

Subwellenlängenstrukturen zur Erzeugung allgemeiner Phasenfunktionen

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Physikalisch-Astronomischen Fakultät
der Friedrich-Schiller-Universität Jena

von M. Sc. Sebastian Linß
geboren am 25. August 1993 in Hansestadt Seehausen (Altmark)

Gutachter:

1. Prof. Dr. Uwe Zeitner, Jena/München
2. Prof. Dr. Norbert Lindlein, Erlangen
3. Prof. Dr. Thomas Zentgraf, Paderborn

Tag der Disputation: 23. Januar 2024

Sub wavelength structures for the generation of general phase functions

Optical technologies are important to modern society in many ways. Data transmission through optical fibers, lithographic production of semiconductor components, and sensors for autonomous driving are just a few examples. The general trend toward miniaturization also includes optical components, e. g. for smartphone cameras. The usage of diffractive optical elements (DOEs) offers possibilities for almost arbitrary manipulation of light within small available spaces. One popular approach is the use of metamaterials. Their operating principle is often based on sub-wavelength structures, which cannot be resolved by the light, and can therefore have properties that are impossible for classical materials. Here, mainly effective-index structures of dielectric materials have been considered. Depending on the geometry, light propagation through them can be described by an effective index of refraction.

A DOE can be characterized by its reflection and transmission function describing how amplitude and phase of the incident light are locally influenced. Computer-generated holograms (CGHs), a special type of DOEs, can be used to generate almost arbitrary intensity distributions by their diffraction pattern. Ideally, phase elements are used to maximize the efficiency. A CGH phase profile is typically calculated by an iterative Fourier transform algorithm (IFTA). A promising approach for the realization of a general phase function is the use of effective-index structures with a binary geometric profile, but the capability to generate arbitrary phase steps. This allows the design of high-efficient CGHs with comparatively low manufacturing effort.

The simulation of light propagation through DOEs can be quite challenging. Typically, small regions of individual structures are calculated using rigorous solvers of Maxwell's equations under periodic boundary conditions. The obtained results are then applied to the entire element, assuming that the local structure is periodically continued. For a general manipulation of light propagation, the use of different structures is unavoidable and thus the violation of local periodicity. For small deflection angles, the resulting inaccuracy of the calculated field is often acceptable. At larger deflection angles, the deviations increase and prevent a precise prediction of the optical function. A rigorous simulation of the entire element is often too difficult due to the enormous computational requirements. Existing approaches include the simulation of subdomains using

rigorous methods or the usage of neural networks, but refer so far only to problems with limited complexity, e. g. due to rotational symmetry. The aim of this work is to overcome the limitations of established algorithms. For this purpose, new methods for the optical simulation of general DOEs have been developed: On the one hand, the use of suitable approximations allows a significant reduction of the computational effort to solve Maxwell's equations. On the other hand, physical simulations can be bypassed by machine learning to predict the optical function. In addition, the application of the new methods to design and optimization tasks is demonstrated.

For the approximate calculation of light propagation in general microstructures, a further developed version of the beam propagation method (BPM) is presented, which is several orders of magnitude faster than rigorous methods, while achieving a high agreement to their results. The accuracy of the method decreases with increasing index contrast of the materials and larger propagation angles. However, reliable results can be produced for glass surfaces in air at moderate angles. Due to its speed, it allows the investigation of large simulation areas, for example to perform high resolution scattering simulations. Another new approach, called polynomial field approximation (PFA), has been developed to predict the optical behavior even more efficiently, particularly for regularly arranged microstructures like CGHs. Instead of the actual element, a limited set of training data with similar structures but random distribution is simulated. Based on this data, a model can be formed which assigns the expected transmission to the structures and their local environment. The resulting model can quickly predict the optical function of arbitrary distributions of the respective structures, especially for very large elements.

For the design of CGHs, the polynomial structure approximation (PSA) is introduced to solve the inverse problem of computing a structure generating a given field distribution. Reliable results could be achieved for binary elements, while the generalization to effective-index structures was not successful, because an arbitrary field does not necessarily have an exact counterpart among the available structures. The design results of established methods could be improved by simulating the farfield, e. g. by the PFA, identifying systematic deviations from the target distribution, and correcting them in a subsequent calculation. Furthermore, the IFTA was directly combined with the PFA, for binary elements also with the PSA. In this way, an improvement over the pre-corrected design could be observed, especially for effective-index CGHs. A binary beamsplitter was fabricated and experimentally investigated. The results confirm that the assumption of local periodicity is unsuitable for accurate predictions of the optical function at large deflection angles, while the PFA showed good agreement up to angles of about 50° . With vectorial training data, based on rigorous methods, even larger deflection angles can be reliably calculated by the PFA.

The presented methods can be used to optimize a variety of DOEs, for example metalenses or beamsplitters with high NA. In particular, the design of pattern generators for three-dimensional measurement of objects, face recognition, or camera calibration are promising application areas.

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen	5
2.1	Propagation von Licht	5
2.1.1	Maxwell-Gleichungen in optischen Materialien	5
2.1.2	Propagation im homogenen Raum	7
2.1.3	Berechnung des Fernfeldes	8
2.1.4	Periodische Strukturen	9
2.1.5	Polarisation	10
2.1.6	Energie elektromagnetischer Felder	10
2.2	Optische Simulation von Mikrostrukturen	11
2.2.1	Rigorese Simulationsverfahren	11
2.2.2	Semi-rigorese Simulationsverfahren	14
2.2.3	Analyse von optischen Elementen durch Näherungen	16
2.2.4	Vergleich des Rechenaufwandes	19
2.3	Design von diffraktiven optischen Elementen (DOEs)	20
2.3.1	Varianten der Berechnung	21
2.3.2	Computergenerierte Hologramme (CGHs)	22
2.3.3	Iterativer Fourier-Transformations-Algorithmus (IFTA)	25
3	Effiziente Simulation von optischen Mikrostrukturen	31
3.1	Modifikation der Beam Propagation Method	31
3.1.1	Anpassung an große Öffnungswinkel	31
3.1.2	Effiziente Berechnung	33
3.1.3	Randbedingungen	35
3.1.4	Grenzflächen zwischen Schichten	36
3.1.5	Bidirektionalität	37
3.2	Anwendungsbeispiele	38
3.2.1	Basiszellen von Effektiv-Index-Strukturen	38
3.2.2	Effektiv-Index-Metalinse	41

3.2.3	Polarisations-Strahlteiler	46
3.2.4	Streulichtanalyse für Beugungsgitter	50
4	Analyse der optischen Funktion durch Polynomapproximation	55
4.1	Approximation des Fernfeldes	55
4.1.1	Polynomielle Beschreibung des Nahfeldes	56
4.1.2	Superposition des Fernfeldes	57
4.2	Bestimmung der Polynomkoeffizienten	59
4.2.1	Idee der Approximation	59
4.2.2	Auswahl von Regressionsparametern	60
4.2.3	Berechnung der Modell-Parameter	62
4.3	Anwendungsbeispiele	65
4.3.1	Außeraxialer Strahlteiler	66
4.3.2	Gitterprojektion	68
4.3.3	Simulation von Siliziumstrukturen für Strahlteiler	71
4.3.4	Analyse eines Weitwinkel-Strahlteilers	74
5	Semi-rigoroses Design von CGHs	79
5.1	Anpassung des IFTA	79
5.1.1	Phasenprofile für gleichmäßige Fernfeldverteilungen	79
5.1.2	Modifikation der Zielverteilung	81
5.2	Design und experimentelle Untersuchung von binären CGHs	82
5.2.1	Bestimmung der Struktur zu einem Phasenprofil	82
5.2.2	Verwendung von Polynomapproximationen im IFTA	84
5.2.3	Berechnung eines Weitwinkel-Strahlteilers	85
5.2.4	Herstellung der Strahlteiler	89
5.2.5	Experimentelle Untersuchung	90
5.3	Design von CGHs mit Effektiv-Index-Strukturen	93
5.3.1	Nutzung von Polynomapproximationen bei mehrstufigen Elementen	93
5.3.2	Strahlteiler mit mehrstufigem Phasenprofil	94
6	Zusammenfassung und Ausblick	99
	Literaturverzeichnis	101
	Abkürzungsverzeichnis	111
	Symbolverzeichnis	113

A	Anhang	117
A.1	Mathematische Grundlagen	117
A.1.1	Differentialoperatoren	117
A.1.2	Spezielle Funktionen und Distributionen	118
A.1.3	Approximation von Funktionen durch Polynome	118
A.1.4	Fourier-Transformation	120
A.2	Simulation von optischen Feldern	123
A.2.1	Fernfeldberechnung	123
A.2.2	Beam Propagation Method	127
A.2.3	Polynomielle Approximation des Nahfeldes	131
A.2.4	Simulationsparameter	134
B	Liste der wissenschaftlichen Veröffentlichungen und Vorträge	137
C	Lebenslauf	139
D	Danksagung	141
E	Ehrenwörtliche Erklärung	143

1 Einleitung

Optische Technologien durchdringen die moderne Gesellschaft auf vielfältige Weise. Die Datenübertragung durch Glasfasern, lithografische Herstellung von Halbleiterbauteilen und Sensorik für autonomes Fahren sind nur einige Beispiele. Der allgemeine Trend zur Miniaturisierung von Komponenten umfasst dabei auch die optischen Bauteile, wofür die leistungsfähigen Kameras in Smartphones ein prägnantes Beispiel sind. Die Verwendung diffraktiver optischer Elemente (DOEs), deren Funktionsweise auf Beugungseffekten beruht, verspricht Möglichkeiten zur nahezu beliebigen Manipulation des Lichtes innerhalb kleinster Bauräume [1, 2]. Ein vielbeachteter Ansatz ist dabei die Nutzung von Metamaterialien oder -oberflächen. Deren Funktionsprinzip beruht häufig auf Subwellenlängenstrukturen, also Strukturen die kleiner als die Wellenlänge des Lichtes sind und deshalb durch dieses nicht aufgelöst werden. Sie können daher Eigenschaften besitzen, die für klassische Materialien unmöglich sind [3, 4]. Im Rahmen dieser Arbeit werden vor allem Effektiv-Index-Strukturen aus dielektrischen Materialien betrachtet. Abhängig von der Geometrie lässt sich die Lichtausbreitung durch diese Strukturen mit einem effektiven Brechungsindex zwischen dem des Dielektrikums und dem des umgebenden Mediums beschreiben [5].

Die Funktionsweise eines DOEs kann durch seine Reflexions- bzw. Transmissionsfunktion charakterisiert werden. Diese beschreibt, wie Amplitude und Phase des einfallenden Lichtes durch das Element lokal beeinflusst werden. Da die Amplitude ohne aktive Verstärkung nur abnehmen kann, ist eine Manipulation der Phase bei geringer Abschwächung der Amplitude von Interesse, um Elemente mit hoher Effizienz zu realisieren. Eine typische Anwendung für Metastrukturen ist die Erzeugung eines DOEs mit der Funktionsweise einer Linse, welches dann als Metalinse bezeichnet wird [6, 7]. Inzwischen existiert eine große Vielfalt derartiger Elemente mit beispielsweise einer hohen numerischen Apertur [8], schaltbarer Fokusslänge [9] oder korrigierter Dispersion [10, 11].

Eine spezielle Kategorie von DOEs sind computergenerierte Hologramme (CGHs) [12]. Deren Funktionsweise ist von klassischen Hologrammen inspiriert, welche die Amplitude und Phase eines realen Objektes fotografisch aufzeichnen und so eine Rekonstruktion dessen dreidimensionaler Beschaffenheit erlauben [13]. Die Berechnung der Transmissionsfunktion eines entsprechenden Hologramms ist allerdings nicht auf reale Objekte beschränkt. Vielmehr kann durch das Beugungsbild eines CGHs eine nahezu beliebige Intensitätsverteilung des Lichtes erzeugt werden. Voraussetzung ist die möglichst exakte Nachbildung der berechneten Transmission durch ein DOE, wobei zur Maximierung der Effizienz idealerweise auf Phasenelemente zurückgegriffen wird. CGHs

1 Einleitung

werden in vielen Bereichen angewendet, z. B. zur optischen Verifikation [14, 15], Vermessung von Freiform-Elementen [16] oder Musterprojektion für Gesichts- und Gestenerkennung [17].

Es existieren vielfältige Möglichkeiten zur Realisierung einer allgemeinen Phasenfunktion. Eine weite Verbreitung haben binäre Elemente, bei denen zwei Höhenstufen in einer Substratoberfläche erzeugt werden, die zwei verschiedenen Phasenwerten entsprechen. Derartige Strukturen lassen sich vergleichsweise einfach herstellen und durch Abformung günstig vervielfältigen. Sie sind aber in ihrer Effizienz beschränkt und erzeugen prinzipiell ein punktsymmetrisches Beugungsbild, was für viele Anwendungen unerwünscht ist. Zur Vermeidung dieser Effekte sind Elemente mit mehr als zwei Phasenstufen notwendig. Die Herstellung von Profilen mit mehreren Höhenstufen ist allerdings deutlich schwieriger [18]. Ein vielversprechender Ansatz ist die Nutzung von Effektiv-Index-Strukturen, welche zwar gemäß ihrer Geometrie ein binäres Profil haben, durch die Subwellenlängengröße der Strukturen aber beliebige Phasenstufen erzeugen können [19, 20]. Diese Vorgehensweise erlaubt das Design von CGHs mit hoher Effizienz, ohne dass die Herstellung wesentlich aufwendiger wird.

Die Simulation der Lichtausbreitung durch CGHs bzw. DOEs im Allgemeinen kann große Herausforderungen mit sich bringen. Typischerweise werden kleine Gebiete mit einzelnen Strukturen mithilfe von rigorosen Lösungsverfahren der Maxwell-Gleichungen berechnet, wobei man periodische Randbedingungen nutzt. Die gewonnenen Resultate werden dann auf das gesamte Element übertragen, unter der Annahme, dass die Struktur lokal periodisch fortgesetzt wird [21]. Für eine allgemeine Manipulation der Lichtausbreitung ist die Verwendung verschiedener Strukturen unvermeidlich und damit auch deren Aufeinandertreffen sowie die Verletzung der lokalen Periodizität. Bei kleinen Ablenkwinkeln ist die resultierende Ungenauigkeit der berechneten Feldverteilungen häufig akzeptabel und die Vorgehensweise kann erfolgreich zum Design verwendet werden. Degegen nehmen bei größeren Ablenkwinkeln die Abweichungen zu, sodass auf diese Weise keine präzise Vorhersage der optischen Funktion möglich ist. Eine rigorose Simulation des gesamten Elementes ist aufgrund des enormen Rechenbedarfs oft nur schwer möglich. Bestehende Lösungsansätze umfassen die Simulation einzelner Teilbereiche durch rigorose Verfahren [16, 22–24] oder die Nutzung neuronaler Netze [25, 26]. Diese beziehen sich bislang aber nur auf Probleme mit eingeschränkter Komplexität z. B. aufgrund von Rotationssymmetrie. Ziel dieser Arbeit ist die Überwindung der Limitationen etablierter Methoden. Dazu wurden neue Verfahren für die optische Simulation allgemeiner DOEs entwickelt: Zum Einen erlaubt die Nutzung geeigneter Approximationen eine erhebliche Reduktion des Rechenaufwandes zur Lösung der Maxwell-Gleichungen bei linearer Skalierung mit der Elementgröße, zum Anderen kann die physikalische Simulation durch maschinelles Lernen zur Vorhersage der optischen Funktion umgangen werden. Darüber hinaus wird die Anwendung der neuen Methoden für Design- und Optimierungsaufgaben demonstriert.

In Kap. 2 werden zunächst die theoretischen Grundlagen der Lichtausbreitung geschildert,

bevor ein Überblick über die hier verwendeten Simulationsverfahren gegeben wird. Außerdem werden Methoden zum allgemeinen Design von diffraktiven Elementen und speziell für CGHs dargelegt. Im folgenden Kap. 3 wird eine Weiterentwicklung der *Beam Propagation Method* (Strahlpropagationsmethode, BPM) vorgestellt, welche eine deutlich höhere Geschwindigkeit im Vergleich zu rigorosen Verfahren aufweist und gleichzeitig eine hohe Genauigkeit besitzt. Sie wird zur Analyse verschiedener diffraktiver Elemente genutzt, um die Ergebnisse mit den Resultaten rigoroser Berechnungen zu vergleichen. Ein weiterer neuer Ansatz, die *Polynomial Field Approximation* (polynomielle Approximation des Feldes, PFA), wird in Kap. 4 erläutert. Anstatt das eigentliche Element zu simulieren, werden Trainingsdaten mit gleichartigen Strukturen, aber zufälliger Verteilung erzeugt und deren Transmissionsverhalten simuliert, z. B. mittels der BPM. Auf Basis dieser Daten kann ein Modell gebildet werden, welches den Strukturen und ihrer lokalen Umgebung das zu erwartende Lichtfeld zuordnet. Das entstehende Modell kann die optische Funktion von beliebigen Verteilungen der Strukturen und insbesondere von sehr großen Elementen schnell vorhersagen. Die Anwendung der Methode wird für die Simulation verschiedener CGHs demonstriert. Aufgrund ihrer hohen Geschwindigkeit lässt sich die PFA auch für das Design von CGHs nutzen, siehe die Beschreibung in Kap. 5. Dabei werden zunächst binäre Elemente betrachtet und die Simulationsergebnisse mit der experimentellen Untersuchung eines hergestellten CGHs verglichen. Schließlich werden Möglichkeiten zum Design von CGHs aus Effektiv-Index-Strukturen unter Verwendung der geschilderten Simulationsverfahren vorgestellt. Abschließend werden die vorgestellten Inhalte in Kap. 6 zusammengefasst und es wird ein Ausblick auf mögliche Erweiterungen und Anwendungen gegeben.

2 Grundlagen

Das Verhalten elektromagnetischer Strahlung lässt sich grundsätzlich mithilfe der Theorie des Elektromagnetismus vorhersagen. Die Anwendung auf reale Probleme kann aber erhebliche Schwierigkeiten verursachen, sodass je nach Anwendungsfall verschiedene Herangehensweisen nötig sind, um Aussagen über die tatsächliche Ausbreitung des Lichtes treffen zu können. Hier werden zunächst die physikalischen Grundlagen elektromagnetischer Felder und deren Ausbreitung geschildert. Danach werden verschiedene Verfahren zur numerischen Berechnung derselben vorgestellt, bevor ein Überblick über existierende Verfahren zum Design von diffraktiven Elementen gegeben wird.

2.1 Propagation von Licht

Innerhalb der Theorie des Elektromagnetismus wird die Ausbreitung von Licht durch das Verhalten der Vektorfelder des elektrischen und magnetischen Feldes beschrieben [27, 28]. Die allgemeine Grundlage dafür bilden die *Maxwell-Gleichungen*, welche im Folgenden mit ihrer Anpassung auf die Anwendung in der Optik dargelegt werden. Weiterhin werden verschiedene Aspekte der Lichtausbreitung vorgestellt, die zur Untersuchung von DOEs relevant sind.

2.1.1 Maxwell-Gleichungen in optischen Materialien

In quellenfreien Medien erfolgt die Ausbreitung der Felder in Abhängigkeit von Ort $\mathbf{r} = (x, y, z)^\top$ und Zeit t gemäß den Maxwell-Gleichungen:

$$\nabla \times \mathcal{E}(\mathbf{r}, t) = -\frac{\partial}{\partial t} \mathcal{B}(\mathbf{r}, t) , \quad \nabla \cdot \mathcal{D}(\mathbf{r}, t) = 0 , \quad (2.1)$$

$$\nabla \times \mathcal{H}(\mathbf{r}, t) = \frac{\partial}{\partial t} \mathcal{D}(\mathbf{r}, t) , \quad \nabla \cdot \mathcal{B}(\mathbf{r}, t) = 0 . \quad (2.2)$$

Darin bezeichnen \mathcal{E} und \mathcal{D} die elektrische Feldstärke und Flussdichte sowie \mathcal{H} und \mathcal{B} die entsprechenden magnetischen Größen. Fettgedruckte Variablen kennzeichnen dreikomponentige Vektoren, z. B. $\mathcal{E} = (\mathcal{E}_x, \mathcal{E}_y, \mathcal{E}_z)^\top$. Die Differentialoperatoren werden durch den Nabla-Operator ∇ ausgedrückt. Eine explizite Darstellung ist im Anhang, Abschnitt A.1.1, zu finden. Mithilfe der Fourier-Transformation (FT) lassen sich die zeitabhängigen Größen in winkelfrequenzabhängige Größen überführen, für Details siehe Abschnitt A.1.4.1 im Anhang. Dabei werden aus den reellwertigen, zeitabhängigen Größen, wie $\mathcal{E}(\mathbf{r}, t)$, komplexwertige Größen, wie $\mathbf{E}(\mathbf{r}, \omega)$, in Abhängigkeit

2 Grundlagen

von der Kreisfrequenz ω . Man bezeichnet den Betrag auch als Amplitude bzw. das Argument als Phase der komplexen Feldkomponenten. Die Maxwell-Gleichungen nehmen dann folgende Gestalt an:

$$\nabla \times \mathbf{E}(\mathbf{r}, \omega) = i\omega \mathbf{B}(\mathbf{r}, \omega) , \quad \nabla \cdot \mathbf{D}(\mathbf{r}, \omega) = 0 , \quad (2.3)$$

$$\nabla \times \mathbf{H}(\mathbf{r}, \omega) = -i\omega \mathbf{D}(\mathbf{r}, \omega) , \quad \nabla \cdot \mathbf{B}(\mathbf{r}, \omega) = 0 . \quad (2.4)$$

Betrachtet man nur ein festes ω , so beschreiben diese Gleichungen ein monochromatisches Feld und die zeitabhängigen Größen ergeben sich beispielsweise als

$$\mathcal{E}(\mathbf{r}, t) = \Re \{ \mathbf{E}(\mathbf{r}, \omega) \exp(-i\omega t) \} \quad (2.5)$$

für das elektrische Feld. Das Feld schwingt demnach harmonisch und erreicht nach jeweils einer Periodendauer $T = 2\pi/\omega$ wieder denselben Zustand.

In isotropen und linearen Materialien sind \mathbf{E} und \mathbf{D} über die elektrische Feldkonstante ε_0 und die relative Permittivität ε direkt miteinander verknüpft. Analog gilt das durch die magnetische Feldkonstante μ_0 und die relative Permeabilität μ für \mathbf{H} und \mathbf{B} :

$$\mathbf{D}(\mathbf{r}, \omega) = \varepsilon_0 \varepsilon(\mathbf{r}, \omega) \mathbf{E}(\mathbf{r}, \omega) , \quad \mathbf{B}(\mathbf{r}, \omega) = \mu_0 \mu(\mathbf{r}, \omega) \mathbf{H}(\mathbf{r}, \omega) . \quad (2.6)$$

Für viele optische Materialien ist $\mu \approx 1$ und kann daher vernachlässigt werden. Damit ergibt sich die in der Optik übliche Formulierung der Maxwell-Gleichungen:

$$\nabla \times \mathbf{E}(\mathbf{r}, \omega) = i\omega \mu_0 \mathbf{H}(\mathbf{r}, \omega) , \quad \nabla \cdot (\varepsilon(\mathbf{r}, \omega) \mathbf{E}(\mathbf{r}, \omega)) = 0 , \quad (2.7)$$

$$\nabla \times \mathbf{H}(\mathbf{r}, \omega) = -i\omega \varepsilon_0 \varepsilon(\mathbf{r}, \omega) \mathbf{E}(\mathbf{r}, \omega) , \quad \nabla \cdot \mathbf{H}(\mathbf{r}, \omega) = 0 . \quad (2.8)$$

Durch Anwendung eines weiteren Rotationsoperators $\nabla \times$ auf den ersten Teil von Gl. (2.7) und Ausnutzung der Identität $\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \Delta \mathbf{E}$ erhält man die Wellengleichung für das elektrische Feld:

$$\Delta \mathbf{E}(\mathbf{r}, \omega) + \frac{\omega^2}{c_0^2} \varepsilon(\mathbf{r}, \omega) \mathbf{E}(\mathbf{r}, \omega) = \nabla(\nabla \cdot \mathbf{E}(\mathbf{r}, \omega)) = -\nabla \left(\frac{\nabla \varepsilon(\mathbf{r}, \omega)}{\varepsilon(\mathbf{r}, \omega)} \mathbf{E}(\mathbf{r}, \omega) \right) . \quad (2.9)$$

Dabei bezeichnet Δ den komponentenweise angewendeten Laplace-Operator. Es ist $c_0 = 1/\sqrt{\varepsilon_0 \mu_0}$ die Vakuumlichtgeschwindigkeit. Die während einer Schwingungsperiode bei dieser Geschwindigkeit zurückgelegte Wegstrecke definiert die Vakuumwellenlänge $\lambda_0 = c_0 T = 2\pi c_0 / \omega$. Die Wellengleichung für das Magnetfeld,

$$\Delta \mathbf{H}(\mathbf{r}, \omega) + \frac{\omega^2}{c_0^2} \varepsilon(\mathbf{r}, \omega) \mathbf{H}(\mathbf{r}, \omega) = -\frac{\nabla \varepsilon(\mathbf{r}, \omega)}{\varepsilon(\mathbf{r}, \omega)} \times (\nabla \times \mathbf{H}(\mathbf{r}, \omega)) , \quad (2.10)$$

kann auf analoge Weise gewonnen werden. Da \mathbf{E} und \mathbf{H} unmittelbar voneinander abhängen, ist es hinreichend, eine der beiden Größen zu bestimmen.

Aus den Maxwell-Gleichungen ergeben sich Übergangs- oder Kontinuitätsbedingungen elektromagnetischer Felder zwischen verschiedenen Medien, z. B. für eine Grenzfläche in der Ebene $z = 0$. Unabhängig von den auftretenden Materialien sind die tangentialen Komponenten des \mathbf{E} - und \mathbf{H} -Feldes sowie die normalen Komponenten des \mathbf{D} - und \mathbf{B} -Feldes stetig [28]. Für einen infinitesimalen Abstand δz gilt also z. B.

$$\begin{aligned} E_x(x, y, -\delta z, \omega) &= E_x(x, y, \delta z, \omega) , & H_x(x, y, -\delta z, \omega) &= H_x(x, y, \delta z, \omega) , \\ E_y(x, y, -\delta z, \omega) &= E_y(x, y, \delta z, \omega) , & H_y(x, y, -\delta z, \omega) &= H_y(x, y, \delta z, \omega) , \\ D_z(x, y, -\delta z, \omega) &= D_z(x, y, \delta z, \omega) , & B_z(x, y, -\delta z, \omega) &= B_z(x, y, \delta z, \omega) . \end{aligned} \quad (2.11)$$

2.1.2 Propagation im homogenen Raum

Im homogenen Raum ist die Permittivität nicht länger ortsabhängig, sodass die rechte Seite der Wellengleichung aufgrund des Divergenzgesetzes in Gl. (2.7) verschwindet. Damit erhält man die Wellengleichung für den homogenen Raum

$$\Delta \mathbf{E}(\mathbf{r}, \omega) + \frac{\omega^2}{c_0^2} \varepsilon(\omega) \mathbf{E}(\mathbf{r}, \omega) = 0 . \quad (2.12)$$

Darin sind die Vektorkomponenten des Feldes unabhängig voneinander und lassen sich ebenso durch die skalare Wellengleichung (*Helmholtz-Gleichung*)

$$\Delta u(\mathbf{r}, \omega) + \frac{\omega^2}{c_0^2} \varepsilon(\omega) u(\mathbf{r}, \omega) = 0 \quad (2.13)$$

vollständig beschreiben, worin u eine beliebige Feldkomponente repräsentiert. Die allgemeine Lösung dieser Gleichung lässt sich als Überlagerung von ebenen Wellen der Gestalt

$$u(\mathbf{r}, \omega) = u_0 \exp(\mathbf{i} \mathbf{k} \mathbf{r}) , \quad \text{mit } |\mathbf{k}|^2 = \varepsilon(\omega) k_0^2 , \quad (2.14)$$

darstellen. Dabei nennt man \mathbf{k} den Wellenvektor der ebenen Welle und $k_0 = \omega/c_0 = 2\pi/\lambda_0$ die Vakuumwellenzahl. Weiterhin ist $n(\omega) = \sqrt{\varepsilon}$ der Brechungsindex des Mediums und definiert die Wellenzahl $k(\omega) = n k_0 = |\mathbf{k}|$ sowie die Lichtgeschwindigkeit in einem Medium $c(\omega) = c_0/n$. Zur Erfüllung der Divergenzgleichungen müssen elektrisches und magnetisches Feld einer ebenen Welle $\mathbf{E}(\mathbf{r}, \omega) = \mathbf{E}_0 \exp(\mathbf{i} \mathbf{k} \mathbf{r})$ orthogonal zum Wellenvektor stehen:

$$\mathbf{E}_0 \cdot \mathbf{k} = 0 , \quad \mathbf{H}_0 \cdot \mathbf{k} = 0 . \quad (2.15)$$

Elektrisches und magnetisches Feld sind dabei mittels

$$\mathbf{E}_0 = -\frac{1}{c_0 \varepsilon_0 \varepsilon(\omega) k_0} \mathbf{k} \times \mathbf{H}_0 , \quad \mathbf{H}_0 = \frac{1}{c_0 \mu_0 k_0} \mathbf{k} \times \mathbf{E}_0 \quad (2.16)$$

direkt durch den Wellenvektor verknüpft.

2 Grundlagen

Ist eine monochromatische Feldverteilung $u(\mathbf{r}, \omega) = u(\mathbf{r})$ in einer Ebene $z = z_0$ bekannt, so kann die Zerlegung in ebene Wellen genutzt werden, um die Verteilung im gesamten Raum zu ermitteln. Zunächst verwendet man eine zweidimensionale (2D-) FT um das Winkelspektrum der Feldverteilung zu erhalten:

$$U(k_x, k_y; z_0) = \mathcal{F} \{u(x, y, z_0)\} , \quad (2.17)$$

wie in Gl. (A.33) im Anhang definiert. Die FT der Helmholtz-Gleichung (2.13) ist

$$\left(\frac{\partial^2}{\partial z^2} - k_x^2 - k_y^2 + \varepsilon k_0^2 \right) U(k_x, k_y; z) = 0 . \quad (2.18)$$

Die Lösungen dieser Gleichung lassen sich mittels der Übertragungsfunktion G als

$$U(k_x, k_y; z) = U(k_x, k_y; z_0) G(k_x, k_y; z - z_0) , \quad (2.19)$$

$$G(k_x, k_y; z - z_0) = \exp(ik_z(z - z_0)) , \quad \text{mit } k_z^2 = k^2 - k_x^2 - k_y^2 , \quad (2.20)$$

darstellen. Für $k_z > 0$ wird dabei eine in positiver z -Richtung propagierende Welle beschrieben, für $k_z < 0$ propagiert sie in entgegengesetzter Richtung. Ist $k_x^2 + k_y^2 > k^2$, so wird k_z eine imaginäre Größe und beschreibt eine evaneszente Welle, deren Amplitude in Propagationsrichtung exponentiell abklingt. Aus dem Winkelspektrum in einer beliebigen Ebene ergibt sich durch inverse FT mit

$$u(x, y, z) = \mathcal{F}^{-1} \{U(k_x, k_y; z)\} = \mathcal{F}^{-1} \{G(k_x, k_y; z - z_0) \mathcal{F} \{u(x, y, z_0)\}\} \quad (2.21)$$

wieder das Feld im Ortsraum. Diese Zerlegung wird als *Angular Spectrum of Plane Waves* (Winkelspektrum ebener Wellen, ASPW) bezeichnet. Sie ermöglicht die exakte Lösung der Wellengleichung im homogenen Raum. Im Allgemeinen sind unendlich viele Werte zur exakten Charakterisierung von U nötig. Im Falle einer periodischen Feldverteilung u ist U jedoch nur an diskreten Punkten von 0 verschieden. Zudem können evaneszente Ordnungen vernachlässigt werden, wenn die Distanz $|z - z_0|$ hinreichend groß ist. Dann genügt eine endliche Anzahl an Werten zur exakten Beschreibung des Feldes.

2.1.3 Berechnung des Fernfeldes

Für die Analyse von diffraktiven optischen Elementen (DOEs) ist häufig folgende Konfiguration von Bedeutung: Aus Richtung $z < 0$ trifft Licht auf ein Element, das sich an der Position $z = 0$ befindet. Von Interesse ist dann die Ausbreitung des transmittierten Lichtes im Halbraum $z > 0$ bzw. des reflektierten bei $z < 0$ insbesondere in Entfernungen, die deutlich größer als die Abmessungen des Elementes bzw. die Wellenlänge des Lichtes sind. Man spricht vom *Fernfeld*, falls die Fresnel-Zahl,

$$N_F = \frac{a^2}{\lambda z} = \frac{a^2 k}{2\pi z} , \quad (2.22)$$

deutlich kleiner als eins ist: $N_F \ll 1$. Dabei bezeichnet a die charakteristische Größe des Elementes. In nicht-paraxialer *Fraunhofer-Näherung* lässt sich das Fernfeld als

$$u(x, y, z) \approx -\frac{ikz}{2\pi r^2} \exp(ikr) U\left(k\frac{x}{r}, k\frac{y}{r}; 0\right) \quad (2.23)$$

berechnen [27]. Eine Herleitung ist in Abschnitt A.2.1.1 im Anhang dargestellt. Die Feldverteilung im Fernfeld ist demnach direkt proportional zur FT des Feldes unmittelbar am Element, weshalb es in vielen Fällen genügt, die reine FT zu betrachten und vereinfachend als Fernfeld zu bezeichnen.

2.1.4 Periodische Strukturen

Eine periodische Struktur mit Perioden von w_x und w_y in x - bzw. y -Richtung ist durch eine Brechungsindexverteilung der Form

$$n(x, y, z) = n(x + K_x w_x, y + K_y w_y, z), \quad K_x, K_y \in \mathbb{N} \quad (2.24)$$

gekennzeichnet, siehe Abb. 2.1(a). Trifft eine ebene Welle mit dem Wellenvektor $\mathbf{k} = (k_x, k_y, k_z)^\top$ aus einem Bereich mit Brechungsindex n_S auf eine solche Struktur, so entstehen durch Interferenz neue Wellen, die sich sowohl in Propagationsrichtung, ins Medium mit Brechungsindex n_C , als auch in entgegengesetzter Richtung ausbreiten. Die entstehenden Beugungsordnungen (m_x, m_y) haben Wellenvektoren von

$$\mathbf{k}_{m_x m_y}^\pm = \begin{pmatrix} k_{x, m_x} \\ k_{y, m_y} \\ k_{z, m_x m_y}^\pm \end{pmatrix} = \begin{pmatrix} k_x + \frac{2\pi m_x}{w_x} \\ k_y + \frac{2\pi m_y}{w_y} \\ \pm \sqrt{n_{C/S}^2 k_0^2 - k_{x, m_x}^2 - k_{y, m_y}^2} \end{pmatrix}. \quad (2.25)$$

Das obere Vorzeichen steht dabei für transmittierte, das untere für reflektierte Wellen. Ist die Struktur invariant in y -Richtung $n(x, y, z) = n(x, z)$ und der Einfall erfolgt in der xz -Ebene, siehe Abb. 2.1(b,c), ergibt sich die *Gittergleichung* [28]

$$n_{C/S} \sin \vartheta_{C/S, m_x} = n_S \sin \vartheta + \frac{m_x \lambda_0}{w_x}. \quad (2.26)$$

für $k_x = nk_0 \sin \vartheta$. Der Index S ist dabei für die transmittierten, der Index C für die reflektierten Anteile zu beachten.

Nur die Ordnungen, für die $|k_{x, m_x}| < n_{C/S} k_0$ gilt, können im homogenen Raum propagieren. Wenn die Periode w_x hinreichend klein ist,

$$w_x < \frac{\lambda_0}{n_{C/S} + n_S |\sin \vartheta|}, \quad (2.27)$$

trifft das nur für die Ordnung $m_x = 0$ zu. In diesem Fall findet keine Beugung mehr statt. Das Licht propagiert durch die periodische Struktur wie durch ein homogenes Medium mit einem Brechungsindex n_{eff} , dem sogenannten *effektiven Brechungsindex* [5, 29]. Dieser hängt von der genauen Form der periodischen Struktur ab und liegt in der Regel zwischen den Brechungsindizes der vorkommenden Materialien. Derartige Strukturen werden als *Effektiv-Index-Strukturen* bzw. *Effektiv-Medium* bezeichnet.

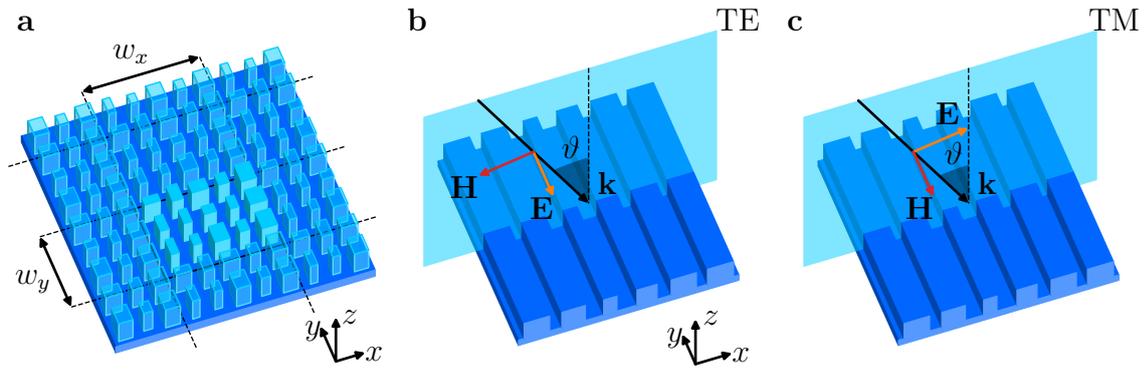


Abb. 2.1: Periodische Struktur mit Hervorhebung einer Elementarzelle (a), in y -Richtung invariante Struktur mit Zerlegung der Felder in TE- (b) und TM-Polarisation (c) und Darstellung der Einfallsebene in hellblau.

2.1.5 Polarisation

Bei Invarianz der betrachteten Felder und Materialeigenschaften in einer Raumrichtung, hier der y -Richtung, vereinfachen sich die Gleichungen zur Beschreibung der Propagation und die Felder lassen sich durch zwei unabhängige Komponenten darstellen, z. B. mit der Wellengleichung (2.9) bzw. (2.10):

$$\Delta E_y(\mathbf{r}, \omega) + \frac{\omega^2}{c_0^2} \varepsilon(\mathbf{r}, \omega) E_y(\mathbf{r}, \omega) = 0, \quad (2.28)$$

$$\Delta H_y(\mathbf{r}, \omega) + \frac{\omega^2}{c_0^2} \varepsilon(\mathbf{r}, \omega) H_y(\mathbf{r}, \omega) = \frac{\nabla \varepsilon(\mathbf{r}, \omega) \cdot \nabla H_y(\mathbf{r}, \omega)}{\varepsilon(\mathbf{r}, \omega)}. \quad (2.29)$$

Die anderen Feldkomponenten ergeben sich über die Maxwell-Gleichungen direkt aus E_y und H_y . Man bezeichnet den Anteil des Feldes, für den das elektrische Feld senkrecht zur Einfallsebene steht, als TE- (transversal elektrisch) polarisiert und denjenigen, für den das magnetische Feld senkrecht zur Einfallsebene steht, als TM- (transversal magnetisch) polarisiert, was hier gerade den Komponenten E_y bzw. H_y entspricht, siehe Abb. 2.1(b,c). Diese Zerlegung einer ebenen Welle kann im homogenen Raum in Bezug auf eine Einfallsebene stets durchgeführt werden. Bei Invarianz in einer Raumrichtung verhalten sich beide Polarisationen auch an einer Grenzfläche unabhängig voneinander. Im Allgemeinen findet aber eine Kopplung der Komponenten statt.

2.1.6 Energie elektromagnetischer Felder

Der *Poynting-Vektor* \mathcal{S} beschreibt die Leistungsdichte des elektromagnetischen Feldes [28] und ist definiert als

$$\mathcal{S}(\mathbf{r}, t) = \mathcal{E}(\mathbf{r}, t) \times \mathcal{H}(\mathbf{r}, t). \quad (2.30)$$

Diese Größe kann aufgrund der hohen Frequenz von optischen Feldern typischerweise nicht direkt detektiert werden, sodass man stattdessen deren zeitliches Mittel z. B. über eine Periode als Maß

verwendet. Für ein monochromatisches Feld ergibt sich:

$$\langle \mathcal{S}(\mathbf{r}, t) \rangle_T = \frac{1}{T} \int_0^T \mathcal{S}(\mathbf{r}, t) dt = \frac{1}{2} \Re \{ \mathbf{E}(\mathbf{r}, \omega) \times \mathbf{H}^*(\mathbf{r}, \omega) \} . \quad (2.31)$$

Darin bezeichnet \mathbf{H}^* die komplexe Konjugation von \mathbf{H} . Der Betrag des zeitlichen Mittelwertes wird als *Bestrahlungsstärke* oder *Intensität* $I = |\langle \mathcal{S} \rangle_T|$ bezeichnet. Für eine propagierende ebene Welle $\mathbf{E}(\mathbf{r}, \omega) = \mathbf{E}_0 \exp(i\mathbf{k}\mathbf{r})$, vereinfacht sich die Intensität zu

$$I = \frac{\varepsilon_0 c_0 n}{2} |\mathbf{E}_0|^2 = \frac{\mu_0 c_0}{2n} |\mathbf{H}_0|^2 . \quad (2.32)$$

Die Leistung, die auf eine Fläche trifft und damit eine mögliche Messgröße darstellt, lässt sich als

$$P = \int_A I(\mathbf{r}) \mathbf{e} \cdot \mathbf{n}(\mathbf{r}) dA \quad (2.33)$$

berechnen. Darin ist \mathbf{e} der Einheitsvektor in Ausbreitungsrichtung und \mathbf{n} der Normalenvektor der Oberfläche A .

Als Beugungseffizienz η einer Beugungsordnung (m_x, m_y) wird das Verhältnis aus eingestrahelter und in diese Ordnung gebeugter Leistung bezeichnet:

$$\eta_{m_x m_y} = \frac{P_{m_x m_y}}{P_{\text{in}}} . \quad (2.34)$$

Für ein diffraktives Element in der Ebene $z = 0$ entspricht das dem Verhältnis der Intensitäten,

$$\eta_{m_x m_y} = \frac{|e_{z, m_x m_y}| I(k_{x, m_x}, k_{y, m_y})}{|e_{z, \text{in}}| I_{\text{in}}} = \frac{|k_{z, m_x m_y}| |\mathbf{E}_0(k_{x, m_x}, k_{y, m_y})|^2}{|k_{z, \text{in}}| |\mathbf{E}_{\text{in}}|^2} , \quad (2.35)$$

unter Beachtung der jeweiligen Ausbreitungsrichtung.

2.2 Optische Simulation von Mikrostrukturen

Für die Berechnung der Ausbreitung von Licht gibt es eine Vielzahl von Varianten. Diejenigen, welche die Maxwell-Gleichungen ohne weitere Näherungen lösen, werden als rigorose Verfahren bezeichnet. Sie sind typischerweise durch einen hohen numerischen Aufwand gekennzeichnet. Andere Verfahren nutzen in unterschiedlichem Maß verschiedene Näherungen, um die Rechenzeit zu reduzieren. Unter diesen sollen die semi-rigorosen Verfahren hervorgehoben werden, welche insbesondere die Beugungseffekte an Subwellenlängenstrukturen noch gut modellieren können.

2.2.1 Rigorose Simulationsverfahren

Zwei näherungsfreie Verfahren, die im Rahmen dieser Arbeit zum Vergleich der Simulationen genutzt wurden, sind die *Finite Difference Time Domain Method* (Finite-Differenzen-Methode im Zeitbereich, FDTD) und die *Rigorous Coupled Wave Analysis* (rigorose Analyse gekoppelter Wellen, RCWA). Beide zeichnen sich durch eine hohe Genauigkeit der simulierten Felder aus.

2.2.1.1 Finite Difference Time Domain Method

In der FDTD verwendet man die Maxwell-Gleichungen im Zeitbereich (2.1), (2.2) für optische Materialien:

$$\frac{\partial}{\partial t} \mathcal{H}(\mathbf{r}, t) = -\frac{1}{\mu_0} \nabla \times \mathcal{E}(\mathbf{r}, t) , \quad (2.36)$$

$$\frac{\partial}{\partial t} \mathcal{E}(\mathbf{r}, t) = \frac{1}{\varepsilon_0 \varepsilon(\mathbf{r})} \nabla \times \mathcal{H}(\mathbf{r}, t) - \mathbf{j}_{\text{makr}}(\mathbf{r}, t) . \quad (2.37)$$

Die Stromdichte \mathbf{j}_{makr} dient dabei zur Erzeugung der Felder. Die Berechnung erfolgt mittels der direkten Diskretisierung von Ort und Zeit. So wird aus der Differentialgleichung für die \mathcal{H}_x -Komponente,

$$\frac{\partial}{\partial t} \mathcal{H}_x(x, y, z, t) = -\frac{1}{\mu_0} \left(\frac{\partial}{\partial y} \mathcal{E}_z(x, y, z, t) - \frac{\partial}{\partial z} \mathcal{E}_y(x, y, z, t) \right) , \quad (2.38)$$

die diskretisierte Form

$$\begin{aligned} \mathcal{H}_x(x, y, z, t + \delta t) = \mathcal{H}_x(x, y, z, t) - \frac{\delta t}{\mu_0} & \left(\frac{\mathcal{E}_z(x, y + \delta y, z, t) - \mathcal{E}_z(x, y, z, t)}{\delta y} \right. \\ & \left. - \frac{\mathcal{E}_y(x, y, z + \delta z, t) - \mathcal{E}_y(x, y, z, t)}{\delta z} \right) . \end{aligned} \quad (2.39)$$

Somit können aus den bekannten Feldern zu einem Zeitpunkt t die Felder in einem nächsten Zeitschritt errechnet werden. Durch die Anwendung von symmetrischen Differenzenquotienten werden Diskretisierungsfehler minimiert [30]. Üblicherweise setzt man zu Beginn einer Simulation sämtliche Felder auf null und gibt eine Stromdichte zur Anregung vor. Dann läuft die Simulation entweder bis zu einem eingeschwungenen Zustand oder man ermittelt mit einer FT das Verhalten bei einer bestimmten Frequenz. Die zeitliche Diskretisierung muss dabei klein genug gewählt werden, um die Ausbreitung des Lichtes vollständig zu erfassen. Dazu muss

$$\delta t^2 \leq \frac{1}{c^2} \left(\frac{1}{\delta x^2} + \frac{1}{\delta y^2} + \frac{1}{\delta z^2} \right)^{-1} \quad (2.40)$$

gelten. Bei einer Verfeinerung der räumlichen Diskretisierung müssen also auch die Zeitschritte entsprechend verkleinert werden. Zudem hat die Methode einen großen Speicherbedarf, denn die Feldkomponenten aller Raumpunkte werden in jedem Schritt benötigt. Für die einzelnen Raumrichtungen können verschiedene Randbedingungen gewählt werden, welche die Behandlung der Felder an den Grenzen des Simulationsgebietes festlegen. Im Rahmen dieser Arbeit handelt es sich dabei in der Regel um periodische Randbedingungen oder sogenannte *perfectly matched layer* (perfekt angepasste Schichten, PML); das sind künstliche Materialien, die durch die Wahl ihres Brechungsindex eine perfekt absorbierende Grenzschicht nachbilden.

2.2.1.2 Rigorous Coupled Wave Analysis

Die RCWA löst die Maxwell-Gleichungen im Frequenzraum, (2.7) und (2.8), für periodische Strukturen. Dadurch können alle auftretenden Felder und Materialeigenschaften in Fourier-Reihen dargestellt werden [31–34]. So wird z. B. das elektrische Feld im homogenen Raum mit Brechungsindex n als

$$\mathbf{E}(\mathbf{r}) = \sum_{m_x, m_y} \mathbf{E}_{m_x m_y}^+ \exp(\mathbf{i} \mathbf{k}_{m_x m_y}^+ \mathbf{r}) + \sum_{m_x, m_y} \mathbf{E}_{m_x m_y}^- \exp(\mathbf{i} \mathbf{k}_{m_x m_y}^- \mathbf{r}) \quad (2.41)$$

dargestellt. Dabei ist $\mathbf{k}_{m_x m_y}^\pm$ der Wellenvektor gemäß Gl. (2.25) für eine Strukturperiode von w_x bzw. w_y , wobei

$$k_{z, m_x m_y}^\pm = \begin{cases} \pm \sqrt{n^2 k_0^2 - k_{x, m_x}^2 - k_{y, m_y}^2}, & \text{für } k_{x, m_x}^2 + k_{y, m_y}^2 \leq n^2 k_0^2, \\ \pm i \sqrt{k_{x, m_x}^2 + k_{y, m_y}^2 - n^2 k_0^2}, & \text{sonst.} \end{cases} \quad (2.42)$$

Mit diesem Ansatz wird die Überlagerung aus allen für diese Periode möglichen vorwärts (+) und rückwärts (−) propagierenden Feldern beschrieben, wobei sowohl propagierende Anteile (k_z reell), als auch evaneszente Anteile (k_z imaginär) berücksichtigt werden. Bei einem inhomogenen Brechungsindex lässt sich das Feld im Allgemeinen als Summe von Fourier-Komponenten $\tilde{\mathbf{E}}_{m_x m_y}$,

$$\mathbf{E}(\mathbf{r}) = \sum_{m_x, m_y} \tilde{\mathbf{E}}_{m_x m_y}(z) \exp(\mathbf{i}(x k_{x, m_x} + y k_{y, m_y})) \quad (2.43)$$

darstellen und auf analoge Weise gleichfalls das Magnetfeld auf Basis von $\tilde{\mathbf{H}}_{m_x m_y}(z)$. Ebenso lässt sich die Permeabilität als Fourier-Reihe ausdrücken. Dabei nimmt man typischerweise deren Invarianz in z -Richtung an: $\varepsilon(\mathbf{r}) = \varepsilon(x, y)$. Um beliebige Strukturen beschreiben zu können, zerlegt man den Simulationsbereich gegebenenfalls in Schichten, in denen diese Annahme näherungsweise erfüllt ist und erhält damit

$$\varepsilon(x, y) = \sum_{m_x, m_y} \varepsilon_{m_x m_y} \exp(\mathbf{i}(x k_{x, m_x} + y k_{y, m_y})) \quad (2.44)$$

Die Maxwell-Gleichungen (2.7) und (2.8) gelten dann für jeden der Fourier-Terme, z. B. für die x -Komponente der Gleichungen:

$$i k_{y, m_y} \tilde{E}_{z, m_x m_y} - \frac{\partial}{\partial z} \tilde{E}_{y, m_x m_y} = i \omega \mu_0 \tilde{H}_{x, m_x m_y} \quad (2.45)$$

$$i k_{y, m_y} \tilde{H}_{z, m_x m_y} - \frac{\partial}{\partial z} \tilde{H}_{y, m_x m_y} = -i \omega \varepsilon_0 \sum_{m'_x, m'_y} \varepsilon_{m_x - m'_x, m_y - m'_y} \tilde{E}_{x, m'_x m'_y} \quad (2.46)$$

Es entsteht ein gekoppeltes Differentialgleichungssystem für dessen Lösung zunächst die Eigenwerte der Koeffizientenmatrix bestimmt werden müssen. Daraus erhält man die Fundamentalmoden der inhomogenen Schicht, analog zu den ebenen Wellen im homogenen Raum. Die zugehörigen Amplituden ergeben sich aus den Übergangsbedingungen (2.11) an den Grenzflächen zwischen den

jeweiligen Schichten. Diese müssen wiederum für jeden der Fourier-Terme separat erfüllt werden. Als Lösung des entstehenden Gleichungssystems erhält man insbesondere direkt die Amplituden der transmittierten bzw. reflektierten Ordnungen im homogenen Raum, $\mathbf{E}_{m_x m_y}^\pm$.

Der Vorteil der RCWA besteht darin, dass sie die Maxwell-Gleichungen bereits mit einer kleinen Anzahl an Ordnungen annähernd exakt lösen kann: Dazu müssen alle propagierenden und einige evaneszente Ordnungen berücksichtigt werden. Alle weiteren Ordnungen fallen in ihrer Amplitude zu schnell ab, um einen Beitrag zum Ergebnis zu liefern. Das macht die RCWA insbesondere für die Berechnung von eindimensionalen (1D-) Gittern mit Perioden von höchstens einigen Wellenlängen zu einem sehr effizienten Werkzeug. Andererseits werden für größere Simulationsgebiete zunehmend mehr Ordnungen benötigt, was die Lösung des Eigenwertproblems sehr aufwendig macht. Insbesondere für 2D-Strukturen bleibt die praktische Anwendbarkeit daher auf kleine Simulationsgebiete beschränkt.

2.2.2 Semi-rigorese Simulationsverfahren

Unter bestimmten Voraussetzungen lassen sich die Maxwell-Gleichungen vereinfachen, sodass sie mit geringerem Zeitaufwand dennoch exakt oder näherungsweise gelöst werden können. Eine typische Annahme ist dabei die Invarianz des Brechungsindex in Ausbreitungsrichtung des Lichtes, was hier die z -Richtung sei. Für ein endliches optisches Element lässt sich diese Bedingung zumindest stückweise erfüllen, indem man die Betrachtung in Schichten zerlegt, in denen die Bedingung jeweils erfüllt ist. Dann kann die Ausbreitung des Lichtes zumindest innerhalb der Schichten gut beschrieben werden, wobei auch die Übergänge zwischen den Schichten durch geeignete Verfahren einbezogen werden können. In diesem Abschnitt sollen beispielhaft die BPM und die *Wave Propagation Method* (Wellenpropagationsmethode, WPM) vorgestellt werden.

2.2.2.1 Beam Propagation Method (BPM)

Die BPM nutzt eine Diskretisierung der Wellengleichung, um die Propagation in Ausbreitungsrichtung zu beschreiben. Die gezeigte Herleitung folgt der vektoriellen Darstellung in Ref. [35]. Ausgehend von der Wellengleichung (2.9) betrachtet man zunächst nur die transversalen Komponenten, $\mathbf{E}_\perp = (E_x, E_y)^\top$, zur genauen Notation siehe Gl. (A.7) im Anhang:

$$\Delta \mathbf{E}_\perp + n^2 k_0^2 \mathbf{E}_\perp = -\nabla_\perp \left(\frac{1}{n^2} \left(\nabla_\perp n^2 \cdot \mathbf{E}_\perp + \frac{\partial n^2}{\partial z} E_z \right) \right), \quad (2.47)$$

worin $\partial n^2 / \partial z$ allerdings aufgrund der Invarianz in z -Richtung verschwindet. Wählt man nun einen Referenzindex \bar{n} zur Einführung einer langsam veränderlichen Amplitude $\mathbf{E}_\perp = \hat{\mathbf{E}}_\perp \exp(i\bar{n}k_0 z)$, so ergibt sich

$$\frac{\partial^2}{\partial z^2} \hat{\mathbf{E}}_\perp + 2i\bar{n}k_0 \frac{\partial}{\partial z} \hat{\mathbf{E}}_\perp + \Delta_\perp \hat{\mathbf{E}}_\perp + (n^2 - \bar{n}^2) k_0^2 \hat{\mathbf{E}}_\perp + \nabla_\perp \left(\frac{1}{n^2} \nabla_\perp n^2 \cdot \hat{\mathbf{E}}_\perp \right) = 0. \quad (2.48)$$

Im Rahmen dieser Arbeit wird der Referenzindex \bar{n} stets als arithmetisches Mittel des höchsten bzw. niedrigsten Brechungsindex im aktuellen Simulationsbereich gewählt. Die obige Gleichung lässt sich als System gekoppelter Differentialgleichungen schreiben:

$$\frac{\partial^2}{\partial z^2} \begin{pmatrix} \hat{E}_x \\ \hat{E}_y \end{pmatrix} + 2i\bar{n}k_0 \frac{\partial}{\partial z} \begin{pmatrix} \hat{E}_x \\ \hat{E}_y \end{pmatrix} + 2\bar{n}^2 k_0^2 \begin{pmatrix} \mathcal{A}_{xx} & \mathcal{A}_{xy} \\ \mathcal{A}_{yx} & \mathcal{A}_{yy} \end{pmatrix} \begin{pmatrix} \hat{E}_x \\ \hat{E}_y \end{pmatrix} = 0 . \quad (2.49)$$

Darin sind die Komponenten von \mathcal{A} Differentialoperatoren 2. Ordnung, die in Abschnitt A.2.2.1 im Anhang detailliert angegeben werden. In Umgebungen mit homogenem Brechungsindex handelt es sich bei \mathcal{A}_{xx} und \mathcal{A}_{yy} im Wesentlichen um Laplace-Operatoren, andernfalls garantieren sie die Erfüllung der Übergangsbedingungen (2.11). Für eine semi-vektorielle Betrachtung vernachlässigt man die Komponenten \mathcal{A}_{xy} , \mathcal{A}_{yx} . Damit entstehen zwei unabhängige Gleichungen für die x - und y -Komponente, die deren unterschiedliches Verhalten beschreiben können, während jegliche Kopplung vernachlässigt wird. Eine ähnliche Gleichung kann für das magnetische Feld gewonnen werden, siehe ebenfalls Abschnitt A.2.2.1.

Zur Lösung der Gleichung kann diese formal separiert werden, wie hier für \hat{E}_y :

$$\left(\frac{\partial}{\partial z} - i\bar{n}k_0 (\sqrt{1 + 2\mathcal{A}_{yy}} - 1) \right) \left(\frac{\partial}{\partial z} + i\bar{n}k_0 (\sqrt{1 + 2\mathcal{A}_{yy}} + 1) \right) \hat{E}_y = 0 , \quad (2.50)$$

worin der erste Faktor ein vorwärts (in z -Richtung) propagierendes Feld und der zweite ein rückwärts propagierendes Feld beschreibt. In paraxialer Näherung setzt man $\sqrt{1 + 2\mathcal{A}_{yy}} \approx 1 + \mathcal{A}_{yy}$, was der ursprünglichen Formulierung der BPM entspricht [36]. Damit ergibt sich für den vorwärts propagierenden Fall:

$$\frac{\partial}{\partial z} \hat{E}_y = i\bar{n}k_0 \mathcal{A}_{yy} \hat{E}_y . \quad (2.51)$$

Diese Gleichung lässt sich durch Diskretisierung iterativ lösen, indem auf der rechten Seite das Feld an der Position z eingesetzt wird, um das Feld an einer Position $z + \delta z$ zu berechnen. Im Unterschied zur FDTD muss man dabei in jedem Schritt nur die Felder in einer Ebene betrachten und speichern, was den Rechenaufwand deutlich reduziert. Zudem ist durch die Verwendung einer langsam veränderlichen Amplitude eine gröbere Diskretisierung in Propagationsrichtung bei gleicher Genauigkeit möglich. Limitiert wird die Methode durch die Gültigkeit der paraxialen Näherung [37]: Einerseits sind nur begrenzte transversale Änderungen des Feldes erlaubt, andererseits müssen die Indexkontraste gering sein, damit der Ausdruck $n - \bar{n}$ für alle auftretenden Materialien hinreichend klein ist. Diese Näherung kann zur Erweiterung des Einsatzbereiches der Methode durch genauere Approximationen ersetzt werden, was in Kap. 3 ausführlich beschrieben wird.

