Delft University of Technology

Inverse Electromagnetics for EUV mask metrology and inspection

Ansuinelli, P.

DOI
10.4233/uuid:7c5d0326-c293-4f5a-a926-edf7358f6510

Publication date
2022

Document Version
Final published version

Citation (APA)
Ansuinelli, P. (2022). *Inverse Electromagnetics for EUV mask metrology and inspection*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:7c5d0326-c293-4f5a-a926-edf7358f6510

Inverse Electromagnetics for EUV mask metrology and inspection

Inverse Electromagnetics for EUV mask metrology and inspection

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
Chair of the Board for Doctorates,
to be defended publicly on
Wednesday 9th of March 2022 at 12.30

by

Paolo ANSUINELLI

Master of Science in Nanotechnology Engineering,
Sapienza University of Rome, Italy,
born in Rome, Italy.

This dissertation has been approved by the promotors.

Promotor: Prof.dr. W.M.J.M. Coene
Promotor: Prof.dr. H.P. Urbach

Composition of the doctoral committee:

Rector Magnificus,	chairman
Prof.dr. W.M.J.M. Coene,	Delft University of Technology, promotor
Prof.dr. H.P. Urbach,	Delft University of Technology, promotor
Independent members:	
Prof.dr. M.H.G. Verhaegen,	Delft University of Technology
Prof.dr. G. Vdovine,	Delft University of Technology
Prof.dr. P. Thibault,	University of Trieste, Italy
Dr. A. Erdmann,	Fraunhofer Institute IISB, Germany
Dr. Y. Ekinci,	Paul Scherrer Institute, Switzerland
Prof.dr. B. Rieger,	Delft University of Technology, reserve member

The research presented in this thesis was funded by the European Commission, Horizon 2020 (H2020) programme, through the Marie Skłodowska-Curie Actions (MSCA) via the NOLOSS project (675745).

Keywords: inverse problems, scatterometry, phase retrieval, EUV mask inspection and metrology
Printed by: ProefschriftMaken, Utrecht

ISBN 978-94-6423-700-9

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Alla mia famiglia
To my family

Contents

Summary

The importance of inverse problems is paramount in science and physics because their solution provides information about parameters that cannot be directly observed. This thesis discusses and details the application of a few inverse methods in optical imaging, metrology and inspection of lithographic targets, particularly patterned structures on top of an extreme ultraviolet (EUV) lithographic mask.

Chapter 1 is an introductory chapter. The defining aspects and characteristics of inverse problems are presented, explained and illustrated with examples. The concept of *ill–posedness* is discussed and its cure, regularization theory, is described and introduced. Particular emphasis is placed on the importance of the inclusion of *a–priori* information in inverse problem solving. The distinction among variational and statistical methods is outlined and relevant references are highlighted. We conclude this chapter by discussing a few optimization methods required for the minimization of a given loss functional.

Chapter 2 focuses on scatterometry, the industrial workhorse for fast, non–destructive metrology of periodic structures realized in the lithographic process. We discuss the three fundamental building blocks of scatterometry: data acquisition, modeling of light matter interaction and methods to solve the underlying inverse problem. We subsequently focus on the problem of dimensionality reduction as a cure for the *curse of dimensionality*. After describing a few fundamental methods for model reduction we introduce a novel algorithm for automatic feature selection. We discuss the application of this method to periodic and isolated scatterers and we compare its outcome to the one given by sensitivity analysis methods.

Chapter 3 is dedicated to the introduction of the problem of extreme ultraviolet (EUV) mask imaging and inspection. We discuss a few existing solutions for the imaging of EUV masks and we subsequently focus on a particular approach: imaging via iterative *phase retrieval* algorithms. We present and discuss the physical origin of the phase problem and we detail its mathematical formulation. We focus on the algorithms and discuss single intensity phase retrieval methods and ptychography, a phase retrieval method where one processes multiple diffraction patterns acquired by means of partially overlapping successive illuminations of a scattering object by a certain illumination function, which is typically a focused light field or probe.. We apply these methods to the reconstruction of EUV mask layouts and we reason that a possible improvement may be obtained by including *a priori* information in the phase retrieval optimization algorithm.

Chapter 4 focuses on developing methods to improve the ptychographic inspection of EUV masks by including prior information in the optimization algorithm. We discuss how the nominal layout of the EUV mask can be used as an input to a rigorous forward Maxwell solver to obtain a precise estimation of the exit wave for the nominal mask and how this can be subsequently employed as a prior dur-

ing the reconstruction of the defective masks. We demonstrate the feasibility of this method and we show that the inclusion of this physically sound prior results in the reconstruction of finer and subwavelength extrusion and intrusion absorber defects in EUV mask layouts. Different update rules, obtained by employing variational and statistical methods, are given and discussed. We further analyze and compare the introduced "hand–crafted" prior with popular sparsity priors (total variation and Lasso). We end the chapter by noting that further improvements in the reconstruction may be obtained by developing more sophisticated approaches to the ptychographic imaging of mask layouts. One of these is polarization sensitive imaging.

Chapter 5 discusses an approach for polarization sensitive imaging of semiconductor structures. We show how to mathematically describe the light matter interaction for the considered structures and how the recorded intensity is directly related, in this particular case, to the Jones matrix of the scatterer. We present results obtained with this approach for a cluster of nanostructures whose critical dimensions are of the order of the wavelength and discuss our findings.

Chapter 6 presents ideas for future work. We introduce a method that can potentially yield subwavelength reconstruction of targets: total internal reflection phase retrieval and ptychography. After a brief discussion on the physics of this method we highlight potential research directions.

Finally, we offer a closing and brief summary of the findings of this thesis in (7).

1

Introduction

In the words of the mathematician J.B. Keller [1] "two problems are *inverses* of one another if the formulation of each involves all or part of the solution of the other. Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other has never been studied and is not so well understood. In such cases, the former is called the *direct problem*, while the latter is the *inverse problem*". For a physicist the distinction among direct and inverse problem can be thought of in terms of causality. In the direct problem (often also referred to as the forward problem) we follow the cause–effect chain meaning that we try to establish or to compute the consequences of some known causes, while in the inverse problem we "walk backwards" and we try to retrieve unknown causes from known consequences. Solving the inverse problem is usually more challenging than solving the forward one. For instance in electromagnetics we can avail ourselves of a theorem known as the uniqueness theorem. Loosely stated it says that if we have a source of light in a certain medium, and we know the boundary conditions that are to be satisfied by the resulting field, then we can safely compute the field distribution that is the unique solution – the *physical* solution – of the boundary value problem. Unfortunately this theorem does not hold when read backwards: given a certain measured field distribution and the knowledge of the boundary conditions we cannot uniquely determine the distribution of the light source which has generated the observed field.

Generally stated, inverse problems are challenging because they are ill–posed. A mathematical problem is said to be *ill–posed* when it is not *well–posed*. According to Hadamard [2], given two Banach spaces $\mathcal{B}_1$ and $\mathcal{B}_2$ and two vectors $x \in \mathcal{B}_1$ and $y \in \mathcal{B}_2$, an operator equation

$$F(x) = y \tag{1.1}$$

with $F : \mathcal{B}_1 \to \mathcal{B}_2$, is well–posed when the following criteria are simultaneously fulfilled:

- for any $y \in \mathcal{B}_2$, a solution $x \in \mathcal{B}_1$ exists that satisfies:
- the solution x is unique; and
- the solution is stable with respect to perturbations in y: if $Fx_* = y_*$ and $Fx = y$, then $x \to x_*$ when $y \to y_*$.

1.1. Examples

We consider two examples that can illustrate the difficulties that arise when treating inverse problems and how they are related to the Hadamard conditions presented above.

1.1.1. Example 1: inverse scattering

We consider the interaction of a known incident monochromatic light field with a certain object [3]. We suppose that the scattering object is embedded within a certain background material. We assume, without loss of generality, that both the background material and the scatterer are isotropic. We introduce the Cartesian position vector $\mathbf{r} = [x \quad y \quad z]^T$ and the relative electric permittivities of the background and of the scattering media, $\epsilon_b = n_b^2$ and $\epsilon_s(\mathbf{r}) = n_s^2$, respectively. If we call Ω the region of space that is occupied by the scatterer we define:

$$\epsilon(\mathbf{r}) = \begin{cases} \epsilon_b, & \text{for } \mathbf{r} \notin \Omega \\ \epsilon_s(\mathbf{r}), & \text{for } \mathbf{r} \in \Omega. \end{cases} \tag{1.2}$$

Further, we introduce

$$\begin{aligned} k_b &= \frac{2\pi n_b}{\lambda}, \\ n(\mathbf{r}) &= n_b + n_s(\mathbf{r}), \\ k(\mathbf{r}) &= k_b \frac{n(\mathbf{r})}{n_b} = k_b\left(1 + \frac{n_s(\mathbf{r})}{n_b}\right), \end{aligned} \tag{1.3}$$

where k_b is the wavenumber in the background material, λ is the wavelength in vacuum and $k(\mathbf{r})$ is the wave number that depends of the locally varying index of refraction $n(\mathbf{r})$. Assuming the electric field to be time harmonic, and assuming that $n_s(\mathbf{r}) = 0$, the scalar total field $U(\mathbf{r})$ is a solution of the homogeneous Helmholtz equation:

$$(\nabla^2 + k_b^2)U(\mathbf{r}) = 0. \tag{1.4}$$

When $n_s(\mathbf{r}) \neq 0$ Eq. (1.4) can be rewritten as the inhomogeneous Helmholtz equation. We start by defining $D(\mathbf{r})$ as the support of the scatterer:

$$D(\mathbf{r}) = \begin{cases} 0, & \text{for } \mathbf{r} \notin \Omega \\ 1, & \text{for } \mathbf{r} \in \Omega. \end{cases} \tag{1.5}$$

Given a generic position vector $\mathbf{r}$ and by noting that $D(\mathbf{r})$ is a binary function (i.e. $(1 - D)^n = 1 - D$) we can rewrite the wavevector, and its square, as:

$$\begin{aligned} k(\mathbf{r}) &= k_b(1 - D(\mathbf{r})) + k_s(\mathbf{r})D(\mathbf{r}), \\ k^2(\mathbf{r}) &= k_b^2(1 - D(\mathbf{r})) + k_s^2(\mathbf{r})D(\mathbf{r}). \end{aligned} \tag{1.6}$$

Given Eq. (1.6) the Helmholtz equation can be rewritten

$$\begin{aligned} (\nabla^2 + k^2(\mathbf{r}))U(\mathbf{r}) &= 0, \\ (\nabla^2 + k_b^2)U(\mathbf{r}) &= k_b^2\left(1 - \frac{k_s^2(\mathbf{r})}{k_b^2}\right)D(\mathbf{r})U(\mathbf{r}), \\ (\nabla^2 + k_b^2)U(\mathbf{r}) &= \chi(\mathbf{r})U(\mathbf{r}), \\ \text{with} \quad \chi(\mathbf{r}) &= k_b^2\left[1 - \left(\frac{n(\mathbf{r})}{n_b}\right)^2\right]D(\mathbf{r}). \end{aligned} \tag{1.7}$$

where $\chi(\mathbf{r})$ is the contrast function given by the scatterer.

Eq. (1.7) states that the scatterer acts as a source for the electromagnetic field. Owing to the linearity of Maxwell's equations the field originated by a certain object can be thought of as the superposition of the fields originated by elementary impulses. We introduce the Green function, $G(\mathbf{r},\mathbf{r}')$, as the solution of

$$(\nabla^2 + k_b^2)G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}'), \tag{1.8}$$

where $\delta(\mathbf{r})$ is the Dirac delta function defined to have the property that

$$\int \delta(\mathbf{r}-\mathbf{r}')f(\mathbf{r}')d^3r' = f(\mathbf{r}). \tag{1.9}$$

In free space it holds that

$$G(\mathbf{r},\mathbf{r}') = \frac{e^{ik_b|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}. \tag{1.10}$$

We now define the incident field, $U_{in}(\mathbf{r})$, as the field that is present when there is no scatterer. By this definition $U_{in}(\mathbf{r})$ satisfies the homogeneous Helmholtz equation (1.4). The total field is then given by the summation of the incident and the scattered fields:

$$U(\mathbf{r}) = U_{in}(\mathbf{r}) + U_s(\mathbf{r}). \tag{1.11}$$

We can now derive from Eqs. (1.4) – (1.11) the expressions for both the scattered and the total fields due to sources in Ω as observed at a certain position $\mathbf{r}$:

$$\begin{aligned} U_s(\mathbf{r}) &= \int_\Omega G(\mathbf{r}-\mathbf{r}')\chi(\mathbf{r}')U(\mathbf{r}')d^3r', \\ U(\mathbf{r}) &= U_{in}(\mathbf{r}) + \int_\Omega G(\mathbf{r}-\mathbf{r}')\chi(\mathbf{r}')U(\mathbf{r}')d^3r'. \end{aligned} \tag{1.12}$$

Equation (1.12) is an implicit equation for the scattered field and it has no generic analytical solution. The field generated by scatterers of a given geometry is usually computed numerically using dedicated solvers.

In the inverse scattering problem we usually measure the intensity or the power of the scattered field Eq. (1.12) in the far zone and we apply an inversion method to retrieve from it the optical contrast $\chi(\mathbf{r})$. The algorithms that attempt to solve this problem often assume that the Born approximation is valid. The Born approximation is the first iteration of the Born series which is obtained by iteratively replacing the field $U(\mathbf{r})$ in the Lippmann–Schwinger equation – Eq. (1.12) – with its successive evaluations. Letting $n = 1, 2, 3, ..., \infty$, the Born series evaluates the field at the nth iteration as:

$$U_n(\mathbf{r}) = U_{in}(\mathbf{r}) + \int_\Omega G(\mathbf{r}-\mathbf{r}')\chi(\mathbf{r}')U_{n-1}(\mathbf{r}')d^3r'. \tag{1.13}$$

At the first iteration the total field $U(\mathbf{r})$ in the integral in Eq. (1.12) is replaced by the incident field $U_{in}(\mathbf{r})$ and

$$\begin{aligned} U_B(\mathbf{r}) &= U_{in}(\mathbf{r}) + U_{B,s}(\mathbf{r}), \\ U_{B,s}(\mathbf{r}) &= \int_\Omega G(\mathbf{r}-\mathbf{r}')\chi(\mathbf{r}')U_{in}(\mathbf{r}')d^3r', \end{aligned} \tag{1.14}$$

where $U_B(\mathbf{r})$ and $U_{B,s}(\mathbf{r})$ are the total field and the scattered field evaluated in the first Born approximation respectively. This approximation is valid only for weak scattering potentials or, equivalently, for small scattering cross–sections.
If we now assume that the wavelength of the incident field is much bigger than the characteristic length of the scatterer, $\lambda >> d$, then we can assume that U_{in} is approximately constant all over the scatterer, assuming a value approximately equal to U_0. It follows that the scattered field in Eq. (1.14) depends mainly on the product of the optical contrast and the volume occupied by the sources. This, intuitively speaking, means that the scattered field given by a certain scattering object of optical contrast $\chi(\mathbf{r})$ extending over a volume Ω could be quite close to the one given by an object of smaller optical contrast but bigger volume. This in principle makes the retrieval of the scattering contrast extremely challenging as multiple combinations of volume and scattering potential exist that adequately fit the far–field dataset. The problem is then ill–posed as it violates at least one of the conditions for well–posedness in (1.1) in that the solution of the problem should be unique. Authors in [4] solved for this ambiguity by including *a-priori* information about the optical contrast to be retrieved. They demonstrated that lifting this ambiguity has immediate effect on the reconstruction quality and that this affects the resolution of the optical instrument.

1.1.2. Example 2: linear inverse problems

We consider an equation of the form

$$F\mathbf{x} = \mathbf{y}^\delta, \tag{1.15}$$

where we are interested in retrieving $\mathbf{x}$ from noisy measurements $\mathbf{y}^\delta$. We assume, for simplicity's sake, that $F \in \mathbb{R}^{mxm}$ is a real matrix with strictly positive decreasing singular values and that it is invertible. We denote by $\mathbf{y}^\delta \in \mathbb{R}^m$ the data, contaminated with additive noise. We further denote by $\mathbf{y}$ the unknown error–free data vector:

$$\mathbf{y}^\delta = \mathbf{y} + \boldsymbol{\delta}, \tag{1.16}$$

where $\boldsymbol{\delta}$ is the error vector. The error–free linear system of equation then reads:

$$F\mathbf{x}_{tr} = \mathbf{y}, \tag{1.17}$$

where $\mathbf{x}_{tr} \in \mathbb{R}^m$ is the input vector associated with the error–free output $\mathbf{y}$. The solution of Eq. (1.15)

$$\mathbf{x} = F^{-1}\mathbf{y}^\delta = F^{-1}(\mathbf{y} + \boldsymbol{\delta}) = \mathbf{x}_{tr} + F^{-1}\boldsymbol{\delta}, \tag{1.18}$$

is usually dominated by the back–propagated error, $F^{-1}\boldsymbol{\delta}$, and therefore does not carry any physical meaning. This behaviour can be understood in terms of the singular value decomposition (SVD) of F. If $F = U\ \mathrm{diag}(\sigma_i)\ V^T$ then

$$F^{-1}\mathbf{y}^\delta = V\ \mathrm{diag}(\sigma_i^{-1})\ U^T\mathbf{y}^\delta = \mathbf{x}_{tr} + \sum_{i=1}^{m} \sigma_i^{-1}(\mathbf{u}_i^T\boldsymbol{\delta})\mathbf{v}_i, \tag{1.19}$$

Where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the column vectors of U and V respectively. The presence of small singular values causes amplification of the noise and due to the high condition number – i.e. high ratio among the biggest and the smallest singular value of the matrix – of F the reconstruction is unstable with respect to perturbations in the data (noise). A matrix with high condition number is said to be *ill–conditioned*. In order to overcome this instability we need to filter out small singular values from the matrix F. However it shall be noted that this filtering corresponds to a low pass filtering operation because the small singular values are associated with the high frequency information of the sought signal. This is important to realize especially in optical sciences when we attempt to retrieve an unknown quantity form measured data. Propagation of light from the object to the detector implies a loss of information, and the inverse problem also requires to sacrifice a part of the information in favour of the possibility to, at the very least, retrieve something intelligible.

The two examples in Sections 1.1.1 and 1.1.2 should have clarified two instances in which the Hadamard conditions for well–posedness are not satisfied. We have noticed how ill–posed solutions can arise and how this can somehow happen due to a lack of information about the object we seek to retrieve. It appears natural to compensate for this information deficit by introducing in the problem further *a–priori* information. This is the task tackled by regularization methods. In the next sections we introduce tools necessary to understand and tackle inverse problems. The main references related to the next sections are [5–8].

1.2. Elements of regularization theory.

We briefly discuss various ways to regularize a solution.

Regularization by filtering. We start by recalling Eq. (1.19). We noticed that fluctuations in the data causes instability in the reconstruction due to the presence of small singular values in the forward operator F. A way to overcome this instability is to alter the expression in (1.19) by introducing a regularization filter $w_\alpha(\sigma_i^2)$, depending on the *regularization parameter* α, for which $w_\alpha(\sigma^2)\sigma^{-1} \to 0$ as $\sigma \to 0$. Thus we obtain an approximation of $\mathbf{x}_{tr}$:

$$\begin{aligned} \mathbf{x}_\alpha &= V \,\mathrm{diag}(w_\alpha(\sigma_i^2)\sigma_i^{-1})\; U^T \mathbf{y}^\delta \\ &= \sum_{i=1}^{m} w_\alpha(\sigma_i^2)\sigma_i^{-1}(\mathbf{u}_i^T \mathbf{y}^\delta)\mathbf{v}_i. \end{aligned} \tag{1.20}$$

A possible choice for $w_\alpha(\sigma^2)$ is a cut–off filter (so called truncated SVD):

$$w_\alpha(\sigma^2) = \begin{cases} 1 & \text{if } \sigma^2 > \alpha, \\ 0 & \text{if } \sigma^2 \leq \alpha. \end{cases} \tag{1.21}$$

Another example is given by the Tikhonov filter:

$$w_\alpha(\sigma^2) = \frac{\sigma^2}{\sigma^2 + \alpha}, \tag{1.22}$$

with corresponding solution:

$$\begin{aligned}\mathbf{x}_\alpha &= \sum_{i=1}^{m} \frac{\sigma_i(\mathbf{u}_i^T \mathbf{y}^\delta)}{\sigma_i^2 + \alpha} \\ &= (F^T F + \alpha I)^{-1} F^T \mathbf{y}^\delta,\end{aligned} \tag{1.23}$$

where α in Eqs. (1.20)–(1.23) is called *regularization parameter*. It should be clear at this point that the value chosen for α when solving the inverse problem is extremely important. A value that is too low leads to an over–oscillatory behaviour of the solution with respect to fluctuations in the data due to the noise, while assigning an exaggerated value to α alters the operator F^{-1} to the point where the retrieved reconstruction is too far away from the correct one. The solution is then said to be *biased*. If we denote the error in the solution space as

$$\begin{aligned}\mathbf{e}_\alpha &= \mathbf{x}_\alpha - \mathbf{x}_{true} \\ &= \mathbf{e}_\alpha^{bias} + \mathbf{e}_\alpha^\delta,\end{aligned} \tag{1.24}$$

then we can see that the error is comprised by two parts. The error introduced by the truncation of the small singular values of the operator F^{-1} is named $\mathbf{e}_\alpha^{bias}$ while the error caused by the amplification of the noise during the back–propagation is called $\mathbf{e}_\alpha^\delta$. The value of the regularization parameter α must be tuned so to guarantee that the retrieved solution is stable with respect to the noise and that the solution is not excessively smooth (i.e. the back–propagator F^{-1} has not been excessively altered by the truncation of its singular values). The machine learning community refers to this aspect of the data fitting as the *bias–variance tradeoff*. Without going into details we just point out that a multitude of methods are available to properly select the regularization parameter α. For a more detailed analysis of the error and a discussion about the regularization parameter selection methods please refer to [5, 7].

Variational regularization methods. We can introduce these methods by noting that Eq. (1.23) is equivalent to the following problem:

$$\text{minimize} \quad ||F\mathbf{x} - \mathbf{y}^\delta||^2 + \alpha||\mathbf{x}||^2. \tag{1.25}$$

where $||\cdot||$ denotes the ℓ_2 norm. We start by rewriting Eq. (1.25)

$$\text{minimize} \quad \left\| \begin{pmatrix} F \\ \sqrt{\alpha}I \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{y}^\delta \\ 0 \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} F\mathbf{x} - \mathbf{y}^\delta \\ \sqrt{\alpha}\mathbf{x} \end{pmatrix} \right\|^2. \tag{1.26}$$

The solution by the normal equations is:

$$\begin{pmatrix} F^T & \sqrt{\alpha}I \end{pmatrix} \begin{pmatrix} F \\ \sqrt{\alpha}I \end{pmatrix} \mathbf{x}_\alpha = \begin{pmatrix} F^T & \sqrt{\alpha}I \end{pmatrix} \begin{pmatrix} \mathbf{y}^\delta \\ 0 \end{pmatrix}, \tag{1.27}$$

and therefore

$$\mathbf{x}_\alpha = (F^T F + \alpha I)^{-1} F^T \mathbf{y}^\delta. \tag{1.28}$$

The advantage is that the problem in Eq. (1.25) does not require the computation of an SVD and this can be advantageous for problems that involve the retrieval of many parameters. Generally speaking the variational formulation of regularized inverse problems seeks to minimize a data loss functional, $\mathcal{L}(\mathbf{x}, \mathbf{y}^\delta)$, penalized by a regularizer $R_\alpha(\mathbf{x})$

$$\mathbf{x}_\alpha = \arg\min_{\mathbf{x}} \{\mathcal{L}(\mathbf{x}, \mathbf{y}^\delta) + R_\alpha(\mathbf{x})\}, \tag{1.29}$$

where the loss functional $\mathcal{L}(\mathbf{x}, \mathbf{y}^\delta)$ is a scalar that quantifies how much two elements in the data space are close to each other and the regularizer, $R_\alpha(\mathbf{x})$, is a scalar that quantifies how much the solution possesses certain desired characteristics and α is the regularization parameter. It is customary to introduce all the *a-priori* information about the sought signal in R_α. Classical examples of R_α are Tikhonov, total variation (TV) and lasso regularization.

It is worthwhile noting that Eq. (1.25) is a particular instance of the more generic problem

$$\text{minimize} \quad ||F\mathbf{x} - \mathbf{y}^\delta||^2 + ||L_\alpha \mathbf{x}||^2. \tag{1.30}$$

where L_α has been chosen as $L_\alpha = \alpha I$, however there are examples in which the regularization matrix, L_α, has been deliberately tuned to satisfy certain properties of the sought solution (see for instance [9]).

Discretization as regularization. Here the level of discretization of the forward operator controls the accuracy of the forward operator but it also stabilizes the reconstruction, see for instance [10].

Iterative regularization. The core idea of these methods is to use the iteration count as a regularization parameter. This is due to the fact that ill–posedness of the inverse problem leads to semiconvergent behaviour. If we define a data misfit

$$\mathcal{L}(\mathbf{x}, \mathbf{y}^\delta) = ||F\mathbf{x} - \mathbf{y}^\delta||^2, \tag{1.31}$$

we can observe that too few iterations yield an overly smooth approximate solution and too many iterations lead to highly oscillatory reconstructions. Terminating the iterations at the appropriate iteration count can grant a satisfactory solution.

A different aspect of these methods is highlighted in [11] where the author stresses out that in these family of methods the cost functional to be minimized is refined iteratively at each step and that this can be beneficial in some cases.

Statistical regularization. Statistical regularization theory is a framework for statistical inference in inverse problems. Both the data and the model parameters are assumed to be a realization of certain random variables which are described by certain probability functions. In variational methods the data can be thought of as a particular realization of a random variable, but the model parameters to be retrieved are usually not assumed to be described by a certain statistics [7]. In Bayesian inference the information we have about certain model parameters $\mathbf{x}$ is influenced by both an *a-priori* knowledge about these parameters and by observations of the misfit of the output of the statistical model given the parameters and the data. In mathematical terms:

$$\begin{aligned}\pi_{post}(\mathbf{x}|\mathbf{y}^\delta) &= \frac{L(\mathbf{x})\pi_{prior}(\mathbf{x})}{\int L(\mathbf{x})\pi_{prior}(\mathbf{x})d\mathbf{x}}, \\ L(\mathbf{x}) &:= p_{\mathbf{y}}(\mathbf{y}^\delta, \mathbf{x}),\end{aligned} \tag{1.32}$$

where π_{post} denotes the posterior probability, the prior distribution $\pi_{prior}(\mathbf{x})$ describes the prior knowledge and $L(\mathbf{x})$ is the likelihood function that is the probability that $\mathbf{y} = \mathbf{y}^\delta$ for a certain input $\mathbf{x}$.
The computation of the entire posterior distribution is computationally prohibitive for many applications and a more practical approach consists of retrieving only an estimator from the posterior distribution that is statistically relevant. Two important point estimators – points which belong to the posterior distribution and that hold statistical relevance – are introduced below.

Maximum a–posteriori (MAP) estimator. This estimator maximizes the posterior probability Eq.(1.32):

$$\mathbf{x}_{MAP} = \arg\max_{\mathbf{x}} \pi_{post}(\mathbf{x}|\mathbf{y}^\delta) = \arg\min_{\mathbf{x}}\{-\log L(\mathbf{x}) - \log \pi_{prior}(\mathbf{x})\}. \tag{1.33}$$

Eq. (1.33) shows that in Bayesian approaches the regularization term requires the knowledge of a statistical prior distribution of the sought signal. We also note that in case in which the prior is a Gaussian probability distribution, the MAP-estimator is equivalent to the Tikhonov regularization of Eq. (1.30).

Maximum likelihood (ML) estimator. This is the estimator that maximizes $L(\mathbf{x})$. In other words it is the value of the parameter $\mathbf{x}$ that, when used as input in the statistical model, minimizes the discrepancy among the measured data, $\mathbf{y}^\delta$, and the model output $\mathbf{y}$. It can be retrieved by letting the second term in Eq. (1.33) equal to zero:

$$\mathbf{x}_{ML} = \arg\max_{\mathbf{x}} L(\mathbf{x}) = \arg\min_{\mathbf{x}} -\log L(\mathbf{x}). \tag{1.34}$$

Estimating $\mathbf{x}_{ML}$ requires only access to the data likelihood, which is known when a model for the noise is available. This estimation scheme does not require the knowledge of a prior distribution which may not be available. On the other hand this implies that the reconstruction is not regularized, and this is problematic if the problem is ill–posed.
Notice that variational methods can be interpreted, in finite dimensional problems, as particular instances of statistical MAP or ML methods. In both the variational and statistical methods presented above we ended up with a problem which consists of minimizing or maximizing a certain functional. In the next section we discuss some methods that do this.

1.3. A few optimization methods

Optimization is a vast topic, here we briefly introduce a few important methods that are discussed in literature and that are used in this thesis.

1.3.1. definitions and descent methods

We focus on the problem of finding a minimizer for a certain cost functional, $\mathcal{L}(\mathbf{x}) = F(\mathbf{x})$, where $F(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ depends nonlinearly from $\mathbf{x}$:

$$\text{minimize} \quad F(\mathbf{x}). \tag{1.35}$$

We denote by J or by ∇F the gradient

$$J = \nabla F = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_n}, \end{bmatrix} \tag{1.36}$$

and with H the Hessian

$$H = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \cdots & \frac{\partial^2 F}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n x_1} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix}. \tag{1.37}$$

It is convenient to introduce the Taylor expansion around a certain point $\mathbf{x}$. The methods introduced here can be derived by stopping the expansion at the quadratic term:

$$\begin{aligned} F(\mathbf{x} + \Delta\mathbf{x}) &= F(\mathbf{x}) + \Delta\mathbf{x}^T J + \Delta\mathbf{x}^T H \Delta\mathbf{x} + \mathcal{O}(\Delta\mathbf{x}^3) \\ &\approx F(\mathbf{x}) + \Delta\mathbf{x}^T J \quad \text{for } \Delta\mathbf{x} \text{ sufficiently small}. \end{aligned} \tag{1.38}$$

The sufficient conditions for a local minimum of an unconstrained minimization problem are that J is zero and that H is positive definite. A *search direction* $\Delta\mathbf{x}$ is called a *descent direction* if $F(\mathbf{x} + \Delta\mathbf{x}) < F(\mathbf{x})$. If we assume that the first order approximation in Eq. (1.38) holds then this implies the condition $\Delta\mathbf{x}^T J < 0$. The process of choosing how far to move from the point $\mathbf{x}$ in the descent direction to have an optimal decrease of the loss functional is called *line search*.
Many descent methods share a similar structure, where the search direction is taken to have the form $\Delta\mathbf{x} = -B^{-1} J$, where B is a matrix with:

- B = I in the steepest descent,
- B = H in the Newton's method, and
- B ≈ H, in the Quasi–Newton method.

The steepest descent method, which will be employed later on, is outlined in more details below.

Steepest descent method. This method consists in noting that the gradient of a given function corresponds to the direction of the steepest ascent. It follows that the "best" (locally) decrease direction is given by the negative gradient. In an iterative scheme the new iterate is then given by

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \nabla F_n. \tag{1.39}$$

where $F_n = F(\mathbf{x}_n)$. The choice of the update rule Eq. (1.39) can be combined with a line search:

$$\begin{aligned} \Delta\mathbf{x}_n &= -\nabla F_n, \\ \mu_n &= \arg\min_{\mu>0} F(\mathbf{x}_n + \mu\Delta\mathbf{x}_n), \\ \mathbf{x}_{n+1} &= \mathbf{x}_n + \mu_n\Delta\mathbf{x}_n. \end{aligned} \tag{1.40}$$

This method converges, but can be slow during the last stages of the optimization. For this reason it is sometimes clustered with a method that is a better choice at the last iterations. One of such methods is Newton's method.

1.3.2. Methods for nonlinear least squares

Here we focus on the following problem:

$$\text{minimize} \quad \mathcal{L}(\mathbf{x}) = F(\mathbf{x}) = \frac{1}{2}||f(\mathbf{x})||^2 = \frac{1}{2}\sum_{i=1}^{m}(f_i(\mathbf{x}))^2, \tag{1.41}$$

with $f : \mathbb{R}^n \to \mathbb{R}^m$ and $m \geq n$. This problem can be tackled using dedicated algorithms that enjoy favorable convergence properties without requiring the implementation of second order derivatives. The Jacobian J in Eq. (1.36) is here defined as

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}. \tag{1.42}$$

We consider two methods that solve the problem described in Eq. (1.41)

Gauß–Newton method. This method can be derived by replacing f in Eq. (1.41) by its linear approximation:

$$\begin{aligned} \Delta\mathbf{x}_n &= \arg\min_{\Delta\mathbf{x}_n} ||f(\mathbf{x}_n) + \Delta\mathbf{x}_n^T J||^2, \\ \mathbf{x}_{n+1} &= \mathbf{x}_n + \mu_n\Delta\mathbf{x}_n, \end{aligned} \tag{1.43}$$

where μ_n is computed by the line search. The usual solution of Eq. (1.43) is via the equations: $\Delta\mathbf{x} = -(J^TJ)^{-1}J^Tf$. A necessary condition for the algorithm above to converge is that J is full rank at each step. This limitation is lifted in the method discussed below.

