Fast rigorous simulation of mask diffraction using the waveguide method with parallelized decomposition technique

Feng Shao, Peter Evanschitzky, David Reibold, and Andreas Erdmann
Fraunhofer Institute of Integrated Systems and Device Technology (Fraunhofer IISB)
Schottkystrasse 10, 91058 Erlangen, Germany
feng.shao@iisb.fraunhofer.de, peter.evanschitzky@iisb.fraunhofer.de
david.reibold@iisb.fraunhofer.de, andreas.erdmann@iisb.fraunhofer.de

ABSTRACT

A new and optimized electromagnetic field (EMF) solver based on the waveguide method with a decomposition technique for rigorous optical and extreme ultraviolet (EUV) mask near field simulations is presented. The implemented software algorithm enables full three dimensional (full 3D) mask simulations as well as three dimensional mask simulations based on a parallelized decomposition technique (Q3D, "Q" stands for "quasi"). After a short introduction to the waveguide method and to an optimized mask description, the basis of the decomposition technique and its parallelization are presented. Subsequently the capabilities of the new electromagnetic field solver are demonstrated by simulations of advanced optical and EUV imaging problems. Simulations of larger sized mask areas and of standard sized defective EUV mask areas using the decomposition technique are shown. Finally, a further reduction of computation time using parallelization is demonstrated.

Keywords: Waveguide EMF solver, fast three dimensional simulation, parallelized decomposition technique

1 INTRODUCTION

As the next generation of lithography technology aims at semiconductor manufacturing for the 32nm node in 2010 and beyond, rigorous simulation of mask diffraction is highly required by lithographers to simulate such small features with the required accuracy. Faster simulation speed is also demanded due to the significant impact of mask topography effects on the optical proximity correction (OPC). To meet these requirements, a fast rigorous electromagnetic field solver for EUV and optical mask simulations based on the waveguide method has been developed at Fraunhofer IISB. Previous studies have demonstrated a very good performance in terms of computation time and convergence for both 3D EUV and 3D optical cases [1]. However, the simulated mask sizes are still restricted to be smaller than desired due to the general complexity of rigorous 3D computations. To be able to simulate larger masks (e.g. 4µm x 4µm at a wavelength of 13.5nm, 10µm x 10µm at a wavelength of 193nm) in an acceptable time, a decomposition technique based on the waveguide method has been developed and extended to a parallel version.
In the next section the basic idea of the waveguide approach with a newly developed mask description method is presented. The basis of the decomposition technique and its parallelization are explained. Section 3 shows several simulation examples to demonstrate the performance of the implemented software. The paper concludes with a summary.

2 SIMULATION MODEL

The simulation of partial coherent projection systems for lithographic imaging consists of two main parts: First, the mask near field or diffraction spectrum has to be computed. In this paper we employ the waveguide method and decomposition techniques for an efficient simulation of mask diffraction spectra. In the second step a vector extension of Fourier optics is used to compute the image formation for a given mask diffraction spectrum. In this section, a short description of the waveguide method with a new mask representation is presented. Afterwards, the decomposition technique and its parallelization are explained. Further details of the waveguide approach, the EUV multilayer model and the imaging system can be found in the paper by Evanschitzky et al. [1].

2.1 Waveguide method

The basic idea of the waveguide method is as the following: The 3D mask geometry is sliced into x-y parallel layers with constant optical properties within the material in the z-direction, which is perpendicular with respect to the mask plane (see Figure 1). Then the material distributions and the electromagnetic fields in each layer are expanded into Fourier series. The Fourier series are truncated by the number M of the modes to be computed for a specified accuracy. In this step a very important convergence optimization has to be included in the computation of the truncated Fourier series for optical masks [2]. In each slice or layer, the application of this Fourier representation of the material properties and of the electromagnetic fields to the Maxwell equations results to an eigenvalue problem. The numerical solution of this eigenvalue problem provides the propagating and evanescent modes inside this slice or layer. After that, all layers are coupled by applying the boundary conditions between each pair of neighbored layers. Finally, the reflected and transmitted plane waves or diffraction orders at the top and bottom side of the mask are computed.
The mask is always assumed to be a periodic structure. To compute totally isolated features, a mask with a sufficiently large period has to be simulated to approximate the non-periodic condition. The described method can be applied to the simulation of two dimensional (e.g. lines) and of three dimensional (e.g. post absorbers) features. Further mathematical details of the waveguide method can be found in the paper by Lucas et al. [3] and a technical description including simulation examples is given in [1]. Further information on other modal methods, which are similar to the waveguide method, can be found in the paper by Moharam et al. [4] and by Turunen et al. [5].
The computation time for the waveguide method is proportional to $A \cdot M^3$. Here A specifies the number of layers consisting of more than one material after the mask slicing. M is the overall number of modes used for the computation and can be expressed by $M=(2N_x+1)\cdot(2N_y+1)$. N_x, N_y specify the number of positive (and negative) diffraction orders in x- and y-direction, respectively. According to all investigations performed so far the following rule of thumb can be established to achieve a sufficient accuracy: $N_{x,y}=b_{x,y}/2\lambda$ for EUV masks and $N_{x,y}=3b_{x,y}/\lambda$ for optical masks ($b_{x,y}$ is the real mask period in x- and y-direction, λ is the illumination wavelength). The difference of the factor in the formulas of orders is due to the different refractive indices of materials used in EUV masks and optical masks.

