Model order reduction for the time-harmonic Maxwell equation applied to complex nanostructures
Model order reduction for the time-harmonic Maxwell equation applied to complex nanostructures

Martin Hammerschmidt\textsuperscript{a}, Sven Herrmann\textsuperscript{a}, Jan Pomplun\textsuperscript{b}, Sven Burger\textsuperscript{a,b}, Frank Schmidt\textsuperscript{a,b}

\textsuperscript{a}Zuse Institute Berlin, Takustraße 7, 14195 Berlin, Germany
\textsuperscript{b}JCMwave GmbH, Bolivarallee 22, 14050 Berlin, Germany

This paper is made available as an electronic preprint with permission of SPIE and will be published in Proc. SPIE 9742 (2016). The full citation reads:
Copyright 2016 Society of Photo Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic electronic or print reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

ABSTRACT

Fields such as optical metrology and computational lithography require fast and efficient methods for solving the time-harmonic Maxwell’s equation. Highly accurate geometrical modeling and numerical accuracy at low computational costs are a prerequisite for any simulation study of complex nano-structured photonic devices. We present a reduced basis method (RBM) for the time-harmonic electromagnetic scattering problem based on the $hp$-adaptive finite element solver JCMsuite capable of handling geometric and non-geometric parameter dependencies allowing for online evaluations in milliseconds. We apply the RBM to compute light-scattering at optical wavelengths off periodic arrays of fin field-effect transistors (FinFETs) where geometrical properties such as the width and height of the fin and gate can vary in a large range.

Keywords: finite element method, rigorous optical modeling, reduced basis method, reduced order models, electromagnetic field solver, fin field-effect transistors

1. INTRODUCTION AND PROBLEM SETTING

In many fields of current research and development involving complex nano-structured or photonic devices, the demand for accurate and fast solutions of the electromagnetic scattering problem arises. Fields such as optical critical dimension metrology or computational lithography rely on efficient methods to solve the time-harmonic Maxwell’s equation. The feature size and complexity of the structures makes the computations of solutions of the Maxwell’s equations challenging as the required highly-accurate geometrical modeling usually leads to a large number of degrees of freedom. This limits either the numerical accuracy or the total number of computations to be executed. The finite element method (FEM)\textsuperscript{3} is ideally suited to resolve structures with large discrepancies in feature sizes as it allows for various geometrical elements to be combined into a single mesh. Furthermore, the use of locally adapted polynomial degrees allows to tailor the required accuracy locally taking source and material parameters into account. This strategy is known as $hp$-adaptivity.\textsuperscript{2}

In metrology applications geometrical parameters are identified via solution of an inverse problem. A fabricated structure is illuminated from multiple directions and the reflectance is recorded. Subsequently a parameterized numerical model of the optical setup is simulated to identify model parameters by comparing model output and measurement. The estimation of the model parameters through this inverse process is tedious and requires easily thousands of electromagnetic field simulations which have to be executed with high accuracies. In

\textsuperscript{3}Further author information: (Send correspondence to Martin Hammerschmidt)
Martin Hammerschmidt: E-mail: hammerschmidt@zib.de, Telephone: +49 30 84185-149
In this context reduced order models can be used to speed up computations drastically, provided they approximate the solution of the full problem accurately enough.

In this work we present a reduced basis method (RBM) for geometrical dimensions of an array of fin field-effect transistors (FinFETs) as an example for a model order reduction strategy. We briefly review the fundamentals of the underlying finite element discretization and the reduced basis methods in Sections 2 and 3. In Section 4 we present a case study of the FinFET as a model order reduction (MOR) method for complex nano-structures.

2. FINITE ELEMENT METHOD

In this section we briefly review some important aspects and the key ideas of the finite element method. For details we refer to the extensive literature on the subject.¹–³ We introduce notation and concepts to be used in the section dedicated to the reduced basis method.