Levenberg–Marquardt method. This method is essentially a damped Gauß–Newton method. The update rule Eq. (1.43) is substituted by

$$\begin{aligned} \min_{\Delta\mathbf{x}_n} \quad & ||f(\mathbf{x}_n) + \Delta\mathbf{x}_n^T J||^2, \\ \text{s.t.} \quad & ||\Delta\mathbf{x}_n||^2 \leq \delta, \end{aligned} \tag{1.44}$$

where δ is a positive number, assumed known, such that the model is sufficiently accurate inside a ball with radius δ, centered at $\mathbf{x}_n$.
By introducing the Lagrangian multiplier μ, we can rephrase the minimization problem Eq. (1.44) as a variational problem that is similar to Eq. (1.25) which admits a solution of the following form:

$$\Delta\mathbf{x}_n = -(J^T J + \mu_n I)^{-1} J^T f \quad \text{with} \quad \mu_n \geq 0, \tag{1.45}$$

where the parameter μ influences the direction and the size of the step and it is changed along the iterations in the attempt to grant optimal convergence [6].

1.4. Summary

In this chapter we introduced a few fundamental concepts that are of paramount importance in the field of inverse problems. Moving from these premises we shall focus on technical problems that arise in mask metrology and inspection for extreme ultraviolet (EUV) lithography which are relevant and interesting. We rely on the content of this chapter to devise new inversion schemes and introduce ideas that can tighten and improve the inspection and metrology of targets on EUV wafers and reticles.

References

[1] J. B. Keller, *Inverse problems,* The American Mathematical Monthly **83**, 107 (1976), https://doi.org/10.1080/00029890.1976.11994053 .

[2] H. Jacques, *Lectures on Cauchy's Problems in Linear Partial Differential Equations* (1923).

[3] P. Müller, M. Schürmann, and J. Guck, *The theory of diffraction tomography,* (2015), arXiv:1507.00466 [q-bio.QM] .

[4] T. Zhang, C. Godavarthi, P. C. Chaumet, G. Maire, H. Giovannini, A. Talneau, M. Allain, K. Belkebir, and A. Sentenac, *Far-field diffraction microscopy at $\lambda/10$ resolution,* Optica **3**, 609 (2016).

[5] C. R. Vogel, *Computational Methods for Inverse Problems* (Society for Industrial and Applied Mathematics, 2002) https://epubs.siam.org/doi/pdf/10.1137/1.9780898717570 .

[6] K. Madsen, H. Nielsen, and O. Tingleff, *Methods for non-linear least squares problems (2nd ed.),* (2004).

[7] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, *Solving inverse problems using data-driven models,* Acta Numerica **28**, 1–174 (2019).

[8] J. Eriksson, *Optimization and Regularization of Nonlinear Least Squares Problems*, Ph.D. thesis, Umeå University (1996).

[9] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo, *Inverse problems with structural prior information,* Inverse Problems **15**, 713 (1999).

[10] B. Kaltenbacher, A. Kirchner, and B. Vexler, *Adaptive discretizations for the choice of a tikhonov regularization parameter in nonlinear inverse problems,* Inverse Problems **27**, 125008 (2011).

[11] M. R. Charest and P. Milanfar, *On iterative regularization and its application,* IEEE Transactions on Circuits and Systems for Video Technology **18**, 406 (2008).

2

An inverse problem in EUV mask metrology: EUV scatterometry

Parts of this chapter have been published in *Applied Optics* **58**, 5916-5923 (2019) [1]

Scatterometry is the industrial workhorse for fast, quantitative and non–destructive optical metrology of periodic semiconductor structures on a surface. Examples of such periodic structures are line/space photoresist or absorber gratings and periodic arrays of squares/holes. The periodicity of the sample dictates the diffraction of the incoming light field into a finite set of diffraction orders. A measured dataset in scatterometry consists of one or more measured diffraction orders, whose direction of propagation is given by the grating equation:

$$\sin\theta_i + \sin\theta_n = n\frac{\lambda}{\Lambda}, \tag{2.1}$$

where θ_i is the angle of incidence, θ_n is the angular location of the n^{th} diffraction order, λ is the wavelength and Λ is the pitch of the grating. The field above the grating can then be written, for the case of a reflection setups, in terms of the Rayleigh series:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_{inc}(\mathbf{r}) + \mathbf{E}_s(\mathbf{r}) = \mathbf{E}_{inc}(\mathbf{r}) + \sum_n \mathbf{R}_n e^{i(k_{xn}x + k_{yn}y + k_{zn}z)}, \tag{2.2}$$

where $\mathbf{E}(\mathbf{r})$ is the total field above the grating, $\mathbf{E}_{inc}(\mathbf{r})$ is the incident field, $\mathbf{E}_s(\mathbf{r})$ is the scattered (diffracted) field, $\mathbf{R}_n$ is the reflection coefficient of the n^{th} diffraction order and k_{xn}, k_{yn} and k_{zn} are the components of the wavevector of the n^{th} diffraction order, with $k = \frac{2\pi}{\lambda}$. Notice that in a transmission setup the reflection coefficient needs to be substituted by the transmission coefficient. As stated by Eq. (2.1) the components of the diffracted vector **k** are linked to the ones of the incident wavevector by the Bragg's law. If we assume that the grating is periodic in the x direction, then:

$$\begin{aligned} k_{xn} &= k_x + \frac{2\pi n}{\Lambda}, \\ k_{yn} &= k_y, \\ k_{zn} &= \sqrt{k^2 - k_{xn}^2 - k_{yn}^2}. \end{aligned} \tag{2.3}$$

It can be seen from Eq. (2.2) that the measured diffraction intensities are $\mathbf{I}_n = |\mathbf{R}_n|^2$. It is common to measure the 0^{th} order which has a high intensity and consequently provides a high signal to noise ratio. Depending on the wavelength, the pitch and the angle of incidence, higher diffraction orders can fall in the numerical aperture (NA) – i.e. the acceptance angle – of the detector and can be measured [2], however it should be noticed that for short pitches only a few propagating diffraction orders can exist (Eq. 2.1). The fraction of incident light that gets diffracted into any diffraction order encodes information about overlay and structural parameters of the 3D profile of the grating such as height, critical dimension (CD) and side–wall angle (SWA) [3], the latter information being the aim of the 3D profilometry of gratings. In the usual scatterometric approach one attempts to retrieve this structural information by doing the following:

- measure a set of diffracted intensities as a function – for instance – of wavelength or angular locations by, for instance, scanning the detector angle at a fixed angle of incidence. This set of intensities can comprise only the zeroth order or also the higher orders.
- model the light–sample interaction, usually using a fully rigorous Maxwell solver,
- parameterize the grating, and eventually also the underlying stack, with certain shape parameters, and
- solve the inverse problem by fitting the acquired data set. This can be done in real–time, letting the optimization routine run while the industrial production process is ongoing, or by searching through a previously assembled library of simulations for different possible geometries.

In what follows we give a brief overview of these topics. For a more thorough overview please refer to [4].

A few scatterometry setups. We describe a few important scatterometers, classified according to their *modus operandi*, that have been discussed in literature.

Angular scatterometers. In this configuration the light generated by a source passes through an optical scanning system that refracts and focuses the light upon the target. The optical system can be designed so that a discrete set of angles of incidence, θ_i are available. In a sequential measurement the detector can be moved so to measure any of the diffracted orders, Eq. (2.2), as the angle of incidence is changed while in a parallel acquisition all of the diffraction orders are simultaneously measured [2]. We can then retrieve, for instance, the variations in the diffraction efficiency of a certain order or of a set of orders with respect to the angle of incidence [5, 6]. Alternatively one could measure the angle–dependent scattering intensities for a fixed angle of illumination in the so called *fixed–angle* scatterometers. When the pitch of the grating approaches the wavelength only a few orders can be measured, and one may rather wish to study only the efficiency of the specular reflection with respect to the angle of incidence. This is the $2-\Theta$ configuration [7].

Spectral scatterometers. There are cases in which one may wish to keep the angle of incidence fixed. In such cases one could illuminate with a broadband radiation and filter the light at the detector side with a spectrometer. Because the angular locations of the diffracted modes depend on the wavelength λ, the grating scatters the incoming light at different locations for different wavelengths, and one can retrieve the diffraction intensities per wavelength at a specific location of, for instance, a line array detector. One can then retrieve a data–set by measuring the diffraction efficiency of a certain order or of a set of orders with respect to the wavelength [6].

Coherent Fourier scatterometry. In this scatterometric setup a high NA optics focuses a coherent beam on the grating. When the pitch is sufficiently large compared to the wavelength more than one diffraction order ends up in the collection NA leading to overlapping diffraction orders [2]. The use of coherent light allows one to retrieve phase information – intended as the phase difference among overlapping orders – that can be disentangled by scanning the spot with respect to the grating. In [8] authors demonstrated that this phase information can improve the sensitivity of the scatterometer with respect to its incoherent version. The method has also been successfully combined with interferometry [9] and with phase retrieval algorithms [10].

The modeling of light matter interaction. For very simple grating structures, when the wavelength is much smaller than the grating pitch, it may be possible to simulate the light–matter interaction via the scalar diffraction theory. However in the vast majority of cases a fully rigorous solver is employed. A thorough discussion on the computational aspects of forward electromagnetic methods is beyond the purpose of this thesis. Here we limit ourselves to the basic introduction of a few methods.

Rigorous coupled wave analysis (RCWA). The Rigorous Coupled Wave Analysis (RCWA) is a computational method for periodic structures that belongs to the class of modal expansion methods [11]. In RCWA, a grating geometry is divided into staircases/slabs typically along the direction orthogonal to the surface of the patterned sample. One solves for the field at each slab interface and enforces boundary conditions at each slab in order to link the solutions of consecutive slabs to each other, which in the end enables one to compute the diffraction efficiencies. Because of the fact that the grating is divided into slices this method is not easily adaptable to the case of a grating with oblique interfaces. This method has been extended to the case in which the sample is aperiodic by using perfectly matched layers (PMLs) [12, 13].

Finite element methods (FEM). In FEM solvers one starts from eliminating the magnetic field from the Maxwell's equations of time harmonic fields to obtain the Helmholtz equation for the electric field:

$$\nabla \times \nabla \times \mathbf{E} - \omega^2 \mu \epsilon \mathbf{E} = 0, \tag{2.4}$$

where impressed sources have been neglected and ϵ and μ are the permittivity and permeability tensors. The scattering problem is solved by truncating the computational domain into an interior domain Ω_{CD} and an exterior domain. The incoming field is added as a light source in Eq. (2.4) and the infinite domain through which the field in principle diffuses is terminated by the perfectly matched layer (PML). The PML, which dampens the field that is entering its domain, occupies a region slightly bigger then Ω_{CD} and inside the PML the scattered field is used as unknown.

In the inner boundary the field boundary conditions are that the tangential components of the total electromagnetic field $\mathbf{E}$ and $\mathbf{H}$ should be continuous and on the outer boundary the tangential component of $\mathbf{E}$ or $\mathbf{H}$ are set to zero. Inside the PML complex stretching coordinates are defined by means of which the light field enters the PML without getting reflected and is then absorbed [11]. Once the boundary conditions are in place the FEM expresses the field as a summation of ansatz polynomials of finite support over discretization elements with a certain mesh size [14]. Eq. (2.4) is then reformulated into a weak form to yield a linear system of equations

$$AE = b, \tag{2.5}$$

where b contains contributions of the incoming field, A is a sparse matrix and E contains the coefficients of the ansatz functions. The system of equations in Eq. (2.5) can be solved directly for small systems or iteratively, for instance via a preconditioned method, for larger systems.

Integral equations methods. Volume integral equations (VIE) and surface integral equations (SIE) methods for finding the solution to Maxwell's equations are based on the equivalence principles according to which a scatterer can be removed and replaced by a certain source distribution that radiates the same EM field [15]. One starts from dividing the space in a computational domain, Ω_{cd}, and an external domain. We assume, in the case of the SIE, that Ω_{cd} is isotropic and homogeneous and that the incoming field is defined in this domain. In such cases one can show that the total field inside Ω_{cd} is given by the primary fields plus the contribution given by equivalent surface current densities which are the rotated tangential fields components:

$$\mathbf{J}_S(\mathbf{r}) = \mathbf{n}(\mathbf{r}) \times \mathbf{H}(\mathbf{r}), \quad \text{and} \quad \mathbf{M}_S(\mathbf{r}) = \mathbf{E}(\mathbf{r}) \times \mathbf{n}(\mathbf{r}), \tag{2.6}$$

where $\mathbf{n}(\mathbf{r})$ is the normal to the considered surface. The unknown fields are the tangential components of the electric or magnetic fields and can be computed by solving a linear system of equations. This method is difficult to implement for the case of inhomogenous and/or anisotropic material.
In such cases the VIE method can be implemented. In this case the anisotropic material with permittivity and permeability tensors $\boldsymbol{\epsilon}_r$ and $\boldsymbol{\mu}_r$ are embedded in a certain background material with material parameters ϵ and μ. In the VIE method the equivalent polarization current densities are defined as

$$\mathbf{J}_V(\mathbf{r}) = -i\omega\epsilon\boldsymbol{\Delta\epsilon}_r(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}), \quad \text{and} \quad \mathbf{M}_V(\mathbf{r}) = -i\omega\mu\boldsymbol{\Delta\mu}_r(\mathbf{r}) \cdot \mathbf{H}(\mathbf{r}), \tag{2.7}$$

where $\boldsymbol{\Delta\epsilon}_r(\mathbf{r}) = \boldsymbol{\epsilon}_r - I$ and $\boldsymbol{\Delta\mu}_r(\mathbf{r}) = \boldsymbol{\mu}_r - I$ with I being the identity matrix. The scattered field is then an integral over the contribution of elementary dipoles in Ω_{cd} and one can arrive to a linear set of equation in the unknowns. An advantage of this method is that one only meshes the region where the nontrivial scatterers are located so the non–patterned areas of a layer stack either above or below the grating do not need to be numerically evaluated. A disadvantage is that the system matrix is large and full.

The inverse problem. The other fundamental building block of scatterometry is the fitting of the acquired dataset. One usually defines a cost functional, which reads:

$$\mathcal{L}(\mathbf{p}, \mathbf{y}^{\delta}) = \frac{||\mathbf{y}^{\delta} - F(\mathbf{p})||^2}{2\boldsymbol{\sigma}^2}, \tag{2.8}$$

where $\mathbf{y}^{\delta}$ is the noisy data, $F(\mathbf{p})$ is the model evaluation given parameter vector $\mathbf{p}$ and $\boldsymbol{\sigma}$ contains the uncertainty about the measured data. One then proceeds to iteratively adjust $\mathbf{p}$ to minimize $\mathcal{L}(\mathbf{p}, \mathbf{y}^{\delta})$ thus finding the set of parameters that should be close to the correct ones. Popular algorithms in this sense are the Gauß–Newton – Eq. (1.42) – and the Levenberg–Marquardt – Eq. (1.43) – [3, 16], although other interesting choices are available [17]. Another popular approach consists of pre–assembling a library of simulations where the optimal parameters are retrieved by looking at the simulated data that better resembles the measurand [18].

Usually one is interested in a few parameters, for instance the critical dimensions (CD), height and side–wall–angle (SWA) of the grating. The scatterometric approach tends to be cumbersome and slow when many parameters are to be adjusted. This is predominantly because of two reasons:

- the minimization of Eq. (2.8) requires multiple evaluations of many similar systems. This implies multiple computationally intensive rigorous simulations.
- The features to be retrieved can be correlated. Correlation causes ill–posedness and the presence of multiple local minima as it is difficult to distinguish the signature of one feature from some of the others when analyzing the far field pattern. In such cases regularization methods need to be employed because the problem is ill–conditioned. It should also be noticed that the presence of many features manifests itself by introducing in the optimization landscape many local minima with higher error than the global minimum and saddle points that might be mistakenly interpreted as a local minimum [19].

These and other challenges that arise when exploring high–dimensional data–sets or when minimizing cost functionals over high dimensional spaces are referred to as *curse of dimensionality*.
The industrial need for *real–time* computations and optimization in high dimensional spaces is a driving force for research in computational methods in electromagnetics and in optimization. In the following section we briefly discuss various approaches to reduce the problem of the excessive dimensionality of the search space and we introduce the approach we devised to mitigate the difficulties that arise in the inverse problem of scatterometry when many features are to be retrieved.

2.1. Dimensionality reduction techniques

A list of few important methods for dimensionality reduction in multidimensional spaces are reported below.

Singular value decomposition (SVD) based methods. Usually these methods consist of applying an SVD to a symmetric/hermitian (semi)positive–definite matrix. The eigendecomposition of such matrices identifies a full set of orthogonal eigenvectors that represent the basis for a vector space. One could then think of identifying the eigenvectors which contain most of the information/variation of the data or of the model. Methods based on the singular value decomposition are often applied by the machine learning community to reduce the dimensionality of the datasets for training purposes or to analyze and understand the data better. These methods can also be applied in the model space to simplify the model by retaining its most important degree of freedom. An example is in [20], where authors noticed that the feature of the model they were using showed interdependencies. They considered a linear inverse problem of the form

$$F\mathbf{x} = \mathbf{y}, \tag{2.9}$$

with $F : \mathbb{R}^n \to \mathbb{R}^{\mathbb{m}}$. They proceeded to build up a library of possible input states $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ constrained by prior information. X was subsequently used to build up a covariance matrix:

$$\begin{aligned} \mathbf{C} &= \frac{1}{N}\sum_{l=1}^{N}(\mathbf{x}_l - \boldsymbol{\mu})(\mathbf{x}_l - \boldsymbol{\mu})^T, \\ \boldsymbol{\mu} &:= \frac{1}{N}\sum_{l=1}^{N}\mathbf{x}_l. \end{aligned} \tag{2.10}$$

As $\mathbf{C}$ is by construction symmetric and semi–definite positive it can be diagonalized and it has orthogonal eigenvectors and real non–negative eigenvalues. They used the principal component analysis (PCA) to retain in the model only a subset of the eigenvectors (principal components) $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$, $p < n$, that provide an accurate low dimensional representation of the original space. Any model in the starting space is then represented as $\mathbf{x} = \boldsymbol{\mu} + \sum_{k=1}^{p} \alpha_k \mathbf{v}_k$ and by substitution in Eq. (2.9) one is left to solve a linear system of equations in the reduced set of uncorrelated unknowns $\boldsymbol{\alpha}$.

Another application of SVD in the model space is in [21] where authors decomposed the covariance matrix to identify and rank parameters based on their importance so to retain in the model only the most relevant of them. Another interesting example of the use of the SVD for the purpose of model reduction is in [22], where authors decomposed the covariance matrix of the gradient vector to find the eigendirections where the input function of interest varied more strongly. They proceeded to construct a response surface using only the most important of the rotated coordinates identified by the eigenvectors. Such *active subspace* can, in some instances, be an adequately informative low rank representation of the input space.

Regression based methods. This is a family of methods that reduces the dimensionality of the model by enforcing a few coefficients in a regression problem to zero. Examples are forward selection [23], lasso [24] and elastic–net [25] regression. The main idea is to exclude from the model those degrees of freedom that to do not contribute to improve the fitting too much. We will use one of these methods in the next chapter.

Sensitivity based methods. In model based approaches one could assess the importance of a parameter by looking at how much it affects the output. The most inexpensive sensitivity analysis methods are the local ones, where a certain input is perturbed from its nominal position and the impact on the output is observed. This method however does not immediately reveal any secondary effect or any interdependencies among parameters and the local sensitivities may change drastically depending on the particular evaluation point in the parameter space. Global sensitivity analysis methods, such as variance decomposition methods [26, 27], provide the means to tackle these issues, however they are expensive to evaluate and require the approximation of a high dimensional integral with multiple expensive Monte Carlo methods. Screening methods are less expensive to evaluate than global sensitivity analysis methods and give indication on the global importance of a parameter. A particularly successful screening method, outlined in more details in the next section, is the Morris design [28, 29].

We have summarized a few methods for dimensionality reduction in the model space. In the next chapter we devise an algorithm for feature selection which is applicable in the nonlinear regression case and we compare it to the Morris design.

2.2. Automatic feature selection in EUV scatterometry

We are concerned with the retrieval of certain unknown parameters of an object from measured data and with the understanding of which of these unknowns are most relevant for an appropriate description of the object. The purpose is to simplify the model, leaving only its most important parameters as unknowns.

One could tackle the problem using sensitivity analysis methods which "allow to study how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input or may be used to determine the most contributing input variables to an output behavior, or to ascertain some interaction effects within the model" [30]. However for certain applications, especially for the industrial ones, it may be preferable to make this sort of modeling decision automatic, as one often wishes to keep the user from performing any kind of mathematical operation on a certain system. In 1996, Tibshirani introduced a penalized regression, the lasso [24], which is able to automatically select the most important inputs in a model by shrinking the regression coefficients of the least relevant ones to exactly zero. Further research has demonstrated that combining the ℓ_2 and ℓ_1 norm penalties in the regression can help to overcome some limitations of the lasso [25]. Even though these algorithms have been developed to solve linear regression problems, their use can be extended to the nonlinear case. In nonlinear regression the aim is to minimize the cost functional:

$$\mathcal{L}(\mathbf{p}, \mathbf{y}^\delta) = \frac{||\mathbf{y}^\delta - F(\mathbf{p})||^2}{2\boldsymbol{\sigma}^2}. \tag{2.11}$$

In Eq. (2.11) $||\cdot||$ denotes the euclidean norm, i.e. the l_2 norm, $\mathbf{y}^\delta$ represents the noisy data, $F(\mathbf{p})$ is the model evaluation given the parameter vector $\mathbf{p}$ and $\boldsymbol{\sigma}$ is the vector containing the uncertainties about the measured data. We assume for the variance of the measured value [3]:

$$\boldsymbol{\sigma}(\lambda)^2 = [a \cdot \mathbf{E}(\lambda)]^2 + b_g^2, \tag{2.12}$$

where $\mathbf{E}$ is the vector that contains the diffraction efficiencies (which is the fraction of incident power that is diffracted in a certain direction), λ denotes the wavelength, a is a constant assumed to be equal to 0.05 (i.e. 5% noise relative to the diffraction efficiencies) and b_g is the background noise of the detector, whose contribution becomes dominant for diffracted orders whose efficiency is less than 0.01% [31]. As stated in Chapter 1 problems as these are usually solved by dedicated methods as explained in Section 1.3.2. Here we start recalling the Gauß–Newton update rule, Eq. (1.43)

$$\begin{aligned} \Delta\mathbf{p} &= \underset{\Delta\mathbf{p}_n}{\arg\min} ||\mathbf{y}^\delta - F(\mathbf{p}_n) - \Delta\mathbf{p}^T J_n||^2, \\ \mathbf{p}_{n+1} &= \mathbf{p}_n + \Delta\mathbf{p}. \end{aligned} \tag{2.13}$$

Where, as usual, J_n is the Jacobian of $F(\mathbf{p}_n)$ evaluated at the current iterate and $\mathbf{p}_n$ is the current estimate. When a sufficiently accurate prior, $\mathbf{p}_0$, is available, it is possible to replace $\mathbf{p}_n$ in Eq. (2.13) by $\mathbf{p}_0$, alternatively $\mathbf{p}_0$ can be used as a first guess. As previously discussed, the usual solution of Eq. (2.13) is via the normal equations.

In those cases in which a large number of parameters have to be optimized, one can penalize the incremental vector $\mathbf{\Delta p}$ in Eq. (2.13) with a penalty term that encourages a sparse reconstruction:

$$\mathbf{p}_{n+1} = \mathbf{p}_n + \underset{\mathbf{\Delta p}}{\arg\min} ||\mathbf{\Delta}F - \mathbf{\Delta p}^T J_n||^2 + \gamma P_\alpha(\mathbf{\Delta p}), \tag{2.14}$$

where

$$P_\alpha(\mathbf{\Delta p}) = (1-\alpha)\frac{1}{2}||\mathbf{\Delta p}||_2^2 + \alpha||\mathbf{\Delta p}||_1. \tag{2.15}$$

In (2.14) $\mathbf{\Delta}F = \mathbf{y}^\delta - F(\mathbf{p}_n)$, γ is the regularization parameter to be determined by seeking a balance among the data fitting term and the regularization term, $||\cdot||_p$, with $p = 1, 2$, is the ℓ_1 or ℓ_2 norm and $\alpha \in [0, 1]$ is a parameter that determines the relative strength among the ℓ_2 and the ℓ_1 norm. In all that follows we have chosen $\alpha = 1/2$. The penalized regression in (2.14), known in the literature as the "elastic net" [25], produces a more simplistic model by removing redundant degrees of freedom via a variable selection process. The output, $\mathbf{\Delta p}_n$, of the regression (2.14) will be a vector with some entries which can be equal to zero. By adding a zero offset to some of the entries of $\mathbf{p}_n$, some of the parameters will be fixed to a certain value, resulting in a reduced number of unknowns to be retrieved by the estimation routine.

A further important aspect to be considered while solving Eq. (2.14) lies in the appropriate selection of the regularization parameter. In this work we have applied the L–curve criterion at each iteration. As the value of γ given by the L–curve can change at each iteration, this can result in an oscillatory trend. An heuristic formula that deals with this problem is [32, 33]:

$$\gamma_n = \begin{cases} \epsilon\gamma_{n-1} + (1-\epsilon)\gamma & \text{if } \gamma < \gamma_{n-1}, \\ \gamma_{n-1} & \text{otherwise} \end{cases} \tag{2.16}$$

with $0 < \epsilon < 1$ being a–priori chosen. The algorithm described above is applicable to those inverse problems for which prior information is available and for which it is possible to compute the gradient of the function to be optimized. We stop iterating after the second time in which $\mathcal{L}$ in Eq. (2.11) decreases by a small amount (0.01) [34].

It is interesting to compare the outcome of the method above with the one given by sensitivity analysis methods. A particulary robust and widely used screening method is the Morris design [28].

2.2.1. The Morris design

In the Morris method one aims to classify the inputs in three different categories:

- inputs having negligible effect on the output,

- inputs having significant linear and no cross–coupling effects on the output, and
- inputs having significant nonlinear and/or cross–coupling effects.

The method proceeds as follow:

- the input space is discretized in what Morris called an n–level grid (i.e. each parameter is discretized in n discrete values);
- a step, Δp is defined as a multiple of $\frac{1}{n-1}$;
- the input space is sampled and a starting point (also known as *base point*), in which the model output is evaluated, is selected;
- one moves away from this point by displacing one of the inputs by Δp. At this new location of the input space the model output is again evaluated, and an incremental ratio, named "elementary effect" (which we will formally introduce at a slight later moment below) is computed;
- the procedure described in the point immediately above is repeated for all the parameters, moving one parameter at a time while the others are kept fixed, until each parameter has been displaced. Notice that this creates a "path" in the input space. Morris referred to this path as a "trajectory"
- After each parameter has been displaced one selects randomly a different *base point* and the whole procedure is repeated. This is done a total R times, with R usually equal to 10–20, for R independently generated trajectories.

As mentioned the method is based on the computation of the "elementary effects":

$$E_i^{(j)} = \frac{F(p_1^{(j)}, p_2^{(j)}, \dots, p_i^{(j)} + \Delta p_i^{(j)}, \dots, p_n^{(j)}) - F(\mathbf{p}^{(j)})}{\Delta p_i^{(j)}} \tag{2.17}$$

where $E_i^{(j)}$ is the elementary effect associated to the i–th input and to the j–th trajectory. The statistics of the distributions of the elementary effects in the R trajectories is evaluated using the following estimators:

$$\mu_i^* = \frac{1}{R}\sum_{j=1}^{R} |E_i^{(j)}|, \qquad \sigma_i = \sqrt{\frac{1}{2}\sum_{j=1}^{R}\left(E_i^{(j)} - \frac{1}{R}\sum_{j=1}^{R} E_i^{(j)}\right)^2}. \tag{2.18a,b}$$

The absolute value in (2.18) is used in order to keep close–valued elements of opposite sign from canceling each–other out [29]. A high mean value, μ_i^*, implies an high overall effect of the i–th input over the output and a high spread, σ_i, about the mean implies that the elementary effects relative to this input are significantly different from each other, hence the value of an elementary effect is strongly affected by the choice of the point in the input space at which it is computed. This indicates an input with a non-linear effect on the output, or an input involved in interactions with other inputs. A plot of μ^* against σ allows one to examine the computed values relative to each other and to evaluate the importance of inputs in the model. If a given input has both low μ^* and low σ values then it has low impact on the output and it is not involved in significant non–linear interactions. Hence it can be dropped from the model by fixing it to a certain value within its uncertainty bounds.

It is important to mention that the accuracy of the method depends on an appropriate sampling of the input space. A finer grid (i.e. smaller Δp) in the input space yields an overall more accurate result in terms of the evaluated sensitivities, however a finer grid also implies that many trajectories need to be generated in the input space in order for it to be sampled accurately hence leading to a more computationally intensive procedure. There is no strict rule for the choice of Δp and the selection of its value and the value of the number of trajectories is left to the user. A practical way to select these parameters is to start by retrieving $[\mu_i^*, \sigma_i]$ with a coarse grid and few trajectories and gradually refine the grid and increase the number of trajectories. When the output of the method (i.e. $[\mu_i^*, \sigma_i]$) does not change much with further refinements one can understand which is a suitable value for the Δp and the number of trajectories.

There are important differences among the Morris method and the regression–based methods as applied to the problem of the selection of important parameters in a parameterized model. This is firstly because they explore the input space differently. Even though the Morris design is based on the computation of small steps from one point to the other, it can be considered as a “global” sensitivity analysis method as it explores the entire input space spanned by the parameters of interest. Conversely the regression in Eq. (2.14) could be thought as a “local” method, in the sense that it is looking for a solution in the neighborhood of $\mathbf{p}_n$. Another difference lies in the criteria of importance. In the Morris design a certain variable is considered important when its perturbation significantly affects the output and/or when it is involved in nonlinear effects. On the other hand, the elastic–net tries to remove unimportant parameters solving a regularized regression problem. The metric of the variable selection algorithm is determined by a trade–off among goodness of fit and complexity of the model. Oversimplistic models fail to accurately describe the data and lead to biased solutions, while overcomplicated ones are difficult to interpret and “overfit”, in the sense that they are too sensitive to the noise in the data, leading to poor predictions over future data–sets [35, 36]. The lasso and elastic net are regression methods able to select a simple model, starting from a complicated one, by shrinking some of the regressors to exactly zero. These methods seek to find a simple model, selected from among many, that best captures the data. A way to further understand this is by studying the algorithm that solve the minimization problem (2.14). An example are coordinate–descent algorithms, in which the update rule, besides a scaling factor, is [37]:

$$\Delta p_j = \begin{cases} \sum_{i=1}^{N} r_{ij} + \alpha\gamma, & \text{if } \sum_{i=1}^{N} r_{ij} < -\alpha\gamma \\ 0, & \text{if } -\alpha\gamma < \sum_{i=1}^{N} r_{ij} < \alpha\gamma \\ \sum_{i=1}^{N} r_{ij} - \alpha\gamma, & \text{if } \sum_{i=1}^{N} r_{ij} > \alpha\gamma. \end{cases} \tag{2.19}$$

In Eq. (2.19) $\sum r_{ij}$ represents the sum over the residuals of the linear regression, each weighted by a certain coefficient, obtained while fitting such that the j–th variable is excluded from the model. According to Eq. (2.19) a certain parameter is excluded from the model if its presence does not improve significantly the fitting.

In what follows we have used the following matlab® packages: regtools for the L–curve regularization [38] and "Morris SU sampling", implemented by the group of Rafael Muñoz–Carpena at the University of Florida [39]. Rigorous electromagnetic solutions of the forward problem are computed using the finite element method solver JCMsuite® [40].

2.2.2. Application to gratings on an EUV mask

We apply the algorithm described above to the problem of feature selection to gratings on an EUV mask. As stated in Section 2.2 the problem we are challenged with consists of devising a method for automatic profile selection of parameterized gratings or nanostructures on an EUV mask. The purpose is to simplify the model by eliminating redundant degrees of freedom. This can be useful especially for industrial applications where one often does not want to let the user of a certain system/machine perform modelling operations. We stress that all of the data and results presented in this section are computationally generated.

Fig. 2.1(a) presents the cross section of a grating profile. The grating is parameterized with six parameters that correspond to the X and Y coordinates within the respective layers. The grating is assumed to be symmetric. The EUV radiation illuminates a Mo/Si multilayer coated reflective mask, with a patterned absorber profile on top of it. The angle of incidence is $6°$, for which the multilayer is in resonance, giving a reflectance of 60–70 %. The materials properties are listed in Table 2.1 [16]. The side wall angle (SWA) of the SiO_2 layer is assumed to be equal to the SWA of the TaN layer above. The period of the grating at the level of the mask is 420 nm and the nominal width of the absorber is 140 nm , for a line to space ratio of 1:2. For such configuration only the diffracted orders from -6 to +11 are detectable with sufficient intensity [3]. Figure 2.1(b) reports the recorded (simulated) diffraction efficiencies for the aforementioned settings and for three different wavelengths of the incoming s–polarized light field [3]: $\lambda_1 = 13.398$ nm, $\lambda_2 = 13.664$ nm, $\lambda_3 = 13.931$ nm.

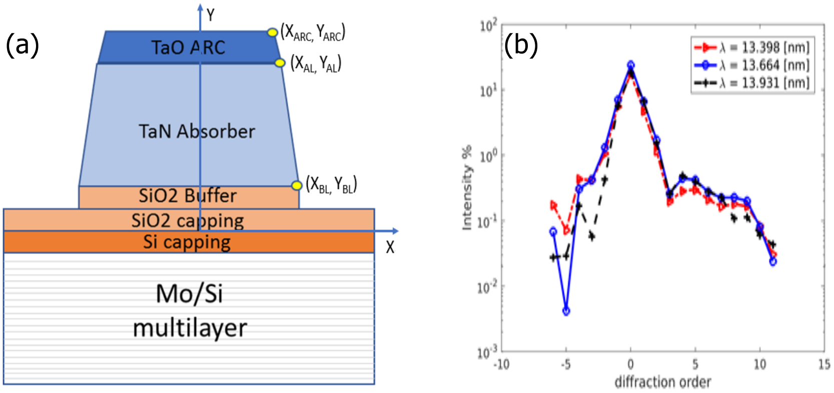

Figure 2.1: a) Symmetric grating with parameterized profile. The independent degrees of freedom are the X and Y coordinates of the yellow points. The grating is assumed to be symmetric. The materials are given in Table 2.1. Angle of incidence is $6°$. Light is s polarized. To preserve clarity the layers thicknesses are not presented in real scale (for the nominal values refer to Table 2.1) b) Diffraction efficiencies are given in percentage. For the given geometry and wavelengths only a subset of the diffraction orders can be detected.

