used for minimization of  $S(\mathbf{p})$  for all the model parameters in the subset.



**Figure 8.10:** Least square regression for a non-linear model. Experimental data error  $\mathbf{r}$  is minimized with respect to expectation plane.

In Figure 8.10 the progress of the least square estimation is shown. On the  $x$  axis, the number of iterations is plotted and on the  $y$  axis the orthogonality criterion is shown. Only four iterations are required for the orthogonality criterion to be satisfied. Even when  $c_p$  of the substrate and IR/EUV ratio are heavily correlated, the algorithm is not hampered and converges rapidly.

**Table 8.2:** Model parameter subset estimates after optimization of residual vector.

PARAMETER	TRUE	ESTIMATE	ERROR (%)
$c_p$	705	674.7	4.30
$G_{zy}$	$1.281 \cdot 10^7$	$1.284 \cdot 10^7$	0.26
IR/EUV ratio	0.625	0.563	10.0

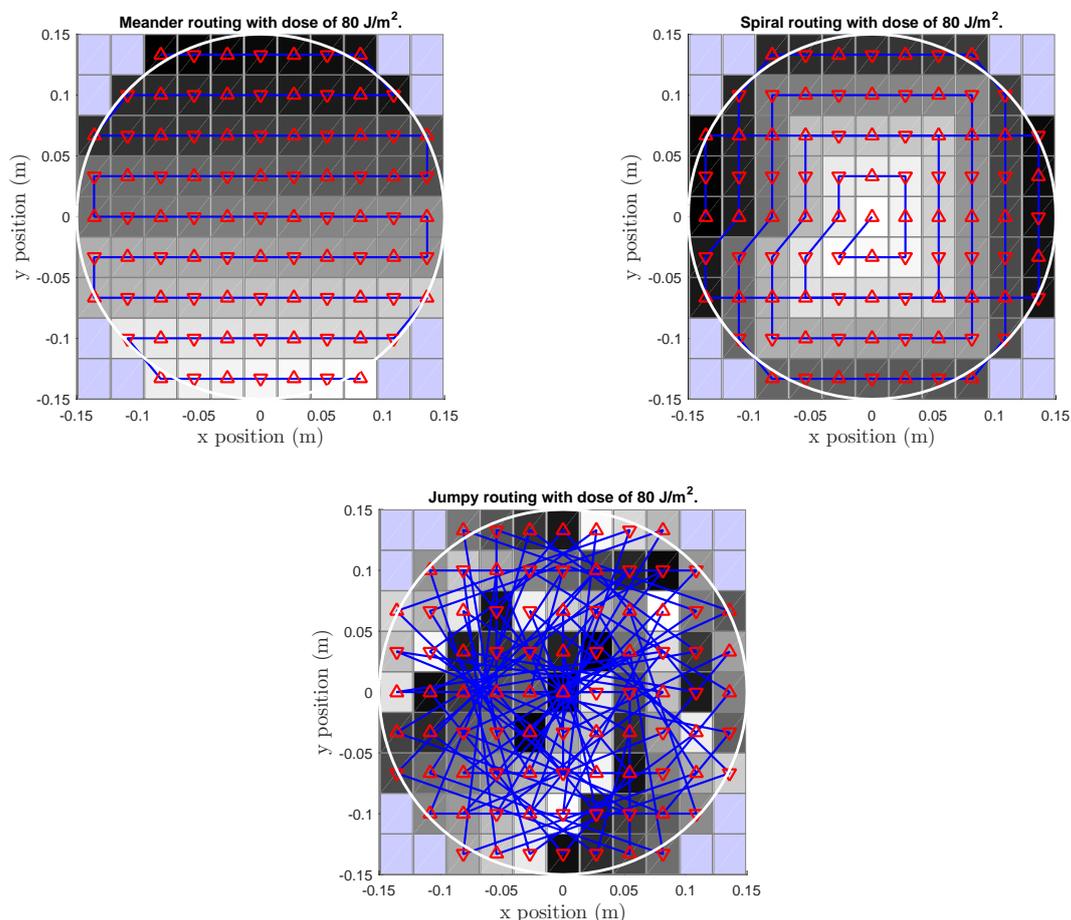
In Table 8.2 the estimates are listed. None of the model parameters are found at their true value, but the tangential stiffness comes close. The  $c_p$  and IR/EUV are the worst estimates with a relative error of 4.30% and 10.0%, respectively. IR/EUV is badly correlated with  $c_p$ , which hampers the ability of finding a good estimate. The IR/EUV estimate lies below its true value, which implies less heat input to the system. Therefore,  $c_p$  must decrease as well in order to keep the temperature level the same. Due to this thermal relation, the parameters are heavily correlated and hamper the least square estimation.

### 8.3 Standard Experiments

In order to improve overlay accuracy, standard experiments are conducted, which are subsequently used for model parameter estimation. In this section, the standard experiments are illustrated and the associated overlay results are discussed.

#### 8.3.1 Routings

There are three different type of experiments used within ASML. The experiments are called: meander routing, spiral routing, and jumpy routing. The meander routing was already explained in Chapter 7 as the nominal experiment, but is repeated for convenience. For all three experiments, the dose is constant during the exposure and has an intensity of  $80 \text{ J m}^{-2}$  and the scan direction is alternating between each field exposure. The meander routing involves of a meander scanning pattern along the fields of the substrate, the spiral routing involves a spiral exposure pattern along the fields, which starts at the left and the jumpy routing jumps in a random order from field to field during exposure. The routing sequence per experiment is illustrated in Figure 8.11.



**Figure 8.11:** Meander, spiral, and jumpy routing. The colours indicate the field order during exposure, with black indicating the start position and white indicating the end position. The red triangular markers indicate the scan direction per field and the fields are connected by the blue line.

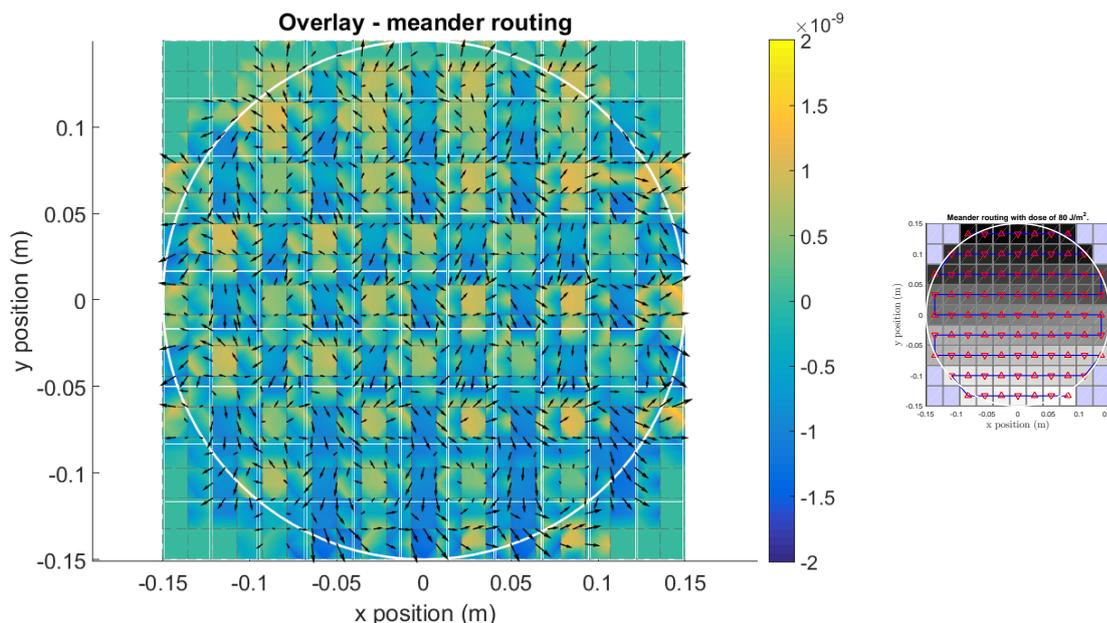
### 8.3.2 Overlay Results

In this subsection, the three standard experiments are used to perform overlay simulations. The nominal model parameter values listed in Table 7.2 and Table 7.3 are used and the overlay results are illustrated in the upcoming figures.

The overlay result for the meander routing was already illustrated in Section 8.1. From Figure 8.12 the meander pattern can be seen. The scan exposure begins at the top left, with a scan motion in positive  $y$  direction. The overlay deformation follows the scan direction, as is indicated by the arrows and the yellow colour. The first row is exposed from left to right and this effect can be seen by the arrows, which are tilted to the right. The second row is exposed from right to left as is illustrated in Figure 8.12, the arrows in the second row follow the direction of the scan motion in  $y$  direction, but they are also tilted to the left, which is the scan motion in  $x$  direction.

The spiral routing is illustrated in Figure 8.13. The same reasoning holds as for the meander overlay results. However, due to the chosen scanning field order and the associated scan motion in  $y$  direction, some fields are hardly deformed. This happens, for instance, when one field is scanned in positive  $y$  direction and right away the field above is scanned in negative  $y$  direction. This latter field is compensated by the lagging heat from the first illuminated field. This first field induces deformations in positive  $y$  direction in the second field as well, but the second illumination in the second field induces negative deformations. Hence the overlay result is approximately zero. For overlay performance this effect might be advantageous, but for parameter estimation this may not be desired. The response is almost zero and noise might become dominant.

The overlay result of the jumpy routing is illustrated in Figure 8.14. There are no significant differences in the magnitude of the deformations between the overlay results. However, the overlay shapes are different and this will influence the performance of model parameter estimation. It is difficult to draw conclusions from the overlay results only about their performance. Additional information is required, such as the fingerprints illustrated in Section 8.1.



**Figure 8.12:** Overlay simulation results with meander routing.

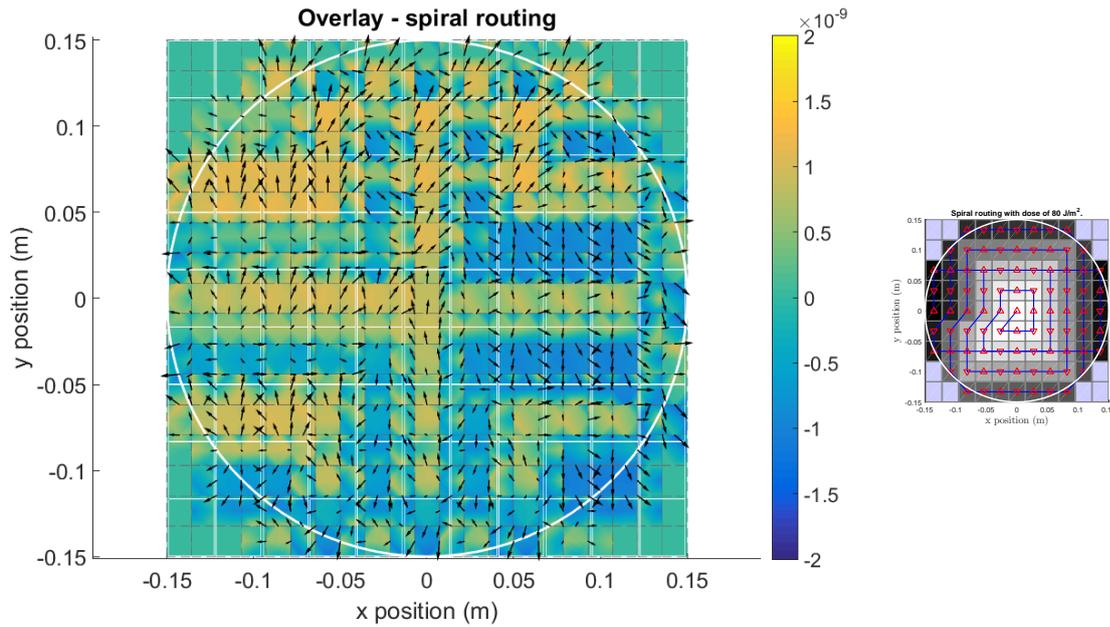


Figure 8.13: Overlay simulation results with spiral routing.

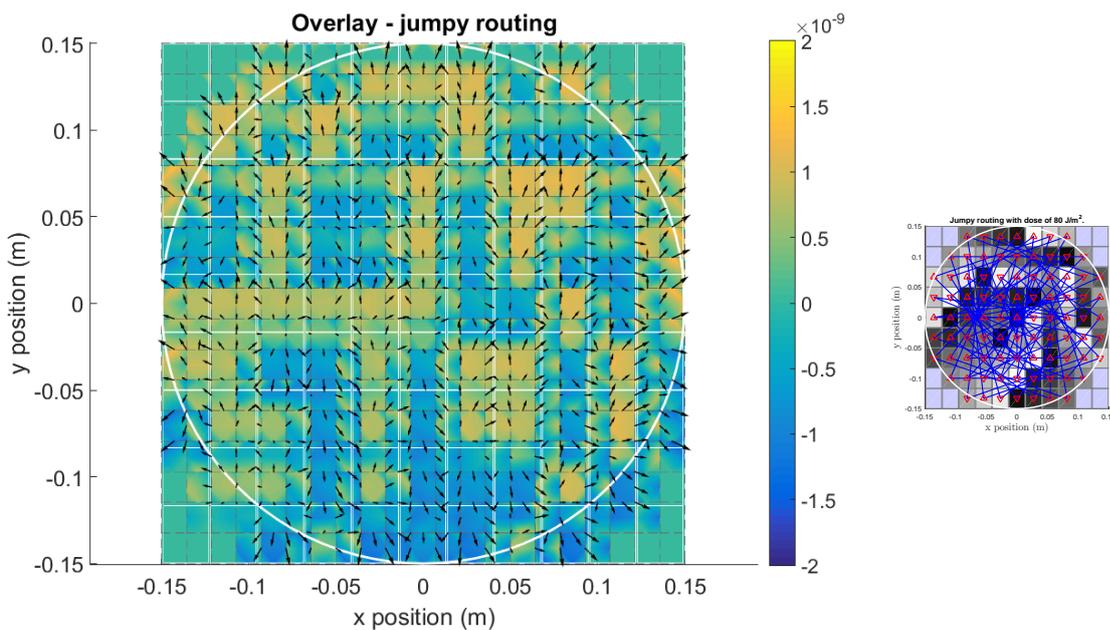


Figure 8.14: Overlay simulation results with jumpy routing.

## 8.4 Model Parameter Estimation

The three standard experiments are investigated for least squares estimation on the model parameter subset. Two cases are examined, where each case involves two model parameters. The IR/EUV ratio is an important model parameter, which is quite uncertain. Therefore, this model parameter is included in both cases. The cases are listed in Table 8.3.

**Table 8.3:** Model parameters included in case 1 and case 2 for experiment design.

PARAMETER	CASE 1	CASE 2
$c_p$	x	
$G_{zy}$		x
IR/EUV ratio	x	x

For each experiment, a simulation with nominal model parameters  $\mathbf{p}$  is performed. When the overlay result  $\mathbf{O}_{\text{ref}}$  is obtained, noise is added in order to represent an experimental measurement. Subsequently, the measured response is used for least squares estimation of the model parameters and is given by:

$$\mathbf{O}_{\text{ref},\eta} = \mathbf{O}_{\text{ref}} + \boldsymbol{\eta} \quad (8.15)$$

Here,  $\boldsymbol{\eta}$  represents the normally distributed noise with mean zero. The standard deviation is  $3\sigma = 0.5$  nm.

The heat capacitance of the substrate and the IR/EUV ratio in case 1 and the tangential burl stiffness and the IR/EUV ratio in case 2 are estimated from the meander, spiral and jumpy routing by means of least squares estimation. In Chapter 8 it was observed that the model is non-linear in the heat capacitance and tangential stiffness and linear in the IR/EUV ratio. Due to the non-linearity of  $c_p$  and  $G_{zy}$ , the Gauss-Newton iterative method with the orthogonality criterion is used for the least squares problem. The least squares problem is given by minimization of:

$$S(\hat{\mathbf{p}}) = \|\mathbf{O}_{\text{ref},\eta} - \mathbf{O}(\hat{\mathbf{p}})\|^2 \quad (8.16)$$

Least squares estimation is performed for case 1 and case 2 for all standard experiments. The initial values are set to  $\hat{\mathbf{p}} = 0.9\mathbf{p}$ . The least squares algorithm converges in 3 to 4 iterations, depending on the case and routing. The convergence rates are illustrated in the figures in appendix A.1. The algorithm converges fast, since the sensitivities are smooth functions with low order of non-linearity. Besides, the algorithm has no trouble finding a least squares solution for all the experiments, even when the model parameters in case 1 are highly correlated. This correlation can be seen in Section 8.1.3. In Table 8.4 the estimates for both cases and all experiments are listed.

The true values are not retrieved, as is expected. Although the error is shown, this does not indicate any statistical information about the model parameters and their uncertainty. For this, more regression analysis is required, but for now it is shown that the least squares estimation is not hampered by the choice of experiment or by the choice of the model parameter subset.

**Table 8.4:** Case 1 and 2. Model parameter estimation for meander routing.

Case 1					Case 2				
<b>MEANDER EXPERIMENT</b>									
ITER.	PAR.	TRUE	EST.	ERR. (%)	ITER.	PAR.	TRUE	EST.	ERR. (%)
4	$c_p$	705	735.6	4.4	4	$G_{zy}$	$1.28 \cdot 10^7$	$1.3 \cdot 10^7$	1.5
	IR/EUV	0.625	0.717	14.7		IR/EUV	0.625	0.629	0.7
<b>SPIRAL EXPERIMENT</b>									
ITER.	PAR.	TRUE	EST.	ERR. (%)	ITER.	PAR.	TRUE	EST.	ERR. (%)
3	$c_p$	705	679.2	3.7	3	$G_{zy}$	$1.28 \cdot 10^7$	$1.24 \cdot 10^7$	3.0
	IR/EUV	0.625	0.570	8.8		IR/EUV	0.625	0.640	2.4
<b>JUMPY EXPERIMENT</b>									
ITER.	PAR.	TRUE	EST.	ERR. (%)	ITER.	PAR.	TRUE	EST.	ERR. (%)
4	$c_p$	705	671.4	4.8	3	$G_{zy}$	$1.28 \cdot 10^7$	$1.11 \cdot 10^7$	13.4
	IR/EUV	0.625	0.456	27.1		IR/EUV	0.625	0.501	19.9

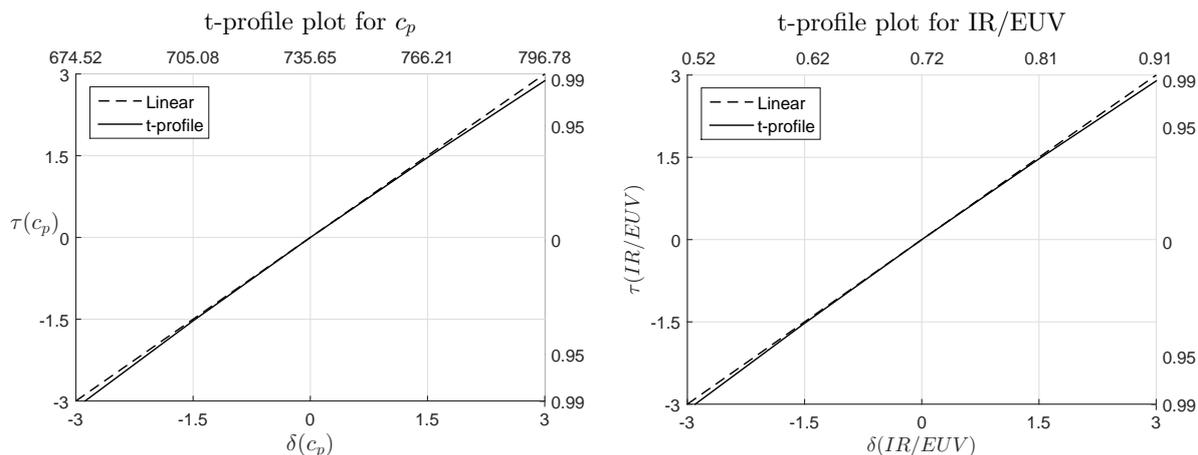
## 8.5 Statistical Investigation

The three standard experiments are subjected to least squares estimation. However, one estimation does not provide additional statistical data about the model parameters. It is known that the model behaves non-linear in its model parameters. Hence, profiling techniques are applied in order to investigate the severity of non-linearity. When this is within certain bounds, linear approximations are sufficient to obtain reliable statistical information about the model parameters.