2.2.2.2 Wave Propagation Method

In der WPM wird die ASPW für inhomogene Indexverteilungen erweitert. Man betrachtet die inhomogene Wellengleichung (2.9), allerdings unter Vernachlässigung der rechten Seite,

$$\Delta u + k_0^2 n^2 u = 0, \quad (2.52)$$

und erhält so auch hier eine skalare Form. Betrachtet man $n(x, y)$ lokal als konstant, lässt sich die Gleichung analog zur ASPW lösen, wobei allerdings im Vergleich zu Gl. (2.21) die Übertragungsfunktion G von x und y abhängt [38]:

$$u(x, y, z + \delta z) = \mathcal{F}^{-1} \{ G(k_x, k_y; x, y, \delta z) \mathcal{F} \{ u(x, y, z) \} \}, \quad (2.53)$$

mit

$$G(k_x, k_y; x, y, \delta z) = \exp \left(i \delta z \sqrt{n^2(x, y) k_0^2 - k_x^2 - k_y^2} \right). \quad (2.54)$$

Wählt man δz hinreichend klein, wird damit die Wellengleichung in Umgebungen, in denen n sich nicht ändert, quasi exakt gelöst. Abweichungen ergeben sich, wenn Bereiche mit verschiedenem n aufeinandertreffen. In einer vektoriellen Formulierung können Grenzflächen zwischen einzelnen Schichten berücksichtigt werden, inklusive des unterschiedlichen Verhaltens der Polarisationskomponenten [39]. Allerdings wird auch dabei für die Propagation innerhalb einer Schicht nur die skalare Wellengleichung gelöst. Es ist möglich, die WPM in einer effizienten Variante zu implementieren, welche die Anwendung der inversen FT nur einmal für jedes Material anstatt für jeden Raumpunkt erfordert [40, 41].

Die Methode ist durch die Verwendung der FT deutlich schneller als rigorose Verfahren. Sie ist besonders zur wellenoptischen Analyse von Strukturen geeignet, die größer als die Wellenlänge sind, wie z. B. Mikrolinsen. Bei Subwellenlängenstrukturen spielen die vernachlässigten polarisationsabhängigen Effekte eine zu große Rolle und führen zu Abweichungen.

2.2.3 Analyse von optischen Elementen durch Näherungen

Um die Funktionsweise von DOEs auf einfache Weise zu beschreiben, kann auf verschiedene Näherungen zurückgegriffen werden. Die Beschreibung erfolgt hier für skalare Felder, da unter Vernachlässigung von Kopplungseffekten die jeweiligen Vektor- bzw. Polarisationskomponenten unabhängig voneinander sind. Ein simpler Ansatz ist die *Thin Element Approximation* (Dünne-Elemente-Approximation, TEA). Dabei nimmt man an, dass sich eine Struktur mit der lokalen Dicke d lokal genauso verhält, wie eine unendlich ausgedehnte Schicht derselben Dicke, die zudem dünn genug ist, dass Beugungseffekte innerhalb der Schicht keine Rolle spielen. Dann kann das Feld nach der Schicht als Multiplikation des Eingangsfeldes mit einer Transmissionsfunktion

$t_d(x, y)$,

$$u(x, y, d) = t_d(x, y)u(x, y, 0) , \quad (2.55)$$

beschrieben werden. Im einfachsten Fall, unter Vernachlässigung von Absorption, ergibt sich die Transmissionsfunktion für die senkrechte Propagation einer ebenen Welle durch die Schicht als

$$t_d(x, y) = \exp(in(x, y)k_0d) \quad (2.56)$$

und beschreibt die akkumulierte Phase aufgrund der Propagationslänge. Es lassen sich ebenfalls Aussagen über die Effizienz treffen, indem obige Formel um die entsprechenden Fresnel-Koeffizienten t_F erweitert wird, z. B. im Falle einer einfachen Transmission:

$$t_d(x, y) = t_F(n_S, n(x, y)) t_F(n(x, y), n_C) \exp(in(x, y)k_0d) , \quad \text{mit } t_F(n_1, n_2) = \frac{2n_1}{n_1 + n_2} . \quad (2.57)$$

Hierbei sei n_S der Brechungsindex im Bereich $z < 0$ und n_C der Brechungsindex für $z > d$, analog zu Abschnitt 2.1.4. Die Fresnel-Koeffizienten t_F bzw. r_F bestimmen den Anteil des transmittierten bzw. reflektierten Lichtes beim Auftreffen auf eine ebene Grenzfläche zwischen Medien mit verschiedenem Brechungsindex. Die Gleichung lässt sich für Mehrfachreflexionen erweitern und erlaubt die Berücksichtigung eines nicht-senkrechten Einfallswinkels. Ein typischer Anwendungsfall sind Höhenprofile: Eine Schicht habe für $z < d(x, y)$ einen Index n_S und für $z > d(x, y)$ dann n_C , wobei $0 \leq d(x, y) \leq \hat{d}$. Setzt man beide Bereiche zusammen, erhält man die Transmissionsfunktion

$$t_{\hat{d}}(x, y) = t_F(n_S, n_C) \exp(in_C k_0 \hat{d} + i(n_S - n_C)k_0 d(x, y)) . \quad (2.58)$$

Die Anwendbarkeit der TEA ist allerdings immer durch die Annahme der beugungsfreien Propagation durch die Schicht limitiert. Damit diese erfüllt ist, muss die Fresnel-Zahl aus Gl. (2.22) deutlich größer als eins sein: $N_F = a^2/(\lambda \hat{d}) \gg 1$. Dabei ist a die charakteristische Größe der Bereiche gleicher Struktureigenschaften und λ die Wellenlänge.

Um Elemente mit kleineren Strukturen zu analysieren, werden diese in Basiszellen mit einer bestimmten Größe zerlegt. Diese Basiszellen, typischerweise mit Ausdehnungen im Bereich der Wellenlänge, können mithilfe von rigorosen Verfahren wie RCWA oder FDTD exakt simuliert werden. Dabei nimmt man typischerweise periodische Randbedingungen in transversaler Richtung an. Aus dem so erhaltenen Feld nach der Struktur der Dicke d bestimmt man die Transmissionsfunktion t_f , die von den Parametern der betrachteten Struktur abhängt, wie z. B. vom geometrischen Füllfaktor f , siehe Abb. 2.2(a-c). Insbesondere wenn die Basiszellen kleiner als die Wellenlänge des Lichtes sind, genügt es dabei, die Mittelwerte der Felder zu betrachten:

$$t_f = \overline{u_f(x, y, d)} / \overline{u(x, y, 0)} . \quad (2.59)$$

Dabei bezeichnet \bar{u} die Mittelwertbildung in x - und y -Richtung. Dieses Vorgehen ist dabei gleichbedeutend mit der Beschränkung auf die 0. Beugungsordnung im Sinne eines effektiven Brechungsindex. Zur Ermittlung der Transmissionsfunktion eines gesamten Elementes werden nun

die Transmissionsfunktionen der jeweiligen Basiszellen entsprechend ihrer Verteilung im Element zusammengesetzt, siehe Abb. 2.2(d-f). Dieses Vorgehen wird als *Local Periodic Approximation* (Näherung lokaler Periodizität, LPA) bezeichnet [21] und häufig zur Beschreibung von DOEs verwendet [19, 42, 43].

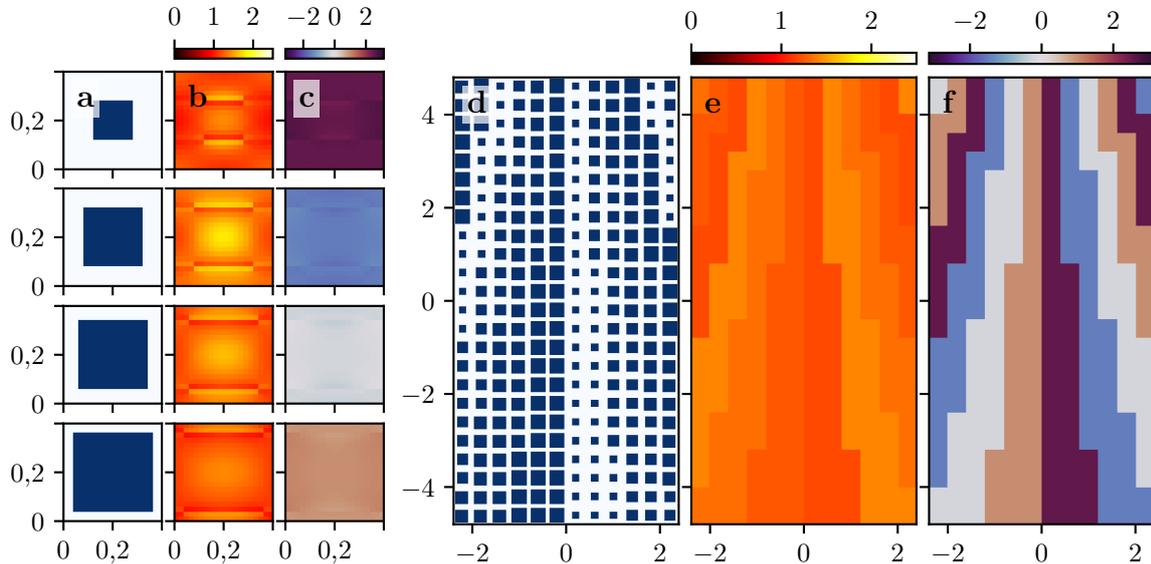


Abb. 2.2: Basiszellen mit Säulen verschiedener Größe (a), zugehörige Amplitude (b) und Phase (c) des berechneten Feldes nach dem Element; Struktur eines DOEs (d) und zugeordnete Transmissionsfunktion gemäß der LPA mit Amplitude (e) und Phase (f). Die Achsenbeschriftung kennzeichnet die Größe in μm , die Amplitude des einfallenden Feldes ist auf eins normiert und die Phase jeweils in rad angegeben.

Die LPA beschreibt das Feld an einer spezifischen Position quasi exakt, solange sich in einer hinreichend großen Umgebung, typischerweise einige Wellenlängen, nur Basiszellen mit den gleichen Strukturen befinden. Haben die benachbarten Basiszellen dagegen abweichende Strukturen, kommt es mitunter zu deutlichen Abweichungen, wie in Abb. 2.3 gezeigt: Im unteren Bereich sind die Strukturen über größere Bereiche gleich und verhalten sich weitestgehend gemäß der LPA, mit Ausnahme der Strukturgrenze im Zentrum. Nach oben hin ändern sich die Strukturen häufiger und die Abweichungen nehmen zu. Das gezeigte Element nutzt Effektiv-Index-Strukturen vergleichbar zu denen in Abschnitt 3.2.2.

Die Abweichungen der LPA sind typischerweise umso stärker, je geringer die Indexkontraste der beteiligten Materialien sind [25]. Selbiges gilt für die TEA, die als Spezialfall der LPA mit homogenen Schichten in den Basiszellen betrachtet werden kann. Beide Näherungen sind daher bei moderaten Indexkontrasten auf Anwendungsfälle mit kleinen Ablenkungswinkeln beschränkt, da dann die Bereiche gleicher bzw. ähnlicher Strukturen hinreichend groß sein können. Für bestimmte Anwendungen ist der Einfluss einer nicht-periodischen Umgebung erwünscht [44, 45], lässt sich für reale Elemente aber nur schwer numerisch beschreiben.

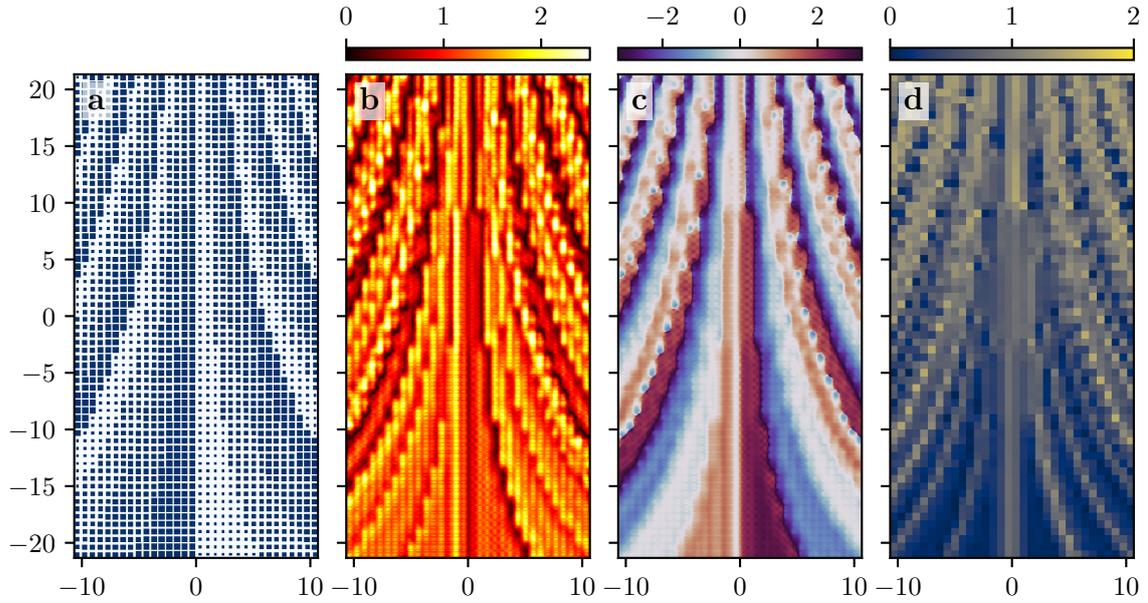


Abb. 2.3: Struktur eines DOEs (a), berechnete Amplitude (b) und Phase (c) nach dem Element, Betrag der Abweichung des Mittelwertes der Simulation von der LPA für jede Zelle (d). Die Längenangaben sind in μm .

2.2.4 Vergleich des Rechenaufwandes

Tab. 2.1: Vergleich der Laufzeiten und des Speicherbedarfes für die Simulation eines DOEs mit verschiedenen Simulationsalgorithmen. Die Definition der Parameter ist im Fließtext gegeben. Für eine fixierte Dicke des DOEs kann die Anzahl der Punkte in Propagationsrichtung N_z gegebenenfalls als konstant betrachtet werden.

Verfahren	Rechenzeit	Speicherbedarf
FDTD	$\mathcal{O}(N_x N_y N_z N_t)$ bzw. $\mathcal{O}(N^4)$	$\mathcal{O}(N_x N_y N_z)$ bzw. $\mathcal{O}(N^3)$
RCWA	$\mathcal{O}(M_x^3 M_y^3)$ bzw. $\mathcal{O}(N^6)$	$\mathcal{O}(M_x^2 M_y^2)$ bzw. $\mathcal{O}(N^4)$
BPM	$\mathcal{O}(N_x N_y N_z)$ bzw. $\mathcal{O}(N^3)$	$\mathcal{O}(N_x N_y)$ bzw. $\mathcal{O}(N^2)$
WPM	$\mathcal{O}(N_x N_y N_z \log N_x N_y)$ bzw. $\mathcal{O}(N^3 \log N)$	$\mathcal{O}(N_x N_y)$ bzw. $\mathcal{O}(N^2)$
TEA	$\mathcal{O}(N_x N_y)$ bzw. $\mathcal{O}(N^2)$	$\mathcal{O}(N_x N_y)$ bzw. $\mathcal{O}(N^2)$
LPA	$\mathcal{O}(N_x N_y)$ bzw. $\mathcal{O}(N^2)$	$\mathcal{O}(N_x N_y)$ bzw. $\mathcal{O}(N^2)$

Zum Vergleich des Rechenzeit- und Speicherbedarfes verschiedener Simulationsverfahren wird die Landau-Notation \mathcal{O} verwendet. Die Laufzeit $T(N)$ eines Algorithmus angewendet auf N Werte ist in der Laufzeitklasse $f(N)$, also $T(N) \in \mathcal{O}(f(N))$, wenn eine Konstante C und ein N_0 existieren, sodass für alle $N > N_0$

$$T(N) \leq C f(N) \quad (2.60)$$

gilt. Es seien N_x, N_y, N_z, N_t die Anzahl der räumlichen und zeitlichen Diskretisierungspunkte, zudem M_x und M_y die Anzahl der berücksichtigten Ordnungen in der RCWA. Zur besseren Vergleichbarkeit werden alle diese Größen auf einen Parameter N zurückgeführt, was sich durch eine geeignete Wahl der Konstante C ergibt. Dabei ist zu beachten, dass für die FDTD die zeitliche Diskretisierung an die räumliche gekoppelt ist und in der RCWA die Zahl der propagierenden Ordnungen linear von der Größe des Simulationsgebietes abhängt und damit indirekt bei festen Diskretisierungsschritten auch von der Zahl der Diskretisierungspunkte in anderen Verfahren. In Tab. 2.1 sind die Rechenzeiten und Speicheranforderungen der verschiedenen Algorithmen zusammengefasst. Bei der FDTD muss das Feld an jedem Raumpunkt für jeden Zeitschritt berechnet werden und das Feld für alle Raumpunkte gespeichert werden. In der BPM wird in jedem Iterationsschritt nur das Feld in einer Ebene betrachtet, sodass sich Laufzeit und Speicher entsprechend reduzieren. Analog ist dies bei der WPM der Fall, wobei sich der logarithmische Faktor aus der Verwendung der *Fast Fourier Transformation* (schnelle Fourier-Transformation, FFT) ergibt [46]. Bei TEA und LPA werden die Felder lediglich an den entsprechenden Punkten nach dem Element ausgewertet. Eine Sonderstellung nimmt die RCWA ein, deren Aufwand im Wesentlichen in der Lösung des auftretenden Eigenwertproblems begründet ist [32]. Es wird deutlich, dass die RCWA für größere Simulationsgebiete ungeeignet ist, da die Laufzeit sehr schnell anwächst. Für eine fixierte Dicke des DOEs kann die Anzahl der Punkte in Propagationsrichtung N_z als konstant betrachtet werden, ebenso wie die zeitliche Diskretisierung N_t für fixierte Diskretisierungsschritte. Dann haben sowohl BPM als auch FDTD eine Laufzeit bzw. einen Speicherbedarf von $\mathcal{O}(N^2)$, skalieren also linear mit der Gesamtgröße des Elementes. Hervorzuheben ist aber, dass die tatsächliche Laufzeit der BPM um mehrere Größenordnungen geringer als die der FDTD ist, siehe Abschnitt 3.2.

2.3 Design von diffraktiven optischen Elementen (DOEs)

Mithilfe von Mikrostrukturen kann das einfallende Licht auf vielfältige Weise beeinflusst werden. Die Betrachtung beschränkt sich im Folgenden auf die Transmission durch ein DOE. Reflektive Elemente lassen sich auf analoge Weise behandeln. Außerdem werden nur skalare Felder betrachtet, denn auch ein vektorielles Feld lässt sich durch Überlagerung der skalaren Komponenten darstellen. Die Funktionsweise des gesamten Elementes wird im Rahmen der TEA durch eine komplexe Transmissionsfunktion t beschrieben, die das einfallende Feld u^{in} in das ausgehende Feld u^{out} überführt:

$$u^{\text{out}}(x, y) = t(x, y)u^{\text{in}}(x, y) . \quad (2.61)$$

Sowohl Amplitude als auch Phase von t können zur Manipulation des Lichtes genutzt werden. Beispiele für reine Amplitudenelemente, die durch ein reelles t beschrieben werden können, sind

Fresnel-Zonenplatten oder Amplitudenhologramme [47]. Aufgrund der Energieerhaltung bedeutet eine Manipulation der Amplitude aber stets eine Verringerung der Transmission, sodass die mit Amplitudenelementen erreichbaren Effizienzen im Allgemeinen geringer als bei Phasenelementen sind. Insofern werden hier nur Phasenelemente betrachtet, für die idealerweise stets $|t| = 1$ gilt. In der Realität beeinflussen jedoch auch derartige DOEs die Amplitude, sodass es sinnvoll ist, diese im Design zu berücksichtigen, wie in Kap. 5 demonstriert wird.

2.3.1 Varianten der Berechnung

Grundsätzlich lassen sich zwei verschiedene Herangehensweisen für das Design von DOEs unterscheiden: Entweder sucht man zunächst eine Transmissionsfunktion, die eine bestimmte Funktionalität gewährleistet und dann nach Strukturen, die eben diese Transmission besitzen, oder man optimiert direkt die Strukturen, bis das Element eine gewollte Funktionsweise aufweist.

Je nach Anwendung lässt sich im ersten Fall die Transmissionsfunktion mitunter auf analytische Weise bestimmen, wie beispielsweise für die in Abschnitt 3.2.2 beschriebene Metalinse. In anderen Fällen muss die Transmissionsfunktion zunächst numerisch berechnet werden, wie bei computergenerierten Hologrammen (CGHs), siehe die folgenden Abschnitte. Für die Umsetzung in ein DOE wählt man nun zunächst eine Gruppe von Strukturen, die als Kandidaten in Betracht kommen und ermittelt deren Transmissionsfunktion durch ein geeignetes Simulationsverfahren, siehe z. B. Abschnitt 3.2.1. Anschließend ordnet man jedem Punkt der diskretisierten Transmissionsfunktion diejenige Struktur zu, welche dieser am nächsten kommt und nimmt dabei implizit an, dass sich die Struktur dort wie in der vorherigen Simulation verhält (LPA). Der Erfolg dieses Vorgehens hängt davon ab, ob die verfügbaren Strukturen die benötigten Werte abdecken und wie gut die LPA erfüllt ist. In Abb. 2.4 werden beispielhaft verschiedene Realisierungsvarianten eines CGHs (a) gezeigt [1, 48, 49]. Die Höhenprofile (b,c) basieren auf der Anwendung der TEA [50, 51], wobei durch das binäre Profil (b) die Phase nur in zwei Stufen erzeugt wird. Durch Ausnutzung des effektiven Brechungsindex erhält man das Profil (d) [19, 52], während in (e) die doppelbrechenden Strukturen für zirkular polarisiertes Licht eine geometrische Phase in Abhängigkeit des Rotationswinkels erzeugen [42, 43, 53–55].

Im zweiten Fall simuliert man die optische Funktion eines gesamten Elements. Ausgehend von einer Startlösung werden die Parameter einzelner Strukturen variiert, um das Ergebnis zu verbessern. Diese Vorgehensweise ist typischerweise numerisch aufwändiger, da stets das gesamte Element betrachtet werden muss, führt aber zu besseren Ergebnissen, da sie nicht auf der Gültigkeit der LPA basiert. Eine derartige Methode wurde zum Design eines Polarisations-Strahlteilers verwendet, siehe Abschnitt 3.2.3.

Das Ziel dieser Arbeit ist es, die Möglichkeiten zum Design von DOEs zu erweitern. Dabei wurden insbesondere solche Fälle betrachtet, in denen die erste Variante aufgrund der Verletzung

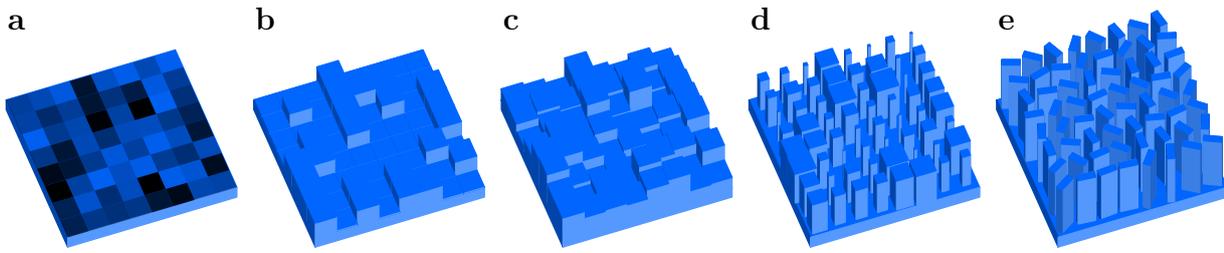


Abb. 2.4: Verschiedene Realisierungsvarianten eines DOEs zur Erzeugung jeweils derselben Transmissionsfunktion: ideales Phasenprofil als dünnes Element (a), binäres Höhenprofil (b), Höhenprofil mit stetigen Höhenstufen (c), Effektiv-Index-Profil (d), doppelbrechendes Profil mit Ausnutzung der geometrischen Phase (e).

der LPA keine qualitativ guten Ergebnisse liefert, wohingegen eine präzise Simulation des gesamten DOEs für die praktische Anwendung zu aufwendig ist. Dazu wurden einerseits schnelle, aber möglichst exakte Simulationsverfahren entwickelt, siehe Kap. 3 und 4, andererseits wurde speziell für die Berechnung von CGHs die Bestimmung der Transmissionsfunktion so angepasst, dass die Ergebnisse dieser Verfahren berücksichtigt werden, siehe Kap. 5.

In der Literatur gibt es verschiedene Ansätze, um den wechselseitigen Einfluss benachbarter Strukturen zu berücksichtigen, z. B. rigorose Simulationen des gesamten Elements in Kombination mit entsprechenden Optimierungsverfahren [22, 56–60], genetische Algorithmen zur Optimierung der Strukturen [61, 62], Zerlegungen des Simulationsgebietes [16, 23, 63] oder neuronale Netze zur Vorhersage der optischen Funktion aus der Strukturgeometrie [25, 64–66]. Diese Verfahren sind aber bislang auf 1D-Elemente beschränkt oder würden einen enormen Rechenaufwand für reale Anwendungen benötigen.

2.3.2 Computergenerierte Hologramme (CGHs)

Unter einem klassischen Hologramm versteht man die analog-fotografische Aufnahme eines Objektes, bei der durch Interferenz mit einer Referenzwelle nicht nur Amplituden- sondern auch Phaseninformationen im Film gespeichert werden. Bei Beleuchtung des Hologramms mit dieser Referenzwelle wird die ursprüngliche Feldverteilung rekonstruiert, sodass ein dreidimensionales Bild des Objektes beobachtet werden kann [13, 67]. Das vom Objekt ausgehende Feld in der Ebene des Hologramms kann auch mittels Computersimulationen berechnet werden. Ein DOE, welches diese Feldverteilung aus einer einfallenden Referenzwelle erzeugt, wird dann als CGH bezeichnet [12, 68, 69]. Dabei ist die Berechnung eines CGHs nicht an real existierende Objekte gebunden. Stattdessen lassen sich nahezu beliebige Feldverteilungen erzeugen. So können beispielsweise Freiformflächen interferometrisch vermessen werden [70, 71]: Durch einen Strahlteiler wird das einfallende Licht aufgespalten, siehe Abb. 2.5(a). Ein Teil läuft durch das CGH und wird so abgelenkt, dass es idealerweise stets senkrecht auf das Prüfobjekt trifft. Das reflektierte Licht

wird am Strahlteiler mit dem Reflex des Referenzspiegels zusammengeführt. Die Überlagerung beider Signale wird vom Interferometer aufgezeichnet. Abweichungen der Prüfobjektes von seiner idealen Form werden dann als Interferenzmuster sichtbar. Eine weitere typische Anwendung ist die Erzeugung einer beliebigen Intensitätsverteilung bei der Projektion ins Fernfeld (b), worauf im Folgenden näher eingegangen wird.

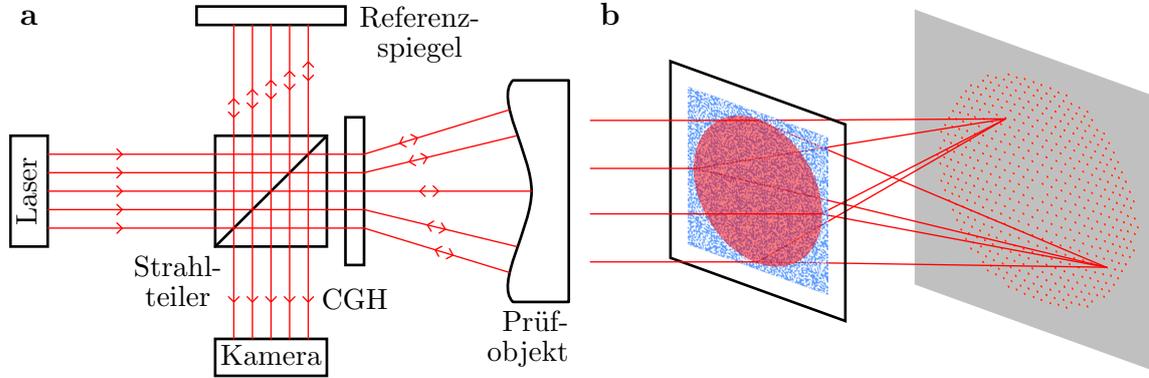


Abb. 2.5: Prinzip der interferometrischen Prüfung von Freiformflächen unter Nutzung eines CGHs (a), Erzeugung einer Intensitätsverteilung im Fernfeld (b).

Im Rahmen dieser Arbeit werden die Felder und Strukturen in diskretisierter Weise beschrieben. Unter einem CGH wird ein DOE verstanden, das aus einer regelmäßigen Anordnung von Mikrostrukturen in einem Gitter besteht. Als ein Pixel wird der einem Wert des berechneten Feldes zugeordnete Teil des Elementes mit der Größe $p_x \times p_y$ bezeichnet. Das gesamte CGH bestehend aus $M_x \times M_y$ Pixeln hat dann eine Gesamtgröße von $M_x p_x \times M_y p_y$. Alle CGHs werden bei der Berechnung als periodisch fortgesetzt angenommen, siehe Abb. 2.6(a). Dafür gibt es zwei wesentliche Gründe: Einerseits ist diese Annahme in der Verwendung der FT in den Berechnungen bereits immanent enthalten. Andererseits ist die Handhabung von periodischen Elementen einfacher, da keine bestimmte Positionierung zur Beleuchtung erforderlich ist. Zur genaueren numerischen Analyse der Strukturen werden feinere Abtastungen innerhalb eines Pixels gewählt, die hier als Subpixel der Größe $\delta x \times \delta y$ bezeichnet werden (b).

Im Folgenden soll der Zusammenhang zwischen Nah- und Fernfeld für solche periodischen Elemente näher erläutert werden. Zur Vereinfachung der Notation wird dabei zunächst ein eindimensionales Element aus M Pixeln der Größe p betrachtet, das mit einem konstanten Feld der Amplitude eins beleuchtet wird. Unter der Annahme eines invarianten Feldes innerhalb eines jeden Pixels ist

$$u(x) = u_m, \quad \text{für } \left(m - \frac{1}{2}\right)p \leq x \leq \left(m + \frac{1}{2}\right)p, \quad (2.62)$$

wobei $m = 0, \dots, M - 1$. Die periodische Fortsetzung lässt sich gemäß Ref. [50] als

$$u(x) = \left[\sum_{m=0}^{M-1} u_m \delta_0(x - mp) \right] * \text{rect}\left(\frac{x}{p}\right) * \sum_{K=-\infty}^{\infty} \delta_0(x - KMp) \quad (2.63)$$

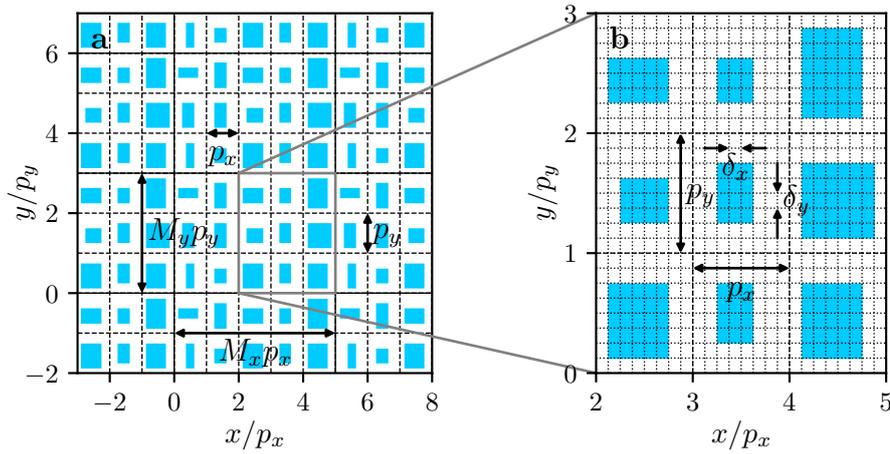


Abb. 2.6: Periodisch fortgesetztes CGH (a), Diskretisierung der Strukturen zur numerischen Berechnung (b).

ausdrücken. Dabei bezeichnet $*$ eine Faltung, δ_0 ist die Dirac'sche Delta-Distribution und rect die Rechteckfunktion. Die jeweiligen Definitionen sind im Anhang zu finden, Abschnitt A.1.2. Der linke Faktor beschreibt die Feldverteilung innerhalb eines Elements, der mittlere Faktor erweitert den Funktionswert auf die Breite der Pixel und der rechte Faktor beschreibt die periodische Fortsetzung mit einer Periode Mp . Die zur Bestimmung des Fernfeldes nutzbare FT dieses Feldes ist

$$U(k_x) = \frac{2\pi}{M} \text{sinc} \left(\frac{k_x p}{2} \right) \left[\sum_{m=0}^{M-1} U_m \delta_0(k_x - k_{x,m}) * \sum_{K=-\infty}^{\infty} \delta_0 \left(k_x - \frac{2\pi K}{p} \right) \right], \quad (2.64)$$

wie in Abschnitt A.2.1.2 hergeleitet wird, wobei U_m die diskrete FT von u_m ist. Deren Definition sowie die der sinc-Funktion werden in den Abschnitten A.1.4.1 und A.1.2 gegeben. Der sinc-Faktor beschreibt eine Modulation durch die Beugung an den einzelnen Pixeln. Die linke Summe beschreibt die Zuordnung von U_m zur m -ten Beugungsordnung mit einer Wellenzahl von $k_{x,m} = 2\pi m / (Mp)$, während die rechte Summe eine periodische Fortsetzung im Frequenzraum beschreibt. In Abb. 2.7(a) ist beispielhaft die periodische Fortsetzung unter Beachtung der Modulation dargestellt. Die einzelnen, diskreten Beugungsordnungen sind in Abb. 2.7(b) zu erkennen. Für viele Anwendungen ist $k_x p \ll 1$ (paraxiale Näherung) und es genügt, die diskrete FT zu betrachten. Andernfalls kann der sinc-Term als Korrekturfaktor berücksichtigt werden, siehe Abschnitt 5.1.2. Die periodische Fortsetzung spielt häufig keine Rolle, da einerseits die höheren Ordnungen durch den sinc-Faktor deutlich unterdrückt werden, andererseits hohe Ordnungen bei kleinen Pixelgrößen evaneszent sind und nicht zum Fernfeld beitragen. Zur anschaulicheren Darstellung werden im Rahmen dieser Arbeit die Ordnungen typischerweise von $-\lfloor M/2 \rfloor$ bis $\lfloor (M-1)/2 \rfloor$ statt von 0 bis $M-1$ dargestellt. Dabei ist $U_{-m} = U_{M-m}$. Für zweidimensionale Elemente ergibt sich das Fernfeld auf ähnliche Weise, siehe Gl. (A.64) im Anhang.

Ist das einfallende Feld nicht konstant, so kann das Feld nach dem Element im Sinne der

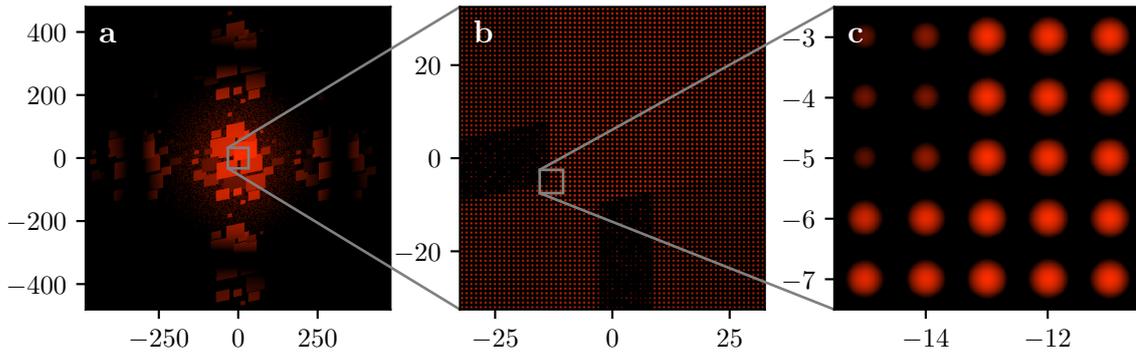


Abb. 2.7: Simulierte Fernfeldintensität eines CGHs zur Erzeugung eines Logos bei Beleuchtung mit einem gaußförmigen Eingangsfeld in verschiedenen Vergrößerungen. Die Skalen der Achsen verdeutlichen die Beugungsordnung. Es wurde eine logarithmische Farbskala verwendet, gemäß dem visuellen Eindruck des menschlichen Auges.

Transmissionsfunktion als Produkt

$$u^{\text{out}}(x, y) = u(x, y)u^{\text{in}}(x, y) \quad (2.65)$$

dargestellt werden. Die FT zur Beschreibung des Fernfeldes ergibt sich gemäß dem Faltungssatz als

$$U^{\text{out}}(k_x, k_y) = \frac{1}{2\pi} U(k_x, k_y) * U^{\text{in}}(k_x, k_y) , \quad (2.66)$$

siehe Gl. (A.43) im Anhang. Wenn sich das einfallende Feld nur auf Skalen ändert, die deutlich größer als eine Periode des CGHs sind, so ist das Winkelspektrum des Eingangsfeldes U^{in} schmaler als der Abstand zwischen zwei Ordnungen des CGHs. Dann beeinflusst das Eingangsfeld die Intensitätsverteilung auf die einzelnen Ordnungen nicht, sondern bestimmt lediglich das Profil der Beugungsordnungen, z. B. im Falle eines gaußförmigen Intensitätsprofils wie in Abb. 2.7(c). Die im Weiteren untersuchten Elemente sind stets für einen derartigen Anwendungsfall konzipiert, weshalb nur das Verhalten beim Einfall einer ebenen Welle betrachtet wird.

2.3.3 Iterativer Fourier-Transformations-Algorithmus (IFTA)

2.3.3.1 Grundform von Gerchberg und Saxton

Der *Iterative Fourier Transform Algorithm* (Iterativer Fourier-Transformations-Algorithmus, IFTA) dient zur Ermittlung eines Phasenprofils, welches ein einfallendes Feld so modifiziert, dass sich nach der weiteren Propagation eine bestimmte Intensitätsverteilung einstellt. In seiner ursprünglichen Formulierung wurde er von Gerchberg und Saxton entwickelt [72], weshalb er auch als *Gerchberg-Saxton-Algorithmus* bezeichnet wird. Im Vergleich zu anderen Algorithmen zeichnet er sich durch eine schnelle Konvergenz aufgrund der gleichzeitigen Optimierung des

2 Grundlagen

gesamten Profils und die allgemeine Anwendbarkeit aus [73]. Zu einem einfallenden skalaren Feld, das in diskretisierter Form $u_{m_x m_y}^{\text{in}}$ gegeben ist, sucht man eine normierte Transmissionsfunktion $t_{m_x m_y} = \exp(i\phi_{m_x m_y})$, sodass man nach Anwendung eines Propagationsoperators \mathcal{P} eine bestimmte Intensität $I_{m_x m_y}$ erhält:

$$\mathcal{P}\left(u_{m_x m_y}^{\text{in}} t_{m_x m_y}\right) = C \sqrt{I_{m_x m_y}} \exp\left(i\Phi_{m_x m_y}\right). \quad (2.67)$$

Darin bezeichnet C einen beliebigen skalaren Faktor und $\Phi_{m_x m_y}$ eine beliebige Phase, welche für die Intensitätsverteilung unerheblich ist. Der Propagationsoperator könnte z. B. auf der ASPW basieren und so die Propagation über eine endliche Distanz im homogenen Raum beschreiben, typischerweise wird aber die Propagation ins Fernfeld oder in die Fourier-Ebene einer Linse nach dem Element betrachtet. In beiden Fällen kann der zugehörige Operator bis auf einen Faktor durch eine FT beschrieben werden: $\mathcal{P} \sim \mathcal{F}$, siehe Gl. (2.64). Darauf wird die Betrachtung im Folgenden beschränkt. Zur Vereinfachung der Notation werden Indizes weggelassen, wenn nicht explizit Bezug darauf genommen wird, wobei alle Größen in diskretisierter Form zu verstehen sind. In seiner Grundform lässt sich der IFTA wie folgt formulieren: Ausgehend von einer zufälligen Phasenfunktion $\phi^{(0)}$ ermittelt man zunächst ein Startfeld $u^{(0)} = u^{\text{in}} \exp(i\phi^{(0)})$. Dann führt man iterativ die folgenden Schritte aus:

1. Propagation ins Fernfeld: $U^{(j)} = \mathcal{F}u^{(j)}$.
2. Ersetzen der Amplitude mithilfe der gewünschten Intensität: $U'^{(j)} = \sqrt{I} \exp\left(i \arg U^{(j)}\right)$.
3. Rückpropagation ins Nahfeld: $u'^{(j)} = \mathcal{F}^{-1}U'^{(j)}$.
4. Ersetzen der Amplitude gemäß des Eingangsfeldes: $u^{(j+1)} = |u^{\text{in}}| \exp\left(i \arg u'^{(j)}\right)$.

Wenn die Abweichung zwischen $|U^{(j)}|$ und \sqrt{I} klein genug ist, gegebenenfalls nach einer geeigneten Skalierung, erhält man die gesuchte Phase als $\phi = \arg(u^{(j+1)}/u^{\text{in}})$. In der gezeigten Form konvergiert der IFTA stets gegen ein lokales Minimum [74]. Im Allgemeinen handelt es sich dabei allerdings nicht um ein globales Minimum und die gefundene Lösung hängt stark von der gewählten Startphase $\phi^{(0)}$ ab. Typischerweise wendet man den Algorithmus auf verschiedene Ausgangswerte an, um ein möglichst gutes Ergebnis zu erzielen. Die Zahl der nötigen Iterationen wird deutlich von der konkreten Zielfunktion bzw. Diskretisierung beeinflusst. Die im Rahmen dieser Arbeit gezeigten Beispiele haben typischerweise 30 bis 60 Iterationen durchlaufen. Im Folgenden werden verschiedene Modifikationen des IFTA vorgestellt, welche die Konvergenz beschleunigen oder bestimmte Eigenschaften der gefundenen Lösung verbessern. Die gesicherte Konvergenz ist dafür in der Regel nicht mehr gegeben.

2.3.3.2 Amplitudenfreiheit

Für viele Anwendungen ist die Intensitätsverteilung in einem gewissen Bereich des Fernfeldes von größerem Interesse als in den umliegenden Bereichen. Es sei

$$\chi_{S,m_x m_y} = \begin{cases} 1, & \text{für } (m_x, m_y) \in S, \\ 0, & \text{für } (m_x, m_y) \notin S, \end{cases} \quad (2.68)$$

die Indikatorfunktion dieses Signalbereiches S . Ein typisches Beispiel ist die gleichmäßige Aufteilung des Eingangsfeldes in eine gewisse Anzahl Ordnungen, sodass $I = \chi_S$ ist. Eine gleichmäßigere Verteilung innerhalb von S kann durch eine Reduktion der Anforderungen an die umliegenden Bereiche erreicht werden. Dazu erlaubt man außerhalb von S Amplitudenfreiheit, d. h. die Fernfeldamplitude wird nur im Bereich von S ersetzt [75, 76]. Der Bereich der Amplitudenfreiheit hat die Indikatorfunktion $\chi_A = 1 - \chi_S$. Schritt 2 der Iteration ändert sich damit zu

$$2a. \text{ Ersetzen für Signalregion: } U^{(j)} = \left(\alpha \chi_S \sqrt{I'} + (1 - \alpha) \chi_A |U^{(j)}| \right) \exp(i \arg U^{(j)}).$$

Über den Faktor α kann dabei der Leistungsanteil im Signalbereich beeinflusst werden. Um die Wirksamkeit dieser Ersetzung zu verbessern, ist eine Skalierung der Zielfunktion I nötig, sodass diese die gleiche Gesamtenergie wie das errechnete Fernfeld $U^{(j)}$ besitzt. Dazu setzt man

$$I'_{m_x m_y} = I_{m_x m_y} \frac{M_x M_y \sum_{m'_x, m'_y} |u_{m'_x m'_y}^{\text{in}}|^2}{\sum_{m'_x, m'_y} I_{m'_x m'_y}}, \quad (2.69)$$

denn aufgrund des Satzes von Plancherel ist $\sum_{m_x, m_y} |U_{m_x m_y}^{(j)}|^2 = M_x M_y \sum_{m_x, m_y} |u_{m_x m_y}^{(j)}|^2$, siehe Gl. (A.45) im Anhang, wobei M_x und M_y die Gesamtzahl der Punkte je Richtung sind.

Amplitudenfreiheit kann auch eingesetzt werden, um eine gleichmäßigere Konvergenz des Algorithmus zu ermöglichen. Große Änderungen zwischen zwei Iterationen lassen sich vermeiden, wenn das Fernfeld nur sukzessive an die Zielverteilung angepasst wird. Auf diese Weise kann z. B. eine kontinuierliche Phasenfunktion erzeugt werden [77]. Der entsprechende Iterationsschritt lautet dann

$$2b. \text{ Ersetzen mit Amplitudenfreiheit: } U^{(j)} = \left(\alpha_j \sqrt{I'} + (1 - \alpha_j) |U^{(j)}| \right) \exp(i \arg U^{(j)}).$$

Mit dem Parameter $0 < \alpha_j \leq 1$ kann so der Anteil der Amplitudenfreiheit innerhalb der Signalregion reguliert werden. Im Laufe der Iterationen wächst α_j von 0 auf 1. Hier wird ebenfalls die normierte Zielfunktion I' verwendet.

Die Skalierung der Zielfunktion kann ebenfalls so gewählt werden, dass die Änderung des Fernfeldes bei der Ersetzung minimiert wird [78], indem α_j als

$$\alpha_j = \frac{\sum_{m_x, m_y} |U_{m_x m_y}^{(j)}| \sqrt{I'_{m_x m_y}}}{\sum_{m_x, m_y} I'_{m_x m_y}} \quad (2.70)$$

berechnet wird. Außerhalb des Signalbereiches ist die Skalierung davon unabhängig wählbar und kann dort gezielt die Intensität unterdrücken. Dieser Bereich hat einen Anteil von

$$\alpha'_j = \frac{\sum_{m_x, m_y} \chi_{A, m_x m_y} |U_{m_x m_y}^{(j)}|^2}{\sum_{m_x, m_y} I'_{m_x m_y}} \quad (2.71)$$

an der gesamten Intensität. Den 2. Schritt der Iteration wählt man nun z. B. als

2c. Ersetzen mit separater Skalierung:

$$U'^{(j)} = \left(\chi_S \alpha'_j \sqrt{I'} + \chi_A \min \left\{ 1, \sqrt{\frac{\alpha_{\max}}{\alpha'_j}} \right\} |U^{(j)}| \right) \exp(i \arg U^{(j)}) .$$

Dann erhöht man einerseits die Korrelation zwischen aktuellem Signal und Zielfunktion und stellt andererseits sicher, dass der Anteil der Energie außerhalb des Signalbereiches höchstens α_{\max} beträgt.

2.3.3.3 Verbesserung der Konvergenz

Die bisher genannten Modifikationen verwendeten eine globale Variation der für die Ersetzung in Schritt 2 genutzten Amplitude. Daneben ist auch ein Eingriff auf der Ebene einzelner Ordnungen denkbar. Dafür wird die Zielfunktion im Signalbereich so angepasst, dass Ordnungen die momentan noch eine zu geringe Amplitude haben, bewusst angehoben bzw. bei zu hoher Amplitude abgesenkt werden [79, 80]. Dann erhält man beispielsweise folgenden Iterationsschritt:

$$2d. \text{ Ersetzen mit Überanpassung: } U'^{(j)} = \left(\chi_S \frac{I'^{\beta/2}}{|U^{(j)}|^\beta} \sqrt{I'} + \chi_A |U^{(j)}| \right) \exp(i \arg U^{(j)}) .$$

Darin ist β eine Konstante. Auf diese Weise kann eine gleichmäßigere Annäherung an die Zielfunktion erreicht werden. Mit größerem β nimmt die Übereinstimmung zu, aber der Algorithmus neigt zunehmend zu Instabilität und konvergiert gegebenenfalls nicht.

2.3.3.4 Diskretisierung der Phasenfunktion

In den bislang geschilderten Varianten erzeugt der IFTA stets eine Verteilung von kontinuierlichen Phasenwerten. Die Herstellung eines Elementes ist jedoch in der Regel auf einige diskrete Phasenwerte $\Phi = \{\phi_1, \dots, \phi_{N_D}\}$, $\phi_j \in [0, 2\pi]$, beschränkt. Um diskrete Werte zu erhalten, können die stetigen Werte auf den jeweils nächstgelegenen der möglichen diskreten Werte projiziert werden, was sich durch einen Projektionsoperator \mathcal{D}_Φ ausdrücken lässt:

$$\mathcal{D}_\Phi \phi := \arg \min_{\phi' \in \Phi \cup \tilde{\Phi}} |\phi - \phi'| , \quad (2.72)$$

mit $\tilde{\Phi} = \{\phi_1 \pm 2\pi, \dots, \phi_{N_D} \pm 2\pi\}$. Diese Schreibweise dient zur korrekten Berücksichtigung der 2π -Periodizität.

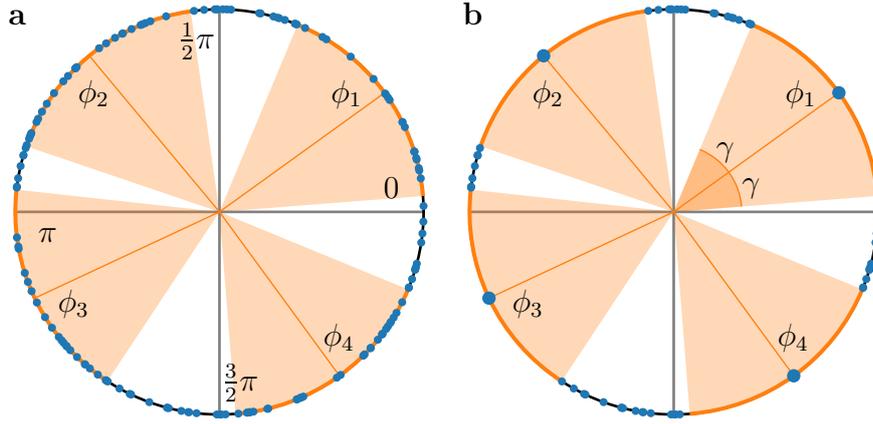


Abb. 2.8: Veranschaulichung der Projektion durch den Operator $\mathcal{D}_{\Phi, \gamma}$: Phasenwerte vor der Projektion (a) und diskrete Phasenpunkte ϕ_1, \dots, ϕ_4 ; Phasenwerte nach der Projektion aller Punkte (b) innerhalb eines Abstandes von γ zu den diskreten Werten (orange Bereiche).

Wendet man eine derartige Diskretisierung direkt im IFTA an, kommt es zur Stagnation des Algorithmus, d. h. die Abweichung zwischen $A^{(j)}$ und \sqrt{I} bleibt deutlich größer als ohne Diskretisierung, vgl. Ref. [78]. In Verallgemeinerung der dort geschilderten Vorgehensweise ist es sinnvoller, diese harte Projektion erst am Ende der Iterationen anzuwenden und sie vorher durch eine abgemilderte Variante zu ersetzen. Dafür wird der Projektionsoperator $\mathcal{D}_{\Phi, \gamma}$ als

$$\mathcal{D}_{\Phi, \gamma} \phi := \begin{cases} \mathcal{D}_{\Phi} \phi, & \text{für } |\phi - \mathcal{D}_{\Phi} \phi| < \gamma \text{ oder } |\phi - \mathcal{D}_{\Phi} \phi \pm 2\pi| < \gamma, \\ \phi, & \text{sonst} \end{cases} \quad (2.73)$$

definiert. Für $\gamma = 0$ belässt dieser Operator den Phasenwinkel bei seinem Wert und falls γ hinreichend groß ist, wird er identisch zu \mathcal{D}_{Φ} , siehe Abb. 2.8. Dabei wird wiederum die Periodizität berücksichtigt. Im Laufe der Iterationen des IFTA wird der Wert von γ schrittweise erhöht, sodass anfänglich nur Phasenwerte, die ohnehin nahe bei den vorgegebenen Werten sind, projiziert werden und erst am Ende Projektionen stattfinden, die mit stärkeren Änderungen einhergehen. So können die Einflüsse der Projektion im Laufe der Iterationen noch ausgeglichen werden. Der 4. Schritt der Iterationen ändert sich damit zu:

4a. Ersetzen der Amplitude bei diskreter Phase: $u^{(j+1)} = u^{\text{in}} \exp \left(i \mathcal{D}_{\Phi, \gamma_j} \arg \frac{u'^{(j)}}{u^{\text{in}}} \right)$.

3 Effiziente Simulation von optischen Mikrostrukturen

Zur effizienten Simulation von Mikrostrukturen wurde die in Abschnitt 2.2.2.1 eingeführte BPM angepasst. Insbesondere wurde der verwendete Differentialoperator akkurater approximiert, was die Berechnung größerer Ablenkwinkel und Indexkontraste ermöglicht. In diesem Kapitel werden zunächst die durchgeführten Modifikationen des Simulationsverfahrens geschildert und anschließend Anwendungsbeispiele vorgestellt. Die Inhalte wurden im Wesentlichen in Ref. [81] veröffentlicht und werden hier in überarbeiteter und erweiterter Form wiedergegeben.

3.1 Modifikation der Beam Propagation Method

Um die BPM für die Simulation von beispielsweise Effektiv-Index-Strukturen nutzbar zu machen, wurde die Berechnung an mehreren Punkten modifiziert. Der wesentliche Vorteil der BPM gegenüber der FDTD liegt in der geringeren Laufzeit, die vor allem darauf basiert, dass nur das stationäre Feld bei einer festen Frequenz berechnet wird, anstatt den zeitlichen Verlauf zu simulieren. Bei sämtlichen Anpassungen muss daher sichergestellt werden, dass sich die Laufzeitklasse der BPM nicht ändert.

3.1.1 Anpassung an große Öffnungswinkel

In Anlehnung an die entsprechende Notation im vorigen Kapitel wird die Propagation eines elektromagnetischen Feldes in z -Richtung durch eine in dieser Richtung invariante Indexverteilung betrachtet. Aus dem an einer Position z bekannten Feld soll dabei das Feld an einer neuen Position $z + \delta z$ berechnet werden. Im Folgenden wird beispielhaft die Berechnung der y -Komponente der langsam veränderlichen Amplitude des elektrischen Feldes beschrieben. Gemäß Gl. (2.50) gilt

$$\frac{\partial}{\partial z} \hat{E}_y = i\bar{n}k_0 \left(\sqrt{1 + 2\mathcal{A}_{yy}} - 1 \right) \hat{E}_y \quad (3.1)$$

für den vorwärts propagierenden Teil des Feldes. In der bislang vorgestellten Version der BPM wird der Differentialoperator $\sqrt{1 + 2\mathcal{A}_{yy}}$ linear genähert, siehe Gl. (2.51). Diese Näherung führt nur bei kleinen Propagationswinkeln und Indexdifferenzen zu nutzbaren Ergebnissen. Zur Gewinnung

einer allgemeinen Lösung kann eine Padé-Approximation des Grades (N_P, N_Q) genutzt werden [82]. Dabei nähert man den Differentialoperator als rationalen Ausdruck

$$\sqrt{1 + 2\mathcal{A}_{yy}} \approx \frac{P(\mathcal{A}_{yy})}{Q(\mathcal{A}_{yy})}, \quad (3.2)$$

worin P und Q Polynome vom Grad N_P bzw. N_Q sind, für Details siehe Abschnitt A.1.3.3 im Anhang. Der vorwärts propagierende Teil des Feldes ergibt sich damit als

$$\frac{\partial}{\partial z} \hat{E}_y = i\bar{n}k_0 \left(Q^{-1}(\mathcal{A}_{yy})P(\mathcal{A}_{yy}) - 1 \right) \hat{E}_y. \quad (3.3)$$

Bei \mathcal{A}_{yy} handelt es sich um einen Differentialoperator zweiter Ordnung, der durch die Anwendung finiter Differenzen numerisch berechnet werden kann. Höhere Potenzen innerhalb der Polynome P, Q entsprechen der wiederholten Anwendung dieses Operators, während Q^{-1} den inversen Operator zu Q kennzeichnet. Die Padé-Approximation der Ordnung $(1, 0)$ ergibt gerade die lineare Näherung wie in Gl. (2.51).

Um das bekannte Feld an einer Position, $\hat{E}_y^{(s)} = \hat{E}_y(z)$, zu einer neuen Position zu propagieren, $\hat{E}_y^{(s+1)} = \hat{E}_y(z + \delta z)$, wird ein Crank-Nicolson-Schema angewendet [82]:

$$\left((1 + \kappa)Q(\mathcal{A}_{yy}) - \kappa P(\mathcal{A}_{yy}) \right) \hat{E}_y^{(s+1)} = \left((1 - \kappa)Q(\mathcal{A}_{yy}) + \kappa P(\mathcal{A}_{yy}) \right) \hat{E}_y^{(s)}, \quad (3.4)$$

worin $\kappa = \frac{1}{2}i\bar{n}k_0\delta z$ ist, siehe auch Abschnitt A.2.2.3 im Anhang. Stellt man \mathcal{A}_{yy} als Matrix zur Berechnung der finiten Differenzen dar, so entspricht dieser Iterationsschritt der Lösung eines linearen Gleichungssystems. Die Polynome in \mathcal{A}_{yy} auf beiden Seiten der Gleichung lassen sich gemäß des Fundamentalsatzes der Algebra als Produkt von linearen Faktoren darstellen [83]:

$$l_0 \left(\prod_{j=1}^{N_P} (1 + l_j \mathcal{A}_{yy}) \right) \hat{E}_y^{(s+1)} = r_0 \left(\prod_{j=1}^{N_P} (1 + r_j \mathcal{A}_{yy}) \right) \hat{E}_y^{(s)}. \quad (3.5)$$

Dabei sei ohne Beschränkung der Allgemeinheit $N_P = N_Q$, während l_j, r_j die negativen Reziproken der Polynomnullstellen bezeichnen. In dieser Darstellung kann Gl. (3.4) durch die N_P -malige Lösung von Gleichungssystemen, welche \mathcal{A}_{yy} nur in linearer Form enthalten, gelöst werden. Auf diese Weise können die Koeffizientenmatrizen in tridiagonaler Form dargestellt werden, was eine wesentlich schnellere Lösung der Gleichungssysteme als für allgemeine Matrizen ermöglicht, siehe den folgenden Abschnitt.