The scattering of monochromatic light by nano-structures is governed by the linear Maxwell’s equations. This set of equations can be rewritten to a single, second order curl-curl equation for the electric field $E$

$$\nabla \times \mu^{-1} \nabla \times E - \omega^2 \varepsilon E = i \omega j. \quad (1)$$

with the permeability and permittivity tensors $\mu$ and $\varepsilon$. $\omega$ is the frequency of the time-harmonic field and the electric current density $j$ models electromagnetic sources within the computational domain $\Omega$. The exterior $\Omega_{\text{ext}} = \mathbb{R}^3 \setminus \Omega$ hosts incoming electric fields which act as sources in the interior $\Omega$. The perfectly matched layer method (PML) is used as a transparent boundary condition at the boundary $\Gamma$ in our solver.⁴

A variational formulation of (1) is the basis for the finite element discretization. It is obtained by multiplying (1) with a vector valued test function $\varphi$ and subsequent integration over $\mathbb{R}^3$. This gives

$$\int_{\mathbb{R}^3} \nabla \times \varphi \mu^{-1} \nabla \times E - \omega^2 \varphi \varepsilon E = i \omega \int_{\mathbb{R}^3} \varphi j, \quad (2)$$

which we abbreviate as $a(\varphi, E)$ with the right hand side $f(\varphi)$ in the following. As the test function $\varphi$ is arbitrary, we demand (2) to hold for any $\varphi$. $H(\text{curl}, \Omega)$ is the proper (infinite) function space¹,² for $E$ and $\varphi$.

The weak formulation of (1) finally reads: Find $E \in H(\text{curl}, \Omega)$ such that:

$$a(\varphi, E) = f(\varphi) \quad \forall \varphi \in H(\text{curl}, \Omega). \quad (3)$$

A numerical solution of (3) can be obtained by restricting the solution space to a finite dimensional subspace $V_h \subset H(\text{curl}, \Omega)$ with $\dim V_h = N < \infty$ where it is sufficient to demand (3) to hold for a basis $\{\varphi_1, \ldots, \varphi_N\}$ of $V_h$. This allows to expand every $E = \sum_{i=1}^{N} \alpha_i \varphi_i$ of $V_h$ in this basis and to restate (3) in $V_h$ as a linear system

$$\sum_{i=1}^{N} a(\varphi_j, \varphi_i) \alpha_i = f(\varphi_j) \quad j = 1, \ldots, N. \quad (4)$$

The structure of the matrix $A = (a(\varphi_j, \varphi_i))_{i,j=1}^{N}$ is ideally sparse, allowing to compute the coefficients $\alpha_i$ with an efficient LU decomposition. Ansatz functions with a local support yield the desired sparsity of $A$.

The general idea of the finite element discretization is to split the computational domain $\Omega$ into small geometrical patches or elements of simple form such as tetrahedrons, prisms or bricks. Polynomials of order $p$ over the patches are chosen as ansatz functions. The order $p$ does not have to be identical on all patches. It can be adapted to the chosen spatial resolution and can be lowered in regions where the spatial resolution is high and increased where it is low. This adaptation yields the $hp$ finite element method⁵ which allows to compute very accurate results of (3) at reasonable computational effort.
3. MODEL ORDER REDUCTION WITH THE REDUCED BASIS METHOD

Model order reduction (MOR) is a class of numerical strategies to reduce computational complexity in mathematical models described by partial differential equations. It is employed to make simulations of large-scale and dynamic systems feasible. Most methods in this class construct surrogate models which can be evaluated in significantly reduced time by trading in complexity for numerical accuracy. The most widely used methods comprises proper orthogonal decomposition\textsuperscript{6–9} for dynamic systems and the reduced basis method (RBM)\textsuperscript{10–14} for stationary problems. In the following we focus on the RBM for the electromagnetic scattering problem (3).

The RBM allows to construct error controlled approximations to input-output relationships. We call a specific configuration of parameters $\mu \in \mathbb{R}^d$ the input. An output quantity $s(\mu) = s(E(\mu))$, derived from a solution $E$ of the parameter dependent problem (3), corresponds to the input. The parameterized electromagnetic scattering problems, with parameters $\mu$ describing geometry, material variations or properties of the illumination, can be solved for every $\mu$ by discretizing (3) with a finite element discretization with $N$ degrees of freedom and subsequent computation of the output of interests, e.g. the Fourier transform of the electric field or the electric field energy contained within a subdomain of $\Omega$. Such a solution is typically referred to as a truth solution. Their computation is quickly rendered infeasible if the number of parameter configurations considered is large or if there is a limit on runtime.

