Solution of the inhomogeneous Maxwell’s equations using a Born series

BENJAMIN KRÜGER,* THOMAS BRENNER, AND ALWIN KIENLE

Institut für Lasertechnologien in der Medizin und Meßtechnik an der Universität Ulm, Helmholtzstr. 12, 89081 Ulm, Germany
*benjamin.krueger@ilm-ulm.de

Abstract: An algorithm for the numerical solution of the inhomogeneous Maxwell’s equations is presented. The algorithm solves the inhomogeneous vector wave equation of the electric field by writing the solution as a convergent Born series. Compared to two dimensional finite difference time domain calculations, solutions showing the same accuracy can be calculated more than three orders of magnitude faster.

© 2017 Optical Society of America

OCIS codes: (260.2110) Electromagnetic optics; (290.5825) Scattering theory; (290.4020) Mie theory.

References and links
1. Introduction

Today there are numerous applications where solutions of Maxwell’s equations in inhomogeneous media are required. These applications include the propagation of electromagnetic waves in biological materials [1, 2], in lithography [3], and in optical metamaterials [4, 5].

While there are analytic solutions for simple geometries such as spheres and infinite cylinders, most real scattering problems need numerical calculations for a realistic description [6–8]. For this purpose several numerical methods have been developed. These methods include the finite difference time domain (FDTD) method [9, 10] where the time integration of Maxwell’s equations is performed on a discrete grid using finite differences to calculate the spatial derivatives. For this method, very small cell sizes have to be used. This restriction can partly be overcome by using the pseudo spectral time domain (PSTD) method [11, 12] where the spatial derivatives are calculated using a Fourier transformation. A further method is the discrete dipole approximation (DDA), [13, 14] where the system is discretized by small polarizable grains. Another method is the solution using finite element methods (FEM) [2, 15].

Recently, Osnabrugge et al. [16] found a solution of the Helmholtz equation of the electric field that uses plane waves as trial functions. The resulting linear system can then be solved using a Born series. The advantage of this method is that only one electric-field distribution needs to be stored at a time. In contrast, Krylov subspace solvers like the generalized minimization of residuals [17] need to store several basis vectors of the Krylov subspace. The Helmholtz equation is an approximation of the wave equation that emerges from Maxwell’s equations. The divergence of the electric field vanishes only in two dimensions for waves with electric field perpendicular to the scattering plane. Thus an accurate solution is possible in two dimensions with one distinct polarization only. In all other cases the Helmholtz equation is an approximation to Maxwell’s equations. This approximation looses accuracy when the divergence of the electric field increases.

Here, we show how the method of Osnabrugge et al. can be extended to solve the full wave equation, i.e. Maxwell’s equations. This extension allows for a better accuracy of the calculated field in systems with a strong spatial change of the refractive index for all polarizations.

This work is organized as follows. In the remaining part of this section the idea of Osnabrugge et al. is shortly sketched and the Helmholtz equation as an approximation of the wave equation that can be derived from Maxwell’s equations is discussed. In the second part we describe the new algorithm with special attention to its comparison to the work of Osnabrugge et al. In the third and fourth part we present numerical results using the described algorithm. These results are compared with an analytical theory for scattering by multiple cylinders [7, 18–20]. Finally, a conclusion is given.

1.1. Maxwell’s equations and the Helmholtz equation

We start with Maxwell’s equations where the electric and magnetic fields are harmonic oscillations in time with a given frequency $\omega$. The values of $\epsilon$ and $\mu$ are constant in time and no current or free charges are present. The fields then become $E(t) = \tilde{E}e^{-i\omega t}$ and $H(t) = \tilde{H}e^{-i\omega t}$. 

We get
\[ \vec{\nabla} \cdot (\epsilon \vec{E}) = 0, \]  
\[ \vec{\nabla} \cdot (\mu \vec{H}) = 0, \]  
\[ \vec{\nabla} \times \vec{E} = i\omega \mu \vec{H}, \text{ and} \]  
\[ \vec{\nabla} \times \vec{H} = -i\omega \epsilon \vec{E}. \]  

Taking the curl of Eq. (3) and inserting Eq. (4) one gets
\[ \vec{\nabla} \times (\vec{\nabla} \times \vec{E}) = \omega^2 \mu \epsilon \vec{E} \]  
assuming that \( \mu \) is not spatially dependent, which is usually assumed, e. g. for biophotonics. This equation can be rewritten as
\[ \vec{\nabla}^2 \vec{E} + \kappa^2 \vec{E} = k^2 \vec{E} \]  
with the spatially dependent wave number \( k \) with \( k^2 = \omega^2 \mu \epsilon \). If the first term on the left-hand side can be neglected, we arrive at the Helmholtz equation. For the Helmholtz equation in Cartesian coordinates the components of the electric field do not mix. Therefore, each component has to fulfill the scalar Helmholtz equation
\[ \vec{\nabla}^2 \psi + k^2 \psi = -s \]  
where \( \psi \) is one of the three components of the electric field and \( s \) is a source term.