2.2 New description method of the mask in horizontal direction (mask plane)

In the waveguide method, Fourier transforms of the material permittivity distribution in all layers are required as the input for the eigenvalue problem. For that purpose, the mask geometry in the layer is represented by non-overlapping rectangles, which can be Fourier transformed analytically.

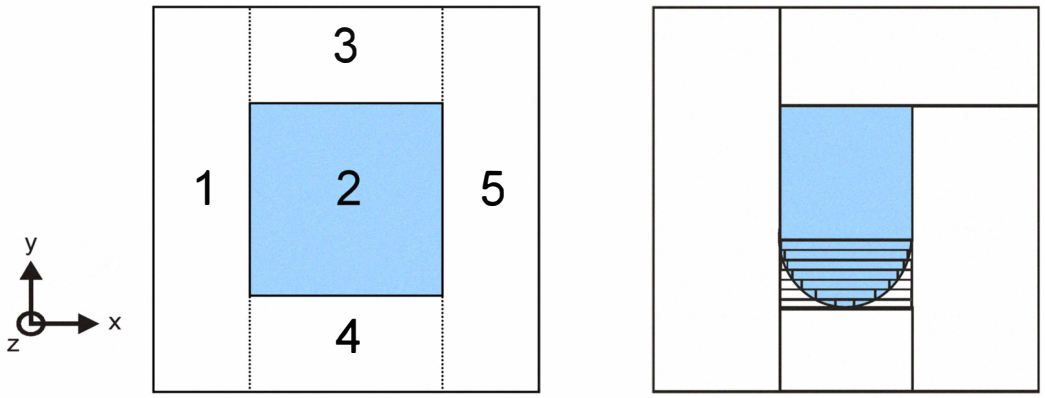

Figure 1: Mask layer geometries and representation by non-overlapped rectangles. Left: square post described by 5 non-overlapped rectangles (labeled 1-5), right: rounded feature described by a series of smaller rectangles.

Figure 1 demonstrates the description of mask layer geometries with two examples. Both of the two mask layers consist of two different materials (in light blue and white color), but of different geometries. The left mask layer consists of a rectangle absorber in the center and is described by 5 non-overlapped rectangles (labeled 1-5). In this case no sampling error occurs because the Fourier series of the material permittivity distribution are computed analytically for the

unchanged structure. The right mask layer has a more complicated geometry. The semicircle is described by a series of smaller rectangles, where a sampling error has to be taken into account related to the minimum size of rectangles. Compared to the previous mask description method based on a typical equidistant FDTD grid discretization (see [1]), the new approach uses only the geometry parameters of (few) rectangles and therefore the memory requirement is much lower. For the lithographically important practical case of mask geometries consisting only of rectangle features (so called Manhattan-type features), this accurate description will cause no sampling error. The improvement resulting from this new mask description in terms of computation time and memory consumption is demonstrated in section 3.1.