Table 2.1: **Nominal layer thicknesses and material properties for the mask in Fig. 2.1 at $\lambda = 13.5$ nm**

layer	thickness [nm]	n	k
ARC TaO	12	0.951	0.003
Absorber TaN	54.9	0.946	0.0326
SiO_2 (buffer)	8	0.97352	0.01608
SiO_2 (oxidation)	1.246	0.97352	0.01608
Capping layer Si	12.536	0.99846	0.00184
$MoSi_2$	0.5	0.96675	0.00446
Mo	2.256	0.91872	0.00672
$MoSi_2$	0.5	0.96675	0.00446
Si	3.077	0.99846	0.00184
Substrate	6.35e6	0.97352	0.01608

We choose our starting point for the regression, $\mathbf{p}_0$, by sampling a uniform prior distribution within the following intervals [41] : $X_{BL} = 70 \pm 7$ nm, $Y_{BL} = 21 \pm 5$ nm, $X_{AL} = 67 \pm 7$ nm, $Y_{AL} = 77 \pm 5$ nm, $X_{ARC} = 65 \pm 7$ nm, $Y_{ARC} = 89 \pm 5$ nm. The data used in the computational experiment is gathered by simulating a grating with parameter values equal to $[X_{BL}, Y_{BL}, X_{AL}, Y_{AL}, X_{ARC}, Y_{ARC}]$ = [70, 21, 67, 77, 65, 89].

In Figures 2.2, 2.3 and 2.4 we report the results of the presented algorithm for different starting prior vectors and different noise levels, particularly we plot:

(a) the elastic net coefficients against the strength of the regularization parameter, and

(b) the normalized local sensitivities evaluated at the end of the estimation routine and defined as $\sum_n \left| \frac{\partial I_n}{\partial p_i} \cdot \frac{p_i}{I_n} \right|$ where I is the computed intensity, p_i is the parameter of interest and the summation is over the diffraction efficiencies. They are a measure of the overall perturbation of the output due to the slight perturbation of a certain parameter.

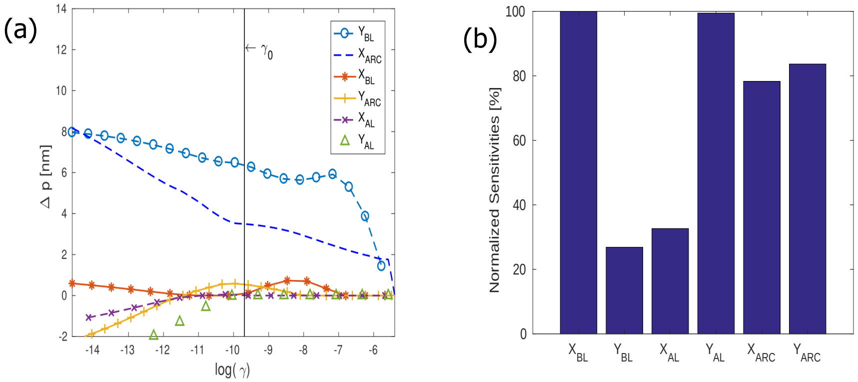

Figure 2.2: Plots at the last iteration of the automatic variable selection algorithm. 5% Gaussian noise is added to the synthetic data. $\mathbf{p}_0$ is [X_{BL}, Y_{BL}, X_{AL}, Y_{AL}, X_{ARC}, Y_{ARC}] = [66.9, 22.47, 73.41, 81.65, 60.21, 93.7] nm. a) Elastic net coefficients as a function of regularization parameter strength. γ_0 is the regularization strength selected according to the criteria in Eq. (2.16). b) Normalized local sensitivities in percentage.

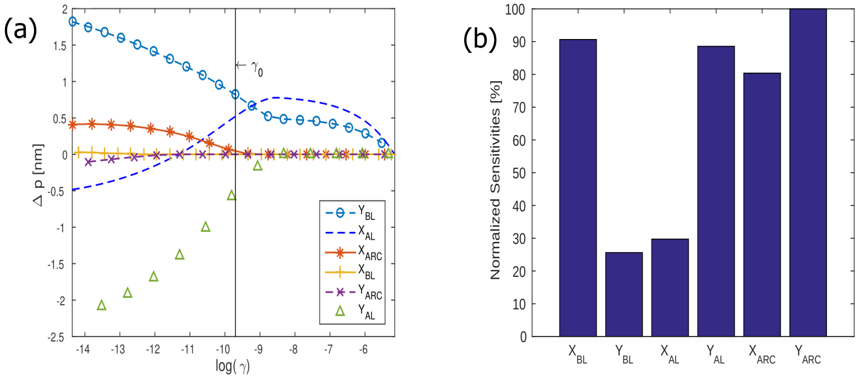

Figure 2.3: Plots at the last iteration of the automatic variable selection algorithm. 10% Gaussian noise is added to the synthetic data. $\mathbf{p}_0$ is [X_{BL}, Y_{BL}, X_{AL}, Y_{AL}, X_{ARC}, Y_{ARC}] = [74.4, 26.05, 61.78, 81.13, 66.85, 84.97] nm. a) Elastic net coefficients as a function of regularization parameter strength. γ_0 is the regularization strength selected according to the criteria in Eq. (2.16). b) Normalized local sensitivities in percentage.

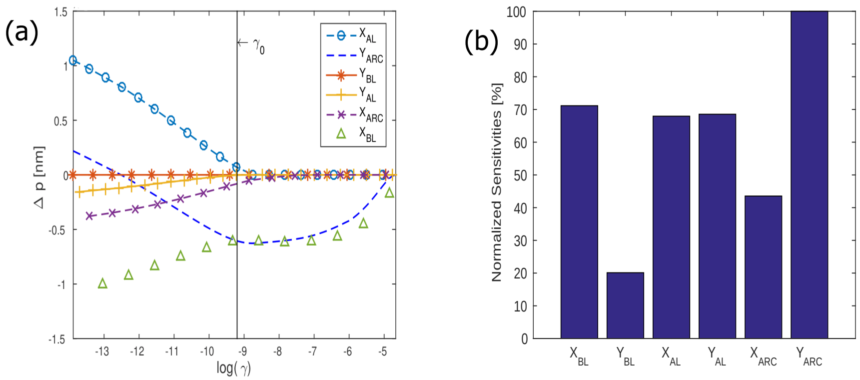

Figure 2.4: Plots at the last iteration of the automatic variable selection algorithm. 15% gaussian noise is added to the synthetic data. $\mathbf{p}_0$ is [X_{BL}, Y_{AL}, X_{AL}, Y_{AL}, X_{ARC}, Y_{ARC}] = [68.15, 23.25, 70.92, 72.81, 71, 91.7] nm. a) Elastic net coefficients as a function of regularization parameter strength. γ_0 is the regularization strength selected according to the criteria in Eq. (2.16). b) Normalized local sensitivities in percentage.

The algorithm converges fast and, once converged, it successfully shrinks some of the entries of $\mathbf{\Delta}_p$ to zero, however these entries depend on the noise level in the data and on the starting point of the optimization. In Figure 2.2 the parameters that were shrunk to exactly zero at the end of the optimization were X_{AL} and Y_{AL}, in Figure 2.3 the selected ones were X_{BL} and Y_{ARC}, while in Figure 2.4 only Y_{BL} was exactly equal to zero. We also notice a change in the local sensitivities, evaluated around the estimated set of parameters $\mathbf{p}_n$, in Figure 2.2(b), 2.3(b) and 2.4(b). Further, a comparison among Figures 2.2, 2.3 (a) and (b) reveals that the parameters that get fixed are not necessarily the ones for which the local sensitivity is the lowest. In other words, the parameters that locally perturb the output the most, may not be the ones that the elastic net locally identifies as important for proper fitting of the data. Also, the ℓ_1 norm strongly biases towards the prior. Hence we used the algorithm only to remove the unnecessary parameters. Once that the redundant variables are identified and fixed by the algorithm the free ones can be estimated. We have done this by employing the same method, but retaining only the ℓ_2 norm penalty in Eq. (2.15) in order to avoid the more aggressive shrink towards $\mathbf{p}_0$ given by the ℓ_1 norm. An example of such estimation, evaluated using the free parameters in Fig. 2.2(a), is given in Table 2.2:

Table 2.2: **Reconstruction results**

parameter	Reconstructed value[nm]	standard deviation [nm]
X_{BL}	67.12	0.8
Y_{BL}	23.65	1.8
X_{ARC}	61.67	0.4
Y_{ARC}	87.85	0.3

where we have approximated the covariance matrix as $(J^T \cdot J)^{-1}\sigma^2$ [34].
In what follows we report our findings for the application of the Morris design to gratings. The input space is discretized in a 12 level grid. We have generated 1000 trajectories and have retained 30 of them which grant the highest "spread" in the input space [29], for a total of $R \cdot (p+1) = 210$ model evaluations. As the model produces 54 outputs - 18 diffraction efficiencies per wavelength - we have analyzed them separately. Plots for an illustrative subset of the diffracted order at $\lambda = 13.398$ nm are shown in Figure 2.5, the other two wavelengths show similar trends.

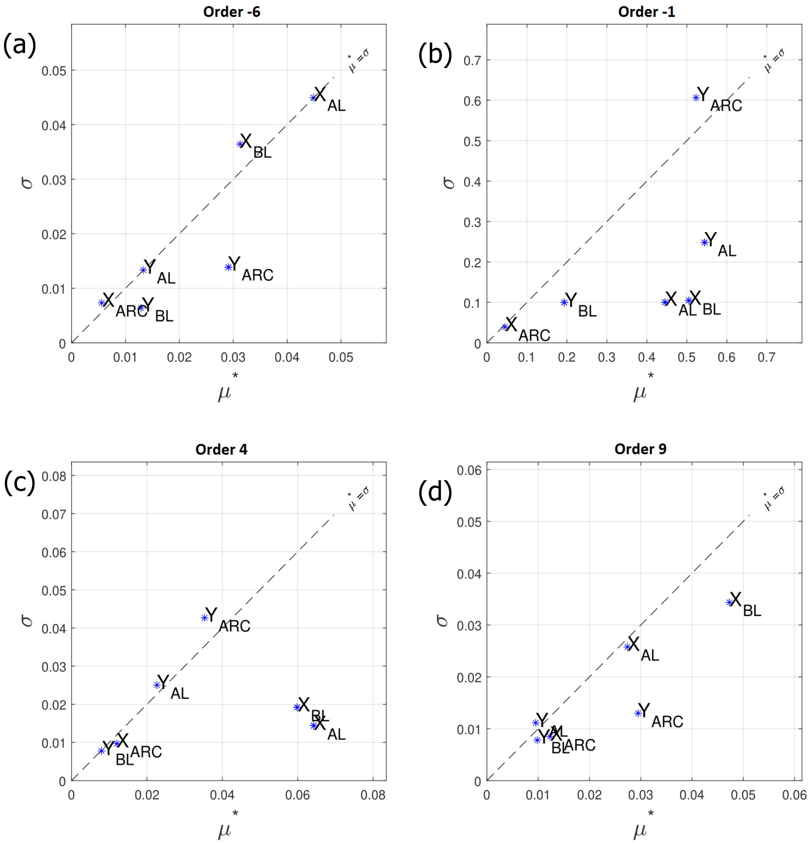

Figure 2.5: Morris plots for four different diffracted orders at $\lambda = 13.398$ nm: a) order -6, b) order -1, c) order 4, d) order 9.

Some conclusions can be drawn from the Morris plots in Figure 2.5:

- The widths of the buffer and of the absorber layer – X_{BL} and X_{AL} – which determine the CDs of the grating, and the thickness of the ARC layer – Y_{ARC} – which determines the amount of incoming power that is transmitted to the grating, are quite separated from the other inputs for most of the diffracted orders. This indicates their importance in the model.

- Y_{AL}, that determines, for a fixed thickness of the buffer layer, the thickness of the absorber, is very important for the orders [-3,4] which are the ones that mostly propagate through the entire height of the absorber. Its importance decreases for orders diffracted at higher angles.

- X_{ARC} and Y_{BL} always appear close the origin of the plot. This indicates that they are the least important inputs in the model and can be considered fixed to a certain value within their uncertainty bounds.

- For all of the diffracted orders some of the inputs are involved in nonlinear effects, which causes them to appear close to the diagonal in Figure 2.5. It is also interesting to note that the degree of nonlinearity or correlation related to a certain input is captured by certain diffraction efficiencies rather than by others. For example, examining the plot for the order -6 (Figure 2.5(a)), X_{BL} and X_{AL} appear to be involved in strong interactions or non–linear behavior. This doesn't appear to be the case for the orders -1 or 4 in Figure 2.5(b–c) .

The observations above are consistent with previous modeling work [3], in which the authors have retained in the model only those parameters that identified top and bottom CDs and the SWA of the grating. However, according to the Morris design, Y_{ARC} should be considered as a free degree of freedom rather than be fixed.
As the Morris design and the elastic net penalty rank the importance of parameters according to different criteria, and as they cover the input space differently, they lead to different results. For instance in Figure 2.2(a) the elastic net penalty fixed the parameters X_{AL} and Y_{AL}. This, according to the Morris design, would have deprived the model of two important variables.
In light of this, a better strategy could be to remove for the model those parameters that are identified as unimportant by both the Morris design and the penalized regression (4). For example, for the case in Figure 2.4(a), one could fix only Y_{BL}. In this way one would retain in the model those parameters that are important for proper fitting of the data and that, at the same time, have a substantial effect over the output.

2.2.3. Application to 3–D scatterers

It is interesting to apply the method developed in Section 2 to the complex case of feature selection for 3–D isolated nanostructures. The model–based approach has been investigated predominantly for 2–D grating profiles and 3–D periodic scatterers but its use for the reconstruction of isolated nanostructures is still to be discussed. The modelling of a 3–D nanostructure is challenging, and understanding how to parametrize a given structure and which features to retain in the model is difficult. In such cases the tools presented above can be particularly useful. We apply the algorithm described in Section 2 to the scatterer in Figure 2.6(a), which is parameterized with 7 parameters. We fit the diffuse scattered intensities, displayed in Figure 2.6(b).

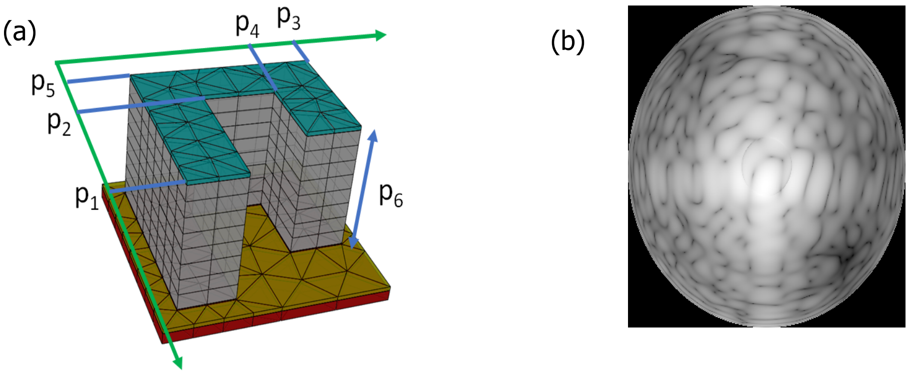

Figure 2.6: a) Scatterer with parameterized profile. The parameter p7, not indicated in figure, is the thickness of the anti–reflective layer. b) Diffuse scattering given by the structure in Fig. 2.6(a).

For the 3D scatterer we use the following data:

Table 2.3: **Layers thicknesses and material properties at $\lambda = 13.5$ nm**

layer	thickness [nm]	n	k
ARC TaBO	2	0.952	0.026
Absorber TaBN	58	0.95	0.031
Ru	0.5	0.88586	0.01727
Ru (Capping layer)	2	0.88586	0.01727
Si	1.8968	0.99888	0.00183
MoSi2	0.7986	0.96908	0.00435
Mo	2.496	0.92347	0.00649
MoSi2	1.8908	0.96908	0.00435

We replace the multilayer with an equivalent substrate that offers, for the given wavelength and angles of incidence, approximately the same reflectance. The incoming light field is a beam with a diameter of about $2\mu m$ and radiating 5e11 photons/sec. The detection NA is 0.5. The computational domain is truncated on all sides by the perfectly matched layers [42]. The meshing setting is such to have a relative error in the far field evaluation of about 1%. The elastic net plots for the isolated scatterer are reported in Figure 2.7.

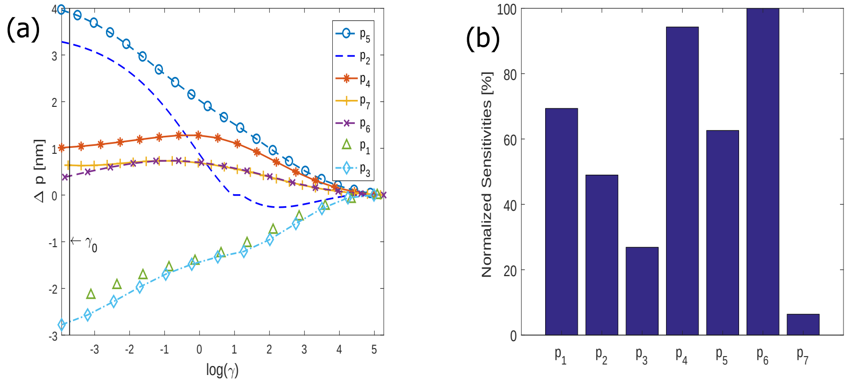

Figure 2.7: a) Elastic net coefficients as a function of the regularization parameter strength. b) Normalized local sensitivities in percentage.

Figure 2.7 reveals that all of the parameters should be kept in the model for a proper fitting of the data, even though the contribution of p_6 and p_7 is quite limited compared to the others. A thorough study of the applicability of model–based reconstruction in 3–D a–periodic case, and related modeling work, is beyond the purpose of this thesis.

2.3. Summary

We have discussed the inverse problem of scatterometry and offered a generic introduction. The focus has been on the reduction of complexity in inverse modeling. We have devised an algorithm for automatic feature selection and compared its outcome with sensitivity analysis methods. These methods can assists the modeler by offering different viewpoints for selection of the appropriate degrees of freedom in a parameterized physical structure.

References

[1] P. Ansuinelli, W. M. J. Coene, and H. P. Urbach, *Automatic feature selection in euv scatterometry,* Appl. Opt. **58**, 5916 (2019).

[2] N. Kumar, P. Petrik, G. K. P. Ramanandan, O. E. Gawhary, S. Roy, S. F. Pereira, W. M. J. Coene, and H. P. Urbach, *Reconstruction of sub-wavelength features and nano-positioning of gratings using coherent fourier scatterometry,* Opt. Express **22**, 24678 (2014).

[3] H. Gross, A. Rathsfeld, F. Scholze, and M. Bär, *Profile reconstruction in extreme ultraviolet (EUV) scatterometry: modeling and uncertainty estimates,* Measurement Science and Technology **20**, 105102 (2009).

[4] M. H. Madsen and P.-E. Hansen, *Scatterometry—fast and robust measure-*

ments of nano-textured surfaces, Surface Topography: Metrology and Properties **4**, 023003 (2016).

[5] C.-H. Ko and Y.-S. Ku, *Overlay measurement using angular scatterometer for the capability of integrated metrology,* Opt. Express **14**, 6001 (2006).

[6] C. Raymond, *Overview of scatterometry applications in high volume silicon manufacturing,* AIP Conference Proceedings **788**, 394 (2005), https://aip.scitation.org/doi/pdf/10.1063/1.2062993 .

[7] C. J. Raymond, S. W. Farrer, and S. Sucher, *Scatterometry for the measurement of metal features,* in *Metrology, Inspection, and Process Control for Microlithography XIV*, Vol. 3998, edited by N. T. Sullivan, International Society for Optics and Photonics (SPIE, 2000) pp. 135 – 146.

[8] O. El Gawhary, N. Kumar, S. Pereira, W. Coene, and H. Urbach, *Performance analysis of coherent optical scatterometry,* Applied Physics B: Lasers and Optics **105**, 775 (2011).

[9] S. Roy, N. Kumar, S. F. Pereira, and H. P. Urbach, *Interferometric coherent fourier scatterometry: a method for obtaining high sensitivity in the optical inverse-grating problem,* Journal of Optics **15**, 075707 (2013).

[10] X. Xu, A. P. Konijnenberg, S. Pereira, and H. P. Urbach, *Phase retrieval of the full vectorial field applied to coherent Fourier scatterometry,* Optics Express **25**, 29574 (2017).

[11] H. Urbach, O. Janssen, S. van Haver, and A. Wachters, *On the modeling of optical systems containing elements of different scales,* Journal of Modern Optics **58**, 496 (2010), https://doi.org/10.1080/09500340.2010.538736 .

[12] P. Lalanne and E. Silberstein, *Fourier-modal methods applied to waveguide computational problems,* Optics letters **25**, 1092 (2000).

[13] M. Pisarenco, *Scattering from finite structures : an extended Fourier modal method*, Ph.D. thesis, Department of Mathematics and Computer Science (2011).

[14] M. Hammerschmidt, M. Weiser, X. G. Santiago, L. Zschiedrich, B. Bodermann, and S. Burger, *Quantifying parameter uncertainties in optical scatterometry using Bayesian inversion,* in *Modeling Aspects in Optical Metrology VI*, Vol. 10330, edited by B. Bodermann, K. Frenner, and R. M. Silver, International Society for Optics and Photonics (SPIE, 2017) pp. 8 – 17.

[15] P. Yla-Oijala, J. Markkanen, S. Jarvenpaa, and S. Kiminki, *Surface and volume integral equation methods for time-harmonic solutions of maxwell's equations,* Progress In Electromagnetics Research **149**, 15 (2014).

[16] B. Bodermann, M. Wurm, A. Diener, F. Scholze, and H. Gross, *Euv and duv scatterometry for cd and edge profile metrology on euv masks,* in *25th European Mask and Lithography Conference* (2009) pp. 1–12.

[17] P.-I. Schneider, X. Garcia Santiago, V. Soltwisch, M. Hammerschmidt, S. Burger, and C. Rockstuhl, *Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction,* ACS Photonics **6**, 2726 (2019), https://doi.org/10.1021/acsphotonics.9b00706 .

[18] X. Chen, S. Liu, C. Zhang, and H. Jiang, *Improved measurement accuracy in optical scatterometry using correction-based library search,* Appl. Opt. **52**, 6726 (2013).

[19] Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, *Identifying and attacking the saddle point problem in high-dimensional non-convex optimization,* NIPS **27** (2014).

[20] J. Fernández-Martínez, T. Mukerji, E. García-Gonzalo, and Z. Fernández-Muñiz, *Uncertainty assessment for inverse problems in high dimensional spaces using particle swarm optimization and model reduction techniques,* Mathematical and Computer Modelling **54**, 2889 (2011).

[21] M. Göhler, J. Mai, and M. Cuntz, *Use of eigendecomposition in a parameter sensitivity analysis of the community land model,* Journal of Geophysical Research **118**, 904 (2013).

[22] P. G. Constantine, E. Dow, and Q. Wang, *Active subspace methods in theory and practice: Applications to kriging surfaces,* SIAM Journal on Scientific Computing **36**, A1500 (2014), https://doi.org/10.1137/130916138 .

[23] F. G. Blanchet, P. Legendre, and D. Borcard, *Forward selection of explanatory variables,* Ecology **89**, 2623 (2008), https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/07-0986.1 .

[24] R. Tibshirani, *Regression shrinkage and selection via the lasso,* Journal of the Royal Statistical Society. Series B (Methodological) **58**, 267 (1996).

[25] H. Zou and T. Hastie, *Regularization and variable selection via the elastic net,* J. R. Statist. Soc. B **67**, 301 (2005).

[26] G. Glen and K. Isaacs, *Estimating sobol sensitivity indices using correlations,* Environmental Modelling & Software **37**, 157 (2012).

[27] N. Farchmin, M. Hammerschmidt, P.-I. Schneider, M. Wurm, B. Bodermann, M. Bär, and S. Heidenreich, *Efficient global sensitivity analysis for silicon line gratings using polynomial chaos,* in *Modeling Aspects in Optical Metrology VII*, Vol. 11057, edited by B. Bodermann, K. Frenner, and R. M. Silver, International Society for Optics and Photonics (SPIE, 2019) pp. 115 – 121.

[28] M. D. Morris, *Factorial sampling plans for preliminary computational experiments,* Technometrics **33**, 161 (1991).

[29] F. Campolongo, J. Cariboni, and A. Saltelli, *An effective screening design for sensitivity analysis of large models,* Environmental Modelling and Software **22**, 1509 (2007).

[30] B. Iooss and P. Lemaître, *A review on global sensitivity analysis methods,* in *Uncertainty management in Simulation–Optimization of Complex Systems: Algorithms and Applications*, edited by G. Dellino and C. Meloni (Springer, 2015) Chap. 5, pp. 101–122.

[31] M.-A. Henn, H. Gross, F. Scholze, M. Wurm, C. Elster, and M. Bär, *A maximum likelihood approach to the inverse problem of scatterometry,* Opt. Express **20**, 12771 (2012).

[32] J. Eriksson, *Optimization and Regularization of Nonlinear Least Squares Problems*, Ph.D. thesis, Ume$\mathring{a}$ University (1996).

[33] A. Doicu, T. Trautmann, and F. Schreier, *Numerical regularization for atmospheric inverse problems* (Springer, 2010).

[34] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing* (Cambridge University Press, New York, NY, USA, 1992).

[35] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics (Springer, 2009).

[36] C. R. Vogel, *Computational Methods for Inverse Problems* (Society for Industrial and Applied Mathematics, 2002) https://epubs.siam.org/doi/pdf/10.1137/1.9780898717570 .

[37] J. Friedman, T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent,* Journal of Statistical Software **33**, 1 (2009).

[38] P. C. Hansen, *Regularization tools – a matlab package for analysis and solution of discrete ill-posed problems,* Numerical Algorithms (1994).

[39] Y. Khare and R. Muñoz-Carpena, *Global sensitivity analysis: Elementary effects method of morris using sampling for uniformity (su) matlab code manual,* `https://abe.ufl.edu/faculty/carpena/software/SUMorris.shtml` (2014).

[40] `https://jcmwave.com/`.

[41] K. Rasmussen, J. B. Kondrup, A. Allard, S. Demeyer, N. Fischer, E. Barton, D. Partridge, L. Wright, M. Bär, H. G. André Fiebach, S. Heidenreich, M.-A. Henn, R. Model, S. Schmelter, G. Kok, and N. Pelevic, *Novel mathematical and statistical approaches to uncertainty evaluation: best practice guide to uncertainty evaluation for computationally expensive models*, Tech. Rep. (Euramet, 2015).

[42] J.-P. Berenger, *A perfectly matched layer for the absorption of electromagnetic waves,* J. Comput. Phys. **114**, 185 (1994).

3

An inverse problem in EUV mask inspection: EUV mask imaging

Parts of this chapter have been published in *Nanoengineering: Fabrication, Properties, Optics, Thin Films, and Devices XVI, International Society for Optics and Photonics (SPIE, 2019)* **11089**, 299-305 [1]

In EUV lithography the small features that get printed on a wafer are the result of the projection of a mask on the wafer. In order to realize a mask a device maker needs to design a chip which is translated into a file format. The file format, which contains the mask layout, is then processed using various equipment to realize the mask. EUV masks consist of a patterned mask blank which is composed of a low thermal expansion substrate – needed to ensure that the absorbed light does not cause distortions in the mask – followed by a series of 40–50 layers of alternating silicon and molybdenum which compose a Bragg mirror followed by a protective layer of ruthenium and an absorber material. The absorber has varying thickness which gives the required pattern of the reflected light and consists of a tantalum boron nitride film topped with an anti-reflective oxide.

The complexity of the EUV masks and of the processes by means of which they are realized makes them vulnerable to defects. Defects on an EUV mask are classified as phase or absorber defects. Phase defects are introduced by the substrate or by particles in the multilayer (ML). Local pits and bumps in the ML cause the ideally flat surface of the mirror to present irregularities of just nanometer height or depth which cause a local intensity drop in the printed image [2]. Absorber defects are caused by the improper patterning of the absorber and can cause a deviation of the printed structures from their desired layout. These defects become critical when they deteriorate the image of the mask on the wafer causing the device not to yield. For this reason the development of mask inspection tools is an important part in the further development of EUV lithography. These inspection tools need to scan the entire mask to locate potential defective sites. The mask blank is firstly inspected. After the inspection the locations of defects are reviewed. The printability of defects is then assessed and one attempts to either avoid the printable defects – a process by means of which the absorber pattern is moved along some degrees of freedom (for instance shift/rotation) to hide the defects beneath the design pattern [3] – or to repair them. The repair process involves an e–beam activated chemically induced etching or deposition process that can compensate for the distortions introduced by the defects [4]. The repair needs to be verified by a further review step to confirm that the mask is defect–free [5]. The mask is subsequently inspected once the absorber has been patterned.

This chapter introduces phase retrieval methods and discusses the framework within which they operate. The methods are used, together with rigorous simulations, to study their applicability for the imaging of patterned EUV masks. We remark that the actinic inspection, which is the inspection done using a wavelength identical to the one used in lithography to print the pattern, thus equal to 13.5 nm, of the masks is particularly attractive. The reason for this is that detecting printable phase defects is difficult when the inspection is carried out with deep ultra–violet (DUV) or e-beam tools – which are at the moment the prevalent solution – because the light at the corresponding wavelength does not penetrate the ML with appreciable amplitude [6]. In order to reveal such defects the inspection should be carried out via an actinic microscope. Further, images obtained using DUV or e-beam are sometimes quite different from the aerial images of a scanner. In other words an actinic microscope can mimic the scanner image. A few solutions for actinic inspection, presented in the literature are listed below.

A few actinic microscopes. Here are reported a few setups for actinic inspection of defects.

A Fresnel zoneplate microscope: The SEMATECH Berkeley actinic inspection tool (AIT). This is a "direct" EUV inspection tool and it employs a bending magnet at the Advanced Light Source synchrotron at Lawrence Berkeley National Laboratory as its source. It employs a system of mirrors to focus the light at 6° on the EUV mask. This is a dual–mode microscope that can be operated in two different modalities [7]. In the *scanning mode* the mask moves continuously in the 2D plane of the mask under the focused EUV spot, of dimensions of about $1\mu m$, while the reflected light intensity is measured by a photo-diode. Both bright-field (specular reflectivity) and dark-field scattering signals can be recorded. In the *imaging mode* a Fresnel zoneplate lens projects an image of the illuminated mask surface, field size of about $40\mu m$ onto an EUV charge–coupled device (CCD) camera with a magnification of approximately 670×, or higher.
This tool was designed to initially emulate the resolution of a 0.25–NA 4x stepper (a tool that prints directly in one illumination step a full image of the chip area on a wafer with a demagnification of four times with respect to the mask reticle), although other numerical aperture vales (0.3 and 0.35) were later made available to the users [8]. Extensive work has been done by the researchers at SEMATECH which has continued with the development of the SHARP microscope for future technology nodes [9, 10].
As this solution directly projects the image at the camera, phase information is not available. Two microscopes that can image the phase of the target are reported below.

Zernike's phase contrast microscopy. A way to retrieve the phase information, or to image phase objects, is by means of phase contrast microscopy [11]. We briefly discuss this method following Goodman [12].
Suppose that a transparent phase object is described by its transmission function

$$t(\mathbf{r}) = \exp(i\Phi(\mathbf{r})), \tag{3.1}$$

and that it is illuminated by a coherent spot which can be approximated as a plane wave. We assume that the magnification is equal to unity and ignore the finite extent of the NA of the lenses and any other non–idealities. We express the phase in Eq. (3.1) as small variations about an average value Φ_0 via a first order Taylor expansion:

$$t(\mathbf{r}) = e^{i\Phi_0}e^{i\Delta\Phi} \approx e^{i\Phi_0}[1 + i\Delta\Phi(\mathbf{r})], \tag{3.2}$$

where higher order terms have been neglected. The first term in Eq. (3.2) represents a strong wave component, the undiffracted or zero–th order beam, that passes through the sample accumulating a phase shift Φ_0, while the second term generates diffracted light that is deflected from the optical axis. The image produced by a brightfield microscope is then written as

$$I_i \approx |1 + i\Delta\Phi|^2 \approx 1, \tag{3.3}$$

where, as done previously, the second order term has been neglected. The diffracted light arising from the phase structure is not observable in the focal plane because it is in phase quadrature with respect to the strong undiffracted beam and it does not lead to appreciable intensity fluctuations in the image plane. Zernike realized this and proceeded to change the interference between the two beams by placing a phase changing plate in the pupil plane (where only the phase of the undiffracted beam is changed). The phase of the undiffracted beam was retarded with either $\pi/2$ or $3\pi/2$ radians with respect to the diffracted beams, for which the intensity of the image becomes either

$$I_i = |\exp(i\pi/2) + i\Delta\Phi|^2 = |i(1 + \Delta\Phi)|^2 \approx 1 + 2\Delta\Phi, \tag{3.4}$$

or

$$I_i = |\exp(i3\pi/2) + i\Delta\Phi|^2 = |-i(1 - \Delta\Phi)|^2 \approx 1 - 2\Delta\Phi. \tag{3.5}$$

The image intensity depends linearly on the phase modulation of the phase object which can now yield appreciable intensity variations at the image plane.
The use of a phase contrast microscope for EUV mask inspection was proposed in [13]. That study demonstrated that the phase contrast method is suitable for the inspection of phase defects (bumps and pits in the ML) and it reported studies on the signal strength of the defects with respect to a through focus scan and on the impact of their size.