Except the source term Eqs. (7) and (6) are the same if \( \vec{\nabla} (\vec{\nabla} \cdot \vec{E}) = 0 \). This expression can be rewritten as
\[ -\vec{\nabla} \left( \vec{E} \cdot \vec{\nabla} \ln(\epsilon) \right) = 0 \]  
by making use of
\[ \vec{\nabla} \cdot \vec{E} = \epsilon \vec{E} - (\epsilon \vec{E}) = \frac{\vec{E} \cdot \vec{\nabla} \ln(\epsilon)}{\epsilon^2} = -\vec{E} \cdot \vec{\nabla} \ln(\epsilon) \]  
and Eq. (1). Plugging this expression into Eq. (6) one finds that the Helmholtz equation is identical to the wave equation if \( \vec{\nabla} \ln(\epsilon) = 0 \) or \( \vec{E} \perp \vec{\nabla} \ln(\epsilon) \). The first condition is identical to the absence of any scatterer in the system. The second condition is for example fulfilled for two-dimensional scattering where the electric field is perpendicular to the scattering plane.

Furthermore it is possible to use the Helmholtz equation as an approximation for some other systems where the term \( \vec{\nabla} (\vec{\nabla} \cdot \vec{E}) \) is small compared to the other two terms in Eq. (6). This is the case if \( \epsilon \) varies slowly on the length scale of a wavelength.

1.2. Numerical solution of the Helmholtz equation using a Born series

Equation (7) can be solved numerically using a Born series as shown by Osnabrugge et al [16]. They write Eq. (7) as
\[ \vec{\nabla}^2 \psi + (k_0^2 + i\delta) \psi = -(k^2 - k_0^2 - i\delta) \psi - s = -v \psi - s, \]  
where \( k_0 \) is a constant background wave number and \( \delta \) is a constant real number. The left-hand side is independent of the position. It is therefore possible to derive a Green’s function \( g(\vec{r}) \) by solving
\[ \vec{\nabla}^2 g(\vec{r}) + (k_0^2 + i\delta) g(\vec{r}) = -\delta_D(\vec{r}), \]
where $\delta_D$ is the Dirac function. After a Fourier transform to $k$ space this equation becomes

$$- (k_x^2 + k_y^2 + k_z^2) \hat{g}(\vec{k}) + (k_0^2 + i\delta) \hat{g}(\vec{k}) = -1,$$

(12)
since a derivative in real space transforms to a multiplication in $k$ space. The Fourier transform of the Green’s function then reads

$$\hat{g}(\vec{k}) = \frac{1}{k_x^2 + k_y^2 + k_z^2 - k_0^2 - i\delta}.$$  

(13)

For the calculation of the Green’s function in real space one needs to distinguish between two-dimensional and three-dimensional systems. The solutions can be found in Eqs. (35) and (39), respectively.

Writing Eq. (10) on a discrete grid, we get

$$\vec{\Psi} = GV\vec{\Psi} + G\vec{S}$$  

(14)

with the matrices $G$ and $V$ representing the discretized Green’s function $g$ and potential $v$, respectively. The vectors $\vec{\Psi}$ and $\vec{S}$ contain the fields $\psi$ and $s$ at every grid point, respectively. This can be rewritten as

$$\vec{\Psi} = (\mathbb{I} - GV)^{-1} G\vec{S}$$  

(15)

or in the Born series

$$\vec{\Psi} = \sum_{n=0}^{\infty} (GV)^n G\vec{S}.$$  

(16)

Here, $\mathbb{I}$ is the identity matrix. Cutting the sum after the first term would give a result similar to the perturbative Born approximation [21]. The problem is that even with higher terms included this series converges only if the magnitudes of all eigenvalues of $GV$ are smaller than one.

To overcome this problem, Osnabrugge et al. introduce the precondioner [16]

$$\gamma = i\delta V.$$  

(17)

It can be seen that $\gamma$ is a diagonal matrix with the entries

$$\frac{i}{\delta}(k^2 - k_0^2 - i\delta).$$  

(18)

Multiplying Eq. (14) with $\gamma$ yields

$$\gamma\vec{\Psi} = \gamma GV\vec{\Psi} + \gamma G\vec{S}.$$  

(19)

This can be solved for $\vec{\Psi}$ as

$$\vec{\Psi} = [\mathbb{I} - (\gamma GV - \gamma + \mathbb{I})]^{-1}\gamma G\vec{S}$$  

(20)

and written as the series

$$\vec{\Psi} = \sum_{n=0}^{\infty} (\gamma GV - \gamma + \mathbb{I})^n \gamma G\vec{S}.$$  

(21)

From this series the iterative solution

$$\vec{\Psi}_{n+1} = (\gamma GV - \gamma + \mathbb{I})\vec{\Psi}_n + \gamma G\vec{S}$$  

(22)

with $\vec{\Psi}_0 = 0$ can be derived.
Osnabrugge et al. found a restriction of [16]

\[ \delta \geq \max_r |k^2(r) - k_0^2| \]  

(23)

for which this iteration will always converge if

\[ U = \mathbb{I} + 2i\delta G \]  

(24)

is a unitary matrix. However from Eq. (18) one can see that if the value of \( k \) in at least one cell fulfills the conditions

\[ \text{Re}(k^2 - k_0^2) = 0 \]  

(25)

and

\[ \text{Im}(k^2 - k_0^2) = \delta \]  

(26)

the matrix \( \gamma \) becomes singular. In this case the multiplication in Eq. (19) changes the solution of the linear system. Thus the restriction in Eq. (23) is replaced with the slightly stricter form

\[ \delta > \max_r |k^2(r) - k_0^2| \]  

(27)

that ensures that the preconditioner is always non-singular.