In der bisher geschilderten Form kann diese Iteration zu numerischen Instabilitäten führen, da durch die Näherung in Gl. (3.2) Polynome mit reellen Koeffizienten entstehen, welche einerseits in der Nähe der Nullstellen von Q divergieren und andererseits den exponentiellen Abfall evaneszenter Wellen nicht modellieren können. Beides lässt sich beheben, indem vor Anwendung der Padé-Approximation eine komplexwertige Transformation durchgeführt wird [84, 85]. Für ein beliebiges ψ gilt

$$\sqrt{1 + 2\mathcal{A}_{yy}} = \exp\left(i\frac{\psi}{2}\right) \sqrt{1 + 2\tilde{\mathcal{A}}_{yy}}, \quad \text{für } \tilde{\mathcal{A}}_{yy} = \exp(-i\psi) \left(\frac{1}{2} + \mathcal{A}_{yy} \right) - \frac{1}{2}. \quad (3.6)$$

Nun kann $\tilde{\mathcal{A}}_{yy}$ als Argument der Padé-Approximation genutzt werden:

$$\sqrt{1 + 2\mathcal{A}_{yy}} \approx \exp\left(i\frac{\psi}{2}\right) \frac{P(\tilde{\mathcal{A}}_{yy})}{Q(\tilde{\mathcal{A}}_{yy})}, \quad (3.7)$$

sodass die Wahl eines $\psi > 0$ einer Verschiebung des Entwicklungspunktes der Padé-Approximation gleichkommt. Im Vergleich zu Gl. (3.5) ändern sich dabei die Koeffizienten l_j, r_j sowie das Argument \mathcal{A}_{yy} , während die Grundstruktur erhalten bleibt. Sinnvollerweise wird ψ aus $[0, \pi/2]$ gewählt, wobei größere Werte tendenziell zu mehr Stabilität, aber verminderter Genauigkeit führen. Bei den im Rahmen dieser Arbeit typischerweise verwendeten Polynomgraden von $N_P = 3, 4$ hat sich ein Wert von $\psi = 0,25\pi$ bzw. $0,35\pi$ als guter Kompromiss herausgestellt, wobei die Ergebnisse nur schwach von der exakten Wahl von ψ abhängen. Zur Vereinfachung der Notation werden im Folgenden die auftretenden Größen in ihrer bisherigen Form notiert, sind jedoch inklusive der geschilderten Transformation zu verstehen. Für die Berechnung der anderen Feldkomponenten $\hat{E}_x, \hat{H}_x, \hat{H}_y$ ist die Vorgehensweise vollständig analog.

3.1.2 Effiziente Berechnung

Wie im vorherigen Abschnitt geschildert, bildet die iterative Lösung von Gleichungen der Form

$$(1 + l_j \mathcal{A}_{yy}) \hat{E}_y^{(s+1)} = (1 + r_j \mathcal{A}_{yy}) \hat{E}_y^{(s)} \quad (3.8)$$

das Kernstück der BPM. In diesem Abschnitt soll die numerische Lösung dieser Gleichungen geschildert werden, wiederum am Beispiel der Komponente \hat{E}_y . Der Differentialoperator \mathcal{A}_{yy} hat dabei folgende Gestalt:

$$\mathcal{A}_{yy} \hat{E}_y = \frac{1}{2\bar{n}^2 k_0^2} \left(\frac{\partial^2}{\partial x^2} \hat{E}_y + \frac{\partial}{\partial y} \left(\frac{1}{n^2} \frac{\partial}{\partial y} (n^2 \hat{E}_y) \right) \right) + \frac{n^2 - \bar{n}^2}{2\bar{n}^2} \hat{E}_y. \quad (3.9)$$

Zunächst wird der 1D-Fall betrachtet. Dazu seien alle Strukturen und Felder in y -Richtung invariant, d. h. alle Ableitungen $\partial/\partial y$ verschwinden, sodass sich der Operator zu

$$\mathcal{A}_{yy} \hat{E}_y = \frac{1}{2\bar{n}^2 k_0^2} \frac{\partial^2}{\partial x^2} \hat{E}_y + \frac{n^2 - \bar{n}^2}{2\bar{n}^2} \hat{E}_y \quad (3.10)$$

vereinfacht. Zur numerischen Lösung werden Operator, Feld und Brechungsindex diskretisiert, d. h. aus $\hat{E}_y(x, z)$ wird $\hat{E}_{y,m}^{(s)}$ und \mathcal{A}_{yy} wird mithilfe zentraler finiter Differenzen als

$$\left(\mathcal{A}_{yy} \hat{E}_y\right)_m^{(s)} = \frac{1}{2\bar{n}^2 k_0^2 \delta x^2} \left(\hat{E}_{y,m+1}^{(s)} - 2\hat{E}_{y,m}^{(s)} + \hat{E}_{y,m-1}^{(s)} \right) + \frac{n_m^2 - \bar{n}^2}{2\bar{n}^2} \hat{E}_{y,m}^{(s)} \quad (3.11)$$

berechnet [35]. Eine vollständige Beschreibung der Diskretisierung wird in den Abschnitten A.2.2.3 und A.2.2.4 im Anhang gegeben. Die Anwendung des Operators entspricht der Multiplikation des Vektors $\hat{E}_y^{(s)}$ mit einer tridiagonalen Matrix. Die Lösung von Gl. (3.8) erfordert also neben einer solchen Matrix-Vektor-Multiplikation nur die Lösung eines linearen Gleichungssystems mit

tridiagonaler Koeffizientenmatrix. Mithilfe des Thomas-Algorithmus [86] kann ein solches System für einen Vektor der Länge N und eine $N \times N$ -Matrix in linearer Laufzeit $\mathcal{O}(N)$ gelöst werden. Ein vollständiger Iterationsschritt gemäß Gl. (3.5) erfordert die N_P -malige Anwendung dieses Algorithmus. Das ist dennoch erheblich schneller als die Lösung eines allgemeinen Gleichungssystems, wie es aus der direkten Lösung von Gl. (3.4) resultieren würde. Deren Laufzeit skaliert gemäß $\mathcal{O}(N^3)$. Zudem ist N_P eine typischerweise einstellige, von der Anzahl der Stützstellen unabhängige Konstante.

Für den 2D-Fall, wenn der Brechungsindex sowohl in x - als auch in y -Richtung variiert, können die Differentialoperatoren nicht mehr in Form einer Tridiagonalmatrix diskretisiert werden, siehe z. B. Gl. (A.99) im Anhang. Die Lösung eines Gleichungssystems mit einer allgemeinen Koeffizientenmatrix ist allerdings erheblich aufwendiger. Deshalb wird ein sogenanntes *alternating direction implicit scheme* (implizites Schema alternierender Richtungen, ADI) zur Berechnung verwendet [87]. Dabei werden multidimensionale Operatoren als Produkt von 1D-Operatoren dargestellt, welche dann nacheinander angewendet werden können. So schreibt man z. B. den Operator \mathcal{A}_{yy} aus Gl. (3.9) als $\mathcal{A}_{yy} = \mathcal{T}_x + \mathcal{T}_y + \mathcal{T}_0$, wobei

$$\mathcal{T}_x = \frac{1}{2\bar{n}^2 k_0^2} \frac{\partial^2}{\partial x^2}, \quad \mathcal{T}_y = \frac{1}{2\bar{n}^2 k_0^2} \frac{\partial}{\partial y} \left(\frac{1}{n^2} \frac{\partial}{\partial y} (n^2 \cdot) \right), \quad \mathcal{T}_0 = \frac{n^2 - \bar{n}^2}{2\bar{n}^2}. \quad (3.12)$$

Für die Anwendung in Gl. (3.5) ist dann

$$\mathcal{L}_j = 1 + l_j \mathcal{A}_{yy} = \underbrace{\left(1 + l_j \mathcal{T}_y + \frac{l_j}{2} \mathcal{T}_0 \right) \left(1 + l_j \mathcal{T}_x + \frac{l_j}{2} \mathcal{T}_0 \right)}_{\text{Produkt von 1D-Operatoren } \bar{\mathcal{L}}_j} - \underbrace{l_j^2 \left(\mathcal{T}_y + \frac{1}{2} \mathcal{T}_0 \right) \left(\mathcal{T}_x + \frac{1}{2} \mathcal{T}_0 \right)}_{\text{Fehlerterm } \tilde{\mathcal{L}}_j} \quad (3.13)$$

eine klassische Faktorisierung [85]. Das für die Berechnung genutzte Produkt $\bar{\mathcal{L}}_j$ weicht also um einen Fehlerterm $\tilde{\mathcal{L}}_j$ vom eigentlichen Operator ab. Ohne Padé-Approximation ist l_j proportional zu δz [88] und wird folglich durch eine geringere Schrittweite reduziert. Bei Verwendung der Padé-Approximation erhält l_j jedoch einen von δz unabhängigen Anteil, während \mathcal{T}_0 für einen inhomogenen Brechungsindex ebenso von 0 verschieden ist. Somit ist der Fehlerterm vom Indexkontrast abhängig und kann nicht vermieden werden. Zur Reduzierung des Fehlers wurde in Ref. [81] eine alternative Faktorisierung vorgeschlagen:

$$\mathcal{L}_j = \left[(1 + l_j \mathcal{T}_0)^{\frac{1}{2}} + l_j \mathcal{T}_y (1 + l_j \mathcal{T}_0)^{-\frac{1}{2}} \right] \left[(1 + l_j \mathcal{T}_0)^{\frac{1}{2}} + l_j (1 + l_j \mathcal{T}_0)^{-\frac{1}{2}} \mathcal{T}_x \right] - l_j^2 \mathcal{T}_y (1 + l_j \mathcal{T}_0)^{-1} \mathcal{T}_x. \quad (3.14)$$

Der Operator $1 + l_j \mathcal{T}_0$ entspricht einer reinen Diagonalmatrix, sodass sich die Wurzel und die Inversen trivial berechnen lassen. In dieser Darstellung ist \mathcal{T}_0 im Fehlerterm nicht separat enthalten, sondern nur in Kombination mit \mathcal{T}_x und \mathcal{T}_y . Demzufolge hat dieser nur dort einen größeren Einfluss, wo sich Feld bzw. Brechungsindex in beiden Richtungen ändern. Insbesondere ergibt diese Faktorisierung den exakten 1D-Operator, falls \mathcal{T}_x oder \mathcal{T}_y verschwinden.

Ein anderer Ansatz zur Vermeidung des Fehlerterms ist die iterative Verbesserung der Lösung, wie in Ref. [85] vorgestellt. Ein ähnlicher Ansatz in Form von Nachiterationen wurde hier genutzt. Die Lösung von Gl. (3.5) erfordert die wiederholte Lösung von Gleichungen der Form

$$(\bar{\mathcal{L}}_j + \tilde{\mathcal{L}}_j) \hat{E}'_y = (\bar{\mathcal{R}}_j + \tilde{\mathcal{R}}_j) \hat{E}_y . \quad (3.15)$$

Zur Vereinfachung der Notation sei \hat{E}_y das Ausgangsfeld und \hat{E}'_y das Ergebnis eines solchen Zwischenschritts. Die Operatoren auf beiden Seiten wurden gemäß Gl. (3.13) zerlegt. Dann wird die Gleichung zunächst mit den 1D-Operatoren gelöst, anschließend wird das Residuum bestimmt und durch iterative Lösung der Gleichung reduziert:

$$\bar{\mathcal{L}}_j \hat{E}_{y,(1)} = \bar{\mathcal{R}}_j \hat{E}_y , \quad \bar{\mathcal{L}}_j \hat{E}_{y,(j'+1)} = \bar{\mathcal{R}}_j \hat{E}_y - \mathcal{L}_j \sum_{j''=1}^{j'} \hat{E}_{y,(j'')} , \quad j' = 1, \dots, N_I - 1 . \quad (3.16)$$

Das Ergebnis ergibt sich als Summe $\hat{E}'_y = \sum_{j'=1}^{N_I} \hat{E}_{y,(j')}$, für eine Gesamtzahl an Nachiterationen von N_I . Bei dieser Verfahrensweise stehen auf der rechten Seite der Gleichung stets Matrix-Vektor-Multiplikationen, auf der linken Seite nur Produkte von 1D-Operatoren, die zunächst für die y -Richtung und dann auf die x -Richtung angewendet werden. Mit einer entsprechenden Umsortierung der Einträge können beide Operationen mithilfe von Tridiagonalmatrizen dargestellt werden, vgl. Abschnitt A.2.2.3 im Anhang, sodass der effiziente Thomas-Algorithmus zur Lösung der Gleichungssysteme genutzt werden kann.

Im Rahmen dieser Arbeit wurden beide Varianten zur Verringerung des Fehlerterms genutzt. Es ist ebenso möglich, Nachiterationen für die alternative Faktorisierung anzuwenden, aber hierbei treten verstärkt numerische Instabilitäten auf. Die Anwendung der Nachiterationen erhöht die Rechenzeit um einen Faktor von $N_I + 1$, wohingegen bei der alternativen Faktorisierung kein zusätzlicher Rechenaufwand entsteht, denn die Berechnung der modifizierten Matrizen erfolgt einmalig und ist daher für die Gesamtlaufzeit unerheblich.

3.1.3 Randbedingungen

Die Propagation der BPM berücksichtigt keine Randbedingungen in z -Richtung, verwendet dort also inhärent offene Randbedingungen, während in transversaler Richtung verschiedene Möglichkeiten bestehen. Im Rahmen dieser Arbeit wurden Bloch-periodische Randbedingungen [89] genutzt. Damit können die häufig ohnehin periodisch angeordneten Strukturen auch bei schrägem Lichteinfall simuliert werden. Bei der Diskretisierung der Operatoren hat dies zusätzliche Einträge in den Matrizen zufolge, siehe Gl. (A.94) und (A.95) im Anhang, welche dadurch ihre rein tridiagonale Gestalt verlieren. Unter Nutzung der Sherman-Morrison-Formel [90] können die entstehenden Gleichungssysteme jedoch weiterhin effizient mit dem Thomas-Algorithmus gelöst werden, siehe die Erklärung in Abschnitt A.2.2.3 im Anhang.

3.1.4 Grenzflächen zwischen Schichten

Bislang wurde die BPM für eine in z -Richtung invariante Struktur beschrieben. Diese Anforderung wird jetzt reduziert und stattdessen ein in z -Richtung stückweise konstantes System betrachtet, d. h. es wird angenommen, dass sich der Brechungsindex durch eine endliche Zahl von Schichten mit jeweils z -invariantem Indexprofil beschreiben lässt. Durch geeignete Übergangsbedingungen an den Grenzflächen dieser Schichten können diese miteinander verbunden werden [91, 92]. Da dabei stets Reflexionen stattfinden, müssen Felder mit entgegengesetzter Propagationsrichtung ebenfalls berücksichtigt werden.

Zunächst soll wiederum der 1D-Fall, d. h. Invarianz in y -Richtung, betrachtet werden. An den Grenzflächen zweier Dielektrika sind die Komponenten E_x , E_y , $1/n^2 E_z$, H_x , H_y und H_z gemäß Gl. (2.11) stetig. Diese Bedingungen müssen für alle Komponenten in der Überlagerung von vorwärts (+) und rückwärts (−) propagierenden Feldern erfüllt sein. Im 1D-Fall folgt aus den Maxwell-Gleichungen (2.7) die Stetigkeit von $\partial E_y / \partial z$ aus der Stetigkeit von H_x . Die Felder an der Grenzfläche zwischen zwei Schichten ergeben sich unter Berücksichtigung der langsam veränderlichen Amplitude als

$$E_{y,1/2}^\pm = \hat{E}_{y,1/2}^\pm \exp(\pm i \bar{n}_{1/2} k_0 z) , \quad \frac{\partial}{\partial z} E_{y,1/2}^\pm = \left(\frac{\partial}{\partial z} \hat{E}_{y,1/2}^\pm \pm i \bar{n}_{1/2} k_0 \hat{E}_{y,1/2}^\pm \right) \exp(\pm i \bar{n}_{1/2} k_0 z) . \quad (3.17)$$

Mithilfe der Padé-Approximation, Gl. (3.3), kann die z -Ableitung als

$$\frac{\partial}{\partial z} E_{y,1/2}^\pm = \pm i \bar{n}_{1/2} k_0 \frac{P(\mathcal{A}_{yy,1/2})}{Q(\mathcal{A}_{yy,1/2})} \hat{E}_{y,1/2}^\pm \exp(\pm i \bar{n}_{1/2} k_0 z) \quad (3.18)$$

ausgedrückt werden. Aufgrund der Übergangsbedingungen gilt also:

$$\hat{E}_{y,1}^+ + \hat{E}_{y,1}^- = \hat{E}_{y,2}^+ + \hat{E}_{y,2}^- , \quad (3.19)$$

$$\bar{n}_1 \frac{P(\mathcal{A}_{yy,1})}{Q(\mathcal{A}_{yy,1})} (\hat{E}_{y,1}^+ - \hat{E}_{y,1}^-) = \bar{n}_2 \frac{P(\mathcal{A}_{yy,2})}{Q(\mathcal{A}_{yy,2})} (\hat{E}_{y,2}^+ - \hat{E}_{y,2}^-) . \quad (3.20)$$

Wenn vorwärts und rückwärts propagierende Felder in einer Schicht bekannt sind, lassen sie sich damit jeweils auch in der anderen Schicht berechnen. Die E_y -Komponente entspricht dabei der TE-Polarisation. Gl. (3.20) erfordert neben einer Faktorisierung der Polynome analog zu Gl. (3.5) nur die wiederholte Lösung linearer Gleichungssysteme mit tridiagonaler Koeffizientenmatrix, wie für die Propagation. Auf ähnliche Weise lässt sich das magnetische Feld berechnen, wie in Abschnitt A.2.2.2 im Anhang dargestellt, um den TM-polarisierten Anteil zu bestimmen.

Im 2D-Fall ist die Erfüllung der Übergangsbedingungen komplizierter, beispielsweise lautet die Maxwell-Gleichung (2.7) für die x -Komponente:

$$\frac{\partial}{\partial y} E_z - \frac{\partial}{\partial z} E_y = -i \omega \mu_0 H_x , \quad (3.21)$$

sodass $\partial E_z/\partial y - \partial E_y/\partial z$ an einer Grenzfläche stetig ist und nicht $\partial E_y/\partial z$ selbst, wie im 1D-Fall. Im Rahmen des verwendeten semi-vektoriellen Ansatzes wird die Kopplung zwischen den Feldkomponenten vernachlässigt. Daher ist es hier naheliegend, $\partial E_z/\partial y$ zu vernachlässigen. Obwohl dabei nicht unmittelbar E_z unberücksichtigt bleibt, sondern lediglich die Änderung in y -Richtung, ist diese Näherung vor allem sinnvoll, wenn E_z klein ist. Das bedeutet für die E_y -Komponente, dass sich die Propagationsrichtung etwa in der xz -Ebene befindet. Dieser Fall kann als Quasi-TE-Polarisation bezeichnet werden. Für die orthogonale Quasi-TM-Polarisation ist es sinnvoller, anstelle der deutlich fehlerbehafteten Komponente E_x die magnetische Komponente H_y zu betrachten, wie in Abschnitt 4.3.4. Für diese gelten die Überlegungen analog und die Näherung ist wiederum gut erfüllt.

Mit der semi-vektoriellen Näherung können die Gleichungen für den Übergang zwischen beiden Schichten auch hier verwendet werden. Der Operator \mathcal{A}_{yy} wird dabei wie für die Propagation in 1D-Operatoren zerlegt, um eine effiziente Lösung zu ermöglichen. Der entstehende Fehlerterm könnte ebenso mit den oben beschriebenen Verfahren reduziert werden, die durchgeführten Berechnungen zeigten jedoch mit der klassischen Faktorisierung die größte Stabilität, welche daher für die Grenzflächen genutzt wurde.

3.1.5 Bidirektionalität

Die BPM ist an sich ein unidirektionales Verfahren. Reflexionen an den Übergängen verschiedener Schichten können jedoch auf die oben beschriebene Weise einbezogen werden. Für die bidirektionale Propagation mittels der BPM durch ein gesamtes Element gibt es verschiedene Ansätze [84, 93, 94]. Hier werden Ideen aus den Ref. [95, 96] genutzt, allerdings in modifizierter Weise, um eine effiziente Berechnung zu ermöglichen. Dafür ergibt sich folgende iterative Verfahrensweise, beispielhaft für das elektrische Feld:

1. Das vorwärts laufende Feld \hat{E}^+ wird in der ersten Schicht, z. B. dem Substrat, definiert und das rückwärts laufende Feld \hat{E}^- auf null gesetzt.
2. \hat{E}^+ wird durch alle Schichten propagiert, wobei an den Grenzflächen jeweils $\hat{E}^- = 0$ angenommen wird.
3. \hat{E}^- wird in der letzten Schicht, z. B. dem Superstrat, auf null gesetzt.
4. \hat{E}^- wird bis zurück zur ersten Schicht propagiert, wobei an den Grenzflächen jeweils \hat{E}^+ aus der vorherigen Iteration genutzt wird.
5. \hat{E}^+ wird in der ersten Schicht auf das Eingangsfeld gesetzt.
6. \hat{E}^+ wird bis zur letzten Schicht propagiert, wobei an den Grenzflächen jeweils \hat{E}^- aus der vorherigen Iteration genutzt wird. Danach wird mit Schritt 3 fortgefahren.

Die separate Behandlung von \hat{E}^+ und \hat{E}^- ist notwendig, damit evaneszente Anteile korrekt berechnet werden, d. h. in der jeweiligen Propagationsrichtung abfallen. Mit dieser Iteration kann die effiziente Berechnung mittels Tridiagonalmatrizen in allen Schritten verwendet werden. Schritt 2 wurde im Vergleich zu Ref. [81] zur Verbesserung der Konvergenz modifiziert und für Schritt 5 die Beschreibung korrigiert. Wie viele Iterationen notwendig sind, hängt von der Stärke der reflektierten Felder ab. Bei schwacher Reflektivität, z. B. an einer einfachen Grenzfläche zwischen Quarzglas und Luft, bringt häufig schon ein einfacher Durchlauf der obigen Schritte plausible Ergebnisse, siehe Abschnitt 3.2.1. Bei höherem Indexkontrast sind oft wenige Wiederholungen ausreichend. Ist die Reflektivität allerdings sehr hoch, z. B. bei Resonatoren, können viele Iterationen zur Verstärkung von Abweichungen führen und das Verfahren instabil machen.

3.2 Anwendungsbeispiele

Im Folgenden werden Simulationsergebnisse der BPM für verschiedene Anwendungsfälle vorgestellt. Zunächst werden einzelne Zellen mit Effektiv-Index-Strukturen betrachtet. Diese sind klein genug, um sie mit rigorosen Verfahren effizient berechnen zu können und ermöglichen einen einfachen Vergleich der erhaltenen Transmissionsfunktionen. Anschließend wird eine aus derartigen Strukturen zusammengesetzte Metalinse analysiert. Auf diese Weise wird demonstriert, dass die BPM zu rigorosen Verfahren vergleichbare Ergebnisse liefert, dabei aber deutlich schneller berechnet werden kann. Das nachfolgende Beispiel eines Polarisations-Strahlteilers zeigt, dass die BPM die vektoriellen Eigenschaften des elektromagnetischen Feldes und die Dispersion berücksichtigt. Ein derartiges Element kann mit Verfahren, die lediglich die skalare Wellengleichung lösen, nicht beschrieben werden. Schließlich wird die BPM zur Simulation von Streulicht an einem Beugungsgitter verwendet. Dafür sind große Simulationsgebiete notwendig, die mit rigorosen Methoden kaum berechnet werden können. Die Modifikation der BPM hat im Rahmen dieser Arbeit insbesondere die Analyse von CGHs zum Ziel. Entsprechende Anwendungsfälle werden in Kap. 4 vorgestellt.

3.2.1 Basiszellen von Effektiv-Index-Strukturen

Zur Realisierung von DOEs ist die gezielte Manipulation der lokalen Reflexion bzw. Transmission von entscheidender Bedeutung. Das kann über die Nutzung von Effektiv-Index-Strukturen erreicht werden. Im Rahmen dieser Arbeit wurden überwiegend regelmäßig angeordnete Säulen bzw. Löcher in Dielektrika betrachtet. Deren effektiver Brechungsindex liegt abhängig von der Strukturgröße zwischen dem des Dielektrikums und dem der Umgebung [5, 52].

Als Beispiel dient hier die Bestimmung der Transmissionsfunktion einer Effektiv-Index-Struktur aus quadratischen Säulen bzw. Löchern auf bzw. in einem Substrat aus Quarzglas (SiO_2). Die

Beleuchtung erfolgt senkrecht aus Richtung des Substrates mit einer ebenen Welle der Wellenlänge $\lambda_0 = 632,8 \text{ nm}$. Die Tiefe der Strukturen d ist so gewählt, dass der Phasenunterschied der Propagation durch das Substrat bzw. durch die umgebende Luft etwa 2π beträgt, d. h. es ist

$$d = \frac{\lambda_0}{n - 1} \approx 1385 \text{ nm} . \quad (3.22)$$

Der Brechungsindex des Substrates ist dabei $n = 1,457$ und der von Luft ist $n = 1$. Es werden quadratische Basiszellen mit Größen von $p = 300, 500, 800 \text{ nm}$ betrachtet. Gemäß Gl. (2.27) handelt es sich nur im ersten Fall um ein tatsächliches Effektiv-Medium, während im zweiten Fall nur weitere Beugungsordnungen in Luft, aber nicht als Reflexion ins Substrat ausgeschlossen sind. Im dritten Fall findet auch in Transmission Beugung statt. Die Breite der Strukturen wird mit dem Füllfaktor f variiert, der hier den Anteil des hochbrechenden Mediums in einer xz - oder yz -Ebene bezeichnet. Die Breite der Säulen ist demnach fp , die der Löcher $p - fp$, siehe Abb. 3.1(a,b). Der Flächenfüllfaktor, der den Anteil des Mediums in einer xy -Ebene kennzeichnet, ergibt sich für Säulen bzw. Löcher als

$$f_S = f^2 , \quad \text{bzw. } f_L = 1 - (1 - f)^2 . \quad (3.23)$$

Das Simulationsgebiet umfasst jeweils eine Basiszelle der Größe $p \times p \times d$, wie in Abb. 3.1(c-h) gekennzeichnet. In x - und y -Richtung werden periodische Randbedingungen angenommen, in z -Richtung offene. Das Eingangsfeld wird als ebene Welle an der Grenzfläche zwischen Substrat und Struktur definiert. Die Transmissionsfunktion für die 0. Ordnung ergibt sich aus dem Mittelwert des Feldes in der Ebene unmittelbar oberhalb der Struktur.

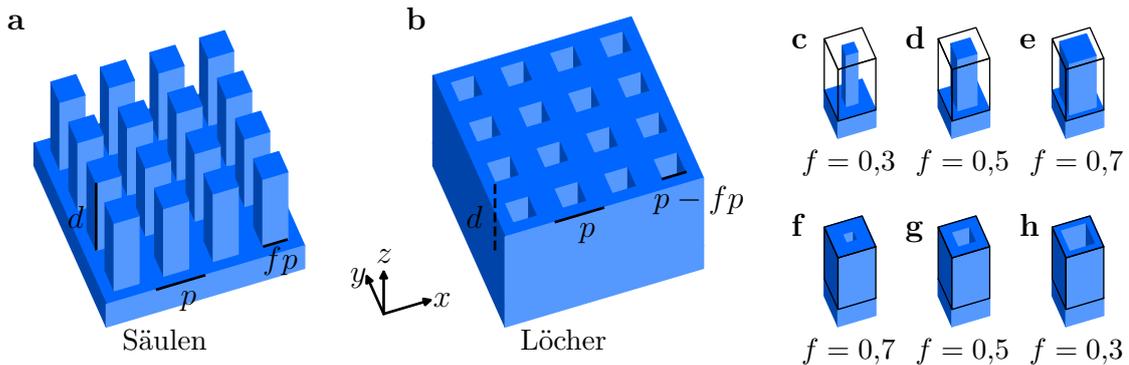


Abb. 3.1: Effektiv-Index-Struktur aus Säulen (a) bzw. Löchern (b) mit einer Strukturtiefe d , einer Periode von p und einem Füllfaktor von f ; Basiszellen mit Säulen (c-e) und Löchern (f-h). Der schwarz umrandete Bereich kennzeichnet das Simulationsgebiet.

Die für die Simulation genutzten Parameter sind in Tab. 3.1 zusammengefasst. Dabei bezeichnen N_x und N_y die Anzahl der Diskretisierungspunkte in der jeweiligen Richtung, es ist also z. B. $\delta x = p/N_x$. Die Anzahl der Iterationen bezeichnet die Zahl der zusätzlichen Durchläufe zur Einbeziehung der Reflexionen. Die Zahl 1 bedeutet dabei genau einen Durchlauf durch das Schema

in Abschnitt 3.1.5. Für die WPM wurde ein ähnliches Iterationsschema angewendet. Für die BPM wurden die beiden Ansätze mit der klassischen und alternativen Faktorisierung gemäß Gl. (3.13) bzw. (3.14) unterschieden. Die Bedeutung der übrigen Parameter wird in den Abschnitten 3.1.1 und 3.1.2 geschildert. Als Referenz dient eine RCWA-Simulation mit 8×8 Ordnungen.

Für die WPM wurde eine eigene Implementierung in vektorieller Formulierung gemäß des in den Ref. [39, 40] geschilderten Algorithmus genutzt. Die BPM wurde in der oben geschilderten Weise in Python implementiert. Zum Vergleich mit rigorosen Verfahren wurde die kommerzielle Software *Lumerical FDTD Solutions* bzw. die hausinterne Software *Moose* für die RCWA genutzt.

Tab. 3.1: Parameter der verwendeten Simulationsverfahren.

Parameter	WPM	BPM 1	BPM 2	FDTD
$N_x = N_y$	40	40	40	40
δz	25 nm	25 nm	25 nm	10 nm
Iterationen	1	1	1	
N_P		3	3	
Faktorisierung		alternativ	klassisch	
ψ		$0,25\pi$	$0,35\pi$	
N_I		0	3	

Die Ergebnisse der Simulation sind in Abb. 3.2 und 3.3 dargestellt. In den Amplitudenverläufen wird sichtbar, dass die semi-rigorosen Verfahren, WPM und BPM, teilweise vom Verlauf der rigorosen Ergebnisse, FDTD und RCWA, abweichen. Dabei sind die Abweichungen der WPM teils stark und die berechneten Amplituden mitunter bedeutend größer als eins, was einerseits unphysikalisch ist und andererseits durch exponentielles Wachstum die Stabilität der Berechnung gefährdet. Die Abweichungen der BPM sind deutlich geringer, wobei im Allgemeinen die Variante BPM 2 dichter an den rigoros berechneten Werten liegt als BPM 1. Die Ergebnisse von FDTD und RCWA sind dabei fast identisch. Für Perioden von 300 und 500 nm ist die Amplitude nahezu homogen und weicht nur aufgrund der Fresnel-Reflexion an den Grenzflächen von eins ab. Bei 800 nm lässt sich dagegen eine größere Variabilität bemerken: Hier treten weitere Beugungsordnungen auf und das Feld wird nicht nur in der 0. Ordnung transmittiert.

Die ermittelten Phasenwerte wurden jeweils als Differenz zur einfachen Grenzfläche mit $f = 0$ dargestellt. Unterschiede zwischen den Simulationsverfahren sind in dieser Darstellung kaum zu erkennen. Es fällt aber auf, dass der Verlauf für kleine Perioden, also 300 nm, fast proportional zum Flächenfüllfaktor f_S bzw. f_L ist, während für größere Perioden Abweichungen davon auftreten. Zur besseren Vergleichbarkeit wurden zudem die Phasenwerte als Differenz zu den RCWA-Ergebnissen dargestellt. Dabei zeigt sich ein ähnliches Bild wie für die Amplituden: Teils deutliche Abweichungen für die WPM, geringere für die BPM und zur Referenz quasi identische Werte bei

der FDTD. Dabei ist die Version BPM2 wiederum meist etwas akkurater als die BPM1.

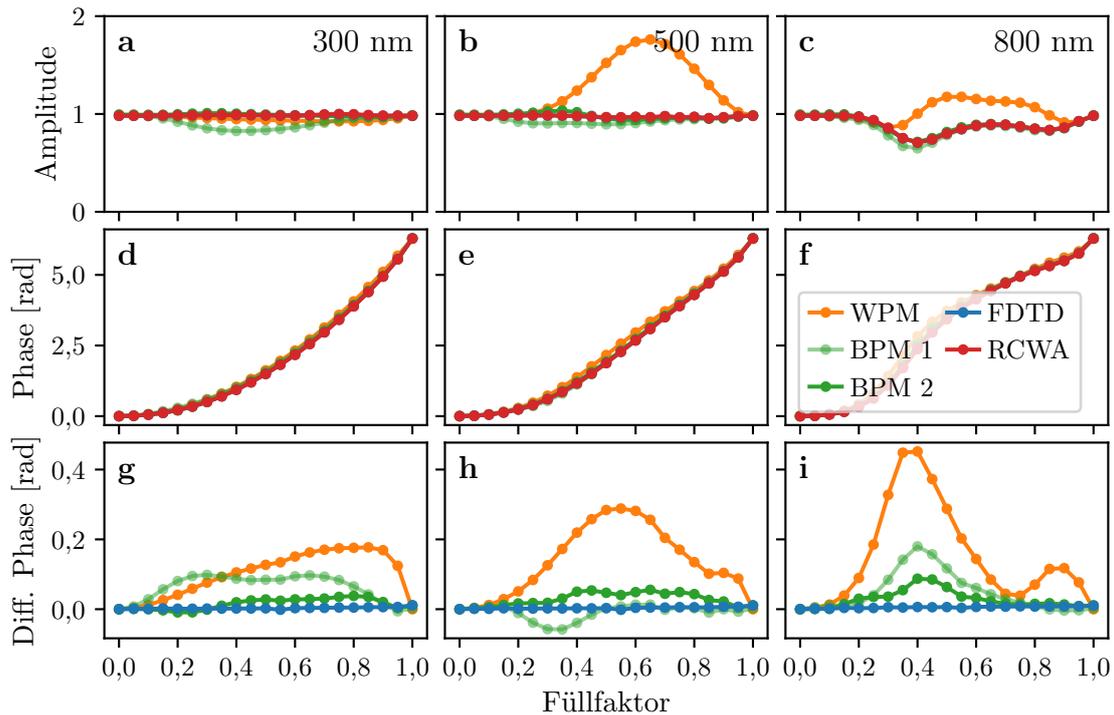


Abb. 3.2: Transmissionsfunktion für Säulenstrukturen mit Perioden von 300 (a,d,g), 500 (b,e,h) und 800 nm (c,f,i). Dargestellt sind jeweils Amplitude (a-c), Phase (d-f), sowie die Differenz zur Phase der RCWA-Simulation (g-i).

Insgesamt ist die WPM für die Simulation derartiger Strukturgrößen eher ungeeignet. Ursächlich dafür dürfte vor allem die skalare Näherung der Wellengleichung (2.52) sein. Die vektorielle Formulierung ändert daran nichts, sie beachtet lediglich die richtungsabhängigen Übergangsbedingungen an den Grenzflächen. Die BPM liefert dagegen plausible Ergebnisse, wobei die Abweichungen in Variante 2 geringer sind. Da diese aufgrund der Nachiterationen aber einen höheren Rechenaufwand hat, kann hier eine Abwägung zwischen Genauigkeit und Rechendauer getroffen werden. Die Ergebnisse der beiden rigorosen Verfahren unterscheiden sich fast nicht. Daher ist es ausreichend, für weitere Beispiele die bei größeren Simulationsgebieten schnellere FDTD zu verwenden. In Abschnitt 4.3.4 wird die Genauigkeit der BPM in Abhängigkeit der auftretenden Ablenkungswinkel eines CGHs diskutiert. Ähnliche Limitationen sind bezüglich der modellierbaren Einfallswinkel zu erwarten.

3.2.2 Effektiv-Index-Metalinse

Das betrachtete Element soll einfallendes kollimiertes Licht im Abstand der Brennweite F fokussieren. Es fungiert demnach als Linse, erreicht diese Wirkung aber nicht durch die Form, sondern durch die Anordnung der Mikrostrukturen, weshalb es als Metalinse bezeichnet wird.

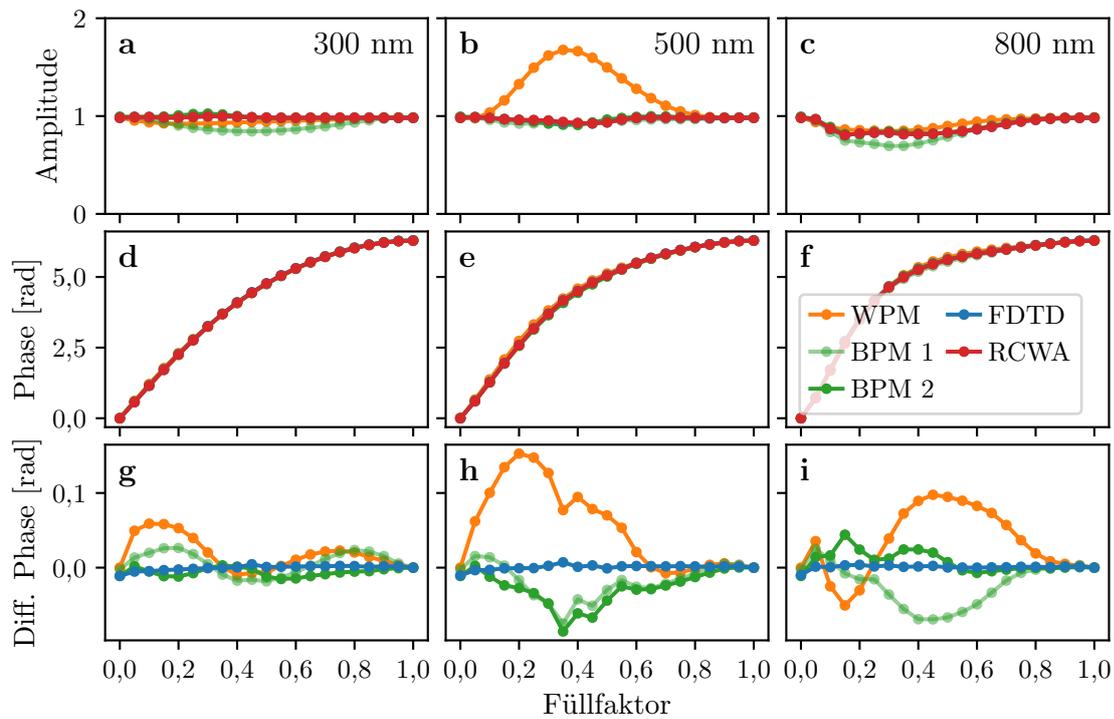


Abb. 3.3: Transmissionsfunktion für Lochstrukturen mit Perioden von 300 (a,d,g), 500 (b,e,h) und 800 nm (c,f,i). Dargestellt sind jeweils Amplitude (a-c), Phase (d-f), sowie die Differenz zur Phase der RCWA-Simulation (g-i).

3.2.2.1 Design

Die Fokussierung erfordert eine Transmissionsfunktion von

$$t_L(x, y) = \exp(i\phi_L(x, y)) = \exp\left(-i\frac{2\pi}{\lambda_0}\sqrt{F^2 + x^2 + y^2}\right) \quad (3.24)$$

für das DOE auf der Rückseite eines planen Substrates [6, 10]. Dabei ist $\lambda_0 = 632,8$ nm die Designwellenlänge. Die Realisierung dieser Transmission erfolgt mithilfe der im vorherigen Abschnitt geschilderten Effektiv-Index-Strukturen aus Quarzglas. Es wurden 6 Strukturen ausgewählt, die möglichst äquidistante Phasenschritte ermöglichen, siehe Abb. 3.4(a-f). Zur Bestimmung der Transmissionsfunktionen wurde eine FDTD-Simulation verwendet. Die Parameter der einzelnen Strukturen sind in Tab. 3.2 zusammengefasst. Die Basiszellen sind quadratisch mit einer Seitenlänge von $p = 400$ nm und einer Strukturtiefe von $d = 1155$ nm, was einer Reduktion um $1/6$ gemäß Gl. (3.22) darstellt, sodass der maximale Phasenunterschied $5\pi/3$ beträgt.

Für die Zuordnung der Strukturen zu einzelnen Positionen werden die idealen Phasenwerte $\phi_L(x, y)$ über die jeweilige Zelle gemittelt und die Basiszelle mit dem nächstgelegenen Phasenwert ausgewählt, siehe Abb. 3.4(g). Dieses Vorgehen entspricht der Anwendung der LPA. Das gesamte Element umfasst 48×48 Zellen und hat eine Fläche von $19,2 \times 19,2 \mu\text{m}^2$ bei einer Fokusslänge von $F = 9,6$ bzw. $19,2 \mu\text{m}$. Die numerische Apertur (NA) der Metalinse beträgt 0,71 bzw. 0,45, sodass eine paraxiale Beschreibung unzulänglich wäre.

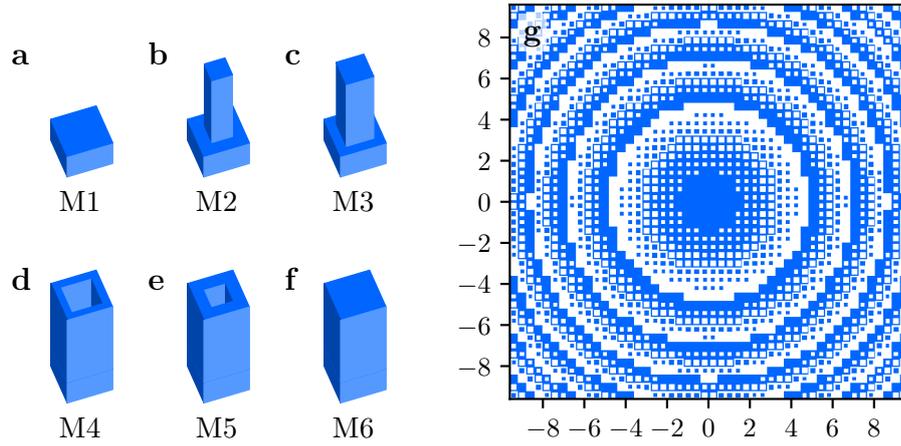


Abb. 3.4: Basiszellen mit verschiedenen Subwellenlängenstrukturen (a-f), Geometrie der simulierten Metalinse mit $F = 19,2 \mu\text{m}$ (g), Längenangaben sind in μm .

Tab. 3.2: Seitenlängen der quadratischen Strukturen für die verschiedenen Phasenschritte der Metalinse. Amplitude und Phase sind in Relation zum Feld im Substrat vor der Grenzfläche angegeben.

Bezeichnung	Typ	Seitenlänge	Amplitude	Phase
M1	keine Struktur	—	1,184	-1,098
M2	Säule	180 nm	1,185	-0,003
M3	Säule	240 nm	1,197	0,883
M4	Loch	280 nm	1,203	1,782
M5	Loch	200 nm	1,185	3,008
M6	volle Schicht	—	1,184	-2,131

3.2.2.2 Simulationsparameter

Die Funktionsweise des DOEs wurde mithilfe der konventionellen LPA, der oben geschilderten BPM und der rigorosen FDTD simuliert. Innerhalb des Substrates wurde die Amplitude auf eins normiert. In allen Berechnungen wurden periodische Randbedingungen in transversaler Richtung angewandt. Für die LPA wurde die mittels FDTD berechnete, polarisationsunabhängige Transmissionsfunktion der Basiszellen genutzt. Deren leicht verschiedene Amplituden wurden berücksichtigt. Die BPM wurde in der Formulierung für das elektrische Feld verwendet, mit einer Diskretisierung von $\delta x = \delta y = 20 \text{ nm}$ bzw. $N_x = N_y = 20$ und $\delta z = 50 \text{ nm}$, Padé-Approximationen der Ordnung $N_P = 4$ und ansonsten den in Tab. 3.1 aufgeführten Parametern für die beiden BPM-Varianten. Bei diesen Werten führt eine weitere Verfeinerung der Parameter zu keiner signifikanten Variation der Ergebnisse. Die FDTD-Simulation wurde mit einer Diskretisierung von $\delta x = \delta y = 20 \text{ nm}$ und $\delta z = 10 \text{ nm}$ durchgeführt. Durch die Nutzung (anti-)symmetrischer Randbedingungen konnte das Simulationsgebiet auf einen Quadranten der xy -Ebene beschränkt werden. In z -Richtung wurden

PML-Randbedingungen genutzt, welche den offenen Randbedingungen der BPM entsprechen.

3.2.2.3 Ergebnisse der Berechnung

In allen Simulationen wurde das elektrische Feld in der Ebene unmittelbar nach dem DOE ermittelt und für die weitere Auswertung genutzt. Die erhaltenen Resultate sind in Abb. 3.5 dargestellt. Es ist deutlich zu erkennen, dass die Amplitude in der LPA nahezu homogen ist (a), wohingegen bei BPM und FDTD Kopplungseffekte auftreten (b,c), insbesondere an den Übergängen von Zellen ohne und mit vollem Füllfaktor (M1 bzw. M6). Der Verlauf der Phase wird durch die LPA im Allgemeinen korrekt wiedergegeben, doch im Detail sind Abweichungen zu den anderen Verfahren vorhanden (d-f). Im direkten Vergleich von BPM und FDTD können nur geringfügige Abweichungen beobachtet werden, welche hauptsächlich an den Ecken der Strukturen auftreten. Dort kommt es in einer rigorosen Simulation zur Kopplung verschiedener Polarisationskomponenten, was in der BPM nicht modelliert wird.

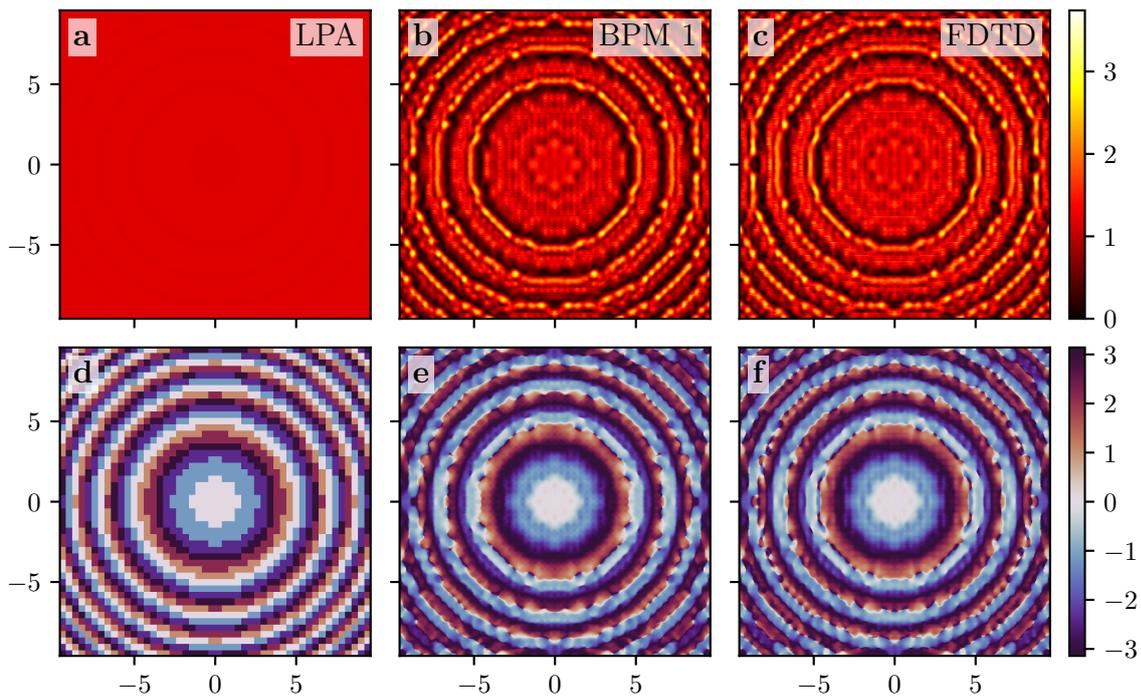


Abb. 3.5: Amplitude (a-c) und Phase (d-f) des elektrischen Feldes nach dem Metalinsen-DOE mit $F = 19,2\mu\text{m}$ gemäß der Simulation mit LPA (a,d), BPM (b,e) und FDTD (c,f). Längen sind in μm angegeben.

Die Qualität der Simulationen kann durch den Vergleich des Fernfeldes von LPA und BPM zu dem der FDTD quantifiziert werden. Das Fernfeld $U(k_x, k_y)$ wird gemäß Gl. (2.23) als FT des Nahfeldes $u(x, y)$ direkt nach dem DOE berechnet. Dann wurde der Korrelationskoeffizient (corr) zwischen zwei Feldern betrachtet, unter Beschränkung auf die im Freiraum propagierenden

Ordnungen:

$$\text{corr}(U_1, U_2) = \frac{\sum_{k_x^2+k_y^2 \leq k_0^2} (U_1(k_x, k_y) - \bar{U}_1) (U_2^*(k_x, k_y) - \bar{U}_2^*)}{\sqrt{\sum_{k_x^2+k_y^2 \leq k_0^2} |U_1(k_x, k_y) - \bar{U}_1|^2 \sum_{k_x^2+k_y^2 \leq k_0^2} |U_2(k_x, k_y) - \bar{U}_2|^2}}. \quad (3.25)$$

Hierbei bezeichnet U^* die komplexe Konjugation und \bar{U} das arithmetische Mittel der propagierenden Ordnungen. Die Korrelation zwischen BPM und FDTD ist dabei in allen Fällen hoch, wie aus Tab. 3.3 ersichtlich wird. Die LPA zeigt dagegen eine deutliche Abweichung, die bei einer kürzeren Brennweite bzw. höheren NA stärker ist. In diesem Fall wechseln die Strukturen häufiger und Kopplungseffekte spielen eine größere Rolle. Die BPM-Simulation wurde unter Verwendung des magnetischen Feldes wiederholt, wobei sich ähnliche Resultate mit etwas geringerer Korrelation ergaben. Mithilfe der ASPW wurde das Strahlprofil im Fokus der Linse berechnet, basierend

Tab. 3.3: Betrag der Korrelationskoeffizienten der Fernfelder des jeweiligen Verfahrens und der FDTD, für die BPM jeweils basierend auf dem elektrischen (el.) und magnetischen (mag.) Feld.

Brennweite	LPA	BPM 1 el.	BPM 2 el.	BPM 1 mag.	BPM 2 mag.
9,6 μm	0,567	0,995	0,996	0,987	0,988
19,2 μm	0,797	0,996	0,997	0,991	0,991

auf dem Nahfeld der verschiedenen Verfahren, siehe Abb. 3.6. Wiederum zeigt sich eine hohe Übereinstimmung von BPM und FDTD, während die LPA signifikant abweicht. Insbesondere die zentrale Amplitude ist deutlich größer, sodass die Effizienz der Linse durch die LPA überschätzt wird.

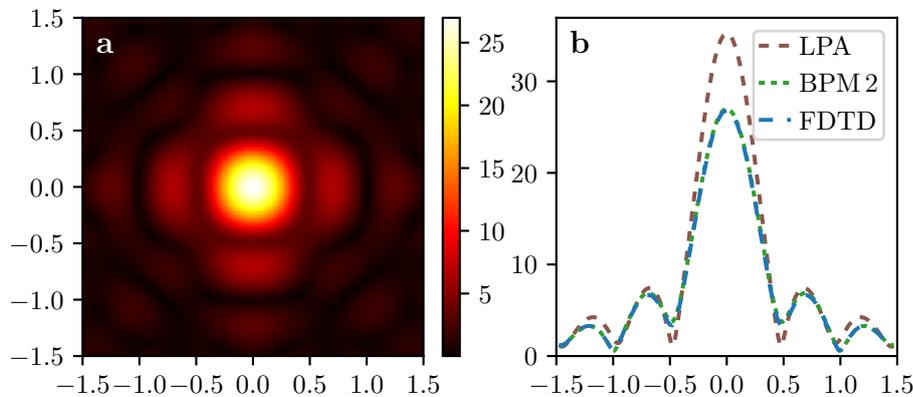


Abb. 3.6: Strahlprofil im Fokus der Metalinse (a) mit Brennweite $F = 9,6 \mu\text{m}$ gemäß BPM-Simulation, Querschnitt des Strahlprofils zum Vergleich der Verfahren (b).

3.2.2.4 Vergleich der Laufzeiten

Zum Vergleich der Laufzeiten von BPM und FDTD wurde ein DOE mit verringerter Größe von 16×16 Zellen berechnet. Die Laufzeiten von BPM1 und BPM2 betragen mit obigen Parametern 13s bzw. 46s bei einer Prozessorgeschwindigkeit von 3,3GHz. Eine entsprechende FDTD-Simulation benötigte auf demselben Computer 2,5 Stunden. Um einen fairen Vergleich durchzuführen, wurde dabei jeweils nur ein einzelner Prozess genutzt und in der FDTD auf die Ausnutzung von (anti-)symmetrischen Randbedingungen verzichtet, die man auch in der BPM verwenden könnte. Demzufolge war die BPM für dieses Beispiel 700- bzw. 200-mal schneller. Für eine weitere Reduktion der Laufzeit können in der BPM etwas größere Parameter genutzt werden, ohne dass die Genauigkeit deutlich abnimmt. Aufgrund der vollständigen Diskretisierung mithilfe von Tridiagonalmatrizen skaliert die Laufzeit linear mit der Anzahl der verwendeten Gitterpunkte.

Die geschilderten Ergebnisse zeigen, dass die BPM die wellenoptische Simulation gesamter DOEs ermöglicht. Dabei wird eine mit rigorosen Verfahren vergleichbare Genauigkeit erzielt, während die Laufzeit um mehrere Größenordnungen geringer ist. Insbesondere werden lokale Variationen der Amplitude beschrieben, die aus der Wechselwirkung benachbarter verschiedenartiger Basiszellen herrühren. Im Gegensatz zu einer skalaren BPM [36, 61] werden dabei polarisationsabhängige Übergangsbedingungen berücksichtigt.

3.2.3 Polarisations-Strahlteiler

Das betrachtete Strahlteiler-DOE soll einfallendes, unpolarisiertes Licht in zwei orthogonale, linear polarisierte Polarisationszustände aufspalten. Das Element wird zunächst auf herkömmliche Weise, basierend auf der LPA, entwickelt. Dafür werden wiederum Effektiv-Index-Strukturen genutzt, die aber im Gegensatz zum vorherigen Beispiel im Allgemeinen eine rechteckige Grundfläche besitzen. Bei der wellenoptischen Analyse fallen Abweichungen zum erwarteten Verhalten auf. Anschließend wurden die Parameter der Strukturen unter Nutzung der BPM optimiert, um ein Design zu erhalten, das auch bei einer rigorosen Simulation die gewünschte Funktionalität liefert.

3.2.3.1 Design

Idealerweise sollte die Phase der Transmissionsfunktion ein lineares Profil entlang der x -Koordinate aufweisen, allerdings mit umgekehrten Vorzeichen je nach Polarisationszustand:

$$\phi_{\text{TE}}(x, y) = +2\pi \frac{x}{Mp}, \quad \phi_{\text{TM}}(x, y) = -2\pi \frac{x}{Mp}. \quad (3.26)$$

Dabei ist p die Größe der quadratischen Basiszellen und M die Anzahl derartiger Zellen, aus denen eine 2π -Periode gebildet wird. Die Polarisationszustände TE und TM werden durch die Komponenten E_y und H_y repräsentiert. Mit einem derartigen Phasenprofil wird TE-polarisiertes Licht vollständig in die $+1$. und TM-polarisiertes in die -1 . Ordnung gebeugt. Für das Design

werden rechteckige Säulen gleicher Höhe aus Aluminiumoxid auf einem Quarzsubstrat verwendet. Ihre Transmissionsfunktion wurde gemäß der LPA für die einzelnen Zellen mit periodischen Randbedingungen berechnet, um Abmessungen zu finden, die sechs Phasenschritte in entgegengesetzter Richtung für TE- und TM-Polarisation ermöglichen ($M = 6$, einfaches Design) [97–101]. Zum Vergleich wurde jede Zelle zweimal platziert ($M = 12$, doppeltes Design), da dann die LPA besser erfüllt ist. Unter Nutzung der BPM zur Berechnung der Beugungseffizienz fand eine Optimierung der Strukturabmessungen statt ($M = 6$, optimiertes Design). Dazu wurde iterativ die Auswirkung kleiner Änderungen der Parameter berechnet und das Design jeweils gemäß der Richtung mit der größten Verbesserung variiert. Die erhaltenen Elemente sind in Abb. 3.7 dargestellt.

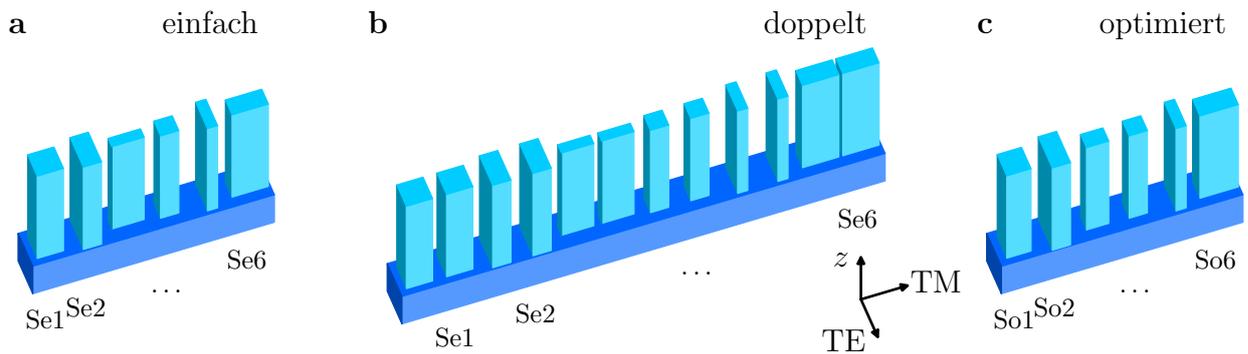


Abb. 3.7: Einheitszellen der verschiedenen Designvarianten des Strahlteilers: einfach (a), doppelt (b) und optimiert (c). Das einfallende Licht propagiert in z -Richtung, wobei das elektrische Feld je nach Polarisation entsprechend ausgerichtet ist.

Wie für die Metalinse wurde eine Wellenlänge von $\lambda_0 = 632,8 \text{ nm}$ und eine quadratische Zellgröße von $p = 400 \text{ nm}$ verwendet. Eine Elementarzelle besteht aus 6×1 (12×1) Zellen mit einer Fläche von $2,4 \times 0,4 \mu\text{m}^2$ ($4,8 \times 0,4 \mu\text{m}^2$). Die Abmessungen der einzelnen Säulen sind in Tab. 3.4 zusammengefasst. Die Strukturtiefe beträgt $d = 1,65 \mu\text{m}$. Wiederum wird eine periodische Fortsetzung der Strukturen in x - und y -Richtung angenommen. Der Brechungsindex bei λ_0 beträgt 1,457 für das Quarzsubstrat und 1,766 für das Aluminiumoxid. Der Ablenkwinkel der transmittierten ± 1 . Ordnung beträgt $\pm 15,3^\circ$ ($\pm 7,6^\circ$).

3.2.3.2 Simulationsparameter

Das Fernfeld des Strahlteilers wurde jeweils aus dem Feld hinter dem Element berechnet, um die Intensitäten der -1 ., 0 . und $+1$. Ordnung für TE- und TM-Polarisation bei Wellenlängen von 500 nm bis 800 nm zu vergleichen. Die Materialdispersion wurde dabei berücksichtigt. Für die LPA wurde eine polarisationsabhängige Transmissionsfunktion mithilfe der FDTD bestimmt. In den BPM-Simulationen wurden die in Tab. 3.1 angegebenen Parameter verwendet bis auf eine Diskretisierung von $\delta z = 50 \text{ nm}$ in Propagationsrichtung. Die Simulation wurde für jede der betrachteten Wellenlängen wiederholt. Für die rigorose Berechnung mittels FDTD wurde eine

Tab. 3.4: Strukturgrößen der rechteckigen Säulen aus Abb. 3.7 für den Polarisations-Strahlteiler.

Bezeichnung	Breite x	Breite y	Bezeichnung	Breite x	Breite y
Se1	260 nm	260 nm	So1	240 nm	260 nm
Se2	180 nm	360 nm	So2	180 nm	380 nm
Se3	320 nm	100 nm	So3	220 nm	160 nm
Se4	180 nm	180 nm	So4	180 nm	180 nm
Se5	100 nm	320 nm	So5	100 nm	320 nm
Se6	360 nm	180 nm	So6	380 nm	200 nm

Diskretisierung von $\delta x = \delta y = \delta z = 20$ nm verwendet. Diese genügt hier, da auch Materialgrenzen in halbem Diskretisierungsabstand korrekt berücksichtigt werden.