2.3 Decomposition technique for the waveguide method

As indicated in section 2.1, the mask simulation time is highly dependent on the overall number of modes M to be computed. As the simulated mask size becomes larger, M also gets larger, and the computation time increases by the order of 6 in the 3D case and by the order of 3 in the 2D case. In practice, the 3D simulation of larger masks (e.g. 4µm x 4µm at a wavelength of 13.5nm, 10µm x 10µm at a wavelength of 193nm) is not feasible in an acceptable time. Therefore, a simplified 3D waveguide computation based on a decomposition technique was developed. This technique replaces the full 3D simulation by a superposition of several independent two dimensional and one dimensional simulations. The result is a significantly reduced computation time. The error resulting from this approach is negligible for feature sizes of 45nm or larger, which has been discussed in more detail in another publication [8]. As explained in section 2.4, such decomposition technique enables a direct and efficient parallelization.
The decomposition technique used in combination with the waveguide method is based on a proposal by Adam et al. [6]. The technique is already implemented and well established at Fraunhofer IISB in combination with the finite-difference time-domain method. A detailed technical description and simulation study can be found in the paper by Adam et al. [6] and by Erdmann et al. [7]. The basic implementation and combination of this decomposition technique with the waveguide method are identical to those with the finite-difference time-domain method. Comparable behavior can be expected for the waveguide method. Figure 2 demonstrates the basic idea of the applied decomposition technique.

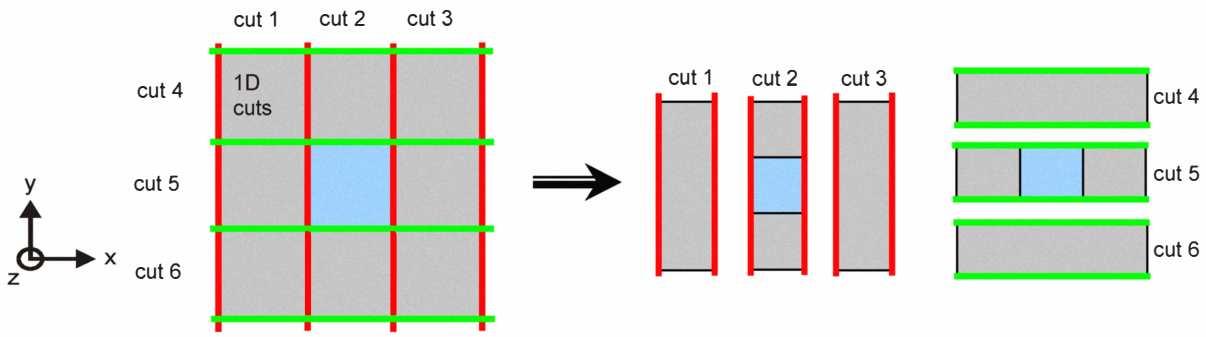

Figure 2: Top view of a 3D mask with a post absorber (light blue) in the center. The 3D mask is decomposed into two dimensional cuts (cut 1-6) with constant dielectric properties in x- and y-direction, respectively. Additional crossed cut areas with constant dielectric properties in both directions (1D cuts) are not displayed here but required in the computation.

The basic idea of the decomposition technique is to separate diffraction effects from mask edges along the x and the y direction. A full 3D mask is split into several cuts along the x and the y direction, which have homogeneous material distribution in direction of the cut. According to the example of Figure 2, three cuts in x-direction (denoted by cut 1, cut 2, cut 3) and 3 cuts in y-direction (denoted by cut 4, cut 5, cut 6) can be found. In the next step the electromagnetic fields in all cuts are computed with the waveguide method under the assumption that the dielectric properties of each cut vary only in y-direction (in case cut 1, cut 2, cut 3) or vary only in x-direction (in case cut 4, cut 5, cut 6). Based on this assumption, the waveguide computation of each cut is degraded from 3D to 2D, which provides a significant speedup of the required numerical simulations. Finally, the transmissions and reflectivities of the mask at all cross-areas of the cuts (denoted 1D cuts) are computed with the waveguide method. The resulting diffraction spectrum of the 3D mask is obtained by a composition of the complex diffraction spectra of the x- and y-configurations and of the mask transmissions and reflectivities of the 1D cuts.

The composition of the partial results is done in the frequency domain. Compared to a space domain composition, the real sizes and positions of all geometrical elements without any discretization or sampling error are taken into account. This provides an additional speedup and accuracy improvement compared to space domain decomposition techniques.