2. Algorithm

In this section our algorithm for the solution of the wave equation is described. For its derivation we follow the ansatz used by Osnabrugge et al. in their calculations for the Helmholtz equation [16]. For the wave equation we now have to solve

\[ \nabla^2 \vec{E} - \nabla(\vec{\nabla} \cdot \vec{E}) + (k_0^2 + i\delta)\vec{E} = -v\vec{E} - \vec{s}. \]  

(28)

Here \( \vec{s} \) is a source term that depends on the incident light. This term will be discussed later and will be calculated in Eq. (61). From Eq. (28) the Green’s function is calculated by solving

\[ \nabla^2 g(\vec{r}) - \nabla(\vec{\nabla} \cdot g(\vec{r})) + (k_0^2 + i\delta)g(\vec{r}) = -\delta_D(\vec{r}) \mathbb{I}, \]  

(29)

where the Green’s function \( g(\vec{r}) \) is a \( 3 \times 3 \) matrix. As shown above for the Helmholtz equation it is convenient to transform Eq. (29) into Fourier space. This yields

\[
\begin{pmatrix}
-k_y^2 & k_x & 0 \\
-k_x^2 & k_y & 0 \\
k_y k_x & -k_x k_y & k_z^2
\end{pmatrix}
\hat{g}(\vec{k}) + (k_0^2 + i\delta)\hat{g}(\vec{k}) = -\mathbb{I},
\]  

(30)

employing the fact that a differentiation in real space becomes a multiplication in \( k \) space.

In the following one needs to distinguish between the two-dimensional and three-dimensional cases.

2.1. Two-dimensional system

In the two-dimensional system there is no change in \( z \) direction. Therefore the derivatives in \( z \) direction vanish and Eq. (30) becomes

\[
\begin{pmatrix}
-k_y^2 & k_x & 0 \\
k_x & -k^2 & 0 \\
0 & 0 & -k_x^2 - k_y^2
\end{pmatrix}
\hat{g}(\vec{k}) + (k_0^2 + i\delta)\hat{g}(\vec{k}) = -\mathbb{I}
\]  

(31)
It becomes readily visible that the system separates into one equation for polarization in the \(xy\) plane and another for the polarization perpendicular to it. The Green’s function is calculated from solving Eq. (31) for \(\hat{g}(\vec{k})\). For the matrix elements one finds

\[
\hat{g}_{m,n}(\vec{k}) = \frac{\delta_{m,n}}{k^2 - k_0^2 - i\delta} + \frac{k_m k_n}{(k^2 - k_0^2 - i\delta)(-k_0^2 - i\delta)}(1 - \delta_{m,z})(1 - \delta_{n,z}),
\]

where \(k^2 = k_x^2 + k_y^2\).

For the transformation back to real space we use

\[
\mathcal{F}^{-1}_{2D} \left[ \frac{1}{k^2 + a^2} \right] = \frac{K_0(ar)}{2\pi}.
\]

Here \(K_n(ar)\) is the modified Bessel function of the second kind of order \(n\) with the parameter \(ar\). The function in Fourier space depends on the square of \(a\). Thus, \(a\) can assume two different values. Without loss of generality we assume that the real part of \(a\) is positive. Setting \(a^2 = -k_0^2 - i\delta\), the first part of Eq. (32) can be transformed. For the second part we use the expression

\[
\mathcal{F}^{-1}_{2D} \left[ \frac{k_m k_n}{(k^2 + a^2) a^2} \right] = -\frac{\partial}{\partial r_m} \frac{\partial}{\partial r_n} \frac{K_0(a\sqrt{x^2 + y^2})}{2\pi a^2},
\]

from which we find the Green’s function

\[
g_{m,n}(\vec{r}) = \frac{K_0(ar)}{2\pi} \delta_{m,n} + \frac{K_1(ar)}{2\pi ar} \delta_{m,n} (1 - \delta_{m,z}) - \frac{K_2(ar)r_m r_n}{2\pi r^2} (1 - \delta_{m,z})(1 - \delta_{n,z})
\]

in real space. The first term is the Green’s function for the Helmholtz equation.

### 2.2. Three-dimensional system

For the three-dimensional system the Green’s function in \(k\) space is derived from the full Eq. (30) including \(k_z\). The matrix elements are

\[
\hat{g}_{m,n}(\vec{k}) = \frac{\delta_{m,n}}{k^2 - k_0^2 - i\delta} + \frac{k_m k_n}{(k^2 - k_0^2 - i\delta)(-k_0^2 - i\delta)}.
\]