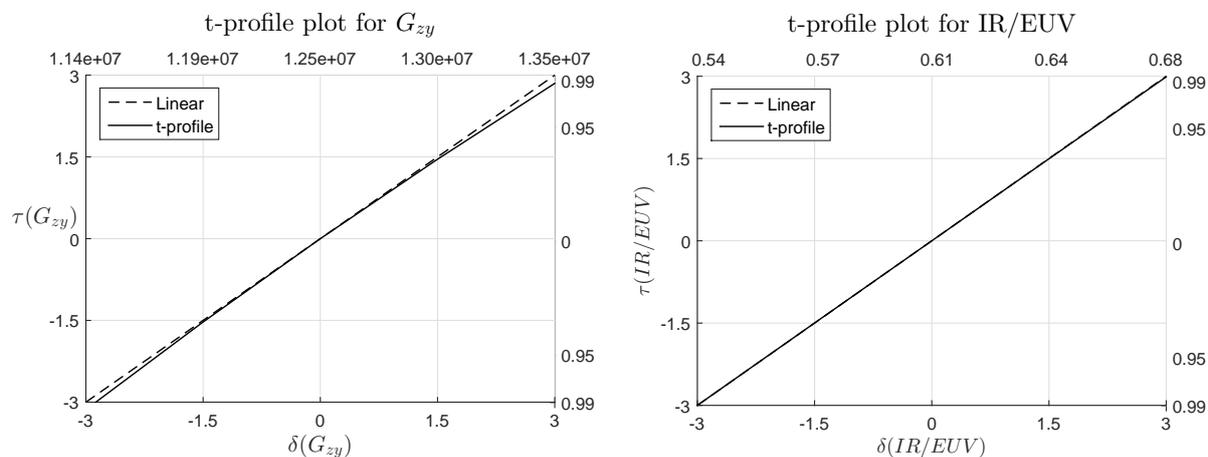
In the following subsections, the profiling techniques are applied to the meander, spiral, and jumpy experiment. At last, a comparison is made between the different experiments and their performance on the accuracy of estimation of the model parameters.

### 8.5.1 T-profiling Standard Experiments

The meander routing is subjected to t-profiling in order to investigate the severity of non-linearity in the confidence response for the model parameters in the associated cases. The results are shown and discussed only, since this technique is extensively described in Chapter 5.



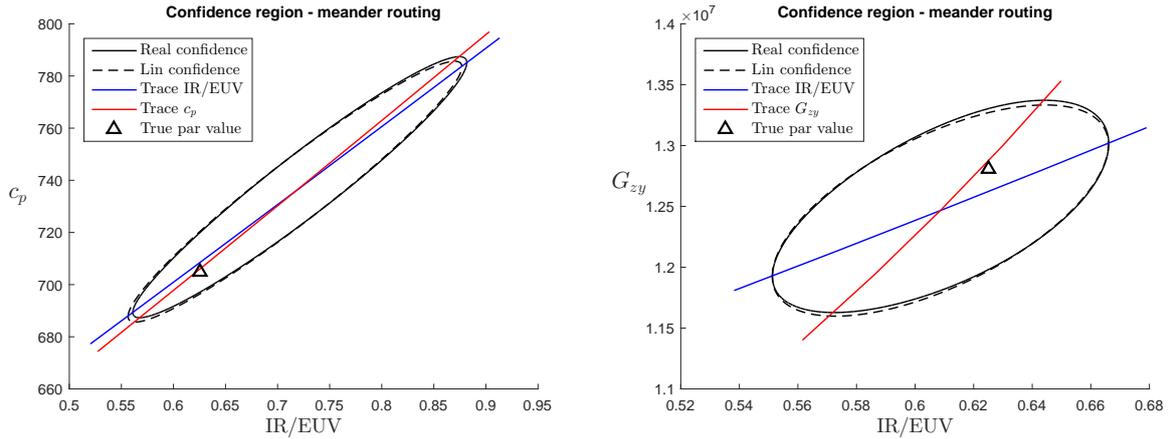
**Figure 8.15:** T-profiling meander routing for parameter case 1. The heat capacitance of the substrate and the IR/EUV ratio behave non-linear in their confidence response due to high correlation.



**Figure 8.16:** T-profiling meander routing for parameter case 2. The tangential burl stiffness behaves non-linear in its confidence response, whereas the IR/EUV ratio behaves almost linear in its confidence response due to low to moderate correlation.

In figures 8.15 and 8.16 the t-profiling is illustrated for the meander routing for model parameter case 1 and 2. From the figures, it is immediately evident that the confidence response behaves almost linear in the model parameters. Small deviations from the (dotted) reference line are visible, indicating the presence of non-linear behaviour. In Figure 8.15 it can be seen that both plots are similar. This can be explained by the fact that the model parameters of case 1 are highly correlated.

When observing Figure 8.16, two different t-profile plots are visible. These parameters are less correlated than the parameters in case 1. The tangential burl stiffness is a non-linear parameter, as was observed in Chapter 8 and the non-linearity can be confirmed by the deviating t-profile plot. The IR/EUV ratio is a linear model parameter with respect to overlay. Since the correlation is low to moderate, this linear behaviour is transferred to the confidence t-profile response, as can be seen in the right figure. Note, due to some correlation of the latter t-profile response, it is not exactly linear but this is hardly visible in the figure.



**Figure 8.17:** Confidence regions of 95% for meander routing for both cases. The linearly approximated confidence regions are sufficient to represent the true confidence regions.

In Figure 8.17, the true 95% confidence region, linearly approximated 95% confidence region and the profile trace vectors are illustrated for the meander routing. The linearly approximated confidence regions are obtained by linearising the model around  $\hat{\mathbf{p}}$ , see equation (5.41). The true confidence regions are sufficient approximated by the linearly confidence regions, as is expected when observing the t-profiling results. For case 1, the profile trace vectors intersect at a sharp angle and they almost coincide. This phenomenon indicates correlation between the model parameters and results in an elongated ellipsoidal confidence region.

The profile trace vectors for parameter case 2 (right figure) indicate less correlation. As a result, the confidence region is less elongated, which results in a smaller confidence region. When comparing both ellipsoids, it can be observed that the confidence bound for IR/EUV is smaller when its value is estimated together with the tangential burl stiffness.

Finally, the true nominal parameter values, indicated by the triangle, lie within both confidence regions.

The same reasoning holds for the t-profiling results of the spiral and jumpy experiment. The associated t-profiling responses and confidences regions can be found in appendix A.3.

### 8.5.2 Routing Performance

In the previous subsection, it was seen that the true confidence regions can be approximated by the linearised confidence region for all 3 experiments. For both parameter cases, these 95% confidence regions are illustrated in Figure 8.18. It should be noted that the associated parameter values are not shown, in order to compare the confidence region per experiment at the point of the estimates.



the sensitivities – or so-called fingerprints – of the model parameter subset are evaluated for each experiment in the following subsections.

In Section 8.1.3, the shapes and the origin of correlation between model parameters is explained and is not repeated in the following subsections, since it is analogous for the spiral and jumpy fingerprints.

### 8.6.1 Meander Routing

The sensitivity information for the meander routing was already obtained in Section 8.1.3. The fingerprints for  $c_p$ ,  $G_{zy}$  and IR/EUV ratio are illustrated in the figures 8.6, 8.7 and 8.8, respectively.

Two essential ingredients are required to obtain a good performing experiment: little correlation between model parameters and a high signal to noise ratio. The correlation matrix in equation (8.17) reveals a correlation, for model parameter case 1, that lies between the correlation value of the remaining experiments. In Figure 8.18 it can be seen that the confidence region of the meander routing lies between the spiral and jumpy routing as well.

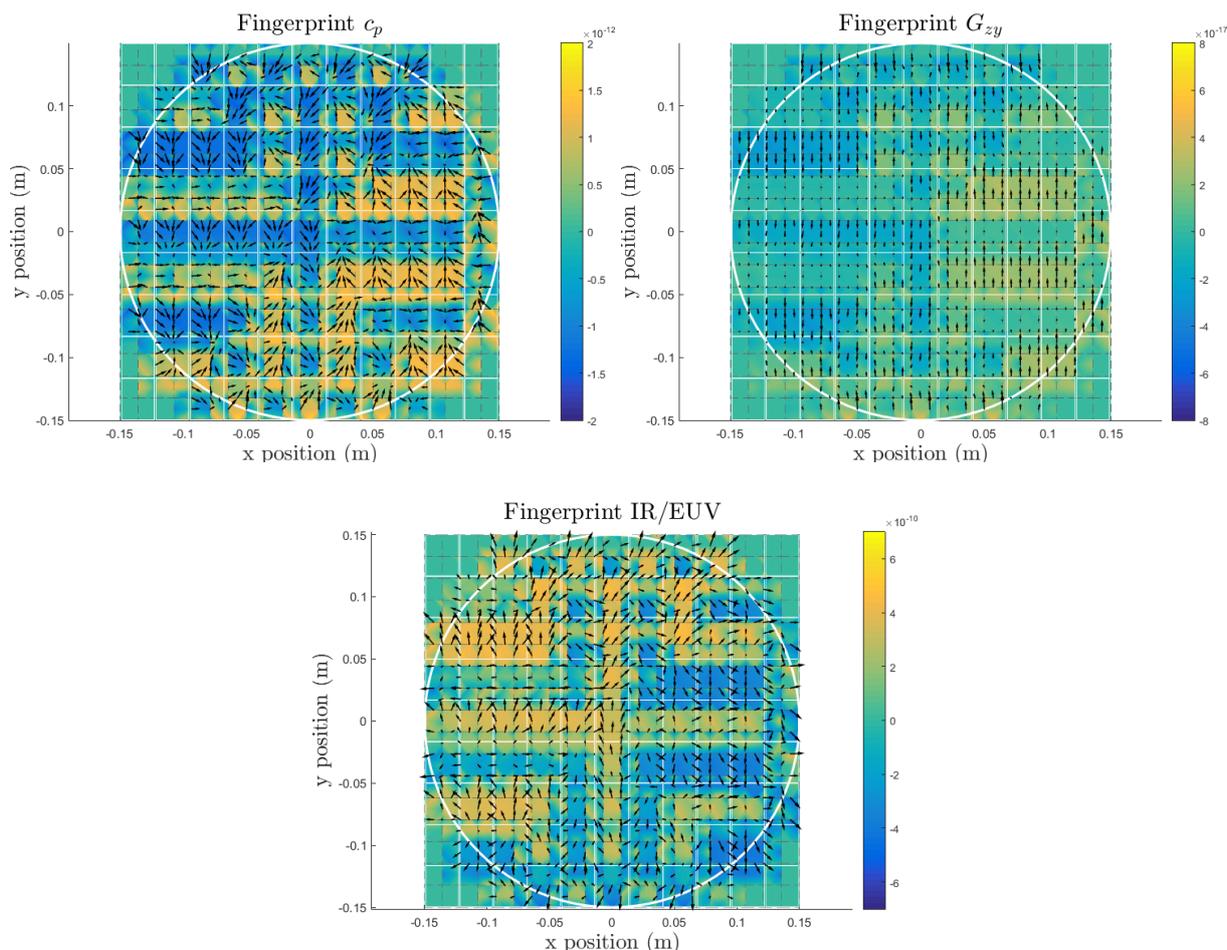
For model parameter case 2 the meander routing is the best performing, since the correlation for the meander routing is the lowest and the meander routing has the smallest confidence region according to Figure 8.18. From the fingerprints, it is obvious to conclude that the reduction of correlation is obtained by the horizontal scanning motion. Horizontal scanning motion introduces more overlay deformations in the horizontal direction compared to the remaining experiments. Therefore, the IR/EUV ratio sensitivity is more pointed in the horizontal direction as well, but the tangential stiffness sensitivity remains in the vertical direction. This makes the fingerprint more orthogonal to each other which has a reduction in correlation as a consequence. There are no indications that the signal to noise ratio is influencing the performance compared to the rest. Hence, the correlation is decisive for model parameter accuracy.

### 8.6.2 Spiral Routing

The spiral routing is performing worse in both model parameter cases, as can be seen in Figure 8.18. The fingerprints are calculated and illustrated in Figure 8.19.

Two essential ingredients are required to obtain a good performing experiment: little correlation between model parameters and a high signal to noise ratio. The correlation matrix in equation (8.17) indicates high correlation for both model parameter cases. Correlation between parameters results in a tilted ellipsoid which is elongated. This results in an increase of the confidence region and therefore a decrease in model parameter accuracy.

From the fingerprint of  $G_{zy}$ , an undesired effect can be seen. Due to the spiral routing and the associated scan motions, some overlay responses are nullified, as is mentioned in Section 8.3.2. Therefore, gaps appear where the sensitivity information in  $G_{zy}$  is hardly present. This nullification reduces the signal to noise ratio, hence resulting in more model parameter uncertainty.



**Figure 8.19:** Fingerprints of  $c_p$ ,  $G_{zy}$ , and IR/EUV for spiral routing.

### 8.6.3 Jumpy Routing

The jumpy routing performs the best in model parameter case 1 and performs the worst in model parameter case 2, as can be seen in Figure 8.18. The fingerprints are calculated and illustrated in Figure 8.20.

Two essential ingredients are required to obtain a good performing experiment: little correlation between model parameters and a high signal to noise ratio. The correlation matrix in equation (8.17) indicates the lowest correlation for model parameter case 1, but the highest correlation for model parameter case 2. From the fingerprints it can be observed that the signal to noise ratio is equivalent to that of the meander routing. Therefore, it can be concluded that the correlation determines the performance of the experiment in terms of the accuracy of the model parameters.

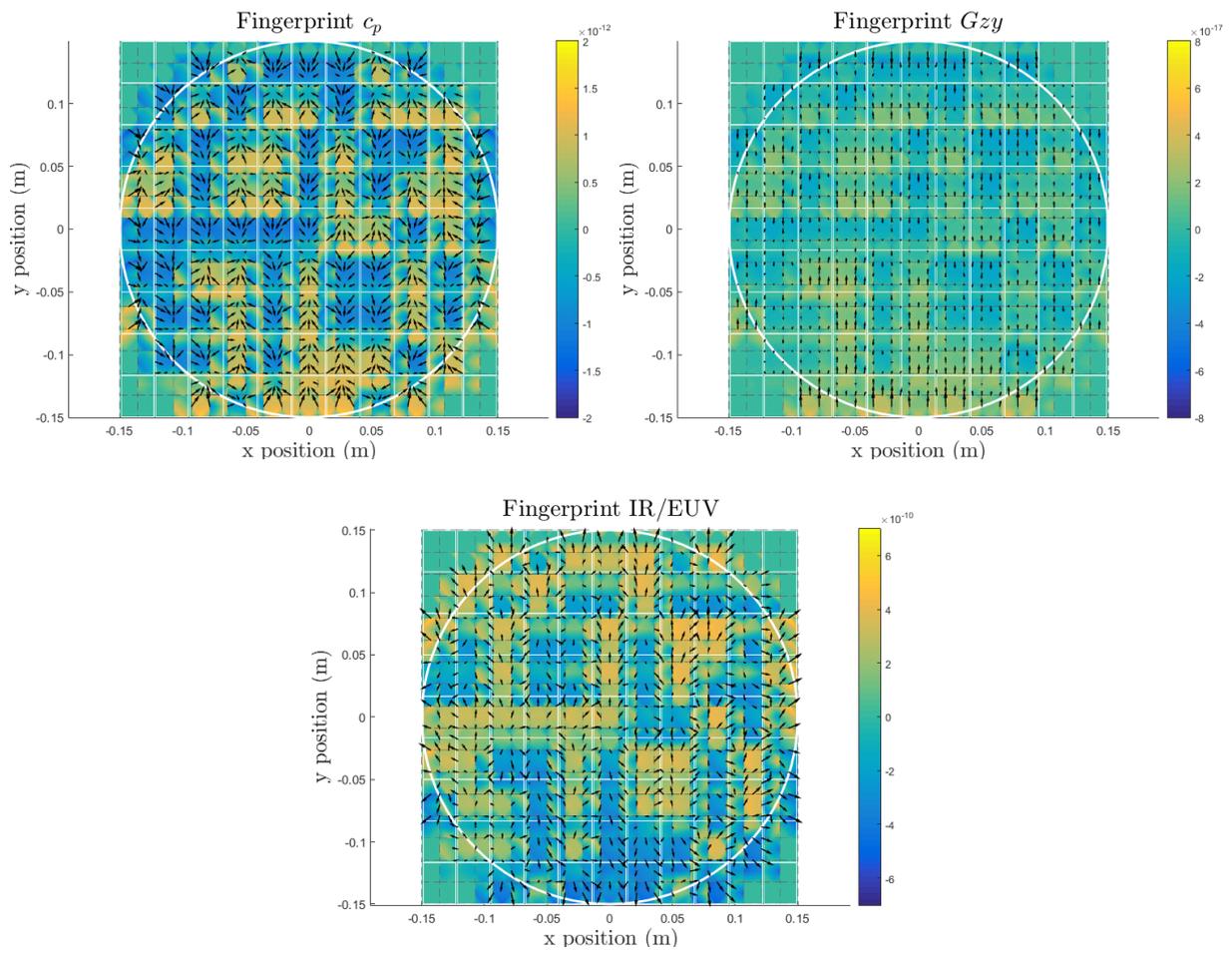


Figure 8.20: Fingerprints of  $c_p$ ,  $G_{zy}$ , and IR/EUV for jumpy routing.



# 9

## Optimal Experiment Design

This chapter is devoted to the optimization of an experiment design, which is subsequently conducted on the simplified thermo-mechanical feedforward model. The input parameters to be optimized are the dose, routing, and scan direction per field.

At first, the synthesis of the input parameters is investigated in Section 9.1 and, subsequently, in Section 9.2 experiment optimization for model parameter case of  $c_p$  and IR/EUV, and  $G_{zy}$  and IR/EUV is executed. The choice of model parameter per case can be found in Section 8.4. When the optimal experiments are obtained, a thorough analysis is performed in Section 9.3 to gain statistical information about the model parameter accuracy and to acquire a deeper understanding of the performance of the optimal experiment. In Section 9.4, the optimal experiments are compared against a brute force approach in order to investigate the efficiency of the used optimization algorithm. Finally in Section 9.5, the conversion of model parameter accuracy to overlay accuracy is given to see the importance of experiment design as a means to improve overlay performance.

### 9.1 Input Parameters for Experiment Design

The model input parameters are parameters that are used to provide the system required information before it can execute a simulation. One can think of model input parameters as the set-up of the environment of the system. The input parameters are stored in  $\varphi$ , which is referred to as the design vector. The design vector spans a space of possible experiments, which is called the design space  $\Phi$ , see Section 4.5. For the simplified model, the input parameters are dose, routing, and scan direction per field and are explained in Section 7.5.

It can be noted that the dose is a continuous input parameter which varies between a lower and an upper limit. These limits can be determined via equation (2.1), where the scan speed  $v_{\text{scan}}$  is the limiting factor. The dose is investigated separately, as will be seen in Section 9.2.2.