3.2.3.3 Ergebnisse der Berechnung

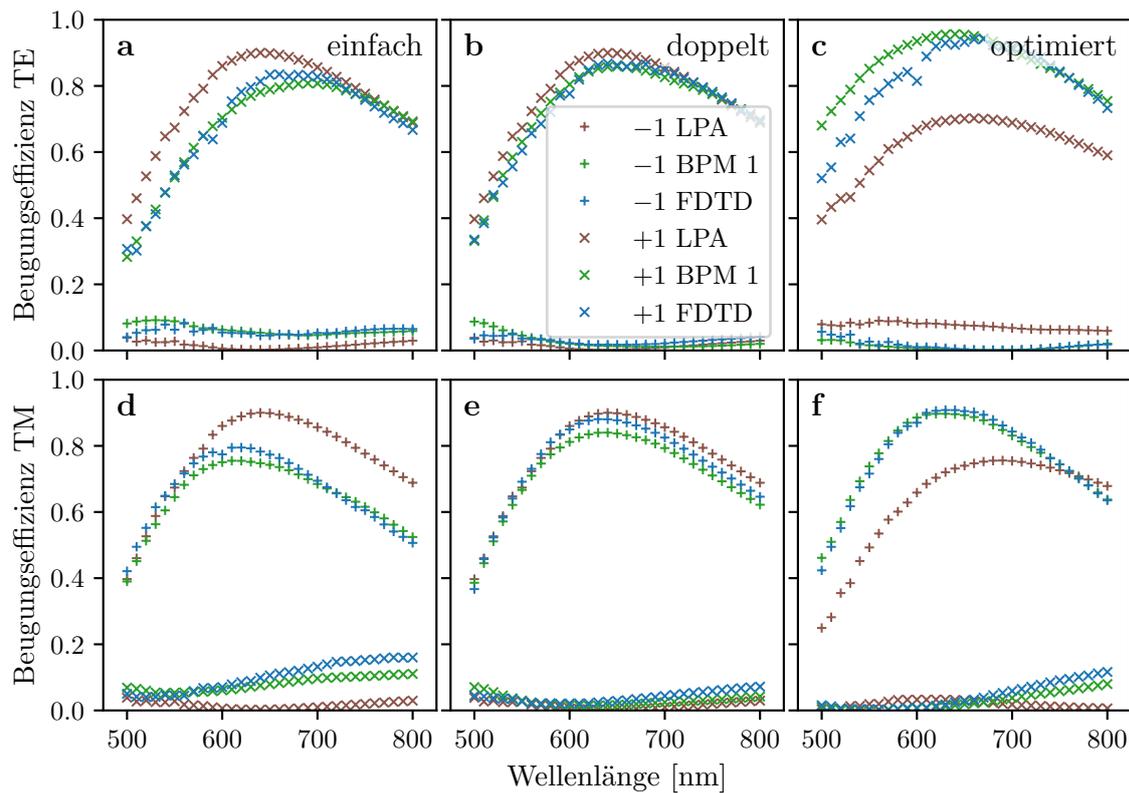


Abb. 3.8: Beugungseffizienz der $-1.$ und $+1.$ Ordnung für einfallendes Licht in TE- (a-c) bzw. TM-Polarisation (d-f) und die verschiedenen Designvarianten einfach (a,d), doppelt (b,e) und optimiert (c,f). Verglichen werden jeweils die Ergebnisse der Simulation mittels LPA, BPM 1 und FDTD. Die Effizienzen sind auf die gesamte transmittierte Leistung normiert.

Aus Abb. 3.8 lässt sich entnehmen, dass das betrachtete DOE seine Funktionalität im We-

sentlichen erfüllt: TE-polarisiertes Licht wird größtenteils in die +1. Ordnung gebeugt (a-c), TM-polarisiertes in die -1. (d-f). Das jeweilige Maximum bzw. Minimum ist meist in der Nähe der Designwellenlänge von 632,8 nm. Die für das einfache Design aus der LPA erwartete Qualität kann durch die FDTD allerdings nicht bestätigt werden. Insbesondere ist die maximale Effizienz in den jeweiligen Ordnungen geringer als durch die LPA vorhergesagt. Andererseits gibt es eine gute Übereinstimmung zwischen BPM und FDTD. Daran zeigt sich, dass der herkömmliche Designprozess fehlerbehaftet ist, da er die Wechselwirkung verschiedenartiger Strukturen nicht korrekt berücksichtigt. Im Fall des doppelten Designs führt die LPA zu praktisch den gleichen Ergebnissen, die jetzt aber deutlich besser mit BPM und FDTD übereinstimmen. Hier treten bereits weniger Strukturänderungen zwischen benachbarten Zellen auf, sodass die Anwendung der LPA eher gerechtfertigt ist. Schließlich zeigt das optimierte Design eine hohe Effizienz in den BPM- und FDTD-Simulationen, während die LPA abweichende Resultate hervorbringt. Die in der Optimierung mithilfe der BPM erwartete Leistungsfähigkeit wird so bestätigt. Bei etwa 600 nm kann in den FDTD-Daten ein Abfall beobachtet werden, im Gegensatz zur BPM. Die Ursache dafür könnte ein resonantes Verhalten sein, das sich in der BPM nur schwer abbilden lässt.

Aus den erhaltenen Beugungseffizienzen η lässt sich die Polarisierungseffizienz η_P berechnen, z. B. als

$$\eta_{P,+1} = \frac{|\eta_{+1}^{\text{TE}} - \eta_{+1}^{\text{TM}}|}{|\eta_{+1}^{\text{TE}} + \eta_{+1}^{\text{TM}}|}. \quad (3.27)$$

für die +1. Ordnung. Die Ergebnisse sind in Tab. 3.5 dargestellt. Sie unterstützen die bereits geschilderten Beobachtungen: Mit der LPA wird die Leistungsfähigkeit des einfachen Designs überschätzt, ebenso in geringerem Ausmaß die des doppelten, während das optimierte Design unterschätzt wird. Dagegen stimmen die Resultate von BPM und FDTD deutlich besser überein, wobei die Abweichung für die Variante BPM 1 etwas geringer als für die BPM 2 ist. Damit wird die gegenüber dem einfachen Design verbesserte Effizienz des optimierten Designs bestätigt.

Tab. 3.5: Polarisierungseffizienz der verschiedenen Designs bei einer Wellenlänge von 630 nm.

Design	+1. Ordnung (TE)				-1. Ordnung (TM)			
	LPA	BPM 1	BPM 2	FDTD	LPA	BPM 1	BPM 2	FDTD
einfach	0,99	0,83	0,75	0,80	0,99	0,86	0,81	0,88
doppelt	0,99	0,97	0,92	0,95	0,99	0,97	0,91	0,96
optimiert	0,91	0,98	0,94	0,96	0,80	0,99	0,99	0,98

Die Simulation eines derartigen polarisationsabhängigen DOEs ist mit skalaren Methoden nicht möglich. Das gezeigte Beispiel ist relativ klein und kann auch mit rigorosen Verfahren analysiert werden, allerdings bietet die reduzierte Laufzeit der BPM für die durchgeführte Optimierung

bereits einen deutlichen Vorteil. Zudem lässt sich die BPM auf deutlich größere Simulationsgebiete anwenden.

3.2.4 Streulichtanalyse für Beugungsgitter

Für die Berechnung und insbesondere Optimierung von Beugungsgittern ist nur ein 1D-Simulationsgebiet mit einer Ausdehnung in der Größenordnung der Wellenlänge erforderlich. Diese Aufgabe kann mit rigorosen Methoden, wie der RCWA, mit hoher Präzision gelöst werden. Bei der Herstellung der Gitter kommt es aber zu Abweichungen vom Design, die entsprechend mit einer Änderung der optischen Eigenschaften einhergehen. So führt die Kantenrauheit zur Entstehung von Streulicht [102, 103], welches z. B. bei Anwendungen in der Spektroskopie fehlerhafte Signale hervorrufen kann. Um diesen Effekt numerisch zu untersuchen und die Auswirkung von verschiedenen Ausprägungen der Rauheit zu vergleichen, ist eine 2D-Simulation notwendig, da die lokalen Abweichungen über die gesamte Fläche eines Elementes variieren. Die Größe der zu betrachtenden Gebiete erlaubt die rigorose Simulation nur unter Einsatz von massiven Rechenkapazitäten [104]. Aufgrund ihrer geringeren Rechenzeit ist die BPM für derartige Untersuchungen allerdings gut geeignet.

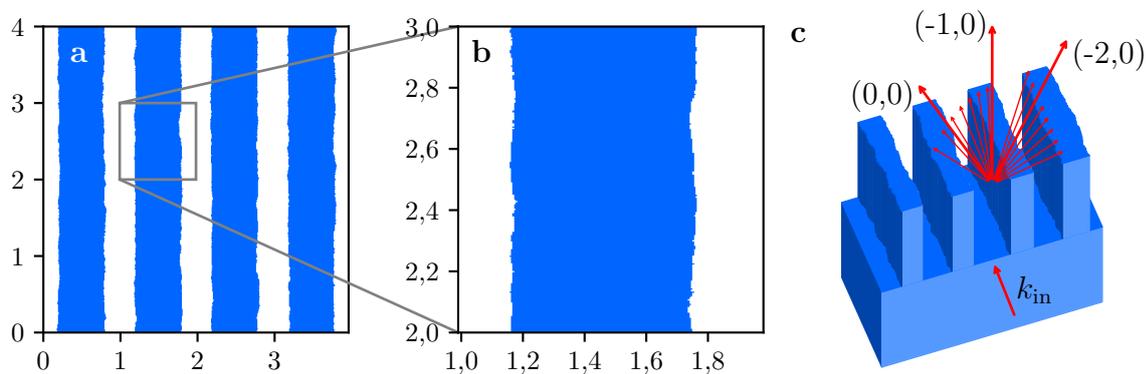


Abb. 3.9: Einheitszelle mit künstlicher Kantenrauheit und Abmessungen in μm (a,b), schematische Darstellung des Gitters und der Lage der Beugungsordnungen (c). Die Hauptbeugungsordnungen des Gitters sind beschriftet, ein Teil der weiteren Ordnungen wird durch die kürzeren Pfeile angedeutet.

Da die genaue Form der Kanten eines Gitters das Ergebnis eines stochastischen Prozesses ist, weisen die Strukturen keine Periodizität auf, sodass eine Beschreibung der gesamten beleuchteten Fläche nötig wäre, um den Einfluss der Kantenrauheit vollständig zu simulieren. Um die numerische Berechnung zu ermöglichen, wird ein Modellsystem betrachtet: Eine endliche Einheitszelle der Größe $w_x \times w_y$ wird definiert, die einige Gitterperioden in x -Richtung und eine ähnliche Ausdehnung in y -Richtung umfasst, wobei periodische Randbedingungen angenommen werden. Eine gewisse Rauheit der Gitterkanten wird zufallsbasiert vorgegeben, siehe Abb. 3.9.

Unabhängig vom Simulationsverfahren erzeugt dieses Modellsystem ein periodisches Nahfeld, aus dem sich z. B. gemäß Gl. (2.23) das Fernfeld berechnen lässt. Für ein ideales Gitter haben dabei nur die Hauptbeugungsordnungen der Gittergleichung (2.26) eine von null verschiedene Amplitude. Mit der künstlichen Kantenrauheit wird ein gewisser Teil des Lichtes auch in die anderen Beugungsordnungen gebeugt. Aufgrund der Periodizität des Modellsystems erlangt man auf diese Weise nur an diskreten Punkten eine Information über das Streulicht. Bei einer Vergrößerung des Simulationsgebietes erhöht sich die Anzahl der Ordnungen, deren Fernfeld beschrieben wird, sodass sich genauere Erkenntnisse über die Verteilung des Streulichtes gewinnen lassen. Die genaue Lage der Ordnungen m_x, m_y wird dabei nur durch die künstlich eingeführte Größe der Einheitszelle bestimmt und hat keine physikalische Bedeutung. Stattdessen erwartet man eine kontinuierliche Verteilung des Streulichtes. Um einen realistischen Eindruck zu erhalten, rechnet man daher die Effizienz der einzelnen Ordnungen $\eta_{m_x m_y}$ in eine auf den Raumwinkel bezogene Größe um, das sogenannte *Angular Resolved Scattering* (winkel aufgelöste Streuung, ARS):

$$\text{ARS}_{m_x m_y} = \frac{n^2 w_x w_y}{\lambda_0^2} \cos \vartheta_{m_x m_y} \eta_{m_x m_y} . \quad (3.28)$$

Darin ist n der Brechungsindex des umgebenden Mediums, λ_0 die Vakuumwellenlänge und $\vartheta_{m_x m_y}$ der Polarwinkel der jeweiligen Beugungsordnung.

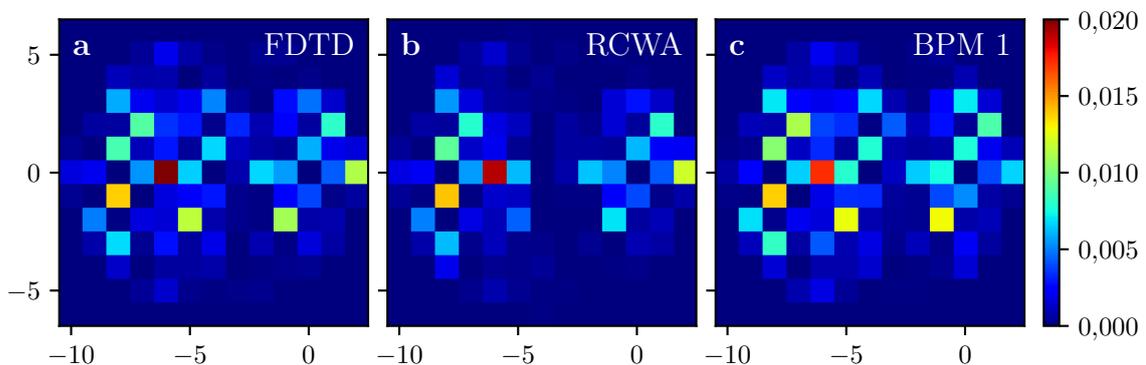


Abb. 3.10: Fernfeld (ARS) mit Kantenrauheit für kleine Einheitszelle gemäß Simulation mittels FDTD (a), RCWA (b) und BPM (c).

Eine beispielhafte Berechnung des Streulichtes wurde für ein Quarzglasgitter mit einer Periode von $p_x = 0,99 \mu\text{m}$ durchgeführt. Es besitzt eine Gittertiefe von $d = 2 \mu\text{m}$ und wird rückseitig mit Licht der Wellenlänge $\lambda_0 = 0,633 \mu\text{m}$ in TE-Polarisation unter einem Einfallswinkel von $\vartheta_{\text{in}} = 26^\circ$ bestrahlt. Dabei ergeben sich 3 Hauptbeugungsordnungen, siehe Abb. 3.9(c). Zunächst wurde eine vergleichsweise kleine Einheitszelle mit Abmessungen von $w_x = 4p_x = 3,96 \mu\text{m}$ und $w_y = 4 \mu\text{m}$ betrachtet. Diese lässt sich mit den rigorosen Methoden FDTD und RCWA simulieren, was einen Vergleich mit den Ergebnissen der BPM ermöglicht. Die Simulation wurde jeweils für 40 Einheitszellen mit zufällig erzeugter Kantenrauheit durchgeführt und das gemittelte Fernfeld in Abb. 3.10 dargestellt. Die Hauptbeugungsordnungen des Gitters sind dabei ausgeblendet, da

sie um einige Größenordnungen stärker sind. Es ist gut zu erkennen, dass die grundsätzliche Verteilung für alle Verfahren ähnlich ist. Im Detail unterscheiden sich die Ergebnisse, wobei die rigorosen Verfahren aufgrund der Diskretisierung bzw. Anzahl der berücksichtigten Ordnungen ebenfalls nicht exakt übereinstimmen. Die Abweichungen der BPM sind dabei nicht wesentlich größer als die zwischen FDTD und RCWA. Für die FDTD-Simulation wurde eine Diskretisierung von $\delta x = \delta y = \delta z = 20$ nm genutzt und für die BPM $\delta x = \delta y = 10$ nm, $\delta z = 20$ nm und ansonsten die in Tab. 3.1 gegebenen Parameter für die Variante BPM 1.

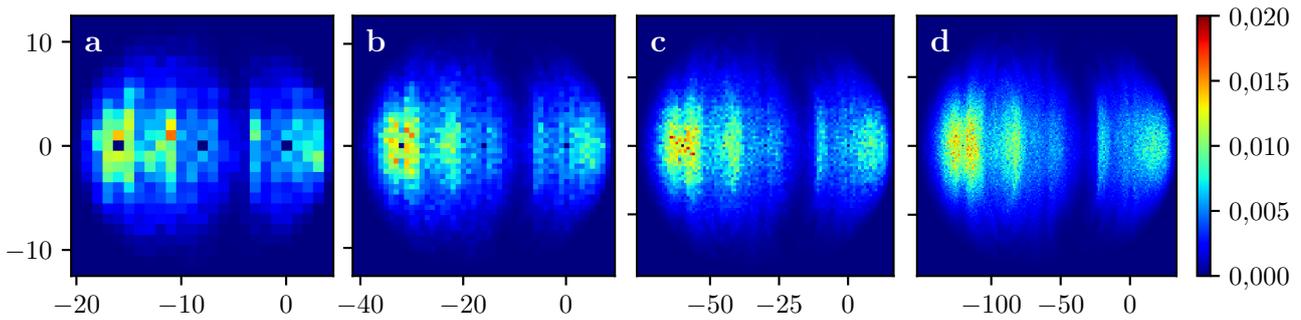


Abb. 3.11: Fernfeld (ARS) mit Kantenrauheit gemäß BPM-Simulation für Einheitszellen mit Größen von $7,92 \times 8 \mu\text{m}^2$ (a), $15,84 \times 16 \mu\text{m}^2$ (b), $29,7 \times 30 \mu\text{m}^2$ (c) und $59,4 \times 60 \mu\text{m}^2$ (d).

Die BPM-Simulation wurde für größere Einheitszellen wiederholt, siehe Abb. 3.11, wobei jeweils die Parameter zur Erzeugung der Kantenrauheit identisch sind. Damit lässt sich die Streulichtverteilung mit zunehmend höherer Auflösung darstellen, wodurch erst ein realistischer Eindruck der tatsächlichen Verteilung entsteht. Die Diskretisierung wurde hier mit $\delta x = 20$ nm, $\delta y = 40$ nm, $\delta z = 80$ nm etwas gröber gewählt, was die Ergebnisse aber nur geringfügig beeinflusst.

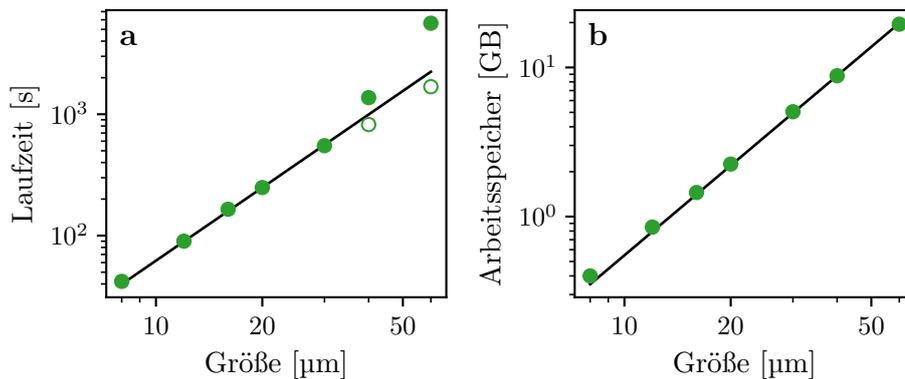


Abb. 3.12: Laufzeit der gesamten Simulation (a) und Speicherbedarf der einzelnen Prozesse (b) der BPM-Simulation von je 40 Elementen in Abhängigkeit von der Größe der Einheitszelle. Die ungefüllten Kreise kennzeichnen die extrapolierte Laufzeit für ausreichenden Arbeitsspeicher. Die Geraden beschreiben eine quadratische Abhängigkeit von der Größe.

Wiederum erfolgte eine Simulation von jeweils 40 verschiedene Einheitszellen, wobei jeweils zehn Berechnungen parallel durchgeführt wurden. Die benötigte Laufzeit und der jeweilige Speicherbedarf sind in Abb. 3.12 dargestellt. Für die größten Einheitszellen musste die Zahl der Prozesse aufgrund des begrenzten Arbeitsspeichers reduziert werden, weshalb im Diagramm zusätzlich die für einen ausreichenden Arbeitsspeicher extrapolierte Laufzeit angegeben ist. Aus der doppelt logarithmischen Darstellung lässt sich entnehmen, dass der genutzte Algorithmus in Bezug auf die transversale Ausdehnung tatsächlich einen Laufzeit- und Speicherbedarf von $\mathcal{O}(N^2)$ besitzt. Damit wird der effiziente Einsatz der BPM für große Simulationsgebiete ermöglicht.

4 Analyse der optischen Funktion durch Polynomapproximation

In diesem Kapitel werden die optischen Eigenschaften von DOEs untersucht, die aus Mikrostrukturen in einem regelmäßigen Gitter zusammengesetzt sind, wie beispielsweise CGHs oder Metalinsen. Zur Vereinfachung beschränkt sich die Betrachtung dabei auf Elemente, die aus rechteckigen Pixeln bestehen, deren Struktur sich mittels eines einzigen Parameters beschreiben lässt. Das könnte die angestrebte Phase der Transmissionsfunktion sein, aber auch eine geometrische Größe wie die Ätztiefe für ein Höhenprofil, der Füllfaktor für eine Effektiv-Index-Struktur oder der Rotationswinkel bei Nutzung der geometrischen Phase, vgl. Abb. 2.4. Beim Design eines solchen Elementes entstehen oft gleiche oder zumindest ähnliche Anordnungen der Strukturen an verschiedenen Stellen des gesamten Elementes. Ziel der folgenden Betrachtungen ist es, das Feld hinter solchen Anordnungen einmalig physikalisch zu berechnen und es dann auf geeignete Weise mit den Parametern der umgebenden Pixel zu verknüpfen, sodass eine Zuordnung zu allen gleichen oder ähnlichen Konstellationen innerhalb des Elementes möglich wird. Voraussetzung dafür ist eine Lokalität der optischen Funktion, nämlich dass das Feld nur von den Strukturen in einer begrenzten Umgebung abhängig ist. Der vorgestellte Ansatz, die PFA, wurde in Ref. [105] veröffentlicht und wird hier in modifizierter Weise wiedergegeben.

4.1 Approximation des Fernfeldes

Für die Simulation von Subwellenlängenstrukturen ist typischerweise eine feine Diskretisierung, deutlich unterhalb der Wellenlänge, notwendig. Damit geht eine entsprechend feine Darstellung des Feldes einher. Variationen auf Skalen unterhalb der Wellenlänge entsprechen aber Raumfrequenzen, die sich bei einer weiteren Propagation im Freiraum evaneszent verhalten, sind also für die Berechnung des Fernfeldes bedeutungslos. In diesem Abschnitt wird betrachtet, wie die große Informationsmenge der hochaufgelösten Darstellung des Nahfeldes reduziert werden kann, ohne dass für das Fernfeld relevante Informationen verloren gehen. Zur Vereinfachung der Notation wird hier der 1D-Fall analysiert. Für die entsprechenden 2D-Formeln wird jeweils auf den Anhang verwiesen.

4.1.1 Polynomielle Beschreibung des Nahfeldes

Es wird ein 1D-Element betrachtet, das invariant in y -Richtung ist und durch eine Brechungsindexverteilung $n(x, z)$ im Bereich $0 \leq z \leq d$ beschrieben werden kann, während der Raum für $z < 0$ und $z > d$ homogen ist. Unter Einfall einer ebenen Welle aus dem Bereich $z < 0$ ergibt sich ein bestimmtes Feld an der Position $z = d$. Jede Komponente dieses elektrischen bzw. magnetischen Feldes lässt sich als skalares Feld $u(x)$ beschreiben, welches für jeden Pixel m der Größe p durch Legendre-Polynome [106] bis zum Grad N_L approximiert werden kann:

$$u(x) \approx \sum_{j=0}^{N_L} l_{mj} L_j \left(\frac{2(x - mp)}{p} \right), \quad \text{für } \left(m - \frac{1}{2}\right)p \leq x \leq \left(m + \frac{1}{2}\right)p, \quad (4.1)$$

vergleiche Abschnitt A.1.3.1 im Anhang. Dabei werden die jeweiligen Koeffizienten durch

$$l_{mj} = \frac{2j + 1}{2} \int_{-1}^1 u \left(\left(m + \frac{x'}{2}\right)p \right) L_j(x') dx' \quad (4.2)$$

bestimmt. Da die Legendre-Polynome auf dem Intervall $[-1,1]$ ein Orthogonalsystem bilden, hängen die Koeffizienten l_{mj} nicht vom maximalen Grad ab. So kann durch das Hinzufügen weiterer Polynome eine bessere Approximation erlangt werden, ohne die bisherigen Koeffizienten neu berechnen zu müssen. Indem die Approximation für alle Pixel des Elementes separat durchgeführt wird, lässt sich das Feld stückweise als Polynom beschreiben, sodass das gesamte Feld mithilfe weniger Koeffizienten je Pixel dargestellt werden kann.

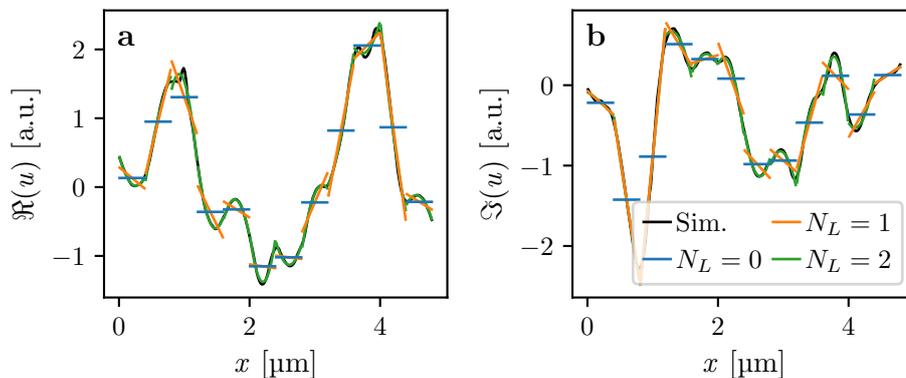


Abb. 4.1: Realteil (a) und Imaginärteil (b) des simulierten Feldes hinter einem 1D-Element und zugehörige Polynomapproximationen für $N_L = 0, 1, 2$.

Wie Abb. 4.1(a) zeigt, wird der Verlauf des Feldes mit steigendem Polynomgrad zunehmend besser approximiert. Hier lässt sich bereits für $N_L = 2$ kaum noch ein Unterschied zum nicht approximierten Feld feststellen. Im Allgemeinen hängt die Anzahl der notwendigen Polynome vom Verhältnis zwischen Pixelgröße und Wellenlänge p/λ_0 ab, siehe Abschnitt 4.1.2. In Abb. 4.2 wird ein direkter Vergleich des berechneten und approximierten Feldes eines 2D-Elementes für $N_L = 2$ dargestellt. Es ist gut zu erkennen, dass das Feld allgemein gut approximiert wird und

nur in den Bereichen, wo aufgrund von Materialgrenzen starke Amplitudenvariationen auftreten, größere Abweichungen vorkommen. Diese sind jedoch meist kleinräumig und verhalten sich bei der weiteren Propagation evaneszent.

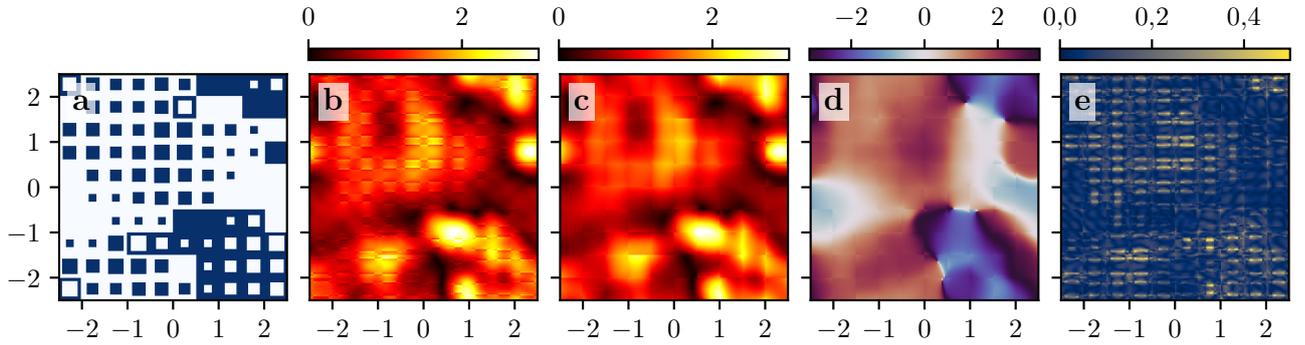


Abb. 4.2: Struktur eines Effektiv-Index-CGHs (a), Amplitude des Feldes hinter dem Element gemäß BPM-Simulation (b), Amplitude und Phase der zugehörigen Polynomapproximation für $N_L = 2$ (c,d), Betrag der Differenz zwischen Simulation und Approximation (e).

Für die im nächsten Abschnitt geschilderte Projektion ins Fernfeld wird die Approximation mithilfe von Standardpolynomen,

$$u(x) \approx \sum_{j=0}^{N_L} c_{mj} \left(\frac{x - mp}{p} \right)^j, \quad \text{für } \left(m - \frac{1}{2} \right) p \leq x \leq \left(m + \frac{1}{2} \right) p, \quad (4.3)$$

ausgedrückt, um eine einfachere Darstellung zu erreichen. Die Koeffizienten l_{mj} und c_{mj} können auf eindeutige Weise ineinander umgerechnet werden, siehe Abschnitt A.1.3.2 im Anhang.

4.1.2 Superposition des Fernfeldes

In Abschnitt 2.3.2 wurde die Ausbildung des Fernfeldes nach einem CGH erläutert. Dabei wurde von einem konstanten Feld hinter jedem Pixel ausgegangen. In Verallgemeinerung von Gl. (2.63) wird das Feld hinter jedem Pixel hier als Polynomapproximation

$$u(x) = \sum_{j=0}^{N_L} \left[\sum_{m=0}^{M-1} c_{mj} \delta_0(x - mp) \right] * \left(\frac{x}{p} \right)^j \text{rect} \left(\frac{x}{p} \right) * \sum_{K=-\infty}^{\infty} \delta_0(x - KMp), \quad (4.4)$$

beschrieben, wobei das gesamte Element aus M Pixeln wiederum periodisch fortgesetzt wird. Zur Bestimmung des Fernfeldes wird die FT dieses Ausdrucks verwendet:

$$U(k_x) = \frac{2\pi}{M} \sum_{j=0}^{N_L} \left(\frac{i}{2} \right)^j \text{sinc}^{(j)} \left(\frac{k_x p}{2} \right) \left[\sum_{m=0}^{M-1} C_{mj} \delta_0(k_x - k_{x,m}) * \sum_{K=-\infty}^{\infty} \delta_0 \left(k_x - \frac{2\pi K}{p} \right) \right]. \quad (4.5)$$

Dafür nutzt man aus, dass Monome wie x^j durch die FT in j -te Ableitungen der sinc-Funktion, $\text{sinc}^{(j)}$, übergehen. Die Koeffizienten C_{mj} ergeben sich durch die diskrete FT der c_{mj} , für $j =$

$0, \dots, N_L$. Eine detaillierte Herleitung und die Darstellung für den 2D-Fall werden im Anhang, Abschnitt A.2.1.3, wiedergegeben.

Die resultierende Abweichung des mittels Legendre-Polynomen approximierten Nahfeldes eines 2D-Elementes ist in Abb. 4.3(a) in Abhängigkeit von N_L für verschiedene Werte des Verhältnisses p/λ_0 dargestellt, hier jedoch für 2D-Elemente, basierend auf dem in Abschnitt 4.2.3 geschilderten Datensatz. Als Maß wird dabei die Standardabweichung (sd) der Differenz $u_D = u - u_S$ aus approximiertem und simuliertem Feld u_S genutzt:

$$sd(u_D) = \sqrt{\frac{1}{N} \sum_{x,y} |u_D(x,y) - \overline{u_D}|^2}, \quad (4.6)$$

wobei N die Gesamtzahl der Feldpunkte und $\overline{u_D}$ das arithmetische Mittel über das gesamte Element darstellt. Es ist klar zu erkennen, dass mit einem kleinen Polynomgrad Felder hinter kleineren Pixeln besser approximiert werden, da das Feld hier innerhalb eines Pixels weniger variiert. In allen Fällen nimmt die Abweichung mit zunehmendem Polynomgrad deutlich ab. Vor allem bei dem kleinen Verhältnis von $1/4$ wird diese Abnahme zunehmend schwächer, da hier Diskontinuitäten an Materialgrenzen das Feld dominieren.

Für die Analyse von CGHs ist insbesondere das Fernfeld von Interesse. Propagiert man die berechneten Koeffizienten der Elemente des Datensatzes ins Fernfeld, kann ebenfalls die Standardabweichung zum Ergebnis der Simulationsdaten bestimmt werden, siehe Abb. 4.3(b). Es zeigt sich, verglichen mit dem Nahfeld, eine deutlich schnellere Abnahme der Abweichung mit zunehmendem Polynomgrad, was den evaneszenten Charakter der Abweichungen im Nahfeld bestätigt. Für geringere Pixelgrößen ist die Abweichung wiederum tendenziell geringer.

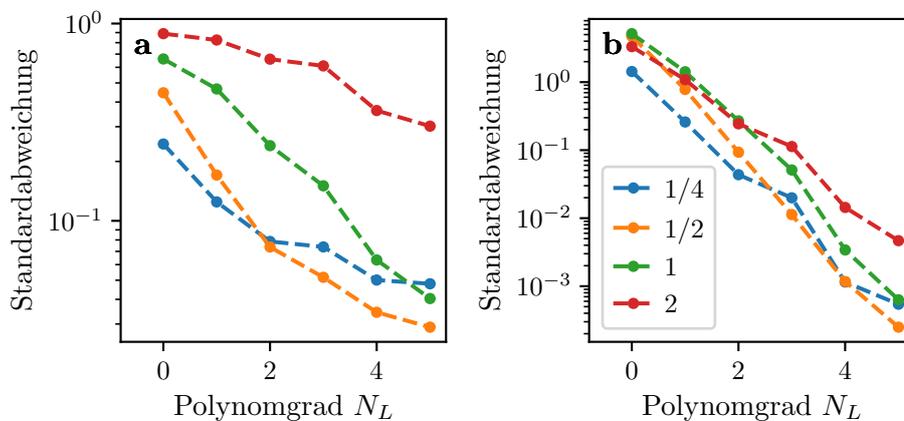


Abb. 4.3: Konvergenz der Approximation bei verschiedenen Verhältnissen von p/λ_0 (siehe Legende) für ein 2D-Element im Nahfeld (a) bzw. Fernfeld (b).

4.2 Bestimmung der Polynomkoeffizienten

Mit der im vorherigen Abschnitt geschilderten Vorgehensweise kann das Fernfeld eines Elements auf Basis von Polynomkoeffizienten für das Feld nach den einzelnen Pixeln dargestellt werden. Diese Koeffizienten lassen sich auf effiziente Weise direkt aus den entsprechenden Strukturparametern der Pixel berechnen, wie in diesem Abschnitt dargelegt wird.

4.2.1 Idee der Approximation

Beispielhaft soll hier das Feld hinter einem auf Effektiv-Index-Strukturen basierenden CGH bei homogener Beleuchtung betrachtet werden. Dieses besteht aus einer regelmäßigen Anordnung von Subwellenlängenstrukturen. Die Struktur in jedem Pixel lasse sich durch einen einzelnen Parameter beschreiben, beispielsweise den Füllfaktor $f \in [0, 1]$. Dabei bietet es sich an, den in Gl. (3.23) eingeführten Flächenfüllfaktor zu verwenden. Dann kann das Feld hinter dem gesamten Element als Funktion dieser Füllfaktoren dargestellt werden. Es ist naheliegend, dass das Feld bei einem gewissen Pixel nicht von der Struktur des gesamten Elementes abhängig ist, sondern nur von einer gewissen Umgebung. In Abb. 4.4 ist dargestellt, dass ähnliche lokale Strukturen zu ähnlichen Feldern führen. Anstatt die genaue Feldverteilung hinter einem Pixel zu betrachten, wird die Bestimmung der entsprechenden Legendre-Koeffizienten der Polynomapproximation angestrebt. Da die Legendre-Polynome orthogonal sind, können die Koeffizienten für jede Ordnung separat bestimmt werden. Für ein 1D-CGH entspricht das der Suche nach einer Funktion F_j , sodass sich der Koeffizient l_{mj} eines Pixels m aus den Füllfaktoren $(f_{m-D}, f_{m-D+1}, \dots, f_{m+D})$ in dessen Umgebung approximieren lässt:

$$l_{mj} \approx F_j(f_{m-D}, \dots, f_{m+D}) . \quad (4.7)$$

Darin enthalten ist die Annahme, dass die Füllfaktoren der Pixel, die weiter als D Positionen entfernt sind, keinen Einfluss auf den Koeffizienten haben. Als Ansatz zur Bestimmung von F_j wird ein lineares Modell gewählt, d. h. es wird versucht, F_j als Polynom in Abhängigkeit der Füllfaktoren darzustellen. Neben den einfachen Füllfaktoren erlauben dabei höhere Ordnungen, also Terme wie f_m^3 , die detaillierte Nachbildung der Zusammenhänge, während Produkte, wie $f_m f_{m+1}$ die Modellierung der Wechselwirkung zwischen benachbarten Pixeln ermöglichen. Die Koeffizienten der einzelnen Terme werden mittels linearer Regression aus einem Trainingsdatensatz bestimmt. Dadurch erhält man jene Koeffizienten, die zur geringsten quadratischen Abweichung zwischen simuliertem und vorhergesagtem Feld für die Trainingsdaten führen. Ob mit dem Modell sinnvolle Vorhersagen möglich sind, hängt vor allem davon ab, welche Terme in der Regression berücksichtigt werden, siehe den folgenden Abschnitt. Die Güte des Modells kann mithilfe eines von den Trainingsdaten unabhängigen Testdatensatzes überprüft werden.

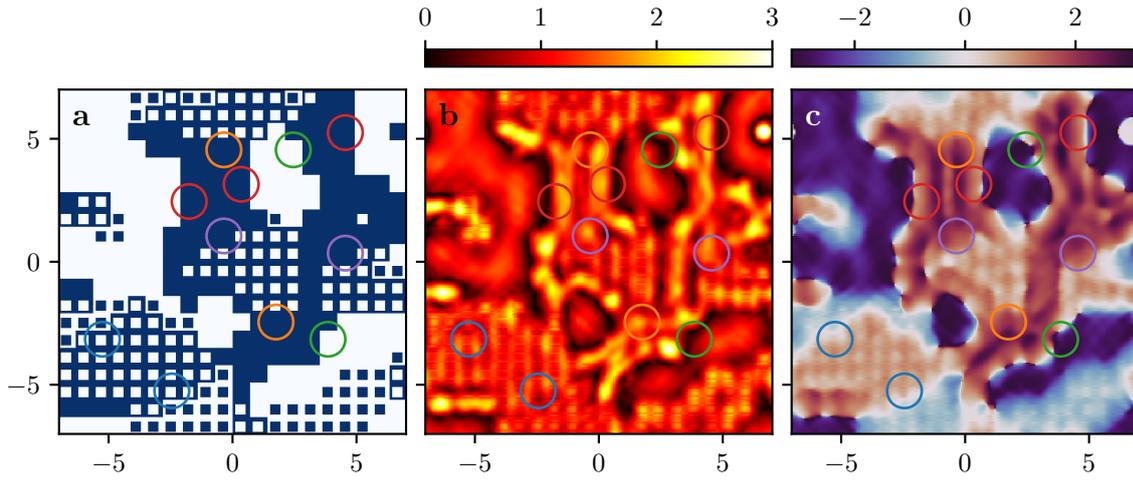


Abb. 4.4: Struktur (a), Amplitude (b) und Phase (c) eines Effektiv-Index-Elements. Bereiche mit ähnlichen Strukturen in den umgebenden Pixeln sind jeweils mit gleichfarbigen Kreisen gekennzeichnet. Die Achsen zeigen die Größe in μm .

4.2.2 Auswahl von Regressionsparametern

Die zur Beschreibung des Nahfeldes genutzten Legendre-Polynome besitzen eine definierte Parität, d. h. sie sind entweder gerade oder ungerade:

$$L_j(-x) = (-1)^j L_j(x) . \quad (4.8)$$

Diese Eigenschaft nutzt man aus, um die für die Regression nützlichen Terme auszuwählen. Angenommen, man invertiert die Verteilung der Füllfaktoren am Zentrum des Pixels m , so werden sich die jeweiligen Koeffizienten gemäß ihrer Parität verhalten:

$$F_j(f_{m+D}, \dots, f_{m-D}) = (-1)^j F_j(f_{m-D}, \dots, f_{m+D}) . \quad (4.9)$$

In dieser Form ist die Gl. für die jeweilige transversale Komponente in TE- und TM-Polarisation gültig. Betrachtet man die anderen Komponenten, ändert sich der Faktor der rechten Seite gegebenenfalls zu $(-1)^{j+1}$, siehe Abschnitt 4.3.4. Zur Erfüllung der Parität in der angegebenen Form müssen die Parameter f_{m-d} und f_{m+d} bei geradem j stets auf gleiche Weise berücksichtigt werden, bei ungeradem j dagegen stets mit umgekehrtem Vorzeichen. Das sollte bei der Erstellung des Modells inhärent berücksichtigt werden, indem man die (anti-)symmetrisierten Parameter

$$f_{m,d}^s = f_{m-d} + f_{m+d} - 1 , \quad f_{m,d}^a = f_{m+d} - f_{m-d} , \quad \text{für } d = 0, \dots, D , \quad (4.10)$$

verwendet. Dabei kennzeichnet s die symmetrischen, a die antisymmetrischen Terme. Die Ausdrücke sind dabei so konstruiert, dass $f_{m,d}^s, f_{m,d}^a \in [-1, 1]$. Dadurch verringert sich die Korrelation zwischen den einzelnen Parametern, was die Berechnung der Regression numerisch stabiler macht. Produkte und Potenzen dieser Parameter können in Multi-Index-Schreibweise ausgedrückt werden:

$$\left(\mathbf{f}_{m,D}^s\right)^{e^s} = \prod_{d=0}^D \left(f_{m,d}^s\right)^{e_d^s} , \quad \left(\mathbf{f}_{m,D}^a\right)^{e^a} = \prod_{d=1}^D \left(f_{m,d}^a\right)^{e_d^a} , \quad (4.11)$$

wobei $\mathbf{f}_{m,D}^s = (f_{m,0}^s, \dots, f_{m,D}^s)$, $\mathbf{f}_{m,D}^a = (f_{m,1}^a, \dots, f_{m,D}^a)$, $\mathbf{e}^s = (e_0^s, \dots, e_D^s)$ und $\mathbf{e}^a = (e_1^a, \dots, e_D^a)$ für $e_d^{a/s} \in \mathbb{N}$. Es ist stets $f_{m,0}^a = 0$, weshalb dieser Term ausgelassen wird. Ein Produkt aus symmetrischen Termen ist wiederum symmetrisch, während die Parität eines Produktes aus asymmetrischen Termen von den Exponenten \mathbf{e}^a abhängt. Zur Unterscheidung wird eine Paritätsfunktion p festgelegt:

$$p(\mathbf{e}^a) := \prod_{d=1}^D (-1)^{e_d^a} = (-1)^{\sum_{d=1}^D e_d^a}. \quad (4.12)$$

Trotz Beachtung der Parität und mit der Beschränkung auf einen begrenzten Abstand gibt es noch unendlich viele Terme, die im Modell berücksichtigt werden könnten. Um deren Zahl zu begrenzen, wird ein Termgewicht eingeführt, welches einem Term in Abhängigkeit seiner Exponenten eine Zahl zuordnet. Für das lineare Modell werden dann alle Terme mit einem Termgewicht unterhalb einer festzulegenden Schranke W verwendet. Für die Definition des Termgewichtes wird angenommen, dass der Einfluss der Füllfaktoren in der Umgebung mit steigendem Abstand abnimmt. Daher erhalten der zentrale Pixel und direkte Nachbarn geringere Gewichte und können infolgedessen mit höheren Potenzen berücksichtigt werden als weiter entfernte. Das Termgewicht wird in Abhängigkeit der Exponenten als

$$w(\mathbf{e}^s) := \sum_{d=0}^D e_d^s (1+d), \quad w(\mathbf{e}^a) := \sum_{d=1}^D e_d^a (1+d) \quad (4.13)$$

definiert. Die Wahl dieser Definition ist heuristisch und ohne Anspruch, optimal zu sein. Im Rahmen der Untersuchungen zur Entwicklung der PFA, erwies sie sich allerdings im Allgemeinen als sinnvolle Wahl. Damit kann nun das lineare Modell formuliert werden: Darin werden alle Terme aus Kombinationen der benachbarten Füllfaktoren bis zu einem Abstand D berücksichtigt, deren Termgewicht höchstens W beträgt:

$$F_j(\mathbf{f}_{m,D}^s, \mathbf{f}_{m,D}^a) = \sum_{\substack{\mathbf{e}^s, \mathbf{e}^a: p(\mathbf{e}^a) = \pm 1 \\ w(\mathbf{e}^s) + w(\mathbf{e}^a) \leq W}} \beta_{\mathbf{e}^s, \mathbf{e}^a} (\mathbf{f}_{m,D}^s)^{\mathbf{e}^s} (\mathbf{f}_{m,D}^a)^{\mathbf{e}^a}. \quad (4.14)$$

Dabei sind $\mathbf{e}^s \in \mathbb{N}^{D+1}$, $\mathbf{e}^a \in \mathbb{N}^D$. Das Vorzeichen für die Parität muss dabei zur Parität von F_j gewählt werden. Die Summe ist endlich, da aufgrund der Definition der Gewichtungsfunktion stets $0 \leq e_d^{s/a} \leq W$ gilt. Die Koeffizienten $\beta_{\mathbf{e}^s, \mathbf{e}^a}$ lassen sich mithilfe linearer Regression bestimmen. Die Verallgemeinerung auf zwei Dimensionen ist in Abschnitt A.2.3.1 im Anhang angegeben, neben einigen Beispielen für die auf diese Weise konstruierten Terme in Abschnitt A.2.3.2.

Wenn das zu analysierende Element eine rein binäre Struktur hat, d. h. $f_m \in \{0, 1\}$ und $f_{m,d}^{a/s} = \pm 1$, reduziert sich die Zahl der in der Summe von Gl. (4.14) zu berücksichtigenden Terme weiter, da bestimmte Terme nicht länger unabhängig sind, z. B. ist stets $(f_{m,d}^{a/s})^2 = 1$.

4.2.3 Berechnung der Modell-Parameter

Um ein lineares Modell in der geschilderten Weise erstellen zu können, müssen Trainingsdaten generiert werden. Dazu werden zufällige Verteilungen erzeugt, die aus den gleichen Strukturen zusammengesetzt sind, wie die später zu analysierenden Elemente. Beispielsweise werden für die Analyse eines Effektiv-Index-CGHs bestehend aus N_E verschiedenen Basiszellen mit den zugehörigen Füllfaktoren zufällige Verteilungen dieser N_E verschiedenen Füllfaktoren erzeugt, während Wellenlänge, Strukturtiefe und Größe der Zellen fest sind. Dabei können mehrere einzelne Elemente anstelle eines großen genutzt werden, solange jedes groß genug ist, um die volle Reichweite der Wechselwirkungen zu berücksichtigen, d. h. in den transversalen Dimensionen mind. $2D + 1$ Pixel enthält. Damit das Modell später sinnvolle Vorhersagen treffen kann, müssen die Trainingsdaten ähnliche Verteilungen wie das zu analysierende Element enthalten. Dazu können z. B. Teile der Verteilungen in den Trainingsdaten geglättet werden, damit Bereiche mit nur gleichen oder ähnlichen Füllfaktoren in der Nachbarschaft enthalten sind, was bei einer rein zufälligen Auswahl unwahrscheinlich ist. Mit welchem Verfahren die Strukturen der Trainingsdaten simuliert werden, ist unerheblich. Für die vorgestellten Anwendungen wurde meist die BPM aufgrund der geringen Rechenzeit genutzt, während in den Abschnitten 4.3.3 und 4.3.4 die FDTD zur Analyse hochbrechender Strukturen bzw. zum Vergleich der Genauigkeit verwendet wurde. Das simulierte Feld hinter jedem Pixel wird gemäß Gl. (4.1) durch Legendre-Polynome approximiert.

Als Ergebnis der Simulation seien die Legendre-Koeffizienten als jeweils ein Vektor $\mathbf{l}_j = (l_{1j}, \dots, l_{Mj})^\top$ zu den Füllfaktoren $\mathbf{f} = (f_1, \dots, f_M)^\top$ gegeben, wobei M die Gesamtzahl aller simulierten Pixel ist und separate Elemente gegebenenfalls nacheinander aufgeführt werden. Dazu konstruiert man eine $M \times N_K$ -Matrix $\hat{\mathbf{f}}$ aus den Füllfaktoren, die in jeder Zeile alle N_K für einen Pixel m konstruierten Terme gemäß Gl. (4.14) enthält, z. B. für gerade Parität

$$\hat{\mathbf{f}} = \begin{pmatrix} 1 & f_1 & \cdots & (\mathbf{f}_{1,D}^s)^{e^s} & (\mathbf{f}_{1,D}^a)^{e^a} & \cdots \\ \vdots & \vdots & & \vdots & & \\ 1 & f_M & \cdots & (\mathbf{f}_{M,D}^s)^{e^s} & (\mathbf{f}_{M,D}^a)^{e^a} & \cdots \end{pmatrix}. \quad (4.15)$$

Dann lässt sich das lineare Modell als Matrixgleichung

$$\mathbf{l}_j = \hat{\mathbf{f}}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4.16)$$

schreiben, mit den Modellkoeffizienten $\boldsymbol{\beta} = (\dots, \beta_{e^s, e^a}, \dots)^\top$ und einem Residuum $\boldsymbol{\epsilon}$. Bestimmt man die Modellkoeffizienten gemäß einer linearen Regression als

$$\boldsymbol{\beta} = \left(\hat{\mathbf{f}}^\top \hat{\mathbf{f}} \right)^{-1} \hat{\mathbf{f}}^\top \mathbf{l}_j, \quad (4.17)$$

so wird das Residuum, genauer die Summe seiner Quadrate $\sum_{m=1}^M |\epsilon_m|^2$, minimiert, d. h. man erhält die bestmögliche Approximation von \mathbf{l}_j durch $\hat{\mathbf{f}}\boldsymbol{\beta}$ im Sinne der quadratischen Abweichung [107].

Dabei ist $\hat{\mathbf{f}}^T \hat{\mathbf{f}}$ nur invertierbar, wenn die gemäß Gl. (4.14) gebildeten Terme linear unabhängig sind. Im 2D-Fall ist die Vorgehensweise analog, indem die Füllfaktoren und Legendre-Koeffizienten der untersuchten Elemente jeweils als eindimensionale Vektoren geschrieben werden und bei der Konstruktion der Matrix $\hat{\mathbf{f}}$ auf eine korrekte Sortierung der entsprechenden Terme geachtet wird.

Das auf diese Weise berechnete Modell ist zwar optimal für den spezifischen Datensatz und die ausgewählten Terme, aber damit ist nicht garantiert, dass es sich auf unbekannte Elemente sinnvoll anwenden lässt. Insbesondere besteht die Gefahr einer Überanpassung, d. h. das Modell kann tatsächlich die in den Trainingsdaten enthaltenen Elemente hervorragend approximieren, liefert aber im Allgemeinen keine guten Ergebnisse, weil die ermittelten Koeffizienten eher spezielle Strukturen der Trainingsdaten als physikalische Abhängigkeiten widerspiegeln. Das geschieht typischerweise, wenn sich die Anzahl der Modellkoeffizienten der Anzahl der in den Trainingsdaten enthaltenen Pixel nähert. Dementsprechend sind für die Erzeugung eines Modells mit vielen Koeffizienten mehr Trainingsdaten erforderlich. Zur Abschätzung der Qualität des Modells wird zusätzlich zum Trainingsdatensatz ein Testdatensatz erzeugt, der eine ähnliche Verteilung der Strukturen aufweist. Als Maß für die Bewertung des Modells wird die Standardabweichung des Residuums ϵ genutzt. Berechnet man diese in Abhängigkeit des maximalen Termgewichtes W und der maximalen Distanz D , so nimmt $\text{sd}(\epsilon)$ für wachsendes W bzw. D immer weiter ab, da immer mehr Koeffizienten für das Modell angepasst werden können. Dagegen beobachtet man für das Residuum der Testdaten $\epsilon' = \mathbf{l}'_j - \hat{\mathbf{f}}' \boldsymbol{\beta}$ zunächst ebenfalls eine Abnahme, bis ein Minimum erreicht ist. Darüber hinaus führen weitere Terme nicht zu einer Verbesserung des Modells, sondern approximieren nur den Trainingsdatensatz besser. Um ein möglichst genaues Modell zu erhalten, sind die Parameter D und W nahe an diesem Optimum zu wählen. Legt man sie dagegen auf geringere Werte fest, so verringert sich die Rechenzeit, gegebenenfalls ohne die Genauigkeit deutlich zu verschlechtern.

Zur Illustration der geschilderten Abhängigkeit wurden ein Trainingsdatensatz mit 64 Elementen aus 20×20 Pixeln, sowie ein Testdatensatz mit 64 Elementen aus 10×10 Pixeln erstellt. Die Elemente enthalten jeweils zehn Arten von Effektiv-Index-Strukturen in Quarzglas, vgl. Abschnitt 3.2.1. Diese wurden zufällig auf die quadratischen Pixel verteilt. Neben voll und nicht abgetragenen Pixeln sind dies quadratische Säulen bzw. Löcher mit Seitenlängen von $1/3$, $1/2$, $7/12$ oder $2/3$ der Pixelgröße p bei einer Höhe von 975 nm. Für verschiedene Pixelgrößen wurden separate Datensätze angelegt, während die Wellenlänge von $\lambda_0 = 500$ nm festgehalten wurde. Zur Simulation der Elemente wurde die BPM unter Nutzung der in Tab. A.9 im Anhang angegebenen Parameter verwendet. In Abb. 4.5(a) ist zu erkennen, dass der Verlauf der Standardabweichung für Trainings- und Testdaten zunächst nahezu identisch ist, bis für die Testdaten etwa bei $W = 10$ ein Optimum erreicht ist und sie im weiteren Verlauf wieder ansteigt. Das erreichte Minimum sinkt für größere Distanzen D , sodass man davon ausgehen kann, dass hier auch Pixel in einem Abstand von 4 noch

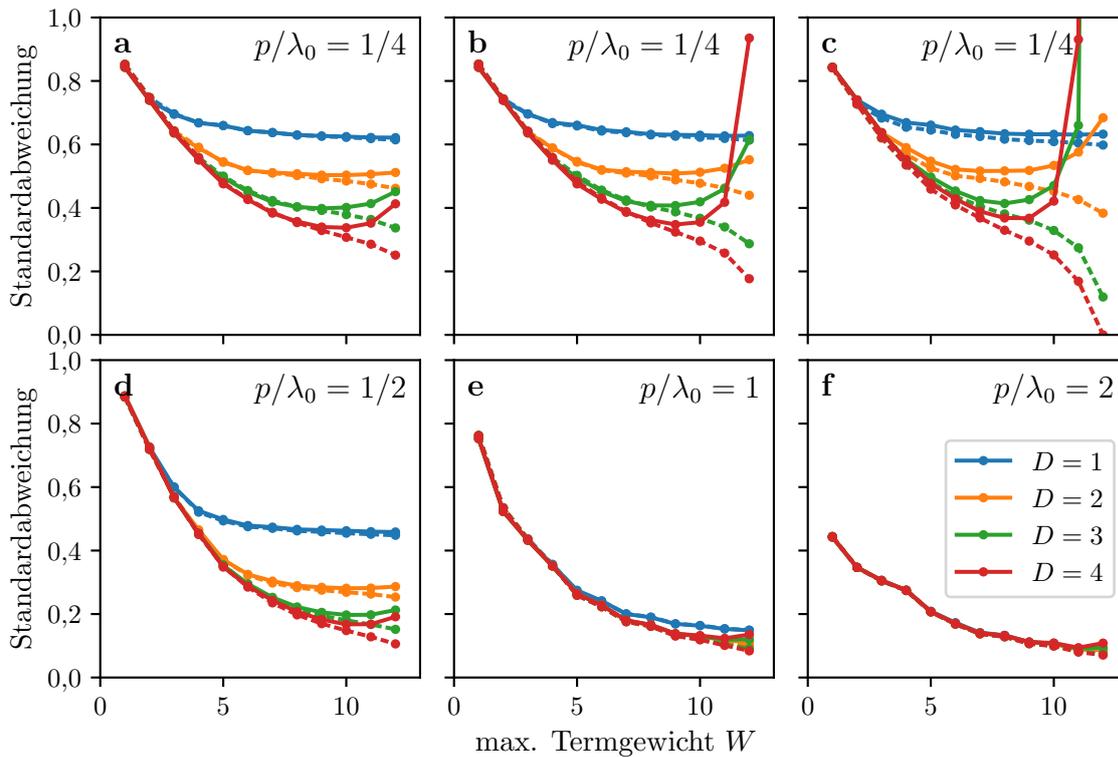


Abb. 4.5: Standardabweichung des Residuums der Testdaten $sd(\epsilon')$ (durchgezogen) und Trainingsdaten $sd(\epsilon)$ (gestrichelt) für den Koeffizienten l_{00} in Abhängigkeit des Termgewichtes W und maximalen Abstandes D für verschiedene Verhältnisse p/λ_0 (a,d-f), Verhalten der Standardabweichung bei Reduktion des Datensatzes auf die Hälfte bzw. ein Viertel der Einträge (b,c).

einen merklichen Einfluss auf das Nahfeld haben. Reduziert man die Menge der Trainingsdaten auf die Hälfte bzw. ein Viertel (b,c), so ist das Minimum größer und wird für ein kleineres W und damit eine geringere Zahl an Koeffizienten erreicht. Für größere Pixelgrößen (d-f) fällt auf, dass die Unterschiede bezüglich der Distanz abnehmen, d. h. innerhalb der Reichweite der Wechselwirkungen liegen weniger Pixel. Zudem wird das erreichte Minimum an sich kleiner. Demzufolge lassen sich die Koeffizienten mit weniger Parametern und genauer abschätzen, allerdings muss in diesem Fall der maximale Polynomgrad N_L eventuell erhöht werden, vgl. Abb. 4.3.