Phase retrieval and ptychography. This class of methods will be presented more thoroughly in the next part of this chapter and a detailed description of these methods will be given later on. Here it should be mentioned that lensless imaging has been proposed and studied experimentally as a possible solution for actinic inspection of EUV masks for instance in [14–16], where authors used synchrotron light and a zone plate for the coherent illumination of the mask.

The remainder of this chapter (and of this thesis) is focused on iterative phase retrieval methods. This class of methods is particularly attractive in that

- they enable the retrieval of phase information on top of the amplitude one, which may lead to a higher sensitivity in the detection of defects,
- they could be employed as a technological solution for the metrology of phase shifts imparted by phase–shift masks which may be adopted as a technological solution for next–generation EUV lithography, and
- they reduce experimental complexity to a minimum (although this comes at the cost of an increased computational burden).

In what follows we shall discuss and familiarize with phase retrieval methods while introducing results that will be necessary for the development of the algorithms that will be discussed in the next chapter. We will focus exclusively on iterative methods. For a broader discussion of methods for imaging of the phase please refer to [17].

3.1. Iterative phase retrieval methods

The *phase problem* appears in a variety of fields among which microscopy, wave front sensing and crystallography and originates in cases in which it is not possible to record the phase in experimental set–ups. In the usual case an object, described by a certain complex-valued transmission or reflection function $f(\mathbf{r})$, is illuminated by a coherent beam and the diffracted light is recorded in the far zone where – as we described in Section 1.1.1 – the Fourier spectrum, $F(\mathbf{k})$, is recorded in intensity. The problem is then to retrieve the phase of $f(\mathbf{r})$ from the sole amplitude information. In what follows we discuss this problem. We focus firstly on generic physical requirements that ought to be satisfied to tackle the phase problem and we subsequently discuss the mathematical aspects of phase retrieval.

3.1.1. Sampling and coherency requirements.

Sampling requirements. An early observation about the phase problem was made in 1952 by Sayre [18] where the author noted that while we wish to reconstruct $f(\mathbf{r})$, we should keep in mind that the sampling is done on the intensity of its Fourier transform, $|F(\mathbf{k})|^2$, rather than on its complex–valued amplitude. Because of the autocorrelation theorem the inverse Fourier transform of the intensity in the momentum space is actually related to the autocorrelation of the object function rather than to the object function itself:

$$|F(\mathbf{k})|^2 = \mathcal{F}\int f^*(\mathbf{r}')f(\mathbf{r}+\mathbf{r}')d\mathbf{r}'.$$

If we assume that the support of the object is R, than the support of the autocorrelation will be $2R$ and therefore the sampling at the far–field ought to be at twice the rate than for the measurement of the complex field. This means that the pixel spacing in the momentum domain has an upper bound of $\Delta k \leq 1/2R$. It can be shown [12] that in the far field it holds that $\mathbf{k} = [k_x, k_y] = [\frac{x}{\lambda z_0}, \frac{y}{\lambda z_0}]$, therefore it holds that for the sampling on the detector in the far field

$$\Delta x \leq \frac{\lambda z_0}{2R}. \tag{3.6}$$

Because the sampling in Fourier space is dictated by the field–of–view (FOV) in real space of the sample plane, the finer sampling is achieved by zero–padding the region around the object which experimentally means that the space outside the support of the object has been filled with an optically transparent or opaque mask which does not contribute to the diffraction pattern (apart from the undiffracted beam).

A different way to look at the oversampling requirement was presented by Miao et al. in [19]. It was noted that, given that experimentally we retrieve intensity information only, the data consists of:

$$|F_m(\mathbf{k})| = \left|\sum_{n=0}^{N-1} f_n(\mathbf{x}) e^{-i\frac{2\pi}{N}mn}\right| \quad \text{with} \quad m = 0, 1, .., N-1. \tag{3.7}$$

When $f(\mathbf{x})$ is complex valued and no *a–priori* information is available, retrieving the phase from Eq. (3.7) is not possible because this equation is underdetermined by a factor 2. Although Eq. (3.7) refers to the 1D case the concept does not change in higher order dimensions.

Spatial coherence. It is necessary for phase retrieval to be successful that the scattered waves interfere at the detector and that the fringe visibility is not too low. This translates in tight requirements in terms of coherence. It is demonstrated in [20] that a coherence length larger than the extension of the autocorrelation of the object is required to properly sample the Fourier transform of the autocorrelation of the object. This is also understandable by Schell's theorem which states that in the far–field the diffraction pattern is proportional to the Fourier transform of the product among the complex degree of coherence and the autocorrelation of the diffracting structure [21]. Therefore the beam needs to illuminate the isolated object and the degree of coherence must be larger than the size of the autocorrelation of the object.

Temporal coherence. The requirement in the temporal coherence length, $l_{Coh}^T = \lambda^2/\Delta\lambda$, is that it has to be at least as great as the maximum path difference among any pair of interfering rays [20]. This means that the coherence length must be bigger than the path difference between rays from opposite sides of the sample to the outer edge of the detector. It can be shown [17, 20] that this leads to a condition that interrelates the coherence length to the width of the object and the maximum scattering angle:

$$l_{Coh}^T > 2R\theta_{MAX}. \tag{3.8}$$

3.1.2. Phase retrieval from a single diffraction pattern.

Mathematically speaking the generic phase problem aims to reconstruct an unknown signal $f(\mathbf{r}) \in \mathbb{C}^n$ satisfying M quadratic constraints

$$\begin{aligned} &\text{find} && f(\mathbf{r}), \\ &\text{such that} && y_k = |\langle \mathbf{a}_k, f(\mathbf{r})\rangle|^2 \quad k = 1, ..., M, \end{aligned} \tag{3.9}$$

where $y_k \in \mathbb{R}$ is the measurement at pixel k, $\mathbf{a}_k \in \mathbb{C}^n$ is a discretized and known operator and $\langle\cdot,\cdot\rangle$ denotes the inner product. Notice that the Fourier phase problem

in Eq. (3.7), which will be the focus of the following work, is a particular instance of Eq. (3.9) and that additional geometric constraints may apply.
Suitable approaches to the problem can be devised by rewriting Eq. (3.9) in different forms.

Alternate Projection based methods. In case in which the forward operator is a Fourier transform we can rewrite the problem in Eq. (3.9) as

$$\begin{aligned} \text{minimize} \quad & ||\mathcal{F}(S(\mathbf{r})f(\mathbf{r})) - b||^2 \\ \text{such that} \quad & |b| = \sqrt{y} \end{aligned} \tag{3.10}$$

where we have introduced the support, $S(\mathbf{r})$, which defines – in the domain in which the object is present – a non–zero window over the entire image:

$$S(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

where Ω is the domain in which the object is present. Eq. (3.10) says that we should seek a complex object function that in real space can be factorized into $b(\mathbf{r}) = S(\mathbf{r})f(\mathbf{r})$ – meaning an object function that in real space occupies a bounded domain – and that in Fourier space has an amplitude given by the observations $\sqrt{y}$. This has led to the proposal of a few algorithms that tackle the phase problem using an alternating projections approach in which one propagates back and forth among real and Fourier space, one iteration minimizing the quadratic functional in Eq. (3.10) by enforcing the support constraint and the next enforcing a magnitude equal to $\sqrt{y}$ to the Fourier transform of the reconstructed signal [22]. Because the projector is an operator that brings a point of a set to the closest point of the other set, this class of algorithms can have a *zig–zag* path in the solution space which may cause convergence issues when approaching the local minimum (Fig. 3.1(a)). Various algorithms have been introduced that attempt to improve convergence, by altering the path of the retrieved signal in the solution space, by changing the projectors (example for the reflector operator Fig. 3.1(b)).
A list that describes the most used projectors can be found in [23]. Below we report, as a particular case, the update rules of the Hybrid Input–Output (HIO) and of the relaxed averaged alternating reflections (RAAR) that will be employed later on:

$$\begin{aligned} f_{n+1}^{HIO}(\mathbf{k}) &= [\mathbf{P_S}\mathbf{P_M} + (\mathbf{I} - \mathbf{P_S})(\mathbf{I} - \beta\mathbf{P_M})]f_n^{HIO}(\mathbf{k}), \\ f_{n+1(\mathbf{k})}^{RAAR} &= [\frac{1}{2}\beta(\mathbf{R_S}\mathbf{R_M} + \mathbf{I}) + (\mathbf{I} - \beta)\mathbf{P_M}]f_n^{RAAR}(\mathbf{k}), \end{aligned} \tag{3.12}$$

where $\mathbf{P_S}$ and $\mathbf{P_M}$ are the projectors that are needed to satisfy the support and magnitude constraints in real and momentum space respectively:

$$\begin{aligned} \mathbf{P_S} &: f(\mathbf{r}) \rightarrow S(\mathbf{r})f(\mathbf{r}), \\ \mathbf{P_M} &: f(\mathbf{r}) \rightarrow \mathcal{F}^{-1}\left\{\frac{F(\mathbf{k})}{|F(\mathbf{k})|}\sqrt{y}\right\} \end{aligned} \tag{3.13}$$

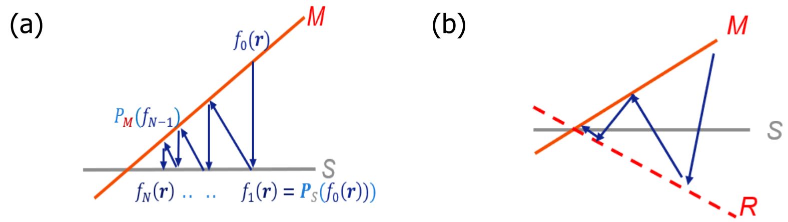

Figure 3.1: a) An illustrative example of the *zig–zag* path of the solution in its space. Many projections take place close to the optimal point that can cause stagnation or long convergence time. b) The reflector projector can mitigate the convergence issues by overshooting the solution twice as far than the support projector.

where y is the recorded intensity and $\mathbf{R} = 2\mathbf{P_S} - \mathbf{I}$ is the reflector operator. In [23] the author noticed heuristically that these methods present a different trajectory in the solution space than the one in Fig. 3.1(a). For instance it was found that the HIO can approach the local minimum in a spiral.

Phase retrieval as a cost functional minimization problem. We could also think to tackle the problem in Eq. (3.10) by directly fitting the diffracted amplitudes:

$$\text{minimize} \quad \mathcal{L}(F, y) \coloneqq |||F(\mathbf{k})| - \sqrt{y}||^2. \tag{3.14}$$

The problem of minimizing a real–valued functional of complex variables is usually tackled with Wirtinger calculus [24]. Particularly Brandwood noticed how the *steepest descent* direction is given by the component of the gradient along the complex conjugate variable [25]. The update rule given by the steepest descent is then given by:

$$f_{n+1}(\mathbf{r}) = f_n(\mathbf{r}) - \mu \nabla_{f^*} \mathcal{L}(F, y), \tag{3.15}$$

where ∇_{f^*} is the component of the gradient along the complex conjugate variable f^*. It can be shown that [17, 26]

$$\nabla_{f^*} \mathcal{L}(f, y) = S(\mathbf{r}) \left[f_n(\mathbf{x}) - \mathcal{F}^{-1} \left(\frac{F(\mathbf{k})}{|F(\mathbf{k})|} \sqrt{y} \right) (\mathbf{r}) \right], \tag{3.16}$$

and one can see that for $\mu = 1$ Eqs. (3.15-3.16) are the update rule of the *error reduction* algorithm [26] which is then given by the direct application of $\mathbf{P_S}$ and $\mathbf{P_M}$ defined in the previous section. The fact that phase retrieval can be cast as a cost functional minimization problem has been instrumental in the development of more advanced optimization algorithms, such as proximal methods [27], gradient descent approaches with dynamic stepsize (Wirtinger flow) [28] and methods based on convex relaxation [29] and sparsity [30] to mention a few. In the remainder of this thesis we will approach the phase retrieval problem predominantly as a cost functional minimization problem.

A note on uniqueness It is worthwhile to notice that the phase problem is non–convex and therefore there are no theoretical guarantees on the convergence to the global minimum. The question of uniqueness was studied in some details in a number of references and it was shown in [31–33] that in the noiseless case, for dimensions $N \geq 2$, a real signal with a finite support can in some cases be uniquely determined – apart from the *trivial* ambiguities of a global phase shift, conjugation and spatial shift – by the magnitude of its Fourier transform. In practice the inclusion of prior information has been shown in some cases to yield uniqueness [34, 35]. Anyway the question of uniqueness remains open and it should be clear that there is no guarantee that the ground truth is retrieved when applying the algorithms on the measured data.

3.1.3. Ptychography

Ptychography is a form of coherent diffractive imaging (CDI) that employs a scanning probe, $P(\mathbf{r})$ to sequentially illuminate an object at J partially overlapping probe positions specified by a position vector $\mathbf{R}_j$. The illuminated object is assumed to be optically thin and it is described by its transmission/reflection function $O(\mathbf{r})$. It is then assumed that it is possible to rewrite the exit–wave at the j–th probe position as

$$\psi_j(\mathbf{r}) = P(\mathbf{r} - \mathbf{R}_j)O(\mathbf{r}), \tag{3.17}$$

and consequently the assumed intensity at the j–th probe position $\mathbf{R}_j$ is modeled as

$$I_j(\mathbf{k}) = |\mathcal{F}(P(\mathbf{r} - \mathbf{R}_j)O(\mathbf{r}))|^2. \tag{3.18}$$

The partial overlap at different probe positions and the scanning of the probe introduce redundancy and translational diversity in the data set which is subsequently computationally processed to retrieve a large field–of–view and a diffraction limited image of the studied object in both amplitude and phase.

It can be shown [17] that the ptychographic phase retrieval problem, in analogy with the single intensity case, can be thought of – per probe position– as

- either the problem of finding the function that lies simultaneously in the domain of the two operators $\mathbf{P_S}$ and $\mathbf{P_M}$ where

$$\begin{aligned} (\mathbf{P_S}f)_j(\mathbf{r}) &= P(\mathbf{r} - \mathbf{R}_j)O(\mathbf{r}) \\ (\mathbf{P_M}f)_j(\mathbf{r}) &= \mathcal{F}^{-1}\left\{\frac{F_j(\mathbf{k})}{|F_j(\mathbf{k})|}\sqrt{I_j(\mathbf{k})}\right\} \end{aligned} \tag{3.19}$$

- or the problem of fitting the diffracted amplitudes:

$$\text{minimize} \quad \mathcal{L}(O, I_j) := \left\| |\mathcal{F}(P(\mathbf{r} - \mathbf{R}_j)O(\mathbf{r}))| - \sqrt{I_j(\mathbf{k})} \right\|^2. \tag{3.20}$$

Because the derivations of Eqs. (3.19, 3.20) are a generalization of what has been presented above and because they are outlined elsewhere [17] the details are omitted here. The derivations will be carried out explicitly in the next chapters at those occasions where it will be necessary.

It is worthwhile to notice that ptychography, a technique that was initially thought of as a means to retrieve the phase of an exit wave by letting the diffracted orders interfere among themselves [36], and has begun in the area of scanning transmission electron microscopy by Rodenburg [37], has developed into an extremely powerful computational imaging technique. On top of the retrieval of the transmission/reflection function of the object [37] it is now possible to reconstruct also the illumination [38–40] together with its eigenmodes when the probe is partially coherent [41, 42]. Experimental efforts and algorithm advancements paved the way towards reconstruction of thick samples [43, 44], three dimensional imaging of biological samples [45] and high resolution imaging with a synthetic high numerical aperture as done with Fourier ptychography at visible [46] and X-ray [47] wavelengths to mention a few.

It has to be emphasized that the only physical assumption made in the ptychographic methods is that the exit wave can be partitioned into the product of the probe and the object as in Eq. (3.17). In contrast with the method employed in scatterometry, presented in the previous chapter, we do not use any physics based and computationally heavy predictive model to infer the output. In this sense it could be argued that ptychography is somehow similar to a *model free* approach where the relevant information is inferred solely from the data, although this comes at the price of accuracy as the forward model is not a fully rigorous one. By relying on an approximate modeling of light–matter interaction, the ptychographic approach provides a very flexible framework for diffraction–limited imaging through phase retrieval which – together with its relatively easy implementation – has been one of the reasons of its wide appeal and success. At the same time one may wonder whether the ptychographic reconstruction could be ameliorated by including *a–priori*, domain specific and physics–based knowledge. Given that the most accurate way to compute light–matter interaction is given by solving Maxwell's equations we can ask ourselves whether a way can be devised to interface the ptychographic method with the rigorous *model–based* methods. This question, on top of the technological interest for actinic inspection of EUV layouts, motivates the research work that follows below and in the next chapter.

3.2. EUV mask feature reconstruction via phase retrieval

In this section we apply some of the concepts and methods that have been introduced in this chapter to the problem of EUV mask imaging. The methods and algorithms are reported in more details below.

Hybrid Input–Output (HIO). As discussed in Subsection 3.1.2 the hybrid input–output (HIO) algorithm [26], makes use of constraints in conjugated planes to drive

the reconstruction. The recorded field amplitude $|F(\mathbf{k})|$ is used as a constraint in the Fourier space, while the extent of the object in real space, or support, is used as a constraint in real space. Information about the dimension of the object in real space is available even without any prior information, as it can be computed by an inverse Fourier of the recorded intensity which yields the auto-correlation of the object function [48]. The algorithm proceeds by iteratively alternating projections between Fourier space and real space.
At the n–th iteration:

1. given the current estimate of the object $f_n(\mathbf{r})$
2. we Fourier transform $f_n(\mathbf{r})$ to get the estimated diffraction field $F_n(\mathbf{k})$.
3. The current guessed Fourier amplitudes $|F_n(\mathbf{k})|$ are replaced by the measured ones, $\sqrt{y}$, while the phase of the current iterate is retained to get an updated diffracted field $F'(\mathbf{k})$.
4. The current Fourier estimate is propagated back to real space by means of an inverse Fourier transform $f'(\mathbf{r})$, and
5. the support constraint, S, is used to get an updated estimate of the object:

$$f_{n+1}(\mathbf{r}) = \begin{cases} f_n'(\mathbf{r}), & \text{if } \mathbf{r} \in S \\ f_n(\mathbf{r}) - \beta f_n'(\mathbf{r}), & \text{otherwise.} \end{cases} \tag{3.21}$$

In Eq. (3.21) the definition of S is slightly different from the case of a real object (an object described by a real index of refraction) because, for the case of objects described by a complex refractive index, S denotes the portion of space where at the same time the object is non–zero and the imaginary part of the retrieved object function is positive [19]. This is a way to include a priori information of the object in the problem in the sense that for all of the materials that we considered in this thesis the index of refraction, $\mathfrak{n} = n + ik$, has a positive k.

Relaxed Averaged Alternating Reflections (RAAR). The RAAR algorithm tries to approach the local solution of the phase problem by projecting the current solution to a closed set which is twice as far. This is done introducing the reflector operator $\mathbf{R} = \mathbf{I} + 2(\mathbf{P} - \mathbf{I})$, where $\mathbf{I}$ is the identity operator and $\mathbf{P}$ is the projection operator as previously explained in Section 3.1.2 [23, 49].
The abve step in the HIO is replaced by:

$$f_{n+1}(\mathbf{r}) = \begin{cases} f_n'(\mathbf{r}), & \text{if } \mathbf{r} \in S \\ \beta f_n(\mathbf{r}) - (1 - 2\beta) f_n'(\mathbf{r}), & \text{otherwise.} \end{cases} \tag{3.22}$$

Ptychography. In ptychography, a set of computed far–field diffraction patterns is processed by certain algorithms to get to an image via phase retrieval. In this work we have used the ptychographical iterative engine (PIE). At the n–th iteration:

1. Given the guessed object $O_{g,n}$,
2. Multiply the current guessed object by the illumination function at position **R** to get the guessed exit wave at the current position
$$\psi_{g,n}(\mathbf{r},\mathbf{R}) = P(\mathbf{r}-\mathbf{R})O_{g,n}(\mathbf{r}), \tag{3.23}$$
3. Fourier transform the exit wave $\psi_{g,n}(\mathbf{r},\mathbf{R})$ to get $\Psi_{g,n}(\mathbf{k},\mathbf{R}) = \mathcal{F}[\psi_{g,n}(\mathbf{r},\mathbf{R})] = |\Psi_{g,n}|e^{i\theta_{g,n}(\mathbf{k},\mathbf{R})}$.
4. Enforce the measured amplitude at the far–field on the current far–field guess, to get a corrected guess:
$$\Psi_{c,n}(\mathbf{k},\mathbf{R}) = |\Psi|e^{i\theta_{g,n}(\mathbf{k},\mathbf{R})}. \tag{3.24}$$
5. Inverse Fourier transform to get a revised exit wave:
$$\psi_{c,n}(\mathbf{r},\mathbf{R}) = \mathcal{F}^{-1}[\Psi_{c,n}(\mathbf{k},\mathbf{R})]. \tag{3.25}$$
6. Update the transmission (or reflection) function of the object in the area covered by the probe by
$$O_{g,n+1} = O_{g,n} + \frac{|P(\mathbf{r}-\mathbf{R})|}{|P_{max}(\mathbf{r}-\mathbf{R})|}\frac{P^*(\mathbf{r}-\mathbf{R})}{(|P^*(\mathbf{r}-\mathbf{R})|^2+\alpha)}\beta(\psi_{c,n}(\mathbf{r},\mathbf{R}) - \psi_{g,n}(\mathbf{r},\mathbf{R})), \tag{3.26}$$
where α prevents division by zero and β is an appropriately chosen feedback parameter.
7. Move the probe to the next position $\mathbf{R}$ so that there is sufficiently overlap between neighboring probes.
8. Repeat 2 to 7 for a sufficient number of iterations.

3.2.1. Numerical Results

Single Intensity phase retrieval

In the simulation study that follows we employ the methods described above to retrieve the phase of nontrivially shaped isolated nanostructures. The x–polarized EUV radiation illuminates a Mo/Si multilayer coated reflective mask, with a patterned absorber profile on top of it. The angle of incidence is 6°, for which the multilayer is in resonance, giving a reflectance of 60–70%. We replace the multilayer with an equivalent substrate that offers, for the given wavelength and angles of incidence, approximately the same reflectance. In order to do so we compute the reflectance of the multilayer via the transfer matrix method (TMM) [50]. In the TMM algorithm

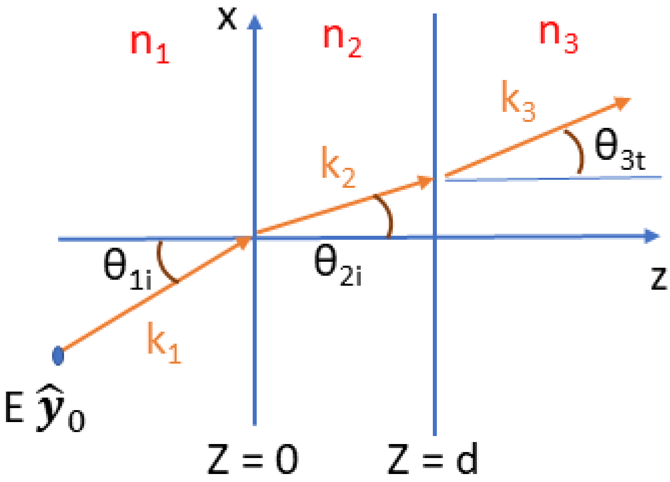

Figure 3.2: One y polarized plane wave illuminated a system made by 3 successive materials and two interfaces at $z = 0$ and $z = d$.

the multilayer response is computed by cascading the responses of single layers which are accounted for in a matrix $\mathcal{M}$. To compute the theoretical response of the multilayer we consider firstly the case of a single layer illuminated by an s polarized (y–polarized) plane wave 3.2 The problem depicted in Fig. 3.2 consists of a boundary value problem which is solved by solving the Helmholtz equation separately at each layer and enforcing boundary conditions (continuity of the tangential component of electric and magnetic field) at the interface between every two layers. We have for the permittivity $\epsilon(z)$:

$$\epsilon(z) = \begin{cases} \epsilon_1, & \text{for } z < 0, \\ \epsilon_2, & \text{for } 0 < z < d, \\ \epsilon_3, & \text{for } z > d. \end{cases} \tag{3.27}$$

.

The materials are further assumed to be isotropic. The generic solution of the electric field, which is assumed to be polarized in the y direction is:

$$E(x, z) = \begin{cases} A_1 e^{i(k_{1x}x + k_{1z}z)} + B_1 e^{i(k_{1x}x - k_{1z}z)}, & \text{for } z < 0, \\ A_2 e^{i(k_{2x}x + k_{2z}z)} + B_2 e^{i(k_{1x}x - k_{2z}z)}, & \text{for } 0 < z < d, \\ A_3 e^{i(k_{3x}x + k_{3z}(z-d))}, & \text{for } z > d, \end{cases} \tag{3.28}$$

where $A_{i=1,2,3}$ are the coefficients for the forward propagating waves and $B_{i=1,2,3}$ are the coefficients for the backwards propagating waves. The amplitude of the magnetic field can be computed by the relation $\mathbf{H}(x, z) = \mathbf{k} \times \mathbf{E}/\omega\mu$ to give:

$$H(x,z) = \begin{cases} \frac{A_1}{\omega\mu}\begin{bmatrix} -k_{1z} \\ 0 \\ k_{1x} \end{bmatrix} e^{i(k_{1x}x+k_{1z}z)} + \frac{B_1}{\omega\mu}\begin{bmatrix} k_{1z} \\ 0 \\ k_{1x} \end{bmatrix} e^{i(k_{1x}x-k_{1z}z)}, & \text{for } z < 0, \\ \frac{A_2}{\omega\mu}\begin{bmatrix} -k_{1z} \\ 0 \\ k_{1x} \end{bmatrix} e^{i(k_{2x}x+k_{2z}z)} + \frac{B_2}{\omega\mu}\begin{bmatrix} k_{2z} \\ 0 \\ k_{2x} \end{bmatrix} e^{i(k_{1x}x-k_{2z}z)}, & \text{for } 0 < z < d, \\ \frac{A_3}{\omega\mu}\begin{bmatrix} -k_{3z} \\ 0 \\ k_{3x} \end{bmatrix} e^{i(k_{3x}x+k_{3z}(z-d))}, & \text{for } z > d. \end{cases} \tag{3.29}$$

Because we assume A_1 and $\mathbf{k}_1$ to be known, and because k_x is continuous across interfaces and k_z is determined by angles which are known via the laws of reflection and refraction the only unknowns in Eqs. (3.28) and (3.29) are $[A_2,\ A_3]$ and $[B_1,\ B_2]$. The four equations necessary to solve the problem are given by the boundary conditions at the two interfaces. At the first interface at $z = 0$ it holds for the electric and magnetic field respectively:

$$\begin{cases} E_1 = A_1 + B_1 = A_2 + B_2, \\ H_1 = \sqrt{\frac{\epsilon_0}{\mu_0}}(A_1 - B_1)n_0 \cos(\theta_{i1}) = \sqrt{\frac{\epsilon_0}{\mu_0}}(A_2 - B_2)n_1 \cos(\theta_{i2}), \end{cases} \tag{3.30}$$

where E_1 and H_1 are the tangential components of the electric and magnetic field at the first interface. At the second interface at $z = d$ it holds for the electric and magnetic field respectively:

$$\begin{cases} E_2 = A_2 e^{ik_{2z}d} + B_2 e^{-ik_{2z}d} = A_3, \\ H_2 = \sqrt{\frac{\epsilon_0}{\mu_0}}(A_2 e^{ik_{2z}d} - B_2 e^{ik_{2z}d})n_1 cos(\theta_{i1}) = \sqrt{\frac{\epsilon_0}{\mu_0}} A_3 n_3 cos(\theta_{t3}), \end{cases} \tag{3.31}$$

where E_2 and H_2 are the tangential components of the electric and magnetic field at the second interface. The system of Equations (3.31) can be solved for A_2 and B_2 which once substituted in the systems of Equations (3.30) yields a matrix that links the fields at the two interfaces:

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = \begin{bmatrix} \cos(k_1 d \cos\theta_{i2}) & (i \sin(k_1 d \cos\theta_{i2})/Y_1) \\ Y_1 i \sin(k_1 d \cos\theta_{i2}) & \cos(k_1 d \cos\theta_{i2}) \end{bmatrix} \begin{bmatrix} E_2 \\ H_2 \end{bmatrix} = \boldsymbol{\mathcal{M}}_1 \begin{bmatrix} E_2 \\ H_2 \end{bmatrix}, \tag{3.32}$$

where we have introduced the complex admittance $Y_1 = \sqrt{\epsilon_0/\mu_0} n_1 \cos(\theta_{i2})$. Because tangential components of the electromagnetic field are continuous across interfaces we can compute the response of the whole multilayer structure as a sequence of products of single–layer matrices:

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = \boldsymbol{\mathcal{M}}_1 \boldsymbol{\mathcal{M}}_2 \ldots \boldsymbol{\mathcal{M}}_n \begin{bmatrix} E_{n+1} \\ H_{n+1} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} E_{n+1} \\ H_{n+1} \end{bmatrix}. \tag{3.33}$$

We come back now at the single layer case. We assume that medium 1 in Eq. (3.27) is air and that medium 3 is a certain substrate and we define the respective admittance $Y_0 = \sqrt{\epsilon_0/\mu_0} n_0 \cos(\theta_{i1})$ and $Y_S = \sqrt{\epsilon_0/\mu_0} n_3 \cos(\theta_{t3})$ so that we can use Eqs. (3.30–3.31) in Eq. (3.32) to obtain

$$\begin{bmatrix} A_1 + B_1 \\ (A_1 - B_1)Y_0 \end{bmatrix} = \boldsymbol{\mathcal{M}}_1 \begin{bmatrix} A_3 \\ A_3 Y_s \end{bmatrix}. \tag{3.34}$$

We can now introduce the reflection and transmission coefficients ,$r = B_1/A_1$ and $t = A_3/A_1$, and expand Eq. (3.34) to find:

$$\begin{cases} 1 + r = m_{11} t + m_{12} Y_s t, \\ (1 - r)Y_0 = m_{21} t + m_{22} Y_s t. \end{cases} \tag{3.35}$$

If we solve for r and t we get :

$$\begin{aligned} r &= \frac{Y_0 m_{11} + Y_0 Y_S m_{12} - m_{21} - Y_S m_{22}}{Y_0 m_{11} + Y_0 Y_S m_{12} + m_{21} + Y_S m_{22}}, \\ t &= \frac{2Y_0}{Y_0 m_{11} + Y_0 Y_S m_{12} + m_{21} + Y_S m_{22}}. \end{aligned} \tag{3.36}$$

Therefore, when we wish to compute r and t of a multilayer we need to know the characteristic matrix of each layer, multiply them in proper order as in Eq. (3.33), and substitute the elements of the resulting matrix in the equations above. The reflectance of the multilayer composed of the materials given in Table 3.2.1 is given in Fig. (3.3).

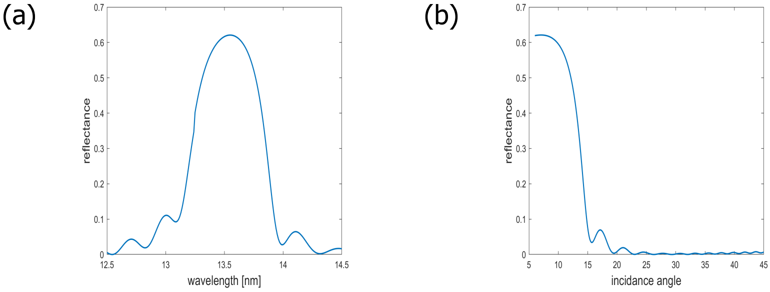

Figure 3.3: Reflectance of the multilayer given by the materials specified in 3.2.1 as a function of (a) wavelength and (b) incidence angle.

The incoming light field is a Gaussian beam, represented in Fig. 3.4, with a 3σ diameter of about $2\mu m$ and radiating 5e11 photons/sec. The detection NA is 0.6. The computational domain, of about $160 \times 160 \times 90 nm^3$ is truncated on all sides by perfectly matched layers and the far field data is sampled to guarantee an abundant oversampling ratio (≈ 10). The materials used for the simulations are reported

in Table 3.2.1. Computations are performed using the rigorous forward Maxwell solver JCMsuite [51]. The rigorous simulation tool used in this case has demanding memory requirements, therefore in order to simplify the computational experiment, we have replaced the detailed multilayer substrate with an equivalent substrate that offers – for the same λ and angle of incidence – about the same reflectance. This was done by fitting the curves in Fig 3.3.
Further we stress that, because of the reflective nature of EUV masks, the computational phase retrieval study is done in reflection. This means that in the results that follow we will retrieve the reflection function of the mask, in particular its amplitude is of interest as we consider the multilayer to be perfectly flat and do not introduce any phase defects in the simulations.
At the k–th iteration, an error metric is used to monitor the phase retrieval[19, 26]:

$$E_k = \left(\frac{\sum_{\mathbf{r} \notin S} |f_k'(\mathbf{r})|^2}{\sum_{\mathbf{r} \in S} |f_k'(\mathbf{r})|^2} \right)^{1/2}, \tag{3.37}$$

where, as described in 3.21, $f_k'(\mathbf{r})$ is the real space estimate obtained at the k–th iteration after enforcing the measured Fourier amplitudes in the Fourier transform of the object. S is the support defined as the locus of points where it holds at the same time that $f(\mathbf{r}) \neq 0$ and the imaginary part of the retrieved object function is positive (i.e. we enforce in the support constraint the knowledge that the materials is lossy). The error metric in Eq. (3.37) is often used in the works of Miao (for instance [19]). It reflects the idea that, while the object is getting reconstructed through multiple iterations, the number of pixels where the retrieved function violates the support constraint should decrease.