To calculate the Green’s function in real space we use the identity

\[
\mathcal{F}^{-1}_{3D} \left[ \frac{1}{k^2 + a^2} \right] = \frac{e^{-ar}}{4\pi r},
\]

where \(a\) is assumed to have a positive real part as for the two-dimensional case. Setting \(a^2 = -k_0^2 - i\delta\) allows us to calculate the inverse Fourier transformation of the first part of Eq. (36). To calculate the inverse Fourier transform of the second part we employ

\[
\mathcal{F}^{-1}_{3D} \left[ \frac{k_m k_n}{(k^2 + a^2) a^2} \right] = -\frac{\partial}{\partial r_m} \frac{\partial}{\partial r_n} \frac{e^{-a\sqrt{x^2 + y^2 + z^2}}}{4\pi a^2 \sqrt{x^2 + y^2 + z^2}}.
\]

This yields

\[
g_{m,n}(\vec{r}) = \delta_{m,n} \frac{e^{-ar}}{4\pi r} + \frac{1 + ar}{a^2 r^2} \delta_{m,n} \frac{e^{-ar}}{4\pi r} - \frac{a^2 r^2 + 3ar + 3 r_m r_n}{a^2 r^2} \frac{e^{-ar}}{r^2} \frac{1}{4\pi r},
\]

where \(a\) is the square root of \(-k_0^2 - i\delta\) with positive real part. As for the two-dimensional case the first term is the Green’s function for the Helmholtz equation.
2.3. Born series

For the calculation of our generalized Born series we follow the calculations of Osnabrugge et al. for the Helmholtz equation. With the aid of the Green’s function for the wave equation one can write

$$\vec{E}(\vec{r}) = \int d\vec{r}' g(\vec{r} - \vec{r}') \left( v(\vec{r}') \vec{E}(\vec{r}') + s(\vec{r}') \right).$$  \hspace{1cm} (40)

In the spatially discretized form this equation reads

$$\Psi = GV\Psi + GS$$ \hspace{1cm} (41)

with a vector $\Psi$ of size $3N$ containing the electric field at each of the $N$ grid points. $G$ is a dense $3N \times 3N$ matrix. This matrix consists of $N^2$ $3 \times 3$ blocks that are $g(\vec{r}_i - \vec{r}_j)$, where $\vec{r}_i$ and $\vec{r}_j$ are the positions of the according grid points. $V$ is a diagonal $3N \times 3N$ matrix.

Similar to the calculations of Osnabrugge et al. we rewrite this expression as

$$\gamma\Psi = \gamma GV\Psi + \gamma GS$$ \hspace{1cm} (42)

with the preconditioner

$$\gamma = \frac{i}{\delta} V$$ \hspace{1cm} (43)

and a value of

$$\delta > \max_{\vec{r}} |k^2 - k_0^2|.$$ \hspace{1cm} (44)

It is worth noting that in contrast to the work of Osnabrugge et al. $\gamma$ is now a $3N \times 3N$ matrix. Equation (42) can be rewritten as

$$\Psi = \left[ I - (\gamma GV - \gamma + I) \right]^{-1} \gamma GS$$ \hspace{1cm} (45)

or in the series expansion

$$\Psi = \sum_{n=0}^{\infty} (\gamma GV - \gamma + I)^n \gamma GS.$$ \hspace{1cm} (46)

From this expression the iterative solution

$$\Psi_{n+1} = (\gamma GV - \gamma + I)\Psi_n + \gamma GS = \Psi_n + \gamma (GV\Psi_n + S) - \Psi_n)$$ \hspace{1cm} (47)

with $\Psi_0 = 0$ can be derived.

For the numerical evaluation one needs to keep in mind that the dense matrix $G$ represents a convolution as can be seen from Eqs. (40) and (41). Since a convolution in real space becomes a local multiplication in Fourier space, one gets

$$\Psi_{n+1} = \Psi_n + \gamma (F^{-1}\hat{G}F(V\hat{\Psi}_n + \hat{S}) - \hat{\Psi}_n),$$ \hspace{1cm} (48)

with a block diagonal matrix

$$\hat{G} = FGF^{-1}.$$ \hspace{1cm} (49)

This matrix has $N$ $3 \times 3$ blocks containing the matrix $\hat{g}(\vec{k})$ for different values of $\vec{k}$. The multiplication with $G$ has a time complexity of $O(N^2)$. Since the multiplication with $\hat{G}$ and the Fourier transforms have time complexities of $O(N)$ and $O(N \log(N))$, respectively, it is faster to use Eq. (48) than Eq. (47) for computations.
2.4. Convergence

The iteration is not necessarily convergent. However, Osnabrugge et al. have proven that this series will always converge if \( \text{Im}(k^2 - k_0^2) \geq 0 \) is fulfilled everywhere in the sample, meaning that there should be no gain in the medium, that the stricter condition \( \text{Im}(k^2 - k_0^2) > 0 \), i.e. absorption, is fulfilled somewhere in the sample, and that the matrix

\[
U = \mathbb{I} + 2i\delta G
\]

is a unitary matrix [16]. For the Helmholtz equation Osnabrugge et al. have shown that \( U \) is unitary. We will now show that this is also the case for the wave equation derived from Maxwell’s equations.

First we note that the Fourier transformation can be written as a unitary matrix. Thus, we are left with the proof that

\[
\hat{U} = \mathcal{F} U \mathcal{F}^{-1}
\]

is a unitary matrix. The matrix \( \hat{U} \) is a block diagonal \( 3N \times 3N \) matrix with \( N \) \( 3 \times 3 \) blocks containing the same matrix

\[
\hat{u}(\vec{k}) = \mathbb{I} + 2i\delta \hat{g}(\vec{k})
\]

for different values of \( \vec{k} \).