The purpose of the reduced basis approximation is to provide a low-dimensional surrogate model to the full problem (3) of high dimension for all parameters $\mu \in \mathcal{D}$ in a parameter space $\mathcal{D} \subseteq \mathbb{R}^d$. The manifold of all solutions $\mathcal{M} = \{E(\mu) \mid \mu \in \mathcal{D}\}$ is assumed to be well approximated by a low dimensional subspace $X_N$ of dimension $N$. Like in the finite element method presented above, we replace the high-dimensional (oftentimes dim $\mathcal{V}_h = N > 10^6$) solution space by a very low dimensional (usually of dimension $N = O(10)$) reduced basis space $X_N$ and find solutions to the parameter dependent PDE in it. Hence we solve the reduced problem: For $\mu \in \mathcal{D}$ find $E_N \in X_N$ such that:

$$a(\varphi, E_N; \mu) = f(\varphi; \mu) \quad \forall \varphi \in X_N.$$  \hfill (5)

A common choice for $X_N$ is to find $N$ linearly independent truth solutions to span $X_N$. We call these snapshots and $\Phi = \{\phi_i\}_{i=1}^N$ where $\text{span} \Phi = \text{span}\{E(\mu_i) \mid E(\mu_i) \text{ is a truth solution for } \mu_i, i = 1, \ldots, N\}$ a reduced basis.

3.1 Offline-online decomposition

The efficiency of the computation hinges on a decomposition into two phases where the numerical operations involving the number of degrees of freedom $N$ of the underlying finite element solution are executed in the first phase. Subsequently, for every parameter configuration only very low dimensional systems have to be assembled and solved. We call this the online phase and the first phase the offline step. The decomposition works as follows:

A solution $E(\mu) = \sum_{i=1}^N \alpha_i \phi_i$, expanded in the reduced basis $\Phi$, is entered into (5). The resulting linear system

$$\sum_{i=1}^N \alpha_i a(\phi_j, \phi_i; \mu) = f(\varphi_j; \mu) \quad j = 1, \ldots, N$$  \hfill (6)

depends on $\mu$, but is only of dimension $N$. Given the sesquilinear form $a(\cdot, \cdot)$ and the linear form $f(\cdot)$ are affine in the parameter, i.e. $a(\phi, \xi; \mu) = \sum_{q=1}^Q \theta_q(\mu) a_q(\phi, \xi)$ and $f(\varphi; \mu) = \sum_{q=1}^Q \theta_q(\mu) f_q(\phi)$, assembly and solution of (6) only depend on $N$ and $Q$ - not on $N$. The $N$-dependent matrices $(A^q)_{i,j} = a_q(\phi_i, \phi_j)$ and vectors $(f^q)_i = f_q(\phi_i)$ can be precomputed. Unfortunately, most parameterizations of the electromagnetic scattering problem are not affine in the parameter. However, the empirical interpolation method\textsuperscript{15,16} allows to construct an affine approximation to it and has been successfully applied to a number of applications\textsuperscript{17,18}.

In summary: During the construction or offline phase $N$ truth solutions are computed to span $X_N$ and the matrices $A^q$ and vectors $f^q$ are computed. In the $N$-independent online phase for a given $\mu$ the system (6) is assembled and solved.
3.2 Basis construction

The quality of the reduced approximation depends on the choice of truth solutions employed to span $X_N$ and naturally, we want to choose $X_N$ optimally. This is not possible a priori. Building $X_N$ self-adaptively allows to place snapshots optimally, if the approximation error over $\mathcal{D}$ is estimated and controlled appropriately. This yields a sequence of reduced basis spaces $X_N$ which contain in each step the maximal available information about the manifold of all solutions $\mathcal{M}$ for any given $N$. The optimization problem to place the next snapshot is solved with a so-called Greedy selection of the worst resolved parameter location within a (finite dimensional) training set $\mathcal{D}_{\text{train}} \subset \mathcal{D}$. This requires rigorous a posteriori error estimators to assess the approximation error. These are available for the electromagnetic scattering problem and usually based on the dual norm of the residuum.$^{12,13}$

4. MOR FOR COMPLEX NANOSTRUCTURES - A FINFET CASE STUDY

In the section we present the construction and evaluation results of a reduced model for a complex nanostructure. We present the optical model in detail first, comment on the construction of the reduced model, before presenting evaluation results of the reduced model and an error analysis of the reduced model. The structure under investigation has been studied extensively before. We previously demonstrated the efficacy of the $hp$-finite element method$^5$ for this structure and presented a reduced basis for the corner rounding radii.$^{19}$