Aus den approximierten Koeffizienten der Testdaten wurden das Nah- und Fernfeld der jeweiligen Elemente berechnet und mit den jeweiligen Simulationsergebnissen verglichen, siehe Abb. 4.6. Dabei wurden die Parameter der PFA für die verschiedenen Pixelgrößen stets im Bereich des Optimums gewählt und zwar als $D = 4, 4, 2, 1$ bzw. $W = 9, 10, 13, 20$ für aufsteigende Pixelgröße. Mit $N_L = 0$ zeigt sich wie in Abb. 4.3 eine geringere Abweichung bei kleineren Pixelgrößen. Mit steigendem N_L verbessert sich zwar die Approximation, allerdings nur geringfügig und umso weniger, je kleiner die Pixel sind. Die Ursache hierfür sind die Abweichungen des Modells bei Koeffizienten geringer Ordnung, v.a. l_{00} . Diese können durch andere Ordnungen nicht ausgeglichen werden. Bei größeren Pixeln sind die Abweichungen der einzelnen Ordnungen tendenziell geringer

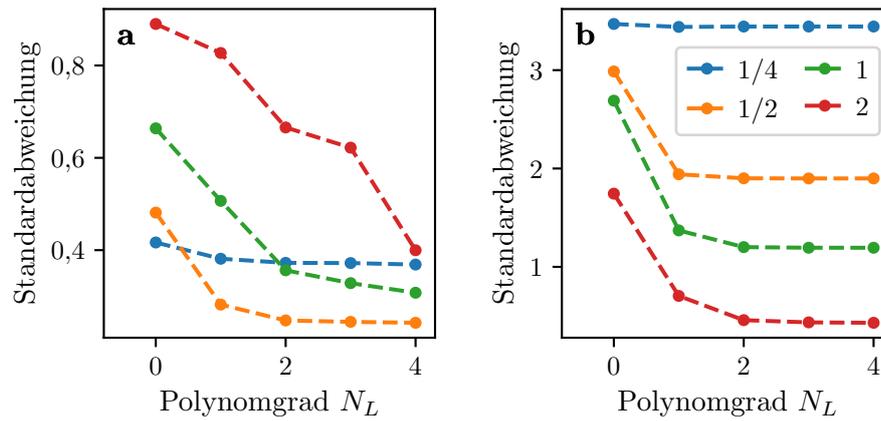


Abb. 4.6: Abweichung des mittels der PFA bestimmten Nahfeldes (a) bzw. Fernfeldes (b) in Abhängigkeit des verwendeten Polynomgrades N_L für verschiedene Verhältnisse p/λ_0 .

und höhere Ordnungen spielen eine stärkere Rolle, sodass sich hier am ehesten eine Reduktion der Abweichung für ein höheres N_L ergibt. Im Fernfeld tritt dieses Zusammenspiel noch deutlicher zutage: Es wird maßgeblich durch die 0. Ordnung beeinflusst, sodass sich jenseits von $N_L = 2$ kaum noch eine Verbesserung erzielen lässt. Das Fernfeld von Elementen mit größeren Pixeln kann tendenziell besser approximiert werden.

4.3 Anwendungsbeispiele

Die PFA wurde in Python implementiert. Die folgenden Beispiele demonstrieren ihre Anwendung zur Analyse von CGHs basierend auf Effektiv-Index-Strukturen. Zur Berechnung der 2D-Phasenprofile wurde ein gewöhnlicher IFTA verwendet. Diesem wurde gemäß der LPA eine Struktur zugeordnet, wobei der Füllfaktor jedes Pixels so gewählt wurde, dass er eine gute Approximation der gewünschten Phase erzielt. Zur Simulation der Trainingsdaten wurde die in Kap. 3 beschriebene BPM benutzt und für den Vergleich zu rigorosen Ergebnissen die kommerzielle Software *Lumerical FDTD*.

Mit dem ersten Beispiel, einem außeraxialen Strahlteiler, soll demonstriert werden, wie die Wechselwirkung benachbarter Pixel mit zunehmendem Ablenkwinkel stärker wird. Dieser Effekt lässt sich mit der PFA beschreiben, im Gegensatz zur LPA. Das zweite Beispiel, ein CGH zur Projektion eines Gittermusters, soll vor allem die Anwendbarkeit der PFA auf sehr große Simulationsgebiete deutlich machen. Im dritten Beispiel werden Trainingsdaten mittels der FDTD erzeugt, sodass die Simulation eines Elementes mit hochbrechenden Strukturen ermöglicht wird. Abschließend wird ein Strahlteiler mit großen Ablenkwinkeln mithilfe verschiedener Methoden analysiert, um Aussagen über deren Genauigkeit zu gewinnen.

4.3.1 Außeraxialer Strahlteiler

Für das Design eines Strahlteiler-CGHs wurde zunächst durch einen Standard-IFTA mit Amplitudenfreiheit, siehe Iterationsschritt 2a aus Abschnitt 2.3.3, ein Phasenprofil mit 100×100 Pixeln für einen 11×11 -Strahlteiler errechnet, welcher das einfallende Licht gleichmäßig in die Ordnungen $m_{x/y} = -5, \dots, 5$ beugen soll. Dabei wurde Amplitudenfreiheit für alle Beugungsordnungen außerhalb eines zentralen Bereiches von 17×17 Ordnungen genutzt. Anschließend wurde ein lineares Profil zu der ermittelten Phase addiert. Dieses führt zu einer Verschiebung der Fourier-Transformierten, d. h. im Fernfeld lässt sich eine zusätzliche Ablenkung aller Ordnungen beobachten. Im Rahmen der LPA ergibt sich damit für beliebige Ablenkwinkel ein geeignetes Phasenprofil. Tatsächlich führt die zusätzliche Ablenkung aber zu stärkeren Unterschieden in der Phase benachbarter Pixel, wodurch der Einfluss von Kopplungseffekten zunimmt und die optische Funktion des Elementes gestört wird. Durch Variation des Ablenkwinkels können die Auswirkungen der Kopplung gezielt verändert und untersucht werden.

Das CGH wurde für eine Wellenlänge von 633 nm ausgelegt. Es besteht aus 100×100 Pixeln mit einer Größe von $400 \times 400 \text{ nm}^2$, hat also eine Gesamtgröße von $40 \times 40 \mu\text{m}^2$. Verschiedene Phasenstufen werden mithilfe von effektiven Medien erzeugt, wie in Abschnitt 3.2.2 beschrieben. Die Quarzglasstrukturen haben eine Höhe von $1,385 \mu\text{m}$, entsprechend einem maximalen Phasenunterschied von 2π . Zwischenwerte werden durch quadratische Säulen bzw. Löcher mit Seitenlängen von 80, 120, 160, 180, 200, 220, 240, 260 bzw. 280 nm erreicht. Auf diese Weise können 20 näherungsweise äquidistante Phasenstufen erzeugt werden, die mithilfe der BPM berechnet wurden.

Für die Berechnung der PFA-Trainingsdaten wurden 64 zufällige CGHs mit 20×20 Pixeln erzeugt, wobei die Verteilungen unterschiedlich stark geglättet wurden. Deren Nahfeld wurde mittels der BPM unter Verwendung der in Tab. A.9 im Anhang angegebenen Parameter berechnet. Für die Auswahl der Terme wurde eine maximale Distanz von $D = 4$ und ein maximales Termgewicht von $W = 11$ genutzt. Dann enthält das lineare Modell für 2D, s. Gl. (A.110) im Anhang, für jeden Legendre-Koeffizienten 4386, 3878 bzw. 3490 Terme, je nach dessen Parität in x - bzw. y -Richtung.

Mit zunehmendem Ablenkwinkel können verschiedene Effekte beobachtet werden, siehe Abb. 4.7: Das Fernfeld eines idealen IFTA-Phasenprofils wird durch eine lineare Phase ausschließlich verschoben. In Abhängigkeit des Ablenkwinkels ändert sich zwar die Zuordnung der Strukturen zu den Pixeln, aber im Wesentlichen bleibt die gewünschte Intensitätsverteilung erhalten. Somit beobachtet man im Rahmen der LPA ebenfalls hauptsächlich eine Verschiebung im Fernfeld (a-c). Unter Beachtung der sinc-Modulation aus Gl. (2.64) wird das Fernfeld von einer zunehmenden Abschwächung der höheren Ordnungen überlagert. Dieser Effekt kann im Rahmen des IFTA durch eine Vorkorrektur der Zielverteilung berücksichtigt werden und stellt keinen spezifischen Nachteil

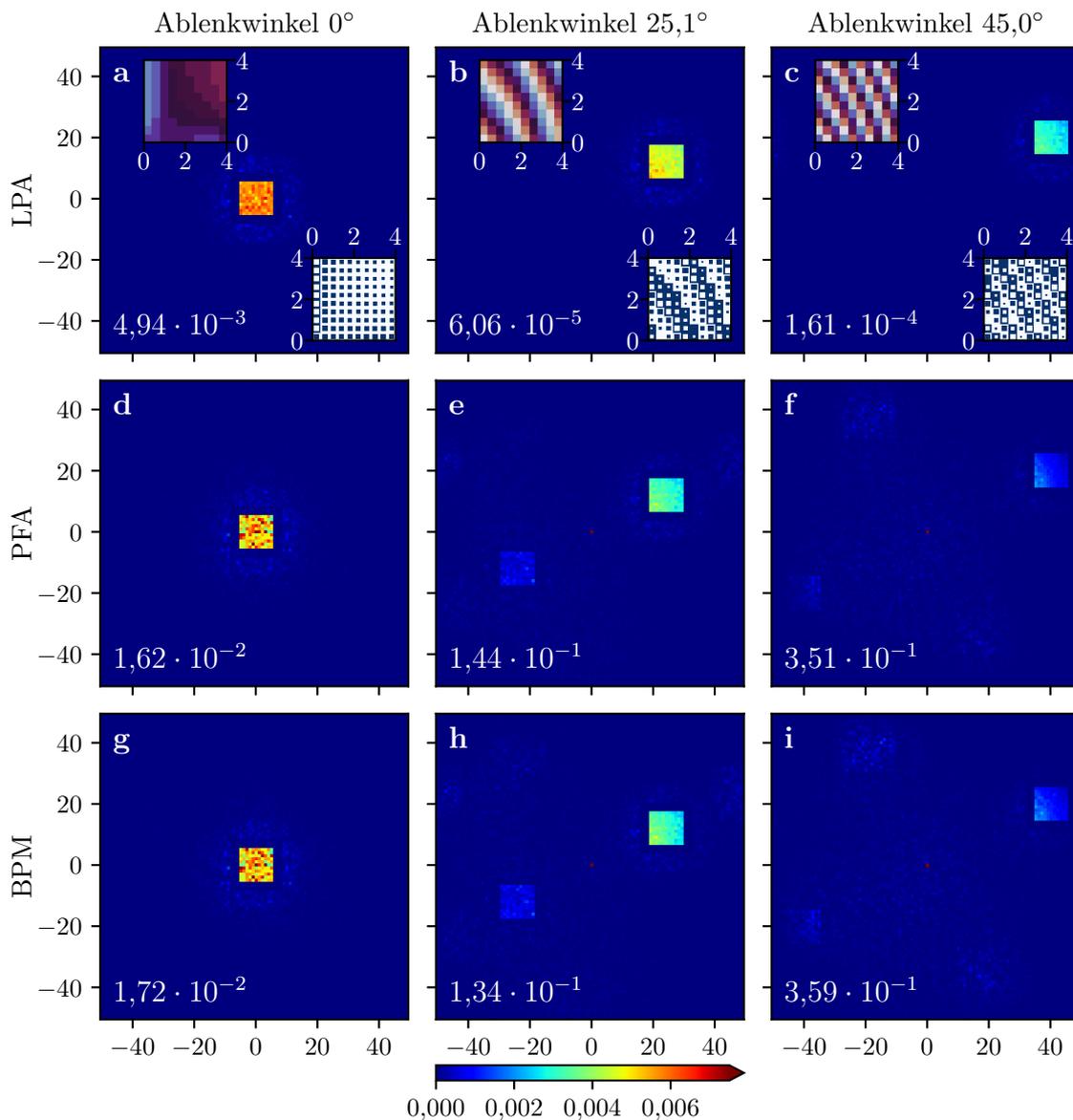


Abb. 4.7: Intensitätsverteilung im Fernfeld für ein Strahlteiler-CGH mit verschiedenen Ablenkwinkeln für die zentrale Beugungsordnung (von links nach rechts), gemäß der Berechnung mittels LPA (a-c), PFA (d-f) bzw. BPM (g-i). Die Achsenbeschriftungen kennzeichnen die Beugungsordnungen. Die eingefügten Diagramme zeigen jeweils einen Ausschnitt von 10×10 Pixel aus dem Phasenprofil (oben links) und der Struktur (unten rechts) mit Längenangaben in μm . Die Zahlenwerte geben die Intensität in der 0. Ordnung an.

der LPA dar. Unter Verwendung der PFA können allerdings weitere Effekte bei zunehmendem Ablenkwinkel beobachtet werden: die 0. Ordnung wird verstärkt und die Homogenität der Nutzordnungen reduziert, während die Intensität in den dazu punktsymmetrischen Ordnungen steigt (d-f). Alle diese Effekte weisen eine hohe Übereinstimmung zu einer entsprechenden BPM-Simulation auf (g-i).

Als Maß für einen quantitativen Vergleich der simulierten Fernfelder wird die Korrelation gemäß Gl. (3.25) genutzt, allerdings unter Berücksichtigung aller Ordnungen. Die Ergebnisse sind in Tab.

4.1 zusammengefasst. Sie zeigen, dass die Genauigkeit der LPA mit zunehmendem Ablenkwinkel deutlich abnimmt, während es für die PFA lediglich zu einer leichten Verschlechterung kommt. Vergleicht man die Abhängigkeit der PFA vom maximalen Grad der Polynome N_L , so sieht man eine deutliche Verbesserung für $N_L = 1$ gegenüber $N_L = 0$. Die Erweiterung auf $N_L = 2$ erhöht die Genauigkeit dagegen nur geringfügig. In Abb. 4.7 sind die Ergebnisse für $N_L = 2$ dargestellt.

Tab. 4.1: Betrag der Korrelationskoeffizienten für die Simulationsergebnissen von LPA und BPM bzw. PFA und BPM. Die Resultate der PFA werden nach dem maximalen Grad der Polynomapproximation N_L unterschieden. Der angegebene Ablenkwinkel bezieht sich jeweils auf die zentrale Beugungsordnung.

Ablenkung	LPA	PFA $N_L = 0$	PFA $N_L = 1$	PFA $N_L = 2$
0,0°	0,923	0,989	0,995	0,995
8,1°	0,879	0,981	0,994	0,995
16,4°	0,800	0,969	0,991	0,993
25,1°	0,700	0,951	0,986	0,988
34,5°	0,590	0,946	0,986	0,988
45,0°	0,414	0,948	0,986	0,988

Bezüglich der Winkelabhängigkeit zeigen die Ergebnisse das erwartete Verhalten: für kleine Ablenkwinkel kann das Fernfeld durch die LPA mit einer gewissen Genauigkeit ermittelt werden, siehe Abb. 4.7(a,d,g). Daher ist sie in diesem Fall zur Berechnung von CGHs geeignet. Mit zunehmendem Ablenkwinkel sind jedoch lokal stärkere Änderungen der Phase notwendig und die Gültigkeit der LPA nimmt ab, sodass genauere Verfahren zur Bestimmung des Fernfeldes nötig werden. Die PFA kann dabei die Ergebnisse der BPM nahezu reproduzieren.

4.3.2 Gitterprojektion

Die hier untersuchten CGHs projizieren jeweils ein quadratisches Gitter ins Fernfeld. Sie unterscheiden sich in der Anzahl der Pixel bzw. der Gesamtgröße einer Einheitszelle. Bei einer größeren Anzahl der Beugungsordnungen können die Gitterlinien mehr einzelne Punkte enthalten, sodass sie gleichmäßiger erscheinen. Ein derartiges Element kann mithilfe eines klassischen IFTA unter Nutzung der LPA berechnet werden, allerdings unter Vernachlässigung der Kopplungseffekte benachbarter Pixel. Zu deren Analyse sind (semi-)rigorose Verfahren notwendig, die entsprechend enorme Rechenkapazitäten benötigen. Dieses Problem kann durch die schnellen und präzisen Fernfeldprojektionen der LPA umgangen werden.

Für das Fernfeld des CGHs wurde eine symmetrische Intensitätsverteilung gemäß Abb. 4.8(k,l) angestrebt, wobei dort nur der erste Quadrant abgebildet ist. Das Design ist für eine Wellenlänge

von 532 nm ausgelegt. Die sinc-Modulation wurde in der Zielverteilung berücksichtigt, siehe Gl. (5.4). Zur Berechnung der Phasenfunktion wurde wie im vorherigen Beispiel ein klassischer IFTA genutzt, wobei Amplitudenfreiheit außerhalb des Gitterbereiches verwendet wurde. Der Ablenkwinkel der höchsten genutzten Beugungsordnungen beträgt $39,7^\circ$ in horizontaler und vertikaler sowie $64,5^\circ$ in diagonaler Richtung. Die betrachteten Elemente haben eine Pixelgröße von $300 \times 300 \text{ nm}^2$ und nutzen, wie im vorigen Abschnitt, eine Kombination von quadratischen Säulen und Löchern auf einem Quarzglassubstrat zur Realisierung verschiedener Phasenstufen. Die Seitenlängen der Strukturen sind 100, 140, 180 bzw. 200 nm. Ihre Höhe beträgt jeweils $1,04 \mu\text{m}$. Im Folgenden werden Ergebnisse für Elemente mit Größen von 25×25 Pixeln ($7,5 \times 7,5 \mu\text{m}^2$) bis zu 1600×1600 Pixeln ($480 \times 480 \mu\text{m}^2$) präsentiert, wobei diese Einheitszellen stets periodisch fortgesetzt werden.

Zur Berechnung der PFA wurde die gleiche Menge an Trainingsdaten wie im vorherigen Abschnitt genutzt. Ebenso wurde die maximale Distanz $D = 4$ beibehalten, aber das zulässige Termgewicht auf $W = 9$ verringert. Dadurch reduziert sich Zahl der unabhängigen Terme des linearen Modells auf 1093, 894 bzw. 751 in Abhängigkeit von der Parität der Koeffizienten. Mit dieser Festlegung erlangt man eine höhere Geschwindigkeit der Berechnungen, während die Genauigkeit leicht abnimmt. Die Simulation der Trainingsdaten mittels BPM, siehe die Parameter in Tab. A.9 im Anhang, benötigte 360 s und die Erstellung des Modells 8 s. Die angegebenen Laufzeiten beziehen sich jeweils auf dasselbe System mit einer Taktung von 3,3 GHz und 16 Prozessoren, welche die parallele Berechnung von jeweils 16 Elementen ermöglichten.

Das CGH mit der kleinsten Einheitszelle erzeugt zwar gemäß der LPA die gewünschte Fernfeldverteilung, siehe Abb. 4.8(a). Bei der Simulation mit präziseren Verfahren zeigen sich jedoch deutliche Abweichungen vom angestrebten Verhalten (d,g,j), hier lässt sich das Gitter kaum erkennen. Die Simulationsparameter sind in Tab. A.9 im Anhang angegeben. Andererseits stimmen die Ergebnisse von PFA, BPM und FDTD gut überein, auch bezüglich der Intensität in den unerwünschten Nebenordnungen. Mit einer größeren Einheitszelle wird das Gitter im Fernfeld deutlich besser erzeugt (b), aber in der Simulation mittels PFA bzw. BPM (e,h) ist die Intensität der Gitterlinien deutlich inhomogener als gemäß der LPA. An den Ecken des Gitters lässt sich zudem eine Abschwächung beobachten. Diese wird nicht durch die sinc-Modulation bedingt, welche im Design korrigiert wurde, sondern ist auf den stärkeren Einfluss benachbarter Strukturen bei größeren Ablenkwinkeln zurückzuführen. Hier zeigt sich ebenso eine hohe Übereinstimmung zwischen PFA und BPM. Ein Vergleich mit der FDTD würde zu viel Rechenzeit benötigen. Bei einer weiteren Erhöhung der Pixelzahl (c,f,i) zeigen die Beugungsordnungen des Gitters für alle Verfahren meist eine höhere Intensität als die Nebenordnungen. Zu beachten sind aber deutliche Unterschiede zwischen einzelnen Nutzordnungen und Abschwächungen an den Ecken.

Die Ergebnisse der PFA weisen eine hohe Korrelation von 0,987 zu denen der BPM auf, nahezu

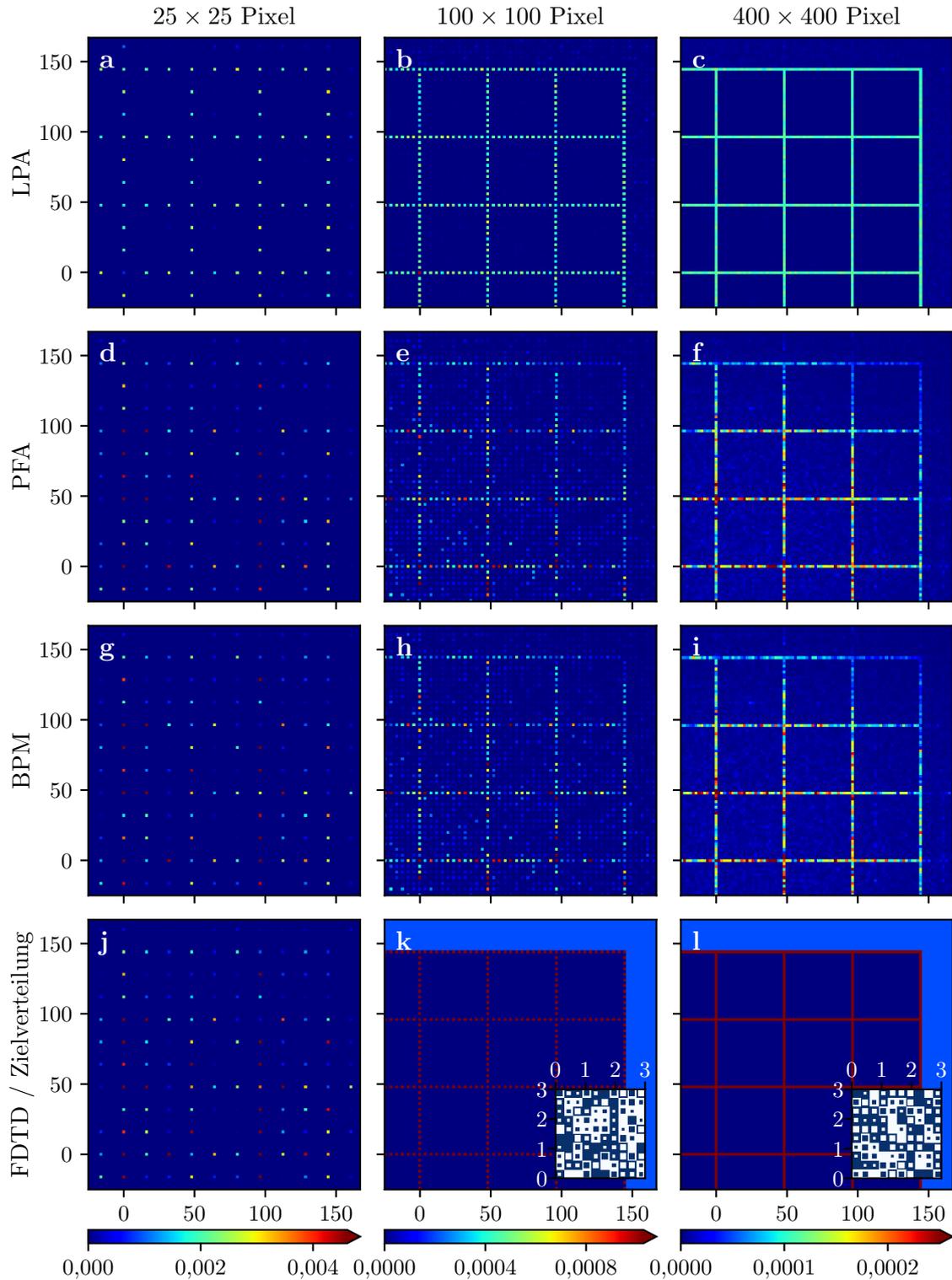


Abb. 4.8: Fernfeldintensitätsverteilung von CGHs zur Gitterprojektion mit Einheitszellen verschiedener Größe (von links nach rechts), gemäß einer Berechnung mittels LPA (a-c), PFA (d-f), BPM (g-i) und FDTD (j); Zielverteilung (k,l), mit Amplitudenfreiheitsbereich in hellblau. Die Achsenbeschriftungen kennzeichnen die Beugungsordnungen, bezogen auf ein Element mit einer Auflösung von 400×400 Pixeln. Die eingefügten Diagramme zeigen einen Ausschnitt der Struktur von 10×10 Pixeln, mit Längenangaben in μm .

unabhängig von der Zahl der Pixel des CGH-Designs. Mit Werten von 0,50 ... 0,55 ist diese für die LPA deutlich geringer. Für das kleinste Beispiel beträgt die Korrelation von PFA und BPM zur FDTD 0,983 bzw. 0,991. Hiermit zeigt sich wiederum die hohe Genauigkeit der PFA, während die LPA starke Abweichungen mit sich bringt.

Tab. 4.2: Vergleich der Laufzeit verschiedener Algorithmen für unterschiedliche Größen der CGH-Einheitszellen. Für die FDTD wurden acht parallele Prozesse verwendet, ansonsten jeweils ein einzelner. Zudem wird eine Abschätzung des genutzten Arbeitsspeichers angegeben. Die Werte in Klammern beziehen sich auf unterteilte Simulationsgebiete.

Größe CGH [Pixel]	Größe CGH [μm]	Laufzeit PFA [s]	Arbeitsspeicher PFA [GB]	Laufzeit BPM [s]	Arbeitsspeicher BPM [GB]	Laufzeit FDTD [s]
25×25	$7,5 \times 7,5$	0,2	<0,2	9,8	0,7	4019
50×50	15×15	0,6	<0,2	40,3	2,5	
100×100	30×30	1,9	0,4	149	9,7	
200×200	60×60	6,5	1,4	660	38,3	
400×400	120×120	27,3	(1,8)	2870	(41,5)	
800×800	240×240	109	(1,9)			
1600×1600	480×480	440	(2,7)			

Die Laufzeit der verschiedenen Berechnungsmethoden wird in Tab. 4.2 für verschiedene Größen der Einheitszelle verglichen. Dabei wurde ausschließlich für die FDTD eine Parallelisierung genutzt. Bei der PFA und BPM wurde das Simulationsgebiet großer Elemente zerlegt, um den Speicherbedarf zu reduzieren. Falls ausreichend Arbeitsspeicher vorhanden ist, könnte das ebenso zur parallelen Berechnung der Teilbereiche genutzt werden, sodass sich die Rechendauer reduziert. Die Laufzeit der PFA hängt nahezu linear von der CGH-Größe ab und ist deutlich geringer als die der BPM, welche wiederum wesentlich schneller als die FDTD ist.

Die vorgestellten Resultate zeigen, dass ein LPA-basiertes Design, insbesondere bei einer kleinen Elementgröße, für die vorgegebene Intensitätsverteilung nicht zielführend ist. Andererseits wird die Anwendung existierender Berechnungsmethoden für großflächige Elemente durch den numerischen Aufwand beschränkt. Dagegen können mit der vorgestellten PFA sogar großflächige Elemente in einer akzeptablen Zeit mit hoher Genauigkeit analysiert werden. Für kleinere Beispiele konnte eine hohe Übereinstimmung mit den Ergebnissen anderer Verfahren nachgewiesen werden.

4.3.3 Simulation von Siliziumstrukturen für Strahlteiler

Mithilfe der in den vorherigen Beispielen betrachteten Effektiv-Index-Strukturen lassen sich prinzipiell beliebige Phasenprofile erzeugen. Die Herstellung ist allerdings herausfordernd, da

typischerweise ein hohes Aspektverhältnis, das Verhältnis zwischen Höhe und Breite der Strukturen, nötig ist. Da zur Erzielung eines effektiven Index die Ausdehnung der Basiszellen auf Subwellenlängengröße festgelegt ist und andererseits die Strukturhöhe durch Gl. (3.22), lässt sich dieses Verhältnis durch die Auswahl der Strukturen für ein gegebenes Material nicht grundlegend verringern. Ein naheliegender Ansatz ist die Nutzung von hochbrechenden Materialien. Als Beispiel werden hier Säulen aus Silizium auf einem Quarzglassubstrat betrachtet [57, 108–110]. Für eine Wellenlänge von 1064 nm betragen die Brechungsindizes 3,553 für die Säulen und 1,450 für das Substrat. Für einen Phasenhub von 2π ist eine Strukturhöhe von nur 417 nm notwendig. Damit sich das elektromagnetische Feld innerhalb des DOEs streng gemäß einem effektiven Index ausbreitet, muss Gl. (2.27) für den Brechungsindex der Säulen erfüllt sein. Daraus ergeben sich Zellgrößen deutlich unterhalb der Wellenlänge, was dem Vorteil der reduzierten Höhe entgegensteht. Tatsächlich beobachtet man z. B. für eine Zellgröße von $600 \times 600 \text{ nm}^2$ bei quadratischen Säulen bis zu einer gewissen Breite einen gleichmäßigen Zusammenhang zwischen Füllfaktor und Phase der Transmissionsfunktion. Darüber hinaus kommt es aber zu Resonanzeffekten. Innerhalb der Säulen breiten sich die Felder wie in einem Wellenleiter aus: Solange die Säulen schmal genug sind, existiert nur eine transversale Eigenmode und der effektive Index hängt monoton von der Breite der Säulen bzw. dem Füllfaktor ab. Treten dagegen weitere transversale Moden auf, führt die Kopplung zwischen den Moden zu einer starken Abhängigkeit der Transmission von der geometrischen Struktur. Dieser Effekt sollte vermieden werden, da er einerseits die in der Herstellung akzeptablen Formabweichungen deutlich reduziert und andererseits die Erstellung eines linearen Modells für die PFA erschwert. Als Ausweg bietet sich eine Vergrößerung der Strukturhöhe an, damit trotz limitierter Säulenbreite die nötigen Phasenwerte realisiert werden können.

Die Anwendung der PFA auf hochbrechende Strukturen wird hier am Beispiel eines CGHs zur Strahlteilung demonstriert. Es nutzt die im vorigen Abschnitt geschilderten Parameter, mit einer auf 600 nm vergrößerten Strukturhöhe. Zur Realisierung verschiedener Phasenstufen werden quadratische Siliziumsäulen mit Breiten von 192,6; 204,0 bzw. 219,6 nm verwendet. Zusammen mit Zellen ohne Säule ergeben sich Phasenschritte von etwa $\pi/2$, siehe Abb. 4.9(a-d). Die Gesamtgröße des CGHs beträgt 100×100 Pixel ($60 \times 60 \mu\text{m}^2$). Als Zielfunktion wird eine gleichmäßige Verteilung auf alle Ordnungen mit einem Ablenkwinkel von höchstens 30° vorgegeben. Die Bestimmung der Phasenfunktion erfolgte mit dem originalen IFTA gemäß Abschnitt 5.2.3. Ein Ausschnitt des berechneten Elementes wird in Abb. 4.9(e) gezeigt.

Für die Simulation der Strukturen mittels FDTD wurde die im Anhang, Abschnitt A.2.4, angegebene Diskretisierung verwendet, sowohl für die Trainingsdaten als auch für das gesamte Element. In Abb. 4.9(f,g) ist das simulierte Nahfeld beispielhaft dargestellt. Die vier verschiedenen Phasenwerte sind deutlich erkennbar. An den Übergängen zwischen verschiedenen Strukturen

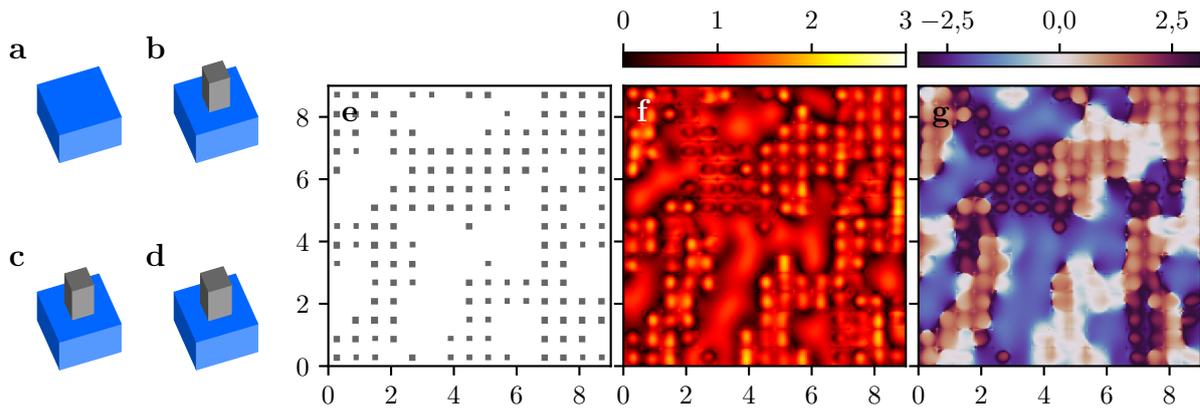


Abb. 4.9: Basiszellen mit Siliziumsäulen (a-d), Ausschnitt der Struktur (e), sowie von Amplitude (f) und Phase (g) des Nahfeldes eines darauf basierenden Strahlteiler-CGHs. Die Achsenbeschriftung verdeutlicht die Größe in μm .

treten wie in vorherigen Beispielen deutliche Variationen in der Amplitude auf. Aufgrund des hohen Brechungsindex kommt es im Vergleich zu Quarzglasstrukturen, wie z. B. in Abb. 4.4, zu einer stärkeren Lokalisation der Amplitude in den Säulen bzw. im Zentrum der Pixel. Eine Simulation mittels BPM ist aufgrund des hohen Indexkontrastes nicht möglich. Für die Auswertung wurde ausschließlich die y -Komponente des elektrischen Feldes verwendet, was aufgrund der moderaten Ablenkungswinkel eine plausible Näherung darstellt. Der Umfang des PFA-Trainingsdatensatzes entspricht dem der vorherigen Beispiele. Das Modell wurde mit einer maximalen Distanz von $D = 4$ und einem Termgewicht von $W = 10$ erstellt.

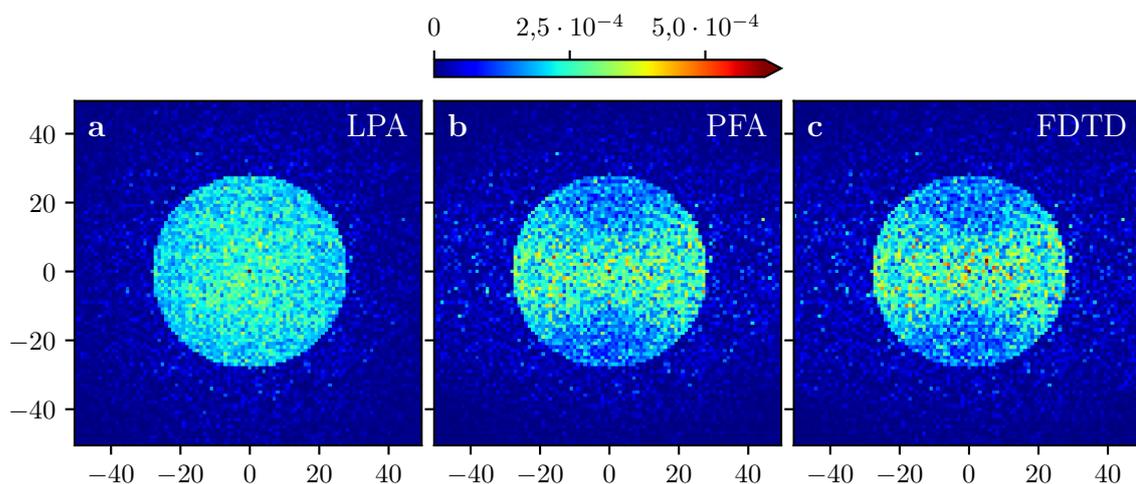


Abb. 4.10: Intensitätsverteilung im Fernfeld für das Strahlteiler-CGH basierend auf Siliziumsäulen. Die Berechnung erfolgte mittels LPA (a), PFA (b) bzw. FDTD (c). Die Achsen kennzeichnen die Beugungsordnungen.

Zum Vergleich der Simulationsverfahren sind die berechneten Fernfelder in Abb. 4.10 dargestellt. Die Ergebnisse der LPA (a) zeigen erwartungsgemäß eine weitestgehend gleichmäßige Verteilung

der Intensität innerhalb des Kreises, der dem vorgegebenen Ablenkwinkel von 30° entspricht. In der PFA-Simulation (b) ist eine deutliche Abschwächung der Nutzordnungen in vertikaler Richtung erkennbar. Die Brechung der Symmetrie wird durch eine Polarisationsabhängigkeit verursacht, welche in der LPA nicht berücksichtigt werden kann. Das Resultat der FDTD (c) zeigt eine hohe Übereinstimmung mit dem der PFA. Der Betrag der Korrelation gemäß Gl. (3.25) beträgt zwischen LPA und FDTD 0,957 und zwischen PFA und FDTD 0,991. Dieses Beispiel verdeutlicht die Vielfältigkeit der Anwendungsmöglichkeiten der PFA, für die lediglich geeignete Trainingsdaten benötigt werden. Insbesondere können hochbrechende Materialien simuliert werden, welche eine Reduzierung des Aspektverhältnisses der Strukturen erlauben und so vorteilhaft für die Herstellung sein können.

4.3.4 Analyse eines Weitwinkel-Strahlteilers

Im Gegensatz zur BPM erhält man bei der Simulation mittels der FDTD Ergebnisse für alle Feldkomponenten des elektrischen und magnetischen Feldes. Daher kann auch ein separates PFA-Modell für jede der Komponenten erstellt werden. Dabei ist die korrekte Beachtung der Symmetriebedingungen zu gewährleisten. Für eine einfallende ebene Welle mit in y -Richtung polarisiertem elektrischem Feld verhalten sich die Komponenten E_y , E_z und H_x des transmittierten Feldes symmetrisch bzgl. der Inversion der Strukturen in x -Richtung, d. h. Gl. (4.9) gilt in der angegebenen Weise. Die Komponenten E_x , H_y und H_z verhalten sich antisymmetrisch, d. h. der enthaltene Faktor ist als $(-1)^{j+1}$ zu wählen. Für eine Inversion in y -Richtung sind entsprechend E_y , H_x und H_z symmetrisch bzw. E_x , E_z und H_y antisymmetrisch.

Zur Berechnung der Beugungseffizienz gemäß Gl. (2.35) ist die Kenntnis der Amplitude $|E_0|$ der ebenen Welle in Beugungsrichtung notwendig. Es sei

$$\mathbf{k} = nk_0 (\cos \varphi \sin \vartheta, \sin \varphi \sin \vartheta, \cos \vartheta)^\top \quad (4.18)$$

der Wellenvektor einer Beugungsordnung mit Polarwinkel ϑ und Azimutwinkel φ , siehe Abb. 4.11(a). Dann gilt für das Quadrat der Amplitude:

$$|E_0|^2 = |E_x|^2 + |E_y|^2 + |E_z|^2, \quad (4.19)$$

bzw. unter Nutzung der Divergenzgleichung (2.7) zur Bestimmung von E_z :

$$|E_0|^2 = |E_x|^2 + |E_y|^2 + \tan^2 \vartheta |\cos \varphi E_x + \sin \varphi E_y|^2. \quad (4.20)$$

In der BPM wird die Kopplung zwischen Komponenten vernachlässigt und für einfallendes y -polarisiertes Licht gilt entsprechend:

$$|E_0|^2 \approx (1 + \sin^2 \varphi \tan^2 \vartheta) |E_y|^2. \quad (4.21)$$

Die Amplitude des magnetischen Feldes besitzt eine analoge Darstellung. Für eine hohe Genauigkeit bietet es sich an, zur Berechnung der Beugungseffizienz jeweils die Feldkomponente zu nutzen, welche den größeren Winkel zur Propagationsrichtung hat, z. B. hier für senkrechten Einfall aus Luft

$$\eta \approx \begin{cases} \frac{\cos \vartheta}{n} (1 + \sin^2 \varphi \tan^2 \vartheta) \frac{|E_y|^2}{|E_{\text{in}}|^2}, & \text{für } |\sin \varphi| \leq \frac{1}{2}, \\ n \cos \vartheta (1 + \cos^2 \varphi \tan^2 \vartheta) \frac{|H_x|^2}{|H_{\text{in}}|^2}, & \text{sonst.} \end{cases} \quad (4.22)$$

Anhand eines weiteren Strahlteiler-CGHs soll im Folgenden die Genauigkeit der verschiedenen Simulationsverfahren verglichen werden. Dieses ist als binäres Element realisiert, d. h. es gibt nur die Füllfaktoren null und eins. Bei der Auswahl der Terme für das lineare Modell muss daher beachtet werden, dass f und f^2 identisch sind, wodurch sich die Anzahl der unabhängigen Terme reduziert. Die Verteilung der Nutzordnungen bildet eine Spirale mit

$$\vartheta_m = \vartheta_{\text{max}} \sqrt{\frac{m - 1/2}{M}}, \quad \varphi_m = \frac{\pi m}{\Phi_0}, \quad m = 1, \dots, M, \quad (4.23)$$

nach, wobei $\Phi_0 = (1 + \sqrt{5})/2$, $\vartheta_{\text{max}} = 75^\circ$ und $M = 1414$. Auf diese Weise wird eine gleichmäßige Verteilung der Ordnungen bis zum maximalen Ablenkwinkel erreicht. Die Gesamtzahl der Nutzordnungen ist $2M$ aufgrund der Punktsymmetrie für binäre Elemente. Für die Berechnung des CGHs sei auf die Vorgehensweise in Abschnitt 5.2.3 für den originalen IFTA und eine gleichmäßige Verteilung auf einer Sphäre verwiesen. Der Fokus soll hier auf der Analyse der Beugungseffizienz für ein gegebenes CGH liegen. Das CGH besteht aus 256×256 quadratischen Pixeln mit einer Seitenlänge von 325 nm und wird als periodisch fortgesetzt betrachtet. Die Tiefe der in das Quarzglassubstrat geätzten Strukturen beträgt 692 nm , entsprechend einem Phasenunterschied von π . Die Beleuchtung erfolgt senkrecht mit einer ebenen Welle der Wellenlänge $632,8 \text{ nm}$.

Für die Modellerstellung wurden 64 zufällige CGHs mit 20×20 Pixeln mittels FDTD und BPM simuliert. Zur Auswertung wurden PFA-Modelle mit den Parametern $D = 4$, $W = 10$ aus den FDTD-Trainingsdaten und mit den Parametern $D = 4$, $W = 11$ aus den BPM-Trainingsdaten erzeugt, für die Simulationsparameter siehe Tab. A.9 im Anhang. Als Vergleichsmaß für die Ergebnisse wird die Korrelation gemäß Gl. (3.25) berechnet, wobei die Summen ausschließlich über die Nutzordnungen gebildet werden.

Auf Basis des elektrischen Feldes der FDTD-Simulation wurde die Beugungseffizienz des CGHs mithilfe von Gl. (4.19) und (4.20) berechnet (η_{E1}, η_{E2}) und ebenso auf Basis des magnetischen Feldes (η_{H1}, η_{H2}). Idealerweise liefern diese Vorgehensweisen identische Ergebnisse, tatsächlich zeigen sich aber Abweichungen, siehe Abb. 4.11. Auffällig ist dabei, dass η_{H2} und η_{E2} gut übereinstimmen (b), ähnliches gilt für η_{H1} . Die Korrelation ist jeweils größer als 0,9997. Andererseits weicht η_{E1} merklich von η_{E2} ab (c) und erreicht nur eine Korrelation von ca. 0,998 zu den übrigen. Die Abweichungen von E_z werden möglicherweise durch die Diskretisierung in der FDTD

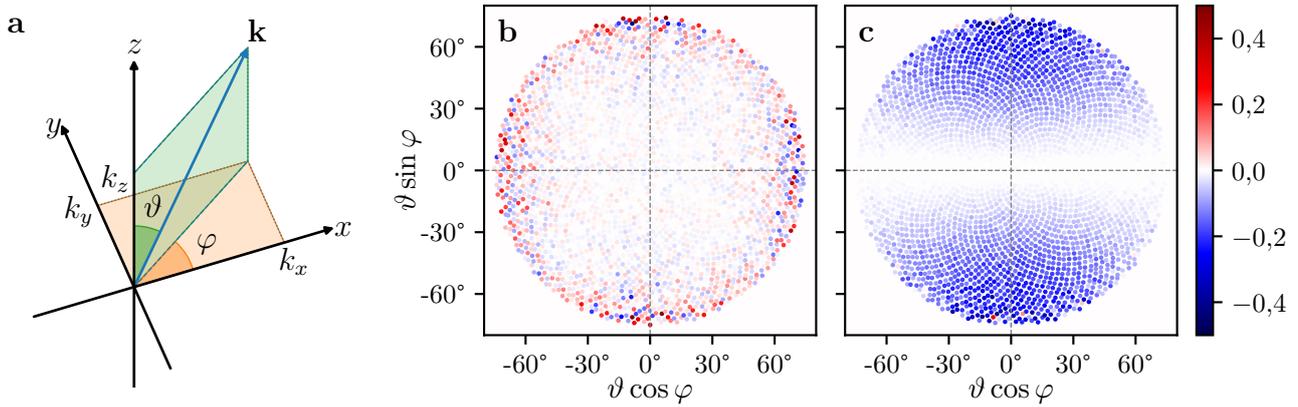


Abb. 4.11: Definition der sphärischen Koordinaten (a), relative Abweichung der berechneten Beugungseffizienz $(\eta_{H2} - \eta_{E2})/\eta_{E2}$ (b) bzw. $(\eta_{E1} - \eta_{E2})/\eta_{E2}$ (c). Die Nutzordnungen werden gemäß der Projektion auf eine Kugel dargestellt.

verursacht. Im Folgenden wird stets die gemäß Gl. (4.22) aus η_{E2} und η_{H2} zusammengesetzte Beugungseffizienz η_{ref} als Referenz verwendet .

Tab. 4.3: Betrag der Korrelationskoeffizienten für die mit verschiedenen Methoden berechnete Beugungseffizienz. Die Ergebnisse von E - und H -Feld wurden analog zu Gl. (4.22) zusammengesetzt. Als Referenz dient jeweils das Ergebnis der FDTD-Simulation gemäß Gl. (4.20).

Methode	FDTD-Poly	FDTD-PFA	FDTD-Appr	BPM	BPM-PFA	TEA
Gl.	(4.20)	(4.20)	(4.21)	(4.21)	(4.21)	(4.21)
Korrelation	0,999 98	0,9969	0,965	0,982	0,987	0,942

In Tab. 4.3 wird die Korrelation der Resultate verschiedener Methoden zur Referenz dargestellt. Zunächst wird nur die in Abschnitt 4.1 beschriebene polynomielle Näherung auf die FDTD-Ergebnisse angewendet (FDTD-Poly). Die Ergebnisse sind quasi identisch. Für die PFA basierend auf FDTD-Daten (FDTD-PFA) ergibt sich eine vergleichsweise hohe Korrelation. Die Qualität der Trainingsdaten kann also in einem gewissen Rahmen verlässlich abgebildet werden. Vernachlässigt man die jeweils kleinere Feldkomponente (FDTD-Appr), nutzt also Gl. (4.21) statt (4.20), so verschlechtert sich erwartungsgemäß die Korrelation. Diese Näherung ist in der BPM inhärent enthalten, weshalb keine deutlich besseren Resultate zu erwarten sind. Tatsächlich liegt die beobachtete Korrelation hier sogar etwas höher, weil sich entgegengesetzte Abweichungen aufheben können. Die auf BPM-Daten basierende PFA (BPM-PFA) erreicht eine ähnlich hohe Korrelation. Deren Abweichungen werden folglich hauptsächlich durch die Vernachlässigung der Kopplung innerhalb der BPM verursacht und nicht durch das Fitmodell an sich. Die TEA weist im betrachteten Fall eine deutlich größere Abweichung auf.

Ergänzend zum globalen Vergleich der berechneten Beugungseffizienz wird nun die Abhängigkeit

der Genauigkeit der einzelnen Verfahren vom Ablenkwinkel der jeweiligen Beugungsordnung betrachtet. Dazu wurde die relative Abweichung $(\eta - \eta_{\text{ref}})/\eta_{\text{ref}}$ der Nutzordnungen berechnet und in Abhängigkeit des Polarwinkels ϑ als gleitender Mittelwert aufgetragen, siehe Abb. 4.12. Aufgrund der spiralförmigen Zielverteilung entspricht das gleichzeitig einer Mittelung über den Azimutwinkel φ . Die polynomielle Näherung (FDTD-Poly) verursacht eine geringe Abweichung, die erst bei großen Ablenkwinkeln ansteigt. Bei der FDTD-PFA gibt es gewisse Abweichungen, die mit zunehmendem Winkel ansteigen, aber auch für große Winkel noch verlässliche Aussagen erlauben. Der Verlauf dieser Kurve hängt von den verwendeten Trainingsdaten und den Parametern der PFA ab. Unter Vernachlässigung der Kopplung (FDTD-Appr) nimmt die Abweichung in Abhängigkeit des Winkels deutlich zu. Für die BPM steigt die Abweichung zunächst nur langsam an, bis sie die Linie des vorigen Graphen erreicht und diesem dann folgt. Für große Winkel limitiert demnach die Vernachlässigung der Kopplung die Genauigkeit der BPM, bei kleineren Winkeln andere Effekte wie die Padé-Approximation. Die BPM-PFA zeigt einen ähnlichen Verlauf wie die BPM selbst. Die Abweichungen werden hier also im Wesentlichen durch die Trainingsdaten, weniger durch das lineare Modell der PFA verursacht. Die TEA weist über den gesamten Verlauf deutlich größere Abweichungen auf.

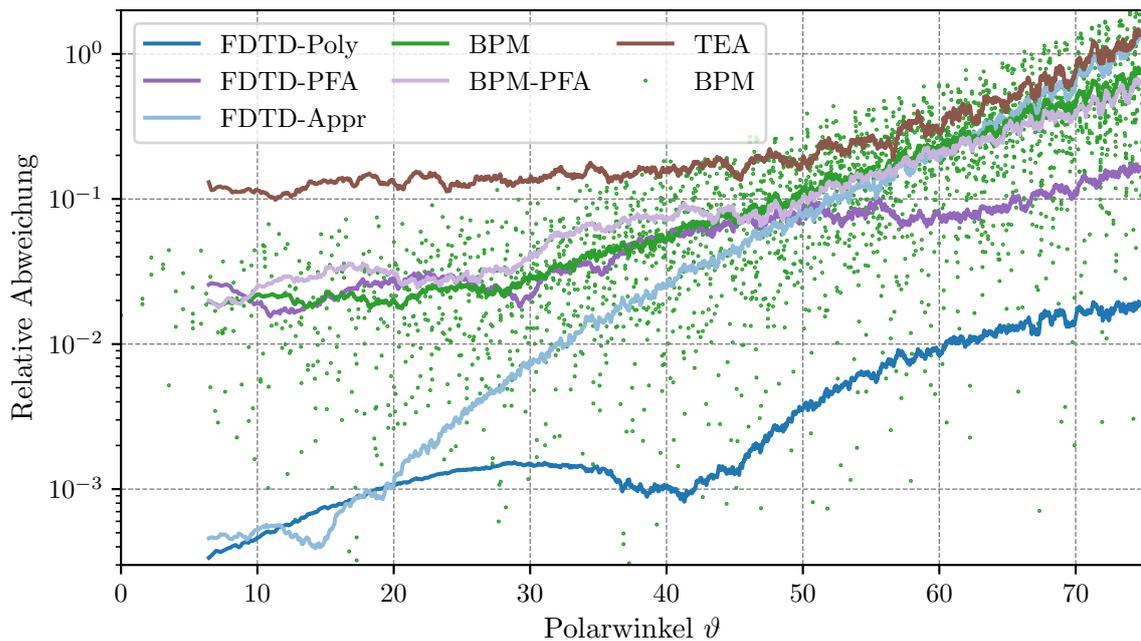


Abb. 4.12: Betrag der relativen Abweichung der berechneten Beugungseffizienz der Nutzordnungen gemäß den Methoden aus Tab. 4.3 in Abhängigkeit des Ablenkwinkels ϑ . Die Graphen zeigen einen gleitenden Mittelwert. Für die BPM werden zusätzlich die Werte der einzelnen Ordnungen dargestellt.

Die Abhängigkeit der Abweichung von φ und ϑ wird in Abb. 4.13 für die einzelnen Ordnungen dargestellt. In der FDTD-PFA erkennt man im Zentrum eine große Übereinstimmung, nach außen kommt es zunächst zu positiven, bei größeren Winkeln zu negativen Abweichungen. Die

Abhängigkeit von φ ist eher schwach ausgeprägt. Bei FDTD-Appr sind die Abweichungen entlang der Achsen klein, da es dort kaum Kopplung zwischen den Komponenten gibt. In diagonaler Richtung gibt es, nach außen zunehmend, dagegen deutlich positive Abweichungen. Für die BPM kommt es bei großen Winkeln zu negativen Abweichungen entlang der Achsen, die sich mit dem vorher beschriebenen Effekt teilweise aufheben. Die Abweichung der BPM-PFA wird maßgeblich durch die Trainingsdaten beeinflusst und weist daher eine ähnliche Verteilung auf.

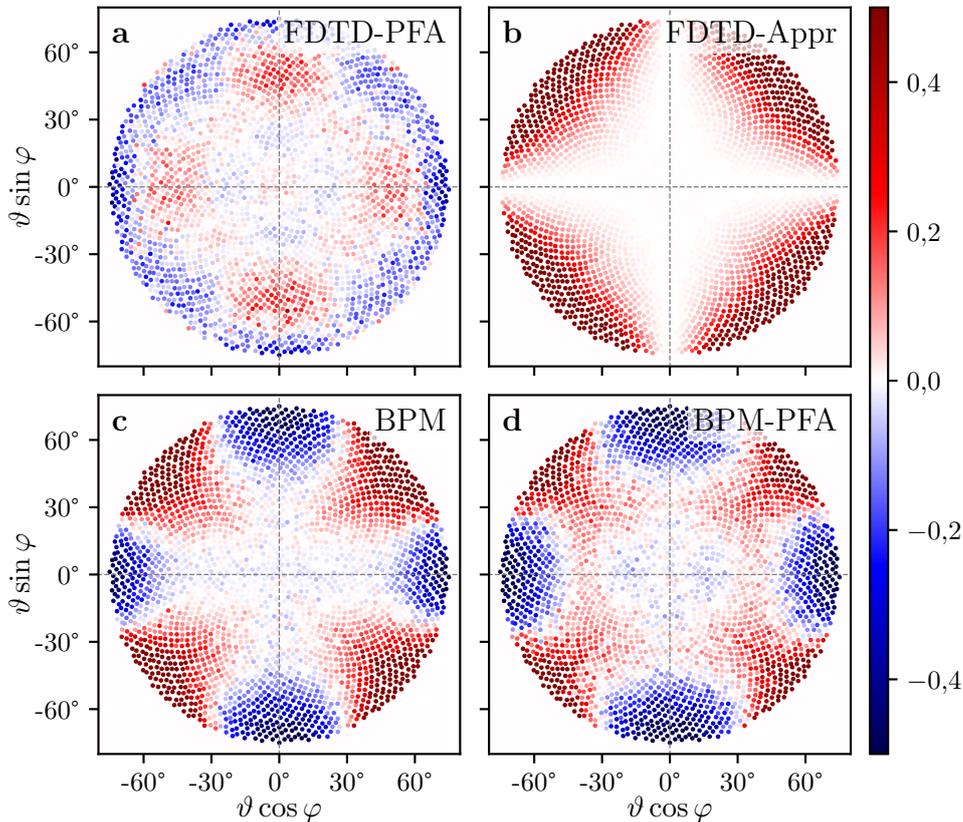


Abb. 4.13: Relative Abweichung der berechneten Beugungseffizienz in den Nutzordnungen gemäß der Projektion auf eine Sphäre für die Methoden FDTD-PFA (a), FDTD-Appr (b), BPM (c) und BPM-PFA (d).

Die vorgestellten Ergebnisse beziehen sich zwar auf ein einzelnes DOE und dessen Zielverteilung, dieses ist aber so gewählt, dass sich möglichst allgemein gültige Aussagen treffen lassen. Tatsächlich erhält man für andere DOEs ähnliche Resultate, insbesondere unabhängig vom maximalen Ablenkwinkel. Insgesamt lässt sich feststellen, dass die BPM bis zu Ablenkwinkeln von etwa 30° sehr gute Resultate liefert und die Ergebnisse bis etwa 50° noch akzeptabel erscheinen. Bei größeren Winkeln führt allein die Vernachlässigung der Kopplung schon zu deutlichen Abweichungen. Mithilfe der PFA lassen sich ebenfalls präzise Ergebnisse erzielen. Die Güte der Approximation wird dabei maßgeblich durch die Auswahl der Trainingsdaten und die Anzahl der Modellparameter beeinflusst. Im gezeigten Fall erzeugt die BPM-PFA ähnlich gute Resultate wie die BPM selbst, mit der FDTD-PFA können sogar bei Winkeln bis etwa 65° gute Ergebnisse erzielt werden.

5 Semi-rigoroses Design von CGHs

In den vorigen Kapiteln wurden effiziente Algorithmen zur Analyse von DOEs vorgestellt, welche insbesondere die Wechselwirkung benachbarter Subwellenlängenstrukturen berücksichtigen. Im Folgenden wird die Anwendung dieser Verfahren auf das Design von CGHs beschrieben. Dazu wird die schnelle Fernfeldberechnung der PFA mit einem IFTA verknüpft. Die BPM dient dabei zur effizienten Erzeugung von Trainingsdaten für die Modelle. Ziel dieses Kapitels ist das Design eines CGHs, dessen Berechnung mit herkömmlichen Designverfahren nicht in gleicher Qualität möglich ist. Dazu werden zunächst Variationen des IFTA eingeführt, die als Referenz für den Vergleich der neu entwickelten Verfahren dienen. Anschließend wird die Verknüpfung von IFTA und PFA zum Design von binären CGHs beschrieben, wobei ein Verfahren zur direkten Berechnung der Struktur für die Erzeugung eines bestimmten Nahfeldes verwendet wird. Die berechneten Elemente wurden hergestellt und experimentell untersucht. Diese Ergebnisse sind in Ref. [111] veröffentlicht und werden hier in überarbeiteter Form wiedergegeben. Schließlich wird eine Methode zum Design nicht-binärer CGHs mit diskreten Phasenstufen unter Nutzung der PFA geschildert.

5.1 Anpassung des IFTA

Für den experimentellen Nachweis der Vorteile des neuen Algorithmus ist es hilfreich, möglichst gleichmäßige Intensitätsverteilungen zu realisieren. Abweichungen von der gewünschten Verteilung lassen sich so direkt auf Unterschiede zwischen verschiedenen Design- bzw. Simulationsverfahren zurückführen. Daher wurden die genutzten Berechnungsvarianten darauf optimiert, eine hohe Übereinstimmung mit der Vorgabe zu erreichen. Zunächst werden hier Varianten vorgestellt, die ohne eine physikalische Simulation des CGHs innerhalb der Iterationen auskommen.

5.1.1 Phasenprofile für gleichmäßige Fernfeldverteilungen

Aus den bereits vorgestellten Varianten und Modifikationen des IFTA wurde ein Iterationsschema zusammengestellt, welches die Berechnung von CGHs mit diskreten Phasenstufen und möglichst hoher Übereinstimmung zwischen gewünschter und realisierter Intensitätsverteilung ermöglicht. Es ähnelt einer Kombination der in Ref. [112] empfohlenen Varianten. Ausgehend von einem Startfeld $u^{(0)} = u^{\text{in}} \exp(i\phi^{(0)})$ werden die in Abschnitt 2.3.3 angegebenen Schritte 1, 2c kombiniert

mit 2d, 3 und 4a verwendet:

1. Propagation ins Fernfeld: $U^{(j)} = \mathcal{F} \{u^{(j)}\}$.

2. Ersetzen der Amplitude mit Überanpassung und separater Skalierung:

$$U^{(j)} = \left(\chi_S \alpha_S^{(j)} \frac{I'^{\beta_j/2}}{|U^{(j)}|^{\beta_j}} \sqrt{I'} + \chi_A \alpha_A^{(j)} |U^{(j)}| \right) \exp \left(i \arg U^{(j)} \right)$$

3. Rückpropagation ins Nahfeld: $u'^{(j)} = \mathcal{F}^{-1} \{U^{(j)}\}$.

4. Ersetzen der Amplitude bei diskreter Phase: $u^{(j+1)} = u^{\text{in}} \exp \left(i \mathcal{D}_{\Phi, \gamma_j} \arg \frac{u'^{(j)}}{u^{\text{in}}} \right)$.