A block diagonal matrix is unitary if all the blocks are unitary transformations. Thus, we are left with showing that \( \hat{u}(\vec{k}) \) is a unitary matrix. An easy way to do so is to calculate the eigenvalues of \( \hat{g} \). Using spherical coordinates with \( (k_x, k_y, k_z) = |k|(\sin(\theta) \cos(\phi), \sin(\theta) \sin(\phi), \cos(\theta)) \), we introduce the rotation matrix

\[
R = \begin{pmatrix}
\cos(\theta) & \cos(\theta) \sin(\phi) & -\sin(\theta) \\
-\sin(\phi) & \cos(\phi) & 0 \\
\sin(\theta) \cos(\phi) & \sin(\theta) \sin(\phi) & \cos(\theta)
\end{pmatrix}.
\]

With this matrix we get

\[
\hat{g} = R^T \begin{pmatrix}
\frac{1}{k^2 - k_0^2 - i\delta} & 0 & 0 \\
0 & \frac{1}{k^2 - k_0^2 - i\delta} & 0 \\
0 & 0 & \frac{1}{k^2 - k_0^2 - i\delta}
\end{pmatrix} R
\]

\[
= \frac{i}{2\delta} R^T \mathbb{I} R - \frac{i}{2\delta} R^T \begin{pmatrix}
\frac{k^2 - k_0^2 + i\delta}{k^2 - k_0^2 - i\delta} & 0 & 0 \\
0 & \frac{k^2 - k_0^2 + i\delta}{k^2 - k_0^2 - i\delta} & 0 \\
0 & 0 & \frac{-k_0^2 + i\delta}{k_0^2 - i\delta}
\end{pmatrix} R. \tag{54}
\]

It is worth noting that for the two-dimensional case \( \vec{k} \) is always in the \( xy \) plane and thus \( \sin(\theta) = 1 \). We now calculate

\[
\hat{u} = R^T \begin{pmatrix}
\frac{k^2 - k_0^2 + i\delta}{k^2 - k_0^2 - i\delta} & 0 & 0 \\
0 & \frac{k^2 - k_0^2 + i\delta}{k^2 - k_0^2 - i\delta} & 0 \\
0 & 0 & \frac{-k_0^2 + i\delta}{k_0^2 - i\delta}
\end{pmatrix} R. \tag{55}
\]

This consists of two rotations and one unitary matrix. Therefore, \( \hat{u} \) is unitary as well.
2.5. Truncating condition

In Eq. (46) the resulting vector $\vec{\Psi}$ is written as an infinite sum. Therefore, to obtain the exact solution one needs to apply the iteration in Eq. (48) an infinite number of times. Therefore, we need a condition that helps us to decide whether the iteration can be truncated at the current state or not.

During the update of Eq. (48) one needs to calculate the value

$$\vec{r}_n = \mathcal{F}^{-1}(\hat{G}F(\mathcal{V}\vec{\Psi}_n + \vec{S}) - \vec{\Psi}_n), \quad (56)$$

that is the residual of Eq. (41) for $\vec{\Psi}_n$. Thus the calculation of the new value $\vec{\Psi}_{n+1}$ gives us the current residual $\vec{r}_n$ for free. This value can now be used to judge whether the iteration has been converged or further iterations are needed.

We have defined two conditions. The absolute error is given by the maximum of the residual over all cells. For the relative error we divide the residual by the field in the cell before finding the maximum. In the data shown in this paper we have used a maximum error of 0.001 and a relative error of 0.01. Stricter values do not affect the accuracies derived in Sec. 4.

2.6. Sources

In Eq. (28) we have defined a source term that includes the incident light. To calculate this term we write the electric field as the sum of an incident field $\vec{E}_i$ that comes from external sources and a field $\vec{E}_s$ that stems from the scattering in the system. Then Eq. (6) reads

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}_s) - \vec{\nabla}^2 \vec{E}_s = k^2(\vec{E}_i + \vec{E}_s). \quad (57)$$

The incident field alone needs to fulfill the wave equation

$$\vec{\nabla}^2 \vec{E}_i + k_m^2 \vec{E}_i = 0 \quad (58)$$

for the medium that surrounds the simulated system. Here $k_m$ is the constant wave number in the medium. In these two equations we have employed the fact that the divergence of $\vec{E}_i$ vanishes as can be seen from Eq. (8), keeping in mind that $\epsilon$ is constant in the medium.