Table 3. Layers thicknesses and material properties at $\lambda = 13.5$ nm

layer	thickness [nm]	n	k
ARC TaBO	2	0.952	0.026
Absorber TaBN	58	0.95	0.031
Ru	0.5	0.88586	0.01727
Ru (Capping layer)	2	0.88586	0.01727
Si	1.8968	0.99888	0.00183
MoSi2	0.7986	0.96908	0.00435
Mo	2.496	0.92347	0.00649
MoSi2	1.8908	0.96908	0.00435

In what follows we considered structures of increasing complexity. Basically we simulated a single isolated nanostructure given by a patterned absorber where the patterning has been changed so to give structures of increasing geometrical complexity.
Figure 3.5 reports the results for an I–shaped structure, Figure 3.6 the ones for an L–shaped structure while in Figure 3.7 we considered a U–shaped structure. The pixel size in the real space has been computed by the theoretical limit $\delta x = \frac{\lambda}{2NA} \approx 11$nm. The images in Figs. 3.5 – 3.7 have been obtained by solving the inverse problem a few times, selecting the 4 reconstructions with the lowest error and aligning and

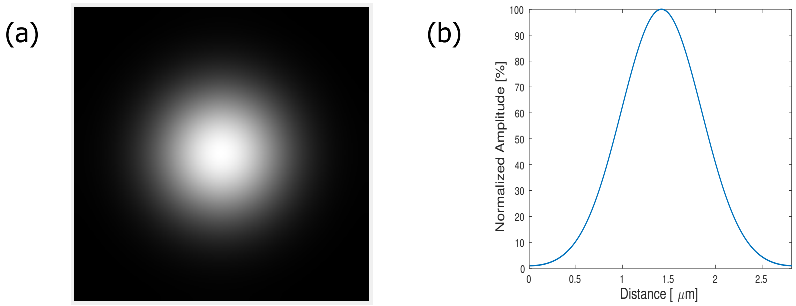

Figure 3.4: Probe used for the computational experiment. (a) full probe and (b) cross section.

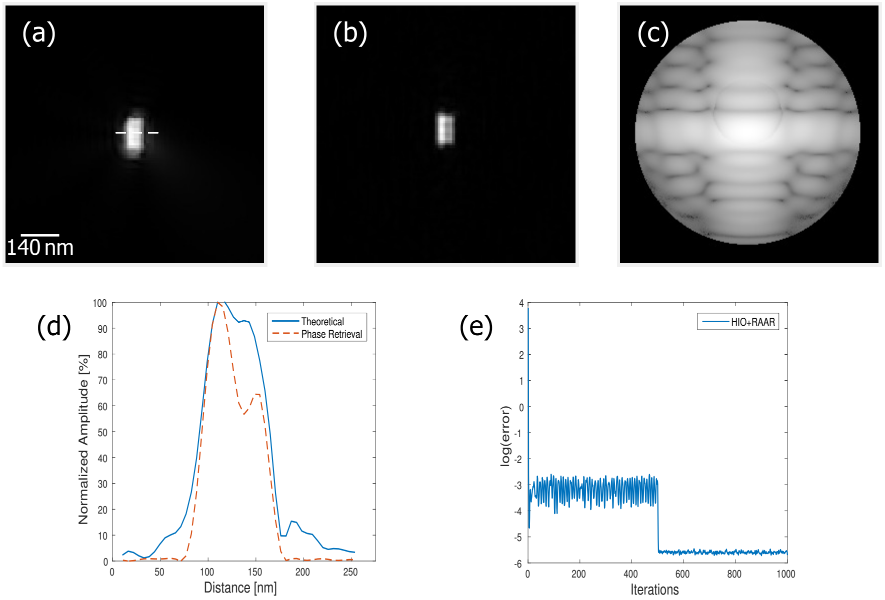

Figure 3.5: Reconstruction – magnitude – for an I–shaped nanostructure. (a) theoretical object (amplitude) obtained by direct inversion of the computed far–field; (b) reconstructed object (amplitude); (c) far field intensities; (d) cross–section, along the white line in (a), of the theoretical and the reconstructed amplitude of the object reflection function and (e) error according to Eq. (3.37).

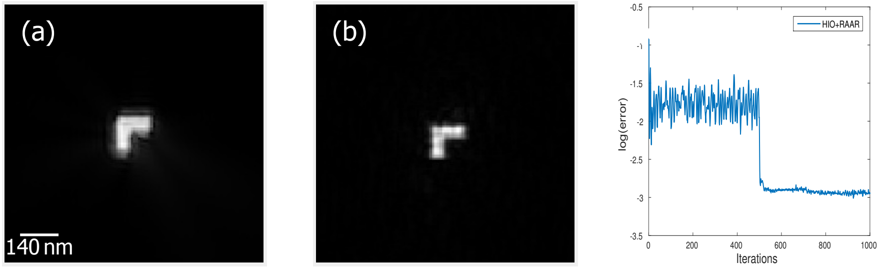

Figure 3.6: Reconstruction – magnitude – for an L–shaped nanostructure. (a) theoretical object (amplitude); (b) reconstructed object (amplitude) and (c) error according to Eq. (3.37).

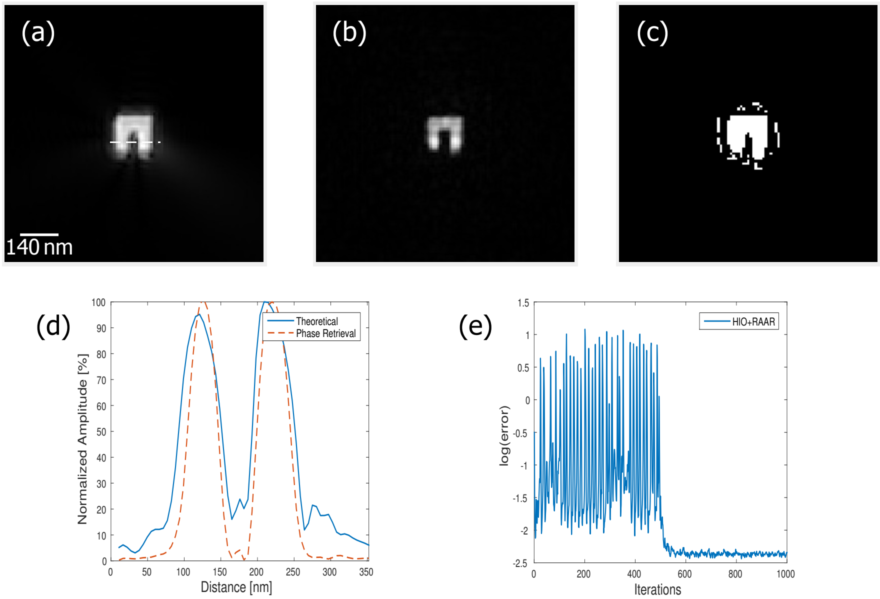

Figure 3.7: Reconstruction –magnitude – for an U–shaped nanostructure. (a) theoretical object (amplitude); (b) reconstructed object (amplitude); (c) support used for the reconstruction; (d) cross–section, along the line in (a), of the theoretical and the reconstructed amplitude of the reflection function of the object and (e) error according to Eq. (3.37).

averaging them. The feedback parameter β was heuristically set to 0.86. We have added noise for a SNR = 90 dB. The cross sections in Figs. 3.6 and 3.7 (d) reveal that the phase retrieval reconstruction mimics closely, albeit with some imperfections, the theoretical object. The retrieval of the I and L–shaped structures did not require any prior information about the precise shape of the object and we could retrieve the object reflection function by using a loose square support of about 20 pixel side. Notice that using a tighter support is generally favorable in that it results in a faster convergence. Reconstructing the U–shaped structure was more involved and we used prior information about the object, which is generally available in lithography, to generate a suitable support. In this case this was done by finding the edges of Fig 3.7(a) and filling them with ones. This support is depicted in Fig 3.7(c).

Ptychography

In what follows we describe the simulation results for actinic ($\lambda = 13.5$ nm) inspection of a specific EUV mask layout. To generate the data for the numerical ptychographic experiment we have used a rigorous volume–integral Maxwell Solver [52, 53]. The solver is formulated for the problem of scattering from periodic objects hence, in order to avoid crosstalk among adjacent cells, we have opted for a supercell approach. The cell size is $\Lambda = 3.5\mu m$. The sampling in the far field and in the illuminating NA equals $\frac{2\pi}{\Lambda}$. Although the lateral dimensions of the supercell are of the order of hundreds times the wavelength, the computational complexity and the memory requirement necessary to solve the forward problem are maintained relatively low, of the order of $\mathcal{O}(N \log N)$ where N is the number of unknowns (the coefficients of the Fourier series to be computed). The probe is assumed to be a Gaussian beam with a 3σ amplitude of about 1.5 μm and it is described by its angular spectrum. The scattered far–field is evaluated, for each of the plane waves which compose the illumination, in parallel on a multicore high–performance–computing (HPC) cluster. The output far–field that results from the interaction of the probe with the object is then given by the weighed coherent superposition of the individual contributions. The ptychographic scans are performed shifting the object in 5 positions over a distance of $0.2\mu m$ in positive and negative x-direction and y-direction, respectively. The probe is polarized in the x direction – parallel to the horizontal axis of the supercell – by proper linear combination of s and p polarization states. The collection NA is 0.6, close to the value (0.54) used for an identical wavelength in [54].

Although the sample is very complex, with a thickness of the structures which is about 5 to 6 times the wavelength, and although the data has been generated by rigorously solving Maxwell' s equations, the scalar ptychographic modeling approach delivers a high quality reconstruction (Fig. 3.8).

Some of the distances among structures are of the order of the wavelength and they appear quite resolved in the final image of Figs. 3.8 (a–c). In spite of the fact that the data–set was generated using a rigorous solver it was possible to obtain a high quality reconstruction using a the ptychographic approach, without recurring to any cumbersome *model–based* approach. In other words it was possible to interpret the output of the rigorous solver in terms of the ptychographic object reflection

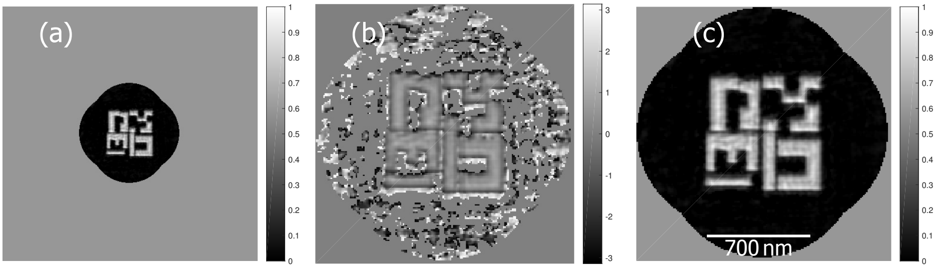

Figure 3.8: (a) Ptychographic reconstruction – magnitude – of an ensemble of nanostructures. (b) Phase of the reconstruction. (c) Zoom of (a).

function modeling. If this is true one may speculate that it should be possible to use the Maxwell solver as an additional building block in ptychography, by means of which one could include further information within this *model–free* CDI method. This speculation is the driving force of the study presented in the next chapter.

References

[1] P. Ansuinelli, W. Coene, and P. Urbach, *EUV mask feature reconstruction via phase retrieval,* in *Nanoengineering: Fabrication, Properties, Optics, Thin Films, and Devices XVI*, Vol. 11089, edited by B. Panchapakesan and A.-J. Attias, International Society for Optics and Photonics (SPIE, 2019) pp. 299 – 305.

[2] R. Jonckheere, *Overcoming EUV mask blank defects: what we can, and what we should,* in *Photomask Japan 2017: XXIV Symposium on Photomask and Next-Generation Lithography Mask Technology*, Vol. 10454, edited by K. Takehisa, International Society for Optics and Photonics (SPIE, 2017) pp. 127 – 139.

[3] A. A. Kagalwalla and P. Gupta, *Comprehensive defect avoidance framework for mitigating EUV mask defects,* in *Extreme Ultraviolet (EUV) Lithography V*, Vol. 9048, edited by O. R. W. II and E. M. Panning, International Society for Optics and Photonics (SPIE, 2014) pp. 212 – 222.

[4] K. A. Goldberg, A. Barty, P. Seidel, K. Edinger, R. Fettig, P. Kearney, H. H. M.D., and O. R. W. II, *EUV and non-EUV inspection of reticle defect repair sites,* in *Emerging Lithographic Technologies XI*, Vol. 6517, edited by M. J. Lercel, International Society for Optics and Photonics (SPIE, 2007) pp. 114 – 120.

[5] D. Hellweg, M. Dietzel, R. Capelli, C. Wolke, G. Kersteen, M. Koch, and R. Gehrke, *Actinic review of EUV masks: challenges and achievements in delivering the perfect mask for EUV production,* in *Photomask Technology 2017*, Vol. 10451, edited by P. D. Buck and E. E. Gallagher, International Society for Optics and Photonics (SPIE, 2017) pp. 115 – 125.

[6] M. Goldstein and P. Naulleau, *Actinic microscope for extreme ultraviolet lithography photomask inspection and review,* Opt. Express **20**, 15752 (2012).

[7] K. A. Goldberg, P. P. Naulleau, A. Barty, S. B. Rekawa, C. D. Kemp, R. F. Gunion, F. Salmassi, E. M. Gullikson, E. H. Anderson, and H.-S. H. M.D., *Performance of actinic EUVL mask imaging using a zoneplate microscope,* in *Photomask Technology 2007*, Vol. 6730, edited by R. J. Naber and H. Kawahira, International Society for Optics and Photonics (SPIE, 2007) pp. 1708 – 1719.

[8] K. A. Goldberg, I. Mochi, P. P. Naulleau, H. Han, and S. Huh, *Benchmarking EUV mask inspection beyond 0.25 NA,* in *Photomask Technology 2008*, Vol. 7122, edited by H. Kawahira and L. S. Zurbrick, International Society for Optics and Photonics (SPIE, 2008) pp. 774 – 781.

[9] M. P. Benk, K. A. Goldberg, A. Wojdyla, C. N. Anderson, F. Salmassi, P. P. Naulleau, and M. Kocsis, *Demonstration of 22-nm half pitch resolution on the sharp euv microscope,* Journal of Vacuum Science & Technology B **33**, 06FE01 (2015), https://doi.org/10.1116/1.4929509 .

[10] K. A. Goldberg, M. P. Benk, A. Wojdyla, I. Mochi, S. B. Rekawa, A. P. Allezy, M. R. Dickinson, C. W. Cork, W. Chao, D. J. Zehm, J. B. Macdougall, P. P. Naulleau, and A. Rudack, *Actinic mask imaging: recent results and future directions from the SHARP EUV microscope,* in *Extreme Ultraviolet (EUV) Lithography V*, Vol. 9048, edited by O. R. W. II and E. M. Panning, International Society for Optics and Photonics (SPIE, 2014) pp. 249 – 258.

[11] F. Zernike, *Das phasenkontrastverfahren bei der mikroskopischen beobachtung,* Z Tech. Phys. **16**, 454 (1935).

[12] J. W. Goodman, *Introduction to Fourier optics* (Roberts & Company, 2005).

[13] Y.-G. Wang, R. Miyakawa, A. Neureuther, and P. Naulleau, *Zernike phase contrast microscope for EUV mask inspection,* in *Extreme Ultraviolet (EUV) Lithography V*, Vol. 9048, edited by O. R. W. II and E. M. Panning, International Society for Optics and Photonics (SPIE, 2014) pp. 266 – 273.

[14] I. Mochi, P. Helfenstein, R. Rajeev, S. Fernandez, D. Kazazis, S. Yoshitake, and Y. Ekinci, *Actinic inspection of EUV reticles with arbitrary pattern design,* in *International Conference on Extreme Ultraviolet Lithography 2017*, Vol. 10450, edited by P. A. Gargini, P. P. Naulleau, K. G. Ronse, and T. Itani, International Society for Optics and Photonics (SPIE, 2017) pp. 28 – 37.

[15] R. Rajendran, I. Mochi, P. Helfenstein, I. Mohacsi, S. Redford, A. Mozzanica, B. Schmitt, S. Yoshitake, and Y. Ekinci, *Towards a stand-alone high-throughput EUV actinic photomask inspection tool: RESCAN,* in *Metrology, Inspection, and Process Control for Microlithography XXXI*, Vol. 10145, edited by M. I. Sanchez, International Society for Optics and Photonics (SPIE, 2017) pp. 199 – 210.

[16] P. Helfenstein, R. Rajeev, I. Mochi, A. Kleibert, C. A. F. Vaz, and Y. Ekinci, *Beam drift and partial probe coherence effects in euv reflective-mode coherent diffractive imaging,* Opt. Express **26**, 12242 (2018).

[17] S. Konijnenberg, *Computational methods for phase retrieval: Non-iterative methods, Ptychography, and Diffractive Shearing Interferometry*, Ph.D. thesis, Delft University of Technology (2019).

[18] D. Sayre, *Some implications of a theorem due to Shannon,* Acta Crystallographica **5**, 843 (1952).

[19] J. Miao, D. Sayre, and H. N. Chapman, *Phase retrieval from the magnitude of the fourier transforms of nonperiodic objects,* J. Opt. Soc. Am. A **15**, 1662 (1998).

[20] J. Spence, U. Weierstall, and M. Howells, *Coherence and sampling requirements for diffractive imaging,* Ultramicroscopy **101**, 149 (2004).

[21] J. W. Goodman, *Statistical optics* (1985).

[22] F. Fogel, I. Waldspurger, and A. d'Aspremont, *Phase retrieval for imaging problems,* Mathematical Programming Computation **8**, 311 (2016).

[23] S. Marchesini, *Invited article: A unified evaluation of iterative projection algorithms for phase retrieval,* Review of Scientific Instruments **78**, 011301 (2007), https://doi.org/10.1063/1.2403783 .

[24] K. Kreutz-Delgado, *The complex gradient operator and the cr-calculus,* (2009), arXiv:0906.4835 [math.OC] .

[25] D. H. Brandwood, *A complex gradient operator and its application in adaptive array theory,* IEE Proceedings F - Communications, Radar and Signal Processing **130**, 11 (1983).

[26] J. R. Fienup, *Phase retrieval algorithms: a comparison,* Appl. Opt. **21**, 2758 (1982).

[27] H. Chang, S. Marchesini, Y. Lou, and T. Zeng, *Variational phase retrieval with globally convergent preconditioned proximal algorithm,* SIAM Journal on Imaging Sciences **11**, 56 (2018), https://doi.org/10.1137/17M1120439 .

[28] E. J. Candès, X. Li, and M. Soltanolkotabi, *Phase retrieval via wirtinger flow: Theory and algorithms,* IEEE Transactions on Information Theory **61**, 1985 (2015).

[29] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion,* SIAM Journal on Imaging Sciences **6**, 199 (2013), https://doi.org/10.1137/110848074 .

[30] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E.-B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. Eldar, O. Cohen, and M. Segev, *Sparsity-based single-shot subwavelength coherent diffractive imaging,* Nature materials **11**, 455 (2012).

[31] Y. Bruck and L. Sodin, *On the ambiguity of the image reconstruction problem,* Optics Communications **30**, 304 (1979).

[32] M. Hayes, *The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform,* IEEE Transactions on Acoustics, Speech, and Signal Processing **30**, 140 (1982).

[33] B. RHT, *Fourier phase problems are uniquely solvable in more than one dimension. i: Underlying theory,* OPTIK (STUTTG.); ISSN 0030-4026; DEU; DA. 1982; VOL. 61; NO 3; PP. 247-262; ABS. GER; BIBL. 19 REF. (1982).

[34] P. Van Hove, M. Hayes, Jae Lim, and A. Oppenheim, *Signal reconstruction from signed fourier transform magnitude,* IEEE Transactions on Acoustics, Speech, and Signal Processing **31**, 1286 (1983).

[35] K. Jaganathan, S. Oymak, and B. Hassibi, *Sparse phase retrieval: Uniqueness guarantees and recovery algorithms,* IEEE Transactions on Signal Processing **65**, 2402 (2017).

[36] J. Rodenburg, *Ptychography and related diffractive imaging methods,* (Elsevier, 2008) pp. 87 – 184.

[37] J. M. Rodenburg and H. M. L. Faulkner, *A phase retrieval algorithm for shifting illumination,* Applied Physics Letters **85**, 4795 (2004), https://doi.org/10.1063/1.1823034 .

[38] A. M. Maiden and J. M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging,* Ultramicroscopy **109**, 1256 (2009).

[39] A. Maiden, D. Johnson, and P. Li, *Further improvements to the ptychographical iterative engine,* Optica **4**, 736 (2017).

[40] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, *Probe retrieval in ptychographic coherent diffractive imaging,* Ultramicroscopy **109**, 338 (2009).

[41] D. J. Batey, D. Claus, and J. M. Rodenburg, *Information multiplexing in ptychography,* Ultramicroscopy **138**, 13 (2014).

[42] P. Thibault and A. Menzel, *Reconstructing state mixtures from diffraction measurements,* Nature **494**, 68 (2013).

[43] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg, *Ptychographic transmission microscopy in three dimensions using a multi-slice approach,* J. Opt. Soc. Am. A **29**, 1606 (2012).

[44] L. Tian and L. Waller, *3d intensity and phase imaging from light field measurements in an led array microscope,* Optica **2**, 104 (2015).

[45] M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C. M. Kewish, R. Wepf, O. Bunk, and F. Pfeiffer, *Ptychographic x-ray computed tomography at the nanoscale,* Nature **467**, 436–439 (2010).

[46] H. Zhang, S. Jiang, J. Liao, J. Deng, J. Liu, Y. Zhang, and G. Zheng, *Near-field fourier ptychography: super-resolution phase retrieval via speckle illumination,* Opt. Express **27**, 7498 (2019).

[47] K. Wakonig, A. Diaz, A. Bonnin, M. Stampanoni, A. Bergamaschi, J. Ihli, M. Guizar-Sicairos, and A. Menzel, *X-ray fourier ptychography,* Science Advances **5** (2019), 10.1126/sciadv.aav0282, https://advances.sciencemag.org/content/5/2/eaav0282.full.pdf .

[48] S. Marchesini, H. He, H. Chapman, S. Hau-Riege, A. Noy, M. Howells, U. Weierstall, and J. Spence, *X-ray image reconstruction from a diffraction pattern alone,* (2003).

[49] D. R. Luke, *Relaxed averaged alternating reflections for diffraction imaging,* Inverse Problems **21**, 37 (2004).

[50] E. Hecht, *Optics* (Pearson, 2012).

[51] `https://jcmwave.com/`.

[52] M. C. van Beurden, *Fast convergence with spectral volume integral equation for crossed block-shaped gratings with improved material interface conditions,* J. Opt. Soc. Am. A **28**, 2269 (2011).

[53] M. C. van Beurden, *A spectral volume integral equation method for arbitrary bi-periodic gratings with explicit Fourier factorization,* Progress In Electromagnetics Research B **36**, 133 (2012).

[54] D. Gardner, M. Tanksalvala, E. Shanblatt, X. Zhang, B. Galloway, C. Porter, R. Karl, C. Bevis, D. Adams, M. Murnane, and G. Mancini, *Subwavelength coherent imaging of periodic samples using a 13.5 nm tabletop high-harmonic light source,* Nature Photonics **11** (2017), 10.1038/nphoton.2017.33.

4

Improving the ptychographic inspection of EUV reticles by including prior information

We have articulated in Section 1.1.1 that *a–priori* information is important when solving an inverse problem, because it can mitigate the ill–conditioning of the problem and it restricts the search space towards appropriate solutions. We have noticed in the previous chapter that the phase problem is a non–convex optimization problem and therefore multiple local minima or saddle points are present in the optimization landscape. A further remark was that the sole physical assumption made in ptychography is that light matter interaction can be modeled as the product of a probe function and an object function, the latter either being an object transmission function or an object reflection function. Therefore one might consider ptychography as a sort of *model free* method where the details of the object get retrieved solely by processing the data and by assuming that the exit wave can be partitioned as the product of the probe and the object, without the need of interfacing with a sophisticated physics based simulation tool as we did in Chapter 2.

Inspired by these observations, this chapter presents and discloses inversion methods that enrich the usual ptychographic approach by incorporating *a–priori* information which is generated by means of fully rigorous Maxwell solvers. The two different, and seemingly incompatible, mathematical and physical models – the 3D, model-based, fully rigorous and vectorial one and the 2D model free and scalar ptychographic model – are interfaced and combined. Their joint use is found to be beneficial for the imaging of an intricate EUV absorber layout and to lead to the detection of smaller absorber defects with a signal–to–noise (SNR) ratio which

---

Parts of this chapter have been published in *Applied Optics* **59**, 5937–5947 (2020) [1] and in *Extreme Ultraviolet lithography 2020, International Society for Optics and Photonics (SPIE, 2019)* **11517**, 13–21 [2] is higher than the one computed with the standard ptychographic iterative engine (PIE).

4.1. A variational method for the inclusion of prior information in ptychography

As we have seen in Section 3.1.3, Eq. (3.20), ptychography can be seen as a cost functional minimization problem in which, considering the j–th probe position $\mathbf{R}_j$ with probe $P(\mathbf{r}-\mathbf{R}_j)$, one seeks a certain object $O_j(\mathbf{r})$ that best fits the j–th recorded diffraction intensity $I_j(\mathbf{k})$:

$$\text{minimize} \quad \mathcal{L}(O_j, I_j) := \left|\left|\mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_j(\mathbf{r})) - \sqrt{I_j(\mathbf{k})}\right|\right|^2 . \tag{4.1}$$

In Eq. (4.1) and in all that follows we assume the probe to be known. The PIE proceeds in a sequential fashion and the complete object, $O(\mathbf{r})$, is reconstructed by the sequential reconstruction of its multiple views [3]. Notice that in Eq. (4.1) we did not include a summation over the probe positions because the algorithms we have used in this thesis optimize the object one probe position at a time. As previously stated the phase problem is a non–convex optimization problem – meaning that there can be multiple local minima – in which the reconstruction is obtained solely by processing the data set, therefore one might wonder whether the inclusion of physically sound prior knowledge in the algorithm can lead to a more satisfactory solution. Prior information is commonly included introducing a regularization term to the data discrepancy functional. In this case:

$$O_j^\alpha(\mathbf{r}) := \arg\min\{\mathcal{L}(O_j, I_j) + R^\alpha(O_j)\} := \arg\min\{\mathcal{L}^\alpha(O_j, I_j)\}, \tag{4.2}$$

where $R^\alpha(O_j)$ is the regularizer, which depends upon the regularization parameter α, $\mathcal{L}^\alpha(O_j, I_j)$ is the regularized cost functional at the j–th probe position and $O_j^\alpha(\mathbf{r})$ is the associated object update. $R^\alpha(O_j)$ is intended as a penalty term in variational approaches or as the negative logarithm of a prior probability distribution from the Bayesian viewpoint [4, 5]. Different regularizers (priors) have been studied in the field of image reconstruction, for instance structural priors [6], sparsity promoting priors [7] or edge–preserving priors. An important example of the latter is total variation (TV), in which $R(O) = ||\nabla O(\mathbf{r})||_1$. TV is commonly employed due to its effectiveness in smoothing noise – by favouring images that have a sparse gradient – while preserving edges. In recent years there have been a number of studies reporting implementations of TV regularized ptychography [8–10]. In this work we have taken a different approach and instead of considering a regularizer that promotes generic properties of the reconstructed image (like its sparsity or the sparsity of its gradient), the regularizer $R^\alpha(O_j)$ has been designed to promote adhesion to a given prior image (object):

$$R^\alpha(O_j) = \alpha \left|\left|O_j(\mathbf{r}) - O_{p,j}(\mathbf{r})\right|\right|^2 , \tag{4.3}$$

where $O_{p,j}(\mathbf{r})$ is the prior of the object at the j–th probe position. Crafting a "good" prior is challenging. This task requires one to properly account for the physics that contributes to the process of image formation. For instance it is known that at EUV wavelengths waveguiding effects and diffraction of light along the thickness of the absorber materials have an impact on image formation [11]. Further, the phase of the scattered wave $\Phi = \mathbf{k} \cdot \mathbf{r}$ accumulates linearly as photons pass through an object and the rate of accumulation depends on the refractive index which changes from material to material [12]. This is important to realize when one wants to compute an *a–priori* map for the phase of a wave scattered by nanostructures that consist of a sequence of layers of different materials and thicknesses. Such a situation applies to the multilayer Bragg–reflector and to the absorber layers of an EUV mask. These aspects should be properly encoded in the prior, in order for it to be reliable and accurate.

These physical aspects and the intrinsically 3D thickness effects can be duly accounted for by recurring to three dimensional fully rigorous simulations. Forward Maxwell solvers can compute the complex field, which is a rigorous solution in terms of amplitude and phase of Maxwell's equations, for a given 3D scattering geometry [13, 14]. It is important to notice that light–matter interaction is modeled differently in the rigorous Maxwell solvers than in ptychography. On the one hand the rigorous electromagnetic solvers provide an accurate solution of Maxwell's equation, on the other hand ptychography models light–matter interaction in terms of the 2D "probe–times–object" approximation. This fundamental difference in the physical models could be a reason of concern when intermixing the use of the rigorous solvers with ptychographic algorithms. In other words, a certain rigorously computed complex–valued far–field Ψ^{Maxw} can be used in ptychographic algorithms only when it can be interpreted in terms of the 2D ptychographic approximation of light matter interaction:

$$\Psi^{Maxw}_{p,j} \approx \mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_{p,j}(\mathbf{r})), \tag{4.4}$$

where $\Psi^{Maxw}_{p,j}$ is the far–field, amplitude and phase, as computed by the forward Maxwell solver, for the nominal – *a priori* known – scattering geometry on the mask. Notice that although ptychography assumes the object function to be two dimensional, $\Psi^{Maxw}_{p,j}$ is computed by rigorous 3D simulations. In other words, the 2D object reflection function comprises genuine 3D information about the object. Notice further that Eq. (4.4) effectively links and combines two fundamentally different models when solving the inverse problem. Bearing Eq. (4.4) in mind we proceed to write, at the j–th probe position, the ptychographic cost functional inclusive of the prior term by using Eqs. (4.1–4.3):

$$\begin{aligned} \mathcal{L}^{\alpha}(O_j, I_j) &:= \mathcal{L}(O_j, I_j) + R^{\alpha}(O_j) = \\ &= \left|\left| |\mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_j(\mathbf{r}))| - \sqrt{I_j(\mathbf{k})} \right|\right|^2 + \alpha \left|\left| O_j(\mathbf{r}) - O_{p,j}(\mathbf{r}) \right|\right|^2, \end{aligned} \tag{4.5}$$

where α regulates the interplay among the two terms on the right hand side of Eq. (4.5). The second term in (4.5) penalizes large deviations of the reconstructed object from the prior of the object. Eq. (4.5) can be minimized analytically using Wirtinger calculus to yield the update rule that can be used in a gradient descent algorithm. The gradient of Eq. (4.5) with respect to O^* is:

$$\nabla\mathcal{L}^{\alpha}(O_j, I_j) = P_j^*\left[\mathcal{F}^{-1}(\Psi_j - \Psi_{c,j})\right] + \alpha(O_j - O_{p,j}), \tag{4.6}$$

and, by defining

$$\begin{aligned} \nabla\mathcal{L}_1 &\coloneqq P_j^*\left[\mathcal{F}^{-1}(\Psi_j - \Psi_{c,j})\right], \\ \nabla\mathcal{L}_2^{\alpha} &\coloneqq \alpha(O_j - O_{p,j}), \end{aligned} \tag{4.7}$$

we can rewrite Eq. (4.6) as

$$\nabla\mathcal{L}^{\alpha}(O_j, I_j) = \nabla\mathcal{L}_1 + \nabla\mathcal{L}_2^{\alpha}. \tag{4.8}$$

In Eq. (4.6) $\Psi_j = \mathcal{F}(P_j O_j)$ is the guessed wavefront in the momentum space, $P_j = P(\mathbf{r}-\mathbf{R}_j)$ and $\Psi_{c,j}$ is the revised wavefront obtained by enforcing the measured amplitudes on the far–field guess while keeping the phase unchanged. We focus now on $\nabla\mathcal{L}_2^{\alpha}$ and, with the use of Eq. (4.4), we obtain:

$$\begin{aligned} \nabla\mathcal{L}_2^{\alpha} &= \alpha(O_j - O_{p,j}) = \alpha\frac{P_j^*}{(|P_j|^2)}(P_j O_j - P_j O_{p,j}) = \\ &= \alpha\frac{P_j^*}{|P_j|^2}\mathcal{F}^{-1}(\mathcal{F}(P_j O_j) - \mathcal{F}(P_j O_{p,j})) = \alpha\frac{P_j^*}{|P_j|^2}\mathcal{F}^{-1}(\Psi_j - \Psi_{j,p}^{Mawx}). \end{aligned} \tag{4.9}$$

The steepest descent update, at iteration n is:

$$O_{j,n+1} = O_{j,n} - \beta\nabla\mathcal{L}^{a}, \tag{4.10}$$

for some value of β. With Eqs. (4.6), (4.9) it follows that the part of the object that is illuminated by the probe at the j–th probe position, $O(\mathbf{r})$, is updated as follows:

$$\begin{aligned} O_{n+1} = O_n &+ \beta\frac{|P_j|}{|P_{j,max}|}\frac{P_j^*}{(|P_j|^2 + c)}\mathcal{F}^{-1}(\Psi_{c,j,n} - \Psi_{j,n}) \\ &+ \alpha'\frac{|P_j|}{|P_{j,max}|}\frac{P_j^*}{(|P_j|^2 + c)}\mathcal{F}^{-1}(\Psi_{p,j}^{Maxw} - \Psi_{j,n}), \end{aligned} \tag{4.11}$$

where $\alpha' = \beta\alpha$. In Eq. (4.11) c is a small constant that prevents division by zero and $|P_j|/|P_{j,max}|$ is a scaling factor – present in the PIE – that makes the update less reliable where the probe is dim. This scaling factor increases the influence of those areas of the sample that have been strongly illuminated while attenuating the errors which could arise from those areas that have been weakly illuminated. Although the prior $O_{p,j}$ does not appear explicitly in (4.11), it is contained implicitly in its second term as $O_{p,j} = P_j^*\frac{\mathcal{F}^{-1}(\Psi_{p,j}^{Maxw})}{|P_j|^2}$. Notice that, besides α' and β, the scaling factors in (4.11) are the same for the two terms. This is in order to avoid overenforcing the prior in those pixels where the probe is dim. For $\alpha = 0$, Eq. (4.11) yields

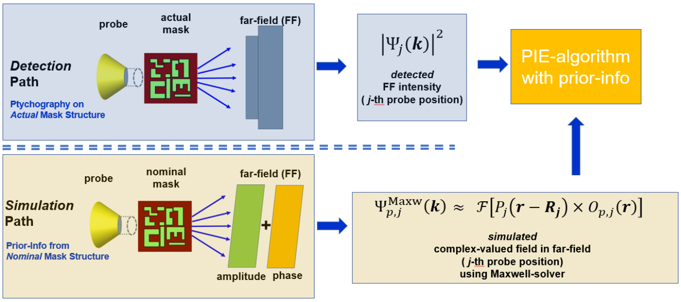

Figure 4.1: A schematic that illustrates the approach discussed in this work. In the detection path – highlighted in blue – the actual defective mask is illuminated by the probe and the diffracted light is measured in the far field. In the simulation path – highlighted in yellow – the probe illuminates the nominal defect–free mask and the far–field is computed, in amplitude and phase, via the forward Maxwell solver. The computationally generated far–field is interpreted in terms of the 2D ptychographic approximation of light matter interaction (Eq. (4.5)) and contains information about the prior object $O_p(\mathbf{r})$. The simulated and detected far–field patterns are subsequently employed in the reconstruction algorithm and in the update rule (4.11).

the same update rule of the regular PIE. As argued before, the algorithm proceeds in a sequential fashion, and the complete object, $O(\mathbf{r})$, is reconstructed by the sequential reconstruction of its multiple j–th views, according to Eq. (4.11).