The routing and the associated scan direction per field are a different type of input parameters. These input parameters can have integer values only. The scan direction can only take values of 1 or -1, indicating upward- or downward scan direction, respectively. The routing is a combinatorial problem, since all fields must be addressed once during an exposure sequence, but cannot be exposed more than once. Standard gradient-based optimization algorithms are not applicable as it is difficult to obtain sensitivity information of the discrete integer combinatorial problem.

Since gradient-based optimization techniques are not applicable, direct optimization algorithms might be an alternative. This type of algorithm searches the surface of the objective function by evaluating it at a certain point, make a perturbation in the input parameters (which can be discrete), evaluate the objective function again. When the perturbation has resulted in

a reduction of the objective value, the algorithm accepts that point and repeats the process from that point until there are no further improvements. There exist several direct methods, which can make clever steps, but due to the non-convexity and the immense design space it is expected that this process is very inefficient.

A different solution approach might be a genetic algorithm. Although its working principles are not explained in detail, it works by using populations of various designs in the design space and introduce random genetic manipulation between these designs (Garcia, 1999). It is inspired by the evolution theory of Darwin. This algorithm is capable of handling discrete integer values and it might be possible to find a global optimum. However, due to the fact that it uses populations of various designs, makes it computationally very cumbersome in terms of storage.

At last, the simulated annealing algorithm can be used for the optimization of the experiment. Simulated annealing is a direct optimization method and interacts directly onto the objective function surface. In contrast to the other direct methods, this one includes randomness in the acceptance of a new search direction. Hence, it is theoretical possible to find a global optimum as is for the genetic algorithm. It is expected that convergence of the simulated annealing algorithm is in the same order as the previous direct methods, but the advantages is the ability to overcome the severe non-convex design space. Besides, this algorithm requires considerably less computational effort in contrast to the genetic algorithm.

In the following subsection, the working principle of the simulated annealing algorithm is explained in more detail and a description is given on how to use this algorithm on the simplified feedforward model.

### 9.1.1 Simulated Annealing

The Simulated Annealing (SA) algorithm is a heuristic method for obtaining good solutions to combinatorial optimisation problems. Since finding the best optimal experiment is impossible or impractical, this heuristic method can be used to speed up the process of finding a satisfactory solution.

The SA algorithm is inspired by the physical annealing process in materials. Physical annealing refers to the process of finding low energy states of a solid by initially melting the substance and subsequently lowering the temperature at a slow rate. In this way, the atoms of the solid are restructured. If the cooling is not done slowly, the resulting solid will not attain the ground state, but will be frozen into a metastable locally optimal structure.

In comparison with the optimisation algorithm, the different states of the solid correspond to the different feasible solutions to the combinatorial optimisation problem of the routing. The energy of the system corresponds to the objective function to be minimised. These objective functions are defined in Section 4.5. During the optimisation, a temperature function  $T(t)$  is used to represent the cooling effect. At the beginning of the optimisation the temperature is high and worse objective values can be accepted according the acceptance function  $\exp(-\delta/T)$ . This allows considerable variation between experiment states and makes it possible to overcome local minima. The acceptance of worse objective values gradually decreases during cooling, as is evident from the acceptance function. As a result, the experiment state becomes more "frozen" and eventually a (sub)optimal solution is found.

In Eglese (1989) a pseudo-code of the SA algorithm is given, which is illustrated in algorithm 1. At first, an initial experiment state  $i$  is selected from the design space  $\Phi$ , which represents a feasible routing and scan direction setting. Next, an initial temperature  $T$  is chosen. This temperature is analogous to the temperature of the physical annealing. There are no descriptions for selecting a good initial temperature, but for a rule of thumb: chose  $T$  twice the value of  $\delta$  for which the difference in objective values is still acceptable. For example, if values of  $\delta = 2$  are still accepted as differences between objective values (see line 8 in algorithm 1), the initial temperature is set to  $T = 4$ . In this way, there is a 60% chance that the new experiment state

$j$  is accepted. When the initial temperature is selected, the algorithm starts with generating a new experiment state  $j$ , which is a neighbour of state  $i$ . A proper neighbourhood function is necessary and will be explained in Section 9.1.1.1. The new state  $j$  is used to compare the change in objective function with respect to state  $i$ . When the new state leads to a decrease of the objective function, the new state  $j$  is accepted. Otherwise, a uniform random number between 0 and 1 is drawn and is subjected to the acceptance function. When the random number is below the acceptance function, the new state is still accepted. The inner loop of algorithm 1 determines how many neighbourhood moves are to be attempted at each temperature, where  $N(t)$  determines the amount of attempts per temperature. The outer loop is responsible for the decrease in temperature. There exist various temperature  $T(t)$  functions (Romeijn and Smith, 1993; Szu and Hartley, 1987), but a standard function is given below:

$$T(t) = \alpha^t T \quad t = 1, 2, \dots \quad (9.1)$$

Here,  $T$  is the initial temperature and the parameter  $\alpha$  determines the cooling rate. According to Eglese (1989) typical values of  $\alpha$  used in practice lie between 0.8 and 0.99.

---

**Algorithm 1:** Simulated Annealing algorithm in pseudo-code

---

```

1 Select an initial state  $i \in \Phi$ ;
2 Select an initial temperature  $T > 0$ ;
3 Set temperature change counter  $t = 0$ ;
4 while stopping criterion false do
5     Set repetition counter  $n = 0$ ;
6     while  $n < N(t)$  do
7         Generate state  $j$ , a neighbour of  $i$ ;
8         Calculate  $\delta = \phi(j) - \phi(i)$ ;
9         if  $\delta < 0$  then
10            |  $i \leftarrow j$ ;
11         else if  $random(0,1) < exp(-\delta/T)$  then
12            |  $i \leftarrow j$ ;
13         end
14          $n \leftarrow n + 1$ ;
15     end
16      $t \leftarrow t + 1$ ;
17      $T \leftarrow T(t)$ 
18 end

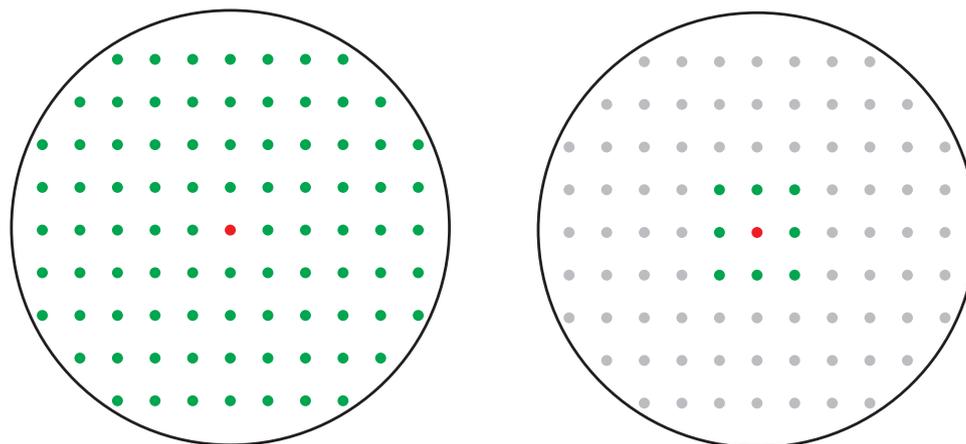
```

---

### 9.1.1.1 Neighbourhood Functions

In order to generate a new experiment state  $j$  in the simulated annealing algorithm, a suitable neighbourhood function is required. Obtaining a suitable neighbourhood function can become quite a challenge and may involve deeper understanding of the system. In general, randomizations of state variables are used, for instance randomize the order of the routing sequence.

When observing the analyses in Chapter 8, it was observed that there exists an unimaginable number of routing settings that are all performing roughly equivalently, but are located far from the global optimum. Only a specific shape is able to enhance model parameter accuracy. It is not smart to build a neighbourhood function that continuously randomizes the routing sequence and scan directions, since a great amount of luck is required to obtain a good performing routing. For the experiment optimization of the thermo-mechanical feedforward model, two neighbourhood functions are developed which are illustrated in Figure 9.1.



**Figure 9.1:** Geometrical interpretation of neighbourhood functions. The red dot indicates first randomly selected field, the green dots represent available fields for the second selection.

The dots in the figure represent the fields. For both neighbourhood functions, one field is selected randomly from all possible fields. This field is marked as the red dot in the figure. For the neighbourhood function illustrated in the left, a second field is chosen randomly from the remaining fields represented by the green dots. As is observable, this second field can be picked all over the substrate. When the two fields are selected, their positions in the routing order vector are exchanged. This is also known as a two swap-opt. For the scan direction, two fields are picked in the same manner. Subsequently, a random number is drawn which decides the scan direction upward or downward.

The neighbourhood function illustrated on the right side of the figure is slightly different. The selection of the first field is equivalent as the previous neighbourhood function and is indicated by the red dot. However, selection of the second field is restricted to fields nearby the first selected field as is indicated by the green dots. This allows variations of the routing which affects small area's at the substrate. The scan directions are chosen identical as before, but with the nearby selection procedure.

Both neighbourhood functions will result in different routings. The first neighbourhood function will most likely generate a routing similar as the jumpy routing, whereas the second neighbourhood function with the nearby search will most likely generate a more ordered routing, like the spiral or meander routing. It was seen that this could make a significant influence per model parameter case. It should be noted that variations of the mentioned neighbourhood functions are possible. More fields could be selected, or the neighbourhood function can be changed during the optimisation. For instance starting the optimization with the first neighbourhood function and when convergence is almost achieved, switch to the nearby search neighbourhood function. This allows for a improved local optimum which is hard to obtain by spatially large permutations of the first neighbourhood function.

## 9.2 Experiment Optimization

In this section the experiment, used for model parameter estimation, is optimized. At first, suitable objective functions are discussed, followed by an optimization of the dose. Thereafter, the experiment is optimized for all experiment input parameters using the simulated annealing algorithm.

### 9.2.1 Objective Functions

A suitable objective function is important to make the experiment optimization succeed. The objective function must be capable to describe the performance of the experiment by one single scalar value computed from the information matrix  $\mathbf{M}$ .

Both model parameter cases involve only two model parameters. The objective functions used for experiment design are given in Section 4.5 and an overview of their properties are listed in Table 4.2. The D-optimality criterion is chosen as an objective function for its acceptable performance for two model parameters.

It was seen that correlation has a significant impact on parameter accuracy. The ACE1 objective function aims at a reduction of correlation, while constraining the smallest eigenvalue of the information matrix and thus including the E-optimality. This objective function consist of a constraint and due to the discrete optimization problem, it is included by means of a penalty function given by:

$$\tilde{\phi} = \phi + p_{\text{pen}}(\max(0, g))^2, \quad (9.2)$$

where  $p_{\text{pen}}$  is a penalization factor and influences the contribution of the constraint,  $\phi$  is the objective function without constraint and  $g$  the constraint function given by:

$$\begin{aligned} \phi &= \min C_{ij}^2 && \text{with } C_{ij} = \max \mathbf{C}|_{\text{base}} \quad i \neq j \\ g &= 1 - \frac{\lambda}{p_{\text{con}}\lambda_{\text{base}}} \end{aligned} \quad (9.3)$$

Here,  $\lambda$  is the eigenvalue to be constrained,  $\lambda_{\text{base}}$  the base value of the eigenvalue to be constrained and  $p_{\text{con}}$  is a constraint factor which can influence the constraining bound of  $\lambda$ . By using a penalization function, the objective value  $\tilde{\phi}$  grows quadratically when the experiment design lies in the infeasible domain of the design space  $\Phi$ .

### 9.2.2 Dose Optimization Analysis

The dose determines the amount of EUV light exposed to the substrate and can vary per field. As mentioned before, the dose is a continuous parameter which lies between a lower and upper bound determined by the scan speed limitations of the lithography machine. The dose can typically be optimized using a gradient-based optimization and for that, sensitivity information of the dose on routing performance is required. This requires expensive sensitivities of the information matrix, which is built of sensitivities as well, and as the dose can differ per field the gradient-based optimization requires enormous computational power. The impact of variation of the dose per field is investigated in order to attempt a reduction in computational power without loss of experiment performance on model parameter accuracy.

An approach is to construct basis functions by hand and investigate the change in performance of the experiment <sup>1</sup>. The following functions are incorporated:

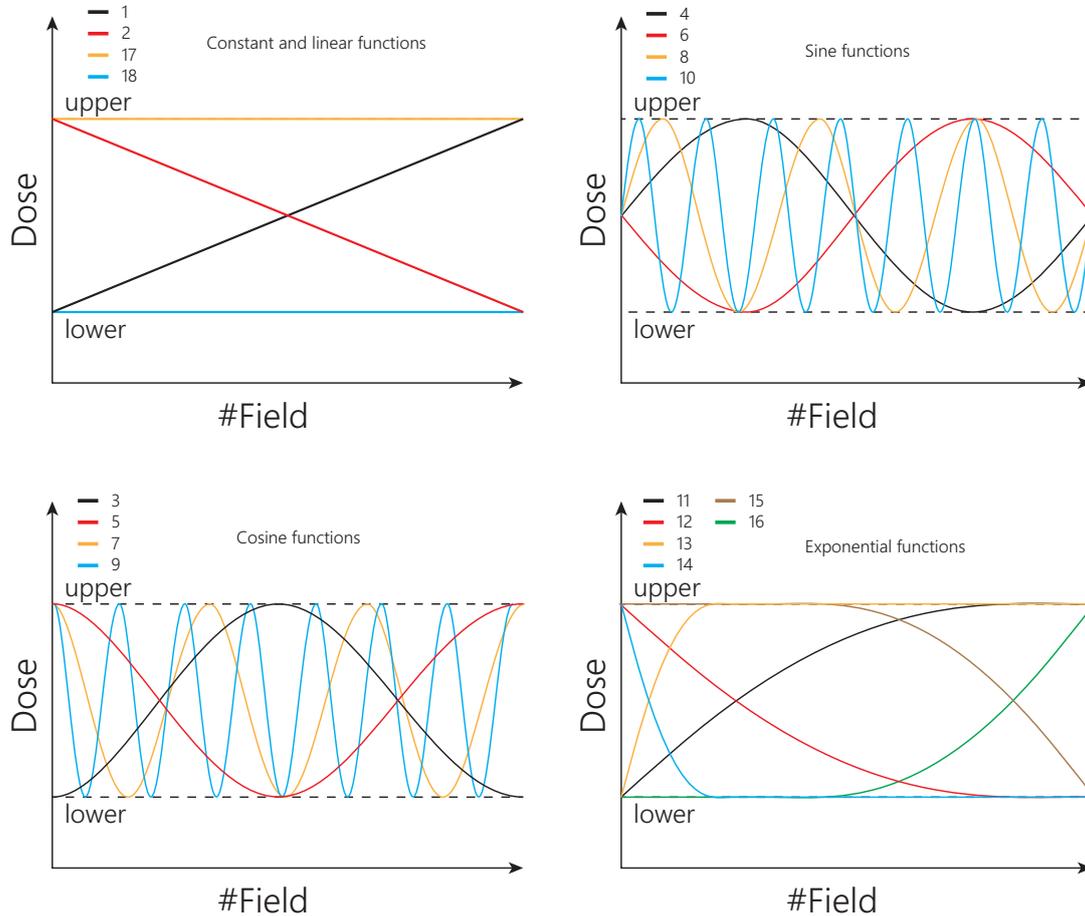
1. Constant functions.
2. Linear functions.
3. Sine and cosine functions.

---

<sup>1</sup>A *singular value decomposition* (SVD) analysis was performed at first by generating multiple randomized dose inputs varying between the lower and upper bound, calculating the D-optimal objective value, scale the dose input vector by the associate objective value and perform the SVD. This was done in order to acquire a set of basis functions that has significant influence on experiment performance. The basis functions can be used to reduce the DOFs of the dose input and hence, reduce the computational effort during gradient-based optimization. Unfortunately, the SVD did not result in any useful dominant basis function. This is caused by the high frequency content of the randomized dose input.

## 4. Exponential functions.

Combinations of the functions, such as a high or low frequency for the geometrical functions, linear increasing or decreasing functions, or rapid or slow exponential functions are made as well. In Figure 9.2 the concept of the basis functions is illustrated, where the dose is a function of the fields on the substrate.



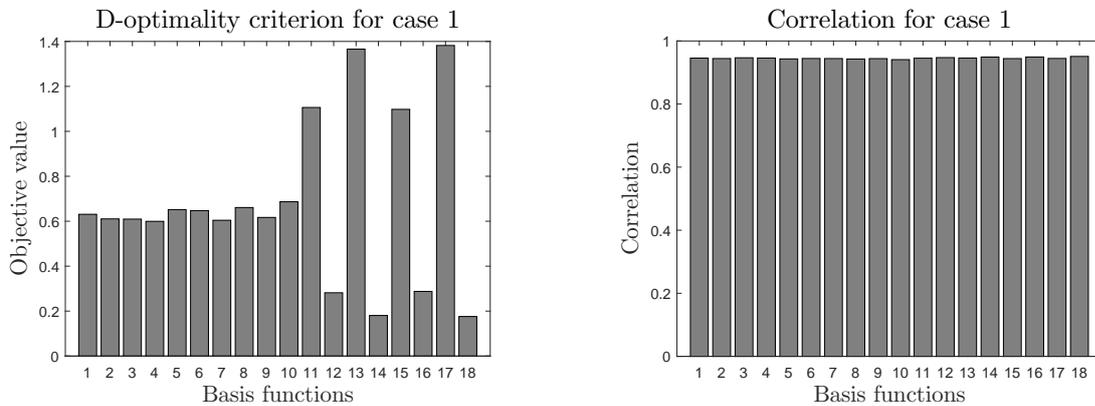
**Figure 9.2:** Basis functions for dose input parameter. Constant functions, linear functions, geometrical functions and exponential functions are investigated.

The basis functions are used to construct a dose input per field and subsequent simulations are performed with the meander routing. Thereafter, the information matrix  $\mathbf{M}$  is constructed for the model parameter subset of case 1 and of case 2. To determine the performance of the experiment for model parameter accuracy, the correlation between the model parameters and the D-optimality criterion is calculated. The result is illustrated in Figure 9.3.

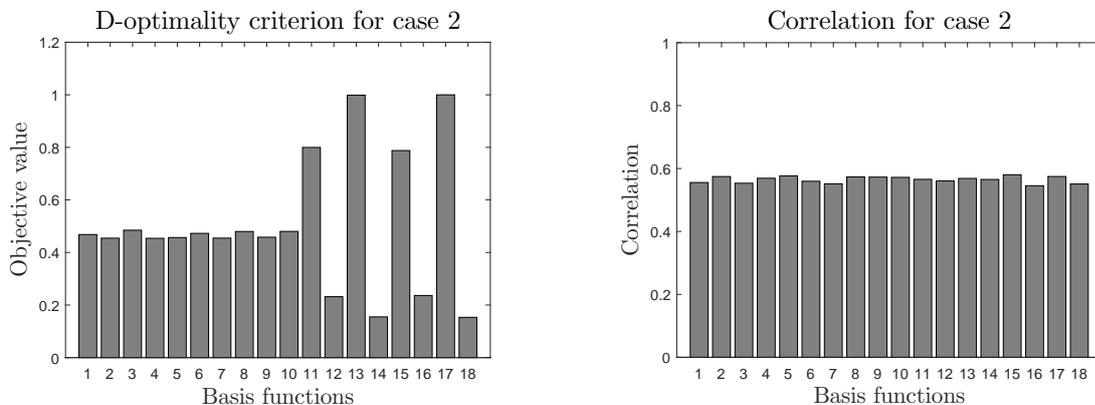
In the left graph, the scaled D-optimality objective value is plotted against the basis functions of the dose. The D-optimality criterion is calculated from the information matrix and a higher objective value indicates more model parameter accuracy, which is desired. The first few basis functions perform roughly the same and they correspond to the linear- and the geometrical functions. The mean dose value for these basis functions is approximately the same. The remaining basis functions do show significant differences in performance of the experiment. It is noticeable that the high objective values are associated to a basis function which involves a high dose for a large number of fields. The best performing dose function is maximum dose for all fields. It may not be a surprise that the worse performing dose function is a minimum dose for all fields. This result can be explained from the fact that a high dose results in higher temperature rises, hence larger overlay deformations. Subsequently, when the overlay

deformations are larger, the signal to noise ratio increases and thus model parameter accuracy is improved.