By plugging Eq. (58) into Eq. (57) one gets

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}_s) - \vec{\nabla}^2 \vec{E}_s = k^2 \vec{E}_s + (k^2 - k_m^2) \vec{E}_i. \quad (59)$$

In a shape that is more in line with Eq. (28) this reads

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}_s) - \vec{\nabla}^2 \vec{E}_s + (k_0^2 + i\delta) \vec{E}_s = -\nu \vec{E}_s - (k_m^2 - k^2) \vec{E}_i. \quad (60)$$

By comparing this with Eq. (28) one easily obtains

$$\vec{s} = (k_m^2 - k^2) \vec{E}_i. \quad (61)$$

2.7. Boundary conditions

As boundary conditions we use the ones introduced by Osnabrugge et al. The simulated system is surrounded by an absorbing region of width $l$, in which the wave number is given by

$$k^2(d) = k_m^2 + \frac{\alpha^2(N - \alpha \delta + 2i\kappa_m d)(\alpha d)^{N-1}}{P_N(d)N!} \quad (62)$$

with the polynomial

$$P_N(d) = \sum_{n=0}^{N} \frac{(\alpha d)^n}{n!} \quad (63)$$

and the distance $d$ from the simulated system.
Fig. 1. (left) Scheme of the anti-aliasing algorithm to assign a refractive index to the simulation cells. The scatterer is a cylinder whose center is not necessarily aligned with the rectangular grid. Its circumference is denoted by the blue line. The grid cells (solid black lines) are divided into subcells (dashed black lines). For each subcell a black dot at its center denotes that the subcell is considered to be inside the cylinder. The effective refractive index of the grid cell (denoted by the gray scale) depends on the number of subcells that are considered to be inside the cylinder. (right) Our test system consists of five cylinders with a radius of 5 μm and a refractive index \( n_2 \) that are placed in a homogeneous medium with a refractive index of \( n_1 \). The near field is calculated in a square that contains all cylinders. From the calculated near field the far field is calculated by a near to far field transformation on the boundary of this square.

### 2.8. Anti-aliasing

Before we start the calculation using our algorithm we need to assign a refractive index to each simulation cell of the grid. One possible method is to assign the refractive index at the center of each cell. To get a smoother surface and to reduce stair casing effects in calculations with round surfaces we decided to calculate the refractive index using an anti-aliasing routine. For this calculation we divide each cell in \( n_s \times n_s \) subcells as depicted in Fig. 1 for \( n_s = 2 \). We then assign to each subcell the refractive index at the center of the respective subcell. Subsequently, the average of the refractive indices in all subcells of one cell is calculated and assigned to the cell. For the results shown in this work we use \( n_s = 10 \).

### 3. Runtime complexity

In this section we calculate the dependence of the simulation time on the size of the system. For simplicity we assume a system with the same size \( L \) and number of cells \( N \) in each spatial direction, i.e. the grid contains \( N^2 \) and \( N^3 \) cells for two and three dimensions, respectively.

#### 3.1. FDTD

In the case of FDTD the time that is needed for the calculation of one simulation step scales as

\[
O(N^D)
\]

with the dimension \( D \) of the system. For the simulation of a time \( t \) with a time step of \( \Delta t \) this yields

\[
O \left( N^D \frac{t}{\Delta t} \right).
\]

(65)
Fig. 2. Near field calculated by the Born series with 10 cells per μm. The subfigures (a)-(c) and (g)-(i) show the amplitude of the electric near field for the first and second set of refractive indices, respectively. Shown are the x ((a) and (g)), the y ((b) and (h)), and the z component ((c) and (i)). Subfigures (d)-(f) and (j)-(l) show the absolute values of the respective differences to the near field calculated analytically. All values are normalized to the amplitude of the incident wave $E_0$. For the sake of illustration the lower end of the colormap has been cut even though there are values that are smaller than the range shown in the color bar.

To ensure the stability of the simulation the time step has to be given by

$$\Delta t = \frac{S}{c} \Delta,$$

where the Courant number $S$ is a constant that depends on the dimension, $c$ is the speed of light in the simulated medium, and $\Delta$ is the grid spacing. To describe the interaction with scatterers accurately we need to ensure that the light is able to travel through the grid $n$ times. The maximum distance in the grid is given by its diagonal. The maximum time we need to
simulate is therefore
\[ t = \frac{n}{c} \sqrt{DL^2} = \frac{n}{c} \sqrt{D}. \]  
(67)

This yields
\[ O \left( \frac{L^{D+1}}{\Delta^{D+1}} \right) \]  
(68)

if \( n \) is assumed to be independent of the system size.

For the determination of a proper value for the grid spacing we note that for small systems the grid spacing needs to scale linearly with the wavelength to keep the error from the finite
Fig. 4. Calculated far field for the five cylinders shown in Fig. 1 with a refractive index $n_1 = 1.0$ of the medium and a refractive index $n_2 = 1.2$ of the cylinders. The figures show the components parallel ((a) and (b)) and perpendicular ((c) and (d)) to the scattering plane. A magnification of the region of backward scattering is shown in (e) and (f). The figures compare the results from the analytical theory in (a), (c), and (e) with the present algorithm using a cell size of 100 nm and in (b), (d), and (f) with the FDTD calculation using a cell size of 20 nm.

Differences constant. In addition to this error we have to take into account that in FDTD simulations there is a numerical dispersion relation that is different to the analytical one. The error due to the occurring phase lag becomes maximal at the boundaries of the simulated area. This maximum dispersion determines the accuracy for large systems. The numerical wave number $k_n$ can be calculated as [10]

$$k_n = \frac{2}{\Delta} \sin^{-1} \left( \frac{1}{S} \sin \left( \frac{\pi S \Delta}{\lambda} \right) \right) \approx k_a + \frac{\pi^3(1-S^2)\Delta^2}{3\lambda^3}$$

(69)

with the analytical wave number $k_a = 2\pi/\lambda$. We note that the value of $k_n$ depends on the
Fig. 5. Similar data as in Fig. 4 but for a refractive index $n_1 = 1.36$ of the medium and a refractive index $n_2 = 1.39$ of the cylinders. The figures compare the results from the analytical theory in (a), (c), and (e) with the present algorithm using a cell size of 100 nm and in (b), (d), and (f) with the FDTD calculation using a cell size of 12.5 nm.