A layout of this approach is given in Fig. 4.1.

The inclusion of prior information as a regularizer in Eq. (4.5) is preferable the alternative where one uses $O_p(\mathbf{r})$ as a starting guess in the optimization. There are various reasons for this:

- instead of being used exclusively as a starting guess in the optimization, the prior in Eq. (4.11) is present at every iteration of the optimization, therefore preventing the reconstruction from diverging towards an “unphysical” solution. Further,

- the presence of the quadratic term in Eq. (4.5) improves the conditioning of the problem. This stabilizes the inversion with respect to the noise and improves the performance of the iterative method. Further, the regularizer aids in creating a better model by achieving a balance in the bias–variance trade–off for a proper selection of α [15]. Values of α which are too small make the reconstruction too sensitive to the noise, however setting α to a value that is too high biases the reconstruction towards the prior, yielding a poor fit. A proper value of α achieves a balance among these two cases.

In what follows, we will show that the inclusion of the regularization term, Eq. (4.3), in the cost functional (4.5) via the use of the accurate physical–analytical models yields a better reconstruction with respect to the standard case in which the reflection function $O(\mathbf{r})$ is retrieved solely by processing the intensity data–set as with the standard PIE approach.

4.1.1. Results

We have applied the method outlined above to reconstruct the patterned absorber of a 3D EUV mask. Table 4.1 reports the materials and the thicknesses of the layers used in this work. The mask comprises 40 layers of Si-MoSi2-Mo-MoSi2.

Table 4.1: **Layers thicknesses and Materials at $\lambda = 13.5$ nm**

layer	thickness [nm]	n	k
ARC TaBO	2	0.952	0.026
Absorber TaBN	58	0.95	0.031
Ru	0.5	0.88586	0.01727
Ru (Capping layer)	2	0.88586	0.01727
Si	1.8968	0.99888	0.00183
MoSi2	0.7986	0.96908	0.00435
Mo	2.496	0.92347	0.00649
MoSi2	1.8908	0.96908	0.00435

Four EUV masks have been considered in this study:

- the “nominal” mask. This is the cell as given by prior information. This cell does not contain any information about the defects. This is the cell used in the simulation path depicted in yellow in Fig. 4.1.
- The defect–free “actual” mask. This is the cell that mimics the “reference” mask which is close to the prior but not exactly the same. In order to account for this difference the actual cell has been generated from the prior cell, displacing the sides of the polygons over a distance of 1–5 nm. This cell is displayed in Fig. 4.2(a).
- Two “actual defective” or programmed defects masks. Consistently with the practice in EUV mask defectivity studies we have perturbed the actual mask, at known locations, with additive and subtractive features (extrusions and intrusion defects) (Figs. 4.2(b–c)). This is the cell used in the detection path depicted in blue in Fig. 4.1.

In Fig. (4.2) we do not show the nominal mask because, as mentioned previously, the actual mask has been generated from the prior mask (4.2(a)), displacing the sides of the polygons over a distance of 1–5 nm. This means that the prior and actual mask would look indistinguishable from above as displayed in Fig. (4.2).
The size of the defects in Figs. 4.2(b,c) is the same on a given polygon, and it changes from polygon to polygon. The number and the side length of the squares that constitute the rough extrusions/intrusions on a certain polygon are the following: [*number of squares, side length*] = [3, 16 nm], [6, 12 nm], [7, 9 nm], [7, 6 nm]. Such sizes have been chosen in accordance to the theoretical Abbe resolution limit imposed by the NA. Because the collection NA has been chosen to be equal to 0.6 in this case, the theoretical achievable resolution limit is $\lambda/2NA = 11$ nm. If the collection NA was to be smaller the defects would have been made bigger accordingly.

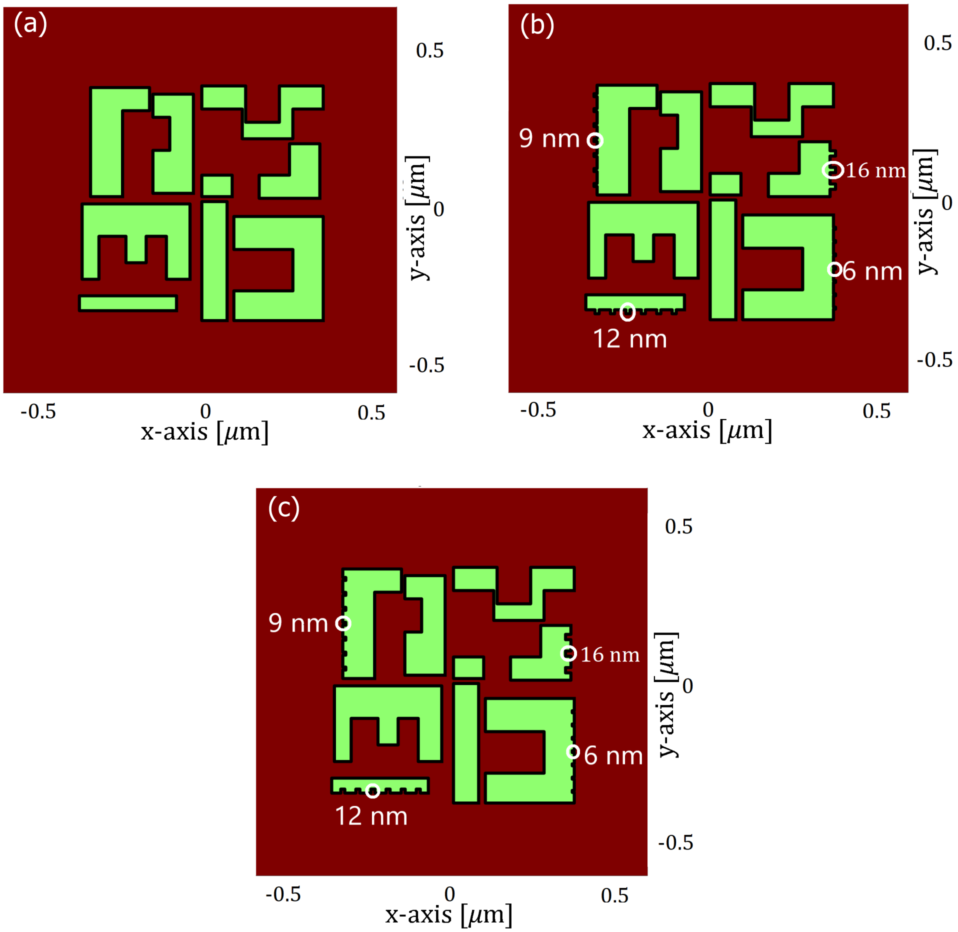

Figure 4.2: Top views of EUV mask layouts. a) Actual cell, b) programmed defect mask (extrusions), c) programmed defect mask (intrusions). The sidelength of the single defect is specified in the figure.

To understand whether the inclusion of prior information yields any benefit for our specific application we have carried out a computational die–to–database comparison [16]. This is a procedure in which one compares the images collected from the defective reticle with the images that are generated from the reticle design data. In this case this is done comparing the reconstruction of the defective mask with the reconstruction of the defect–free actual mask. The two reconstructions are subtracted one from the other to identify the defects at their locations. The impact of the defects is quantified by a certain figure of merit. In what follows we will use the defect SNR – called SNR_δ – defined as [16]:

$$SNR_\delta = \frac{\bar{A}_d - \bar{A}_a}{std(A_a)}, \tag{4.12}$$

where $\bar{A}_d$ is the average magnitude of the defected area, $\bar{A}_a$ is the average magnitude of the whole difference image – where the object is present – and $std(A_a)$ is the standard deviation of the latter area. The definition of the defect SNR does not depend directly on the defect size (as long as the total area occupied by the image of the defects is the same) and it express the intensity of the defect signal with respect to the dark noisy background. This investigation is done using the standard PIE and the PIE with prior where the prior is given by the nominal mask, where we use Eq. (4.11) as update rule to reconstruct both the actual mask (Fig. 4.2(a)) and the defected masks (Fig. 4.2(b–c)).
As outlined above four data sets have been computed:

- one complex data set, in amplitude and phase, for the nominal mask. This is the far field $\Psi_{p,j}^{Maxw}$, related to the nominal mask, we referred to in Eqs. (4.3) and (4.11).
- One intensity–only data set for the actual defect–free mask in Fig. 4.2(a). This data–set has been corrupted with noise to emulate measured data.
- Two intensity–only data sets for the defected masks, one for the extrusions in Fig. 4.2(b) and one for the intrusions in Fig. 4.2(c). This data sets have also been corrupted with noise to emulate measured data.

The simulations set–up was discussed before, in Section 3.2.1, and is reported here for the convenience of the reader.
All the four data sets mentioned above have been generated via a fully rigorous 3D simulations using a volume–integral Maxwell solver [17, 18] and taking account of the actual multi-layer, i.e. without recurring to an effective reflection function for the substrate. The solver is formulated for the problem of scattering from periodic objects hence, in order to avoid cross–talk among adjacent cells, we have opted for a supercell approach. The cell is a square with lateral dimension $\Lambda = 3.5\mu m$. The sampling in the far field and in the illuminating NA equals $\frac{2\pi}{\Lambda}$. Although the lateral dimensions of the supercell are of the order of hundreds times the wavelength, the computational complexity and the memory requirement necessary to solve the forward problem are maintained relatively low, of the order of $\mathcal{O}(N \log N)$

where N is the number of unknowns which are the coefficient of the Fourier modes in the Fourier series expansion of the electric field. The probe is assumed to be a Gaussian beam with a 3σ amplitude of about 1.5 μm and it is described by its angular spectrum. The scattered far–field is evaluated, for each of the plane waves which compose the illumination, in parallel on a multicore HPC cluster. The output far–field that results from the interaction of the probe with the object is then given by the weighed coherent superposition of the individual contributions. The ptychographic scans are performed shifting the object of 0.2 μm, in 5 positions, inside the supercell. This grants a probe overlap of about 85%. The probe is polarized in the x direction – parallel to the horizontal axis of the supercell – by proper linear combination of s and p polarization states. The collection NA is 0.6, close to the value (0.54) used for an identical wavelength in [19]. Fig. 4.3(a,b) illustrates the probe, its cross–section and one of the acquired diffraction patterns for the mask in Fig. 4.3(a).

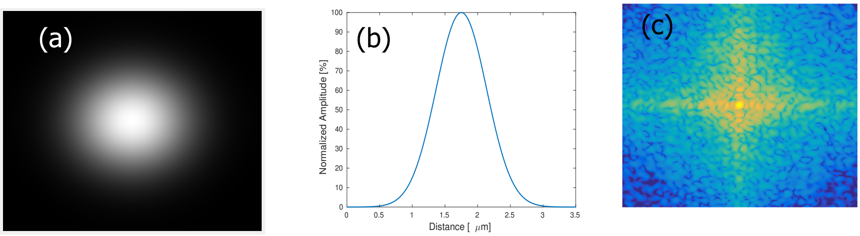

Figure 4.3: a) Gaussian probe used in the computations. b) Cross–section of (a). c) Diffraction pattern (log scale).

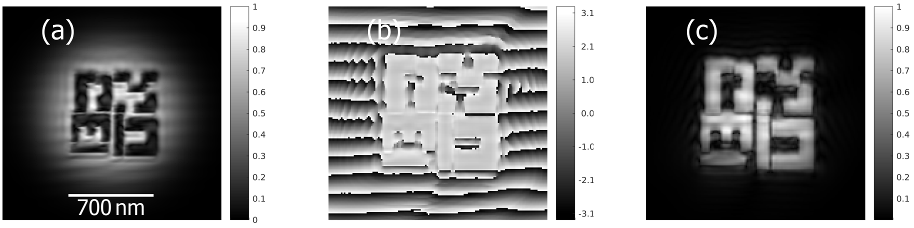

Figure 4.4: Zoom of the nominal reflection function associated to the a–priori known mask geometry as retrieved from the fully rigorous simulations for an incoming beam incident at an angle of 6° on the structures and polarized along the x axis, a) amplitude of nominal reflection function, b) phase and c) amplitude after and removal of the strong unscattered beam. The fringes in (b) are caused by the projection of the term $e^{i\mathbf{k}\cdot\mathbf{r}}$ over the x–y plane related to the 6° angle of incidence of the probe.

Fig. 4.4(a,b) shows the prior for the central probe position. White Gaussian noise has been added to the synthetically generated data for an SNR = 110 dB and the regularization parameter α has been fixed to 2e-2. This value was chosen by trial and error, attempting the use of few values of α and selecting the one for which the defects were better visible in the resulting imaged.

Extrusion defects

The amplitude of the object function reconstructed with ptychography corresponding to the patterned absorber depicted in Fig. 4.2(b), without and with prior respectively, is depicted in Fig. 4.5(a,b), while the phase is shown in Fig. 4.5(d,e). The object error, at iteration n, has been computed as the relative deviation of the reconstructed object, $O_n(\mathbf{r})$, from the theoretical object $O_t(\mathbf{r})$:

$$E_{O,n} = \frac{\sum_{\mathbf{r}} |O_t(\mathbf{r}) - \gamma O_n(\mathbf{r})|^2}{\sum_{\mathbf{r}} |O_t(\mathbf{r})|^2}, \tag{4.13}$$

where γ is a parameter that compensates for phase ambiguities [20].

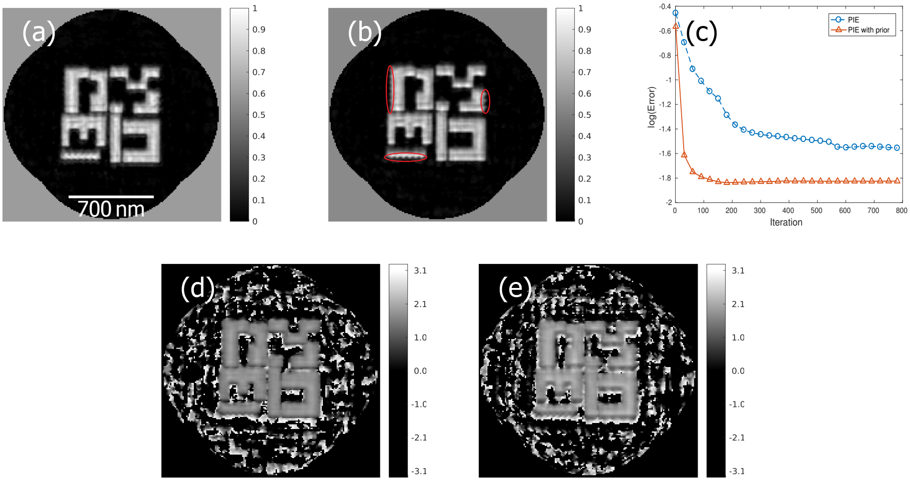

Figure 4.5: Ptychographic reconstructions: a) amplitude of the reflection function of the object as given by the PIE; b) amplitude of the reflection function of the object given by the new algorithm PIE with prior term included. The rough extrusions are highlighted in the red circles;c) error; d) phase given by the PIE and e) phase given by the PIE with prior.

The result of the reconstructions is highlighted in Figs. 4.5(a–d). As will be further discussed and shown below, the rough extrusions, highlighted in Fig 4.5(b), are better resolved in Figs. 4.5(b,e) – which show the amplitude and phase of the object reflection function obtained with the inclusion of the prior term – than in Figs. 4.5(a,d) which are the amplitude and phase of the reconstruction obtained with the standard PIE. The error, defined in Eq. (4.13) and shown in Fig 4.5(c) shows an overall better reconstruction and convergence when the prior is included in the optimization algorithm. Practically, we have observed the algorithm to converge in about a third of the iterations of the regular PIE where the starting point of the reconstruction was a transparent object with random phase. As stated before a quantitative assessment of the improvement in terms of defect inspection can be obtained in a die–to–database comparison by computing difference images. Those are defined as the magnitude of the difference among the retrieved ptychographic reconstruction of the objects in Figs. 4.2(a,b):

$$\Delta O(\mathbf{r}) = |O_a(\mathbf{r}) - O_\delta(\mathbf{r})|, \tag{4.14}$$

with $O_a(\mathbf{r})$ and $O_\delta(\mathbf{r})$ being the reconstructed reflection functions related to the actual mask in Fig. 4.2(a) and to the defected actual mask in Fig. 4.2(b–c). In what follows we will refer to Eq. (4.14) as the object difference metric. Fig. 4.6 shows $\Delta O(\mathbf{r})$ defined in (4.14), obtained when reconstructing $O_a(\mathbf{r})$ and $O_\delta(\mathbf{r})$ using prior information – by the ptychographic update rule Eq. 4.11 – and the standard PIE. Here, and in all that follows, the objects have been aligned before their subtraction.

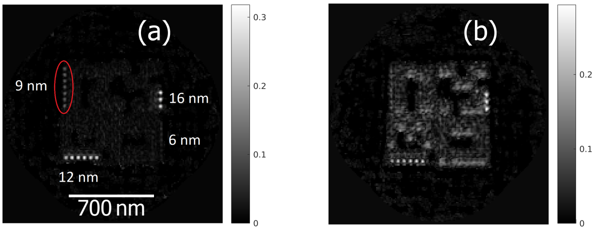

Figure 4.6: a) and b) show the object difference metric ΔO in Eq. (4.14) obtained for extrusions. a) makes use of prior while in b) we employed the standard PIE. The indicated sizes in a) are the side lengths of each of the blocks of the extrusions.

All the defects are better resolved in Fig. 4.6(a) than in Fig. 4.6(b), and their signature appear to be stronger in the difference image 4.6(a). Particularly, the rough defects of 9 nm size are not distinguishable in Fig. 4.6(b), however they are detectable in Fig. 4.6(a), as highlighted in the red circle. We found the finest details, the rough extrusions of 6 nm size, to be absent in the reconstruction in Figs. 4.5(a–b), and in the difference images in Figs. 4.6(a–b). The side of the polygon over which these defects are located appears to be smooth in the reconstructed image. This computational experiment highlights that subwavelength ptychographic imaging is possible to a certain extent, however the spatial periodicities below the theoretical limit of about $\lambda/2$ cannot be reconstructed. Since the collection NA is equal to 0.6, the Abbe limit is 11 nm which is slightly above the size, 9nm, of the smallest defect we managed to resolve. A comparison of the retrieved defect SNR (i.e. SNR$_\delta$) for Figs. 4.6(a,b) is reported in Table 4.2:
Table 4.2 highlights a steep improvement in the detectability of defects when incorporating the prior term in the reconstruction algorithm. The value N/A means that the defects are not visible in the difference image. A cross section of the reconstructions in Fig. 4.5(a–b), taken on the location of the defects, is shown in Figure 4.7.
The cross sections in Fig. 4.7 generally have a more pronounced peak to valley ratio when the prior is included and, in case of the highest resolvable information

Table 4.2: **Extrusion defects SNR**

Defect Size [nm]	Number of defects	SNR_δ PIE with prior	SNR_δ PIE
16	3	3.8	1.8
12	6	5.3	1.2
9	7	1.6	N/A

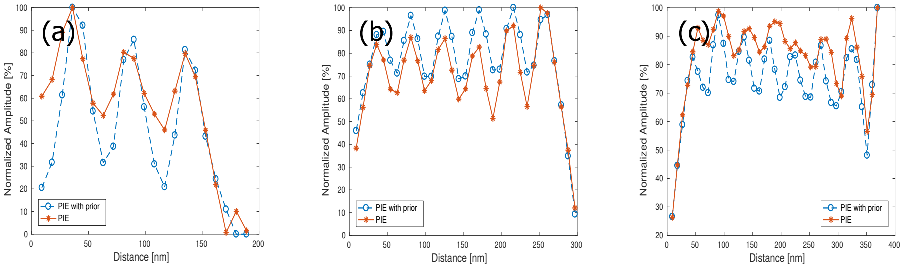

Figure 4.7: Cross section of the reconstruction at the defect sites.a) 3 defects of 16 nm sidelength; b) 6 defects of 12 nm sidelength; c) 7 defects of 9 nm sidelength.

of 9 nm (Fig. 4.7(c)) below Abbe's resolution limit of 11nm, the periodicity of the signal is more evident, while it is lost – in the central part of the plot – in the case of the standard PIE.

Intrusion defects

The amplitude of the reconstructed object reflection function relative to the patterned absorber layout in Fig. 4.2(c), obtained by means of the ptychographic algorithm without and with prior respectively, is depicted in Fig. 4.8(a,b), while the phase is shown in Figs. 4.8(d,e).

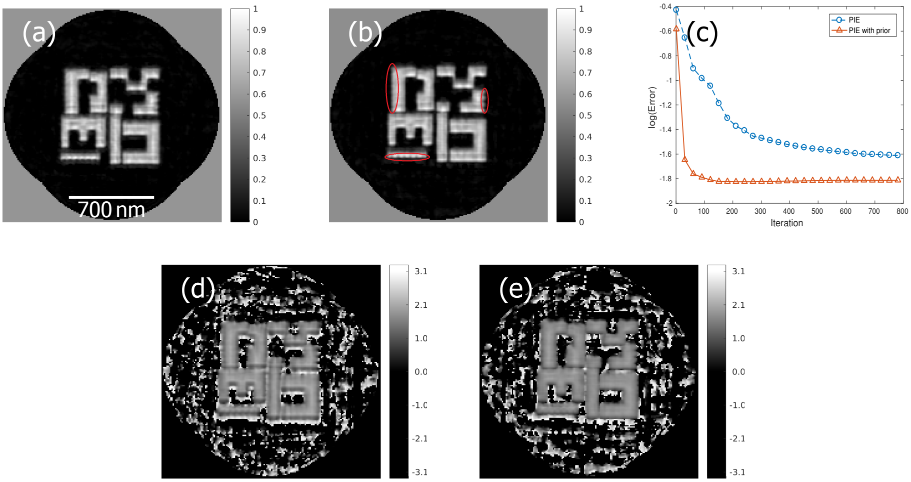

Figure 4.8: Ptychographic reconstructions: a) amplitude of the reflection function of the object as given by the PIE; b) amplitude given by the PIE with prior included. The rough intrusions are highlighted in the red circles; d) phase given by the PIE and e) phase given by the PIE with prior; c) error.

Fig. 4.9 shows the object difference metric $\Delta O(\mathbf{r})$ defined in Eq. (4.14), obtained using the update rule Eq. (4.11) and the standard PIE.

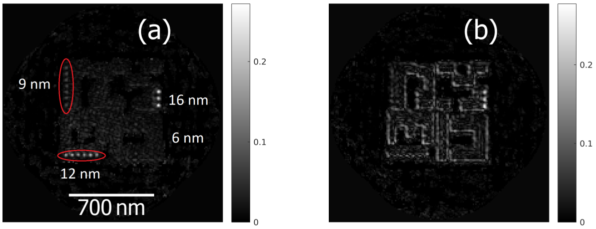

Figure 4.9: a) and b) show ΔO in Eq. (12) obtained for intrusions. a) makes use of prior while in b) we employed the standard PIE. The sizes in a) are the side lengths of each of the blocks of the intrusions.

The comparison of the defect SNR, called SNR$_\delta$, for Figs. 4.9(a,b) is reported in Table 4.3:

Table 4.3: **Intrusion defects SNR**

Defect Size [nm]	Number of defects	SNR$_\delta$ with prior	SNR$_\delta$ PIE
16	3	3.8	1.5
12	6	3.4	N/A
9	7	1	N/A

All the intrusion defects have decreased SNR$_\delta$ with respect to extrusion defects of the same size, and even with the inclusion of the prior term the 9 nm size intrusions are difficult to image. It should be noticed that the imaging of small defects is also hampered by the presence of guided modes that propagate in the gaps of the absorber, making these defects harder to image with high contrast [11]. Fig. 4.10 illustrates the cross section of the reconstructions in Figs. 4.8(a–b), taken at the location of the defects:

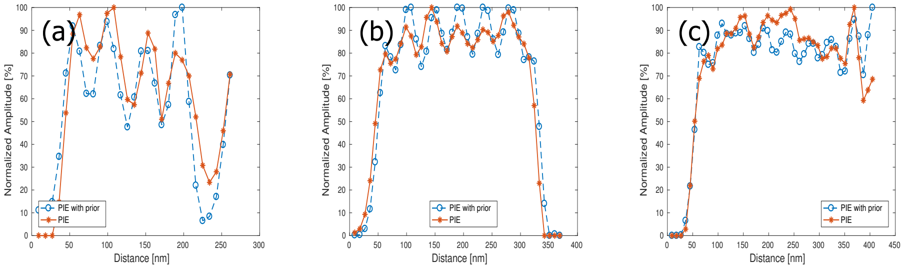

Figure 4.10: Cross section of the reconstruction at the defect sites.a) 3 defects of 16 nm sidelength; b) 6 defects of 12 nm sidelength; c) 7 defects of 9 nm sidelength.

4.1.2. Effect of the number of probe positions, SNR, initial guess and a comparison with TV regularization

Increasing the number of probe positions

Ptychography achieves a very robust reconstruction by exploiting translational diversity and redundancy in the data set. In the study presented earlier the data set included five probe positions. As the object is, in this case, entirely covered by the probe there is a high degree of redundancy in the data in spite of the few probe positions. However it can be interesting to see whether increasing the number of probe positions allows one to get to a reconstruction which is as good as in the case in which the prior term is included. We have performed this study for the case of the extrusions type defects. We have used 9 probe positions and the reconstruction has been carried out using the PIE. The 9 positions constitute a 3–by–3 grid which span, in steps of $200nm$ along the x and y directions, a square whose sidelength is $400nm$. The reconstruction and the difference image is shown in Fig. 4.11.

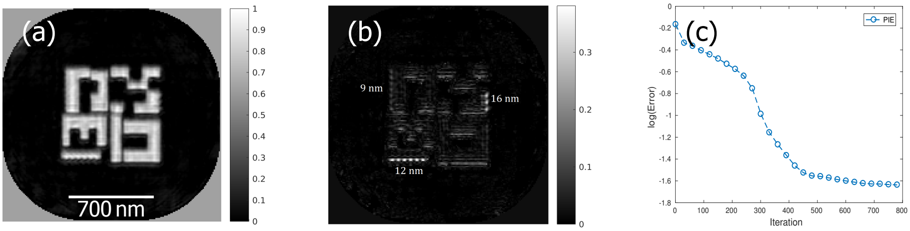

Figure 4.11: Ptychographic reconstructions: a) magnitude as given by the PIE; b) difference image given by PIE; c) error.

Fig. 4.11(b) reports an overall better reconstruction of the defects with respect to Fig. 4.6(b). Still the reconstruction is not as satisfactory as the one obtained with 5 probe positions and the inclusion of the prior term. This is highlighted in Table 4.4.

Table 4.4: **Extrusion defects SNR**

Defect Size [nm]	SNR with prior (5 probe positions)	SNR PIE (9 probe positions)
16	3.8	3
12	5.3	2.5
9	1.6	0.6

Decreasing SNR

The effect of the noise is studied here. A decrease of the SNR constitutes a problem for the retrieval of fine features that weakly scatter light, because their signature could be below the noise. A workaround could be to increase the radiation dose, but this could in turn damage the sample. The role of the prior as a regularizer is helpful in this, as it stabilizes the inversion and enables one to achieve a better reconstruction when the SNR decreases. Here we have decreased the SNR from 110 dB to 100 dB and 90 dB. The results, for the extrusion defects, are given in Figs. 4.12 and 4.13.

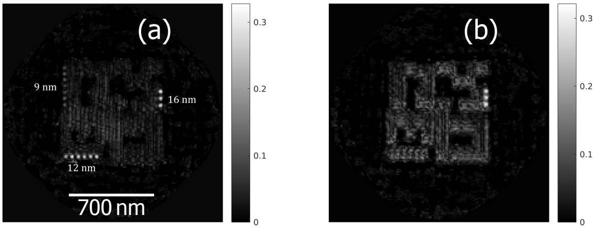

Figure 4.12: a) and b) show the object difference metric ΔO in Eq. 4.14 obtained when SNR = 100 dB. a) PIE with prior; b) PIE.

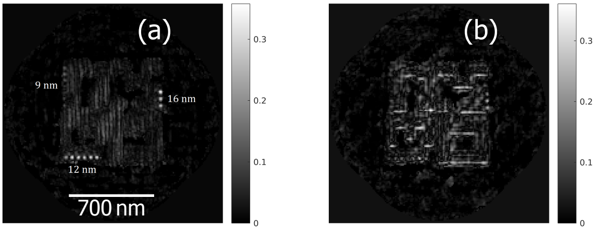

Figure 4.13: a) and b) show the object difference metric ΔO in Eq. 4.14 obtained when SNR = 90 dB. a) PIE with prior; b) PIE.

All of the defects are still visible in Fig. 4.12(a), while in Fig. 4.13(b) only the bigger ones are visible. In Fig. 4.13 the 16 and 12 nm size defects are visible when the prior is applied as a regularizer, but none of the defects can be detected by the PIE.

Using the prior as initial guess

The advantages of having a regularizer in Eq. (4.5) rather than using the prior only as a starting point were outlined in Section 4.1. Here we compare the amplitude of the difference images associated to extrusions type defects and obtained by using three different flavours of the ptychographic algorithm. Specifically we compare the results obtained by reconstructing the reflection function of the object by using:

- the update rule Eq. (4.11),
- the standard ptychographic iterative engine (PIE), and
- the PIE where the first guess of the object is the full object reflection function associated to the nominal mask.

Fig. 4.14 highlights results for an SNR = 110 dB. Fig. 4.15 shows results for an SNR = 100 dB.

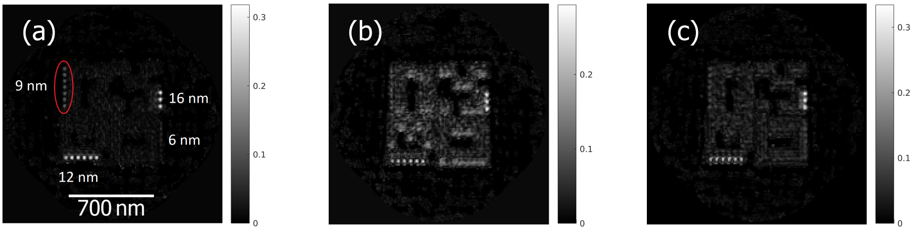

Figure 4.14: Difference images SNR 110 dB: a) Update rule Eq. (4.11); b) PIE; c) PIE with prior as starting guess.

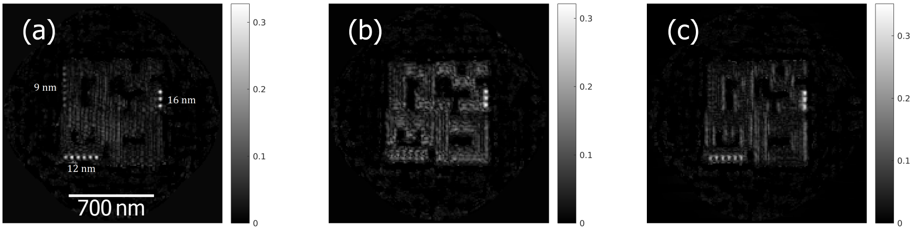

Figure 4.15: Difference images SNR 100 dB: a) Update rule Eq. (4.11); b) PIE; c) PIE with prior as starting guess.

As it can be noticed, a proper starting guess in the standard PIE yields a better reconstruction, however the absence of the regularizer – that stabilizes the reconstruction with respect to the noise and that promotes the retrieval of a better fit via a bias–variance trade–off – has a negative impact on the reconstruction.