In the right graph, the correlation is plotted against the basis functions. It is observable that the correlation is hardly affected by the variation of the dose inputs and, therefore, it can be concluded that maximum dose for all fields will give the best performance for the model parameters of case 1. Note, the shown analysis is performed with the meander routing, but the remaining standard routings resulted in equivalent outcomes.



**Figure 9.3:** Dose input analysis for model parameter case 1. The left graph illustrates the D-optimality objective value against the basis functions. The right graph illustrates the correlation of the model parameter against the basis functions. The best performing dose input consists of maximum dose for all fields.



**Figure 9.4:** Dose input analysis for model parameter case 2. The left graph illustrates the D-optimality objective value against the basis functions. The right graph illustrates the correlation of the model parameter against the basis functions. The best performing dose input consists of maximum dose for all fields.

The analysis is performed on the model parameter subset of case 2 as well. The results are illustrated in Figure 9.4. Similar characteristics are observed for the D-optimality objective value and for the correlation between the model parameters. The correlation is slightly more affected by the variation of the dose input in contrast to case 1. It is noticeable that minimum dose for all fields results in the lowest correlation, but this effect does not dominate the performance of the experiment. It can be concluded that maximum dose for all fields will give the best

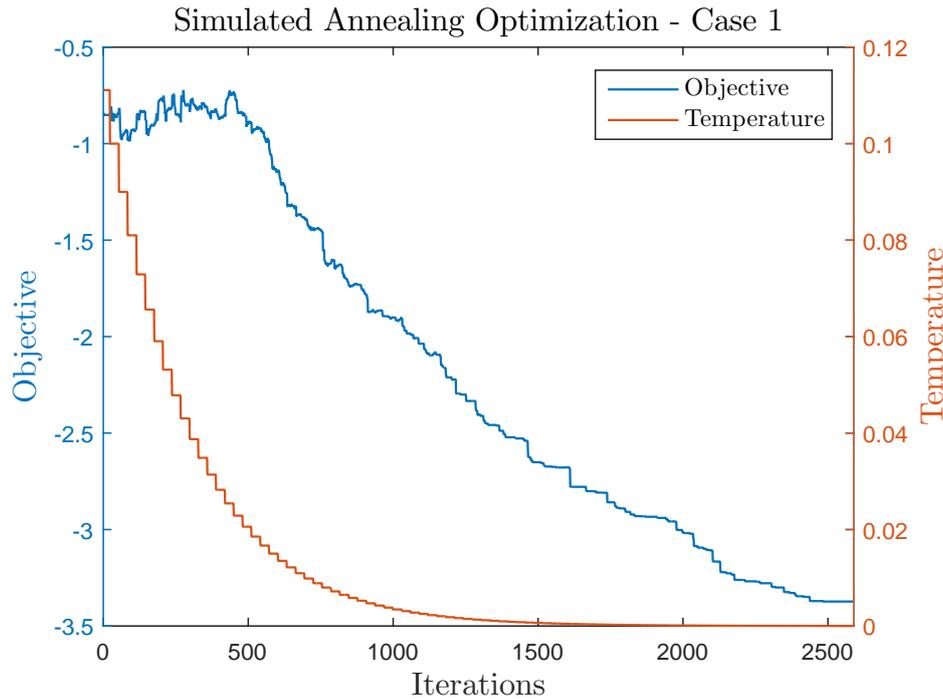
performance for the model parameters of case 2. Note, the shown analysis is performed with the meander routing, but the remaining standard routings resulted in equivalent outcomes. The maximum dose is set to a fixed value of  $80 \text{ J m}^{-2}$  during optimization of the routing and scan direction for both cases.

### 9.2.3 Experiment Optimization

The SA algorithm is used to find an optimal experiment in the context of optimal model parameter estimation. The pseudo-code is given in algorithm 1 and the input parameters routing and scan direction are used. In the previous subsection, it was found that maximum dose for all fields resulted in the best experiment performance. Since there are two cases of the model parameter subset, see Table 8.3, two experiment optimizations are executed.

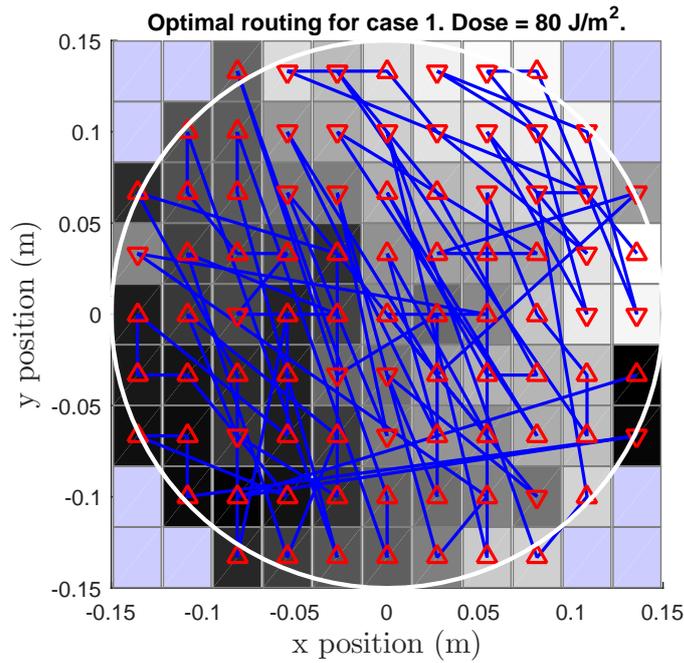
**Optimization for Case 1** In Chapter 8, the standard experiments were examined on their performance. For the model parameters of case 1, i.e., the heat capacitance  $c_p$  and the IR/EUV ratio, the jumpy routing was observed to be the best performing experiment in the context of model parameter estimation and the associated accuracy. This experiment consists of a random permutation of the exposure sequence (routing). Therefore, the initial state of the routing and scan directions for the experiment optimization are randomly generated. The neighbourhood function, which allows permutations over the whole surface of the substrate, is suitable to maintain random behaviour and therefore finding an optimal jumpy routing. For this reason it is used in the SA algorithm.

In Chapter 8 it was seen that the correlation between  $c_p$  and IR/EUV ratio is hardly influenced by the choice of the routing. From Chapter 8 it is evident that the correlation is expected due to the thermal characteristics of both parameters. It is not useful to optimize the routing for a reduction in correlation of the model parameters. Hence, the D-optimality criterion is chosen as objective function  $\phi$  in the SA algorithm. The objective function is scaled by the objective value for the nominal jumpy routing. In this manner, it is easy to see the performance enhancement with respect to the best performing standard experiment during optimization. It must be noted that the objective function is multiplied by -1, since the optimization is a minimization of the objective value. The initial temperature is set to  $T = 0.11$ , which indicates that negative fluctuations of 0.055 in the objective function have a chance of approximately 60% to be accepted. From experience, it is known that the objective value changes in the same order when perturbing the experiment randomly.



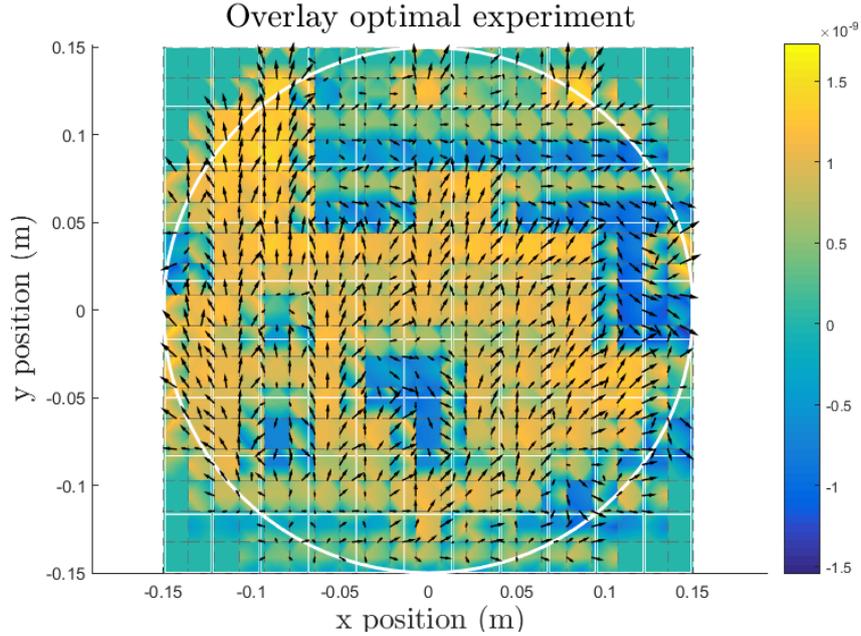
**Figure 9.5:** Simulated annealing optimization for case 1 with D-optimality criterion as the objective function. The blue line indicates the scaled objective value and the red line indicates the temperature during the simulated annealing process.

The simulated annealing optimization progress is illustrated in Figure 9.5. The blue line represents the objective value, which is calculated with the D-optimality criterion. The red line represents the temperature in the SA algorithm and is the most important aspect of the SA algorithm. In the beginning of the optimization process, the temperature is high and worse objective values might be accepted. In the figure, it is observable that the objective function does not start to decrease up to 500 iterations. During these iterations, the solution of the experiment state can overcome local minima. After 500 iterations, the temperature has gradually cooled down to a level where the algorithm is more restricted to accept improved objective values only. Between 500 and 1500 iterations, there is still a possibility to escape local minima and it can be seen that worse objective values are still accepted. However, the algorithm is converging in general. After 1500 iterations, the temperature is decreased such that the algorithm behaves like a direct search method. It is not possible to escape local minima and after 2500 iterations, the algorithm has converged. It is most likely that the found minimum is not the global optimum, but a local optimum instead. Nonetheless, as was mentioned in Section 9.1 the SA algorithm is used to speed up the process of finding a satisfactory solution. The algorithm has converged around an objective value of approximately -3.2, which indicates a reduction 3.2 of the confidence area with respect to the jumpy routing.



**Figure 9.6:** The optimal experiment for model parameters of case 1 established from the SA optimization algorithm.

In Figure 9.6 the optimal experiment in the context of parameter estimation of  $c_p$  and IR/EUV is illustrated. This experiment is established from the SA optimization algorithm. The routing is not ordered like the meander or spiral routing and this is expected due to the choice of the neighbourhood function. The distance between two field exposures is across the whole surface of the substrate in a jumpy fashion. However, the routing does not seem arbitrarily random, since the colours indicate an evidently scanning pattern from the lower left- to the upper right corner of the substrate. From the figure it can be observed that the scan directions are mostly directed in positive  $y$  direction by the optimization algorithm. The optimal experiment is simulated with nominal model parameters and the overlay response is illustrated in Figure 9.7.



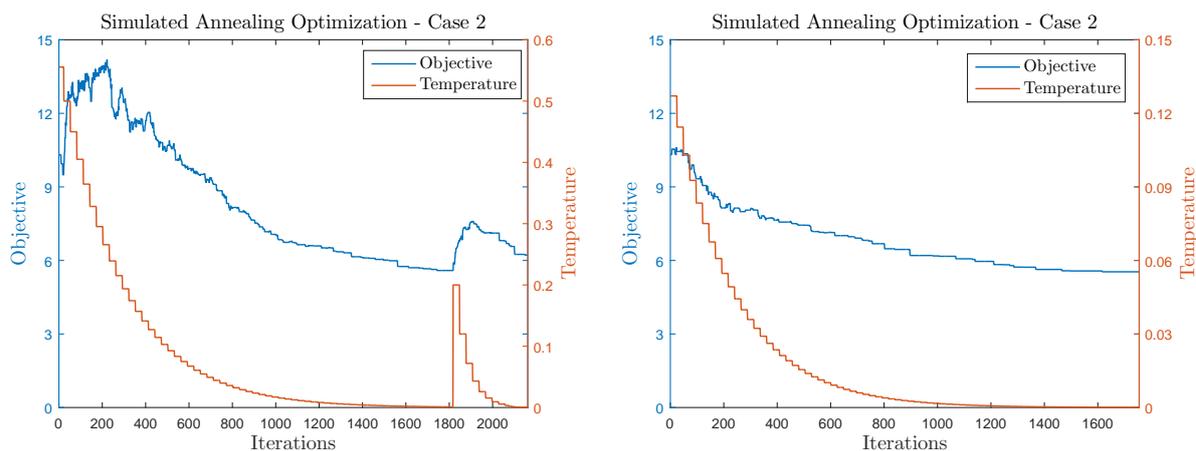
**Figure 9.7:** Overlay response of optimal experiment for parameters from case 1.

The overlay clearly illustrates the scanning motion from the lower left- to the upper right corner of the substrate. The overlay response is spread from the lower left to the upper right as well. In the beginning of the scanning sequence, overlay response is primarily directed in the positive  $y$  direction. When the exposure continues, the overlay response is directed more to the right. It is not possible to gain a deeper understanding of the optimal experiment from the overlay response only. In Section 9.3, a thorough analysis is performed on the optimal experiment.

**Optimization for Case 2** In Chapter 8 the standard experiments were examined on their performance. For the model parameters of case 2, the tangential stiffness  $G_{zy}$  and the IR/EUV ratio, the meander routing was observed to be the best performing experiment in the context of model parameter estimation and the associated accuracy. This experiment consists of an ordered permutation of exposure sequence (routing). In Section 8.6 it was concluded that horizontal scanning motions are beneficial for the correlation of the model parameters in case 2. Therefore, the initial state of the routing and scan direction for the experiment optimization is the meander routing, since this routing involves a horizontal scanning sequence<sup>2</sup>. The neighbourhood function which allows permutations nearby a selected field on the substrate is suitable to maintain ordered behaviour, see Figure 9.1.

For this optimization it is useful to optimize the routing for a reduction in correlation of the model parameters, but improvement of the eigenvalues is desired as well. Hence, the ACE1 optimality criterion with constraint is chosen as objective function  $\phi$  in the SA algorithm. An eigenvalue constraint of  $p_{\text{con}} = 2$  is used, which implies a constraint of the eigenvalue 2 times smaller than the reference eigenvalue  $\lambda_{\text{base}}$ , see equation (9.3). In this way, the least informative eigenvalue will be enhanced as well. The penalty factor  $p_{\text{pen}}$  is dependent on the type of problem and the desired level of contribution, but is set to  $p_{\text{pen}} = 100$  in this case. The initial temperature is set to  $T = 0.55$ , which indicates that negative fluctuations of 0.275 in the objective function have a chance of approximately 60% to be accepted. From experience, it is known that the objective value changes in the same order.

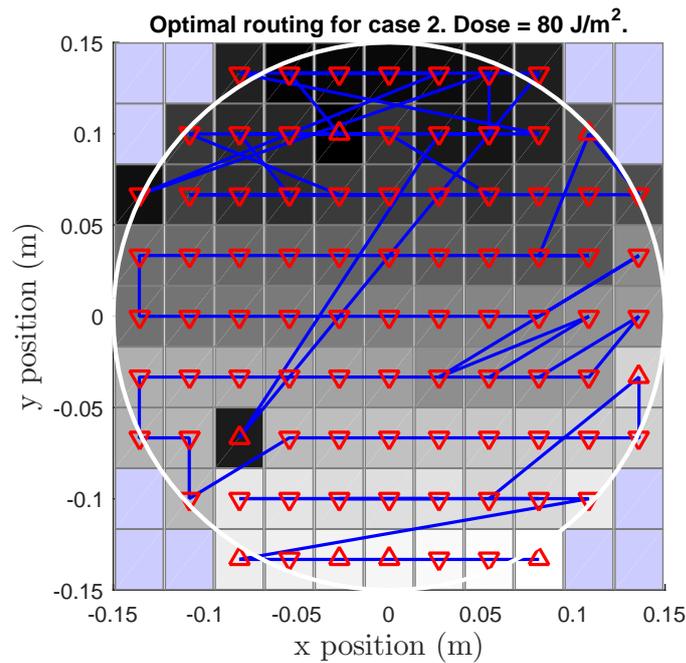
<sup>2</sup>An randomly generated initial state was tried as well, but without success. This is expected from by observing the performance of the jumpy experiment seen in Section 8.5.



**Figure 9.8:** Simulated annealing optimization for case 2 with ACE1-optimality criterion as the objective function. The blue line indicates the objective value and the red line indicates the temperature during the simulated annealing process. The left figure illustrates the first attempt with nearby neighbourhood function, the right figure illustrates the second attempt with the neighbourhood function that allows permutations across the whole substrate.

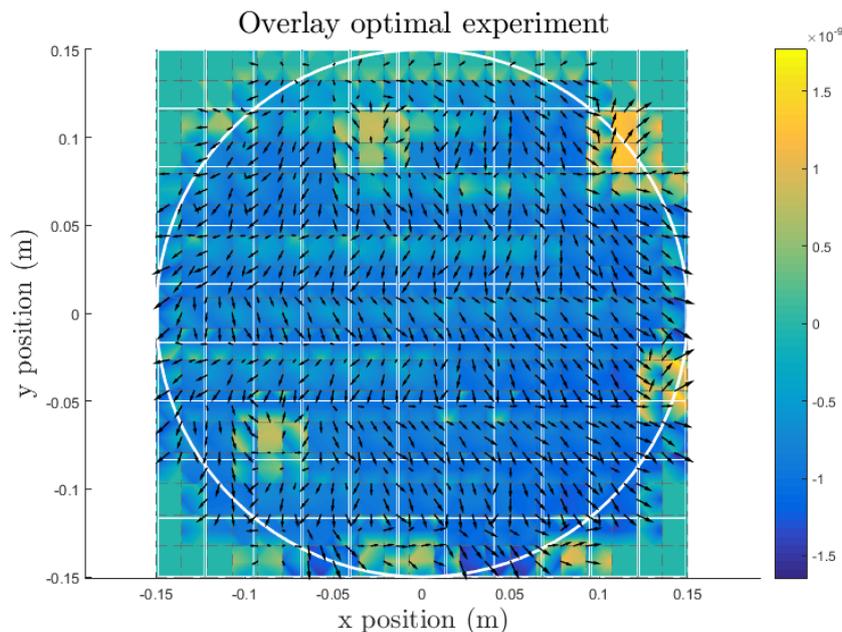
The simulated annealing optimization progress is illustrated in Figure 9.8. Two optimization attempts are performed. The left figure represents the first optimization attempt with the ACE1 optimality criterion and the nearby neighbourhood function. Initially, the temperature is high and worse objective values may be accepted as new experiment design. This behaviour is observable from the figure. After approximately 500 iterations the algorithm starts to converge and stops improving around 1800 iterations. The objective function has a value higher than 1, indicating that the constraint in (9.3) is not satisfied. This becomes clear when considering that the addition of the constraint is zero when the constraint is not active and the correlation squared lies between 0 and 1. Just after 1800 iterations, a re-heating is executed. Re-heating is a technique in order to try to escape from a local minimum. The temperature is increased, which allows worse objective values to be accepted. The figure shows the increase of the temperature and a rise of the objective function. Although this technique is often successful, it is not able to enhance the experiment design. This is mostly caused by the choice of the neighbourhood function.

A second attempt is made with the remaining neighbourhood function which allows permutations across the whole substrate. The initial temperature is lowered to  $T = 0.13$  in order to compensate for large permutations in the experiment design. This is done to remain an approximately ordered routing, since it was seen that a jumpy routing is not beneficial. The right figure illustrates the optimization progress. It is noticeable that the optimization starts converging immediately and only small deviations of worse objective values are accepted. This is logical due to the lower temperature. After approximately 1500 iterations, the SA algorithm is converged and a slightly improved objective value is found when compared with the first attempt.