Darin bezeichnen $\chi_{S/A}$ die charakteristischen Funktionen des Signal- und Amplitudenfreiheitsbereiches, I' die gemäß Gl. (2.69) normierte Zielverteilung der Intensität und $\mathcal{D}_{\Phi, \gamma_j}$ den Projektionsoperator auf die diskreten Phasenwerte in Φ aus Gl. (2.73). Die Faktoren für die Ersetzung der Amplitude bestimmen sich als

$$\alpha_S^{(j)} = \frac{\sum_{m_x, m_y} I'_{m_x m_y}^{(\beta_j+1)/2} |U_{m_x m_y}^{(j)}|^{1-\beta_j}}{\sum_{m_x, m_y} I'_{m_x m_y}^{\beta_j+1} |U_{m_x m_y}^{(j)}|^{-2\beta_j}}, \quad (5.1)$$

gemäß Gl. (2.70) unter Berücksichtigung der Überanpassung, bzw.

$$\alpha_A^{(j)} = \min \left\{ 1, \sqrt{\frac{\alpha_{\max}}{\alpha'_j}} \right\}, \quad (5.2)$$

mit α'_j wie in Gl. (2.71) und einem maximalen Intensitätsanteil von α_{\max} außerhalb des Signalbereiches. Der Parameter β_j steigt im Lauf der Iterationen von 0 auf z. B. 2, während γ_j beginnend von 0 solange erhöht wird, bis die Phasenwerte vollständig diskretisiert werden. Dadurch hat die Iteration anfänglich keine Korrektur der Amplitude bei Nutzung einer stetigen Phase und ändert sich kontinuierlich hin zu einer starken Korrektur bei diskreten Phasenwerten. Das Phasenprofil des CGHs ergibt sich nach einigen Iterationen als $\phi = \arg(u^{(j+1)}/u^{\text{in}})$. Das vorgestellte Iterationsschema ist für die Erzeugung gleichmäßiger Verteilungen nützlich und wird daher hier als homogenisierender IFTA bezeichnet. Für ein binäres Element, wie es im folgenden Abschnitt betrachtet wird, ist $\Phi = \{-\pi/2, \pi/2\}$ und die Projektion der Phase vereinfacht sich zu

$$\mathcal{D}_{\Phi, \gamma} \phi := \begin{cases} -\frac{\pi}{2}, & \text{für } \left| \phi + \frac{\pi}{2} \right| < \gamma, \\ \frac{\pi}{2}, & \text{für } \left| \phi - \frac{\pi}{2} \right| < \gamma, \\ \phi, & \text{sonst,} \end{cases} \quad (5.3)$$

wobei γ maximal auf $\pi/2$ erhöht wird.

5.1.2 Modifikation der Zielverteilung

Für die möglichst genaue Realisierung der Intensitätsverteilung ist die sinc-Modulation aus Gl. (2.64) zu beachten. Dazu kann eine modifizierte Zielverteilung genutzt werden:

$$I_{S,m_x m_y} := \frac{I_{m_x m_y}}{S_{00,m_x m_y}^2}, \quad s_{00,m_x m_y} := \text{sinc}\left(\frac{\pi m_x}{M_x}\right) \text{sinc}\left(\frac{\pi m_y}{M_y}\right), \quad (5.4)$$

wobei M_x und M_y die Gesamtzahl der Pixel in der jeweiligen Dimension sind. Das auf Basis von I_S berechnete Phasenprofil korrigiert dann die Modulation im Fernfeld.

Ein solches Profil wird beispielsweise von einer Effektiv-Index-Struktur erzeugt und kann unter Nutzung der LPA berechnet werden, wie in Abschnitt 3.2.2. Analysiert man die optische Funktion eines solchen CGHs, z. B. mittels der PFA, lassen sich systematische Abweichungen zur gewünschten Zielverteilung beobachten, vgl. Abschnitt 4.3.2. Beispielsweise kann es zu einer weiteren Abschwächung höherer Ordnungen kommen. Dieser Effekt lässt sich durch eine weitere Modifikation der Zielverteilung korrigieren. Dazu berechnet man zunächst einige CGHs und analysiert deren Fernfelder mittels der PFA. Das Verhältnis der resultierenden Intensitätsverteilungen $I'_{m_x m_y}$ zur Zielverteilung kann durch Legendre-Polynome approximiert werden:

$$\frac{I'_{m_x m_y}}{I_{S,m_x m_y}} \approx I_{P,m_x m_y} = \sum_{j_x=0}^{N_L} \sum_{j_y=0}^{N_L-j_x} l_{j_x j_y} L_{j_x}\left(\frac{\pi m_x \lambda}{M_x p_x}\right) L_{j_y}\left(\frac{\pi m_y \lambda}{M_y p_y}\right). \quad (5.5)$$

Dabei sind p_x, p_y die Größen eines Pixels und λ die Wellenlänge. Es finden nur die Beugungsordnungen innerhalb des Signalbereiches, $(m_x, m_y) \in S$, Berücksichtigung. Details zu den orthogonalen Legendre-Polynomen sind im Anhang, Abschnitt A.1.3.1 angegeben. Der maximale Grad der Polynome N_L wird so gewählt, dass eine weitere Erhöhung die resultierende Abweichung nur wenig verringert. Die Verwendung verschiedener CGHs stellt sicher, dass der approximierte Verlauf nicht von einem konkreten Phasenprofil, sondern durch die verwendeten Strukturen und die geforderte Verteilung an sich bestimmt wird. Das geschilderte Prinzip ist in Abb. 5.1 dargestellt.

Nutzt man die Zielverteilung

$$I_{C,m_x m_y} := \frac{I_{S,m_x m_y}}{I_{P,m_x m_y}} \quad (5.6)$$

für den IFTA, so wird die Abschwächung der höheren Ordnungen zusätzlich zur sinc-Modulation kompensiert. Abweichungen des simulierten Fernfeldes, die nur einzelne Ordnungen betreffen, können auf diese Weise dagegen nicht berücksichtigt werden.

Die Beugungseffizienz einer Ordnung (m_x, m_y) ist proportional zur Intensität bzw. zum Quadrat der Fernfeldamplitude multipliziert mit dem Kosinus des Polarwinkels, siehe Gl. (2.35):

$$\eta_{m_x m_y} \propto \cos \vartheta_{m_x m_y} I_{m_x m_y} \propto \cos \vartheta_{m_x m_y} |U_{m_x m_y}|^2. \quad (5.7)$$

Wenn sich die Zielverteilung auf die Leistung anstatt auf die Intensität in den Beugungsordnungen beziehen soll, so ist der Kosinus-Faktor zusätzlich zur sinc-Modulation für die Zielverteilung zu

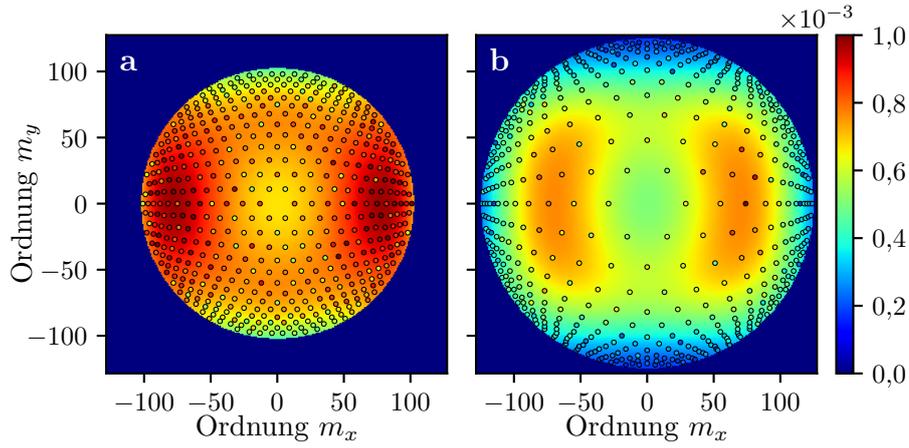


Abb. 5.1: Simulierte Intensität I' im Fernfeld eines Strahlteiler-CGHs und zugehörige Polynomapproximation $I_P I_S$ für $N_L = 3$. Die kleinen Kreise zeigen die Intensität der Nutzordnungen, der Hintergrund stellt den Verlauf der Approximation dar. Die gezeigten Beispiele entsprechen den in Tab. 5.1 beschriebenen Elementen A1 (a) und B1 (b).

berücksichtigen:

$$I_{P,m_x m_y} := \frac{I_{S,m_x m_y}}{\cos \vartheta_{m_x m_y}} . \quad (5.8)$$

5.2 Design und experimentelle Untersuchung von binären CGHs

Die Nutzung von semi-rigorosen Simulationen für die Iterationsschritte erfordert weitere Anpassungen des IFTA. Bei den hier betrachteten binären Elementen hat jeder Pixel entweder ein vollständig hohes ($f = 1$) oder tiefes Profil ($f = 0$). Der Höhenunterschied zwischen beiden ist so gewählt, dass die Phasendifferenz für die Designwellenlänge π beträgt.

5.2.1 Bestimmung der Struktur zu einem Phasenprofil

Mit der PFA wurde ein effizientes Verfahren zur Berechnung des Feldes nach einer Struktur unter Beachtung der umgebenden Pixel vorgestellt. Die Zuordnung einer Struktur zu einer berechneten Phase basiert aber in den bislang vorgestellten IFTA-Varianten stets auf der LPA. Um in dieser Zuordnung ebenfalls den Einfluss der Umgebung einzubeziehen, wurde ein Berechnungsverfahren entwickelt, das in Anlehnung an die PFA als *Polynomial Structure Approximation* (polynomielle Approximation der Struktur, PSA) bezeichnet wird. Wie in Gl. (4.14) für die PFA wird hier ein lineares Modell erstellt, um die Geometrieparameter der Pixel mit den Legendre-Koeffizienten des Feldes zu verknüpfen. Allerdings dienen die Koeffizienten des Feldes diesmal als Argumente zur Berechnung der Füllfaktoren. Dabei werden ausschließlich die Koeffizienten 0. Ordnung

genutzt, welche den Mittelwerten der Felder innerhalb eines Pixels entsprechen. Die Verwendung höherer Ordnungen ist denkbar, hat aber keine Vorteile in den durchgeführten Untersuchungen gebracht. Da es sich um komplexe Größen handelt, wurden Real- und Imaginärteil separat in der Modellerstellung berücksichtigt. Im Allgemeinen ergibt sich auf diese Weise eine reelle Zahl f , die für $f > 1/2$ als hohes, andernfalls als tiefes Profil interpretiert wird.

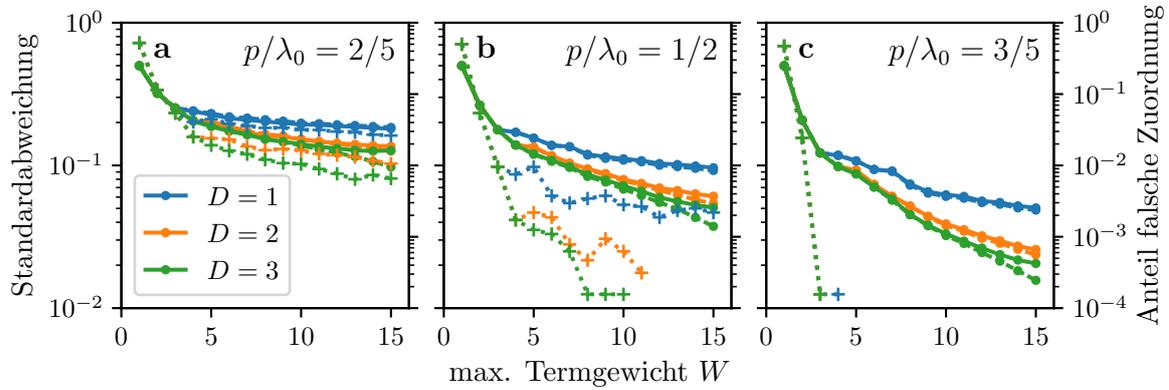


Abb. 5.2: Standardabweichung des Residuums der Testdaten (durchgezogen, linke Skala) und Trainingsdaten (gestrichelt, linke Skala) für den Füllfaktor in Abhängigkeit des Termgewichtes W und maximalen Abstandes D für verschiedene Verhältnisse p/λ_0 , Anteil der falsch zugeordneten Pixel der Testdaten (gepunktet, rechte Skala), insofern größer als 0.

Für die PSA ist, wie für die PFA, eine geeignete Auswahl der im Modell berücksichtigten Terme notwendig, insbesondere durch die Parameter Wechselwirkungsreichweite D und maximales Termgewicht W . Im Vergleich zur PFA, Gl. (4.13), wurde die Berechnung des Termgewichtes in Abhängigkeit der jeweiligen Exponenten leicht variiert:

$$w(\mathbf{e}^s) := \sum_{d=0}^D e_d^s (2 + d) , \quad w(\mathbf{e}^a) := \sum_{d=1}^D e_d^a (2 + d) . \quad (5.9)$$

Auf diese Weise werden bei ähnlicher Termanzahl weiter entfernte Koeffizienten anstelle höherer Potenzen einbezogen. Zur Bewertung des Verfahrens kann wiederum die Standardabweichung des Residuums aus der linearen Regression herangezogen werden. Für einen Trainings- und Testdatensatz, analog zu dem in Abschnitt 4.2.3, allerdings mit einer rein binären Struktur, sind die Ergebnisse in Abb. 5.2 dargestellt. Erkennbar ist eine deutliche Abnahme der Standardabweichung mit steigendem Termgewicht W , also höherer Anzahl der Terme. Der Kurvenverlauf ist dabei für Trainings- und Testdaten weitestgehend ähnlich und eine deutliche Überanpassung erst im weiteren Verlauf zu erwarten. Die Abweichung sinkt bei einem größeren Verhältnis p/λ_0 deutlich schneller. In den gezeigten Fällen ist $D = 2$ eine plausible Wahl für den maximalen Abstand, eine weitere Erhöhung bringt nur geringfügige Verbesserungen. Der Anteil der durch die Approximation korrekt zu den Höhenstufen zugeordneten Pixel ist ein weiteres Qualitätskriterium. Bei kleiner Pixelgröße (a) werden für größeres W ca. 1% der Pixel falsch zugeordnet, im Fall (b) werden

schon ab etwa $W = 10$ alle Pixel der Testdaten korrekt erkannt, für noch größere Pixel bereits ab $W = 3$, also mit Modellen mit nur wenigen Koeffizienten. Bemerkenswert ist, dass die Güte der Modelle hier sehr deutlich von der Pixelgröße abhängt: Für Verhältnisse von $p/\lambda_0 < 0,4$ wird die Zuordnung zunehmend schwierig, während sie für $p/\lambda_0 > 0,6$ quasi perfekt funktioniert.

Mithilfe der PSA kann auf zuverlässige Weise die binäre Struktur bestimmt werden, die eine bestimmte Feldverteilung transmittiert. Allerdings enthalten die Trainingsdaten nur Felder hinter realen Strukturen. Wendet man die PSA auf eine beliebige Feldverteilung an, ergeben sich nicht unbedingt sinnvolle Ergebnisse, da im Allgemeinen keine binäre Struktur existiert, die diese Verteilung gut approximiert. Deutlich wird dies an Ergebnissen für den Füllfaktor, die stark von 0 oder 1 abweichen.

5.2.2 Verwendung von Polynomapproximationen im IFTA

Um semi-rigore Simulationsverfahren direkt in den Designprozess einzubeziehen, wurden die Propagationsschritte variiert. Das Nahfeld eines binären Elementes wird mittels der PFA berechnet, unter Beachtung der nötigen Reduktion der Termanzahl für binäre Elemente. Die Propagation ins Fernfeld erfolgt, wie in Abschnitt 4.1.2, als separate FT der Polynomkoeffizienten $c_{j_x j_y}$ der jeweiligen Ordnung (j_x, j_y) unter Beachtung der zugehörigen Modulation $s_{j_x j_y}$, siehe Gl. (4.5) bzw. (A.72) im Anhang, wobei s_{00} identisch zu dem in Gl. (5.4) angegebenen Faktor ist. Die Amplitude zur Ersetzung wird aus der Superposition der Ordnungen im Fernfeld ermittelt. Dann wird die Amplitude der 0. Polynomordnung C_{00} ersetzt und durch inverse FT ins Nahfeld zurück propagiert. Zu diesem Feld wird nun mittels der PSA eine Struktur errechnet, zunächst repräsentiert durch einen Wert f für jeden Pixel. Eine harte Projektion auf binäre Werte $f' \in \{0, 1\}$ würde zu einer Stagnation der Iteration führen, wie für binäre Phasenwerte. Ähnlich wie im herkömmlichen IFTA wird stattdessen eine vom Parameter γ abhängige Projektion genutzt, die nur für $|f - 1/2| \geq \gamma$ zu binären Werten führt, wobei γ zu Beginn der Iteration z. B. auf 1 gesetzt wird und in deren Verlauf bis auf 0 reduziert wird, im Gegensatz zum Vorgehen in Abschnitt 5.1.1. Nur für diejenigen Pixel mit klarer Zuordnung, d. h. binärem f , wird das Nahfeld direkt mittels der PFA bestimmt. Andernfalls wird die Phase des Koeffizienten der 0. Ordnung c_{00} durch die Phase des zurück propagierten Feldes ersetzt, wie im ursprünglichen IFTA. Die Koeffizienten höherer Ordnung bleiben unverändert. Auf diese Weise erhöht sich die Zahl der Freiheitsgrade, insbesondere für die anfänglichen Iterationsschritte.

Als Ausgangspunkt des Designs dient eine zufällige binäre Struktur f' bzw. die mittels PFA berechneten Koeffizienten des zugehörigen Nahfeldes $c_{j_x j_y}^{(0)} = u^{\text{in}} \cdot \text{PFA}(f')$. Zusammengefasst ergibt sich das folgende Iterationsschema für den modifizierten binären IFTA:

1. Propagation ins Fernfeld und Superposition: $C_{j_x j_y}^{(j)} = \mathcal{F} \{ c_{j_x j_y}^{(j)} \}$, $U^{(j)} = \sum_{j_x, j_y} s_{j_x j_y} C_{j_x j_y}^{(j)}$;

2. Berechnung der neuen Fernfeldamplitude: $|U^{(j)}| = \chi_S \alpha_S^{(j)} \frac{I^{\beta_j/2}}{|U^{(j)}|^{\beta_j}} \sqrt{I} + \chi_A \alpha_A^{(j)} |U^{(j)}|$;
3. Ersetzen der Amplitude der 0. Ordnung: $C_{00}^{\prime(j)} = C_{00}^{(j)} \frac{|U^{(j)}|}{|U^{(j)}|}$;
4. Rückpropagation ins Nahfeld: $c_{00}^{\prime(j)} = \mathcal{F}^{-1} \{C_{00}^{\prime(j)}\}$;
5. Approximation der Struktur: $f^{(j)} = \text{PSA} \left(\frac{c_{00}^{\prime(j)}}{u^{\text{in}}} \right)$;
6. binäre Projektion: $f^{(j)} = \begin{cases} 1, & \text{für } f^{(j)} \geq 1/2, \\ 0, & \text{sonst;} \end{cases}$
7. Berechnung des Nahfeldes: $c_{j_x j_y}^{\prime(j)} = u^{\text{in}} \cdot \text{PFA} (f^{(j)})$;
8. Ersetzen der Phase in 0. Ordnung: $c_{00}^{(j+1)} = \begin{cases} |c_{00}^{\prime(j)}| \exp(i \arg c_{00}^{\prime(j)}), & \text{für } |f^{(j)} - 1/2| < \gamma_j, \\ c_{00}^{\prime(j)}, & \text{sonst;} \end{cases}$
 bzw. $c_{j_x j_y}^{(j+1)} = c_{j_x j_y}^{\prime(j)}$, falls $j_x > 0$ oder $j_y > 0$.

Für die multiplikative Verknüpfung von Eingangsfeld und PFA bzw. PSA wurde dabei vorausgesetzt, dass sich das Eingangsfeld innerhalb eines Pixels höchstens geringfügig ändert, wie beispielsweise bei der Beleuchtung des CGHs mit einem kollimierten Gaußstrahl, der einen Durchmesser deutlich oberhalb der Pixelgröße aufweist.

5.2.3 Berechnung eines Weitwinkel-Strahlteilers

Um die Überlegenheit des modifizierten IFTA-Algorithmus zu demonstrieren, wurden CGHs untersucht, die als Strahlteiler mit großen Öffnungswinkeln fungieren. Die experimentelle Untersuchung kann so auf die Messung der Beugungseffizienz einiger weniger Ordnungen beschränkt werden, anstatt die gesamte Fernfeldverteilung präzise analysieren zu müssen. Als Zielverteilung wurde eine hexagonale Anordnung von 516 Beugungsordnungen gewählt, in welche das einfallende Licht möglichst gleichmäßig aufgeteilt werden soll. Diese Anforderung ist für potentielle Anwendungen entscheidend, damit in einer fotografischen Aufnahme der Intensitätsverteilung die gewollten Nutzordnungen sicher von anderen Beugungsordnungen, die zwangsläufig ebenfalls eine gewisse Intensität aufweisen, unterschieden werden können. Erwartungsgemäß sollten die mit einem herkömmlichen IFTA entworfenen CGHs bei einer semi-rigorous Simulation deutliche Abweichungen von der Zielverteilung zeigen, insbesondere bei großen Ablenkungswinkeln. Die CGHs wurden mit Öffnungswinkeln von 100° (A) bzw. 140° (B) entworfen. Bei diesen Winkeln tritt eine deutliche Verzeichnung der Position der Beugungsordnungen auf. Die Auswahl der Beugungsordnungen in

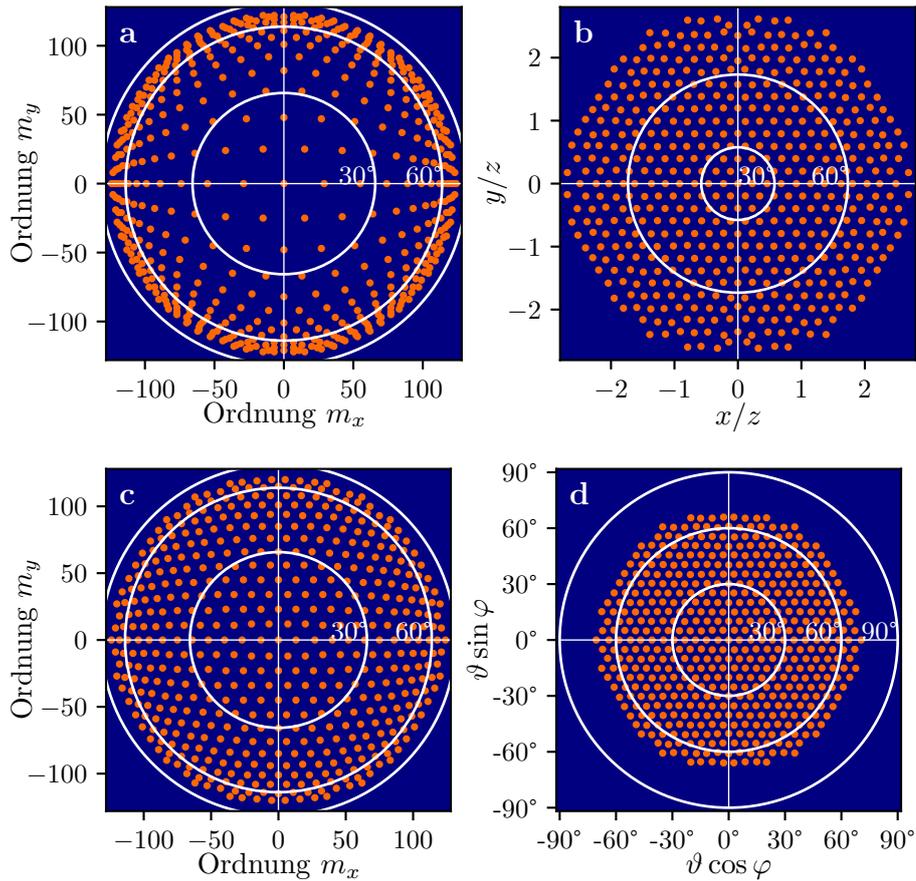


Abb. 5.3: Nutzordnungen der Zielverteilung für die Designvariante E (a), Projektion auf eine Ebene (b), Nutzordnungen für Variante S (c), radial äquidistante Darstellung der Projektion auf eine Sphäre (d). Die Kreise zeigen die Lage der angegebenen Ablenkwinkel.

der Zielverteilung erfolgte dergestalt, dass sich ein gleichmäßiges Muster bei der Projektion auf eine Ebene (E) bzw. Sphäre (S) ergibt, siehe Abb. 5.3.

Die hier vorgestellten CGHs wurden für eine Wellenlänge von 632,8 nm berechnet und bestehen aus 256×256 quadratischen Pixeln in periodischer Fortsetzung zur Vereinfachung der experimentellen Untersuchung. Die Seitenlänge der Pixel beträgt 325 nm, sodass sich eine Gesamtgröße der Einheitszellen von $83,2 \times 83,2 \mu\text{m}^2$ ergibt. Die Strukturen sollen mit einer Tiefe von 692 nm in Quarzglas geätzt werden, was gemäß der TEA genau einem Phasenunterschied von π zwischen geätzten und nicht geätzten Pixeln entspricht. Die geschilderten Algorithmen wurden genutzt, um verschiedene CGHs zu entwerfen: Als Referenz basierend auf herkömmlichen Verfahren wurden zwölf Iterationen des originalen IFTA (IO), siehe Abschnitt 2.3.3.1, mit 48 Iterationen des homogenisierenden IFTA (IH), Abschnitt 5.1.1, kombiniert. Auf diese Weise wird zunächst eine hohe Effizienz in der Zielverteilung realisiert und dann eine gleichmäßige Verteilung basierend auf einem binären Profil. Dabei wurden die enthaltenen Parameter folgendermaßen festgelegt: $\alpha_{\text{max}} = 0,2$, β_j steigt linear von 0 auf 2 und γ_j linear von 0 auf $\pi/2$. Dieselbe Kombination der Algorithmen diente auch zur Berechnung mit einer gemäß Gl. (5.6) korrigierten Zielverteilung (IO*/IH*). Der

modifizierte binäre IFTA (IB) aus dem vorherigen Abschnitt wurde mit 30 Iterationen direkt auf die Zielverteilung angewendet. Der Wert für γ_j wird dabei linear von 1 auf 0 gesenkt, die anderen Parameter sind identisch zu vorher. Die getroffene Wahl von γ_j erzwingt stets die Konvergenz zu einem binären Element in der vorgegebenen Zahl von Schritten. In allen Fällen ließ sich keine signifikante Verbesserung der Resultate durch eine höhere Zahl an Iterationen erreichen. Durch die Nutzung der effizienten PFA bzw. PSA kann in jedem Iterationsschritt des modifizierten IFTA eine semi-rigore Analyse des gesamten Elementes durchgeführt werden, während die Laufzeit mit ca. 5 min auf einem Desktop-PC gering bleibt. Die folgenden Maße wurden zum Vergleich der Ergebnisse berechnet:

$$\text{Effizienz:} \quad \frac{\sum \chi_S |U|^2}{\sum |U|^2}, \quad (5.10)$$

$$\text{Homogenität:} \quad \frac{\min_S \{|U|^2\}}{\max_S \{|U|^2\}}, \quad (5.11)$$

$$\text{SNR [dB]:} \quad -10 \log_{10} \left(1 - \frac{(\sum \chi_S \sqrt{I} |U|)^2}{\sum \chi_S |U|^2 \sum \chi_S I} \right). \quad (5.12)$$

Dabei ist die Summation jeweils über alle Ordnungen (m_x, m_y) zu verstehen, Minimum und Maximum beziehen sich auf den Signalbereich S und χ_S ist dessen Indikatorfunktion. Die Abkürzung SNR bezeichnet das *Signal-Rausch-Verhältnis* (signal-to-noise ratio). Für jedes Design wurden jeweils zehn verschiedene zufällige Startphasen ausgewählt und das CGH mit der höchsten Homogenität gemäß der PFA-Simulation ausgewählt.

Tab. 5.1: Parameter verschiedener CGH-Designs zur Erzeugung eines gleichmäßigen Musters bei Projektion auf eine Ebene (E) bzw. Sphäre (S), bestimmt auf Grundlage von PFA-Simulationen. Für die Berechnung wurden der originale IFTA (IO), der homogenisierende IFTA (IH), bzw. der modifizierte binäre IFTA (IB) verwendet. Der Stern (*) kennzeichnet eine vorkorrigierte Zielfunktion.

Name	Öffnungswinkel	Projektion	Algorithmus	Homogenität	Effizienz	SNR
A1	100°	E	IO+IH	0,410	0,443	22,4
A2	100°	E	IB	0,477	0,434	24,5
A3	100°	S	IO*+IH*	0,487	0,461	25,1
A4	100°	S	IB	0,523	0,460	24,7
B1	140°	E	IO+IH	0,185	0,260	15,6
B2	140°	E	IB	0,443	0,239	25,1
B3	140°	S	IO*+IH*	0,497	0,317	24,7
B4	140°	S	IB	0,507	0,301	24,7

Die Eigenschaften der berechneten CGHs sind in Tab. 5.1 zusammengefasst. Im Allgemeinen

zeigen die Elemente für die Projektion auf eine Ebene (E) schlechtere Resultate als solche für die Projektion auf eine Sphäre (S), denn dabei werden in der Zielfunktion tendenziell höhere Ordnungen adressiert. Mit dem konventionellen Designansatz (A1/B1) kann insbesondere für größere Öffnungswinkel keine hohe Homogenität erzielt werden, weil die höheren Ordnungen gemäß der PFA-Simulation im Vergleich zur TEA abgeschwächt sind. Durch die vorkorrigierte Zielfunktion (A3/B3) ist eine Kompensation dieses Effektes möglich. Mit dem modifizierten binären IFTA kann die Homogenität weiter verbessert werden, während die Effizienz vergleichbar bleibt. Grundsätzlich verringert sich die Effizienz der Elemente, je mehr Intensität in hohe Ordnungen abgelenkt werden soll. Das SNR verhält sich ähnlich wie die Homogenität, ist aber weniger störanfällig durch einzelne Ordnungen. Es ist für das konventionelle Design (A1/B1) geringer und erreicht ansonsten ähnliche Werte.

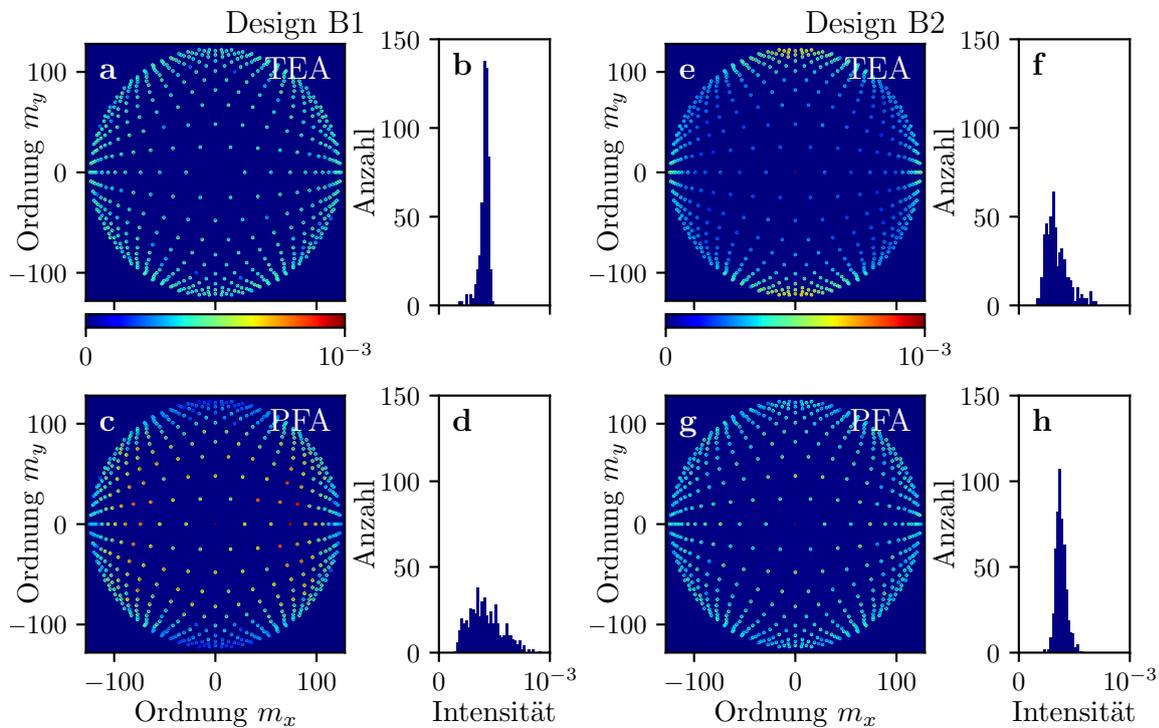


Abb. 5.4: Simuliertes Fernfeld der Designs B1 (a-d) und B2 (e-h) unter Nutzung von TEA bzw. PFA. Die Unterabb. a, c, e und g zeigen die räumliche Verteilung der Intensität im Fernfeld. In den anderen wird ein Histogramm der Intensität in den Nutzordnungen dargestellt.

In Abb. 5.4 wird das Fernfeld der Designs B1 und B2 vergleichend für eine Simulation mittels TEA bzw. PFA dargestellt. Für B1 zeigt das Ergebnis der TEA (a) eine recht homogene Verteilung, deutlich erkennbar am schmalen Peak im Histogramm (b). Wird dasselbe Element mit der genaueren PFA analysiert (c,d) werden Unterschiede in der Intensität der Nutzordnungen sichtbar, insbesondere in Form einer Abschwächung höherer Ordnungen. Umgekehrt zeigt sich bei B2 zwar eine Überhöhung der höheren Ordnungen in der TEA-Simulation (e,f), im Rahmen

der PFA wird damit aber eine gleichmäßige Verteilung erreicht (g,h), da die zu erwartenden Abweichungen kompensiert werden. Dieses Beispiel verdeutlicht, wie durch die Nutzung der PFA im Designalgorithmus die Leistungsfähigkeit der CGHs verbessert werden kann. Für die Designs B3 und B4 würden die entsprechenden Darstellungen ähnlich wie für B2 aussehen. Bei kleineren Öffnungswinkeln (A1 bis A4) sind die Unterschiede zwischen TEA und PFA geringer, aber weiterhin erkennbar.

5.2.4 Herstellung der Strahlteiler

Zur Herstellung der binären Strukturen auf einem Quarzglassubstrat wurden lithografische Standardverfahren genutzt [113], siehe Abb. 5.5. Zunächst wurde das Substrat mit einer 80 nm dicken Chromschicht und 300 nm eines Elektronenstrahllacks (FEP 171, Fuji Film) überschichtet (a). Durch einen Elektronenstrahlschreiber (Vistec SB350 OS) wurden die Strukturen belichtet (b) und durch die Entwicklung im Lack erzeugt (c). Danach wurden sie mittels Ionenstrahlätzen in die Chromschicht übertragen (d), damit diese als Maske für den folgenden Ätzprozess wirkt. Anschließend erfolgte das Tiefätzen ins Quarzglas durch reaktives Ionenätzen auf Basis eines induktiv gekoppelten Plasmas (e). Schließlich wurden die Reste der Chrommaske nasschemisch entfernt (f). Die verschiedenen Designs befinden sich auf demselben Substrat, sodass identische Herstellungsbedingungen gewährleistet waren. Die berechneten Einheitszellen wurden periodisch fortgesetzt, um eine Fläche von $5 \times 5 \text{ mm}^2$ zu füllen, was die optische Charakterisierung vereinfacht.

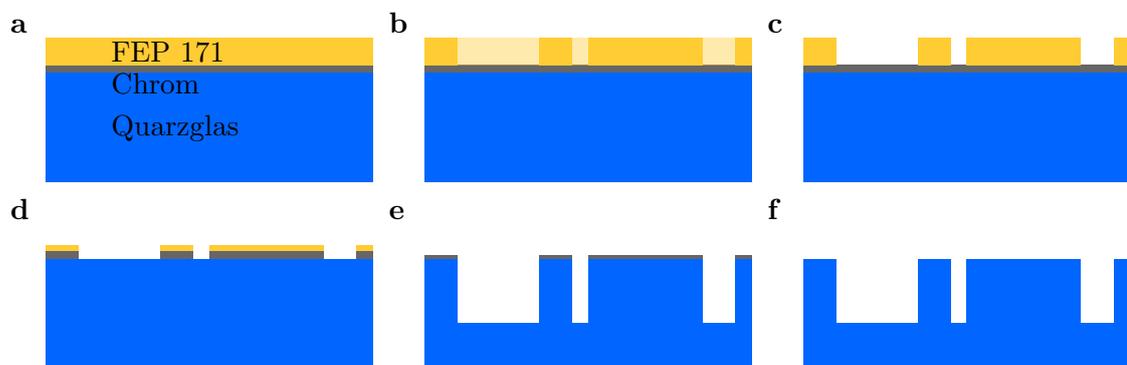


Abb. 5.5: Schematische Darstellung des Herstellungsprozesses für binäre CGHs: Beschichtung des Substrates (a), Elektronenstrahl-Belichtung (b), Entwicklung (c), Chromätzen (d), Quarzglasätzen (e), Ablösung von Chrom und Lack (f).

In Abb. 5.6 werden elektronenmikroskopische Aufnahmen der hergestellten Strukturen gezeigt. Es ist gut erkennbar, dass die binären Strukturen mit sehr hoher Qualität reproduziert wurden (a,b). Allerdings zeigt sich abweichend vom Design, verursacht durch den Ätzprozess, eine Rundung der Kanten (b-d). Dadurch wird die optische Funktion der Elemente beeinflusst. Um den Einfluss dieser Formabweichung einschätzen zu können, wurden Simulationen mit verschieden starker

Ausprägung des Effektes durchgeführt, siehe den folgenden Abschnitt.

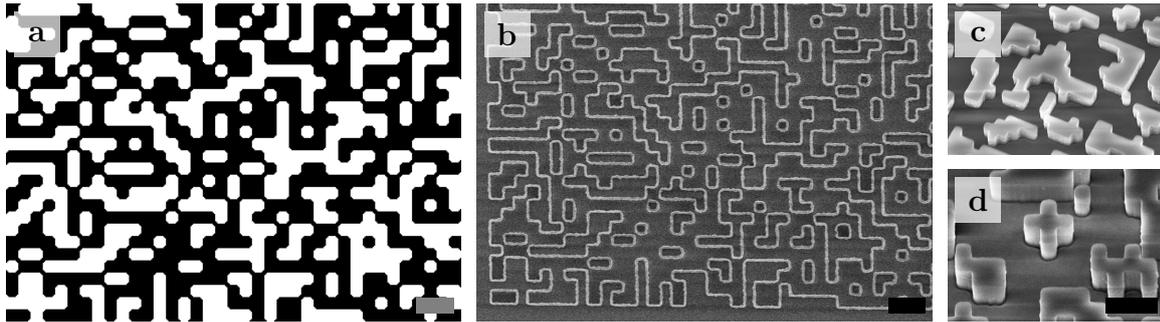


Abb. 5.6: Ausschnitt der berechneten Strukturen unter Beachtung der Kantenrundung (a), Aufnahme derselben Position eines hergestellten Elementes durch ein Raster-Elektronen-Mikroskop (REM, b), weitere REM-Bilder hergestellter Elemente (c,d). Der abgebildete Maßstab kennzeichnet jeweils eine Länge von 1 μm .

5.2.5 Experimentelle Untersuchung

Beleuchtet man die hergestellten CGHs mit einer Laserdiode, in diesem Fall mit einer Wellenlänge von 635 nm, so lässt sich im Fernfeld eine Intensitätsverteilung wie in Abb. 5.7 beobachten. Nahe der Bildmitte ist die 0. Ordnung als heller Punkt zu erkennen. In der rechten unteren Ecke sind einige Reflexionen an der hergestellten Probe sichtbar, welche nicht Teil des Beugungsbildes sind. Das Fernfeld zeigt im Allgemeinen die erwartete Verteilung. Die Beugungsordnungen zwischen den Nutzordnungen lassen sich im Bild fast nicht erkennen, erscheinen aber bei der Betrachtung mit bloßem Auge als deutlich schwächere Punkte. Intensitätsunterschiede zwischen den Nutzordnungen lassen sich in der Darstellung kaum ausmachen.

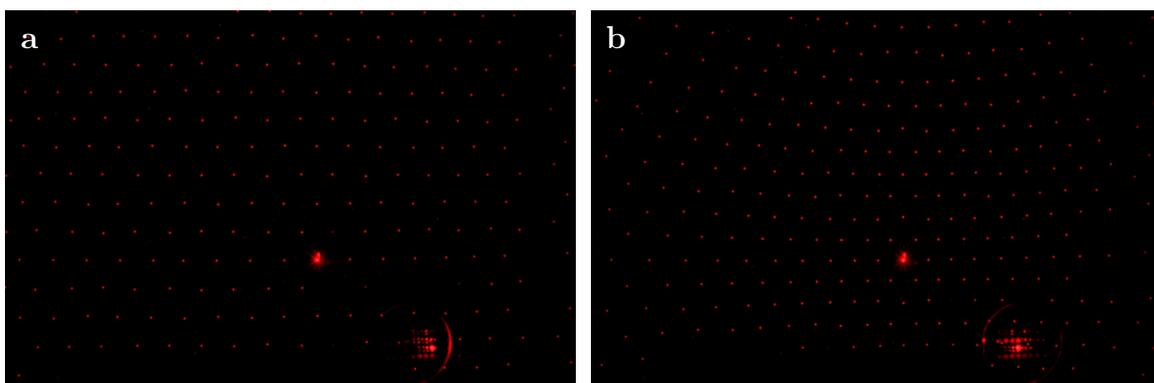


Abb. 5.7: Fotografische Aufnahme des Fernfeldes der Elemente A1 (a) und A3 (b). Der Bildausschnitt zeigt lediglich einen Teil des vollständigen Beugungsbildes.

Für die Vermessung der Intensität in den Beugungsordnungen wurden zwei verschiedene Aufbauten verwendet. Im Aufbau M1 wird die Probe unter senkrechtem Einfall durch einen kollimierten Laser mit kontrollierter Polarisationsrichtung beleuchtet. Dieser hat einen Strahldurchmesser von

2 mm und eine Wellenlänge von 632,8 nm. Die Leistung in einer bestimmten Beugungsordnung wird durch eine Ulbricht-Kugel gemessen, welche in einer goniometrischen Halterung montiert ist. So kann der vermessene Ablenkwinkel variiert werden. Mithilfe eines Strahlteilers direkt hinter der Quelle wird ein Referenzsignal gemessen, sodass Leistungsschwankungen der Quelle berücksichtigt werden können. Zur Reduzierung des Rauschens wird der Laserstrahl durch ein Chopper-Rad moduliert und ein Lock-in-Verstärker für die Aufzeichnung des Signals verwendet. Beim Aufbau M2 handelt es sich um eine vereinfachte Version. Er besteht ebenfalls aus einer Lichtquelle und einer Ulbricht-Kugel in einer goniometrischen Montage, verzichtet aber auf eine Referenzmessung und die Lock-in-Verstärkung und ist daher weniger genau. Er wurde insbesondere für die Messungen bei hohen Ablenkwinkeln genutzt. Die Wellenlänge des Lasers beträgt hier 635 nm. Vergleichende Simulationen zeigten, dass diese kleine Abweichung der Wellenlänge nur geringfügige Auswirkungen auf die Intensitätsverteilung der Nutzordnungen hat.

Zur Untersuchung der Intensitätsvariationen wurde das Fernfeld entlang der horizontalen Achse von Abb. 5.3 vermessen und die resultierende Leistung in den Nutzordnungen gemäß Gl. (5.8) bestimmt. Im Designprozess wurde eine homogene Intensität angestrebt, d. h. ohne Berücksichtigung des Kosinus-Faktors. Beispielhaft zeigt Abb. 5.8 die Messergebnisse für einige der CGHs. Zum Vergleich sind Simulationsergebnisse der PFA dargestellt, wobei das elektrische Feld für die Analyse der TE-Polarisation und das magnetische für die TM-Polarisation genutzt wurde. Der Hauptgrund für Unterschiede zwischen Messung und Simulation sind vermutlich Abweichungen der realen Struktur der Elemente vom Design. Um den Einfluss gewisser Herstellungsfehler zu untersuchen, wurden Simulationen mit veränderten Parametern durchgeführt. Eine abweichende Ätztiefe bewirkt eine Umverteilung der Leistung zwischen 0. Ordnung und den Nutzordnungen, aber beeinflusst kaum die relative Verteilung innerhalb der Nutzordnungen. Ein ähnliches Verhalten zeigt sich, wenn die transversalen Strukturgrößen bei konstanter Pixelgröße global skaliert werden. Diese Abweichung kann auftreten, wenn der Größenunterschied zwischen belichteten und tiefgeätzten Bereichen nicht exakt berücksichtigt wird. Andererseits führt Kantenrundung zu höherer Intensität in achsennahen Ordnungen, während die 0. Ordnung abgeschwächt wird und höhere Ordnungen weitestgehend unbeeinflusst bleiben. Dieser Effekt lässt sich generell beobachten, wird aber in Abb. 5.8(b,d) besonders deutlich.

Die im Fernfeld gemessenen Intensitätsverteilungen lassen sich im Allgemeinen durch die PFA nachvollziehen. Bei kleineren Ablenkwinkeln (a,c) gibt es nur vergleichsweise geringe Unterschiede zwischen TEA und PFA. Bei größeren Winkeln (b,d) sind diese jedoch deutlicher und treten ebenso für achsennahe Ordnungen auf. Aufgrund der stärkeren lokalen Strukturänderungen ist die TEA hier insgesamt nicht in der Lage, das Fernfeld präzise zu beschreiben, nicht nur für hohe Ordnungen. Gemäß der TEA sollte das Fernfeld eines binären Elementes symmetrisch zur 0. Ordnung sein. In der Messung bestätigt sich diese Symmetrie nicht, wie es gemäß Ref. [114] für

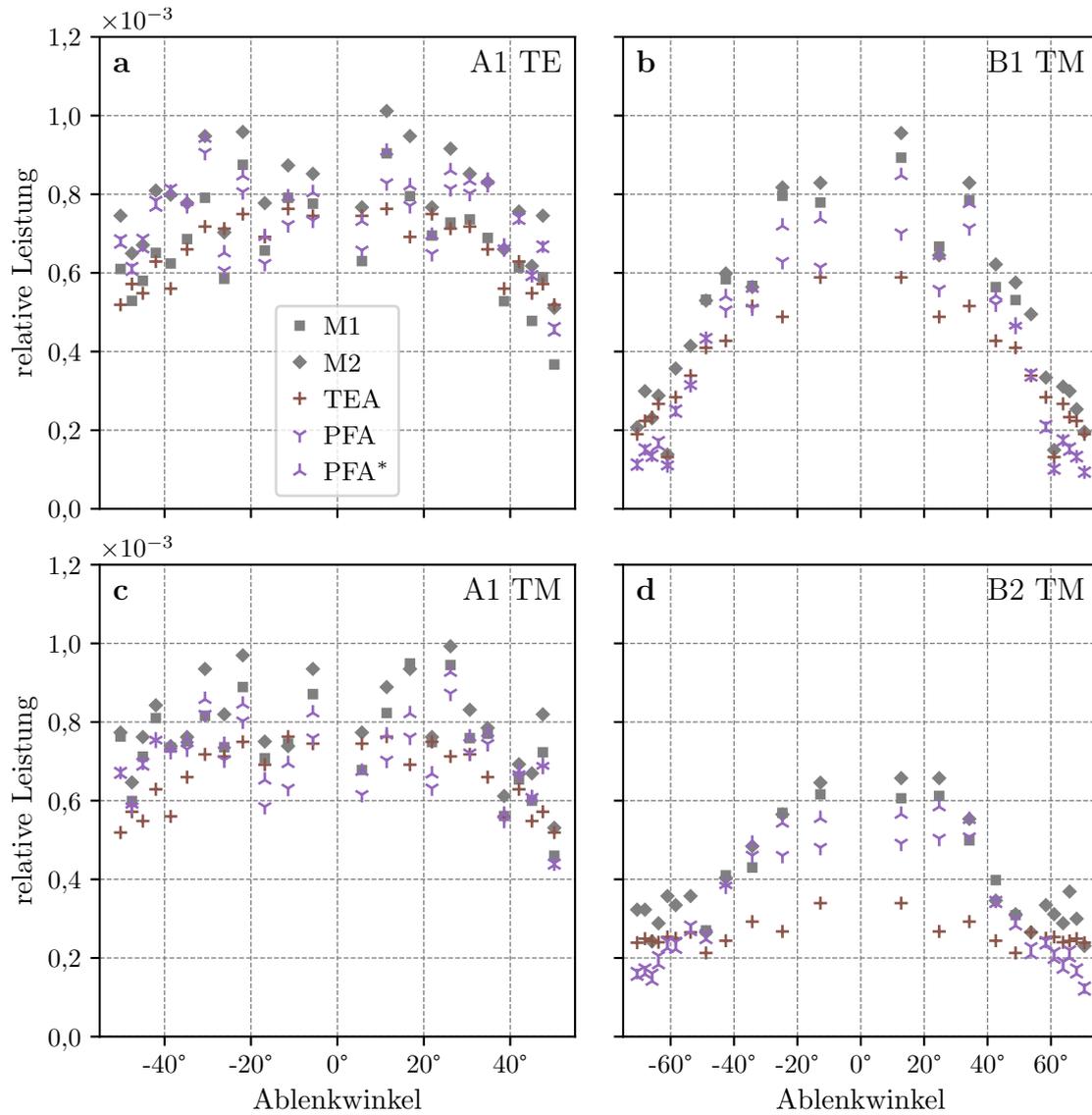


Abb. 5.8: Vergleich zwischen simulierter und berechneter Leistung im Fernfeld des Designs A1 für TE- und TM-Polarisation (a,c), sowie für TM-Polarisation von B1 und B2 (b,d). Die Simulation wurde mittels TEA und PFA für die idealen Strukturen durchgeführt, bei PFA* mit gerundeten Kanten. Für die Messungen wurden die in Abschnitt 5.2.5 beschriebenen Aufbauten M1 bzw. M2 genutzt. Die Messergebnisse haben eine Unsicherheit von ca. 5 % für M1 bzw. 10 % für M2.

Strukturgrößen unterhalb der Wellenlänge zu erwarten ist. Die Brechung der Symmetrie deckt sich allerdings mit den Resultaten der PFA. Zwischen den Messungen M1 und M2 lässt sich eine gewisse systematische Abweichung beobachten, die vermutlich aus der Messung des Referenzsignals resultiert. Im Aufbau M1 wurde dazu ein Graufilter verwendet, dessen tatsächliche Transmission vom Nennwert abweichen kann, während bei M2 dafür der Messbereich gewechselt werden musste. Andererseits können Formabweichungen Änderungen der Simulationsergebnisse bedingen, die weitestgehend einer Skalierung entsprechen. Der Fokus des Vergleichs liegt daher nicht auf der absoluten Höhe, sondern auf dem relativen Verlauf der Leistung in den Beugungsordnungen. Für

beide Polarisationsrichtungen ergibt sich insgesamt eine ähnliche Verteilung. Einzelne Beugungsordnungen zeigen dennoch Abweichungen, welche ebenso durch die PFA vorhergesagt werden können (a,c). Bei hohen Ordnungen gibt es eine Abschwächung aufgrund des Kosinus-Faktors, insbesondere für die Elemente mit großen Ablenkwinkeln (b,d). Bis zu einem Winkel von etwa 50° folgen die Messwerte dabei in guter Übereinstimmung dem erwarteten Verlauf. Bei größeren Winkeln ist die gemessene Leistung signifikant größer als mit der PFA berechnet. Der Grund für diese Abweichung liegt vermutlich in den Trainingsdaten: Die zu deren Erstellung genutzte BPM ist aufgrund der inhärenten Approximationen nur für begrenzte Ablenkwinkel hinreichend genau, vgl. Abschnitt 4.3.4. Weitere Fehlerquellen sind in der Simulation unberücksichtigte Mehrfachreflexionen innerhalb des Substrates, der Einfall weiterer Beugungsordnungen außer der zu vermessenden Nutzordnung in die Ulbricht-Kugel und zusätzliche Strukturabweichungen wie geneigte Flanken oder die Rundung der oberen und unteren Kanten der hergestellten Strukturen.

5.3 Design von CGHs mit Effektiv-Index-Strukturen

Die bisher vorgestellte Verwendung von semi-rigorosen Simulationen für das iterative Design lässt sich nicht trivial auf mehrstufige Elemente übertragen. Im Folgenden wird eine alternative Vorgehensweisen erläutert. Anschließend werden die geschilderten Varianten anhand des Designs eines Strahlteiler-CGHs verglichen.

5.3.1 Nutzung von Polynomapproximationen bei mehrstufigen Elementen

Die Anwendung der PSA zur Berechnung einer Struktur mit mehr als zwei Phasenstufen ist grundsätzlich möglich, auch wenn dafür eine größere Anzahl an Modelltermen als im binären Fall erforderlich ist. Die Projektion des resultierenden Strukturparameters, beispielsweise des Füllfaktors f , auf diskrete Werte ist hierbei allerdings deutlich störungsanfälliger, vor allem bei der Anwendung auf ein beliebiges Feld, welches im Allgemeinen keine genaue Entsprechung in einer Struktur mit diskreten Phasenstufen hat. Eine direkte Verallgemeinerung des im vorigen Abschnitt vorgestellten Ansatzes für nicht-binäre Elemente erwies sich daher als nicht zielführend. Stattdessen wurde ein Iterationsschema entwickelt, das zwar die Nutzung der PFA zur Berechnung der Felder, nicht aber die PSA beinhaltet. Dabei werden die präzise berechnete Amplitude aus der PFA-Simulation und die wie in einem herkömmlichen IFTA berechnete Phase miteinander kombiniert, wobei die Verknüpfung jeweils im Nah- und Fernfeld durchgeführt wird. Auf diese Weise profitiert man von der semi-rigorosen Simulation des Fernfeldes durch die PFA und bewahrt gleichzeitig die Konvergenz des IFTA. Ausgehend von einem Eingangsfeld $u^{(0)} = u^{\text{in}} \exp(i\phi^{(0)})$ wird das folgende Iterationsschema angewendet:

1. Projektion der Phase und Zuordnung der Struktur: $f^{(j)} = \text{LPA} \left(\mathcal{D}_\Phi \arg \frac{u^{(j)}}{u^{\text{in}}} \right)$;
2. Berechnung des Nahfeldes: $c_{j_x j_y}^{(j)} = u^{\text{in}} \cdot \text{PFA} \left(f^{(j)} \right)$;
3. Ersetzen der Amplitude: $u'^{(j)} = |c_{00}^{(j)}| \exp \left(i \arg u^{(j)} \right)$;
4. Propagation ins Fernfeld und Superposition:

$$U^{(j)} = \mathcal{F} \left\{ u'^{(j)} \right\}, C_{j_x j_y}^{(j)} = \mathcal{F} \left\{ c_{j_x j_y}^{(j)} \right\}, U'^{(j)} = \sum s_{j_x j_y} C_{j_x j_y}^{(j)}$$
;
5. Ersetzen der Amplitude mit Homogenisierung und separater Skalierung:

$$U''^{(j)} = \left(\chi_S \alpha_S^{(j)} \frac{I^{\beta_j/2}}{|U'^{(j)}|^{\beta_j}} \sqrt{I'} + \chi_A \alpha_A^{(j)} |U'^{(j)}| \right) \exp \left(i \arg U^{(j)} \right)$$
;
6. Rückpropagation ins Nahfeld: $u''^{(j)} = \mathcal{F}^{-1} \left\{ U'^{(j)} \right\}$;
7. Ersetzen der Amplitude bei diskreter Phase: $u^{(j+1)} = u^{\text{in}} \exp \left(i \mathcal{D}_{\Phi, \gamma_j} \arg \frac{u''^{(j)}}{u^{\text{in}}} \right)$.

Das Ersetzen der Amplitude in Schritt 3 bewirkt eine Gewichtung der Phasenfunktion mit der tatsächlichen Amplitude der jeweiligen Pixel. Die Schritte 5 bis 7 sind identisch zu den Schritten 2 bis 4 des Schemas in Abschnitt 5.1.1, wobei in Schritt 5 zu beachten ist, dass die Phase von U und die Amplitude von U' verwendet werden, wie auch zur Berechnung der Faktoren α_S und α_A . Die Verwendung der PFA-Amplitude entspricht dabei der Korrektur des Fernfeldes wie in Abschnitt 5.1.2, hier aber in jeder Iteration und auf der Ebene einzelner Ordnungen anstelle eines gemittelten Polynoms. Dadurch kann die Abweichung zwischen PFA und LPA im Fernfeld korrigiert werden. Im Folgenden werden die oben beschriebenen Verfahren mit dem hier vorgestellten modifizierten IFTA für diskrete Phasenstufen (ID) verglichen. Zusätzlich wird noch eine Variante ohne die Ersetzung in Schritt 3 einbezogen (ID'), stattdessen wird dort direkt $u'^{(j)} = u^{(j)}$ gesetzt.

5.3.2 Strahlteiler mit mehrstufigem Phasenprofil

Für den Vergleich der Verfahren wurden Strahlteiler-CGHs mit spiralförmigen Zielverteilungen wie in Abschnitt 4.3.4 betrachtet. Bei binären Phasenprofilen folgt die Verteilung der Nutzordnungen Gl. (4.23) mit $M = 40$, ergänzt durch die entsprechenden dazu punktsymmetrischen Ordnungen. Ansonsten wird eine nicht symmetrische Verteilung gemäß

$$\vartheta_m = \vartheta_{\text{max}} \sqrt{\frac{m - 1/2}{M}}, \quad \varphi_m = \frac{2\pi m}{\Phi_0}, \quad m = 1, \dots, M, \quad (5.13)$$

mit $M = 79$ angestrebt. Der maximale Ablenkwinkel ϑ_{max} variiert zwischen 10° und 70° . Die betrachteten Elemente bestehen jeweils aus 64×64 quadratischen Pixeln mit Seitenlängen von 300, 350, 400 oder 500 nm und sind für eine Wellenlänge von $\lambda_0 = 632,8$ nm vorgesehen. Zur Realisierung der berechneten Phasenfunktionen werden binäre Strukturen mit einer Höhe von

692 nm oder Effektiv-Index-Strukturen mit sechs Stufen und 1155 nm in Quarzglas verwendet, wobei die Realisierung durch Säulen- und Lochstrukturen verglichen wird, wie in Abschnitt 3.2.1 vorgestellt. Die Basiszellen sind jeweils so gewählt, dass sich möglichst gleichmäßige Phasenschritte ergeben. Die Füllfaktoren der Effektiv-Index-Strukturen sind in Tab. 5.2 angegeben. Zur Erstellung der PFA-Modelle wurden Trainingsdaten mithilfe der BPM berechnet. Für die Analyse der berechneten Elemente kam ausschließlich die PFA zur Anwendung, da der Fokus hier auf dem Nachweis liegt, dass die IFTA-Varianten in Kombination mit der PFA nützliche Designs erzeugen können. Sollte die Abweichung zu rigorosen Verfahren zu groß sein, kann die Genauigkeit mit besseren PFA-Modellen erhöht werden. Die Berechnung des Fernfeldes erfolgte wie in Abschnitt 4.3.4, um auch große Ablenkwinkel möglichst gut zu berücksichtigen.