2.4 Parallelization of the decomposition technique

To further speed up the simulation of larger 3D EUV and optical masks, a direct and efficient parallelized version of the decomposition technique has been developed at Fraunhofer IISB. The speedup and efficiency of the parallelization is demonstrated by simulation examples for larger EUV masks, optical masks and standard sized defective EUV masks in sections 3.3 and 3.4, respectively.
As the 2D and 1D computations after decomposition are totally independent from each other, the distribution strategy is quite straightforward. The parallelization is implemented in a master-slave mode. First, the master CPU computes the mask geometries for all 2D and 1D cuts. The geometry description of each single cut is transmitted to a slave CPU as an input of the related 2D or 1D waveguide simulation. Spectra of all the decomposed computations are computed by slave CPUs and transmitted back to the master CPU. Finally, the composition of spectra is performed on the master CPU.

3 SIMULATION EXAMPLES

This section starts with some examples to demonstrate typical computation times and memory requirements of full 3D simulations. After that, mask diffraction and imaging simulations of larger mask areas as well as standard sized defective EUV mask areas are presented. Full 3D, Q3D and parallelized Q3D methods are used. Computation time and memory consumption are compared between different methods. For the parallelized Q3D method, a further reduction of computing time is presented and the parallelization efficiency is shown as well. Cross sections and process windows of defective EUV mask images using full 3D and Q3D approaches are compared. Further detailed comparisons of full 3D and Q3D results in terms of cross sections and process windows can be found in the paper by Evanschitzky et al. [8].
If not mentioned otherwise, all sizes in this section are specified on mask scale. The full 3D approach is the most accurate simulation method available and serves as a reference. All simulations are performed on the IISB HPC cluster: Each cluster node has two AMD Opteron 275 Dual Core CPUs and 8GB memory. Sequential simulations are performed on one CPU of a node, and parallel simulations are performed on many CPUs of several nodes.

3.1 Computation time and memory consumption using new mask description

Figure 3 shows the dependencies of simulation time and memory consumption on the number of used diffraction orders for a typical 3D mask simulation. The new mask geometry description is applied in this example.

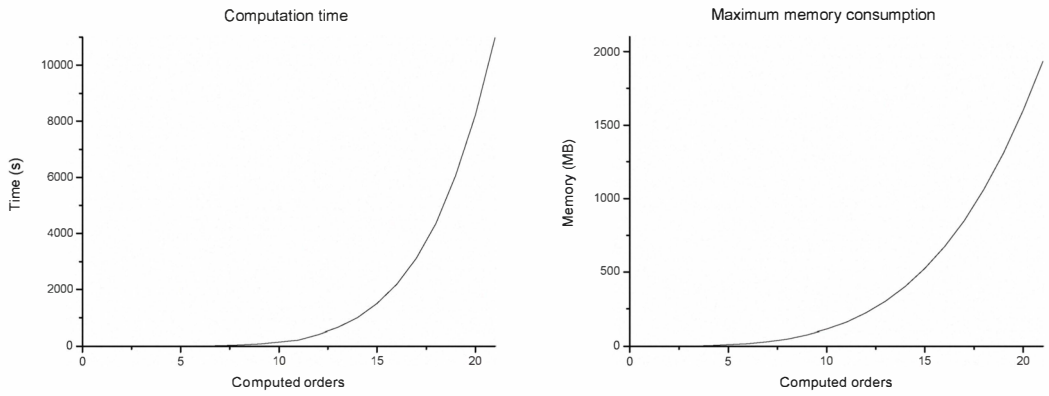

Fig.3: Waveguide simulations of three dimensional optical masks with a 260 nm contact at the center. Illumination with 193nm, 0° incidence angle, polarization 45° with respect to the x-axis. The curves show the overall computation time and maximum memory consumption versus different numbers of diffraction orders (from 2 to 21).

In Figure 3, the simulated mask is a typical 3D optical mask with a 260 nm contact at the center. The contact hole has vertical sidewalls. Therefore only one inhomogeneous layer has to be computed. The computation time is proportional to M^3 where M is the overall number of used diffraction orders. Larger mask sizes / periods require more diffraction orders resulting in increased computation time and memory consumption. Compared to the old equidistant grid mask description approach, the overall computation time for 15 diffracted orders (mask size 0.97µm x 0.97µm) is reduced from 1630 seconds to 1500 seconds. The maximum memory consumption is reduced from 3900MB to 520MB. This improvement is achieved by the new optimized mask description. Specifically, the computation time for the waveguide internal representation of the mask reduces from around 130 seconds to less than 1 second for such simple mask geometry. The memory consumption for the internal representation of the mask geometry is also reduced from about 3400MB to less than 1MB.
For the same mask geometry, the computation of 21 orders using the previous grid discretization approach is not feasible due to the large amount of memory required. However, the new mask description allows the simulation of such a big mask areas with no sampling error and with an acceptable computing time. Since the computation time and memory consumption of this new mask description only depends on the mask geometry complexity, this approach shows very good performance for many practically important mask layouts.