**Figure 9.9:** The optimal experiment for model parameters of case 2 established from the SA optimization algorithm. The neighbourhood function that allows large spatial permutations of the routing is used.

In Figure 9.9 the optimal experiment in the context of parameter estimation of  $G_{zy}$  and IR/EUV is illustrated for the second attempt. The routing is approximately similar to the meander routing, except for a few fields. The meander routing was the initial state of the optimization algorithm and apparently only a few permutations are necessary. Due to the choice of the neighbourhood function the permutations of the routing are spread over the whole substrate. It is noticeable that the scan directions are almost completely directed in the negative  $y$  direction. The optimal experiment is simulated with nominal model parameters and the overlay response is illustrated in Figure 9.10.



**Figure 9.10:** Overlay response of optimal experiment for parameters from case 2.

The overlay clearly illustrates the scanning motion pointed in negative  $y$  direction. The overlay response is spread from top to bottom and in the middle, the overlay response is deflecting to the left or right due to the scan sequence in  $x$  direction per row. It is not possible to gain deeper understanding of the optimal found experiment from the overlay response only. In Section 9.3 a thorough analysis is performed on the optimal experiment.

## 9.3 Optimal Experiment Design Analysis

The optimal experiments are achieved using the SA optimization algorithm. In this section, the optimal experiments are investigated by means of a least squares estimation procedure, followed by t-profiling investigation to investigate non-linear behaviour and obtaining accurate confidence regions. At last, the fingerprints are analysed in order to gain a deeper understanding of the optimization problem and the performance of the experiments.

### 9.3.1 Model Parameter Estimation

Least squares estimation is performed to obtain model parameter estimates for case 1 and case 2. For each case, the optimal experiment established from the SA optimization is used. The experiment is constructed in the same manner as the analyses described in Section 8.4. The initial values for the parameters are set to  $\hat{\mathbf{p}} = 0.9\mathbf{p}$ . The least squares estimation process is illustrated in Figure A.4.

The least squares estimation converges fast and the model parameter estimates are found after 4 iterations. The ease of convergence is expected due to previous analysis of the standard experiments. In Table 9.1 the estimates are listed and as before, the true values are not retrieved. The error is given, but recall that the error does not give any statistical information about the accuracy of the estimates. In the following subsection the statistical information and confidence regions are investigated.

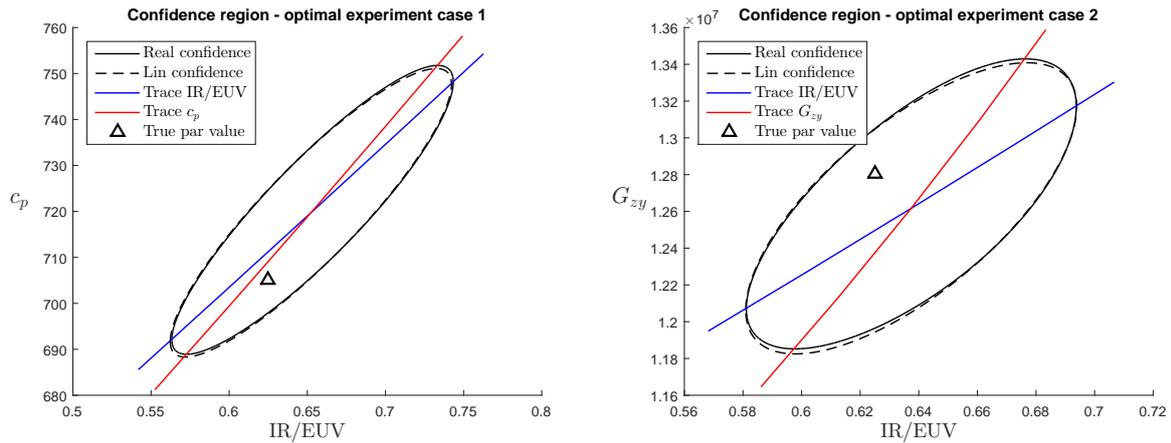
**Table 9.1:** Case 1 and 2. Model parameter estimation for optimal experiments.

Case 1					Case 2				
Optimal experiment 1 (Figure 9.6)					Optimal experiment 2 (Figure 9.9)				
ITER.	PAR.	TRUE	EST.	ERR. (%)	ITER.	PAR.	TRUE	EST.	ERR. (%)
4	$c_p$	705	719.7	2.1	4	$G_{zy}$	$1.28 \cdot 10^7$	$1.26 \cdot 10^7$	1.6
	IR/EUV	0.625	0.652	4.4		IR/EUV	0.625	0.637	1.9

### 9.3.2 Statistical Investigation

Least squares estimates of the model parameters are obtained from the optimal experiments. However, one estimation does not provide additional statistical information about the model parameters. It is known that the model behaves non-linear in its model parameters. Hence, profiling techniques are applied in order to investigate the rate of non-linearity. When this is within certain bounds, linear approximations are sufficient to obtain reliable statistical information about the model parameters.

The profiling techniques are applied to both optimal experiments in the same manner as is described in Section 8.5 and the obtained results are comparable and can be seen in appendix A.4.



**Figure 9.11:** Confidence regions of 95% for both cases with their associated optimal experiment. The left figure represents the confidence region for case 1, the right figure represents the confidence region for case 2. The linearly approximated confidence regions are sufficient to represent the true confidence regions.

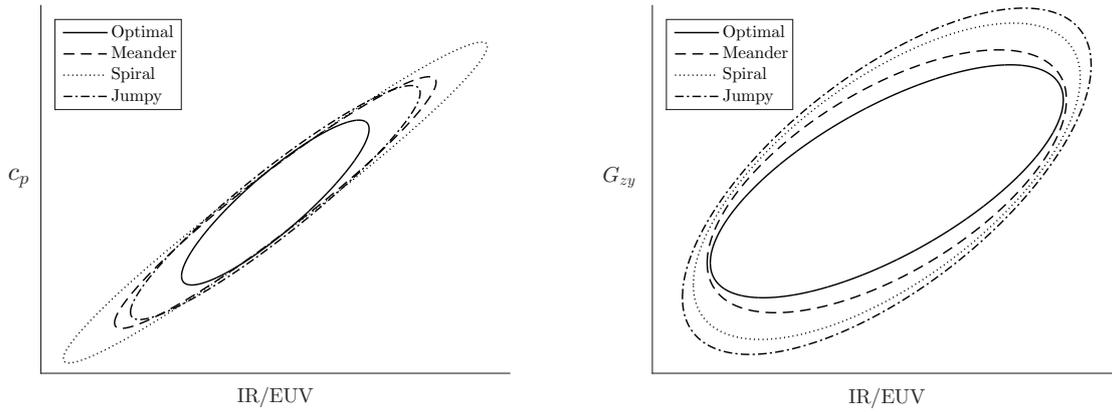
In Figure 9.11 the true 95% confidence region, linearly approximated 95% confidence region, and the profile trace vectors are illustrated for both optimal experiments associated with case 1 and case 2. The linearly approximated confidence regions are obtained by linearising the model around  $\hat{\mathbf{p}}$ . The true confidence regions are well approximated by the linear confidence regions, as is expected from the t-profiling results and previously performed analyses on the standard experiments. For case 1, the profile trace vectors intersect at a sharp angle and they almost coincide. This phenomenon indicates correlation between the model parameters and results in an elongated ellipsoidal confidence region. This was previously seen as well.

The profile trace vectors for parameter case 2 (right figure) indicate less correlation. As a result, the confidence region is less elongated, which results in a smaller confidence region. The confidence bounds for the IR/EUV ratio are correctly approximated by the linearised confidence region, whereas the confidence bounds of  $G_{zy}$  are not exact, but sufficiently approximated by

the linearised confidence region. This is expected when understanding the t-profiling response. When comparing both ellipsoids, it can be observed that the confidence bound for IR/EUV is smaller when its value is estimated together with the tangential burl stiffness.

At last, the true nominal parameter values, indicated by the triangle, lie within both confidence regions.

For both parameter cases, the linearised confidence regions of the optimal experiments and of the standard experiments are illustrated. It should be noted that the associated parameter values are not shown, in order to compare the confidence region per experiment which are derived at the point of the estimates listed in Table 8.4 and 9.1. The 95% confidence regions are illustrated in Figure 9.12.



**Figure 9.12:** Comparison of 95% confidence regions for both parameter cases. The left figure represents the confidence region for case 1, the right figure represents the confidence region for case 2. Both optimal experiments result in the highest accuracy for the model parameters.

The left figure represents the confidence regions for the optimal experiment and the three standard experiments for the model parameters of case 1. From the figure it is evident that the optimized experiment is successful in enhancing the model parameter accuracy. The confidence region is reduced significantly and the confidence bounds are almost reduced by half compared to the best performing routing of the standard experiments. The model parameters are still highly correlated, but this was already expected due to the characteristics of the model parameters. The correlation between the model parameters is calculated and shown below.

$$\mathbf{C}_{\text{jumpy}} = \begin{bmatrix} 1 & -0.91 \\ & 1 \end{bmatrix} \quad \rightarrow \quad \mathbf{C}_{\text{optimal}} = \begin{bmatrix} 1 & -0.89 \\ & 1 \end{bmatrix} \quad (9.4)$$

The optimal experiment is able to reduce the correlation between  $c_p$  and IR/EUV ratio even more than the standard experiment with the jumpy routing, see equation (8.17).

The right figure represents the confidence regions for the optimal experiment and the three standard experiments for the model parameters of case 2. From the figure it is evident that the optimized experiment is successful in enhancing the model parameter, since the confidence region is reduced. However, this is only the case for the tangential burl stiffness. The IR/EUV ratio has similar confidence bounds for the optimal experiment as for the experiment with the meander routing. The latter was the best performing experiment of the standard experiments. The correlation between the model parameters is calculated and shown below.

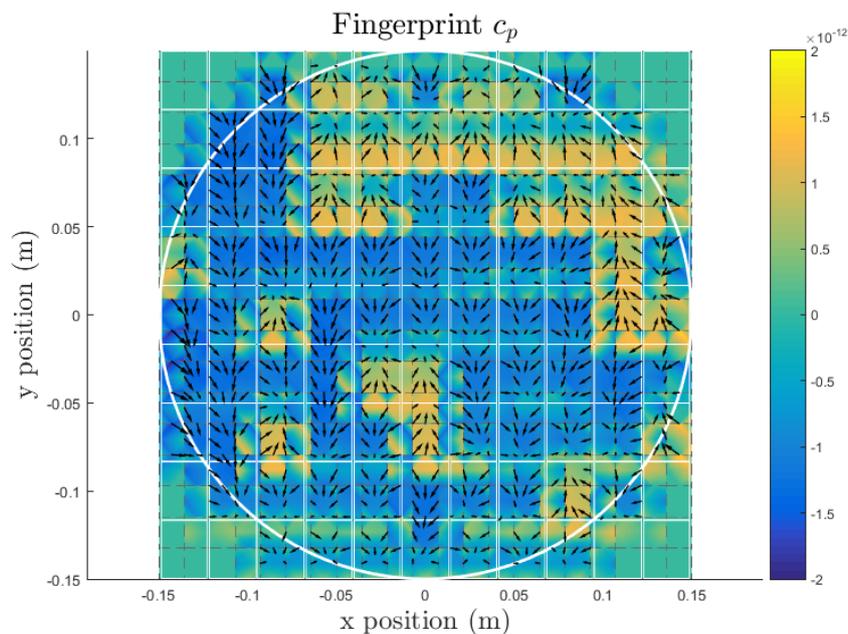
$$\mathbf{C}_{\text{meander}} = \begin{bmatrix} 1 & -0.56 \\ & 1 \end{bmatrix} \quad \rightarrow \quad \mathbf{C}_{\text{optimal}} = \begin{bmatrix} 1 & -0.70 \\ & 1 \end{bmatrix} \quad (9.5)$$

The optimal experiment does not reduce the correlation between the model parameters when compared with equation (8.17). However, the optimal experiment is still able to enhance model parameter accuracy successfully. One disadvantage of this deteriorated correlation is that the least squares estimation may be hampered and thus requires more iterations.

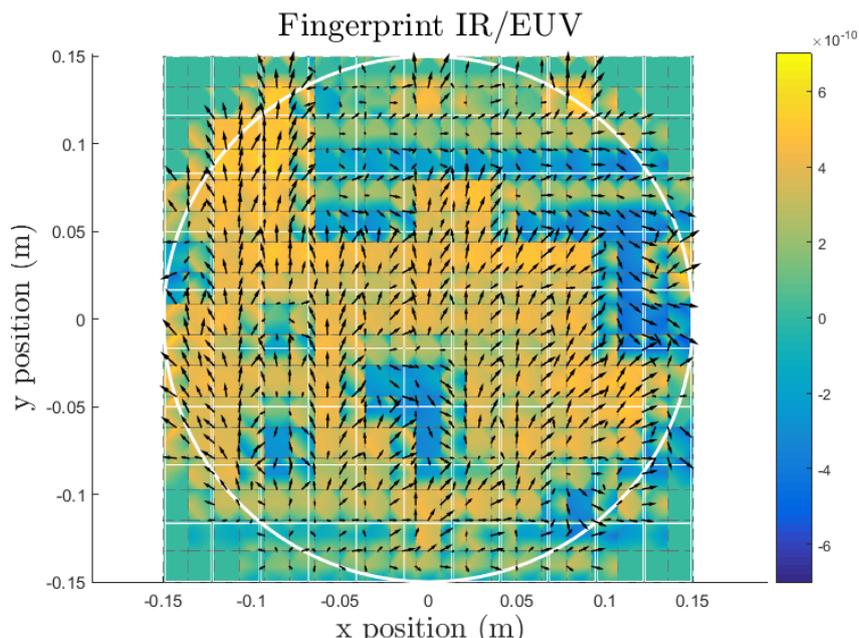
### 9.3.3 Fingerprints of Optimal Experiments

The performance of optimal experiments is evaluated in the previous section. However, it is not known where their performance comes from. In order to investigate this, the fingerprints of the model parameter subset are evaluated for each optimal experiment in the following paragraphs. Note, in Section 8.1.3, the shapes and the origin of correlation between model parameters is already explained.

**Fingerprints Case 1** The established optimal experiment from the SA algorithm for case 1 is illustrated in Figure 9.6. The associated overlay response is constructed and is illustrated in Figure 9.7. In figures 9.13 and 9.14 the fingerprints of  $c_p$  and IR/EUV ratio are illustrated, respectively. Recall that two essential ingredients are required to obtain a good performing experiment: little correlation between model parameters and a high signal to noise ratio.



**Figure 9.13:** Fingerprint, or sensitivity, of optimal experiment for model parameter  $c_p$ .



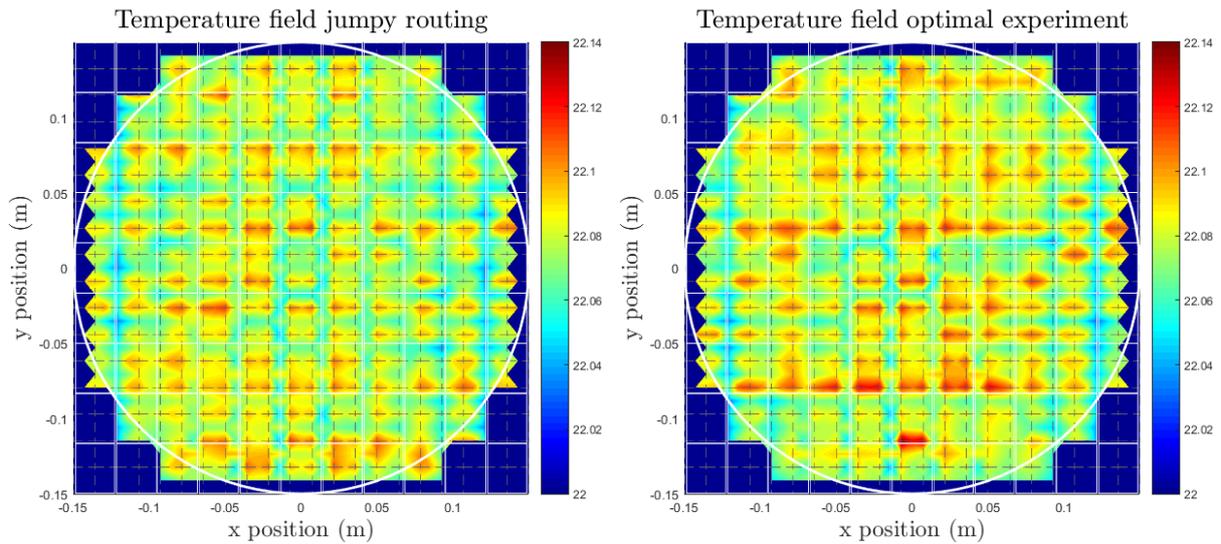
**Figure 9.14:** Fingerprint, or sensitivity, of optimal experiment for model parameter IR/EUV ratio.

The shape of the fingerprint of the heat capacitance  $c_p$  is in reversed direction of the overlay response, which is familiar. The magnitude of the fingerprint is slightly higher when comparing with Figure 8.20, where the fingerprint of  $c_p$  of the best performing standard experiment (jumpy) for this case can be seen. The difference in magnitude is limited and is difficult to see from the colour map.

The shape of the fingerprint of the IR/EUV ratio is identical to the shape of the overlay response, which is familiar as well. However, the magnitude of the fingerprint is significantly higher when comparing with the fingerprint of the IR/EUV ratio of the jumpy routing in Figure 8.20. Apparently, the IR/EUV ratio becomes more sensitive to overlay, which can only be accomplished by an increase in the temperature field. This is achieved due to the skewed routing and the associated geometry of the fields and their scan directions and this is confirmed in Figure 9.15. It can be concluded that the reduction in correlation and the enhancement of sensitivity information of the IR/EUV ratio are accountable for the optimal experiment design.

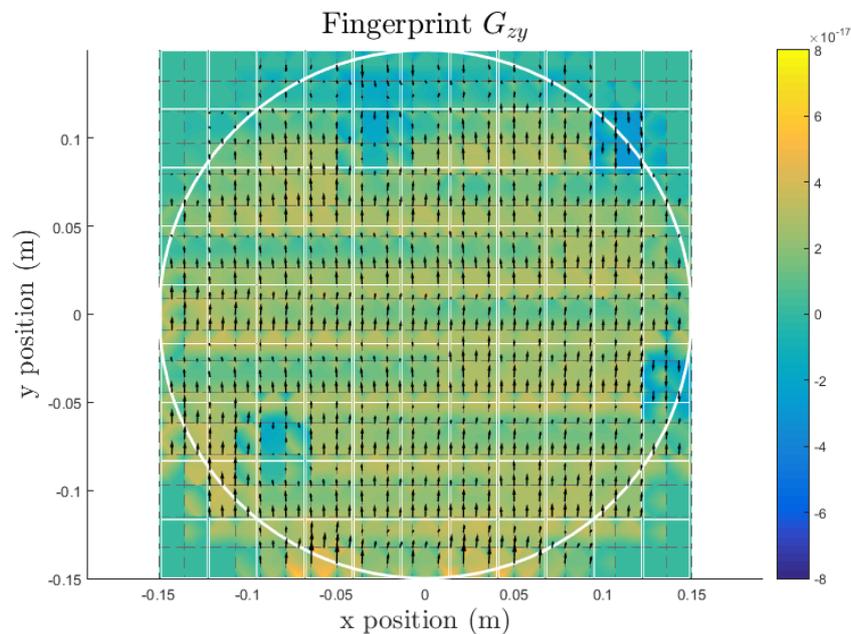
The optimal routing may be improved by understanding that maximum sensitivity response yields in enhanced model parameter accuracy when subjected to least squares estimation. From the IR/EUV fingerprint, it is observable that the sensitivity response of the second row of fields, viewed from the top of the substrate, is less compared to the remaining responses. This response can be improved by adjusting the routing such that the heat is not already dissipated locally when the surrounding fields are exposed.