Tab. 5.2: Füllfaktoren der verwendeten Effektiv-Index-Strukturen für mehrstufige Strahlteiler-CGHs. Werte von 0 oder 1 entsprechen vollständig bzw. nicht tiefgeätzten Basiszellen.

Zellgröße	Säulen	Löcher
300 nm	0; 0,45; 0,65; 0,8; 0,9; 1	0; 0,1; 0,2; 0,35; 0,55; 1
350 nm	0; 0,45; 0,6; 0,75; 0,9; 1	0; 0,1; 0,2; 0,35; 0,55; 1
400 nm	0; 0,45; 0,6; 0,75; 0,9; 1	0; 0,1; 0,2; 0,35; 0,55; 1
500 nm	0; 0,4; 0,6; 0,75; 0,9; 1	0; 0,1; 0,2; 0,3; 0,5; 1

Im Folgenden werden die im vorherigen Abschnitt beschriebenen Varianten ID bzw. ID' und die Varianten IB (Nutzung der PSA) und IH* (korrigierte Zielverteilung) aus Abschnitt 5.2.3 verglichen. Da hier eine große Bandbreite an Elementen betrachtet wird, ist es nicht sinnvoll, identische Parameter für alle Varianten festzulegen. Ziel war es dennoch, möglichst allgemeingültige Parameter mit guten Ergebnissen zu wählen. Die Gesamtzahl der Iterationen beträgt dabei 15, α_{\max} ist 0,1 und β_j steigt linear von 0 auf 3. Führten diese Parameter nicht zur Konvergenz, wurde die Anzahl der Iterationen auf 30 erhöht oder der Anstieg von β_j auf 2 bzw. 1 begrenzt. Bei den Varianten ID, ID' und IH* fand in den ersten drei Schritten keine Homogenisierung bzw. diskrete Phasenprojektion statt, um zunächst die Effizienz zu verbessern. Für jede Variation der Parameter wurden zehn Elemente mit zufälliger Startphase berechnet und jeweils die höchste erreichte Homogenität bzw. das höchste SNR für den Vergleich verwendet.

In Abb. 5.9 sind zunächst die Ergebnisse für das Design von binären Elementen dargestellt. Die verschiedenen Berechnungsvarianten liefern dabei häufig ähnlich gute Resultate in Bezug auf die Homogenität und das SNR. Beide Bewertungskriterien zeigen im Allgemeinen gleichartige Verläufe. Insbesondere kann mit den Varianten ID und ID', die eine Fernfeldberechnung in jedem Schritt beinhalten, hier keine wesentliche Verbesserung gegenüber IH* mit der einmaligen Vorkorrektur der Zielverteilung beobachtet werden. Bei Pixelgrößen von 300 bzw. 500 nm liefert IB etwas bessere Resultate als die übrigen Verfahren. Für große Ablenkwinkel führen ID und IB mitunter

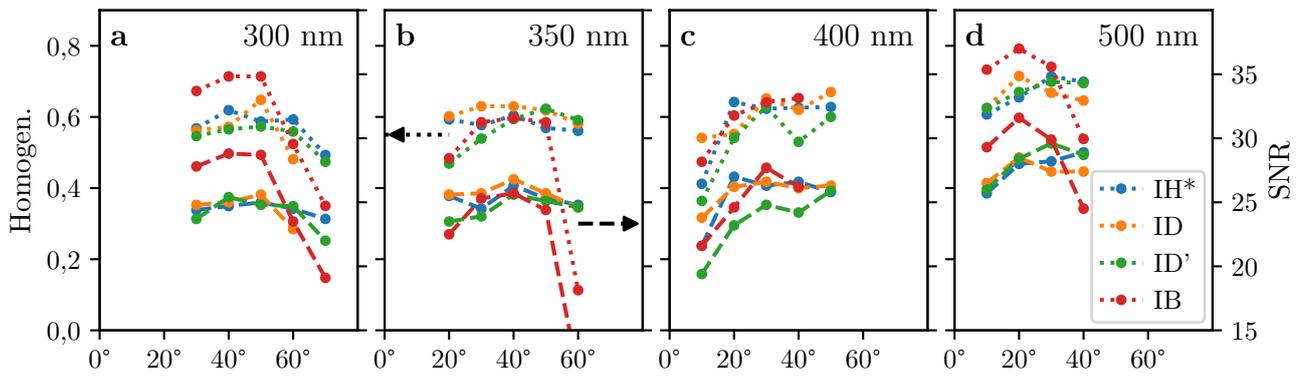


Abb. 5.9: Vergleich der simulierten Fernfelder von binären Strahlteiler-CGHs mit Pixelgrößen von 300 (a), 350 (b), 400 (c) und 500 nm (d). Gezeigt wird jeweils die Homogenität (gepunktet, linke Skala) und das SNR (gestrichelt, rechte Skala). Resultate ohne Konvergenz der Verfahren wurden ausgelassen.

nicht zu Ergebnissen mit hoher Güte, während die weniger abgewandelten Algorithmen IH* und ID' verlässlicher konvergieren. Allgemein werden die Ergebnisse mit zunehmender Pixelgröße besser, vor allem weil damit der Einfluss der umgebenden Pixel abnimmt. Alle Varianten können die bei kleinen Winkeln erzielbaren Ergebnisse auch bei hohen Ablenkwinkeln annähernd erhalten.

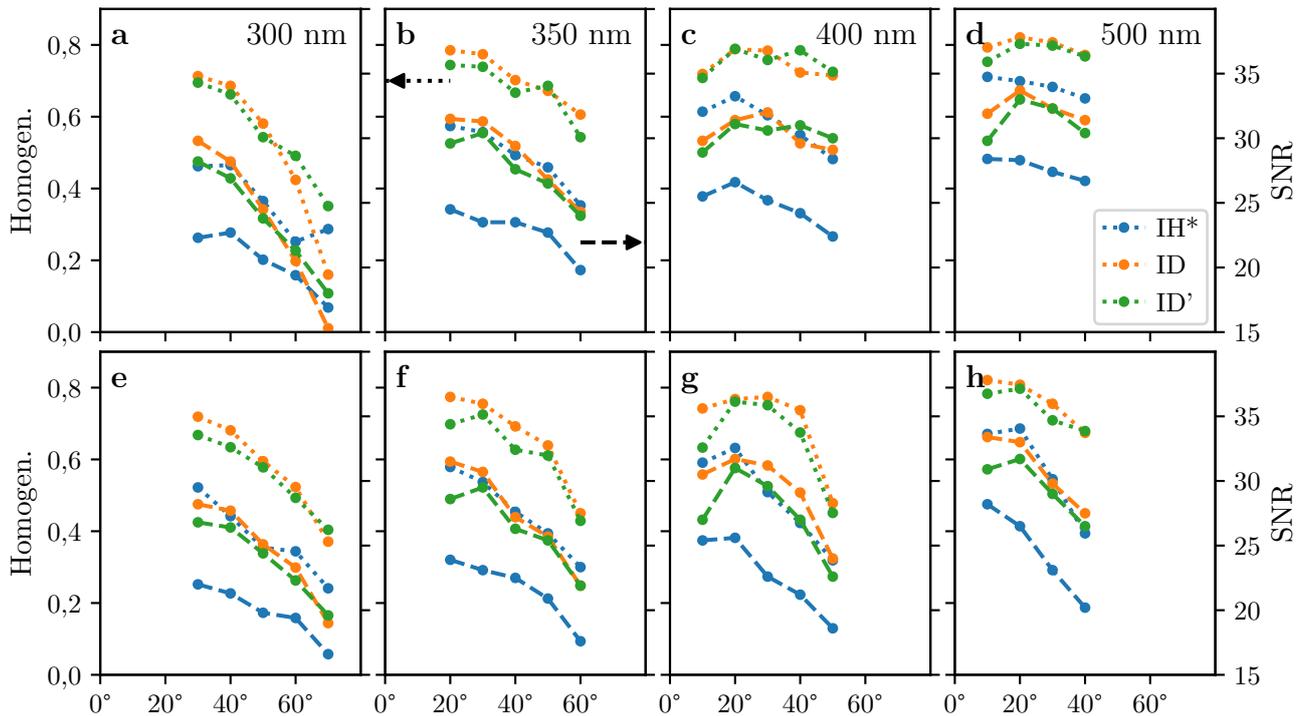


Abb. 5.10: Vergleich der simulierten Fernfelder von Strahlteiler-CGHs aus Effektiv-Index-Strukturen basierend auf Säulen (a-d) bzw. Löchern (e-h) mit Pixelgrößen von 300 (a,e), 350 (b,f), 400 (c,g) und 500 nm (d,h). Gezeigt wird jeweils die Homogenität (gepunktet, linke Skala) und das SNR (gestrichelt, rechte Skala).

Für die Effektiv-Index-Elemente steht die Variante IB nicht zur Verfügung. Die Varianten ID

und ID' erzielen weitestgehend ähnlich gute Resultate, siehe Abb. 5.10. Lediglich bei geringen Ablenkwinkeln ist ID regelmäßig vorteilhaft. Die Variante IH* führt dagegen durchgängig zu Elementen mit geringerer Homogenität. Vor allem die Berechnung des Fernfeldes mittels der PFA in jedem Iterationsschritt scheint hier zu einer besseren Übereinstimmung mit der Zielverteilung zu führen. Das spricht dafür, dass die Nachbarschaftseffekte für die Felder hinter individuellen Pixeln einen deutlichen Einfluss auf das Fernfeld haben. Wie im binären Fall werden die Ergebnisse mit zunehmender Pixelgröße besser. Für große Ablenkwinkel verschlechtert sich allerdings generell die Leistungsfähigkeit der Algorithmen. Mit Säulen (a-d) lassen sich für die untersuchten Elemente bessere Resultate als mit Löchern (e-h) erzielen, was bei Letzteren für einen stärkeren Einfluss der Pixelumgebung spricht.

Zusammenfassend lässt sich feststellen, dass bei Verwendung eines IFTA ohne physikalische Fernfeldberechnung die Vorkorrektur der Zielverteilung wie bei IH* zur Erzielung einer hohen Übereinstimmung in der Regel sinnvoll ist. Systematisch bessere Ergebnisse ließen sich bei symmetrischen, binären Designs mit der Variante IB erzielen. Für Effektiv-Index-Strukturen führten dagegen die Varianten ID oder ID' zu einer deutlichen Verbesserung.

6 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurden verschiedene neue Verfahren zur Simulation von Subwellenlängenstrukturen vorgestellt und Einsatzmöglichkeiten für die Erzeugung beliebiger Phasenfunktionen aufgezeigt. Mithilfe der angepassten BPM können allgemeine Mikrostrukturen simuliert werden. Dabei wird eine hohe Übereinstimmung mit rigorosen Verfahren erreicht. Die unterschiedliche Ausbreitung verschiedener Polarisationskomponenten wird berücksichtigt, deren Kopplung allerdings nicht. Die Genauigkeit der Methode nimmt mit zunehmendem Indexkontrast der Materialien und größeren Propagationswinkeln ab. Allerdings lassen sich für Glasoberflächen an Luft bei moderaten Winkeln verlässliche Ergebnisse produzieren. Das Verfahren ist um einige Größenordnungen schneller als rigorose Methoden und bietet daher die Möglichkeit zur Untersuchung deutlich größerer Simulationsgebiete, wie sie für nicht-periodische Strukturen benötigt werden [45]. Beispielsweise können damit Streulichtsimulationen mit einer durch rigorose Verfahren nicht zugänglichen Auflösung durchgeführt werden.

Für regelmäßig angeordnete Mikrostrukturen kann das optische Verhalten in Nah- und Fernfeld mittels der hier erstmals vorgestellten PFA noch effizienter vorhergesagt werden. Es genügt die physikalische Simulation eines begrenzten Trainingsdatensatzes, um ein Modell zu erstellen, das beliebige Anordnungen der verwendeten Strukturen simulieren kann. Die erzielbare Genauigkeit reicht dabei dicht an das zur Simulation der Trainingsdaten verwendete Verfahren heran. Die Qualität der Ergebnisse wird vor allem durch das Verhältnis von Strukturgrößen und Wellenlänge beeinflusst. Insbesondere für Effektiv-Index-Strukturen, die etwas kleiner als die Wellenlänge sind, lassen sich gute Resultate erzielen. Für das vorgestellte Verfahren müssen die auftretenden Geometrien durch einen Parameter beschreibbar sein, wobei aber auch Kombinationen aus Säulen- und Lochstrukturen eingeschlossen ist. Die Anzahl der Modellparameter ermöglicht eine Abwägung zwischen Geschwindigkeit und Güte der Ergebnisse. Im Allgemeinen erlaubt das Verfahren aber nochmals um mehrere Größenordnungen schnellere Berechnungen als mit der BPM, sodass es hervorragend für Optimierungsaufgaben geeignet ist, welche die wiederholte Simulation ähnlicher Elemente benötigen. Die Berechnung der Trainingsdaten stellt einen einmaligen Aufwand dar und ist zudem leicht parallelisierbar. In der vorgestellten Formulierung kann allerdings nur ein Satz von Strukturen bei festen Parametern wie Ätztiefe oder Gitterabstand verwendet werden. Eine denkbare Erweiterung wäre die Erstellung einer Gruppe von Modellen z. B. für verschiedene Wellenlängen und die anschließende Interpolation der Modellkoeffizienten, um das

dispersive Verhalten simulieren zu können, beispielsweise von mehrfarbigen CGHs [115] oder Metalinsen [10, 11]. Außerdem könnten Strukturen basierend auf mehreren Strukturparametern, wie rechteckige Säulen, untersucht werden. Hierfür ist wiederum eine geeignete Auswahl der genutzten Modellparameter notwendig. Eventuell kann die Berechnung durch neuronale Netze effizienter gestaltet werden als durch einfache lineare Regression.

Für den Design-Prozess von CGHs wurde mit der PSA ein Verfahren vorgestellt, welches die Lösung des inversen Problems, der Berechnung einer Struktur zur Erzeugung einer gegebenen Feldverteilung, ermöglicht. Mit binären Strukturen ließen sich verlässliche Ergebnisse erzeugen, während eine Verallgemeinerung auf z. B. Effektiv-Index-Strukturen nicht einfach möglich ist, vor allem, weil ein beliebiges Feld im Allgemeinen keine direkte Entsprechung unter den verfügbaren Strukturen besitzt. Das Design von CGHs kann verbessert werden, indem zunächst einige Elemente mit etablierten Verfahren berechnet und anschließend z. B. mit der PFA simuliert werden. Die beobachtete systematische Abweichung von der Zielverteilung kann in einer weiteren Berechnung korrigiert werden. Außerdem ist es möglich, die PFA direkt in die Iterationen des IFTA einzubeziehen, für binäre Elemente auch in Kombination mit der PSA. Auf diese Weise lässt sich vor allem für Effektiv-Index-CGHs nochmals eine Verbesserung gegenüber dem vorkorrigierten Design beobachten. Die Algorithmen wurden anhand von Strahlteilern demonstriert, sind aber für allgemeine Zielverteilungen anwendbar. Ein binärer Strahlteiler wurde hergestellt und experimentell untersucht. Dabei bestätigte sich, dass die TEA zur präzisen Vorhersage der optischen Funktion bei großen Ablenkwinkeln ungeeignet ist, während die PFA zumindest bei moderaten Winkeln bis etwa 50° eine gute Übereinstimmung zeigte, die sich noch erhöhen ließ, wenn die Verrundung der Kanten berücksichtigt wurde. Mithilfe von vektoriellen Trainingsdaten, z. B. auf Basis der FDTD, können auch größere Ablenkwinkel durch die PFA verlässlich berechnet werden.

Die präsentierten Verfahren können zur Optimierung einer Vielzahl von DOEs eingesetzt werden, beispielsweise für Metalinsen oder Strahlteiler mit hoher NA. Insbesondere das Design von Punktmustergeneratoren zur dreidimensionalen Vermessung von Objekten [116], Gesichtserkennung oder Kamerakalibrierung sind vielversprechende Anwendungsgebiete. Dabei ist eine präzise Wiedergabe der Zielverteilung notwendig, um die korrekte Erfassung der Messpunkte sicherzustellen. Schnelle und genaue Designverfahren sind auch für die holografische Erzeugung von Videosequenzen [117] sowie im Allgemeinen für die Generierung beliebiger Wellenfronten nützlich.

Literaturverzeichnis

- [1] N. Yu und F. Capasso, “Flat optics with designer metasurfaces”, *Nat. Mater.* **13**, 139–150 (2014).
- [2] A. Arbabi und A. Faraon, “Advances in optical metalenses”, *Nat. Photonics* **17**, 16–25 (2023).
- [3] P. Genevet, F. Capasso, F. Aieta, M. Khorasaninejad und R. Devlin, “Recent advances in planar optics: from plasmonic to dielectric metasurfaces”, *Optica* **4**(1), 139–152 (2017).
- [4] J. Engelberg und U. Levy, “The advantages of metalenses over diffractive lenses”, *Nat. Commun.* **11**, 1991 (2020).
- [5] P. Lalanne und D. Lemerrier-Lalanne, “On the effective medium theory of subwavelength periodic structures”, *J. Mod. Opt.* **43**(10), 2063–2085 (1996).
- [6] F. T. Chen und H. G. Craighead, “Diffractive lens fabricated with mostly zeroth-order gratings”, *Opt. Lett.* **21**(3), 177–179 (1996).
- [7] A. V. Kildishev, A. Boltasseva und V. M. Shalaev, “Planar photonics with metasurfaces”, *Science* **339**(6125), 1232009 (2013).
- [8] C.-Y. Fan, C.-P. Lin und G.-D. J. Su, “Ultrawide-angle and high-efficiency metalens in hexagonal arrangement”, *Sci. Reports* **10**, 15677 (2020).
- [9] L. Chen, Y. Hao, L. Zhao, R. Wu, Y. Liu, Z. Wei, N. Xu, Z. Li und H. Liu, “Multifunctional metalens generation using bilayer all-dielectric metasurfaces”, *Opt. Express* **29**(6), 9332–9345 (2021).
- [10] Z. Li, P. Lin, Y.-W. Huang, J.-S. Park, W. T. Chen, Z. Shi, C.-W. Qiu, J.-X. Cheng und F. Capasso, “Meta-optics achieves rgb-achromatic focusing for virtual reality”, *Sci. Adv.* **7** (5), eabe4458 (2021).
- [11] X. Lu, Y. Guo, M. Pu, Y. Zhang, Z. Li, X. Li, X. Ma und X. Luo, “Broadband achromatic metasurfaces for sub-diffraction focusing in the visible”, *Opt. Express* **29**(4), 5947–5958 (2021).

- [12] B. R. Brown und A. W. Lohmann, “Computer-generated binary holograms”, *IBM J. Res. Dev.* **13**(2), 160–168 (1969).
- [13] E. N. Leith und J. Upatnieks, “Wavefront reconstruction with diffused illumination and three-dimensional objects”, *J. Opt. Soc. Am.* **54**(11), 1295–1301 (1964).
- [14] W. Chen, “Computer-generated hologram using binary phase with an aperture”, *Appl. Opt.* **56**(32), 9126–9131 (2017).
- [15] G. Ruffato, R. Rossi, M. Massari, E. Mafakheri, P. Capaldo und F. Romanato, “Design, fabrication and characterization of computer generated holograms for anti-counterfeiting applications using OAM beams as light decoders”, *Sci. Reports* **7**, 18011 (2017).
- [16] S. J. Byrnes, A. Lenef, F. Aieta und F. Capasso, “Designing large, high-efficiency, high-numerical-aperture, transmissive meta-lenses for visible light”, *Opt. Express* **24**(5), 5110–5124 (2016).
- [17] Z. Zhang, “Microsoft kinect sensor and its effect”, *IEEE MultiMedia* **19**(2), 4–10 (2012).
- [18] J. M. Miller, M. R. Taghizadeh, J. Turunen und N. Ross, “Multilevel-grating array generators: fabrication error analysis and experiments”, *Appl. Opt.* **32**(14), 2519–2525 (1993).
- [19] W. Yu, K. Takahara, T. Konishi, T. Yotsuya und Y. Ichioka, “Fabrication of multilevel phase computer-generated hologram elements based on effective medium theory”, *Appl. Opt.* **39**(20), 3531–3536 (2000).
- [20] W. Eckstein, “Computergenerierte Hologramme auf Basis binärer Subwellenlängenstrukturen”, Dissertation, Friedrich-Schiller-Universität Jena (2015).
- [21] V. Popov, M. Yakovleva, F. Boust, J.-L. Pelouard, F. Pardo und S. N. Burokur, “Designing metagratings via local periodic approximation: From microwaves to infrared”, *Phys. Rev. Appl.* **11**, 044054 (2019).
- [22] M. E. Testorf und M. A. Fiddy, “Efficient optimization of diffractive optical elements based on rigorous diffraction models”, *J. Opt. Soc. Am. A* **18**(11), 2908–2914 (2001).
- [23] Z. Lin und S. G. Johnson, “Overlapping domains for topology optimization of large-area metasurfaces”, *Opt. Express* **27**(22), 32445–32453 (2019).
- [24] T. Phan, D. Sell, E. W. Wang, S. Doshay, K. Edee, J. Yang und J. A. Fan, “High-efficiency, large-area, topology-optimized metasurfaces”, *Light. Sci. & Appl.* **8**, 48 (2019).

- [25] S. An, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, Y. Dong, M. Haerinia, A. M. Agarwal, C. Rivero-Baleine, M. Kang, K. A. Richardson, T. Gu, J. Hu, C. Fowler und H. Zhang, “Deep convolutional neural networks to predict mutual coupling effects in metasurfaces”, *Adv. Opt. Mater.* **10**(3), 2102113 (2022).
- [26] M. V. Zhelyeznyakov, S. Brunton und A. Majumdar, “Deep learning to accelerate scatterer-to-field mapping for inverse design of dielectric metasurfaces”, *ACS Photonics* **8**(2), 481–488 (2021).
- [27] M. Born, E. Wolf, A. Bhatia, P. Clemmow, D. Gabor, A. Stokes, A. Taylor, P. Wayman und W. Wilcock, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Cambridge University, 1999).
- [28] E. Hecht, *Optik* (Oldenbourg Wissenschaftsverlag, 2005).
- [29] J. N. Mait, D. W. Prather und M. S. Mirotznik, “Design of binary subwavelength diffractive lenses by use of zeroth-order effective-medium theory”, *J. Opt. Soc. Am. A* **16**(5), 1157–1167 (1999).
- [30] K. Yee, “Numerical solution of initial boundary value problems involving maxwell’s equations in isotropic media”, *IEEE Transactions on Antennas Propag.* **14**(3), 302–307 (1966).
- [31] M. G. Moharam und T. K. Gaylord, “Rigorous coupled-wave analysis of planar-grating diffraction”, *J. Opt. Soc. Am.* **71**(7), 811–818 (1981).
- [32] M. G. Moharam, E. B. Grann, D. A. Pommet und T. K. Gaylord, “Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings”, *J. Opt. Soc. Am. A* **12**(5), 1068–1076 (1995).
- [33] L. Li, “Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings”, *J. Opt. Soc. Am. A* **13**(5), 1024–1035 (1996).
- [34] Z.-Y. Li und L.-L. Lin, “Photonic band structures solved by a plane-wave-based transfer-matrix method”, *Phys. Rev. E* **67**, 046607 (2003).
- [35] W. Huang und C. Xu, “Simulation of three-dimensional optical waveguides by a full-vector beam propagation method”, *IEEE J. Quantum Electron.* **29**(10), 2639–2649 (1993).
- [36] M. D. Feit und J. A. Fleck, “Light propagation in graded-index optical fibers”, *Appl. Opt.* **17**(24), 3990–3998 (1978).
- [37] R. Scarmozzino, A. Gopinath, R. Pregla und S. Helfert, “Numerical techniques for modeling guided-wave photonic devices”, *IEEE J. Sel. Top. Quantum Electron.* **6**(1), 150–162 (2000).

- [38] K.-H. Brenner und W. Singer, “Light propagation through microlenses: a new simulation method”, *Appl. Opt.* **32**(26), 4984–4988 (1993).
- [39] M. Fertig und K.-H. Brenner, “Vector wave propagation method”, *J. Opt. Soc. Am. A* **27**(4), 709–717 (2010).
- [40] K.-H. Brenner, “A high-speed version of the wave propagation method applied to micro-optics”, in *2017 16th Workshop on Information Optics (WIO)*, (2017), S. 1–3.
- [41] M. Wende, J. Drozella, A. Toulouse und A. M. Herkommer, “Fast algorithm for the simulation of 3d-printed microoptics based on the vector wave propagation method”, *Opt. Express* **30**(22), 40161–40173 (2022).
- [42] M. Khorasaninejad, W. T. Chen, R. C. Devlin, J. Oh, A. Y. Zhu und F. Capasso, “Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging”, *Science* **352**(6290), 1190–1194 (2016).
- [43] J. Borne, D. Panneton, M. Piché und S. Thibault, “Semi-analytical model of the optical properties of a metasurface composed of nanofins”, *J. Opt. Soc. Am. B* **38**(11), 3509–3514 (2021).
- [44] S. Fasold, S. Linß, T. Kawde, M. Falkner, M. Decker, T. Pertsch und I. Staude, “Disorder-enabled pure chirality in bilayer plasmonic metasurfaces”, *ACS Photonics* **5**(5), 1773–1778 (2018).
- [45] D. Arslan, A. Rahimzadegan, S. Fasold, M. Falkner, W. Zhou, M. Kroychuk, C. Rockstuhl, T. Pertsch und I. Staude, “Toward perfect optical diffusers: Dielectric huygens’ metasurfaces with critical positional disorder”, *Adv. Mater.* **34**(5), 2105868 (2022).
- [46] J. W. Cooley und J. W. Tukey, “An algorithm for the machine calculation of complex fourier series”, *Math. Comput.* **19**, 297–301 (1965).
- [47] Z. He, X. Sui, G. Jin, D. Chu und L. Cao, “Optimal quantization for amplitude and phase in computer-generated holography”, *Opt. Express* **29**(1), 119–133 (2021).
- [48] L. Huang, S. Zhang und T. Zentgraf, “Metasurface holography: from fundamentals to applications”, *Nanophotonics* **7**(6), 1169–1190 (2018).
- [49] Q. Jiang, G. Jin und L. Cao, “When metasurface meets hologram: principle and advances”, *Adv. Opt. Photon.* **11**(3), 518–576 (2019).
- [50] T. Kämpfe, E.-B. Kley und A. Tünnermann, “Designing multiplane computer-generated holograms with consideration of the pixel shape and the illumination wave”, *J. Opt. Soc. Am. A* **25**(7), 1609–1622 (2008).

- [51] J. Amako und Y. Shinozaki, “Design and fabrication of a diffractive beam splitter for dual-wavelength and concurrent irradiation of process points”, *Opt. Express* **24**(14), 16111–16122 (2016).
- [52] W. Freese, T. Kämpfe, E.-B. Kley und A. Tünnermann, “Design of binary subwavelength multiphase level computer generated holograms”, *Opt. Lett.* **35**(5), 676–678 (2010).
- [53] D. Wen, F. Yue, G. Li, G. Zheng, K. Chan, S. Chen, M. Chen, K. F. Li, P. W. H. Wong, K. W. Cheah, E. Yue Bun Pun, S. Zhang und X. Chen, “Helicity multiplexed broadband metasurface holograms”, *Nat. Commun.* **6**, 8241 (2015).
- [54] J. P. Balthasar Mueller, N. A. Rubin, R. C. Devlin, B. Groever und F. Capasso, “Metasurface polarization optics: Independent phase control of arbitrary orthogonal states of polarization”, *Phys. Rev. Lett.* **118**, 113901 (2017).
- [55] H. Zhou, B. Sain, Y. Wang, C. Schlickriede, R. Zhao, X. Zhang, Q. Wei, X. Li, L. Huang und T. Zentgraf, “Polarization-encrypted orbital angular momentum multiplexed metasurface holography”, *ACS Nano* **14**(5), 5553–5559 (2020).
- [56] D. C. Kim, A. Hermerschmidt, P. Dyachenko und T. Scharf, “Inverse design and demonstration of high-performance wide-angle diffractive optical elements”, *Opt. Express* **28**(15), 22321–22333 (2020).
- [57] Y. Ni, S. Chen, Y. Wang, Q. Tan, S. Xiao und Y. Yang, “Metasurface for structured light projection over 120° field of view”, *Nano Lett.* **20**(9), 6719–6724 (2020).
- [58] M. M. R. Elsayy, A. Gourdin, M. Binois, R. Duvigneau, D. Felbacq, S. Khadir, P. Genevet und S. Lanteri, “Multiobjective statistical learning optimization of rgb metalens”, *ACS Photonics* **8**(8), 2498–2508 (2021).
- [59] I.-L. Ho, “K-series approximation of vectorial optical fields for designing diffractive optical elements with subwavelength feature sizes”, *Opt. Express* **29**(23), 37211–37224 (2021).
- [60] J. Li, F. Zhang, M. Pu, Y. Guo, X. Li, X. Ma, C. Wang und X. Luo, “Quasi-continuous metasurface beam splitters enabled by vector iterative fourier transform algorithm”, *Materials* **14**(4) (2021).
- [61] W. Eckstein, E.-B. Kley und A. Tünnermann, “Comparison of different simulation methods for effective medium computer generated holograms”, *Opt. Express* **21**(10), 12424–12433 (2013).
- [62] H. Hao, Z. Tingting, S. Qiang und Y. Xiaodong, “Wide angle 2D beam splitter design based on vector diffraction theory”, *Opt. Commun.* **434**, 28–35 (2019).

- [63] K. Donda und R. Hegde, “Optimal design of beam-deflectors using extended unit-cell metagratings”, *Prog. In Electromagn. Res. M* **77**, 83–92 (2019).
- [64] S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. Hu und H. Zhang, “A deep learning approach for objective-driven all-dielectric metasurface design”, *ACS Photonics* **6** (12), 3196–3207 (2019).
- [65] S. An, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, M. Kang, K. A. Richardson, T. Gu, J. Hu, C. Fowler und H. Zhang, “Deep learning modeling approach for metasurfaces with high degrees of freedom”, *Opt. Express* **28**(21), 31932–31942 (2020).
- [66] F. Yang, S. An, M. Y. Shalaginov, H. Zhang, C. Rivero-Baleine, J. Hu und T. Gu, “Design of broadband and wide-field-of-view metalenses”, *Opt. Lett.* **46**(22), 5735–5738 (2021).
- [67] D. Gabor, “A new microscopic principle”, *Nature* **161**(4098), 777–778 (1948).
- [68] A. W. Lohmann und D. P. Paris, “Binary fraunhofer holograms, generated by computer”, *Appl. Opt.* **6**(10), 1739–1748 (1967).
- [69] S. Park, J. Lee, S. Lim, M. Kim, S. Ahn, S. Hwang, S. Jeon, J. Jeong, J. Hahn und H. Kim, “Wide-viewing full-color depthmap computer-generated holograms”, *Opt. Express* **29**(17), 26793–26807 (2021).
- [70] A. J. MacGovern und J. C. Wyant, “Computer generated holograms for testing optical elements”, *Appl. Opt.* **10**(3), 619–624 (1971).
- [71] J. Burge, “Applications of computer-generated holograms for interferometric measurement of large aspheric optics”, *Proc. SPIE - The Int. Soc. for Opt. Eng.* **2576** (1995).
- [72] R. W. Gerchberg und W. O. Saxton, “A practical algorithm for the determination of phase from image and diffraction plane pictures”, *Optik* **35**, 237–246 (1972).
- [73] S. Rothe, P. Daferner, S. Heide, D. Krause, F. Schmieder, N. Koukourakis und J. W. Czarske, “Benchmarking analysis of computer generated holograms for complex wavefront shaping using pixelated phase modulators”, *Opt. Express* **29**(23), 37602–37616 (2021).
- [74] J. R. Fienup, “Phase retrieval algorithms: a comparison”, *Appl. Opt.* **21**(15), 2758–2769 (1982).
- [75] V. Kotlyar, P. Seraphimovich und V. Soifer, “An iterative algorithm for designing diffractive optical elements with regularization”, *Opt. Lasers Eng.* **29**(4), 261–268 (1998).

- [76] M. Pasienski und B. DeMarco, “A high-accuracy algorithm for designing arbitrary holographic atom traps”, *Opt. Express* **16**(3), 2176–2190 (2008).
- [77] K.-H. Brenner, “Method for designing arbitrary two-dimensional continuous phase elements”, *Opt. Lett.* **25**(1), 31–33 (2000).
- [78] F. Wyrowski, “Diffractive optical elements: iterative calculation of quantized, blazed phase structures”, *J. Opt. Soc. Am. A* **7**(6), 961–969 (1990).
- [79] R. D. Leonardo, F. Ianni und G. Ruocco, “Computer generation of optimal holograms for optical trap arrays”, *Opt. Express* **15**(4), 1913–1922 (2007).
- [80] J. S. Liu und M. R. Taghizadeh, “Iterative algorithm for the design of diffractive phase elements for laser beam shaping”, *Opt. Lett.* **27**(16), 1463–1465 (2002).
- [81] S. Linss, D. Michaelis und U. D. Zeitner, “Macroscopic wave-optical simulation of dielectric metasurfaces”, *Opt. Express* **29**(7), 10879–10892 (2021).
- [82] G. R. Hadley, “Wide-angle beam propagation using padé approximant operators”, *Opt. Lett.* **17**(20), 1426–1428 (1992).
- [83] G. R. Hadley, “Multistep method for wide-angle beam propagation”, *Opt. Lett.* **17**(24), 1743–1745 (1992).
- [84] P. L. Ho und Y. Y. Lu, “A stable bidirectional propagation method based on scattering operators”, *IEEE Photonics Technol. Lett.* **13**(12), 1316–1318 (2001).
- [85] E. V. Bekker, P. Sewell, T. M. Benson und A. Vukovic, “Wide-angle alternating-direction implicit finite-difference beam propagation method”, *J. Light. Technol.* **27**(14), 2595–2604 (2009).
- [86] L. H. Thomas, “Elliptic problems in linear difference equations over a network”, *Watson Sci. Comput. Lab. Rep.*, Columbia University, New York (1949).
- [87] J. Yamauchi, T. Ando und H. Nakano, “Beam-propagation analysis of optical fibres by alternating direction implicit method”, *Electron. Lett.* **27**, 1663–1665 (1991).
- [88] J. Shibayama, T. Takahashi, J. Yamauchi und H. Nakano, “A three-dimensional horizontally wide-angle noniterative beam-propagation method based on the alternating-direction implicit scheme”, *IEEE Photonics Technol. Lett.* **18**(5), 661–663 (2006).
- [89] F. Bloch, “Über die Quantenmechanik der Elektronen in Kristallgittern”, *Zeitschrift für Physik* **52**, 555–600 (1929).

- [90] J. Sherman und W. J. Morrison, “Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix”, *The Annals Math. Stat.* **21**(1), 124–127 (1950).
- [91] Y.-P. Chiou und H.-C. Chang, “Analysis of optical waveguide discontinuities using the pade approximants”, *IEEE Photonics Technol. Lett.* **9**(7), 964–966 (1997).
- [92] H. El-Refaei, I. Betty und D. Yevick, “The application of complex pade approximants to reflection at optical waveguide facets”, *IEEE Photonics Technol. Lett.* **12**(2), 158–160 (2000).
- [93] H. El-Refaei, D. Yevick und I. Betty, “Stable and noniterative bidirectional beam propagation method”, *IEEE Photonics Technol. Lett.* **12**(4), 389–391 (2000).
- [94] R. Petruskevicius, “BiBPM modeling of slow wave structures”, *Opt. Quantum Electron.* **39**(4), 407–418 (2007).
- [95] H. Rao, R. Scarmozzino und R. Osgood, “A bidirectional beam propagation method for multiple dielectric interfaces”, *IEEE Photonics Technol. Lett.* **11**(7), 830–832 (1999).
- [96] A. Locatelli, D. Modotto, C. D. Angelis, F. M. Pigozzo und A.-D. Capobianco, “Nonlinear bidirectional beam propagation method based on scattering operators for periodic microstructured waveguides”, *J. Opt. Soc. Am. B* **20**(8), 1724–1731 (2003).
- [97] P. Lalanne, S. Astilean, P. Chavel, E. Cambril und H. Launois, “Design and fabrication of blazed binary diffractive elements with sampling periods smaller than the structural cutoff”, *J. Opt. Soc. Am. A* **16**(5), 1143–1156 (1999).
- [98] J. Yamauchi, K. Sumida und H. Nakano, “Analysis of a polarization splitter with a multilayer filter using a Padé-operator-based power-conserving fourth-order accurate beam-propagation method”, *IEEE Photonics Technol. Lett.* **18**(17), 1858–1860 (2006).
- [99] A. Pors, O. Albrektsen, I. P. Radko und S. I. Bozhevolnyi, “Gap plasmon-based metasurfaces for total control of reflected light”, *Sci. Reports* **3**, 2155 (2013).
- [100] A. Arbabi, Y. Horie, M. Bagheri und A. Faraon, “Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission”, *Nat. Nanotechnol.* **10**, 937–943 (2015).
- [101] F. Cheng, L. Ding, L. Qiu, D. Nikolov, A. Bauer, J. P. Rolland und A. N. Vamivakas, “Polarization-switchable holograms based on efficient, broadband multifunctional metasurfaces in the visible regime”, *Opt. Express* **26**(23), 30678–30688 (2018).

- [102] T. Siefke, M. Heusinger, C. B. R. Hurtado, J. Dickmann, U. Zeitner, A. Tünnermann und S. Kroker, “Line-edge roughness as a challenge for high-performance wire grid polarizers in the far ultraviolet and beyond”, *Opt. Express* **26**(15), 19534–19547 (2018).
- [103] M. Heusinger, D. Michaelis, T. Flügel-Paul und U. D. Zeitner, “Diffuse scattering due to stochastic disturbances of 1D-gratings on the example of line edge roughness”, *Opt. Express* **26**(21), 28104–28118 (2018).
- [104] M. Heusinger, “Untersuchungen zu deterministischem und stochastischem Streulicht in hocheffizienten binären Beugungsgittern”, Dissertation, Friedrich-Schiller-Universität Jena (2019).
- [105] S. Linss, D. Michaelis und U. D. Zeitner, “Ultrafast farfield simulation of non-paraxial computer generated holograms”, *Opt. Express* **30**(8), 13765–13775 (2022).
- [106] G. B. Arfken, H. J. Weber und F. E. Harris, *Mathematical Methods for Physicists (Seventh Edition)* (Academic Press, Boston, 2013).
- [107] F. Pokropp, *Lineare Regression und Varianzanalyse* (Oldenbourg Wissenschaftsverlag, Berlin, Boston, 1994).
- [108] A. Arbabi, Y. Horie, A. J. Ball, M. Bagheri und A. Faraon, “Subwavelength-thick lenses with high numerical apertures and large efficiency based on high-contrast transmitarrays”, *Nat. Commun.* **6**, 7069 (2015).
- [109] H. Kwon, E. Arbabi, S. M. Kamali, M. Faraji-Dana und A. Faraon, “Computational complex optical field imaging using a designed metasurface diffuser”, *Optica* **5**(8), 924–931 (2018).
- [110] M. Faraji-Dana, E. Arbabi, A. Arbabi, S. M. Kamali, H. Kwon und A. Faraon, “Folded planar metasurface spectrometer”, in *Conference on Lasers and Electro-Optics*, (Optica Publishing Group, 2018), S. AF3M.3.
- [111] S. Linß und U. D. Zeitner, “Design of computer-generated holograms based on semi-rigorous simulations of sub-wavelength structures”, *Opt. Express* **31**(7), 11089–11101 (2023).
- [112] O. Ripoll, V. Kettunen und H. P. Herzig, “Review of iterative Fourier-transform algorithms for beam shaping applications”, *Opt. Eng.* **43**(11), 2549 – 2548 (2004).
- [113] U. D. Zeitner und E.-B. Kley, “Advanced lithography for micro-optics”, in *Laser Beam Shaping VII*, Band 6290 F. M. Dickey und D. L. Shealy, Hrsg., International Society for Optics and Photonics (SPIE, 2006), S. 629009.

- [114] G.-N. Nguyen, K. Heggarty, J. L. Meur, A. Bacher und P. Meyrueis, “Diffraction symmetry of binary fourier elements with feature sizes on the order of the illumination wavelength and effects of fabrication errors”, *Opt. Lett.* **42**(24), 5178–5181 (2017).
- [115] O. Kunieda und K. Matsushima, “High-quality full-parallax full-color three-dimensional image reconstructed by stacking large-scale computer-generated volume holograms”, *Appl. Opt.* **58**(34), G104–G111 (2019).
- [116] G. Kim, Y. Kim, J. Yun, S.-W. Moon, S. Kim, J. Kim, J. Park, T. Badloe, I. Kim und J. Rho, “Metasurface-driven full-space structured light for three-dimensional imaging”, *Nat. Commun.* **13**, 5920 (2022).
- [117] J. An, K. Won, Y. Kim, J.-Y. Hong, H. Kim, Y. Kim, H. Song, C. Choi, Y. Kim, J. Seo, A. Morozov, H. Park, S. Hong, S. Hwang, K. Kim und H.-S. Lee, “Slim-panel holographic video display”, *Nat. Commun.* **11**, 5568 (2020).
- [118] I. Sehili und A. Zerroug, “Bivariate legendre approximation”, *Int. J. Appl. Math. Res.* **6**, 125 (2017).

Abkürzungsverzeichnis

Abb.	Abbildung
ARS	Angular resolved scattering (winkelaufgelöste Streuung)
ASPW	Angular spectrum of plane waves (Winkelspektrum ebener Wellen)
BPM	Beam propagation method (Strahlpropagationsmethode)
CGH	Computer-generated hologram (computergeneriertes Hologram)
corr	Korrelationskoeffizient
DOE	diffraktives optisches Element
FDTD	Finite difference time domain method (finite Differenzen im Zeitbereich)
FT	Fourier-Transformation
FFT	Fast Fourier transform (schnelle Fourier-Transformation)
Gl.	Gleichung
IFTA	Iterativer Fourier-Transformations-Algorithmus
Kap.	Kapitel
LPA	Local periodic approximation (Näherung lokaler Periodizität)
PFA	Polynomial field approximation (polynomielle Approximation des Feldes)
PML	Perfectly matched layer (perfekt angepasste Schicht)
PSA	Polynomial structure approximation (polynomielle Approximation der Struktur)
RCWA	Rigorous coupled wave analysis (rigorose Analyse gekoppelter Wellen)
Ref.	Referenz
REM	Raster-Elektronen-Mikroskop
sd	Standard deviation (Standardabweichung)
SNR	Signal-to-noise ratio (Signal-Rausch-Verhältnis)
Tab.	Tabelle
TEA	Thin element approximation (Dünne-Elemente-Approximation)
TE	transversal elektrisch
TM	transversal magnetisch
WPM	Wave propagation method (Wellenpropagationsmethode)
1D/2D	ein-/zweidimensional

Symbolverzeichnis

a	charakteristische Größe eines Elements
A	vorgegebene Amplitude eines CGHs (Zielfunktion)
\mathcal{A}	Operator für das elektrische Feld in der BPM
\mathbf{B}	magnetische Flussdichte
\mathcal{B}	Operator für das magnetische Feld in der BPM
c_0, c	(Vakuum-)Lichtgeschwindigkeit bzw. Koeffizient für Standard-Polynome
C	allgemeine skalare Konstante
d	Schichtdicke bzw. Laufindex für Position
D	maximale Wechselwirkungsreichweite
\mathbf{D}	elektrische Flussdichte
\mathcal{D}_Φ	Projektionsoperator auf diskrete Phasenwerte
e	Exponent im linearen Modell
\mathbf{E}	elektrische Feldstärke
f	Füllfaktor
F	Brennweite bzw. allgemeine Funktion
$\mathcal{F}, \mathcal{F}^{-1}$	Fourier-Transformation, inverse Fourier-Transformation
g, G	Response-, Übertragungsfunktion des homogenen Raumes
\mathbf{H}	magnetische Feldstärke
i	imaginäre Einheit
I	Intensität
j	Laufindex
\mathbf{k}	Wellenvektor
k_0, k	(Vakuum-)Wellenzahl
K	Laufindex für periodische Wiederholungen
l	Koeffizient in der Faktorisierung der Padé-Approximation bzw. in der Approximation durch Legendre-Polynome
ℓ	Koeffizient eines Legendre-Polynoms
L	Legendre-Polynom
\mathcal{L}	Polynom-Faktor der Padé-Approximation

Symbolverzeichnis

m	Laufindex für Summationen und Ordnungen
M	Gesamtzahl der Pixel in einer Dimension
n	Brechungsindex
\bar{n}	Referenzindex
N	Zahl der Stützstellen innerhalb eines Pixels oder Elements
N_D	Zahl der diskreten Phasenwerte
N_E	Zahl der Einheitszellen
N_F	Fresnel-Zahl
N_I	Anzahl der Nachiterationen
N_K	Anzahl der Koeffizienten eines Fitmodells
N_L	Grad der (Legendre-)Polynom-Approximation des Feldes
N_P, N_Q	Ordnung der Padé-Approximation
p	Zellgröße bzw. Paritätsfunktion
P, Q	Polynome der Padé-Approximation
r	Radius (sphärische Koordinaten) bzw. Koeffizient in der Faktorisierung der Padé-Approximation
r_F	Fresnel-Koeffizient (Reflexion)
\mathbf{r}	Ortsvektor
\mathcal{R}	Polynom-Faktor der Padé-Approximation
s	Laufindex für Propagationsschritte oder Korrekturfaktor für Fernfeld-Superposition
S	Signalbereich
\mathbf{S}	Poynting-Vektor
t	Zeit bzw. Transmissionsfunktion
t_F	Fresnel-Koeffizient (Transmission)
T	Periodendauer
\mathcal{T}	linearer Differentialoperator
u	skalares Feld
w	Größe der Einheitszelle bzw. Termgewicht
W	maximales Termgewicht
x, y, z	kartesische Koordinaten
α	Parameter für Amplitudenfreiheit
β	Regressionskoeffizient
δ	Indikator für Diskretisierung infinitesimaler Größen
δ_0	Delta-Distribution
ϵ	Residuum der linearen Regression
ε	relative Permittivität

ε_0	elektrische Feldkonstante
η	Beugungseffizienz
ϑ	Polarwinkel
κ	Konstante in Crank-Nicolson-Schema
λ_0, λ	(Vakuum-)Wellenlänge
μ	relative Permeabilität
μ_0	magnetische Feldkonstante
σ	Symmetrisierung
ϕ	Phase
φ	Azimutwinkel
χ	Indikatorfunktion
ψ	komplexe Rotation für Padé-Approximation
ω	Kreisfrequenz
∇	Nabla-Operator
Δ	Laplace-Operator
*	Faltung
*	(hochgestellt) komplexe Konjugation

A Anhang

A.1 Mathematische Grundlagen

A.1.1 Differentialoperatoren

Es sei $f(x, y, z)$ eine skalare und $\mathbf{F}(x, y, z) = (F_x, F_y, F_z)^\top$ eine vektorwertige Funktion in Abhängigkeit der drei Raumkoordinaten. Der Nabla-Operator lässt sich formal als

$$\nabla := \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)^\top \quad (\text{A.1})$$

schreiben. Damit lassen sich die folgenden Operatoren definieren [106]:

$$\text{Gradient:} \quad \mathbf{grad} f := \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)^\top, \quad (\text{A.2})$$

$$\text{Divergenz:} \quad \mathbf{div} \mathbf{F} := \nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}, \quad (\text{A.3})$$

$$\text{Rotation:} \quad \mathbf{rot} \mathbf{F} := \nabla \times \mathbf{F} = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}, \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}, \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right)^\top, \quad (\text{A.4})$$

$$\text{Laplace:} \quad \Delta f := \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}, \quad (\text{A.5})$$

$$\text{bzw.} \quad \Delta \mathbf{F} := \nabla^2 \mathbf{F} = (\Delta F_x, \Delta F_y, \Delta F_z)^\top. \quad (\text{A.6})$$

In bestimmten Fällen bietet es sich an, die folgende Schreibweise zur Beschreibung eines analogen Verhaltens der ersten beiden Vektorkomponenten zu nutzen: $\mathbf{F}_\perp = (F_x, F_y)^\top$. Diese Schreibweise wird auch für die oben definierten Operatoren angewendet, sodass z. B.

$$\nabla_\perp f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)^\top, \quad \nabla_\perp \cdot \mathbf{F}_\perp = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}, \quad \Delta_\perp f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \quad (\text{A.7})$$

Für mehrfache eindimensionale Ableitungen wird folgende Schreibweise verwendet:

$$f^{(n)}(x) := \frac{\partial^n f}{\partial x^n}, \quad (\text{A.8})$$

für $n \geq 1$, bzw. $f^{(0)}(x) := f(x)$.

A.1.2 Spezielle Funktionen und Distributionen

Faltung: Es seien $f(x), g(x)$ zwei skalare Funktionen. Dann bezeichnet man die Funktion $f * g$, definiert durch

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x')g(x - x') dx' , \quad (\text{A.9})$$

als Faltung von f und g .

Rechteckfunktion: Die Rechteckfunktion wird definiert als

$$\text{rect}(x) := \begin{cases} 1, & \text{für } |x| \leq \frac{1}{2} \\ 0, & \text{für } |x| > \frac{1}{2} \end{cases} . \quad (\text{A.10})$$

Sinc-Funktion: Die Sinc-Funktion wird als Quotient einer Sinusfunktion und deren Argument definiert. Hier wird die nicht-normierte Version verwendet:

$$\text{sinc}(x) := \frac{\sin x}{x} . \quad (\text{A.11})$$

Für $x = 0$ wird die Funktion stetig als $\text{sinc}(0) = 1$ fortgesetzt.

Delta-Distribution: Es bezeichne δ_0 die Dirac'sche Delta-Distribution. Sie wird im Rahmen dieser Arbeit in folgender Notation verwendet:

$$\int_{-\infty}^{\infty} f(x)\delta_0(x) dx := f(0) , \quad (\text{A.12})$$

für eine beliebige Funktion $f(x)$. Insbesondere gilt mit einem $a \in \mathbb{R}$ für die Faltung:

$$f(x) * \delta_0(x - a) = \int_{-\infty}^{\infty} f(x - x')\delta_0(x' - a) dx' = f(x - a) . \quad (\text{A.13})$$

A.1.3 Approximation von Funktionen durch Polynome

A.1.3.1 Legendre-Polynome

Legendre-Polynome können für $x \in [-1, 1]$ in expliziter Form folgendermaßen angegeben werden [106]:

$$L_j(x) = \sum_{j'=0}^{\lfloor j/2 \rfloor} (-1)^{j'} \frac{(2j - 2j')!}{2^j (j - j')! (j - 2j')! j'!} x^{j-2j'} = \sum_{j'=0}^j \ell_{jj'} x^{j'} . \quad (\text{A.14})$$

Darin bezeichnet $\lfloor n/2 \rfloor$ eine Gauß-Klammer, also eine Abrundung auf die nächstkleinere oder gleiche Ganzzahl. Die Polynome erfüllen die Orthogonalitätsrelation

$$\int_{-1}^1 L_j(x)L_{j'}(x) dx = \begin{cases} \frac{2}{2j + 1} & \text{für } j = j' , \\ 0 & \text{sonst} . \end{cases} \quad (\text{A.15})$$

Zudem sind sie auf dem Intervall $[-1, 1]$ vollständig, d. h. jede Funktion kann in diesem Intervall auf eindeutige Weise beliebig genau durch diese Polynome approximiert werden. Die Legendre-Polynome sind definitionsgemäß auf

$$L_j(1) = 1 \quad (\text{A.16})$$

normiert. Beispielhaft seien hier einige Polynome kleinen Grades angegeben, welche im Rahmen dieser Arbeit genutzt wurden:

$$\begin{aligned} L_0(x) &= 1, \\ L_1(x) &= x, \\ L_2(x) &= \frac{3}{2}x^2 - \frac{1}{2}, \\ L_3(x) &= \frac{5}{2}x^3 - \frac{3}{2}x, \\ L_4(x) &= \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8}. \end{aligned} \quad (\text{A.17})$$

Die Polynome lassen sich auf höhere Dimensionen des Argumentes verallgemeinern [118]. Beispielsweise definiert man für zwei Dimensionen:

$$L_{j_x j_y}(x, y) := L_{j_x}(x)L_{j_y}(y). \quad (\text{A.18})$$

Da sowohl die Funktionsdefinition als auch die Integration separabel sind, übertragen sich dabei die Eigenschaften wie Orthogonalität, Normierung und Vollständigkeit für $(x, y) \in [-1, 1]^2$. Insbesondere kann eine beliebige Funktion durch eindeutig bestimmte Koeffizienten approximiert werden, so beispielsweise für das Feld $u(x, y)$ hinter einem Pixel (m_x, m_y) :

$$u(x, y) \approx \sum_{j_x=0}^{N_L} \sum_{j_y=0}^{N_L-j_x} l_{m_x m_y j_x j_y} L_{j_x j_y} \left(\frac{2(x - m_x p_x)}{p_x}, \frac{2(y - m_y p_y)}{p_y} \right), \quad (\text{A.19})$$

für $x \in [(m_x - 1/2)p_x, (m_x + 1/2)p_x]$, $y \in [(m_y - 1/2)p_y, (m_y + 1/2)p_y]$. Der maximale Grad N_L bezieht sich dabei auf die Summe der 1D-Polynomgrade. Die Koeffizienten ergeben sich analog zum 1D-Fall als

$$l_{m_x m_y j_x j_y} = \frac{(2j_x + 1)(2j_y + 1)}{4} \int_{-1}^1 \int_{-1}^1 u \left(\left(m_x + \frac{x'}{2} \right) p_x, \left(m_y + \frac{y'}{2} \right) p_y \right) L_{j_x j_y}(x', y') dx' dy'. \quad (\text{A.20})$$

A.1.3.2 Umrechnung von Polynomkoeffizienten

Für die Approximation des Feldes wird innerhalb der PFA sowohl die Darstellung mithilfe von Legendre- als auch mit Standardpolynomen genutzt, siehe Gl. (4.1) und (4.3):

$$\sum_{j=0}^{N_L} l_j L_j(2x) = \sum_{j=0}^{N_L} \sum_{j'=0}^j l_j \ell_{j j'}(2x)^{j'} = u(x) = \sum_{j=0}^{N_L} c_j x^j, \quad \text{für } x \in \left[-\frac{1}{2}, \frac{1}{2} \right], \quad (\text{A.21})$$

wobei hier zur Vereinfachung der Darstellung auf einen Index zur Kennzeichnung eines Pixels und die entsprechende lineare Transformation des Argumentes verzichtet wurde. Im Folgenden wird gezeigt, wie die Koeffizienten l_j und c_j eindeutig ineinander transformiert werden können. Obige Gleichung kann in Matrixschreibweise folgendermaßen formuliert werden:

$$\mathbf{l}^\top \mathbf{L} \mathbf{D} \mathbf{x} = \mathbf{c}^\top \mathbf{x} , \quad (\text{A.22})$$

mit $\mathbf{x} = (x^0, \dots, x^{N_L})^\top$, $\mathbf{c} = (c_0, \dots, c_{N_L})^\top$, $\mathbf{l} = (l_0, \dots, l_{N_L})^\top$, $\mathbf{L} = (\ell_{jj'})$ und der Diagonalmatrix $\mathbf{D} = (2^j \delta_{jj'})$. Es ergibt sich daher folgende Umrechnung zwischen den Koeffizienten:

$$\mathbf{c} = \mathbf{D} \mathbf{L}^\top \mathbf{l} , \quad \mathbf{l} = \mathbf{L}^{-\top} \mathbf{D}^{-1} \mathbf{c} , \quad \text{mit } \mathbf{L}^{-\top} = (\mathbf{L}^{-1})^\top . \quad (\text{A.23})$$

Für zwei Dimensionen ist die Vorgehensweise aufgrund der Separabilität analog. Hier ist

$$\sum_{j_x=0}^{N_L} \sum_{j_y=0}^{N_L-j_x} l_{j_x j_y} L_{j_x}(2x) L_{j_y}(2y) = u(x, y) = \sum_{j_x=0}^{N_L} \sum_{j_y=0}^{N_L-j_x} c_{j_x j_y} x^{j_x} y^{j_y} , \quad (\text{A.24})$$

für $x, y \in [-1/2, 1/2]$, was wiederum in Matrixschreibweise ausgedrückt werden kann:

$$\mathbf{y}^\top \mathbf{D} \mathbf{L}^\top \mathbf{L} \mathbf{D} \mathbf{x} = \mathbf{y}^\top \mathbf{C} \mathbf{x} . \quad (\text{A.25})$$

Dabei ist $\mathbf{L} = (l_{j_x j_y})$, $\mathbf{C} = (c_{j_x j_y})$. Die Umrechnung ergibt sich dann als

$$\mathbf{C} = \mathbf{D} \mathbf{L}^\top \mathbf{L} \mathbf{D} , \quad \mathbf{L} = \mathbf{L}^{-\top} \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-1} \mathbf{L}^{-1} . \quad (\text{A.26})$$

A.1.3.3 Padé-Approximation

Für eine Funktion $f(x)$ bezeichnet man die rationale Funktion

$$F(x) = \frac{P(x)}{Q(x)} = \frac{\sum_{j=0}^{N_P} c_j x^j}{1 + \sum_{j=1}^{N_Q} c'_j x^j} \quad (\text{A.27})$$

als Padé-Approximation der Ordnung (N_P, N_Q) , wenn der Funktionswert und die ersten $N_P + N_Q$ Ableitungen von f und F an der Stelle $x = 0$ übereinstimmen, d. h.

$$F^{(j)}(0) = f^{(j)}(0) , \quad \text{für } j = 0, \dots, N_P + N_Q . \quad (\text{A.28})$$

Bei der Anwendung auf Gl. (3.1) ist die Funktion $f(x) = \sqrt{1+2x}$ zu approximieren. In Tab. A.1 sind die Ergebnisse für die im Rahmen dieser Arbeit relevanten Ordnungen angegeben.

A.1.4 Fourier-Transformation

Für die Formulierung der FT sind in der Literatur verschiedene Konventionen gebräuchlich, die insbesondere einen Einfluss auf die auftretenden Vorfaktoren in einzelnen Gleichungen haben. Daher werden im folgenden Abschnitt die hier verwendeten Definitionen und die daraus resultierenden Formulierungen der für diese Arbeit relevanten Zusammenhänge vorgestellt.

Tab. A.1: Padé-Approximation $P(x)/Q(x)$ der Funktion $f(x) = \sqrt{1+2x}$ in Abhängigkeit der verwendeten Ordnung (N_P, N_Q) .