3.2 Simulation of larger mask areas

Compared to the full 3D method, the Q3D method enables the simulation of larger mask areas in an acceptable time. Figure 4 and 5 show simulation results for a larger EUV mask area and for a larger optical mask area, respectively. The simplified mask layouts are simulated for demonstration purposes. Another advantage of the Q3D method is to simulate standard sized optical masks and EUV masks in significantly reduced computation time [8].

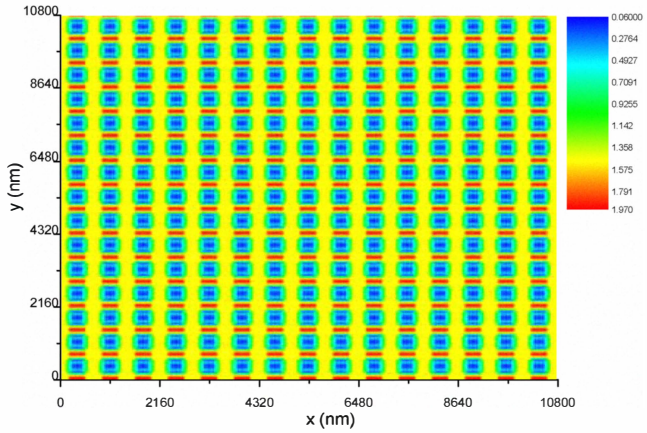

Fig 4: Near field simulation of a larger optical mask area (10.8µm x 10.8µm) with a 360nm x 360nm post absorber array. Illumination with 193nm, 0° incidence angle, X polarized. Simulation time 2010 seconds, memory consumption 76MB.

In Figure 4 the simulated mask area is 10.8µm x 10.8µm with a 360nm x 360nm post absorber array. To achieve a good convergence at the illumination wavelength of 193nm, an overall number of 167 x 167 diffraction orders is required for the near field computation (see section 2.1). Therefore, a full 3D simulation of the mask is not feasible. Using the decomposition technique, the mask geometry is split into 31 cuts in x-direction and 31 cuts in y-direction. The computation time is 2010 seconds and the memory consumption is 76MB.

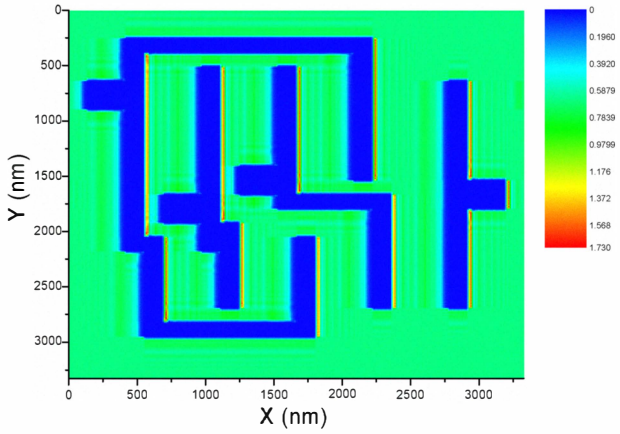

Fig 5: Near field simulation of a larger EUV mask area (3.33µm x 3.33µm) with complex feature geometry. Illumination with a wavelength of 13.5nm, 6° incidence angle, Y polarized. Simulation time 6500 seconds, memory consumption 55MB.

In Figure 5 the simulated mask area is 3.33µm x 3.33µm with a more complex feature geometry. To achieve a good convergence at the illumination wavelength of 13.5nm, an overall number of 116 x 116 diffraction orders is required for the near field computation (see section 2.1). Therefore, a full 3D simulation of the mask is not feasible. Using the decomposition technique, the mask geometry is split into 17 cuts in x direction and 14 cuts in y direction. The computation time is 6500 seconds and the memory consumption is 55MB.