Comparison with TV regularization

Recent works have reported the use of total variation as regularizer for denoising in ptychography [8–10]. Here we compare the results obtained by using the two different regularizers $R^{\alpha}(O) = \alpha||O(\mathbf{r}) - O_p(\mathbf{r})||^2$ and $R^{\alpha_{TV}}(O) = \alpha_{TV}||\nabla O(\mathbf{r})||_1$. Notice that, since $R^{\alpha_{TV}}(O)$ is non–smooth, gradient based approaches are not immediately applicable anymore. One workaround is to replace the ℓ_1 norm by a smooth approximation, another one, employed here, is to solve total variation via the alternating direction method of multipliers (ADMM). We begin by writing the following problem at the j–th probe position:

$$\begin{aligned} \min_{O_j,p} \quad & \left|\left| |\mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_j(\mathbf{r}))| - \sqrt{I_j(\mathbf{k})} \right|\right|^2 + \alpha_{TV}||p||_1, \\ \text{s.t.} \quad & p = \nabla O, \end{aligned} \tag{4.15}$$

The augmented Lagrangian for a real functional of complex variables can be written as [21]:

$$\begin{aligned} \mathcal{L}_\rho(O_j,p,y) = & \left|\left| |\mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_j(\mathbf{r}))| - \sqrt{I_j(\mathbf{k})} \right|\right|^2 \\ & + 2\,\mathrm{Re}(\langle p - \nabla O_j, y\rangle) + \rho||p - \nabla O_j||^2 + \alpha_{TV}||p||_1, \end{aligned} \tag{4.16}$$

where

- $\langle\cdot,\cdot\rangle$ is the inner product over the complex space,
- $\mathrm{Re}(z)$ is the real part of z,
- y is the Lagrangian multiplier. In this case y is to be intended as a function of position, i.e. it is a whole set of Lagrange multipliers, one for every position.
- ρ is the penalty parameter.

Eq. (4.16) can be further simplified scaling the variable y by $1/\rho$ [22]. At iteration n the ADMM solves the following steps:

$$\begin{aligned} O^{n+1} &= \arg\min_{O} \mathcal{L}_\rho(O,p^n,y^n), \\ p^{n+1} &= \arg\min_{p} \mathcal{L}_\rho(O^{n+1},p,y^n), \\ y^{n+1} &= y^n + \rho(p^{n+1} - \nabla O^{n+1}). \end{aligned} \tag{4.17}$$

The derivations necessary to minimize Eq. (4.17) are analogous to the ones that can be found elsewhere [9] and are here omitted. Problem (4.15) is minimized per probe position, the subproblem with respect to O_j is solved via a steepest descent and the subproblem with respect to p has a closed–form solution in the form of a soft–thresholding operator:

$$p^{opt,n} = max\left\{\left|\nabla O_j^n\right| - \frac{y^{n-1}}{\rho}\right\} \mathrm{sign}\left(\nabla O_j^n - \frac{y^{n-1}}{\rho}\right) \tag{4.18}$$

where $p^{opt,n}$ is the optimal value of p at iteration n and sign is the extension of the sign function as applied in the complex space:

$$\mathrm{sign}(z) = \frac{z}{|z|} \tag{4.19}$$

where $z \in \mathbb{C}$.
The regularizer α_{TV} in Eq. 4.16 has been computed by the L–curve[23].
The object difference images for SNR=110 dB and for SNR=90 dB, for extrusions defects, are shown in Figs. 4.16 and 4.17.

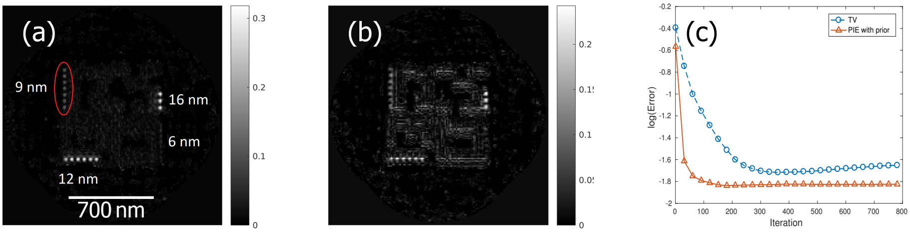

Figure 4.16: Object difference image SNR 110 dB: a) Update rule (4.11); b) TV regularization; c) error, Eq. (4.13).

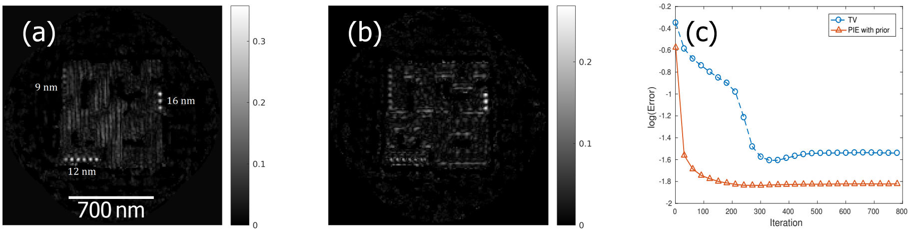

Figure 4.17: Object difference image SNR 90 dB: a) Update rule (4.11); b) TV regularization; c) error, , Eq. (4.13)

Although TV regularization produces results that are qualitatively superior to the PIE (Figs. 4.6(b) and 4.13(b)) the proposed method that includes object prior information in the optimization enables a more robust reconstruction.

Interlude

We have now seen that the inclusion of *a–priori* information – included in the ptychographic method using additional fully three dimensional, model based and domain specific fully rigorous solvers – yields a more robust reconstruction at an increased resolution and that this paves the way for the inspection of finer defects. The question arises on whether further algorithmic developments can yield an even better

reconstruction, possibly revealing the $6nm$ size defects in Figs.4.2(b,c) that we could not manage to resolve so far.
The remaining part of this chapter discloses update rules that have been found to assist in the further improvement of the reconstruction under Poisson noise.

4.2. Refinement strategies for optimal inclusion of prior information in ptychography

It is common practice in the phase retrieval community to cluster different update rules and different methods together in order to avoid stagnation and achieve a better reconstruction [5, 24, 25]. In the following we present cost functionals that we have found to enable better reconstructions when combined with the update rule given by the minimization of Eq. (4.5) in the presence of Poisson noise.

4.2.1. Maximum a posteriori estimation

The tools offered by statistical regularization theory can be employed (cf. Section 1.2) to derive a cost functional that accounts for the knowledge of both the noise model and the prior information. If photon counting is the main source of noise in the physical set–up then the number of photons detected follows the Poissonian statistics. The detected number of photons at the j–th probe position, $n_{m,j}$ is interrelated to the measured intensities $I_{m,j}$ by:

$$n_{m,j} \propto \frac{I_{m,j}}{h\nu}, \quad \text{where} \quad \nu = \frac{v}{\lambda}, \tag{4.20}$$

where h is the Plank's constant, ν is the frequency, λ is the wavelength and v is the speed of light within a certain medium. If we assume the incoming light beam to be monochromatic, the intensity distribution will also be Poissonian and, at the j–th probe position, the probability of measuring a certain number of photons $n_{m,j}$, given a certain object $O_j(\mathbf{r})$, will be given by the likelihood [5]:

$$L(O_j) := p(n_{m,j}, O_j) = \prod_{\mathbf{k}} \frac{I_j(\mathbf{k})^{n_{m,j}(\mathbf{k})}}{n_{m,j}(\mathbf{k})!} e^{-I_j(\mathbf{k})}. \tag{4.21}$$

where $I_j(\mathbf{k})$ is the guessed intensity as given by $I_j = |\mathcal{F}(P(\mathbf{r}-\mathbf{R}_j))O_j(\mathbf{r})|^2$. Following Bayes theorem we can introduce, given Eq. (4.21) and a prior distribution for the model parameters, the posterior:

$$\pi_{post}(O_j|n_{m,j}) = \frac{L(O_j)\pi_{prior}(O_j)}{\int L(O_j)\pi_{prior}(O_j)dO_j}, \tag{4.22}$$

where π_{prior} represents the prior distribution. We assume the prior distribution to be a Gaussian peaked at a certain prior object O_p:

$$\pi_{prior}(O_j) = \frac{e^{-(O_j-O_{p,j})^H\Gamma^{-1}(O_j-O_{p,j})}}{|\pi\Gamma|} = \frac{1}{\pi\sigma^2}\exp\left[\frac{-|O_j - O_{p,j}|^2}{\sigma^2}\right], \tag{4.23}$$

where H denotes the Hermitian transpose and Γ is a Hermitian positive definite matrix that contains the complex variance. We assumed in the right hand side of Eq. (4.23) that Γ is diagonal with equal real elements given by σ^2. We notice that the standard deviation in Eq. (4.23) essentially plays the role of the regularization parameter. The reconstruction of the full posterior distribution in Eq. (4.22) is extremely computationally demanding, and it is common to retrieve from it only point estimators that are statistically relevant. The maximum–a–posteriori (MAP) estimator is defined as the value of the sought object that maximizes Eq. (4.22):

$$\begin{aligned} O_{MAP,j} &= \arg\max_{O_j} \pi_{post}(O_j|n_{m,j}) = \arg\min_{O_j}\{-\log L(O_j) - \log \pi_{prior}(O_j)\} = \\ &= \arg\min_{O_j} \mathcal{L}_{MAP}(O_j, n_{m,j}) \end{aligned} \tag{4.24}$$

The update of the object reflection function at the j–th probe position will be given by:

$$O_{MAP,j} = \arg\min\left\{\sum_{\mathbf{k}}[I_j(\mathbf{k}) - n_{m,j}(\mathbf{k})\log(I_j(\mathbf{k}))] + \sum_{\mathbf{r}}\frac{1}{\sigma^2}|O_j(\mathbf{r}) - O_{p,j}(\mathbf{r})|^2\right\} \tag{4.25}$$

where we have left out constant terms. In Eq. (4.25) $\mathbf{k}$ runs over the pixels in the Fourier space and $\mathbf{r}$ over the pixels in real space. The prior $O_{p,j}$ is the same as the one we used previously in Section 4.1.

Eq. (4.25) can be minimized analytically and, when including Eq. (4.4), the gradient descent rule reads:

$$\begin{aligned} O_{MAP,j,n+1} = O_{j,n} - \beta\nabla\mathcal{L}_{MAP}(O_j, I_j) = O_{j,n} + \beta\frac{|P_j|}{|P_{j,max}|}\frac{P_j^*}{(|P_j|^2+c)}\mathcal{F}^{-1}(\frac{n_{m,j}}{I_j}\Psi_{j,n} - \Psi_{j,n}) \\ + \alpha\frac{|P_j|}{|P_{j,max}|}\frac{P_j^*}{(|P_j|^2+c)}\mathcal{F}^{-1}(\Psi_{p,j}^{Maxw} - \Psi_{j,n}) \end{aligned} \tag{4.26}$$

where $\Psi_{j,n} = \mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_j(\mathbf{r}))$, $\alpha = \beta/\sigma^2$ and the factor $|P_j|^2$ at the denominator needs to be included so that Eq. (4.26) has the dimensions of an object.

4.2.2. Iteratively refined regularizer

In the section above we have changed the update rule to incorporate the knowledge of the noise model and we have introduced an assumed prior distribution for the object reflection function to derive a certain update rule from statistical considerations. Another interesting update could be obtained by employing the update rule that stems from the minimization of Eq. (4.5) followed by the minimizer of

$$\mathcal{L}_{IR}(O_j, I_{m,j}) = \left|\left||\mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_j(\mathbf{r}))| - \sqrt{I_{m,j}(\mathbf{k})}\right|\right|^2 + \alpha\left|\left|O_j(\mathbf{r}) - O_{j,n-1}(\mathbf{r})\right|\right|^2, \tag{4.27}$$

where the subscript IR stresses the fact that the regularization weight is changed iteratively. The reasoning behind this proposal is intuitive. As the optimization proceeds, and once it has reached or is close to convergence, the retrieved solution at the previous iteration should be closer to the ground truth than the prior $O_{p,j}$ in Eq. (4.5). Therefore one could penalize large deviations of the object reflection function with respect to $O_{j,n-1}$ rather than O_p. In other words we are seeking to introduce a regularizer that is closer to the true solution than $O_p(\mathbf{r})$. The update rule in this case will read:

$$\begin{aligned} O_{IR,j,n+1} = O_{j,n} - \beta\nabla\mathcal{L}_{IR}(O,I_j) = O_{j,n} &+ \beta\frac{|P_j|}{|P_{j,max}|}\frac{P_j^*}{(|P_j|^2+c)}\mathcal{F}^{-1}(\Psi_{c,j,n}-\Psi_{j,n}) \\ &+ \alpha'\frac{|P_j|}{|P_{j,max}|}\frac{P_j^*}{(|P_j|^2+c)}\mathcal{F}^{-1}(\Psi_{j,n-1}-\Psi_{j,n}), \end{aligned} \tag{4.28}$$

where $\alpha' = \alpha\beta$ and $\Psi_{c,j,n}$ is the revised or corrected wavefront obtained by replacing the guessed amplitudes with the measured ones.

4.3. Results

In what follows we apply the methods outlined above to the reconstruction of defective EUV mask layouts. The data set is the same as we used previously and it is corrupted with Poisson noise. The materials are the ones we used previously and are reported in Table 4.1.

To understand whether the use of Eqs.(4.26) and (4.28) yields any benefits for our specific application we have repeated the same procedure followed in the previous section, i.e. we have compared the reconstruction of the defected mask with the reconstruction of the defect–free actual mask. The two reconstructions are subtracted one from the other to identify the defects at their locations. The impact of the defects is quantified by the defect SNR, called SNR$_\delta$, as defined in 4.12. In EUV mask defectivity studies this procedure is known as die–to–database inspection. This investigation is done using the PIE with prior, Eq. (4.5), followed by the update given by Eq. (4.26) or the one in Eq. (4.28).

4.3.1. Extrusion defects

Figure 4.18 shows the reconstructed amplitudes and phases of the object functions for extrusion type defects. We used 600 iterations with the method presented in our previous work followed by 200 iterations of either Eq. (4.26) or (4.28). The SNR is 100 dB.

The difference images obtained after subtracting the defect and defect–free amplitudes of the retrieved object function are given below in Fig. 4.19.

The defect SNR, defined in Eq. (4.12), is reported in Table 4.5.

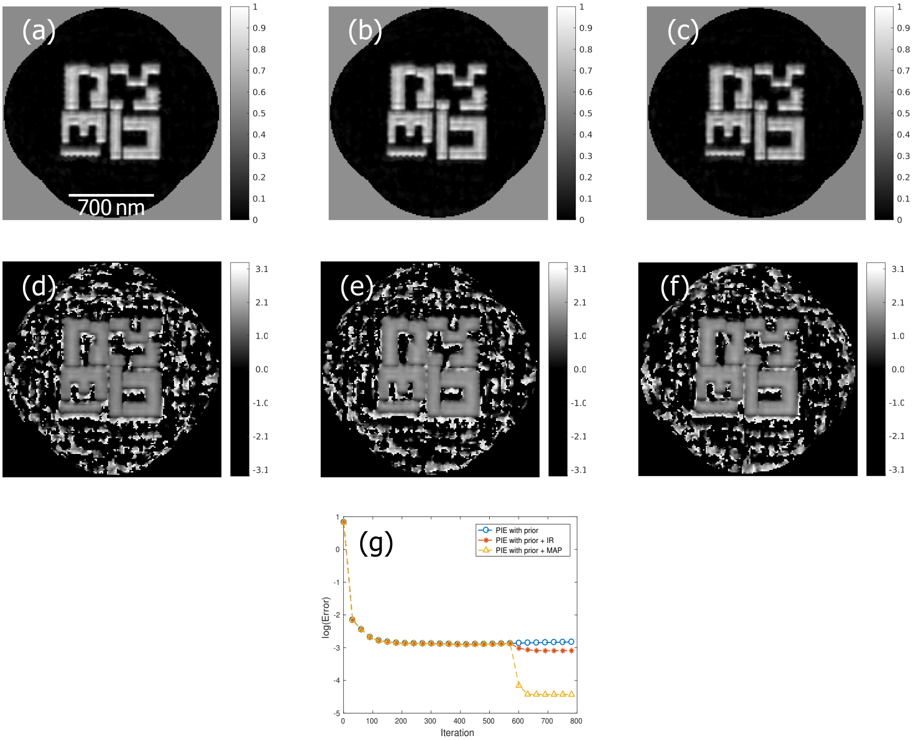

Figure 4.18: Ptychographic reconstructions: a) amplitude as given by the MAP, Eq. (4.26); b) amplitude given by the iteratively refined regularization weight, Eq. (4.28); c) amplitude given by the PIE with prior information; d)–f) phase relative to Figs. a)–c); g) error in the far field

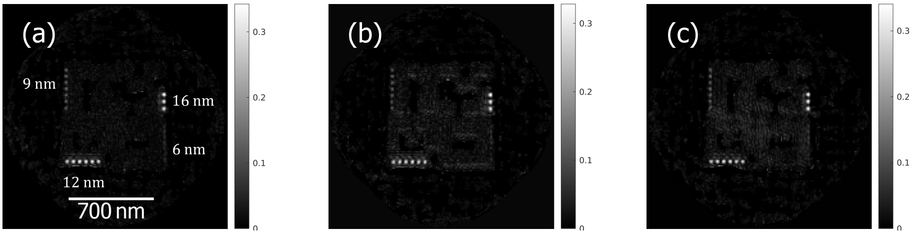

Figure 4.19: Difference images for extrusion type defects. a)MAP, Eq. (4.26), b)IR Eq. (4.28), c) PIE with prior information

Table 4.5: **Extrusions defects SNR$_\delta$**

Defect Size [nm]	MAP (5 probe positions)	IR	Prior PIE
16	5.2	4.8	4.7
12	6.7	5.1	5.2
9	3.5	2	2

The difference images in Fig. 4.19 and Table 4.5 show that the MAP refinements give the best results – in terms of convergence and defects signal to noise ratio – among the three algorithms we discussed. Table 4.5 also shows that further refinements to the reconstruction of the object reflection function obtained by using the iteratively refined regularizer idea discussed above did not lead to a substantial improvement in defect signal to noise ratio with respect to the case in which the reflection function of the object is retrieved by processing the data using the PIE with prior information.

4.3.2. Intrusion defects

Figure 4.20 shows reconstructions in amplitude and phase for intrusion type defects. The reconstruction proceeded as discussed in the previous section.

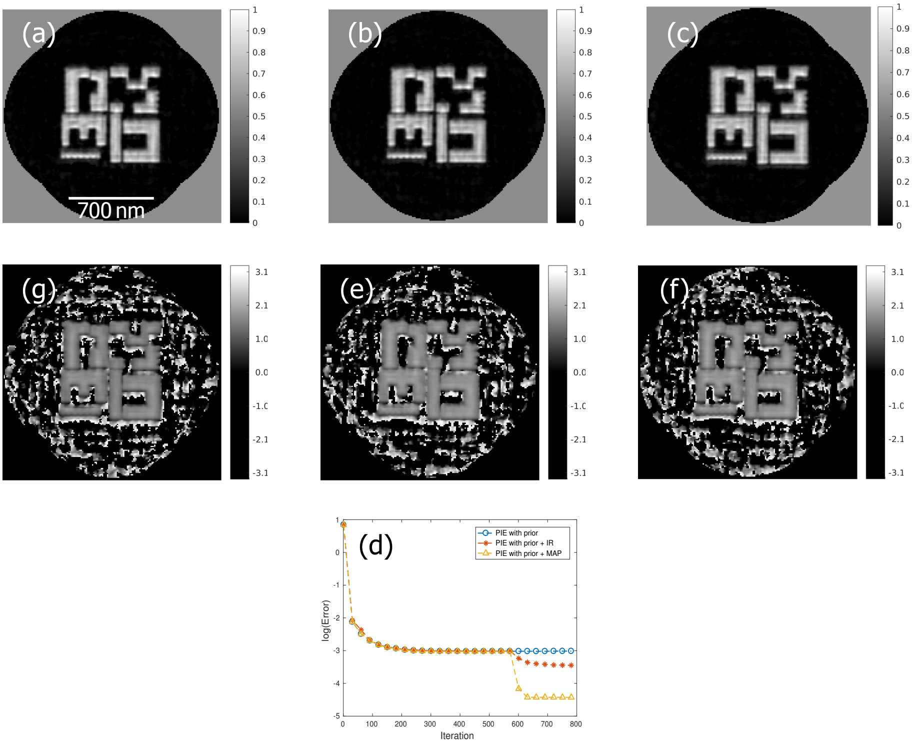

Figure 4.20: Ptychographic reconstructions: a) amplitude as given by the MAP, Eq. (4.26); b) amplitude given by the iteratively refined regularization weight, Eq. (4.28); c) amplitude given by the PIE with prior information; d)–f) phase relative to Figs. a)–c); g) error in the far field

The difference images are given below in Fig 4.21.

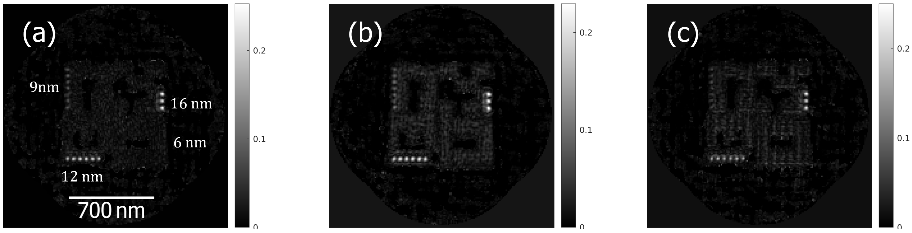

Figure 4.21: Difference images for intrusion type defects. a) MAP, Eq. (4.26), b) IR Eq. (4.28), c) PIE with prior information

The defect SNR is reported below in Table 4.6.
Similarly to the case of the extrusion type defects Table 4.6 shows that the MAP refinements yield an SNR_δ that is higher than the one given by the PIE with prior information and by the iteratively refined regularizer.

Table 4.6: **Intrusion defects SNR$_\delta$**

Defect Size [nm]	MAP (5 probe positions)	IR	Prior PIE
16	5	5.2	3.5
12	5	4.6	3.4
9	3.5	2	2

4.3.3. Comparison with a sparsity prior for the object

As the object covers a small part of the reconstruction matrix it is interesting to compare results with the one given by a sparsity constraint for the object. We solve, per probe position, the following problem via the alternating direction method of multiplier:

$$\begin{aligned} &\min_{O_j,p} \quad \left\| |\mathcal{F}(P(\mathbf{r}-\mathbf{R}_j)O_j(\mathbf{r}))| - \sqrt{I_j(\mathbf{k})} \right\|^2 + \alpha||p||_1, \\ &\text{s.t.} \quad p = O. \end{aligned} \tag{4.29}$$

Results are given, for extrusion type defects, in the figure below, and the defects are better reconstructed when using the full knowledge of the prior object. Another interesting choice could be to look for a sparse reconstruction of the object around O_p, in which case the regularizer would be $||O - O_p||_1$, but this is not discussed in this thesis.

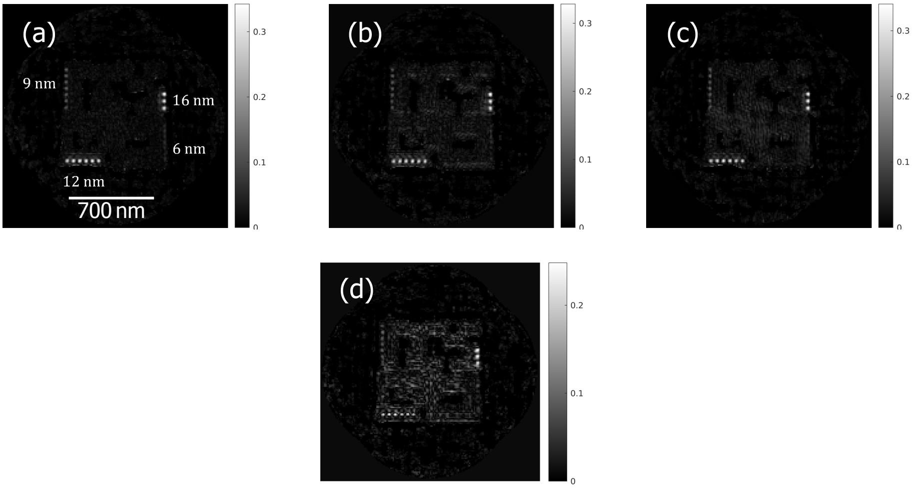

Figure 4.22: Difference images for extrusion type defects. a) MAP, Eq. (4.26), b) IR, Eq. (4.28), c) PIE with prior info,d) Lasso, Eq. (4.29).

Summary

In this chapter we have seen how the ptychographic reconstruction benefits from the inclusion of a prior term which is generated by the additional 3D, model–based and fully rigorous Maxwell solvers. We have presented refinement strategies for the inclusion of *a–priori* information in ptychography which lead to the imaging of smaller defects than the usual case. The algorithmic efforts presented in this chapter paved the way towards the imaging of finer defects and enabled a tighter inspection of the patterned absorber on top of an EUV mask. However the finest defects, of a size smaller than the Abbe limit, proved to be challenging to image. In order to attempt a more satisfactory inspection of the target, we have thought of a different approach that enriches the usual scalar ptychographic method by means of a vectorial formulation of ptychography and which could possibly enable the reconstruction of such fine details (i.e. defects whose size is below the Abbe limit). In the next section we introduce the framework for polarization sensitive imaging of semiconductor samples via vectorial ptychography.

References

[1] P. Ansuinelli, W. M. J. Coene, and H. P. Urbach, *Improved ptychographic inspection of EUV reticles via inclusion of prior information,* Appl. Opt. **59**, 5937 (2020).

[2] P. Ansuinelli, W. M. Coene, and H. P. Urbach, *Refinement strategies for optimal inclusion of prior information in ptychography,* in *Extreme Ultraviolet Lithography 2020*, Vol. 11517, edited by P. P. Naulleau, P. A. Gargini, T. Itani, and K. G. Ronse, International Society for Optics and Photonics (SPIE, 2020) pp. 13 – 21.

[3] J. M. Rodenburg and H. M. L. Faulkner, *A phase retrieval algorithm for shifting illumination,* Applied Physics Letters **85**, 4795 (2004), https://doi.org/10.1063/1.1823034 .

[4] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, *Solving inverse problems using data-driven models,* Acta Numerica **28**, 1–174 (2019).

[5] P. Thibault and M. Guizar-Sicairos, *Maximum-likelihood refinement for coherent diffractive imaging,* New Journal of Physics **14**, 063004 (2012).

[6] Q. Fang, R. Moore, D. Kopans, and D. Boas, *Compositional-prior-guided image reconstruction algorithm for multi-modality imaging,* Biomedical optics express **1**, 223 (2010).

[7] A. Mohammad-Djafari, *Bayesian approach with prior models which enforce sparsity in signal and image processing,* EURASIP Journal on Advances in Signal Processing **2012**, 52 (2012).

[8] H. Chang, P. Enfedaque, J. Zhang, J. Reinhardt, B. Enders, Y.-S. Yu, D. Shapiro, C. G. Schroer, T. Zeng, and S. Marchesini, *Advanced denoising for x-ray ptychography,* Opt. Express **27**, 10395 (2019).

[9] H. Chang, P. Enfedaque, and S. Marchesini, *Iterative Joint Ptychography-Tomography with Total Variation Regularization,* arXiv e-prints , arXiv:1902.05647 (2019), arXiv:1902.05647 [eess.IV] .

[10] V. Nikitin, S. Aslan, Y. Yao, T. Biçer, S. Leyffer, R. Mokso, and D. Gürsoy, *Photon-limited ptychography of 3d objects via bayesian reconstruction,* OSA Continuum **2**, 2948 (2019).

[11] S. Zayko, E. Mönnich, M. Sivis, D. M. Mai, T. Salditt, S. Schäfer, and C. Ropers, *Coherent diffractive imaging beyond the projection approximation: waveguiding at extreme ultraviolet wavelengths.* Optics express **23 15**, 19911 (2015).

[12] J. Rodenburg and A. Maiden, *Ptychography,* in *Springer Handbook of Microscopy*, edited by P. W. Hawkes and J. C. H. Spence (Springer International Publishing, Cham, 2019) pp. 819–904.

[13] X. Wei, A. J. Wachters, and H. P. Urbach, *Finite-element model for three-dimensional optical scattering problems,* J. Opt. Soc. Am. A **24**, 866 (2007).

[14] J. Pomplun, S. Burger, L. Zschiedrich, and F. Schmidt, *Adaptive finite element method for simulation of optical nano structures,* physica status solidi (b) **244**, 3419 (2007).

[15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York Inc., New York, NY, USA, 2001).

[16] I. Mochi, K. A. Goldberg, B. L. Fontaine, A. Tchikoulaeva, and C. Holfeld, *Actinic imaging of native and programmed defects on a full-field mask,* in *Extreme Ultraviolet (EUV) Lithography*, Vol. 7636,, International Society for Optics and Photonics (SPIE, 2010) pp. 425 – 433.

[17] M. C. van Beurden, *Fast convergence with spectral volume integral equation for crossed block-shaped gratings with improved material interface conditions,* J. Opt. Soc. Am. A **28**, 2269 (2011).

[18] M. C. van Beurden, *A spectral volume integral equation method for arbitrary bi-periodic gratings with explicit Fourier factorization,* Progress In Electromagnetics Research B **36**, 133 (2012).

[19] D. Gardner, M. Tanksalvala, E. Shanblatt, X. Zhang, B. Galloway, C. Porter, R. Karl, C. Bevis, D. Adams, M. Murnane, and G. Mancini, *Subwavelength coherent imaging of periodic samples using a 13.5 nm tabletop high-harmonic light source,* Nature Photonics **11** (2017), 10.1038/nphoton.2017.33.

[20] A. M. Maiden and J. M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging,* Ultramicroscopy **109**, 1256 (2009).

[21] L. Li, X. Wang, and G. Wang, *Alternating direction method of multipliers for separable convex optimization of real functions in complex variables,* Mathematical Problems in Engineering **2015**, 1 (2015).

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers,* Found. Trends Mach. Learn. **3**, 1–122 (2011).

[23] P. C. Hansen, *The l-curve and its use in the numerical treatment of inverse problems,* in *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering* (WIT Press, 2000) pp. 119–142.

[24] K. Wakonig, A. Diaz, A. Bonnin, M. Stampanoni, A. Bergamaschi, J. Ihli, M. Guizar-Sicairos, and A. Menzel, *X-ray fourier ptychography,* Science Advances **5** (2019), 10.1126/sciadv.aav0282, https://advances.sciencemag.org/content/5/2/eaav0282.full.pdf .

[25] X. Xu, A. P. Konijnenberg, S. Pereira, and H. P. Urbach, *Phase retrieval of the full vectorial field applied to coherent Fourier scatterometry,* Optics Express **25**, 29574 (2017).

5

Polarization sensitive imaging of semiconductor structures: vectorial ptychography

We have previously seen that the ptychographic method is a scalar method. A vectorial extension of ptychography for the imaging of anisotropic material properties was presented in [1]. In this chapter we derive an explicit formalism for polarization sensitive ptychographic imaging of semiconductor structures. This could be interesting for the following reasons:

- it may be that one of the elements of the Jones matrix, which describes the optical properties of the sample with respect to the polarization of the incoming light, yields a better resolution in the reconstructed image than the other elements, or
- we are interested in retrieving more information about the target (i.e. the full scattering matrix).

5.1. The formalism

The formalism presented below is adapted in order to enforce consistency with the volume integral Maxwell solver that has been used throughout this work. Because the solver is implemented with periodic boundary conditions the isolated structures are simulated via a supercell approach, and we remark that the solver takes as input plane waves which can be polarized along the s or p directions and gives as output the complex amplitudes of the computed scattering orders in the same polarization basis. Eqs. (5.2–5.9) follow from the formalism presented and discussed in [2]. Suppose that the scatterer is located in the (x, y) plane. If we assume the probe to be a scalar quantity we can approximate the incoming Gaussian beam, propagating towards the scatterer in the negative z direction, via its discrete angular spectrum:

$$P(x,y,z) = \sum_{k_x^i,k_y^i} G(k_x^i,k_y^i) exp[i(k_x^i x + k_y^i y + k_z^i z)], \tag{5.1}$$

where:

- $[k_x^i, k_y^i, k_z^i] = \mathbf{k}^i$ are the Cartesian components of a wavevector of an incoming plane wave under a certain angle of incidence, with $k_z^i = -\sqrt{k_0^2 - (k_x^i)^2 - (k_y^i)^2}$ (assuming the probe to be propagating in the direction of negative z) and $k_0 = 2\pi/\lambda$, and
- $G(k_x^i, k_y^i)$ represents a scalar Gaussian scaling function.

Eq. (5.1) states that a light beam that is propagating in the positive/negative z direction can be expressed from its values at $z = 0$ by using two Fourier transforms. First, the field in the plane $z = 0$ must be transformed, and subsequently the field for $z > 0$ is obtained by calculating the inverse Fourier transform of the product $\mathcal{F}(E)(k_x, k_y, 0)e^{\pm izk_z}$ where the sign in the exponential depend on the direction of propagation (positive/negative z). Eq. (5.1) holds only in the scalar approximation, an extension to the vectorial case is given below.

We start by defining the two vectors:

$$\hat{\mathbf{s}}(\mathbf{k}) = \frac{1}{\sqrt{k_x^2 + k_y^2}} \begin{pmatrix} k_y \\ -k_x \\ 0 \end{pmatrix}, \tag{5.2}$$

$$\hat{\mathbf{p}}(\mathbf{k}) = \frac{1}{|\mathbf{k}|\sqrt{k_x^2 + k_y^2}} \begin{pmatrix} k_x k_z \\ k_y k_z \\ -k_x^2 - k_y^2 \end{pmatrix}, \tag{5.3}$$

where

$$|\mathbf{k}| = (\mathbf{k} \cdot \mathbf{k}^*)^{1/2} = (k_x^2 + k_y^2 + |k_z|^2)^{1/2}, \tag{5.4}$$

is the modulus of the wavevector $\mathbf{k}$. It can be shown that the basis given by $[\mathbf{k}/|\mathbf{k}|, \mathbf{s}(\hat{\mathbf{k}}), \mathbf{p}(\hat{\mathbf{k}})]$ is an orthonormal basis of $\mathbb{C}^3$. This means that the electric field, $\mathbf{E}$, can be expressed – in the Fourier space – as a linear combination of such basis vectors. We notice that $\hat{\mathbf{s}}$ and $\hat{\mathbf{p}}$ are given by a combination of Cartesian components.