It is observable as well that some fields are exposed in negative scan direction (lower left corner). This can cause a counterbalancing effect for overlay, hence a reduction in sensitivity information. This counterbalancing effect is probably not seen by the optimizer due to the immense design space. However, manual adjustment of the scan direction at these particular fields is a straightforward modification.

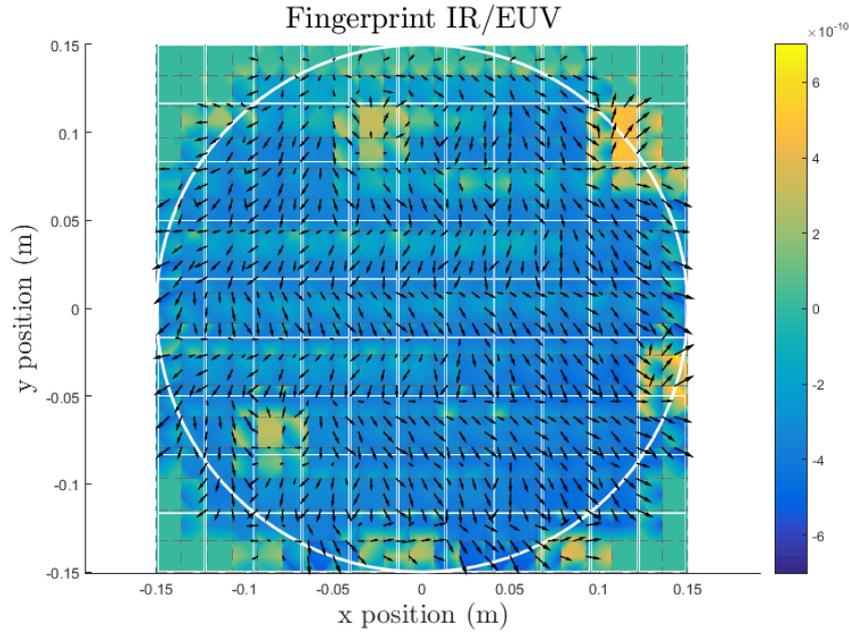


**Figure 9.15:** Temperature field for overlay experiment with jumpy routing and optimal experiment. Due to the routing of the optimal experiment (right figure), the temperature field is slightly higher which results in an increase of the IR/EUV ratio sensitivity.

**Fingerprints Case 2** The optimal experiment from the SA algorithm for case 2 is illustrated in Figure 9.9. The associated overlay response is constructed and is illustrated in Figure 9.10. In figures 9.16 and 9.17 the fingerprints of  $G_{zy}$  and IR/EUV ratio are illustrated, respectively. Recall that two essential ingredients are required to obtain a good performing experiment: little correlation between model parameters and a high signal to noise ratio.



**Figure 9.16:** Fingerprint, or sensitivity, of optimal experiment for model parameter  $G_{zy}$ .



**Figure 9.17:** Fingerprint, or sensitivity, of optimal experiment for model parameter IR/EUV ratio.

The shape of the fingerprint of the tangential burl stiffness  $G_{zy}$  is directed in  $y$  direction and reversed compared to the scan motion, which is familiar. The magnitude of the fingerprint responses is significantly improved when comparing with Figure 8.7, which is the fingerprint of  $G_{zy}$  of the best performing standard experiment (meander) for this case. The difference in magnitude is difficult to see from the colour map. In the meander routing, the sensitivity gets cancelled out by the alternation of the scan direction. This pattern is clearly visible in Figure 8.7. However, due to constant downward scan motions in the optimal experiment, the fingerprint of Figure 9.16 contains more sensitivity information.

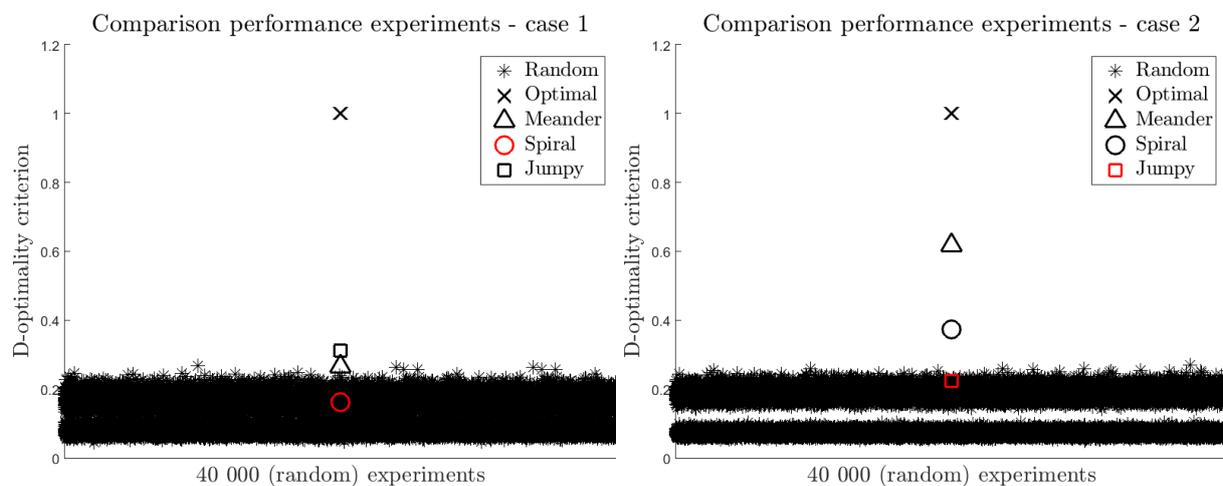
The shape of the fingerprint of the IR/EUV ratio is identical to the shape of the overlay response, which is familiar as well. However, the magnitude of the fingerprint is not significantly higher when comparing with Figure 8.8, which is the fingerprint of the IR/EUV ratio of the meander routing. This effect is seen by the confidence bound for IR/EUV in Figure 9.12. Apparently, the meander experiment is already a good performing experiment. The correlation of this experiment is increased, as was seen in equation (9.5), and can be explained by the fingerprint of IR/EUV. Due to the downward scan motions in  $y$  direction the overlay response is more pointed in the  $y$  direction, hence the IR/EUV ratio sensitivity is more pointed in this direction as well as is seen from the figure. Therefore, the IR/EUV ratio sensitivity becomes less orthogonal to the sensitivity of  $G_{zy}$ , since the latter is purely pointed in the  $y$  direction and thus the correlation is deteriorated. Fortunately, the increase of correlation does not affect the overall performance of the experiment due to the enhancement of the tangential burl sensitivity.

The optimal routing may be improved by understanding that maximum sensitivity response yields in enhanced model parameter accuracy when subjected to least squares estimation. From the  $G_{zy}$  and IR/EUV fingerprint, it is observable that some fields are exposed in positive scan direction (lower left corner). This can cause a counterbalancing effect for overlay, hence a reduction in sensitivity information. Adjusting the scan direction at these particular fields is a straightforward modification which may improve the optimal experiment.

## 9.4 Efficiency Optimization Algorithm Investigation

In the previous sections, the optimal experiments are constructed and a thorough analysis is performed in order to investigate their performance on model parameter estimation. It was seen that both optimal experiments are able to reduce the confidence regions of the model parameters compared to the standard experiments. The Simulated Annealing algorithm was used to obtain the optimal experiments, this algorithm is a direct heuristic method which is able to overcome local minima. However, it is not known how efficient this algorithm performs.

In this section, the efficiency of the SA algorithm is investigated by means of a brute force approach. The High Performance Computing (HPC) cluster of the Technical University Delft is used to generate as many as possible random experiments. This will allow a global search in the design space  $\Phi$ . Subsequently, the randomly generated experiments are compared with the standard and optimal experiments for their performance by means of the D-optimality criterion. This is a simple criterion that determines the area of the confidence region and is a proper choice to investigate the efficiency of the algorithm. In Figure 9.18 the result is illustrated.



**Figure 9.18:** Brute force approach is generating random experiments in order to investigate the efficiency of the SA algorithm.

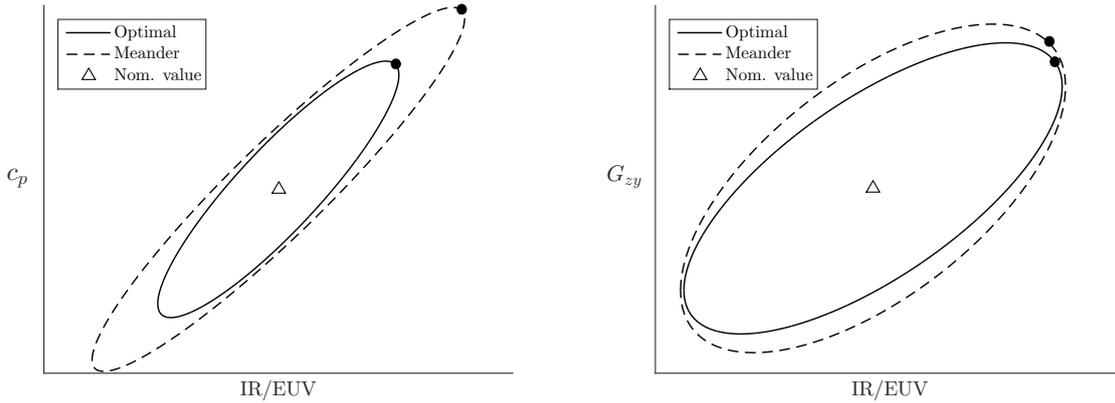
The objective value is scaled by the D-optimality criterion of the optimal experiments. From the figure, it can be seen that the optimal experiment is performing best, followed by the standard experiments according Figure 9.12. It is noticeable that the random generated experiments perform about five times, or more, worse than the optimal experiment. It must be remarked that the brute force approach only covers an infinitesimal small space in the enormous design space  $\Phi$ . However, the performed brute force approach required roughly 40 times more computational effort than the SA algorithm and not even a single experiment comes close to the optimal experiment established from the SA algorithm. From Section 9.1 it is known that only direct methods are applicable for this type of experiment optimization. Therefore, it can be concluded that the SA algorithm is very efficient and successful in finding an optimal experiment design.

From the figure a new phenomenon can be seen as two bands appear wherein the random generated experiments are located according to their D-optimality criterion. When observing the overlay experiments from this upper and lower bound, no direct causes are visible what could cause the appearance of the bands. The routings are randomly generated and seem similar as the jumpy routing and the same holds for the scan directions. The associated fingerprints indicate the differences in D-optimality criteria as alternating magnitudes and counterbalancing overlay response effects are visible. This was seen in previous analyses as well. However, the fingerprints do not indicate a clear characteristic which could explain the fractionated bands.

Further research is necessary which may be useful to acquire even more knowledge about the system, which in turn could be advantageous for experiment design optimization.

## 9.5 Overlay Accuracy

The optimal experiments are found for model parameters of case 1 and 2 and it was seen that the confidence region has been reduced. In this section, the impact on overlay accuracy is investigated by using extreme points on the confidence regions for the best standard and optimal experiment when estimating the model parameters.



**Figure 9.19:** Parameter bounds for optimal experiments compared to best standard experiment. The dots represents the worst joint confidence extremes.

In Figure 9.19 the optimal confidence regions around the nominal parameters  $\mathbf{p}$  are illustrated with the confidence region of the best performing standard experiment per associated model parameter case. The worst joint confidence bounds are determined in order to investigate the accuracy bounds for overlay, see the black dots in the figure. Note, the linearly approximated confidence regions are used as they are sufficient to replace the true confidence regions, as was seen in Section 9.3. In Table 9.2 the nominal (nom.) values, the bounds of the optimal experiment (b. opt.) and the bounds of the best performing standard experiment are listed.

**Table 9.2:** Model parameter bounds for 95% confidence for case 1 and case 2.

CASE 1				CASE 2			
PAR.	NOM.	B. OPT.	B. JUMPY	PAR.	NOM.	B. OPT.	B. MEANDER
$c_p$	705	734.6	748.8	$G_{zy}$	$1.28 \cdot 10^7$	$1.35 \cdot 10^7$	$1.36 \cdot 10^7$
IR/EUV	0.625	0.712	0.762	IR/EUV	0.625	0.679	0.677

In order to determine the improvement on overlay accuracy a nominal simulation is performed with the nominal meander routing. This exposure represents a real scan exposure for, for instance, computer chips. From this simulation, the overlay is determined by choosing the highest deformation from 99.7% of the deformations sorted from low to high. The obtained overlay value is used as a reference  $O_{\text{ref}}$ .

For case 1, one simulation is performed with the same recipe and with the model parameter values from Table 9.2 for the optimal bound and one simulation is performed with the model parameter values from the jumpy bound. Subsequently, the overlay values  $O_{\text{opt}}$  and  $O_{\text{jumpy}}$  are

obtained and the difference in overlay with respect to the reference is calculated according:

$$\begin{aligned}\Delta O_{\text{opt}} &= |O_{\text{ref}} - O_{\text{opt}}| \\ \Delta O_{\text{jumpy}} &= |O_{\text{ref}} - O_{\text{jumpy}}|\end{aligned}\tag{9.6}$$

The values  $\Delta O_{\text{opt}}$  and  $\Delta O_{\text{jumpy}}$  represent the uncertainty bound wherein the overlay may vary due to statistical uncertainty of the model parameters. The percentage difference is determined according:

$$\xi = \left(1 - \frac{\Delta O_{\text{opt}}}{\Delta O_{\text{jumpy}}}\right) 100\%\tag{9.7}$$

The same procedure is executed for model parameter case 2 and the results are listed in Table 9.3.

**Table 9.3:** Reduction overlay uncertainty for model parameter case 1 and case 2.

CASE	$\xi$ (%)
1.	54.4
2.	9.4

Due to the optimal experiment of case 1 the overlay uncertainty is reduced more than half. The optimal experiment design is very effective. For the optimal experiment of case 2 the overlay uncertainty is reduced by approximately one tenth. The difference in result is expected when observing the confidence regions and the individual impact on overlay of the parameters which are given in Figure 8.4.



# 10

## Conclusion and Recommendation

In the present study, experiment optimization of the thermo-mechanical feedforward model in order to enhance model parameter accuracy for improved overlay performance is investigated. The first part of the study involves a thorough literature analysis supported by academic examples in order to acquire the main principles of experiment design. This part includes a review of the currently used thermo-mechanical feedforward model with an associated pre-investigation on model parameter accuracy. Multiple methods are described in order to obtain sensitivity information in an efficient manner and (non)linear regression techniques are described for model parameter estimation with the additional statistical information. At last, the principles of experiment design are explained and elucidated.

The second part of the study involves the application of experiment design on a simplified thermo-mechanical feedforward model. This part includes the description of the simplified thermo-mechanical model, followed by a comprehensive analysis. Thereafter, the standard available experiments are investigated on their behaviour and performance with respect to model parameter accuracy. At last, the experiment is optimized for the simplified thermo-mechanical model in order to improve model parameter accuracy during the estimation process.

### 10.1 Conclusion

By looking back at the research question *How to make the appearance in measurements of the physical parameters subject to calibration more orthogonal via optimization of the experiment that provides these measurements*, the following conclusions can be drawn:

- The main purpose of the research question is to improve model parameter accuracy during calibration by means of an optimized experiment in order to improve overlay performance. In the research question it is assumed that enhancement of orthogonal appearance of the model parameters in the measurements will lead to improved model parameter accuracy during calibration. However, in Chapter 4 and Chapter 9 it was seen that the orthogonal appearance (correlation) is not the only requirement for improved model parameter accuracy. The information matrix  $\mathbf{M}$  is a legit statistical measure for model parameter accuracy which combines correlation and information appearance in experimental measurements.
- Confidence regions of the model parameters can be constructed using the information matrix to obtain statistical information about the model parameters. For a two dimensional parameter case, it is possible to graphically illustrate confidence ellipsoids. These are useful to gain insight into the current accuracy of the model parameters. For higher parameter set dimension, multiple confidence ellipsoids can be constructed.

- Profiling techniques, as described in Chapter 5, are useful to investigate the non-linear behaviour of the confidence response of the model parameters and to determine whether linear approximations are allowable.
- In order to optimize an experiment for maximum model parameter accuracy the information matrix must be subjected to a suitable objective function. In Franceschini and Macchietto (2007) the standard D-, E- and A-optimality criterion are given, which works well for a two dimensional model parameter case. The A-optimality must be handled carefully, since it does not include correlation. In Franceschini and Macchietto (2008) more advanced objective function are described which appears to be very effective in reducing correlation and enhancing information content at the same time.
- Simulated annealing is an efficient direct optimization algorithm in order to optimize the combinatorial problem of the experiment, as is seen in Chapter 9. New intuitive experiment designs are achieved, which have improved performance, with respect to parameter accuracy, when compared with the standard available experiments.

With the usage of the information matrix, a suitable objective function and the simulated annealing algorithm it is possible to acquire an optimal experiment which can increase model parameter accuracy and, therefore, give a reduction of worst-case overlay error of 54.4% and 9.4% for parameter case 1 and 2 respectively, when considering a 95% confidence bound. In Table 10.1, the reduction overlay uncertainty  $\xi$  for model parameter case 1 and case 2 with respect to best performing standard experiments (std. exp.) is listed.

**Table 10.1:** Reduction overlay uncertainty  $\xi$  for model parameter case 1 and case 2 with respect to best performing standard experiments (std. exp.).

CASE	PAR.	OBJ. FNC.	$\xi$ (%)	STD. EXP.
1.	$c_p$ and IR/EUV	D-optimality	54.4	Jumpy
2.	$G_{zy}$ and IR/EUV	ACE1-optimality	9.4	Meander

## 10.2 Recommendation

It is shown that experiment design has potential to improve overlay performance. However, further research is useful in order to successfully optimize an experiment for the thermo-mechanical feedforward model of ASML and it may be useful to investigate more applications of the described experiment design techniques. Recommendation by the author are listed below.

- When using more than two model parameters, the experiment design techniques are still applicable. However, graphical interpretation of the confidence regions is more advantageous to obtain and understand. Multiple graphs of possible model parameter sets must be constructed, or the t-value may be adopted to obtain statistical significance of certain model parameters. The latter allows for easy comparison when the number of model parameters is high. In Franceschini and Macchietto (2008) an helpful example is issued for a three model parameter case.
- The neighbourhood functions of the simulated annealing algorithm described in Section 9.1 are simple and basic. The simulated annealing algorithm significantly benefits from a well designed neighbourhood functions which is able to take model characteristics into account.

This will speed up the converges and will improve the experiment design solutions. Currently only one permutation of a field routing or scan direction was chosen, but choosing more than two fields is possible.