It is important to mention that the simulation time of the Q3D method not only depends on the number of used diffraction orders and on the number of inhomogeneous waveguide layers but also on the number and mutual positions of the absorbers. Therefore the mask size and additionally the complexity of the mask geometry (how many different cuts have to be computed) have to be taken into account for the evaluation of simulation time and memory. A general time estimation, like for full 3D simulations, cannot be given.

3.3 Performance of parallelized Q3D

As seen from the examples in the previous section the computation time for the displayed larger mask areas with the decomposition technique is still in the order of several hours. The simulation time for more complex mask geometries can even become longer. To solve this problem, we propose to use a parallel version of the decomposition technique and to use multiple CPUs. In the following, the performance of parallelized Q3D is demonstrated in terms of speedup and efficiency.

Fig 6: Performance of the parallel version of the decomposition technique for the mask from Figure 4: Computation time versus number of slaves (CPUs).

In Figure 6, the mask from Figure 4 is simulated on the IISB HPC cluster. Master and each slave correspond to one CPU each. The figure shows the dependency of the computation time on the number of slaves (or CPUs). The overall memory consumption on the cluster is given by the sequential memory consumption times the number of CPUs. A comparison of computation time and memory requirement of Full3D, Q3D, parallelized Q3D methods for the simulation of the mask from Figure 4 is listed below:

10.8 µm by 10.8 µm optical mask with a post absorber array, 31cuts by 31cuts:

Full 3D:	simulation time: 2000 days (estimation)
Q3D:	simulation time: 2010 seconds / memory: 76 MB
Parallelized Q3D:	simulation time: 128 seconds (27 slaves) / memory: 76 MB on each CPU (76MB x 27 overall)

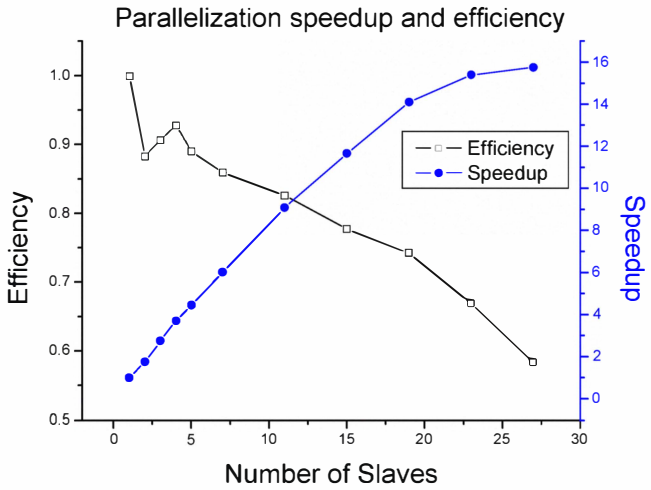

Fig 7: Performance of the parallel version of the decomposition technique for the mask from Fig 5 - speedup and efficiency versus the number of slaves (CPUs).

Figure 7 investigates the speedup and efficiency of the parallelization for the simulation of the mask from Figure 5. Speedup is defined by the ratio between the sequential computation time and computation time for the parallel version. Efficiency is obtained by dividing the speedup by the number of slaves. Using 27 slaves the simulation is around 15.7 times faster than the sequential computation. However the efficiency of parallelization is only 58%. Such loss of efficiency can be attributed to the master-slave communication traffic and to the non-optimized job distribution strategy.

3.4 Simulation of defective EUV masks

As EUV lithography is expected to be a potential approach to reach 32nm and 22nm technology nodes in future, rigorous simulation of the effect of multilayer defects is extremely important. A previous study [1] has investigated the near field of defective EUV masks using the full 3D method. However the simulation time is very long due to the numerous inhomogeneous layers caused by the defects, which also means the mask size to be investigated is highly limited. In this section, sequential and parallelized versions of the Q3D method are applied to simulations of defective EUV masks. Simulated images of masks are shown and process windows and cross sections are compared. Finally, the computation time using different methods are listed and compared.