4.1 Optical model

The nanostructure under investigation is a periodic array of fin field-effect transistors at the 22 nm technology node. The down-scaling of integrated circuits has lead to the development of these small but complex nanostructured 3D architectures.$^{20}$ In Figure 1 (left) a part of the unit cell is depicted as a FEM mesh used for the computations. The main structural components (fin, gate and substrate) are noted as well as the different materials used. The device dimensions$^{20}$ and the material data$^{21}$ are taken from the literature. The fin and gate have pitches of 88 nm and 44 nm respectively. The width and height of the fin are 12.7 nm and 40 nm with
Figure 2: FEM-meshes of the unit cell with varied parameters *height* and *width* of gate and fin respectively to its most extreme values. The arrow-crosses depict the dimension of the initial FEM-mesh as in Figure 1 for comparison.

A side-wall angle of 89.5 deg. It has an undercut of 2.1 nm. The width and height of the gate are 40 nm and 95 nm with an undercut of 2 nm. The remaining layers are visible in Figure 1 as well: SiN (dark red) with 5 nm thickness, a high-k-layer (green) of 2 nm and a TiN layer of 7 nm (gray). The FinFET geometry is discretized with a parameterized mesh using tetrahedral elements which allows to accurately resolve all geometrical features such as the undercuts and non-rectangular side walls. Perfectly matched layers are used to model the transparent boundaries at the top and bottom while periodic boundary conditions are employed in *x* and *y* directions.

The illumination is modeled by a p-polarized plane wave at 390 nm, tilted at 30 deg along the gate direction. The FEM discretization, employing 3rd order ansatz functions, has 181 377 degrees of freedom and is solved in ≈ 1 min of cpu time. The electric intensity of the resulting field is shown as a cross section in Figure 1 on the right.

### 4.2 Reduced basis construction

The reduced model is build for the parameter domain \( \mathcal{D} = [91 \text{ nm}, 99 \text{ nm}] \times [36 \text{ nm}, 44 \text{ nm}] \times [36 \text{ nm}, 44 \text{ nm}] \times [11.7 \text{ nm}, 14.7 \text{ nm}] \) where \( \mu = (h_{\text{gate}}, w_{\text{gate}}, h_{\text{fin}}, w_{\text{fin}}) \in \mathcal{D} \). In Figure 2 FEM-meshes of the unit cell are shown where the indicated parameters are set to their respective minima and maxima while the others are kept at the values noted above. As a guide to the eye, the dimensions of the reference geometry are indicated in each plot. The reduced model is constructed for the Fourier transform of the electromagnetic field in +z direction. The Fourier transform is than used to compute the reflectance of the nanostructure.

The Greedy searches employ an adaptive refinement strategy of the four dimensional parameter space \( \mathcal{D} \subset \mathbb{R}^4 \). An initial training set \( \mathcal{D}_{\text{train}} \) of dimension 64 (3 parameter values in each parameter dimension) is used. The EIM approximation of the right hand side is trivial, whereas the approximation of the system matrix requires 75 snapshots to reach an estimated error of less than \( 5.1 \cdot 10^{-6} \). This comparatively large number reflects the large deformation the FEM mesh undergoes when displaced to the corners of the tensorial parameter space.
Figure 3: Error estimates with increasing reduced basis dimension in a semi-logarithmic plot. The residual error estimate (black solid line) is normalized with the estimate for \( N = 1 \). The error estimate for the Fourier transform (black dashed line) estimates the relative error in the \( L_\infty \) norm. The mean (blue dashed line) and maximum (red dashed line) of the relative errors in this quantity of a set \( \Xi \) of random parameter locations show the same trend.

In the offline phase 38 snapshots are computed to span the reduced basis space. The sequence of snapshots is determined by the Greedy algorithm which in every iteration chooses the worst resolved parameter in the training set to be the next snapshot location. We choose the first snapshot in the center of the parameter domain \( (\mu_1 = (95, 40, 40, 12.7)) \) and observe that the subsequent snapshots are chosen at the corners of the parameter domain. The 29th snapshot location \( (\mu_{29} = (93, 42, 42, 13.95)) \) is the first snapshot not located at one of the corners. This is a very typical behavior of the Greedy algorithm and could be prohibited by not allowing snapshot computation on the boundaries of the parameter domain.