- In Section 9.4 a new phenomena appeared as fractional bands between random generated experiments. Further research is recommended why this characteristics behaviour appears, which in turn may be useful to acquire even more knowledge about the system. This could be advantageous to design a sophisticated neighbourhood function that is used for the experiment design optimization.
- Experiment design is basically a method to improve model parameter information content from a measured responses. The same techniques could be used for sensor designs which are hampered by, for instance, a thermal signal. By optimizing the structure of the sensor according experiment design techniques, the true sensing accuracies of the sensor may be improved.
- In this study, the simulated annealing algorithm is used to optimize the difficult combinatorial experiment to improve model parameter accuracy. However, the same simulated annealing algorithm may be used to find an optimal exposure sequence, when producing chips, which reduces overlay.
- The experiment optimization was performed by a direct solver due to the discrete problem. Gradient-based solvers are beneficial in terms of computer efficiency. One solution approach to convert the discrete optimization problem into a continuous problem is by constructing  $N_{\text{step}}$  step responses in time, where each individual step response is associated to one field and indicating the time when it is exposed. Notice that the time instance of exposure is a continuous parameter. The number  $N_{\text{step}}$  equals the number of exposed fields. The routing is constructed by summing all the step responses and in this manner the time instance at which a field is exposed can be optimized. A penalty function must be added in order to prevent overlapping step responses, since only one field at a time can be exposed. Although this might seem like a convenient solution, the number of possible input parameter settings is unimaginable. Hence, there is a large possibility that the optimization problem is highly non-convex and will hamper gradient-based optimization techniques. From previous analysis, it is shown that only a very small set of input parameters could lead to a significant improvement on parameter estimation. Therefore, it is very unlikely that gradient-based optimization can find a global optimum due to the severe non-convexity, but it is also not likely that the found local optimum comes close to the global optimum. Therefore, this method is rejected as a possible solution for the optimization of the routing and scan directions for now, but further research may be useful.



# Bibliography

- Asprey, S. P. and Naka, Y. (1998). Mathematical problems in fitting kinetic models. *The Society of Chemical Engineers*.
- Atkinson, A. C. and Hunter, W. G. (2012). The design of experiments for parameter estimation. *Taylor Francis*.
- Bard, J. (1974). *Nonlinear Parameter Estimation*. New York: Academic Press.
- Bates, D. M. and Watts, D. G. (1981). A relative offset orthogonality convergence criterion for nonlinear least squares. *Technometrics*.
- Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. John Wiley and Sons.
- Bernaerts, K., Gysemans, K., Minh, T. N., and Impe, J. V. (2004). Optimal experiment design for cardinal values estimation: guidelines for data collection. *Elsevier*.
- Castelijns, H. (2014). Eds wafer heating feedforward model nxe3350. Technical report, Department of Research Mechatronics, ASML, Veldhoven, The Netherlands. Document id: D000229090.
- de Best, J. (2015). Pir: Thermo-mechanical modeling with a reticle heating case study. Technical report, Department of Research Mechatronics, ASML, Veldhoven, The Netherlands. Document id: D000391429.
- Dekking, F., Kraaikamp, C., Lopuhaä, H., and Meester, L. (2005). *A Modern Introduction to Probability and Statistics*. Springer.
- Eglese, R. (1989). Simulated annealing: A tool for operational research. *Elsevier*.
- Franceschini, G. and Macchietto, S. (2007). Model-based design of experiments for parameter precision: State of the art. *Elsevier*.
- Franceschini, G. and Macchietto, S. (2008). Novel anticorrelation criteria for model-based experiment design: Theory and formulations. *AIChE Journal*.
- Garcia, S. (1999). Experimental design optimization and thermophysical parameter estimation of composite materials using genetic algorithms. Master's thesis, Laboratoire de Thermocinetique de l'isitem.
- Goodwin, G. (1987). *Identification: experiment design*. Pergamon Press, Oxford.
- Meulenbroeks, R. (2014). How to derive from ansys a parametric thermo-mechanic-matrix for use in matlab. Technical report, Department of Research Mechatronics, ASML, Veldhoven, The Netherlands. Document id: D000274672.

- Munack, A. and Posten, C. (1989). Design of optimal dynamical experiments for parameter estimation. *American Control Conference*.
- Pauw, D. D. (2005). Optimal experimental design for calibration of bioprocess models: a validated software toolbox. Master's thesis, Ghent University, Belgium.
- Rixen, D. J. (2011a). *Mechanical Analysis for Engineering*. Department of Precision and Microsystems Engineering, Delft University of Technology.
- Rixen, D. J. (2011b). *Numerical methods in Engineering Dynamics*. Department of Precision and Microsystems Engineering.
- Rojas, C. R., Welsh, J. S., Goodwin, G. C., and Feuer, A. (2006). Robust optimal experiment design for system identification. *Elsevier*.
- Romeijn, H. E. and Smith, R. L. (1993). Simulated annealing for constrained global optimization. *Journal of Global Optimization*.
- Rook, R. (2014). Wafer heating model. Technical report, Department of Research Mechatronics, ASML, Veldhoven, The Netherlands.
- Ruckstuhl, A. (2010). Introduction to nonlinear regression.
- Schwab, A. and Delhaes, G. M. (2009). *Multibody Dynamics B - Lecture Notes*. Laboratory for Engineering Mechanics, Delft University of Technology.
- Szu, H. and Hartley, R. (1987). Fast simulated annealing. *Physics Letters A*.
- Tinnemans, P. (2010). Eds lsqnonlin ptin algorithm. Technical report, Department of Research Mechatronics, ASML, Veldhoven, The Netherlands. Document id: D000140314.
- van de Wal, M. (2014). Gid: Overview of nanometer thermal control challenges and thermal expertise for asml scanners. Technical report, Department of Research Mechatronics, ASML, Veldhoven, The Netherlands. Document id: D000276573.
- van der Meulen, S. (2015). Pir: Feasibility study for overlay-based euv thermal feedforward model calibration. Technical report, Department of Research Mechatronics, ASML, Veldhoven, The Netherlands. Document id: D000301068.
- Vanrolleghem, P., Daele, M. V., and Dochain, D. (1995). Practical identifiability of a biokinetic model of activated sludge respiration. *Water Research*.
- Veeke, H. P., Ottjes, J. A., and Lodewijks, G. (2008). *The Delft Systems Approach*. Springer.
- Vuik, C. and Lahaye, D. (2014). *Scientific Computing*. Delft Institute of Applied Mathematics.
- Vuik, C., Vermolen, F., van Gijzen, M., and Vuik, M. (2006). *Numerical Methods for Ordinary Differential Equations*. Delft Academic Press.
- Watts, D. G. (2010). *Essential Numerical Computer Methods*, chapter 2. Academic Press.
- Zullo, L. (1991). Computer aided design of experiments. an engineering approach. Master's thesis, University of London, UK.

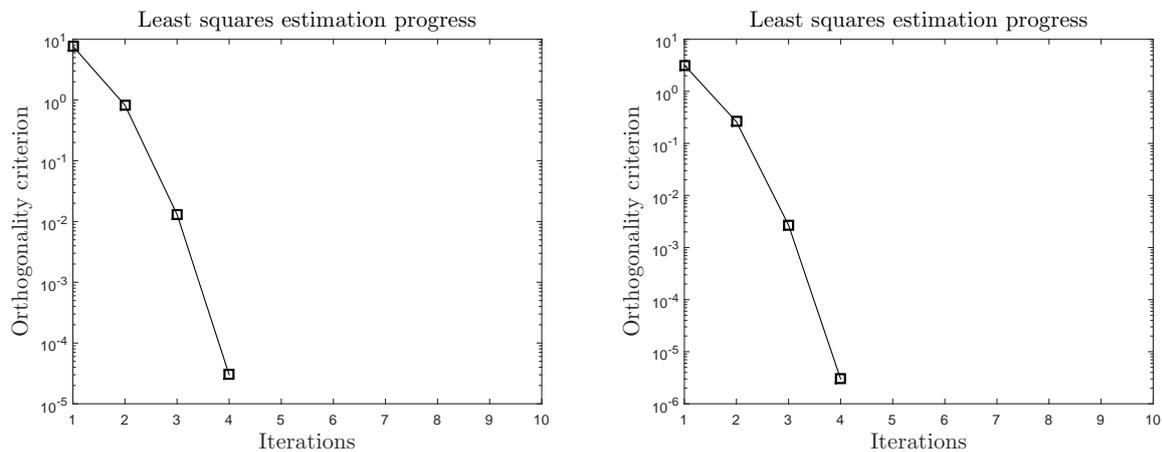
# Appendix



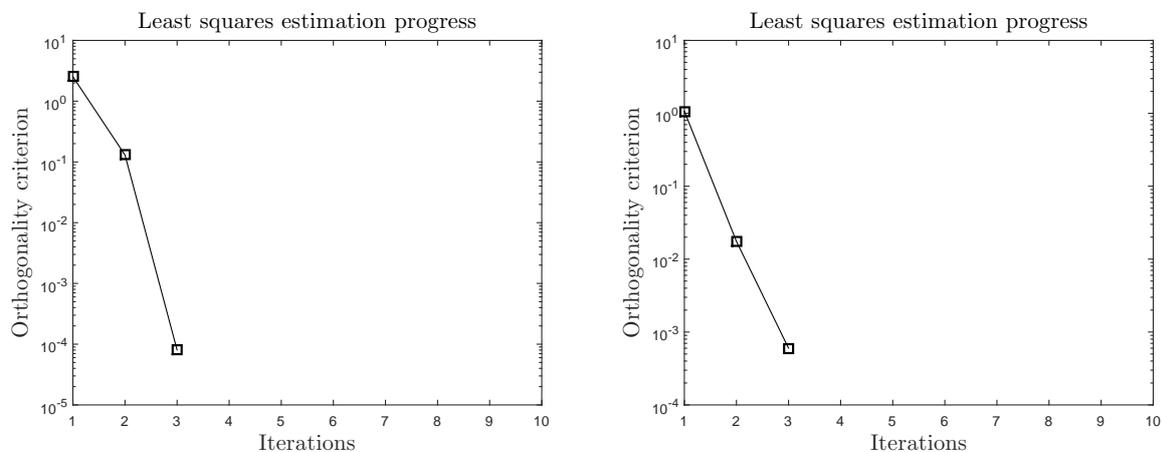
# A

## Figures In Report

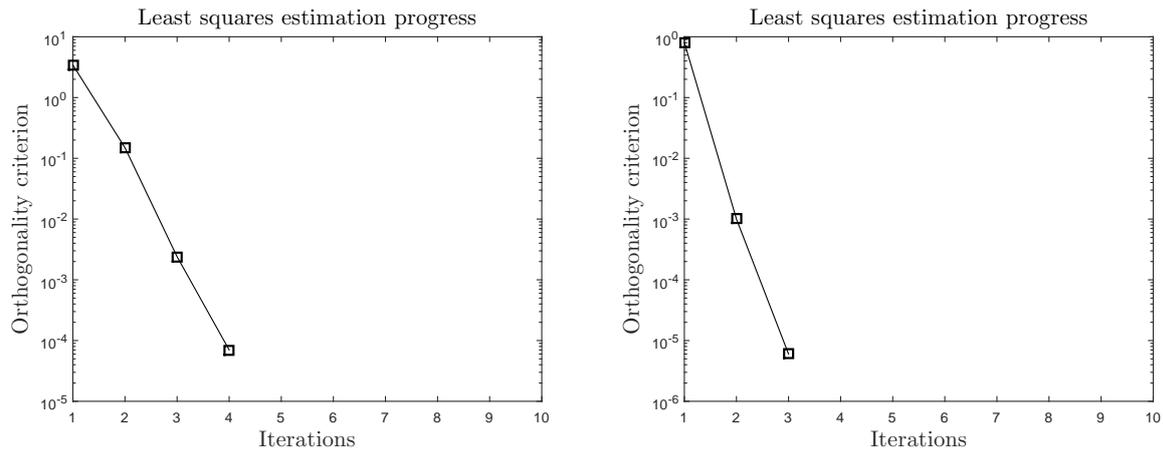
### A.1 Least Squares Estimation Progress Standard Experiments



**Figure A.1:** Least squares estimation progress for meander routing. Left figure represents iteration progress for case 1, the figure on the right indicates the iteration progress for case 2.

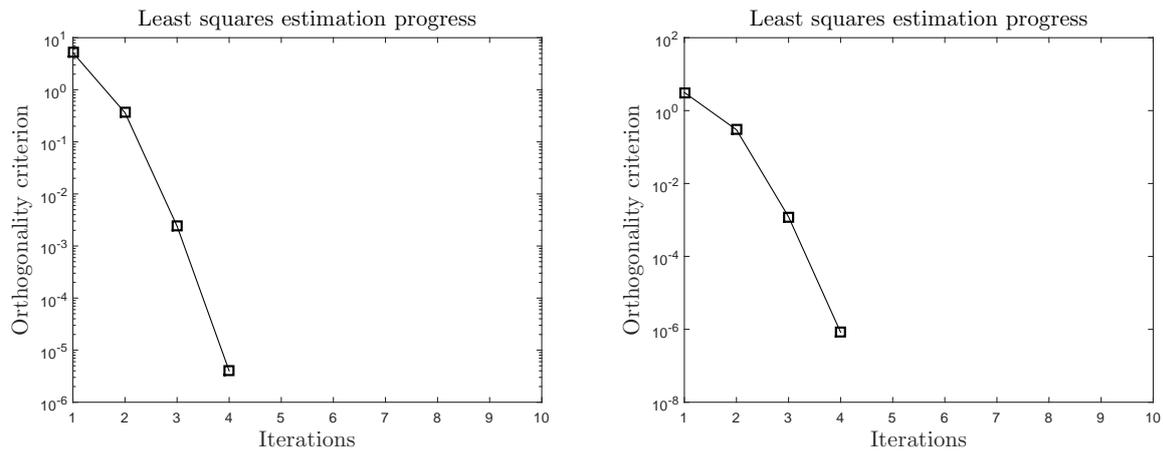


**Figure A.2:** Least squares estimation progress for spiral routing. Left figure represents iteration progress for case 1, the figure on the right indicates the iteration progress for case 2.



**Figure A.3:** Least squares estimation progress for jumpy routing. Left figure represents iteration progress for case 1, the figure on the right indicates the iteration progress for case 2.

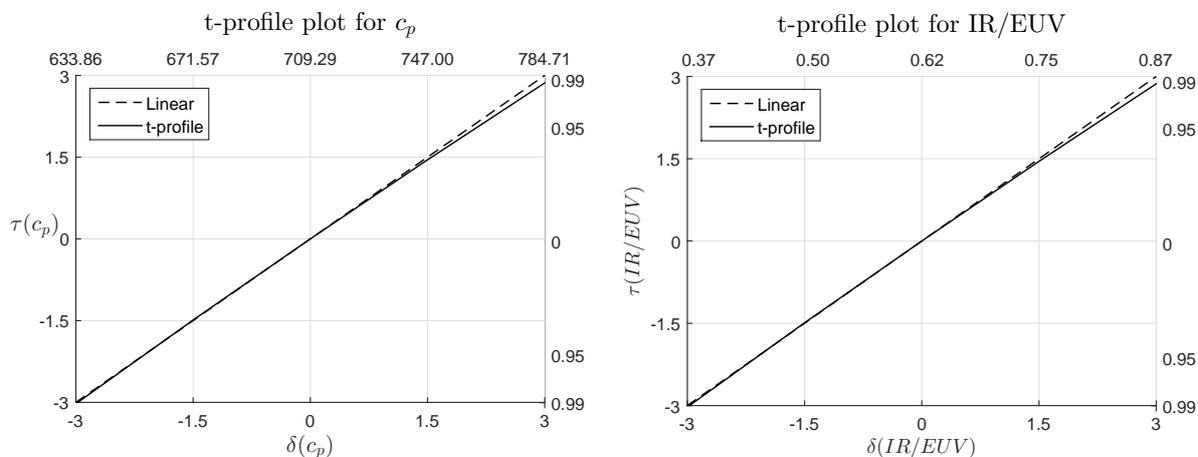
## A.2 Least Squares Estimation Progress Optimal Experiments



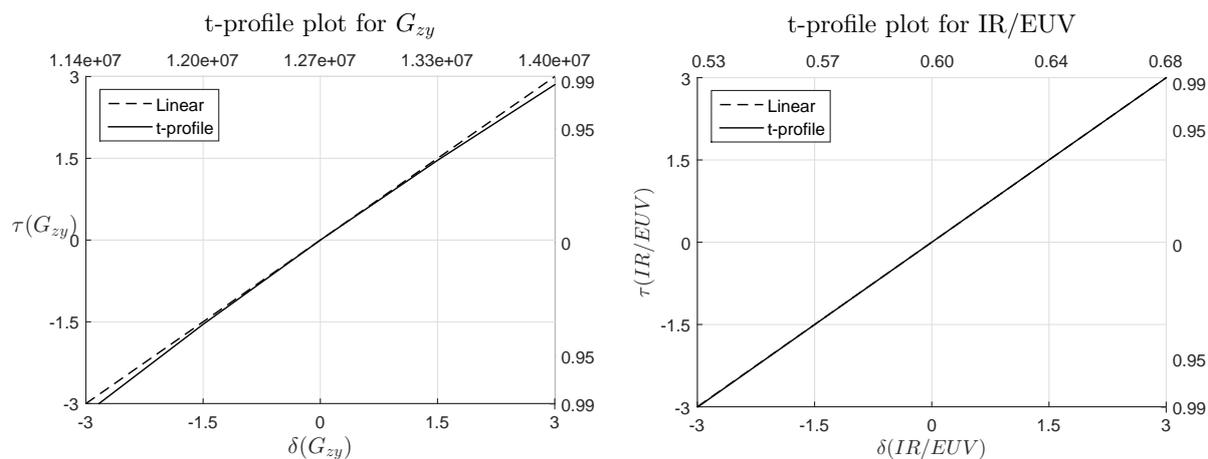
**Figure A.4:** Least squares estimation process for optimal experiments. Left figure represents least squares process for case 1, the figure on the right indicates the least squares process for case 2.

### A.3 T-profiling of Standard Experiments

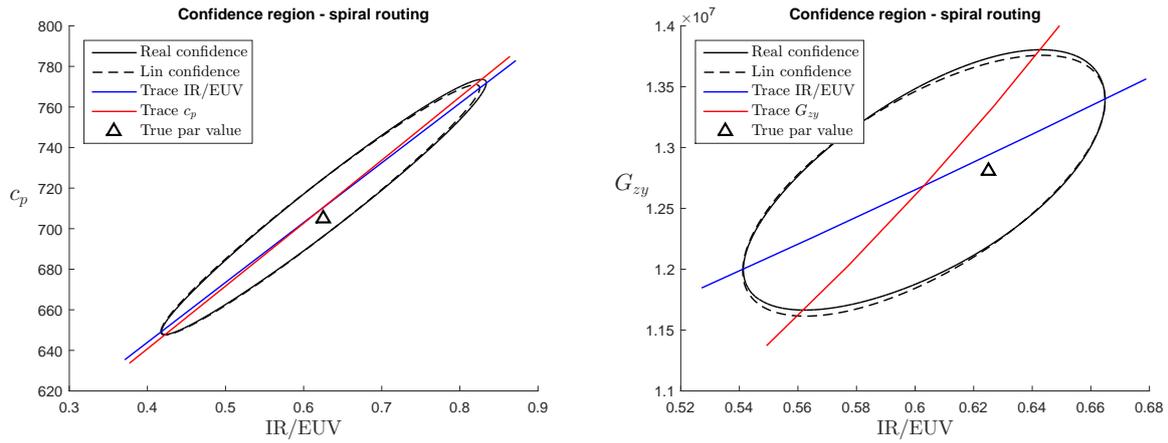
#### A.3.1 Spiral Experiment



**Figure A.5:** T-profiling spiral routing for parameter case 1. The heat capacitance of the substrate and the IR/EUV ratio behave non-linear in their confidence response due to high correlation.