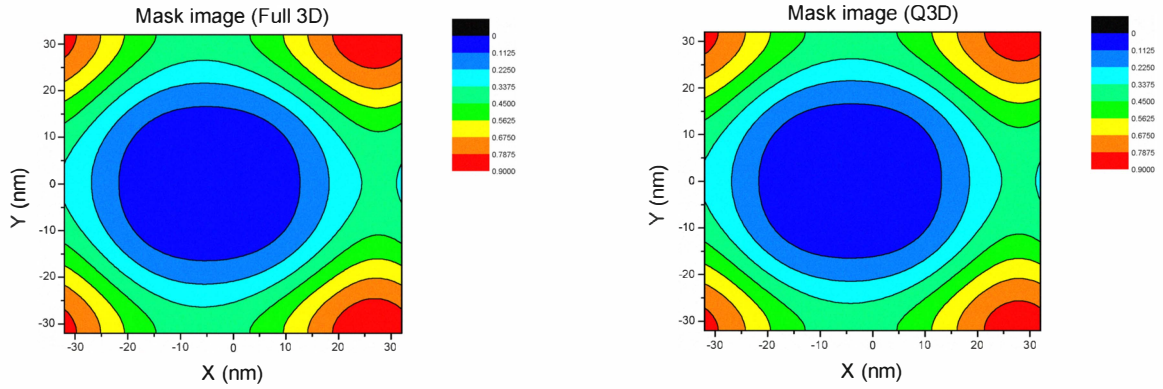

Fig 8: Images of a defective EUV mask (256nm x 256nm mask size with a 128nm x 128nm post absorber – mask scale) at best focus. The mask near fields are computed using the full 3D method (left side) and the Q3D method (right side). The illumination wavelength is 13.5nm with 6° incident angle. Typical imaging settings for EUV masks like a NA of 0.3 and a σ of 0.52 are used to compute images from the diffraction spectra.

Figure 8 presents the images of a defective EUV mask with a 32nm (wafer scale) post absorber at the center. The mask size is 256nm x 256nm, requiring 9 by 9 diffracted orders to be computed (see section 2.1). The multilayer defect is placed in the center of the mask onto the substrate and has a Gaussian shape with a height of 30 nm and a full width at half maximum of 7.5nm (wafer scale). Difference between the images for full 3D and for Q3D simulations can hardly be observed from the figures. The parallelized Q3D method provides the same results as the Q3D method, but in a much shorter computation time. Detailed computation times using the different methods are listed below:

0.256 μm by 0.256 μm defective EUV mask:

Full 3D:	simulation time: 50 hours
Q3D:	simulation time: 5 hours
Parallelized Q3D:	simulation time: 12 minutes (using 27 slaves)

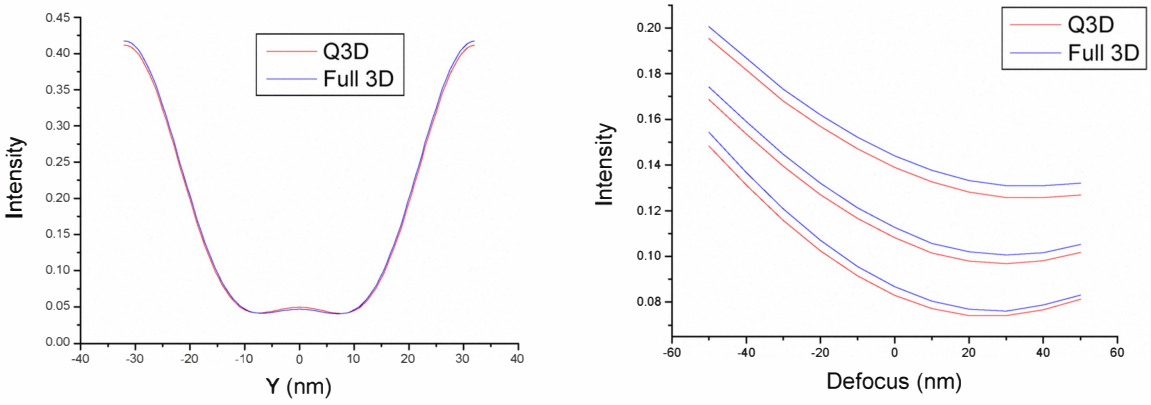

Fig 9: Comparison of cross sections (left side) and process windows (right side) of a defective EUV mask computed with the full 3D and Q3D method, respectively. Mask parameters and imaging conditions as specified in Figure 8.