As explained below Eq. (5.1) the electric field at a distance $z > 0$ can be computed via the angular spectrum by its values at $z = 0$:

$$\mathbf{E}(x,y,z) = \int\int \mathcal{F}(\mathbf{E}(k_x, k_y, z = 0))e^{i\mathbf{k}\cdot\mathbf{r}} dk_x dk_y. \tag{5.5}$$

Because the divergence–free condition in a space $z > 0$ without sources implies $\mathbf{k}\cdot\mathbf{E} = 0$, it follows from Eq. (5.5) that

$$\mathbf{k}\cdot\mathcal{F}(\mathbf{E}(k_x, k_y, 0)) = 0. \tag{5.6}$$

Therefore there exists $A_s(\mathbf{k})$ and $A_p(\mathbf{k})$ such that:

$$\mathcal{F}(\mathbf{E})(k_x, k_y, 0) = A_s(\mathbf{k})\hat{\mathbf{s}}(\mathbf{k}) + A_p(\mathbf{k})\hat{\mathbf{p}}(\mathbf{k}), \tag{5.7}$$

and hence the angular spectrum, Eq. (5.5) can be rewritten as:

$$\mathbf{E}(x, y, z) = \int\int [A_s(\mathbf{k})\hat{\mathbf{s}}(\mathbf{k}) + A_p(\mathbf{k})\hat{\mathbf{p}}(\mathbf{k})]e^{i\mathbf{k}\cdot\mathbf{r}}dk_x dk_y. \tag{5.8}$$

which is the vectorial extension of Eq. (5.1). By using Eqs. (5.2, 5.3) we obtain that the x and y components of the electric field can be computed from the $[\mathbf{s}, \mathbf{p}]$ basis:

$$\begin{aligned}
E_x(\mathbf{r}) &= \int\int\left[A_s(\mathbf{k})\frac{k_y}{\sqrt{k_x^2 + k_y^2}} + A_p(\mathbf{k})\frac{k_x k_z}{|\mathbf{k}|\sqrt{k_x^2 + k_y^2}}\right]e^{i\mathbf{k}\cdot\mathbf{r}}dk_x dk_y \\
&= \int\int E_x(k_x, k_y)e^{i\mathbf{k}\cdot\mathbf{r}}dk_x dk_y, \\
E_y(\mathbf{r}) &= \int\int\left[A_s(\mathbf{k})\frac{-k_x}{\sqrt{k_x^2 + k_y^2}} + A_p(\mathbf{k})\frac{k_y k_z}{|\mathbf{k}|\sqrt{k_x^2 + k_y^2}}\right]e^{i\mathbf{k}\cdot\mathbf{r}}dk_x dk_y \\
&= \int\int E_y(k_x, k_y)e^{i\mathbf{k}\cdot\mathbf{r}}dk_x dk_y,
\end{aligned} \tag{5.9}$$

and, by applying the source–free condition $\mathbf{k}\cdot\mathbf{E} = 0$, the z component of the electric field can be calculated from $[E_x, E_y]$ as:

$$E_z(\mathbf{r}) = \int\int \frac{-(k_x E_x(k_x, k_y) + k_y E_y(k_x, k_y))}{k_z}e^{i\mathbf{k}\cdot\mathbf{r}}dk_x dk_y. \tag{5.10}$$

The formalism presented above allows us to establish a link among the Cartesian coordinates and the $[\hat{\mathbf{s}}, \hat{\mathbf{p}}]$ basis in which the Maxwell solver operates.
We discuss below, in more detail, the steps we went through during the simulations.
We start by noting, from Eqs. (5.9), that:

$$\begin{aligned}
\begin{bmatrix} E_x(k_x, k_y) \\ E_y(k_x, k_y) \end{bmatrix} &= \begin{bmatrix} \frac{k_y}{\sqrt{k_x^2+k_y^2}} & \frac{k_x k_z}{|\mathbf{k}|\sqrt{k_x^2+k_y^2}} \\ \frac{-k_x}{\sqrt{k_x^2+k_y^2}} & \frac{k_y k_z}{|\mathbf{k}|\sqrt{k_x^2+k_y^2}} \end{bmatrix}\begin{bmatrix} A_s(k_x, k_y) \\ A_p(k_x, k_y) \end{bmatrix} = \mathcal{A}\begin{bmatrix} A_s(k_x, k_y) \\ A_p(k_x, k_y) \end{bmatrix}, \\
\mathcal{A} &= \begin{bmatrix} \frac{k_y}{\sqrt{k_x^2+k_y^2}} & \frac{k_x k_z}{|\mathbf{k}|\sqrt{k_x^2+k_y^2}} \\ \frac{-k_x}{\sqrt{k_x^2+k_y^2}} & \frac{k_y k_z}{|\mathbf{k}|\sqrt{k_x^2+k_y^2}} \end{bmatrix}.
\end{aligned} \tag{5.11}$$

We define the probe, in the Cartesian basis, by means of its discrete angular spectrum:

$$P(\mathbf{r}) = \sum_{k_x^i,k_y^i} \begin{bmatrix} G_x \\ G_y \end{bmatrix} e^{i\mathbf{k}^i\cdot\mathbf{r}}. \tag{5.12}$$

In Eq. (5.12) $\mathbf{k}^i$ is the incident wave vector and $[G_x, G_y]$ are the Cartesian coefficients which have been chosen in order to have a Gaussian probe in $[x, y]$.

We now define the scattered field, considering first the scalar case of a single incident plane wave defined by a certain $\mathbf{k}^i$ and polarized along a certain direction. When the plane wave hits the nanostructures on top of the EUV multilayer, multiple diffracted orders are excited and generate a scattered field, $\mathbf{E}_s(\mathbf{r})$, which is given by Eq. (5.13):

$$\mathbf{E}_s(r) = \sum_{k_x^S,k_y^S} [R_s(k_x^S,k_y^S)\hat{\mathbf{s}}(k_x^S,k_y^S) + R_p(k_x^S,k_y^S)\hat{\mathbf{p}}(k_x,k_y)]e^{i\mathbf{k}^S\cdot r}, \tag{5.13}$$

where $[R_s, R_p]$ are the reflection coefficients computed by the Maxwell solver that correspond to a certain scattered order which propagates along the direction given by $\mathbf{k}^s$ which depends on the incident k-vector $\mathbf{k}^i$, and which is defined by:

$$\begin{aligned} k_x^S &= k_x^i + \frac{2\pi m}{\Lambda}, \\ k_y^S &= k_y^i + \frac{2\pi n}{\Lambda}, \\ k_z^S &= \sqrt{k_0^2 - (k_x^S)^2 - (k_y^S)^2}, \end{aligned} \tag{5.14}$$

where Λ is the pitch of the supercell that we have used in our forward simulations (i.e. $\Lambda = 3.5\mu m$).

We now generalize Eq. (5.13) to the case in which the probe is a scalar field defined by multiple plane waves. We consider the probe given by eq. (5.1). If we assume that the response of the scatterer (i.e. the scattered field) is independent from the angle of incidence of the considered plane wave, a change in the angle of incidence will only result in a shift of the scattered far–field. Therefore we can write:

$$\mathbf{E}_s(\mathbf{r}) = \sum_{k_x^S,k_y^S}\sum_{k_x^i,k_y^i} G(k_x^i,k_y^i)[R_s(k_x^S-k_x^i,k_y^S-k_y^i)\hat{\mathbf{s}} + R_p(k_x^S-k_x^i,k_y-k_y^i)\hat{\mathbf{p}}]e^{i\mathbf{k}^S\cdot r}. \tag{5.15}$$

We finally move to the case in which the probe is a vectorial quantity as defined in Eq. (5.12), and by recalling the operator $\mathcal{A}$ in Eq. (5.11) we obtain that the scattered field is given by:

$$\begin{bmatrix} E_{s,x}(\mathbf{r}) \\ E_{s,y}(\mathbf{r}) \end{bmatrix} = \sum_{k_x^S, k_y^S} \sum_{k_x^i, k_y^i} \mathcal{A}(k_x^S, k_y^S) \begin{bmatrix} R_{ss}(k_x^S - k_x^i, k_y^S - k_y^i) & R_{sp}(k_x^S - k_x^i, k_y^S - k_y^i) \\ R_{ps}(k_x^S - k_x^i, k_y^S - k_y^i) & R_{ps}(k_x^S - k_x^i, k_y^S - k_y^i) \end{bmatrix} \cdot \\ \cdot \mathcal{A}^{-1}(k_x^i, k_y^i) \begin{bmatrix} G_x(k_x^i, k_y^i) \\ G_y(k_x^i, k_y^i) \end{bmatrix} e^{i\mathbf{k}^S \cdot \mathbf{r}}. \tag{5.16}$$

We note again that the scattering matrix in Eq. (5.16) has been written in the $[s, p]$ basis because our solver computes the reflection coefficients of the scattering orders in such basis.

We remark that we have written $\mathcal{A}(k_x^S, k_y^S)$ and $\mathcal{A}(k_x^i, k_y^i)$ to stress that the operator $\mathcal{A}$ is generally different for the incident wave and for the scattered diffraction orders. By incorporating the effect of the operator $\mathcal{A}$ in the scattering matrix we obtain

$$\begin{bmatrix} E_{s,x}(\mathbf{r}) \\ E_{s,y}(\mathbf{r}) \end{bmatrix} = \sum_{k_x^S, k_y^S} \sum_{k_x^i, k_y^i} \begin{bmatrix} R_{xx}(k_x^S - k_x^i, k_y^S - k_y^i) & R_{xy}(k_x^S - k_x^i, k_y^S - k_y^i) \\ R_{yx}(k_x^S - k_x^i, k_y^S - k_y^i) & R_{yy}(k_x^S - k_x^i, k_y^S - k_y^i) \end{bmatrix} \cdot \\ \cdot \begin{bmatrix} G_x(k_x^i, k_y^i) \\ G_y(k_x^i, k_y^i) \end{bmatrix} e^{i\mathbf{k}^S \cdot \mathbf{r}}, \tag{5.17}$$

where the input and the output are both given in x and y. We notice that Eq. (5.16, 5.17) has a convolution structure in the Fourier space which, although expressed in a vectorial framework, is similiar to the convolution structure in the Fourier space that is at the core of the usual scalar ptychographic approach.

In the usual ptychographic approach the probe is polarized along either the x or the y direction and we measure the total intensity $I = |E_{s,x}|^2 + |E_{s,y}^T|$ and we proceed to retrieve the object transmission/reflection function by applying the ptychographic algorithm. Here, our aim will be to try to reconstruct all of the elements of the scattering matrix in Eq. (5.17). This can be done by retrieving each component of the scattering matrix separately. Notice that in this case the phase relation among the elements of the scattering matrix would be lost. A way to retain this information consists in polarizing the probe at $45°$ degrees (i.e. $[G_x, G_y] = [1, 1]$) so that, for instance, the intensity component along x would be $|E_{s,x}|^2 \propto |R_{xx}|^2 + |R_{xy}|^2 + 2|R_{xx}||R_{xy}|cos(\Delta\phi)$ where $\Delta\phi$ represents the phase difference among the two terms. In what follows we have not added this rotation, therefore the elements of the scattering matrix are defined up to a relative phase factor and the phase relation among the elements of the scattering matrix is lost.

5.2. Numerical results

In the first instance we applied the approach described in the previous paragraph considering as object the same patterned structures used in the previous chapter. In Fig. 5.1 we report the ptychographic reconstructions for the case in which the probe is polarized in the x and y direction and the total intensity $|E_{s,x}^T|^2 + |E_{s,y}^T|^2$ is detected. Negligible differences were observed among the results given by applying the PIE on the intensity data acquired by illuminating the sample with the two different probe polarizations. The reconstruction of the off–diagonal entries of the scattering matrix was unsuccessful in this case as cross–polarization effects were not pronounced enough to be imaged.

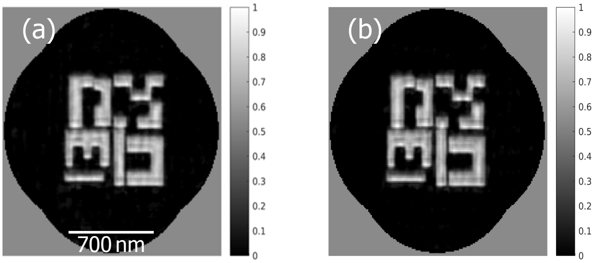

Figure 5.1: Reconstructions. a) Probe polarized in the x direction, total intensity detected. b) Probe polarized in the y direction, total intensity detected.

In order to make polarization effects more pronounced we considered a computational cell which was filled with structures with critical dimensions comparable to the wavelength. The cell was composed of a repetition of a few polygons on a $9x9$ grid and the distance from the center of two adjacent polygons is approximately 80 nm. The retrieved amplitude of the object reflection function associated to the diagonal elements of the scattering matrix and to the total intensity for the two probe polarizations is shown in Fig. 5.2.

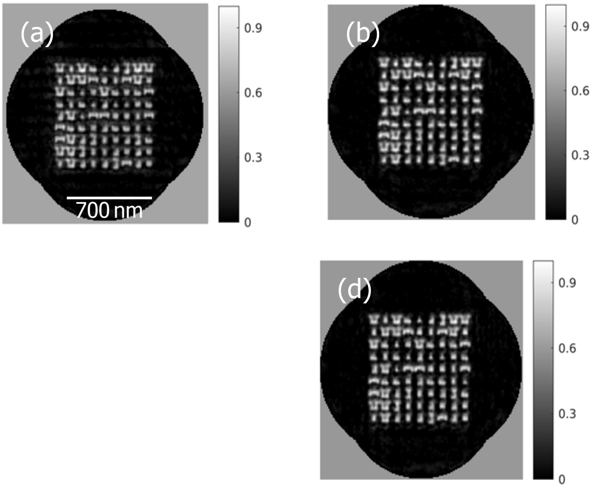

Figure 5.2: Object obtained by processing the dataset in which: a) the probe has been polarized along x and the intensity along x has been recorded (corresponding to R_{xx} in Eq. (5.17), b) the probe has been polarized along x and the total intensity, $|E^T_{s,x}|^2 + |E^T_{s,y}|^2$, has been recorded, c) the probe has been polarized along y and the intensity along y has been recorded (corresponding to R_{yy} in Eq. (5.17), d) the probe has been polarized along y and the total intensity, $|E^T_{s,x}|^2 + |E^T_{s,y}|^2$, has been recorded.

The reconstructions in Fig. 5.2 appeared quite modulated and required considerably more iterations before convergence was reached than the ones in Fig. 5.1. We did not notice major differences, in terms of resolution or reconstruction quality, among the two reconstructions obtained by polarizing the probe in the two different directions $[x, y]$ and recording the total intensity (Figs. 5.2 b,d). Interestingly, a few of the polygons in Figs. 5.2(a,c) – obtained by applying the ptychographic method (PIE) to the intensity given by the two diagonal elements of the matrix in Eq. (5.17) – looks better defined than the ones in Fig. 5.2(b,d) that were obtained processing the total intensity detected for the two incoming probe polarizations. Our attempts to reconstruct the off–axis elements of the polarization matrix were not successful, as cross–polarization effects were not pronounced for this particular case. Similar considerations were found to hold for another sample comprised of fewer polygons of sizes of about the wavelength.

Although we did not manage to reconstruct the full scattering matrix of the object we believe that the approach outlined in this section can be possibly useful for other types of optical set-ups and wavelengths where the optical contrast is more pronounced and were the higher NA can cause more marked polarization effects. Particularly the fact that the reconstructions obtained by processing the detected intensities as given by the diagonal entries of the scattering matrix in Eq. (5.17) (as seen in the images of Figs. 5.2(a,c)) look slightly sharper than the ones obtained by processing the whole detected intensity (Figs. 5.2(b,d)) may be interesting for future experiments. We did not study this approach further.

Summary

This chapter has introduced a formalism for the polarization sensitive imaging of EUV mask layouts. We have shown that it is possible to represent the object with a $2x2$ scattering matrix whose entries are the reflection coefficients generated when illuminating the sample with a x/y (or s/p) polarized probe. This respresent a vectorial extension of the scalar ptychographic approach. We have attempted to image the separate entries of the scattering matrix relative to a sample composed of a few scatterers (with dimensions close to the wavelength) located on top of an EUV mask. We managed only to image the diagonal entries of the matrix as cross–polarization effects were not pronounced for this specific sample and wavelength. We believe this method might be appealing for future studies in other optical setups where polarization effects are more marked.

References

[1] P. Ferrand, M. Allain, and V. Chamard, *Ptychography in anisotropic media,* Opt. Lett. **40**, 5144 (2015).

[2] H. P. Urbach, *Lecture notes on advanced photonics* (2021).

6

Future work

Imaging below the diffraction limit is the predominant challenge of imaging physicists. Such a task requires to retrieve information of the spatial frequencies of the permittivity function that correspond to the evanescent part of the angular spectrum of the scattered field and which therefore do not reach the detector. Hence imaging below the diffraction limit requires a set-up by means of which the information contained in the high spatial frequencies of the permittivity is coupled to the propagating part of the angular spectrum of the scattered field, more precisely to the part that is inside the NA of the detector so that these propagating waves can reach the detector.
Armed with a few of the concepts we presented in this thesis and relying on the work of other research groups we shall propose a (tentative) method for superresolved and wide field–of–view imaging.

Total internal reflection phase retrieval and ptychography.

The concepts that we are about to discuss should in principle be applicable to both single intensity phase retrieval methods (HIO/RAAR ets.) and ptychography. The interest in total internal reflection microscopy stems from the fact that in a conventional far–field scattering experiment the optical contrast is mapped over a sphere (Ewald sphere) in the reciprocal space of radius dictated by the magnitude of the free–space wavenumber, $k_0 = 2\pi/\lambda$. The fact that we cannot directly access the region in the reciprocal space outside the Ewald limiting sphere implies that the resolution of an optical instrument is generally limited to about $\frac{\lambda}{2}$. However, an evanescent field is characterized by the fact that the transverse component of its wave vector is larger than the free–space wavenumber. This means that the accessible region of the reciprocal space of the object increases and includes a usually unexplored part of the Fourier space of the object [1]. In what follows we discuss a possible explanation for the physical mechanism behind the propagation of subwavelength information of the object to the far zone of a total internal reflection microscope. Suppose that we are dealing with the geometry in Fig. 6.1.

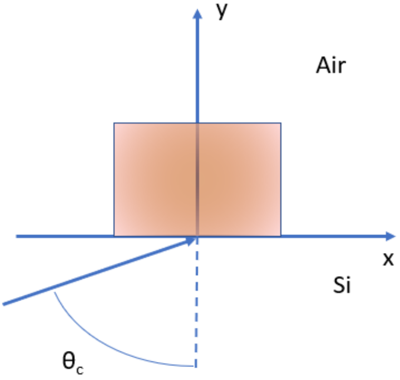

Figure 6.1: Geometry of the experiment: a light field is propagating from Si to Air and hits the interface at an angle $\theta = \theta_c$ therefore creating an evanescent field in Air (not illustrated). A scatterer (block in light red) interacts with evanescent illuminating near field thereby transferring evanescent information into propagating orders.

when the light, coming from a denser medium (Si) to a less dense medium (Air), hits the interface at an angle $\theta \geq \theta_c$ an evanescent wave is excited in the medium 1 (Air). The presence of the scatterer (block coloured in light red in Fig. 6.1) converts the confined evanescent field into propagating orders which hit a detector in the far field. The sub–wavelength imaging performance of TIR microscopes may be attributed to the propagation of subwavelengths details of the object in the excited propagating diffraction orders. To try to understand how this may be happening we consider the simplistic case in which we wish to image an amplitude grating located at $z = 0$. Suppose that this grating has such a tiny period that the kick of momentum due to diffraction is such that – in the "regular" case in which the incoming field, propagating from Si to Air in Fig. 6.1, is incident at an angle $\theta < \theta_c$ – even the first diffraction order would be immediately evanescent. In other words the spatial information of the grating is so fine that we cannot retain it in the far–field in the usual case in which $\theta < \theta_c$. In the assumption considered above the grating is described by a transmission function:

$$O(x) = 1 + A cos(k_x x) \tag{6.1}$$

where $k_x > k_0 = 2\pi/\lambda_0$. Suppose now, without loss of generality, that the incoming light is a plane wave which hits at $\theta \geq \theta_c$. In this case the x component of the wavevector of the incident field is bigger than k_0: $k_x^i > k_0$. The incident field, assumed to be a scalar, is the field that is present when there is no scatterer, and in medium 1 (Air) it is:

$$P(x,z) = P_0 e^{-\alpha|y|} e^{ik_x^i x}. \tag{6.2}$$

In other words in air the field decays away from the interface towards the direction of increasing positive y at a rate given by α and the complex exponential indicates that a surface wave, which moves parallel to the boundary with speed ω/k_x^i, is excited. We do not discuss here the interpretation of this surface wave, for which we refer to [2]. At the plane of the grating ($y = 0$), and assuming that a transmission function approach to the modeling of light–matter interaction is accurate enough, we can evaluate the exit wave just above the grating as:

$$\begin{aligned} P(x)O(x) &= P_0 A e^{ik_x^i x} cos(k_x x) \\ \mathrm{Re}(P(x)O(x)) &= P_0 A cos(k_x^i x) cos(k_x x) = \frac{P_0 A}{2}[cos(k_x^i x - k_x x) + cos(k_x^i x + k_x x)] \end{aligned} \tag{6.3}$$

where with $\mathrm{Re}(z)$ we have indicated the operator that extracts the real part of a given complex number z. When $k_x^i \approx k_x$ the first part of Eq. (6.3) is a low frequency harmonic that can propagate because $|k_x^i - k_x| < k_0$ and that indirectly conveys information about the subwavelength details of the grating. In essence Eq. (6.3) states that we are down–converting high spatial frequencies of the object to base–band so that they can propagate and consequently be detected [3]. According to Eq. (6.3) the physical mechanism that dictates the propagation of subwavelength information of the object from the near to the far zone is the intermodulation among the superoscillatory frequencies of the object and of the illumination. Although Eq. (6.3) has been derived by considering the case of the imaging of a grating, similar observations hold for a generic a–periodic object, which can be expressed in the reciprocal space by a sum of harmonics some of which, containing subwavelength informations, are superoscillatory.
Notice that:

- The probe to be used in ptychography would be an evanescent field that should be retrieved with a near–field probe whose presence would alter the evanescent field itself. Because ptychographic methods allow us to the retrieve accurately the illumination this should not be an issue.

- The objects we consider in phase retrieval are usually comprised by a continuity of spectral components. To detect higher and higher spatial frequencies one could, according to Eq. (6.3), increase k_x^i further by hitting the target at even more grazing incidence from the Si side of Fig. 6.1. In an alternative approach one could interfere two beams that come at opposite directions. As the interference of two sources oscillates as $cos(k_1 x - k_2 x)$ if we pick $k_2 = -k_1$ we would generate a field that oscillates twice as fast. This would further cut the resolution in half. This is the so called standing wave TIR microscopy.

- One could generalize what has been presented here to three dimensional imaging via evanescent ptycho–tomography with the approach in Fig. 6.1 or by means of alternative experimental arrangements (for instance an Otto configuration where an evanescent tail is excited in the neighbourhood of a glass prism [1, 4]).

- The method presented here, if successful, could be employed to other type of applications, for instance for the imaging of cells [5] and can be possibly improved with the use of tailored algorithms similarly to what has been done in this thesis.

In conclusion the idea of total internal reflection phase retrieval and ptychography can possibly provide a framework for the super–resolved and wide–field–of–view lensless imaging of semiconductor or fluorescent samples [6].

References

[1] D. G. Fischer, P. S. Carney, and J. C. Schotland, *Total internal reflection tomography (tirt) for three-dimensional sub-wavelength imaging,* in *Signal Recovery and Synthesis* (Optical Society of America, 2001) p. STuC3.

[2] M. Milosevic, *On the nature of the evanescent wave,* Applied Spectroscopy **67**, 126 (2013), pMID: 23622429, https://doi.org/10.1366/12-06707 .

[3] Z. Liu, S. Durant, H. Lee, Y. Pikus, N. Fang, Y. Xiong, C. Sun, and X. Zhang, *Far-field optical superlens,* Nano letters **7**, 403 (2007).

[4] M. Holler, M. Guizar-Sicairos, E. H. R. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, and G. Aeppli, *High-resolution non-destructive three-dimensional imaging of integrated circuits,* Nature **543**, 402 (2017).

[5] K. Kim, J. Yoon, S. Shin, S. Lee, S.-A. Yang, and Y. Park, *Optical diffraction tomography techniques for the study of cell pathophysiology,* arXiv: Biological Physics (2016).

[6] A. Sentenac, K. Belkebir, H. Giovannini, and P. Chaumet, *Subdiffraction resolution in total internal reflection fluorescence microscopy with a grating substrate,* Optics letters **33**, 255 (2008).

7

Conclusion

The research work outlined in this thesis has been concerned with the development of methods for metrology and inspection of extreme ultraviolet lithographic masks. In Chapter 1 we have outlined the fundamental theoretical and mathematical concepts which are at the core of inverse problems. Moving from these generic foundations we have focused on the inverse problems of scatterometry (Chapter 2) and lensless imaging (Chapters 3 to 5). The research performed in Chapter (2) aimed at finding a feasible solution to the problem of the curse of dimensionality – which originates when too many unknowns are to be retrieved in the inverse problem – in scatterometry. We noticed that the often employed nonlinear least squares routines solve the inverse problem by solving recursively a linear regression problem. This led to the introduction of an optimization algorithm which is a nonlinear extension of a variable selection method known as the "*elastic net*" which is a widely employed method for model simplification in linear regression problems. The application of the novel algorithm to isolated and periodic scattering potentials on an EUV mask revealed that the method converges fast and successfully yields a reduced model that can be subsequently employed during the 3–D profile reconstruction phase. We have compared the outcome of the automatic feature selection algorithm with the results given by one–at–a–time and screening (Morris design) sensitivity analysis methods. A few observations were made that highlighted the differences in the fundamental approach of these methods.

In Chapter (3) we have introduced the phase retrieval problem and the most commonly employed numerical methods that have been proposed to approach the problem. Motivated by the technological need and scientific interest for actinic microscopes we have discussed the application of a few phase retrieval methods for the imaging of isolated and non–trivially shaped nanostructures. In Section (3.2.1) we have shown results for

- the imaging of a single nanostructure via the use of single–intensity phase retrieval algorithms, and

- the imaging of an ensemble of nanostructures using the regular ptychographic–iterative–engine (PIE).

We subsequently shifted the focus of our studies on ptychography. We considered that the only physical assumption in ptychography is that the light–matter interaction can be modeled via a transmission function approach: the exit wave – the field just above the diffracting structures – is assumed to be given by the product among the illumination function (the probe, which we assumed known) and the transmission (or reflection) function of the object which accounts for the way the object interacts with light. This is an approximation of light–matter interaction which we may interpret as a first–order Born approximation. Considering that

- the phase problem is a non–convex optimization problem for which there are no theoretical guarantees on the optimality of the solution, and that
- the ptychographic approach models the light–matter interaction approximately

one could think that a better solution of the phase problem may be achieved by including further and *a–priori* physically sound information in ptychography. This idea was further expanded in Chapter (4) where it was demonstrated that the inclusion of prior information via the use of rigorous Maxwell solvers simulation tools can improve the reconstruction of the object. Specifically, it was shown by means of comprehensive computational studies that the inclusion of a robust and physics–informed prior in the inversion algorithm provides the means for the ptychographic imaging and inspection of smaller and subwavelength defects in a defective EUV absorber layout on top of a Bragg multilayer. The use of variational and statistical methods for the derivation of a suitable update rule to be employed in the optimization scheme was studied and the performance of our methods were discussed in detail. We further presented a comparative study with respect to two examples of sparsity priors, the Lasso and the total–variation regression, that are commonly used in computational methods and that were found to yield a less satisfactory reconstruction than the one obtained with our *physics–informed* prior. The method we deviced improved the inspection of the targets and enabled the imaging of smaller defects (intrusion/extrusions on the patterned absorber). Nonetheless the finest defects, smaller then $\lambda/2$ in size, were not retained in the final retrieved object. The need for a more robust reconstruction and a finer resolution in the retrieved object has driven the further thinking that we presented in the last two chapters.

In Chapter (5) we have presented a vectorial approach for the ptychographic imaging of semiconductor structures (patterned absorber of an EUV mask in this case). This approach can be interesting for polarization sensitive imaging of the considered features when we wish to image the different entries of the scattering matrix in order to verify if one of them is better resolved than the others or if we are interested in reconstructing all of the scattering matrix in order to retrieve more information about the scattering object. We have presented results for fine structures with critical dimensions of the order of the wavelength. We managed to reconstruct the diagonal entries of the scattering matrix but the off–diagonal were too weakly scattering to be imaged and consequently the whole scattering matrix could not be retrieved. The two diagonal entries of the scattering matrix were found to yield, in our case, a similar resolution in the final retrieved image.

This thesis closes with a proposal for a method for superresolved lensless imaging of semiconductor structures: total internal reflection phase retrieval and ptychography. The approach is discussed and the underlying physics is presented. We have reasoned that the superresolved imaging performance associated to total internal reflection microscopes and observed in literature might be understood in terms of intermodulation: superresolution is argued to be given by propagating orders that are the result of a downconversion of high frequency information of the object. This process has also been observation is in line with the findings of different groups

This thesis summarizes the research work done in the past four years and introduces novel concepts, ideas and algorithms that can potentially enable tighter (i.e. better resolved) metrology and inspection of EUV masks and semiconductor structures. The results and the methods presented in this thesis can be particularly interesting for the communities that work in the area of metrology and inspection of target features for the monitoring of the quality of the lithographic patterns at the mask level or wafer level.

Acknowledgements

I joined the Optics group at Delft in what now feels as a somewhat distant 2017. I was moving to a country which was widely unknown to me, predominantly motivated by a sense of curiosity and the desire to see more and experience more than I had already seen and experienced. I still recall the first day at TUD, the warm welcome from Yvonne and the first meeting with Paul where we discussed what courses I could have followed to strengthen my background and tailor my studies towards the topics I was about to step into. The first meeting with Wim happened a week later and there, for the first time, I heard the term *ptychography* and had the first discussion about the fundamental nature of the resolution limit. The same topics would drive this research project in the following four years.

I wish to express my gratitude towards the people that have contributed, in a way or the other, to make the PhD years an engaging and positive time of my life. I like, at first, to address my mentors. Paul, although you were often busy you always welcomed questions and encouraged open scientific discussions among us and also among students. It was always interesting for me to see how a mathematician interprets and understands physics and how deep the connection among these two worlds is. Your deep knowledge and openness are an example for me. Wim, we have spent 4 years discussing in 1 to 1s at TUD first and ASML later. I have learned a lot from you, from both the scientific and the human side. Besides the breadth of your knowledge and the technical discussions your warm and kind personality has provided a firm example of positive leadership and a source of inspiration.

A thank you note to the Optics group staff: Silvania, Aurèle, Omar and Lauryna for involving me in their own courses and for the few discussions that we had during the time I was at TUD. Yvonne, for assisting me numerous times during the PhD years and Roland and Thim for making sure I got out of trouble when something happened at my laptop.

I like to acknowledge Toralf Scharf as the organizer of the NOLOSS project and the ESRs involved, together with the network of PhDs at the Optics group.

My gratitude goes also towards the CMMS group at ASML for making me feel as a part of the team while I was there. The advanced computations performed in this thesis were only possible thanks to the HPC cluster and the sophisticated Maxwell solver which I was allowed to use. I acknowledge ASML for granting me access to these resources.

I want to thank my long–standing friends in Italy and the more recent ones that I met in the Netherlands for their support and for the numerous nights out which helped me to unwind and relax from the hectic PhD period.

Lastly, I like to thank my family and especially my sister and my parents for their unconditional love, continuous support and presence throughout my entire life. You are the greatest blessing of my life.

List of Publications

Journal Publications

6. **P. Ansuinelli**, W. M.J.M. Coene, and H.P. Urbach, Appl. Opt. 59, 5937-5947 (2020). *Improved ptychographic inspection of EUV reticles via inclusion of prior information*, Applied Optics **59**, 5937–5947 (2020).
5. **P. Ansuinelli**, W.M.J.M. Coene, and H.P. Urbach, *Automatic feature selection in EUV scatterometry*, Applied Optics **58**, 5916–5923 (2019).
4. **P. Ansuinelli**, A.G. Schuchinsky, F. Frezza and M.B. Steer, *Passive Intermodulation Due to Conductor Surface Roughness*, IEEE Transactions on Microwave Theory and Techniques **66**, 688–699 (2019).

Conference Proceedings

3. **P. Ansuinelli**, W.M.J.M. Coene, H.P. Urbach, *Refinement strategies for optimal inclusion of prior information in ptychography*, Proc. SPIE , Extreme Ultraviolet Lithography 2020, **11517**, 1151711 (2020).
2. **P. Ansuinelli**, W.M.J.M. Coene, and H.P. Urbach, *EUV mask feature reconstruction via phase retrieval*, Proc. SPIE, Nanoengineering: Fabrication, Properties, Optics, Thin Films, and Devices XVI, **11089**, 110892F (2019).
1. **P. Ansuinelli**, F. Frezza, A.G. Schuchinsky, *PIM generation by rough conductors*, International Workshop on Electromagnetics: Applications and Student Innovation Competition, 44–46 (2017).