The propagation direction of the wave with respect to the grid axes. For simplicity we use the dispersion relation for a wave propagating parallel to one grid axis only. We now aim to keep the maximum phase lag $\phi_{\text{max}}$ constant. For this we approximate $\phi_{\text{max}}$ by the phase lag that occurs while the wave moves through the sample once. The maximum phase lag can thus be written as

$$\phi_{\text{max}} \approx (k_n - k_a) L \approx \frac{\pi^3 (1 - S^2) \Delta^2 L}{3 \lambda^3},$$

(70)

where $\lambda$ is the wavelength. Solving this relation for $\Delta$ yields

$$\Delta \approx \sqrt{\frac{3 \lambda^3 \phi_{\text{max}}}{\pi^3 (1 - S^2) L}}$$

(71)
leading to the conclusion that \( \Delta \) has to scale with the inverse square root of \( L \). Indeed, numerical experiments with a wave traveling along the \( x \) axis of an empty grid with \( S = 0.5 \) and a grid spacing of \( \Delta = 0.2\lambda^{1.5}/L^{0.5} \) have shown that the phase lag over the whole grid is 18 degrees, irrespective of the grid size. We thus get a calculation time that scales with

\[
O \left( \frac{L^{1.5(D+1)}}{\lambda^{1.5(D+1)}} \right).
\]  

(72)

### 3.2. This algorithm

In the calculation of one iteration of the present algorithm, the part with the highest runtime complexity per iteration is the calculation of the Fourier transformation. The runtime for a \( D \) dimensional Fourier transformation is of the order

\[
O(N^D \log(N)).
\]

(73)

Since this algorithm contains no time integration there is no dispersion relation and \( N \) is proportional to \( L/\lambda \).

Osnabrugge et al [16], show that for each additional iteration of the Born series the light travels a distance of about \( 2\delta_0/\delta \). This expression scales linearly with the wavelength. The number of iterations then scales with \( L/\lambda \). We find

\[
O \left( \frac{L^{D+1}}{\lambda^{D+1}} \log \left( \frac{L}{\lambda} \right) \right).
\]

(74)
4. Numerical tests

In this section we present a test of the accuracy and the speed of the present algorithm and compare the results to analytical calculations using an analytical solution [7, 18–20] and FDTD simulations.

4.1. Calculated system

For the test we use a square region with an edge length of 100 μm that is filled with five cylinders with a radius of 5 μm. The background has a refractive index of $n_1$ while the cylinders have a refractive index denoted by $n_2$. The positions of the cylinders are given in Tab. 1. They are positioned so that there is a small space between them and the boundary. The full system is depicted in Fig. 1.

Table 1. Positions of the five cylinders in the calculated system relative to the center of the simulation area. The system is depicted in Fig. 1.

<table>
<thead>
<tr>
<th>cylinder</th>
<th>x position</th>
<th>y position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-44.5 μm</td>
<td>-44.5 μm</td>
</tr>
<tr>
<td>2</td>
<td>44.5 μm</td>
<td>-44.5 μm</td>
</tr>
<tr>
<td>3</td>
<td>44.5 μm</td>
<td>44.5 μm</td>
</tr>
<tr>
<td>4</td>
<td>20.6 μm</td>
<td>-18.0 μm</td>
</tr>
<tr>
<td>5</td>
<td>-10.4 μm</td>
<td>18.1 μm</td>
</tr>
</tbody>
</table>

To get a deeper insight into the effect of the material parameters we performed calculations for two different sets of refractive indices. In the first set with $n_1 = 1$ and $n_2 = 1.2$ the cylinders are situated in vacuum. The second set with $n_1 = 1.36$ and $n_2 = 1.39$ has parameters that are typical for biological tissue [22]. In both cases the incident light is traveling in x direction and has a vacuum wavelength of 1 μm. The incident field is $\vec{E}_i = \vec{E}_0 (0, 1, 1)^T$.

For the boundary conditions in the Born series calculations we used the parameters $\alpha = 0.35/\mu m$, $l = 25 \mu m$, and $N = 4$ as described in Sec. 2.7. A reduction of $l$, i.e. the size of the absorbing region, would make an increase of the damping $\alpha$ necessary to ensure that a wave is sufficiently damped and that no reflections at the edges of the grid occur. But a larger value for $\alpha$ leads to reflections in the transition region where the absorption sets in.

The area that is covered by the absorbing boundaries is $(L + 2l)^2 - L^2 = 12500 \mu m^2$. This area is by a factor of 1.25 larger than the area covered by the region containing the scatterers. This means that for the following two-dimensional benchmarks the size of the whole simulation grid is about two times larger than that used for the finite difference time domain simulations.