Figure 9 displays the comparisons of cross sections and process windows using the full 3D and the Q3D method to compute defective EUV mask near fields. The cross sections are taken in y-direction at x=0. A good agreement of the cross sections is observed. However, the process window obtained with the Q3D method has a small offset from the process window obtained with the full 3D method. This offset, or adaptation factor, is identical with the results of a previous study regarding detailed comparisons of full 3D method and Q3D method [8]. This example demonstrates the potential capability of simulating defective EUV masks in very short time using the parallelized version of the Q3D method.

3.5 Conclusion from the simulations

The waveguide method enables the fast rigorous simulation of standard sized 3D optical and EUV masks. The new mask description approach improves the computation of the waveguide internal mask geometry representation in terms of computation time and memory consumption (see Figure 3).
Decomposition based Q3D waveguide simulations provide the possibility for the rigorous simulation of larger mask areas which are not computable with the full 3D waveguide approach (see Figures 4 and 5).
By introducing parallelization, the computation time of Q3D can be further reduced according to the number of available slave CPUs (see Figure 6 and 7).
Defective EUV masks are also investigated using the full 3D method and the parallelized Q3D method. The improvement of the computation time for rigorous simulations using the parallelized version of Q3D is remarkable. For the specific case investigated in this paper a reduction of computing time from 50 hours using the full 3D method to only 12 minutes for the parallel version of Q3D with 27 slaves was observed. Further research is necessary to investigate the small offset between the process windows which are obtained with the full 3D and the Q3D method, respectively.

4 SUMMARY

A new and optimized waveguide based electromagnetic field solver with a parallelized decomposition technique for the rigorous simulation of optical and EUV mask near fields and diffraction spectra is presented. The decomposition technique based on the waveguide method enables the simulation of larger masks with acceptable computing times. The computing time and memory consumption using this decomposition technique are given for some typical examples. The simulations in this paper have also demonstrated a further speedup by introducing the parallel version of the decomposition technique. Speedup and efficiency of parallelization are studied and corresponding dependencies on the number of available CPUs (or slaves) are shown.
The application of the parallelized decomposition technique is also investigated for the simulation of defective EUV masks. A remarkable speedup using parallelized Q3D is demonstrated. Comparison of simulation results of the Q3D

method with that one of the full 3D method shows a good agreement between image cross sections. However, the small offset between the simulated process windows, which was already observed in [8], has to be investigated in more detail.
In conclusion, the waveguide based decomposition technique with parallelization provides the capability to simulate larger optical and EUV masks in acceptable computation time. It also provides the potential possibility to simulate defective EUV masks as well as etched masks in a very short time with the required accuracy. All presented simulation models are available in the Fraunhofer IISB development and research lithography simulator Dr. LiTHO [9].

ACKNOWLEDGEMENT

The first author, Feng Shao, is member of the Erlangen graduate school in advanced optical technologies (SAOT).

REFERENCES

1. P. Evanschitzky, A. Erdmann, *Fast near field simulation of optical and EUV masks using the waveguide method*, Proc. SPIE 6533, 2007
2. L. Li, *Use of Fourier series in the analysis of discontinuous periodic structures*, J. Opt. Soc. Am. A 13 1870, 1996
3. K. D. Lucas, H. Tanabe, A. Strojwas, *Efficient and rigorous three dimensional model for optical lithography simulation*, J. Opt. Soc. Am. A 13 2187, 1996
4. M. G. Moharam, D. A. Pommet, E. B. Grann, T. K. Gaylord, *Simple implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach*, J. Opt. Soc. Am. A 12 1077, 1995
5. J. Turunen, *Form birefringence limits of Fourier expansion methods in grating theory: errata*, J. Opt. Soc. Am. A 14 2313, 1997
6. K. Adam, A. R. Neureuther, *Simplified models for edge transitions in rigorous mask modeling*, Proc. SPIE 4346, 2001
7. A. Erdmann, C. Kalus, T. Schmöller, A. Wolter, *Efficient Simulation of Light Diffraction from 3-Dimensional EUV-Masks using Field Decomposition Techniques*, Proc. SPIE 5037, 2003
8. P. Evanschitzky, F. Shao, A. Erdmann, D. Reibold, *Simulation of larger mask areas using the waveguide method with fast decomposition technique,* Proc. SPIE 6533, 2007
9. T. Fühner, T. Schnattinger, G. Ardelean, A. Erdmann, *Dr. LiTHO: a development and research lithography simulator*, Proc. SPIE 6520, 2007, see also www.drlitho.com.