(N_P, N_Q)	$P(x)$	$Q(x)$
(1, 0)	$1 + x$	1
(1, 1)	$1 + \frac{3}{2}x$	$1 + \frac{1}{2}x$
(2, 2)	$1 + \frac{5}{2}x + \frac{5}{4}x^2$	$1 + \frac{3}{2}x + \frac{1}{4}x^2$
(3, 3)	$1 + \frac{7}{2}x + \frac{7}{2}x^2 + \frac{7}{8}x^3$	$1 + \frac{5}{2}x + \frac{3}{2}x^2 + \frac{1}{8}x^3$
(4, 4)	$1 + \frac{9}{2}x + \frac{27}{4}x^2 + \frac{15}{4}x^3 + \frac{9}{16}x^4$	$1 + \frac{7}{2}x + \frac{15}{4}x^2 + \frac{5}{4}x^3 + \frac{1}{16}x^4$

A.1.4.1 Definition

Für die Definition der FT gibt es verschiedene Konventionen. Im Rahmen dieser Arbeit wird die folgende Definition verwendet: Zu einer Funktion $u(x)$ bezeichnet man

$$U(k_x) = \mathcal{F}\{u(x)\} := \int_{-\infty}^{\infty} u(x) \exp(-ik_x x) dx \quad (\text{A.29})$$

als Fourier-Transformierte von u und definiert die inverse FT als

$$u(x) = \mathcal{F}^{-1}\{U(k_x)\} := \frac{1}{2\pi} \int_{-\infty}^{\infty} U(k_x) \exp(ik_x x) dk_x . \quad (\text{A.30})$$

Im Falle einer diskreten Funktion u_m wird die diskrete FT bzw. inverse diskrete FT folgendermaßen definiert:

$$U_{m'} = \mathcal{F}\{u_m\} := \sum_{m=0}^{M-1} u_m \exp\left(-2\pi i \frac{mm'}{M}\right) \quad (\text{A.31})$$

$$u_m = \mathcal{F}\{U_{m'}\} := \frac{1}{M} \sum_{m'=0}^{M-1} U_{m'} \exp\left(2\pi i \frac{mm'}{M}\right) , \quad (\text{A.32})$$

wobei $m, m' = 0, \dots, M-1$. Diese Definitionen lassen sich analog auf höhere Dimensionen übertragen. Im für die Berechnung von CGHs relevanten 2D-Fall lauten sie:

$$U(k_x, k_y) = \mathcal{F}\{u(x, y)\} := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) \exp(-i(k_x x + k_y y)) dx dy , \quad (\text{A.33})$$

$$u(x, y) = \mathcal{F}^{-1}\{U(k_x, k_y)\} := \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, k_y) \exp(i(k_x x + k_y y)) dk_x dk_y , \quad (\text{A.34})$$

$$U_{m'_x m'_y} = \mathcal{F}\{u_{m_x m_y}\} := \sum_{m_x=0}^{M_x-1} \sum_{m_y=0}^{M_y-1} u_{m_x m_y} \exp\left(-2\pi i \left(\frac{m_x m'_x}{M_x} + \frac{m_y m'_y}{M_y}\right)\right) , \quad (\text{A.35})$$

$$u_{m_x m_y} = \mathcal{F}\{U_{m'_x m'_y}\} := \frac{1}{M_x M_y} \sum_{m'_x=0}^{M_x-1} \sum_{m'_y=0}^{M_y-1} U_{m'_x m'_y} \exp\left(2\pi i \left(\frac{m_x m'_x}{M_x} + \frac{m_y m'_y}{M_y}\right)\right) , \quad (\text{A.36})$$

mit $m_x, m'_x = 0, \dots, M_x - 1$ bzw. $m_y, m'_y = 0, \dots, M_y - 1$.

Der Übergang von zeitabhängigen Größen, z. B. für das elektrische Feld $\mathcal{E}(t)$, zu winkelfrequenzabhängigen Größen, z. B. $E(\omega)$, wird ebenso durch eine FT vermittelt. Eine gängige Konvention dafür ist

$$E(\omega) = \mathcal{F}_t \{ \mathcal{E}(t) \} := \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{E}(t) \exp(i\omega t) dt \quad (\text{A.37})$$

bzw.

$$\mathcal{E}(t) = \mathcal{F}_t^{-1} \{ E(\omega) \} := \int_{-\infty}^{\infty} E(\omega) \exp(-i\omega t) d\omega . \quad (\text{A.38})$$

Im Unterschied zu obiger Definition sind dabei die Vorfaktoren bzw. Vorzeichen umgekehrt zwischen Hin- und Rücktransformation verteilt, weshalb zur Unterscheidung ein Index zum Symbol hinzugefügt wurde: \mathcal{F}_t . Da es sich bei $\mathcal{E}(t)$ um eine reelle Funktion handelt, gilt

$$E(-\omega) = E(\omega)^* , \quad (\text{A.39})$$

wobei $*$ die komplexe Konjugation bezeichnet. Man beschränkt sich daher in der Interpretation typische auf positive Frequenzen $\omega > 0$ und schreibt die Rücktransformation mathematisch äquivalent als

$$\mathcal{E}(t) = \mathcal{F}_t^{-1} \{ E(\omega) \} = 2\Re \left\{ \int_0^{\infty} E(\omega) \exp(-i\omega t) d\omega \right\} . \quad (\text{A.40})$$

Der Faktor 2 entsteht aus der Überlagerung der Frequenzen ω und $-\omega$.

A.1.4.2 Transformationen von typischen Funktionen

In Tab. A.2 sind die im Rahmen dieser Arbeit relevanten FT einiger Funktionen bzw. Distributionen gegeben.

Tab. A.2: Fourier-Transformation $U = \mathcal{F}\{u\}$ ausgewählter Funktionen und Distributionen, gegebenenfalls für ein $a \in \mathbb{R}$ und ein $m \in \mathbb{N}$.

$u(x)$	$U(k_x)$
$\text{rect}(ax)$	$\frac{1}{ a } \text{sinc} \left(\frac{k_x}{2a} \right)$
$x^m v(x)$	$i^m \frac{d^m}{dk_x^m} V(k_x)$
$\delta_0(x - a)$	$\exp(-iak_x)$
$\sum_{m=-\infty}^{\infty} \delta_0(x - ma)$	$\frac{2\pi}{a} \sum_{m=-\infty}^{\infty} \delta_0 \left(k_x - \frac{2\pi m}{a} \right)$

A.1.4.3 Eigenschaften

Die FT ist linear, d. h. für zwei Funktionen u, v und $a, b \in \mathbb{C}$ gilt:

$$\mathcal{F}\{au + bv\} = a\mathcal{F}\{u\} + b\mathcal{F}\{v\} . \quad (\text{A.41})$$

Für die FT von Produkten und Faltungen zweier Funktionen u, v gilt aufgrund des Faltungstheorems:

$$\mathcal{F}\{uv\} = \frac{1}{2\pi} \mathcal{F}\{u\} * \mathcal{F}\{v\} , \quad (\text{A.42})$$

$$\mathcal{F}\{u * v\} = \mathcal{F}\{u\} \mathcal{F}\{v\} . \quad (\text{A.43})$$

Die Norm einer Funktion u und ihrer FT U stehen gemäß des Satzes von Plancherel in einem festen Verhältnis:

$$\int_{-\infty}^{\infty} |u(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |U(k_x)|^2 dk_x . \quad (\text{A.44})$$

Ein ähnlicher Zusammenhang gilt auch für die diskrete FT:

$$\sum_{m=0}^{M-1} |u_m|^2 = \frac{1}{M} \sum_{m=0}^{M-1} |U_m|^2 . \quad (\text{A.45})$$

A.2 Simulation von optischen Feldern

A.2.1 Fernfeldberechnung

A.2.1.1 Fraunhofer-Näherung

Die Propagation in homogenen Medien lässt sich ausgehend von der ASPW, Gl. (2.21), auch vollständig im Ortsraum darstellen, indem man den Faltungssatz (A.43) und die Responsefunktion $g(x, y, z - z_0) = \mathcal{F}^{-1}\{G(k_x, k_y; z - z_0)\}$ nutzt. Zur Vereinfachung der Notation sei $z_0 = 0$:

$$u(x, y, z) = g(x, y, z) * u(x, y, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x - x', y - y', z) u(x', y', 0) dx' dy' . \quad (\text{A.46})$$

Dabei ist

$$g(x, y, z) = -\frac{1}{2\pi} \frac{\partial}{\partial z} \frac{\exp(ikr)}{r} = -\frac{1}{2\pi} \frac{z}{r^2} \left(ik - \frac{1}{r} \right) \exp(ikr) \quad (\text{A.47})$$

für $r = \sqrt{x^2 + y^2 + z^2}$ [27]. In paraxialer oder Fresnel-Näherung nutzt man $r \approx z + (x^2 + y^2)/(2z)$ und $kz \gg 1$ um folgenden Ausdruck zu erhalten:

$$g(x, y, z) \approx g_F(x, y, z) = -\frac{ik}{2\pi z} \exp(ikz) \exp\left(i \frac{k}{2z} (x^2 + y^2)\right) . \quad (\text{A.48})$$

Im Bereich des Fernfeldes ist die Fresnel-Zahl $N_F = a^2/(\lambda z) \ll 1$. Die charakteristische Größe des Elementes a soll hier in dem Sinne verwendet werden, dass

$$u(x, y, 0) = 0, \quad \text{für } x^2 + y^2 > a^2 \quad (\text{A.49})$$

gilt. In paraxialer Fraunhofer-Näherung ist dann

$$u(x, y, z) \approx \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_F(x - x', y - y', z) u(x', y', 0) dx' dy' \quad (\text{A.50})$$

$$\approx -\frac{ik}{2\pi z} \exp(ikz) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(i\frac{k}{2z}((x - x')^2 + (y - y')^2)\right) u_0(x', y') dx' dy' \quad (\text{A.51})$$

$$\approx -\frac{ik}{2\pi z} \exp(ikz) \exp\left(i\frac{k}{2z}(x^2 + y^2)\right) \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \underbrace{\exp\left(i\frac{k}{2z}(x'^2 + y'^2)\right)}_{\approx 1, \text{ wo } u_0(x', y') \neq 0} \exp\left(-i\frac{k}{z}(xx' + yy')\right) u_0(x', y') dx' dy' \quad (\text{A.52})$$

$$\approx -\frac{ik}{2\pi z} \exp(ikz) \exp\left(i\frac{k}{2z}(x^2 + y^2)\right) U_0\left(k\frac{x}{z}, k\frac{y}{z}\right). \quad (\text{A.53})$$

Ein ähnlicher Ausdruck lässt sich auch unabhängig von der paraxialen Näherung bestimmen, wobei man ebenfalls $kz \gg 1$ annimmt. In nicht-paraxialer Fraunhofer-Näherung lässt sich das Fernfeld ausgehend von Gl. (A.46) damit folgendermaßen berechnen:

$$u(x, y, z) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{z}{r'^2} \left(ik - \frac{1}{r'}\right) \exp(ikr') u(x', y', 0) dx' dy', \quad r' = \sqrt{(x - x')^2 + (y - y')^2 + z^2} \quad (\text{A.54})$$

$$\approx -\frac{ikz}{2\pi r^2} \exp(ikr) \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \underbrace{\left(1 + \frac{i}{kr'}\right)}_{\approx 1, \text{ da } r' > z} \underbrace{\exp\left(i\frac{k}{2r}(x'^2 + y'^2)\right)}_{\approx 1, \text{ da } x'^2 + y'^2 < a^2} \exp\left(-i\frac{k}{r}(xx' + yy')\right) u(x', y', 0) dx' dy' \quad (\text{A.55})$$

$$\approx -\frac{ikz}{2\pi r^2} \exp(ikr) U\left(k\frac{x}{r}, k\frac{y}{r}; 0\right). \quad (\text{A.56})$$

Dabei wurde die Näherung $r' = \sqrt{r^2 - 2xx' - 2yy' + x'^2 + y'^2} \approx r - \frac{xx' + yy'}{r} + \frac{x'^2 + y'^2}{2r}$ genutzt, denn es gilt $|x'| \ll r, |y'| \ll r$, dort wo $|u(x', y', 0)| > 0$ ist.

A.2.1.2 Fernfeld von CGHs

Zur Berechnung der FT des Nahfeldes eines 1D-CGHs gemäß Gl. (2.63) nutzt man verschiedene Eigenschaften der FT aus, was im Folgenden näher ausgeführt wird [50]:

$$U(k_x) = \mathcal{F}\{u(x)\} = \mathcal{F}\left\{\left[\sum_{m=0}^{M-1} u_m \delta_0(x - mp)\right] * \text{rect}\left(\frac{x}{p}\right) * \sum_{K=-\infty}^{\infty} \delta_0(x - KMp)\right\} \quad (\text{A.57})$$

$$= \mathcal{F}\left\{\sum_{m=0}^{M-1} u_m \delta_0(x - mp)\right\} \cdot \mathcal{F}\left\{\text{rect}\left(\frac{x}{p}\right)\right\} \cdot \mathcal{F}\left\{\sum_{K=-\infty}^{\infty} \delta_0(x - KMp)\right\} \quad (\text{A.58})$$

$$= \left[\sum_{m=0}^{M-1} u_m \exp(-impk_x)\right] \cdot \text{psinc}\left(\frac{k_x p}{2}\right) \cdot \frac{2\pi}{Mp} \sum_{K=-\infty}^{\infty} \delta_0\left(k_x - \frac{K}{Mp}\right) \quad (\text{A.59})$$

$$= \frac{2\pi}{M} \text{sinc}\left(\frac{k_x p}{2}\right) \left[\sum_{m=0}^{M-1} u_m \exp\left(-2\pi i \frac{mm'}{M}\right)\right] \cdot \left[\sum_{m'=0}^{M-1} \delta_0\left(k_x - \frac{2\pi m'}{Mp}\right) * \sum_{K=-\infty}^{\infty} \delta_0\left(k_x - \frac{2\pi K}{p}\right)\right] \quad (\text{A.60})$$

$$= \frac{2\pi}{M} \text{sinc}\left(\frac{k_x p}{2}\right) \left[\sum_{m=0}^{M-1} U_m \delta_0(k_x - k_{x,m}) * \sum_{K=-\infty}^{\infty} \delta_0\left(k_x - \frac{2\pi K}{p}\right)\right] \quad (\text{A.61})$$

Dabei verwendet man das Faltungstheorem (A.43), die Identitäten in Tab. A.2, die Faltung mit der Delta-Distribution gemäß Gl. (A.13) und die Definition der diskreten FT (A.31) für $k_{x,m} = 2\pi m/(Mp)$. Als Ergebnis erhält man Gl. (2.64).

Für den 2D-Fall wird nun ein Element aus $M_x \times M_y$ Pixeln der Größe $p_x \times p_y$ betrachtet. Für ein konstantes Feld innerhalb eines jeden Pixels ist

$$u(x, y) = u_{m_x m_y}, \quad \text{für } \left(m_{x/y} - \frac{1}{2}\right) p_{x/y} \leq x/y \leq \left(m_{x/y} + \frac{1}{2}\right) p_{x/y}, \quad (\text{A.62})$$

wobei $m_{x/y} = 0, \dots, M_{x/y} - 1$. Die periodische Fortsetzung ergibt sich gemäß Ref. [50] als

$$u(x, y) = \left[\sum_{m_x=0}^{M_x-1} \sum_{m_y=0}^{M_y-1} u_{m_x m_y} \delta_0(x - m_x p_x) \delta_0(y - m_y p_y)\right] * \text{rect}\left(\frac{x}{p_x}\right) \text{rect}\left(\frac{y}{p_y}\right) * \sum_{K_x=-\infty}^{\infty} \sum_{K_y=-\infty}^{\infty} \delta_0(x - K_x M_x p_x) \delta_0(y - K_y M_y p_y). \quad (\text{A.63})$$

Die FT ergibt sich analog zum 1D-Fall, da das Feld separabel ist:

$$U(k_x, k_y) = \frac{(2\pi)^2}{M_x M_y} \text{sinc}\left(\frac{k_x p_x}{2}\right) \text{sinc}\left(\frac{k_y p_y}{2}\right) \left[\sum_{m_x=0}^{M_x-1} \sum_{m_y=0}^{M_y-1} U_{m_x m_y} \delta_0(k_x - k_{x,m_x}) \delta_0(k_y - k_{y,m_y}) * \sum_{K_x=-\infty}^{\infty} \sum_{K_y=-\infty}^{\infty} \delta_0\left(k_x - \frac{2\pi K_x}{p_x}\right) \delta_0\left(k_y - \frac{2\pi K_y}{p_y}\right)\right] \quad (\text{A.64})$$

Die Wellenzahlen der entsprechenden Beugungsordnungen sind dabei $k_{x,m_x} = 2\pi m_x/(M_x p_x)$ bzw. $k_{y,m_y} = 2\pi m_y/(M_y p_y)$.

A.2.1.3 Fernfeldberechnung mit polynomieller Approximation

Eine Beschreibung des Feldes mithilfe einer polynomiellen Näherung ist, zunächst für den 1D-Fall, in Gl. (4.4) gegeben. Wendet man eine FT auf diese Gleichung an, so erhält man

$$U(k_x) = \mathcal{F}\{u(x)\} = \mathcal{F}\left\{\sum_{j=0}^{N_L} \left[\sum_{m=0}^{M-1} c_{mj} \delta_0(x - mp) \right] * \left(\frac{x}{p}\right)^j \text{rect}\left(\frac{x}{p}\right) * \sum_{K=-\infty}^{\infty} \delta_0(x - KMp) \right\} \quad (\text{A.65})$$

$$= \sum_{j=0}^{N_L} \left[\mathcal{F}\left\{\sum_{m=0}^{M-1} c_{mj} \delta_0(x - mp)\right\} \mathcal{F}\left\{\left(\frac{x}{p}\right)^j \text{rect}\left(\frac{x}{p}\right)\right\} \right] \mathcal{F}\left\{\sum_{K=-\infty}^{\infty} \delta_0(x - KMp)\right\}, \quad (\text{A.66})$$

unter Ausnutzung von Linearität (A.41) und Faltungstheorem (A.43). Für die FT des mittleren Faktors gilt

$$\mathcal{F}\left\{x^j \text{rect}\left(\frac{x}{p}\right)\right\} = \frac{i^j}{2\pi} \frac{d^j}{dk_x^j} \left(p \text{sinc}\left(\frac{k_x p}{2}\right) \right) \quad (\text{A.67})$$

$$= \frac{1}{2\pi p} \left(\frac{ip}{2}\right)^j \text{sinc}^{(j)}\left(\frac{k_x p}{2}\right), \quad (\text{A.68})$$

mit $\text{sinc}^{(j)}$ als j -maliger Ableitung der sinc-Funktion. Die anderen Faktoren sind identisch zur Herleitung im vorherigen Abschnitt, sodass sich schließlich durch Einsetzen Gl. (4.5) ergibt:

$$U(k_x) = \frac{2\pi}{M} \sum_{j=0}^{N_L} \left(\frac{i}{2}\right)^j \text{sinc}^{(j)}\left(\frac{k_x p}{2}\right) \left[\sum_{m=0}^{M-1} C_{mj} \delta_0(k_x - k_{x,m}) * \sum_{K=-\infty}^{\infty} \delta_0\left(k_x - \frac{2\pi K}{p}\right) \right]. \quad (\text{A.69})$$

Für ein 2D-CGH wird das Feld hinter einem Pixel (m_x, m_y) mit Standardpolynomen bis zur Ordnung N_L approximiert, vgl. Gl. (4.3):

$$u(x) \approx \sum_{j_x=0}^{N_L} \sum_{j_y=0}^{N_L-j_x} c_{m_x m_y j_x j_y} \left(\frac{x - m_x p_x}{p_x}\right)^{j_x} \left(\frac{y - m_y p_y}{p_y}\right)^{j_y}, \quad (\text{A.70})$$

für $x \in [(m_x - 1/2)p_x, (m_x + 1/2)p_x]$, $y \in [(m_y - 1/2)p_y, (m_y + 1/2)p_y]$. Das Feld hinter dem gesamten Element lässt sich dann wie in Gl. (4.4) als

$$u(x, y) = \left[\sum_{j_x=0}^{N_L} \sum_{j_y=0}^{N_L-j_x} \left[\sum_{m_x=0}^{M_x-1} \sum_{m_y=0}^{M_y-1} c_{m_x m_y j_x j_y} \delta_0(x - m_x p_x) \delta_0(y - m_y p_y) \right] * \right. \\ \left. * \left(\frac{x}{p_x}\right)^{j_x} \left(\frac{y}{p_y}\right)^{j_y} \text{rect}\left(\frac{x}{p_x}\right) \text{rect}\left(\frac{y}{p_y}\right) \right] * \\ * \sum_{K_x=-\infty}^{\infty} \sum_{K_y=-\infty}^{\infty} \delta_0(x - K_x M_x p_x) \delta_0(y - K_y M_y p_y) \quad (\text{A.71})$$

darstellen. Die FT dieses Ausdrucks ergibt sich analog zum 1D-Fall als Superposition der diskreten

FT der einzelnen Polynomordnungen:

$$\begin{aligned}
U(k_x, k_y) = & \frac{4\pi^2}{M_x M_y} \sum_{j_x=0}^{N_L} \sum_{j_y=0}^{N_L-j_x} \left(\frac{i}{2}\right)^{j_x+j_y} \text{sinc}^{(j_x)}\left(\frac{k_x p_x}{2}\right) \text{sinc}^{(j_y)}\left(\frac{k_y p_y}{2}\right) \cdot \\
& \cdot \left[\sum_{m_x=0}^{M_x-1} \sum_{m_y=0}^{M_y-1} C_{m_x m_y j_x j_y} \delta_0(k_x - k_{x, m_x}) \delta_0(k_y - k_{y, m_y}) * \right. \\
& \left. * \sum_{K_x=-\infty}^{\infty} \sum_{K_y=-\infty}^{\infty} \delta_0\left(k_x - \frac{2\pi K_x}{p_x}\right) \delta_0\left(k_y - \frac{2\pi K_y}{p_y}\right) \right]. \quad (\text{A.72})
\end{aligned}$$

A.2.2 Beam Propagation Method

A.2.2.1 Vektorielle Darstellung der Propagationsgleichungen

Im Rahmen der BPM erhält man folgende Wellengleichung für die transversalen Komponenten des elektrischen Feldes, siehe Gl. (2.48),

$$\frac{\partial^2}{\partial z^2} \hat{\mathbf{E}}_{\perp} + 2i\bar{n}k_0 \frac{\partial}{\partial z} \hat{\mathbf{E}}_{\perp} + \Delta_{\perp} \hat{\mathbf{E}}_{\perp} + (n^2 - \bar{n}^2) k_0^2 \hat{\mathbf{E}}_{\perp} + \nabla_{\perp} \left(\frac{1}{n^2} \nabla_{\perp} n^2 \cdot \hat{\mathbf{E}}_{\perp} \right) = 0. \quad (\text{A.73})$$

bzw. in der Darstellung als gekoppelte Differentialgleichungen (2.49)

$$\frac{\partial^2}{\partial z^2} \begin{pmatrix} \hat{E}_x \\ \hat{E}_y \end{pmatrix} + 2i\bar{n}k_0 \frac{\partial}{\partial z} \begin{pmatrix} \hat{E}_x \\ \hat{E}_y \end{pmatrix} + 2\bar{n}^2 k_0^2 \begin{pmatrix} \mathcal{A}_{xx} & \mathcal{A}_{xy} \\ \mathcal{A}_{yx} & \mathcal{A}_{yy} \end{pmatrix} \begin{pmatrix} \hat{E}_x \\ \hat{E}_y \end{pmatrix} = 0. \quad (\text{A.74})$$

Darin ist

$$\mathcal{A}_{xx} \hat{E}_x = \frac{1}{2\bar{n}^2 k_0^2} \left(\frac{\partial}{\partial x} \left(\frac{1}{n^2} \frac{\partial}{\partial x} (n^2 \hat{E}_x) \right) + \frac{\partial^2}{\partial y^2} \hat{E}_x \right) + \frac{n^2 - \bar{n}^2}{2\bar{n}^2} \hat{E}_x, \quad (\text{A.75})$$

$$\mathcal{A}_{xy} \hat{E}_y = \frac{1}{2\bar{n}^2 k_0^2} \frac{\partial}{\partial x} \left(\frac{1}{n^2} \frac{\partial n^2}{\partial y} \hat{E}_y \right), \quad (\text{A.76})$$

$$\mathcal{A}_{yx} \hat{E}_x = \frac{1}{2\bar{n}^2 k_0^2} \frac{\partial}{\partial y} \left(\frac{1}{n^2} \frac{\partial n^2}{\partial x} \hat{E}_x \right), \quad (\text{A.77})$$

$$\mathcal{A}_{yy} \hat{E}_y = \frac{1}{2\bar{n}^2 k_0^2} \left(\frac{\partial^2}{\partial x^2} \hat{E}_y + \frac{\partial}{\partial y} \left(\frac{1}{n^2} \frac{\partial}{\partial y} (n^2 \hat{E}_y) \right) \right) + \frac{n^2 - \bar{n}^2}{2\bar{n}^2} \hat{E}_y. \quad (\text{A.78})$$

In der Darstellung für das magnetische Feld lautet die Wellengleichung für die transversalen Komponenten:

$$\Delta \mathbf{H}_{\perp} + n^2 k_0^2 \mathbf{H}_{\perp} + \frac{1}{n^2} \left(\frac{\partial n^2}{\partial y} (\nabla \times \mathbf{H})_z - \frac{\partial n^2}{\partial z} (\nabla \times \mathbf{H})_y, \frac{\partial n^2}{\partial z} (\nabla \times \mathbf{H})_x - \frac{\partial n^2}{\partial x} (\nabla \times \mathbf{H})_z \right)^{\top} = 0. \quad (\text{A.79})$$

Nach Einführung einer langsam veränderlichen Amplitude und Vernachlässigung von $\partial n / \partial z$ analog zum elektrischen Feld ergibt sich:

$$\frac{\partial^2}{\partial z^2} \hat{\mathbf{H}}_{\perp} + 2i\bar{n}k_0 \frac{\partial}{\partial z} \hat{\mathbf{H}}_{\perp} + \Delta_{\perp} \hat{\mathbf{H}}_{\perp} + (n^2 - \bar{n}^2) k_0^2 \hat{\mathbf{H}}_{\perp} + \frac{1}{n^2} \left(\frac{\partial}{\partial x} H_y - \frac{\partial}{\partial y} H_x \right) \begin{pmatrix} \frac{\partial n^2}{\partial y} \\ -\frac{\partial n^2}{\partial x} \end{pmatrix}^{\top} = 0, \quad (\text{A.80})$$

was sich wiederum komponentenweise als

$$\frac{\partial^2}{\partial z^2} \begin{pmatrix} \widehat{H}_x \\ \widehat{H}_y \end{pmatrix} + 2i\bar{n}k_0 \frac{\partial}{\partial z} \begin{pmatrix} \widehat{H}_x \\ \widehat{H}_y \end{pmatrix} + 2\bar{n}^2 k_0^2 \begin{pmatrix} \mathcal{B}_{xx} & \mathcal{B}_{xy} \\ \mathcal{B}_{yx} & \mathcal{B}_{yy} \end{pmatrix} \begin{pmatrix} \widehat{H}_x \\ \widehat{H}_y \end{pmatrix} = 0 \quad (\text{A.81})$$

darstellen lässt. Hierbei ist

$$\mathcal{B}_{xx}\widehat{H}_x = \frac{1}{2\bar{n}^2 k_0^2} \left(\frac{\partial^2}{\partial x^2} \widehat{H}_x + n^2 \frac{\partial}{\partial y} \left(\frac{1}{n^2} \frac{\partial}{\partial y} \widehat{H}_x \right) \right) + \frac{n^2 - \bar{n}^2}{2\bar{n}^2} \widehat{H}_x, \quad (\text{A.82})$$

$$\mathcal{B}_{xy}\widehat{H}_y = \frac{1}{2\bar{n}^2 k_0^2} \frac{1}{n^2} \frac{\partial n^2}{\partial y} \frac{\partial}{\partial x} H_y, \quad (\text{A.83})$$

$$\mathcal{B}_{yx}\widehat{H}_x = \frac{1}{2\bar{n}^2 k_0^2} \frac{1}{n^2} \frac{\partial n^2}{\partial x} \frac{\partial}{\partial y} H_x, \quad (\text{A.84})$$

$$\mathcal{B}_{yy}\widehat{H}_y = \frac{1}{2\bar{n}^2 k_0^2} \left(n^2 \frac{\partial}{\partial x} \left(\frac{1}{n^2} \frac{\partial}{\partial x} \widehat{H}_y \right) + \frac{\partial^2}{\partial y^2} \widehat{H}_y \right) + \frac{n^2 - \bar{n}^2}{2\bar{n}^2} \widehat{H}_y. \quad (\text{A.85})$$

A.2.2.2 Übergangsbedingungen an Grenzflächen

Ähnlich wie in Abschnitt 3.1.4 für das elektrische Feld geschildert, lassen sich auch Gleichungen für den Übergang des magnetischen Feldes an Grenzflächen zwischen zwei Schichten finden. Hier wird ebenfalls zunächst der 1D-Fall mit Invarianz in y -Richtung betrachtet. Aufgrund der Maxwell-Gleichungen (2.8) ist neben E_x auch $1/n^2 \partial H_y / \partial z$ stetig. Für die Felder in den jeweiligen Schichten gilt daher

$$H_{y,1/2}^\pm = \widehat{H}_{y,1/2}^\pm \exp(\pm i\bar{n}_{1/2} k_0 z), \quad \frac{\partial}{\partial z} H_{y,1/2}^\pm = \left(\frac{\partial}{\partial z} \widehat{H}_{y,1/2}^\pm \pm i\bar{n}_{1/2} k_0 \widehat{H}_{y,1/2}^\pm \right) \exp(\pm i\bar{n}_{1/2} k_0 z) \quad (\text{A.86})$$

und mit der Padé-Approximation gilt analog

$$\frac{\partial}{\partial z} H_{y,1/2}^\pm = \pm i\bar{n}_{1/2} k_0 \frac{P(\mathcal{B}_{yy,1/2})}{Q(\mathcal{B}_{yy,1/2})} \widehat{H}_{y,1/2}^\pm \exp(\pm i\bar{n}_{1/2} k_0 z). \quad (\text{A.87})$$

Die Erfüllung der Übergangsbedingungen erfordert demnach:

$$\widehat{H}_{y,1}^+ + \widehat{H}_{y,1}^- = \widehat{H}_{y,2}^+ + \widehat{H}_{y,2}^-, \quad (\text{A.88})$$

$$\frac{\bar{n}_1}{n_1^2} \frac{P(\mathcal{B}_{yy,1})}{Q(\mathcal{B}_{yy,1})} (\widehat{H}_{y,1}^+ - \widehat{H}_{y,1}^-) = \frac{\bar{n}_2}{n_2^2} \frac{P(\mathcal{B}_{yy,2})}{Q(\mathcal{B}_{yy,2})} (\widehat{H}_{y,2}^+ - \widehat{H}_{y,2}^-). \quad (\text{A.89})$$

Das entstehende Gleichungssystem lässt sich wie für das elektrische Feld lösen.

A.2.2.3 Diskretisierung von Differentialoperatoren

Für die BPM ist die numerische Berechnung von Differentialoperatoren notwendig. Hier wird beispielhaft ein skalares Feld u betrachtet, das in Anlehnung an Abschnitt 3.1 als $u(x, y, z) = u_{n_x n_y}^{(s)}$ diskretisiert wird. Die Diskretisierung einer Differentialgleichung der Form $\partial u / \partial z = \mathcal{A}u$ mit einem beliebigen Operator \mathcal{A} ergibt sich gemäß dem Crank-Nicolson-Schema als

$$\frac{u^{(s+1)} - u^{(s)}}{\delta z} = \frac{1}{2} (\mathcal{A}u^{(s)} + \mathcal{A}u^{(s+1)}), \quad (\text{A.90})$$

also

$$\left(1 - \frac{\delta z}{2} \mathcal{A}\right) u^{(s+1)} = \left(1 + \frac{\delta z}{2} \mathcal{A}\right) u^{(s)}. \quad (\text{A.91})$$

Diese Formulierung erfordert zwar die Lösung einer impliziten Gleichung, ist aber genauer als eine Berechnung gemäß dem Vorwärts-Euler-Schema, $(u^{(s+1)} - u^{(s)})/\delta z = \mathcal{A}u^{(s)}$. Für die Padé-Approximation ergibt sich Gl. (3.4) auf gleiche Weise aus Gl. (3.3).

Der Operator \mathcal{A} setzt sich im Rahmen dieser Arbeit aus Ableitungen erster und zweiter Ordnung in x und y zusammen. Die Diskretisierung erfolgt mittels zentraler finiter Differenzen:

$$\left(\frac{\partial}{\partial x} u\right)_{n_x n_y} = \frac{u_{n_x+1, n_y} - u_{n_x-1, n_y}}{2\delta x}, \quad \left(\frac{\partial^2}{\partial x^2} u\right)_{n_x n_y} = \frac{u_{n_x+1, n_y} - 2u_{n_x, n_y} + u_{n_x-1, n_y}}{\delta x^2}, \quad (\text{A.92})$$

$$\left(\frac{\partial}{\partial y} u\right)_{n_x n_y} = \frac{u_{n_x, n_y+1} - u_{n_x, n_y-1}}{2\delta y}, \quad \left(\frac{\partial^2}{\partial y^2} u\right)_{n_x n_y} = \frac{u_{n_x, n_y+1} - 2u_{n_x, n_y} + u_{n_x, n_y-1}}{\delta y^2}. \quad (\text{A.93})$$

Diese Operationen lassen sich jeweils als Matrix-Vektor-Multiplikation darstellen, indem man die Felder als Vektoren darstellt, z. B. für festes n_y : $(u_{1, n_y}, u_{2, n_y}, \dots, u_{N_x, n_y})^\top$. Dann ist

$$\begin{pmatrix} (\partial u/\partial x)_{1, n_y} \\ (\partial u/\partial x)_{2, n_y} \\ \vdots \\ \vdots \\ (\partial u/\partial x)_{N_x, n_y} \end{pmatrix} = \frac{1}{2\delta x} \begin{pmatrix} 0 & 1 & & -1 \\ -1 & 0 & 1 & \\ & -1 & 0 & \ddots \\ & & \ddots & \ddots & 1 \\ 1 & & & -1 & 0 \end{pmatrix} \begin{pmatrix} u_{1, n_y} \\ u_{2, n_y} \\ \vdots \\ \vdots \\ u_{N_x, n_y} \end{pmatrix}, \quad (\text{A.94})$$

$$\begin{pmatrix} (\partial^2 u/\partial x^2)_{1, n_y} \\ (\partial^2 u/\partial x^2)_{2, n_y} \\ \vdots \\ \vdots \\ (\partial^2 u/\partial x^2)_{N_x, n_y} \end{pmatrix} = \frac{1}{\delta x^2} \begin{pmatrix} -2 & 1 & & 1 \\ 1 & -2 & 1 & \\ & 1 & -2 & \ddots \\ & & \ddots & \ddots & 1 \\ 1 & & & 1 & -2 \end{pmatrix} \begin{pmatrix} u_{1, n_y} \\ u_{2, n_y} \\ \vdots \\ \vdots \\ u_{N_x, n_y} \end{pmatrix}. \quad (\text{A.95})$$

Dabei sind alle nicht angegebenen Einträge der Matrizen als 0 zu verstehen. Die beiden Einträge abseits der drei Diagonalen aufgrund periodischer Randbedingungen erhöhen die Komplexität der Matrix-Vektor-Multiplikationen nicht. Für die Propagationsschritte müssen allerdings auch lineare Gleichungssysteme mit derartigen Koeffizientenmatrizen gelöst werden. Da es sich nicht um Tridiagonalmatrizen handelt, ist die direkte Anwendung des Thomas-Algorithmus [86] nicht möglich, jedoch lässt sich die Lösung durch Ausnutzung der Sherman-Morrison-Formel [90] auf tridiagonale Gleichungssysteme zurückführen: Für eine Matrix A und zwei Spaltenvektoren v, w kann die Inverse der modifizierten Matrix $A' = A + vw^\top$ durch

$$A'^{-1} = A^{-1} - \frac{A^{-1}vw^\top A^{-1}}{1 + w^\top A^{-1}v} \quad (\text{A.96})$$

Hierbei wurden wiederum periodische Randbedingungen berücksichtigt, zudem zur Vereinfachung $\delta x = \delta y$. Eine effiziente Lösung mit dem Thomas-Algorithmus ist für Gleichungssysteme mit diesen Matrizen nicht möglich.

A.2.2.4 Differentialoperatoren in der BPM

Die in Abschnitt A.2.2.1 gegebenen Operatoren wurden mithilfe von zentralen Differenzenquotienten, siehe Abschnitt A.2.2.3, diskretisiert [35]. Beispielhaft seien hier die Operatoren \mathcal{A}_{yy} und \mathcal{B}_{yy} angegeben:

$$\begin{aligned} (\mathcal{A}_{yy}\widehat{E}_y)_{m_x m_y} &= \frac{n_{m_x m_y}^2 - \bar{n}^2}{2\bar{n}^2} \widehat{E}_{y, m_x m_y} + \frac{1}{2\bar{n}^2 k_0^2 \delta x^2} (\widehat{E}_{y, m_x+1, m_y} - 2\widehat{E}_{y, m_x m_y} + \widehat{E}_{y, m_x-1, m_y}) + \\ &+ \frac{1}{2\bar{n}^2 k_0^2 \delta y^2} \left(- \left(\frac{2n_{m_x m_y}^2}{n_{m_x, m_y+1}^2 + n_{m_x m_y}^2} + \frac{2n_{m_x m_y}^2}{n_{m_x m_y}^2 + n_{m_x, m_y-1}^2} \right) \widehat{E}_{y, m_x m_y} + \right. \\ &\left. + \frac{2n_{m_x, m_y+1}^2}{n_{m_x, m_y+1}^2 + n_{m_x m_y}^2} \widehat{E}_{y, m_x, m_y+1} + \frac{2n_{m_x, m_y-1}^2}{n_{m_x m_y}^2 + n_{m_x, m_y-1}^2} \widehat{E}_{y, m_x, m_y-1} \right), \quad (\text{A.100}) \end{aligned}$$

$$\begin{aligned} (\mathcal{B}_{yy}\widehat{H}_y)_{m_x m_y} &= \frac{n_{m_x m_y}^2 - \bar{n}^2}{2\bar{n}^2} \widehat{H}_{y, m_x m_y} + \frac{1}{2\bar{n}^2 k_0^2 \delta y^2} (\widehat{H}_{y, m_x, m_y+1} - 2\widehat{H}_{y, m_x m_y} + \widehat{H}_{y, m_x, m_y-1}) + \\ &+ \frac{1}{2\bar{n}^2 k_0^2 \delta x^2} \left(- \left(\frac{2n_{m_x m_y}^2}{n_{m_x+1, m_y}^2 + n_{m_x m_y}^2} + \frac{2n_{m_x m_y}^2}{n_{m_x m_y}^2 + n_{m_x-1, m_y}^2} \right) \widehat{H}_{y, m_x m_y} + \right. \\ &\left. + \frac{2n_{m_x m_y}^2}{n_{m_x+1, m_y}^2 + n_{m_x m_y}^2} \widehat{H}_{y, m_x+1, m_y} + \frac{2n_{m_x m_y}^2}{n_{m_x m_y}^2 + n_{m_x-1, m_y}^2} \widehat{H}_{y, m_x-1, m_y} \right). \quad (\text{A.101}) \end{aligned}$$

Die Operatoren \mathcal{A}_{xx} und \mathcal{B}_{xx} ergeben sich durch eine geeignete Vertauschung der Variablen. Kopplungsterme wie \mathcal{A}_{xy} lassen sich auf ähnliche Weise diskretisieren, spielen aber für die Berechnungen im Rahmen dieser Arbeit keine Rolle.

A.2.3 Polynomielle Approximation des Nahfeldes

A.2.3.1 Zweidimensionales Fitmodell

In Abschnitt 4.2.2 wurde die Konstruktion eines linearen Modells für den 1D-Fall geschildert. Hier soll nun das entsprechende Vorgehen für den 2D-Fall erläutert werden. Die meisten Zusammenhänge lassen sich direkt verallgemeinern, wobei die jeweiligen Größen nach x - und y -Richtung unterschieden werden. Zur Beschreibung des Nahfeldes gemäß Gl. (A.19) suchen wir folglich nach einer Funktion $F_{j_x j_y}$ zur Berechnung der Koeffizienten $c_{m_x m_y j_x j_y}$ aus den Füllfaktoren $f_{m_x m_y}$

$$c_{m_x m_y j_x j_y} = F_{j_x j_y} \left(\dots, f_{m_x-1, m_y}, \dots, f_{m_x, m_y-1}, f_{m_x, m_y}, f_{m_x, m_y+1}, \dots, f_{m_x+1, m_y}, \dots \right), \quad (\text{A.102})$$

wobei die Festlegung der zu berücksichtigenden Parameter anders als im 1D-Fall erfolgen muss. Hier werden alle Terme mit einer Manhattan-Distanz von höchstens D Pixeln berücksichtigt, d. h. alle $f_{m_x+d_x, m_y+d_y}$ mit $|d_x| + |d_y| \leq D$. Die Parität der 2D-Legendre-Polynome muss hinsichtlich

der beiden Richtungen unterschieden werden, d. h. es gilt:

$$L_{j_x j_y}(-x, -y) = (-1)^{j_x + j_y} L_{j_x j_y}(x, y) . \quad (\text{A.103})$$

Daher werden die (anti-) symmetrisierten Parameter separat für beide Dimensionen konstruiert:

$$\begin{aligned} f_{m_x m_y, d_x d_y}^{ss} &= \frac{1}{2} \left(f_{m_x + d_x, m_y + d_y} + f_{m_x - d_x, m_y + d_y} + f_{m_x + d_x, m_y - d_y} + f_{m_x - d_x, m_y - d_y} \right) - 1 \\ f_{m_x m_y, d_x d_y}^{as} &= \frac{1}{2} \left(f_{m_x + d_x, m_y + d_y} - f_{m_x - d_x, m_y + d_y} + f_{m_x + d_x, m_y - d_y} - f_{m_x - d_x, m_y - d_y} \right) \\ f_{m_x m_y, d_x d_y}^{sa} &= \frac{1}{2} \left(f_{m_x + d_x, m_y + d_y} + f_{m_x - d_x, m_y + d_y} - f_{m_x + d_x, m_y - d_y} - f_{m_x - d_x, m_y - d_y} \right) \\ f_{m_x m_y, d_x d_y}^{aa} &= \frac{1}{2} \left(f_{m_x + d_x, m_y + d_y} - f_{m_x - d_x, m_y + d_y} - f_{m_x + d_x, m_y - d_y} + f_{m_x - d_x, m_y - d_y} \right) , \end{aligned} \quad (\text{A.104})$$

für $s_x, s_y = 0, \dots, D$. Die auf diese Weise konstruierten Parameter liegen alle im Intervall $[-1, 1]$. Zur Darstellung von Produkten und Potenzen dieser Parameter lässt sich wiederum die Multi-Index-Schreibweise nutzen:

$$\left(\mathbf{f}_{m_x m_y, D}^\sigma \right) \mathbf{e}^\sigma = \prod_{\substack{d_x, d_y \geq 0 \\ d_x + d_y \leq D}} \left(f_{m_x m_y, d_x d_y}^\sigma \right) e_{d_x d_y}^\sigma , \quad (\text{A.105})$$

wobei $\sigma \in \{ss, as, sa, aa\}$, $\mathbf{f}_{m_x m_y, D}^\sigma = (f_{m_x + d_x, m_y + d_y}^\sigma : d_x, d_y \geq 0, d_x + d_y \leq D)$, $\mathbf{e}^\sigma = (e_{d_x d_y}^\sigma : d_x, d_y \geq 0, d_x + d_y \leq D)$, $e_{d_x d_y}^\sigma \in \mathbb{N}$. In dieser Weise ist die Definition nur für $\sigma = ss$ sinnvoll. Für die anderen Symmetrisierungen muss jeweils der Fall $d_x = 0$ bzw. $d_y = 0$ ausgeschlossen werden, da immer

$$f_{m_x m_y, 0, d_y}^{as} = f_{m_x m_y, d_x, 0}^{sa} = f_{m_x m_y, 00}^{aa} = 0 \quad (\text{A.106})$$

gilt. Zur Vereinfachung der Notation wird auf eine explizite Darstellung dieser Ausnahmen verzichtet, sodass im Folgenden stets

$$e_{0, d_y}^{as} = e_{d_x, 0}^{sa} = e_{00}^{aa} = 0 \quad (\text{A.107})$$

mit der Konvention $0^0 = 1$ zu berücksichtigen ist. Zur Bestimmung der Parität wird wiederum über die Exponenten summiert, also z. B.

$$p(\mathbf{e}^\sigma) := \prod_{\substack{d_x, d_y \geq 0 \\ d_x + d_y \leq D}} (-1)^{e_{d_x d_y}^\sigma} = (-1)^{\sum_{d_x, d_y} e_{d_x d_y}^\sigma} . \quad (\text{A.108})$$

Für die Festlegung des Termgewichtes wird die Manhattan-Distanz verwendet:

$$w(\mathbf{e}^\sigma) := \sum_{\substack{d_x, d_y \geq 0 \\ d_x + d_y \leq D}} e_{d_x d_y}^\sigma (1 + d_x + d_y) . \quad (\text{A.109})$$

Bei Produkten aus Termen mit verschiedenen Symmetrisierungen multiplizieren sich die Paritäten, während sich die Termgewichte summieren. Zur Vereinfachung der Notation sei $\mathbf{e} =$

$(\mathbf{e}^{ss}, \mathbf{e}^{as}, \mathbf{e}^{sa}, \mathbf{e}^{aa})$, $p_x(\mathbf{e}) = p(\mathbf{e}^{as})p(\mathbf{e}^{aa})$, $p_y(\mathbf{e}) = p(\mathbf{e}^{sa})p(\mathbf{e}^{aa})$, $w(\mathbf{e}) = \sum_{\sigma} \mathbf{e}^{\sigma}$ und $\mathbf{f}_{m_x m_y, D} = (\mathbf{f}_{m_x m_y, D}^{\sigma} : \sigma \in \{ss, as, sa, aa\})$. Bei der Formulierung des linearen Modells muss nun nach der Parität in x - und y -Richtung unterschieden werden. Es werden alle Terme berücksichtigt, deren Manhattan-Distanz höchstens D und deren Termgewicht höchstens W beträgt:

$$F_{j_x j_y}(\mathbf{f}_{m_x m_y, D}) = \sum_{\substack{\mathbf{e}: p_x(\mathbf{e})=\pm 1 \\ p_y(\mathbf{e})=\pm 1 \\ w(\mathbf{e}) \leq W}} \beta \mathbf{e} \prod_{\sigma} (\mathbf{f}_{m_x m_y, D}^{\sigma})^{\mathbf{e}^{\sigma}}. \quad (\text{A.110})$$

Dabei sind $\mathbf{e}_{d_x d_y}^{\sigma} \in \mathbb{N}$ unter Beachtung der Einschränkung in Gl. (A.107). Das Vorzeichen für die Paritätsfunktionen muss dabei passend zu $F_{j_x j_y}$ gewählt werden.

A.2.3.2 Beispiele für Terme des linearen Modells

Tab. A.3: Terme mit gerader Parität (1D).

w/\hat{d}	0	1	2	3	4
0	1				
1	f_{0s}				
2	f_{0s}^2	f_{1s}			
3	f_{0s}^3	$f_{0s}f_{1s}$	f_{2s}		
4	f_{0s}^4	$f_{0s}^2 f_{1s}, f_{1s}^2, f_{1a}^2$	$f_{0s}f_{2s}$	f_{3s}	
5	f_{0s}^5	$f_{0s}^3 f_{1s}, f_{0s}f_{1s}^2, f_{0s}f_{1a}^2$	$f_{0s}^2 f_{2s}, f_{1s}f_{2s}, f_{1a}f_{2a}$	$f_{0s}f_{3s}$	f_{4s}

Tab. A.4: Terme mit ungerader Parität (1D).

w/\hat{d}	0	1	2	3	4
0					
1					
2		f_{1a}			
3		$f_{0s}f_{1a}$	f_{2a}		
4		$f_{0s}^2 f_{1a}, f_{1s}f_{1a}$	$f_{0s}f_{2a}$	f_{3a}	
5		$f_{0s}^3 f_{1a}, f_{0s}f_{1s}f_{1a}$	$f_{0s}^2 f_{2a}, f_{1s}f_{2a}, f_{1a}f_{2s}$	$f_{0s}f_{3a}$	f_{4a}

Welche Terme für das lineare Modell gemäß Gl. (4.14) bzw. (A.110) berücksichtigt werden, hängt von der Wahl der Größen D und W ab, sowie der jeweiligen Parität. Beispielhaft werden hier die zu berücksichtigenden Terme für kleine Werte von D und W angegeben. Um die Lesbarkeit zu erhöhen, werden die Indizes des zentralen Pixels (m, m_x, m_y) ausgelassen und die Paritäten in den unteren Index geschrieben:

$$f_{m, d}^{\sigma} \rightarrow f_{d\sigma}, \quad f_{m_x m_y, d_x d_y}^{\sigma} \rightarrow f_{d_x d_y \sigma}. \quad (\text{A.111})$$

Für den 1D-Fall sind die Terme in Tab. A.3 und A.4 angegeben, geordnet nach ihrem Termgewicht w und der maximalen Distanz \hat{d} der enthaltenen Parameter. Ein entsprechendes lineares Modell enthält dann alle Terme mit $w \leq W$ und $\hat{d} \leq D$.

Beim 2D-Fall muss die Parität jeweils nach x - und y -Richtung unterschieden werden. Die jeweils möglichen Kombinationen werden in Tab. A.5, A.6, A.7 und A.8 gezeigt.

Tab. A.5: Terme mit gerader/gerader Parität (2D).

w/\hat{d}	0	1	2	3
0	1			
1	f_{00ss}			
2	f_{00ss}^2	f_{01ss}, f_{10ss}		
3	f_{00ss}^3	$f_{00ss}f_{01ss}, f_{00ss}f_{10ss}$	$f_{02ss}, f_{11ss}, f_{20ss}$	
4	f_{00ss}^4	$f_{00ss}^2f_{01ss}, f_{00ss}^2f_{10ss}, f_{01ss}^2,$ $f_{01aa}^2, f_{10ss}^2, f_{10aa}^2, f_{01ss}f_{10ss}$	$f_{00ss}f_{02ss}, f_{00ss}f_{11ss},$ $f_{00ss}f_{20ss}$	$f_{03ss}, f_{12ss}, f_{21ss}, f_{30ss}$

Tab. A.6: Terme mit ungerader/gerader Parität (2D).

w/\hat{d}	0	1	2	3
0				
1				
2		f_{10as}		
3		$f_{00ss}f_{10as}$	f_{11as}, f_{20as}	
4		$f_{00ss}^2f_{10as}, f_{01ss}f_{10as}, f_{10ss}f_{10as}$	$f_{00ss}f_{11as}, f_{00ss}f_{20as}$	$f_{12as}, f_{21as}, f_{30as}$

Tab. A.7: Terme mit gerader/ungerader Parität (2D).

w/\hat{d}	0	1	2	3
0				
1				
2		f_{01sa}		
3		$f_{00ss}f_{01sa}$	f_{02sa}, f_{11sa}	
4		$f_{00ss}^2f_{01sa}, f_{01ss}f_{01sa}, f_{01sa}f_{10ss}$	$f_{00ss}f_{02sa}, f_{00ss}f_{11sa}$	$f_{03sa}, f_{12sa}, f_{21sa}$

A.2.4 Simulationsparameter

In Tab. A.9 sind die in den jeweiligen Anwendungsbeispielen verwendeten Simulationsparameter zusammengetragen. Für die Varianten BPM 1 und BPM 2 sei auf Tab. 3.1 verwiesen.

Tab. A.8: Terme mit ungerader/ungerader Parität (2D).

w/\hat{d}	0	1	2	3
0				
1				
2				
3			f_{11aa}	
4		$f_{01sa}f_{10as}$	$f_{00ss}f_{11aa}$	f_{12aa}, f_{21aa}

Tab. A.9: Simulationsparameter der Beispiele aus Kap. 4

Abschnitt	4.2.3	4.3.1/4.3.2 (PFA-Mod.)	4.3.1/4.3.2 (CGH-Sim.)	4.3.2	4.3.3	4.3.4	4.3.4
Verfahren	BPM 2	BPM 1	BPM 1	FDTD	FDTD	BPM 2	FDTD
λ_0 [nm]	500	633	532	532	1064	632,8	632,8
$\delta_x = \delta_y$ [nm]	$p_x/24$	10	20	20	30	32,5	32,5
δz [nm]	25	25	50	10	15	25	10

B Liste der wissenschaftlichen Veröffentlichungen und Vorträge

Artikel in Fachzeitschriften (peer-reviewed)

- S. Fasold, S. Linß, T. Kawde, M. Falkner, M. Decker, T. Pertsch und I. Staude, “Disorder-enabled pure chirality in bilayer plasmonic metasurfaces”, *ACS Photonics* **5**(5), 1773–1778 (2018).
- S. Linss, D. Michaelis und U. D. Zeitner, “Macroscopic wave-optical simulation of dielectric metasurfaces”, *Opt. Express* **29**(7), 10879–10892 (2021).
- S. Linss, D. Michaelis und U. D. Zeitner, “Ultrafast farfield simulation of non-paraxial computer generated holograms”, *Opt. Express* **30**(8), 13765–13775 (2022).
- S. Linß und U. D. Zeitner, “Design of computer-generated holograms based on semi-rigorous simulations of sub-wavelength structures”, *Opt. Express* **31**(7), 11089–11101 (2023).

Konferenzbeiträge und Vorträge

- S. Fasold, M. Falkner, T. Kawde, S. Linß, M. Decker, I. Staude und T. Pertsch, “Disorder-enabled pure circular dichroism in bilayer plasmonic metasurfaces”, in *2017 European Conference on Lasers and Electro-Optics and European Quantum Electronics Conference*, (Optica Publishing Group, 2017), S. CE_4_5.
- D. Arslan, S. Fasold, A. Rahimzadegan, T. Kawde, S. Linss, N. Abbasirad, M. Falkner, M. Decker, C. Rockstuhl, T. Pertsch und I. Staude, “Tailored structural disorder in optical metasurfaces”, in *CLEO Pacific Rim Conference 2018*, (Optica Publishing Group, 2018), S. W2H.2.
- S. Linß und U. D. Zeitner, “Wave-Optical Design of Wide-Angle Computer-Generated Holograms”, in *27th Microoptics Conference in Jena* (2022).
- S. Linß, D. Michaelis und U. D. Zeitner, “Wave-optical simulation of macroscopic dielectric metasurfaces”, in *Online Workshop on Structured Light and its Applications* (2022).

C Lebenslauf

Persönliche Daten

Name	Sebastian Linß
Geburtsdatum	25.08.1993
Geburtsort	Hansestadt Seehausen (Altmark)

Bildungsweg

08/00 - 07/04	Talschule Jena (Grundschule)
08/04 - 07/12	Carl-Zeiss-Gymnasium Jena
10/12 - 09/15	Physikstudium (B. Sc.), Friedrich-Schiller-Universität Jena
10/13 - 08/16	Mathematikstudium (B. Sc.), Friedrich-Schiller-Universität Jena
04/16 - 11/18	Physikstudium (M. Sc.), Friedrich-Schiller-Universität Jena
01/19 - 01/24	Promotion, Friedrich-Schiller-Universität Jena

Beruflicher Werdegang

10/13 - 10/16	Studentische/wissenschaftliche Hilfskraft Institut für Optik und Quantenelektronik, Jena bzw. Institut für Angewandte Physik, Jena
01/19 - 08/23	Wissenschaftlicher Mitarbeiter Institut für Angewandte Physik, Jena

D Danksagung

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die mich bei meiner Promotion unterstützt und zum Erfolg dieser Arbeit beigetragen haben.

Mein Dank gilt vor allem Prof. Uwe Zeitner für die Betreuung dieses interessanten Themas, für das entgegengebrachte Vertrauen zur eigenständigen Festlegung der inhaltlichen Schwerpunkte, für die anregenden wissenschaftlichen Diskussionen und für die Unterstützung zur Fertigstellung der schriftlichen Arbeit.

Besonders danke ich Dirk Michaelis für die inhaltliche Betreuung zu den Simulationsverfahren, die detaillierten Diskussionen zu den einzelnen Algorithmen und die Identifizierung der Zielstellung einer direkten Berechnung der optischen Funktion aus der Strukturgeometrie unter Beachtung der Wechselwirkung benachbarter Strukturen.

Ein weiterer großer Dank geht an meine Kollegen der Arbeitsgruppe Mikrostrukturtechnologie am Institut für Angewandte Physik in Jena, die mich durch die Herstellung und Charakterisierung meiner Proben unterstützt haben. Dabei möchte ich insbesondere Thomas Käsebier, Holger Schmidt, Natali Sergeev, Daniel Voigt und Werner Rockstroh hervorheben. Außerdem danke ich Thomas Siefke für die Unterstützung bei den Intensitätsmessungen und die Organisation der Probenherstellung sowie den anderen Doktoranden in der Arbeitsgruppe für den interessanten Austausch im Rahmen des Doktorandenseminars.

Ich bedanke mich auch herzlich bei Michael Banasch für die Datengenerierung zur Herstellung der Proben und das Bestreben, die Möglichkeiten des Elektronenstrahlschreibers voll auszunutzen, sowie bei Tino Benkenstein für die Effizienzmessung an den Strahlteilern.

Des Weiteren richtet sich mein Dank an Martin Heusinger für die Einführung in die Thematik des Streulichtes an Gittern und die Unterstützung bei der Anwendung der Simulationsverfahren auf diese Problematik.

Darüber hinaus möchte ich mich bei meinen Kollegen aus unserer Mittagsrunde, vor allem Johannes Hartung, Enrico Hilbert, Sebastian Merx und Johannes Stock bedanken. Unsere Gespräche und Diskussionen waren stets eine willkommene Abwechslung und mitunter wertvolle inhaltliche Anregung.

Schließlich danke ich ganz besonders meiner Frau Judith und meiner Familie für die Unterstützung während der Promotion sowie meiner Frau und meiner Mutter für das Korrekturlesen dieser Arbeit.

E Ehrenwörtliche Erklärung

Hiermit erkläre ich ehrenwörtlich, dass ich die Dissertation selbst angefertigt, keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben habe.

Bei der Auswahl und Auswertung folgenden Materials haben mich die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unterstützt:

- Prof. Dr. Uwe Detlef Zeitner als betreuender Hochschullehrer,
- Dr. Dirk Michaelis (IOF Jena) bei der Modifikation der BPM und als Ideengeber für die optische Analyse der Effektiv-Index-Strukturen,
- die Arbeitsgruppe Mikrostrukturtechnologie des IAP Jena durch die Herstellung der Proben, im Detail Daniel Voigt durch Chrombeschichtung und -ätzen, Michael Banasch (IOF Jena) durch Datengenerierung für die Elektronenstrahlolithografie, Holger Schmidt durch deren Durchführung, Thomas Käsebier durch Quarzglasätzen und Natali Sergeev durch REM-Aufnahmen,
- Tino Benkenstein (IOF Jena) durch die Intensitätsmessung am Aufbau M1,
- Dr. Thomas Siefke (IAP Jena) beim Aufbau des Messplatzes M2,
- Dr. Martin Heusinger (IOF Jena) durch die Bereitstellung von Strukturdaten und RCWA-Simulationen für die Streulichtanalyse.

Die Hilfe einer kommerziellen Promotionsvermittlerin/eines kommerziellen Promotionsvermittlers wurde nicht in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Dissertation wurde noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung wurde bei keiner anderen Hochschule als Dissertation eingereicht.

Die geltende Promotionsordnung der Physikalisch-Astronomischen Fakultät ist mir bekannt.

Ort, Datum

Sebastian Linß