To determine how the simulation time is affected by this increase in size we start from the expression in Eq. (74) and approximate the calculation time $T$ with

$$ T \propto \frac{L^3}{\lambda^3} \log \left( \frac{L}{\lambda} \right). \quad (75) $$

This coarse approximation leads to an increase in simulation time by a factor of 3.7. Thus more efficient boundary conditions could significantly reduce the simulation time of the Born series method in the future.

4.2. Near field

The near field calculated using the present algorithm and the difference to the analytical calculation are shown in Figs. 2 and 3.
It can be seen that the cylinders act like converging lenses. Directly behind the cylinders there is a region where the fields are stronger than the fields of the incident wave. Staying behind the cylinder but increasing the distance beyond the focal length the field strongly decreases. This leads to a strong interaction between cylinder one (lower left corner) and cylinder two (lower right corner) governed by the fact that there is nearly no light directly impinging on the cylinder two due to shading from cylinder one.

In the data from the Born series as shown in Fig. 2, one can see that for the cylinders in vacuum the positions with maximum error are the surfaces of the cylinders, especially at positions where the electric field becomes discontinuous. These positions are for the $x$ and $y$ components at the left/right and at the top/bottom surface of the cylinders, respectively. The $z$ component of the electric field is always continuous, leading to a much smaller maximum error. For the system with refractive indices of biological tissue the error at the surfaces of the cylinders is much smaller since the jump in the refractive index is much smaller. The error is about an order of magnitude smaller, even though the wavelength in the biological tissue is smaller and thus there are less cells per wavelength.

In Figs. 2(a) and 2(g) one finds that there is a sizable $x$ component of the electric field even though the incident field has a vanishing $x$ component. In the calculations using the Helmholtz equation the $x$ component would vanish since in Cartesian coordinates the Helmholtz equation neglects the interaction between different components of the field as discussed in Sec. 1.1.

In the FDTD data in Fig. 3 there is also a region with large errors around the cylinders. In addition, there is an error at the left and right boundaries due to the numerical dispersion of the FDTD.

4.3. Far field

With the present algorithm and FDTD we calculate the far field at a distance of 1 m using a near to far field transformation. In the analytical calculations the far field is directly calculated using a series expansion for large distances. The results for both sets of refractive indices are shown in Figs. 4 and 5. In these figures the results from the present model and the results of FDTD are compared with the analytical results. Assuming that the analytical results are exact, it can be seen that the accuracy of the results from the FDTD calculations is much worse than the accuracy of the results from the present algorithm, even though the FDTD calculations have a much higher resolution.

Figures 4 and 5 show results with different accuracies in the far field. For the comparison of the calculation time of the far fields we aim to compare results with the same quality. For this task we define the accuracy of the far field as

$$A = \frac{1}{2} \text{median}_j \left( \frac{|E_{\parallel}^{\text{num}}(\phi_j)|^2 - |E_{\parallel}^{\text{analyt}}(\phi_j)|^2}{|E_{\parallel}^{\text{analyt}}(\phi_j)|^2} \right) + \frac{1}{2} \text{median}_j \left( \frac{|E_{\perp}^{\text{num}}(\phi_j)|^2 - |E_{\perp}^{\text{analyt}}(\phi_j)|^2}{|E_{\perp}^{\text{analyt}}(\phi_j)|^2} \right),$$

where $E_{\parallel}$ and $E_{\perp}$ are the electric fields parallel and perpendicular to the scattering plane, respectively. The $\phi_j$ are the angles at which the far field is calculated. Here, we use the median instead of the mean to be less sensitive to the large relative errors that occur at positions with very small field amplitude, i.e. the spikes that can be seen in Figs. 4 and 5.

The calculation time is measured on a system with two Intel(R) Xeon(R) E5-2640 v4 CPUs at 2.40 GHz with 10 cores each. The results of this measurement are depicted in Fig. 6. The measured values are fitted with a power law. It can be seen that for an accuracy of 0.1 the present algorithm is about a factor of 5000 faster than the FDTD simulations, depending on the simulated system.
4.4. **Assets and drawbacks of the algorithm**

As it can be seen from Fig. 6, compared to the FDTD the present algorithm performs better for the case of biomaterials than for the cylinders in vacuum. The reason for this is that the difference of the refractive indices in the cylinders and the medium is smaller in the biomaterials. This leads to the fact that smaller values of $\delta$ are possible (see Eq. (44)). From Eqs. (43) and (47) it can be seen that a lower value of $\delta$ directly leads to a larger value of $\gamma$ and thus to an increasing change of the field in each iteration. In general the algorithm is best suited for systems with a small variation in the refractive index.

5. **Conclusion**

In this work we extend the solution of the Helmholtz equation of Osnabrugge et al. to the full vectorial wave equation derived from Maxwell’s equations. The present algorithm calculates the solution of the wave equation in the frequency domain. It allows us to calculate the solution for arbitrary polarizations of the wave, scattering geometry, and refractive indices of the material while it shows its best performance for systems in which the difference between the refractive indices is small. Comparison of our algorithm to FDTD calculations shows better accuracy even for much larger cell sizes. For far fields of comparable quality we find that our algorithm is by almost four orders of magnitude faster.

**Acknowledgments**

The authors like to thank Dominik Reitzle and Dennis Müller for fruitful discussions.