The adaptive refinement of the parameter domain and self-adaptive construction of the reduced model leads to a non-monotonic reduction in the error estimate. This can be observed in Figure 3 where the (normalized) estimated error of the reduced basis approximation and the Fourier transform output quantity are depicted as solid and dashed lines over the reduced basis dimension \( N \). With increasing \( N \) the estimated errors are reduced exponentially in this semi-logarithmic plot. The estimated error converges exponentially up to \( 3.5 \cdot 10^{-6} \). The estimated error of the derived quantity is simultaneously reduced up to \( 2.5 \cdot 10^{-5} \). The different slopes of the curves are due to the different error norms involved (relative \( L_\infty \) norm and \( H(\text{curl}, \Omega) \) norm).

4.3 Reduced basis online evaluation

The reduced model allows for a fast and efficient evaluation in milliseconds in the online phase. We evaluate the model on \( 41^3 \cdot 31 \) points in each dimension of \( \Omega \) equidistant and with a duration of \( \approx 130 \) ms per evaluation. In Figure 4 evaluations using 11 (a) and 36 snapshots (b) are depicted as contour plots of the reflectance in cross-sections through the center of the parameter domain. As evident from these plots, we have a only slight deviations in the results of the output quantity even at just 11 snapshots. The most prominent difference is a feature visible in the lower left plot in Figure 4b which is not present in the preceding plot.
Figure 4: Reflectance over the parameter domain as projection onto the center of two dimensions of the parameter domain each for $N = 11$ (a) and $N = 36$ (b) snapshots.
4.4 Error analysis

In most applications for a reduced basis approximation the error in the output quantity is of more interest than the error of the approximation itself. In the following we investigate these different errors by comparing the reduced basis approximations to reference solutions for a randomly selected set $\Xi \subset D$ of 100 parameter values. For $\mu \in \Xi$ we solve the truth approximation including the empirical interpolation approximation and compare the results to the reduced model.

**Field solution**  The max and mean of of the approximation error over $\Xi$ in the $H(\mathbf{curl}, \Omega)$ norm are shown in Figure 5 as well as a scatter plot relating the estimated error and the error measured in the $H(\mathbf{curl}, \Omega)$ norm. Maximum and mean error exhibit identical convergence trends to the estimated error (note that the error estimate is normalized whereas the actual error is not). With increasing reduced basis dimension error decreases exponentially from $1 \cdot 10^{-1}$ to $1.3 \cdot 10^{-7}$ at most. The mean is about one order of magnitude smaller. The good performance of the error estimator is obvious from the scatter plot. We find that the error and its estimate are highly correlated ($\rho > .97$) over several orders of magnitude.

**Quantity of interest**  In Figure 3 the relative errors in the Fourier transform over $\Xi$ are shown as well as the error estimate (black dashed line). The mean relative error over $\Xi$ is depicted by the blue dashed linear whereas the maximum is shown in red. Both lines follow the error estimate very closely. The errors are reduced from about $2.6 \cdot 10^{-1}$ to $2.3 \cdot 10^{-5}$ with increasing reduced basis dimension.

5. CONCLUSION

The reduced basis method is well suited to produce reduced order models for the time-harmonic electromagnetic scattering problem to be employed in many-query or real-time simulation tasks. Although the required algorithmic preparations are relatively demanding, i.e. the discrete Maxwell operator has to be parameterized in a way that permits effective offline-online decompositions, the self-adaptive reduced basis construction by the Greedy selection strategy allows to build low dimensional approximations with very low approximation errors. In this work, we demonstrated an efficient, error controlled, real-time capable procedure for structures with nanoscale features such as FinFET. We demonstrated, that the errors in both the field solution and the output are satisfyingly controlled and significant speedups in the online computation can be achieved.
ACKNOWLEDGMENTS

The results were obtained at the Berlin Joint Lab for Optical Simulations for Energy Research (BerOSE) of Helmholtz-Zentrum Berlin für Materialien und Energie, Zuse Institute Berlin and Freie Universität Berlin. This research was carried out in the framework of Matheon supported by Einstein Foundation Berlin through ECMath within subprojects SE6 and OT5.

REFERENCES