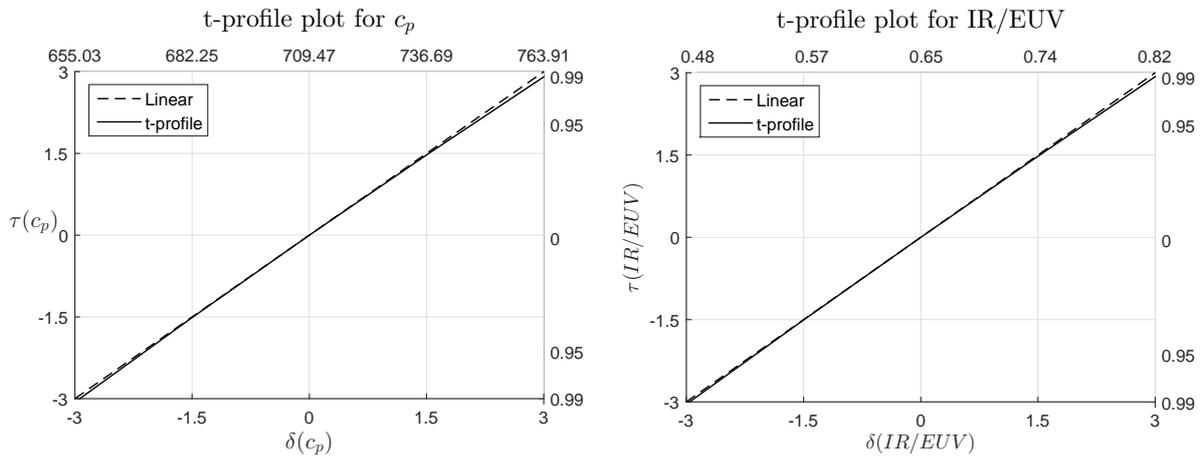


**Figure A.6:** T-profiling spiral routing for parameter case 2. The tangential burl stiffness behaves non-linear in its confidence response, whereas the IR/EUV ratio behaves almost linear in its confidence response due to low to moderate correlation.

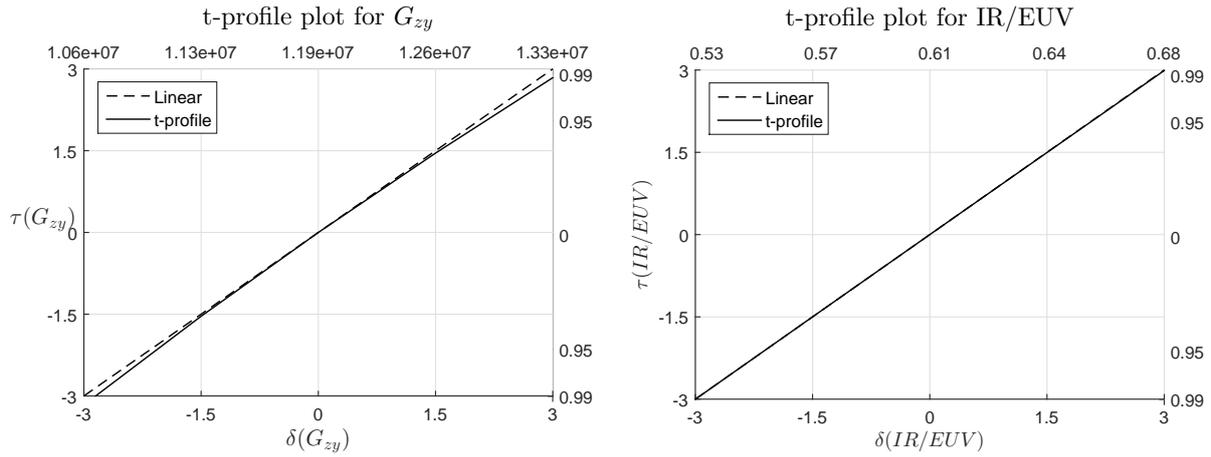


**Figure A.7:** Confidence regions of 95% for spiral routing for both cases. The linearly approximated confidence regions are sufficient to represent the true confidence regions.

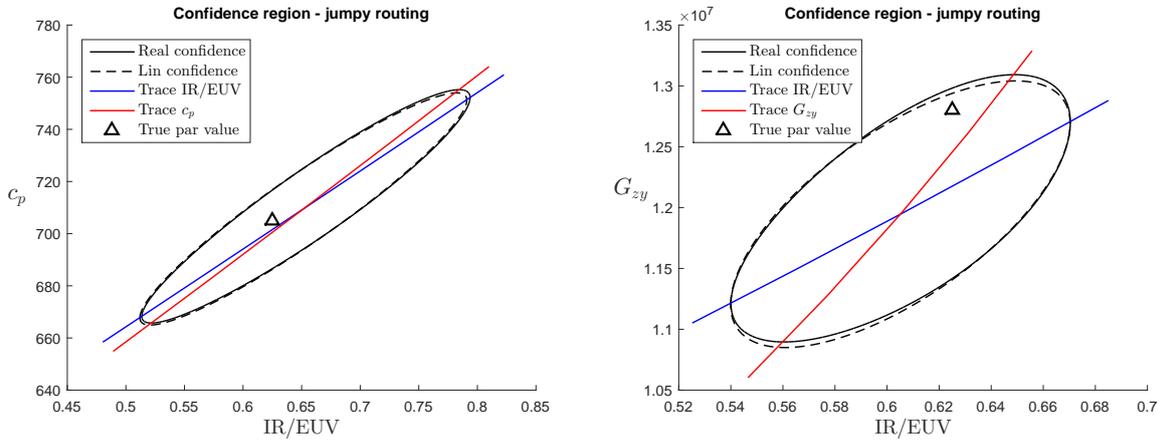
### A.3.2 Jumpy Experiment



**Figure A.8:** T-profiling jumpy routing for parameter case 1. The heat capacitance of the substrate and the IR/EUV ratio behave non-linear in their confidence response due to high correlation.



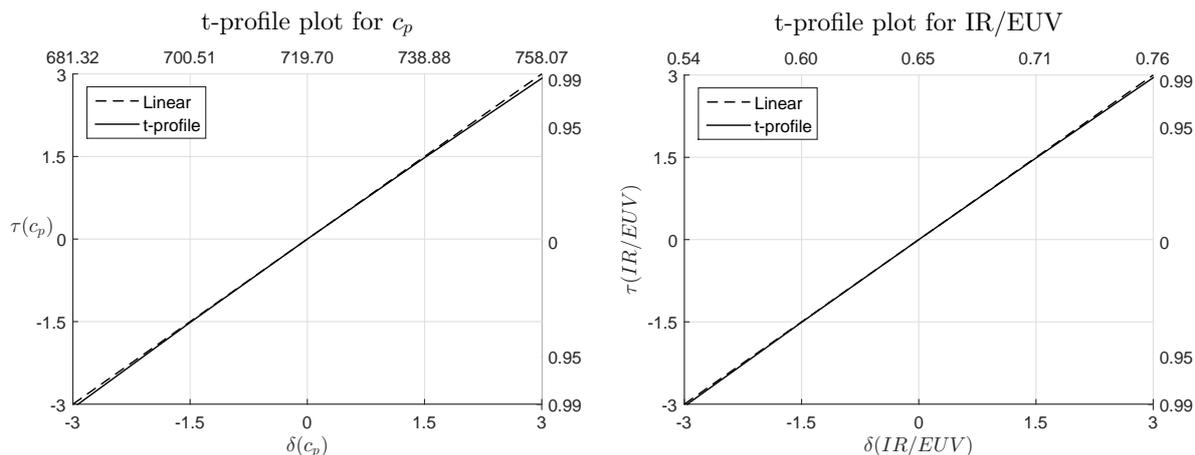
**Figure A.9:** T-profiling jumpy routing for parameter case 2. The tangential burl stiffness behaves non-linear in its confidence response, whereas the IR/EUV ratio behaves almost linear in its confidence response due to low to moderate correlation.



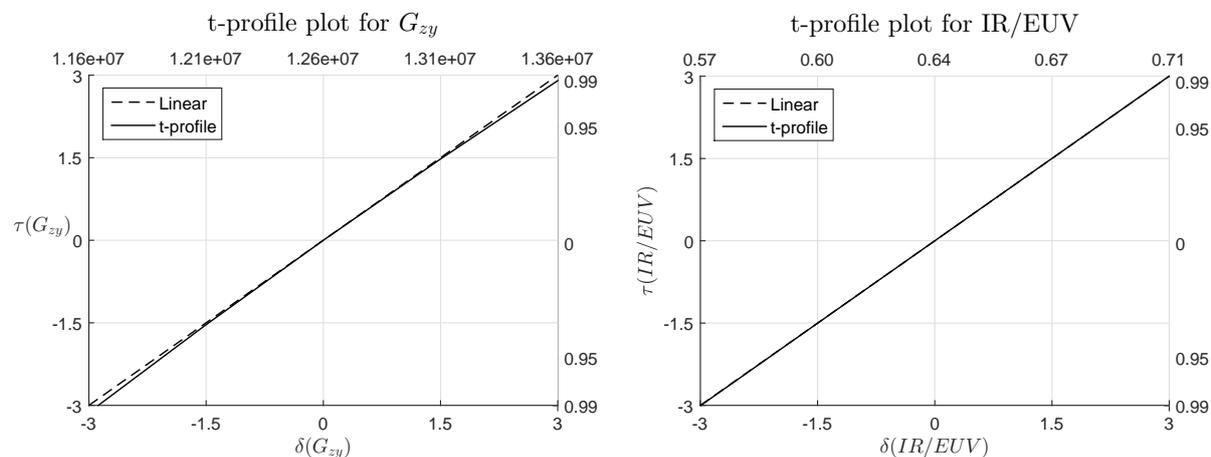
**Figure A.10:** Confidence regions of 95% for jumpy routing for both cases. The linearly approximated confidence regions are sufficient to represent the true confidence regions.

### A.4 T-profiling of Standard Experiments

A comparison is made between the optimal and standard experiments with the associated performance with respect to model parameter accuracy.



**Figure A.11:** T-profiling optimal routing for parameter case 1. The heat capacitance of the substrate and the IR/EUV ratio behaves non-linear in their confidence response due to high correlation.



**Figure A.12:** T-profiling optimal routing for parameter case 2. The tangential burl stiffness behaves non-linear in its confidence response, whereas the IR/EUV ratio behaves linear in its confidence response due to low to moderate correlation.

In the figures A.11 and A.12 the t-profiling is illustrated for both optimal experiments of model parameter case 1 and 2. From the figures, it is immediately evident that the confidence response behaves almost linear in the model parameters. Small deviations from the (dotted) reference line are visible and indicating the presence of non-linear behaviour, but since this effect is so small the non-linearity may be neglected.

When observing Figure A.12, it is evident that the confidence response behaves approximately linear in the model parameters. However, two different t-profile plots are visible. These parameters are less correlated than the parameters in case 1. The tangential burl stiffness is a non-linear parameter, as was observed in Chapter 8 and the non-linearity can be confirmed by the deviating t-profile plot. The IR/EUV ratio is a linear model parameter with respect to overlay. Since the correlation is low to moderate, this linear behaviour is transmitted to the confidence t-profile response, as can be seen in the right figure. Note, due to some correlation of the latter t-profile response, it is not exactly linear but it is hardly visible in the figure.

# B

## Shape Functions

The shape functions of the 20-node brick are given by:

$$N_1(x, y, z) = -\frac{1}{8}(1 - \xi)(1 - \eta)(1 - \mu)(2 + \xi + \eta + \mu)$$

$$N_2(x, y, z) = -\frac{1}{8}(1 + \xi)(1 - \eta)(1 - \mu)(2 - \xi + \eta + \mu)$$

$$N_3(x, y, z) = -\frac{1}{8}(1 + \xi)(1 + \eta)(1 - \mu)(2 - \xi - \eta + \mu)$$

$$N_4(x, y, z) = -\frac{1}{8}(1 - \xi)(1 + \eta)(1 - \mu)(2 + \xi - \eta + \mu)$$

$$N_5(x, y, z) = -\frac{1}{8}(1 - \xi)(1 - \eta)(1 + \mu)(2 + \xi + \eta - \mu)$$

$$N_6(x, y, z) = -\frac{1}{8}(1 + \xi)(1 - \eta)(1 + \mu)(2 - \xi + \eta - \mu)$$

$$N_7(x, y, z) = -\frac{1}{8}(1 + \xi)(1 + \eta)(1 + \mu)(2 - \xi - \eta - \mu)$$

$$N_8(x, y, z) = -\frac{1}{8}(1 - \xi)(1 + \eta)(1 + \mu)(2 + \xi - \eta - \mu)$$

$$N_9(x, y, z) = \frac{1}{4}(1 - \xi)(1 + \xi)(1 - \eta)(1 - \mu)$$

$$N_{10}(x, y, z) = \frac{1}{4}(1 + \xi)(1 - \eta)(1 + \eta)(1 - \mu)$$

$$N_{11}(x, y, z) = \frac{1}{4}(1 - \xi)(1 + \xi)(1 + \eta)(1 - \mu)$$

$$N_{12}(x, y, z) = \frac{1}{4}(1 - \xi)(1 - \eta)(1 + \eta)(1 - \mu)$$

$$N_{13}(x, y, z) = \frac{1}{4}(1 - \xi)(1 + \xi)(1 - \eta)(1 + \mu)$$

$$N_{14}(x, y, z) = \frac{1}{4}(1 + \xi)(1 - \eta)(1 + \eta)(1 + \mu)$$

The shape functions continuous at the next page.

$$N_{15}(x, y, z) = \frac{1}{4}(1 - \xi)(1 + \xi)(1 + \eta)(1 + \mu)$$

$$N_{16}(x, y, z) = \frac{1}{4}(1 - \xi)(1 - \eta)(1 + \eta)(1 + \mu)$$

$$N_{17}(x, y, z) = \frac{1}{4}(1 - \xi)(1 - \eta)(1 - \mu)(1 + \mu)$$

$$N_{18}(x, y, z) = \frac{1}{4}(1 + \xi)(1 - \eta)(1 - \mu)(1 + \mu)$$

$$N_{19}(x, y, z) = \frac{1}{4}(1 + \xi)(1 + \eta)(1 - \mu)(1 + \mu)$$

$$N_{20}(x, y, z) = \frac{1}{4}(1 - \xi)(1 + \eta)(1 - \mu)(1 + \mu)$$

The local dimensionless coordinates  $\xi$ ,  $\eta$  and  $\mu$  are used. In the case of the rectangular cuboids, with a length  $L$ , a width  $W$  and a height  $H$ , the dimensionless coordinates are defined as:

$$\xi = \frac{2x}{L} - 1$$

$$\eta = \frac{2y}{W} - 1$$

$$\mu = \frac{2z}{H} - 1$$

# C

## Interpolating Profile Pair Contours

In Bates and Watts (1988), a method for interpolated profile pair contours is described. This allows for construction of approximated non-linear confidence regions. First, the profile traces  $\tilde{p}_p$  and  $\tilde{p}_q$  are transformed to  $\tau_p$  and  $\tau_q$  coordinates using cubic splines, see Figure C.1. This transformation is possible by using the previous obtained profile parameter values and its associated t-values. Using the  $\tau$  coordinates, the likelihood surface is transformed so that the surface is nearly a paraboloid with elliptical contours. It is easy to interpolate points on these near-ellipses as will be seen. When the model has low rate of non-linearity, the transformed trace vectors in  $\tau$  coordinates are nearly straight lines through the origin.

To interpolate a particular confidence contour, the  $\tau$  coordinates are scaled by dividing by  $\sqrt{MF(M, N - M; \alpha)}$  so that a nominal  $1 - \alpha$  joint confidence contour in the scaled  $\tau$  coordinates is bounded by the square  $-1 \leq \tau_p$  and  $\tau_q \leq 1$ . If the contour were an ellipse, it could be represented in the scaled coordinates in the parametric form:

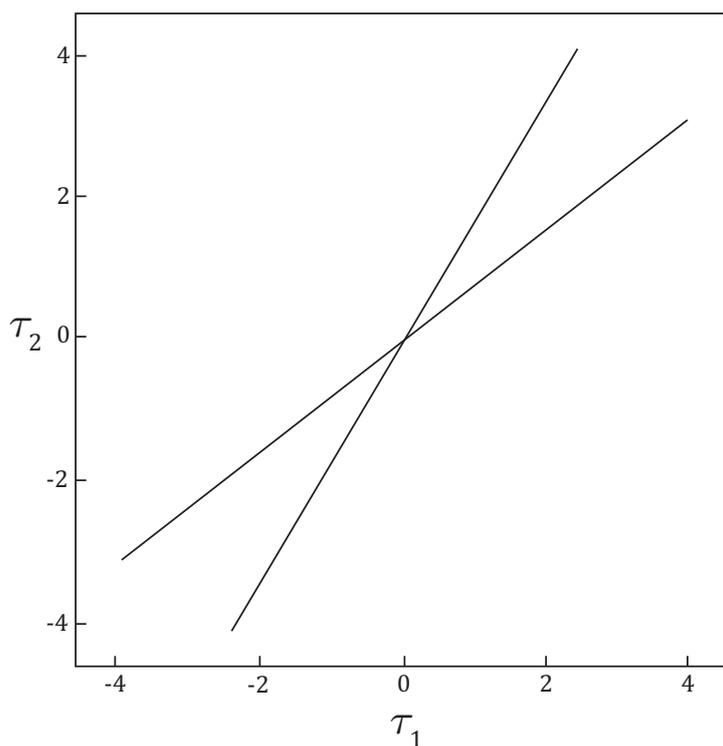
$$\begin{aligned}\tau_p &= \cos(a + d/2) \\ \tau_q &= \cos(a - d/2)\end{aligned}\tag{C.1}$$

Here, the angle  $a$  goes from  $-\pi$  to  $\pi$  and the phase  $d$  is a constant. If the contour is not elliptical, the phase  $d$  is not constant. The interpolation goes as follows: choose the following set of points  $(\tau_{p,r}, \tau_{q,r})$ , with  $r = 1, \dots, 4$ , where this set of points corresponds to the bounds of the confidence region. For instance, point  $r = 1$  is  $(\tau_{p,1}, \tau_{q,1}) = (1.00, 0.80)$  and point  $r = 2$  is  $(\tau_{p,2}, \tau_{q,2}) = (0.79, 1.00)$ , which are derived from figures like Figure C.1. Note, the points are taken from the scaled  $\tau$  coordinates. In order to interpolate between these points, which would give the confidence region, the arccosines  $s_{p,r} = \arccos(\tau_{p,r})$  and  $s_{q,r} = \arccos(\tau_{q,r})$  could be calculated in order to form the averages and differences. Subsequently, the averages and differences are used to obtain the angle and the phase. An example from Bates and Watts (1988) is used to illustrate the principle of selecting the scaled  $\tau$  coordinate points  $r$ , calculating the arccosine, deriving the average and difference and obtaining the angle and phase. The obtained results are listed in Table C.1.

**Table C.1:** Example values from Bates and Watts (1988) to calculate the angle and phase.

Scaled		Arccosine				Angle	Phase
$\tau_1$	$\tau_2$	1	2	AVG.	DIFF.	$a$	$b$
1.000	0.801	0.000	0.641	0.321	-0.641	-0.321	0.641
0.795	1.000	0.651	0.000	0.326	0.651	0.326	0.651
-1.000	-0.762	3.142	2.437	2.789	0.704	2.789	0.704
-0.769	-1.000	2.448	3.142	2.795	-0.693	-2.795	0.693

Even for an ellipse, the differences of the arccosines will vary in sign and the averages of the arccosines will lie between 0 and  $\pi$ , because the arccosine transformation only yields values in the range 0 to  $\pi$ . To obtain suitable values for  $a$  and  $d$  it must be noted that, since  $\cos(-x) = \cos(x)$ , equation (C.1) will yield in the same  $\tau_p$  and  $\tau_q$  if the sign of the average *and* the sign of the difference is reversed. Therefore, reverse the sign of any negative difference and its corresponding average to give  $a$  and  $d$  values, see Table C.1, suitable for interpolation with a periodic spline. When the periodic spline interpolation is performed onto  $a$  and  $d$ , the values can be used to transform back to  $\tau_p$  and  $\tau_q$  using equation (C.1) and hence the confidence contour can be plotted.

**Figure C.1:** Profile traces in the  $\tau$  coordinates.

