

Entwicklung und Verifikation von Algorithmen zur Vereinigung von Methoden der Wellenoptik und des Raytracing

Masterarbeit

an der Fakultät für Elektrotechnik und Informationstechnik

von Tobias Lange

vorgelegt am 29. Januar 2021

Matrikelnummer:	209060
Studiengang:	Elektrotechnik und Informationstechnik
Erstgutachter:	Prof. Dr.-Ing. Dirk Schulz
Zweitgutachter:	M.Sc. Lukas Schulz
betreuendes Institut:	Lehrstuhl für Technologie Optischer Systeme TOS der RWTH Aachen
betreut und koordiniert von:	Dr.-Ing. Annika Völl

Inhalt und Ergebnisse dieser Arbeit sind ausschließlich zum internen Gebrauch bestimmt. Alle Urheberrechte liegen beim Lehrstuhl TOS der RWTH Aachen. Ohne ausdrückliche Genehmigung des Lehrstuhls TOS der RWTH Aachen ist es nicht gestattet, diese Arbeit oder Teile an Dritte weiterzugeben.

Zusammenfassung

Angetrieben von der stetigen Miniaturisierung der Halbleitertechnologie stehen seit einigen Jahren Herstellungsverfahren zur Verfügung, um Oberflächen diverser Materialien in der Größenordnung der sichtbaren Wellenlängen und darunter zu strukturieren. Dies eröffnet ein neues Feld an optischen Bauelementen unter dem Oberbegriff der diffraktiven Optiken, welche sich durch geschickte Auslegung wellenoptische Effekte wie die Beugung zu Nutzen machen und so bisher nicht realisierbare optische Abbildungen induzieren können. Die Herausforderungen dieser Optiken liegen, neben den Herstellungverfahren, in der Auslegung, da zur Bestimmung des Lichtverhaltens an Strukturen in der Größenordnung der Arbeitswellenlänge vollständige elektromagnetische Feldansätze nötig sind. Diese Feldansätze sind in der Simulation maßgeblich rechenintensiver als Simulationsverfahren herkömmlicher Optiken und skalieren bezüglich des Rechenaufwandes zum Teil in der vierten oder fünften Potenz mit der Größe des optischen Bauelementes, was zu Einschränkungen in der simulativen Auslegung von diffraktiven Optiken führt.

Diese Arbeit leistet deshalb einen Beitrag zur numerischen Simulation von Feldern an nanostrukturierten Oberflächen. Mit diesem Ziel wird die vielfach angewandte Rigorous Coupled Wave Analysis (RCWA) in der Form eines Softwarepaketes um die Normalenvektortransformation (NVT) erweitert. Dies optimiert die Konvergenzgeschwindigkeit für beliebig metallisch strukturierte zweidimensional periodische Nanostrukturen maßgeblich. So werden Simulationen über mehrere parallele Eingangsparameterbereiche auf Desktop - Rechenumgebungen möglich und handhabbar.

Ferner wird in dieser Arbeit der neuartige Ansatz der Feldbildsynthese zur Modellbetrachtung der Teilausleuchtung entwickelt. In diesem werden die Beschränkungen der RCWA von der bisher möglichen ebenen Eingangswelle hin zu einer inhomogenen Eingangsbeleuchtung sowie die Annahme der unendlichen periodischen Fortsetzung der Struktur aufgehoben. Der Mehrgewinn liegt hierbei insbesondere in der Anwendung dieser Ansatzes bei der Auslegung von optischen Systemen, in welchen Laser, LEDs oder Projektoren als Lichtquellen Eingangsbeleuchtungen mit endlicher Ausdehnung und verschiedensten Strahlprofilen emittieren. Auf diese Weise wird unter anderem die rigorose Simulation von fokussierten Lichtfeldern in diffraktiven Optiken wie Metalinsen, diffraktiven Linsen und planartechnisch strukturierten Optiken ermöglicht. Indem bei diesem Verfahren die maximale laterale Ausdehnung der zu simulierenden Struktur bis in den Millimeterbereich ragt, ist es vergleichbaren Verfahren mit maximalen Ausdehnungen im mittlerem Mikrometerbereich bei gleicher Rechenkonfiguration überlegen.

Ein weiterer Teil der Arbeit beschäftigt sich mit der Integration der RCWA in das Prinzip des Raytracing. Hierzu wird eine Erweiterung für Physically Based Rendering Technique (pbrt-v3) implementiert. Mittels neu entwickelter Samplingfunktionen und Materialklassen können beliebig strukturierte periodische 2D-Bauelemente unter Berücksichtigung einer vollständigen Wellensimulation in einer virtuellen Szenerie photorealistisch visualisiert werden. Die Methoden dieser Arbeit sollen dabei unterstützen, die Entwicklungsschritte Auslegung, Verifizierung und Testung neuartiger optischer Bauelemente im Interesse von Kostenminimierung und funktioneller Möglichkeiten in den virtuellen Bereich zu verlagern.

Inhaltsverzeichnis

1 Einführung	4
1.1 Motivation	4
1.2 Zielsetzung	6
1.3 Lösungsansatz	6
1.4 Vorbemerkung	7
2 Grundlagen der wellenoptischen Simulation	8
2.1 Optiken mit reiner Lichtbrechung	8
2.2 FDTD	9
2.3 RCWA	11
2.3.1 Normalenvektortransformation	12
2.4 Teilausleuchtung	14
2.5 Raytracing	16
2.5.1 Grundkonzepte des Raytracing	18
2.6 Fazit	19
3 Methodik	21
3.1 Konzept	21
3.2 Normalenvektortransformation	22
3.3 Teilausleuchtung	23
3.4 Raytracing	23
3.5 Verifikation und Vergleich	24
3.6 Bemerkung zur Implementierung	25
4 RCWA - Normalenvektortransformation	26
4.0.1 Formulierung	26
4.0.2 Konvergenz und Verifikation	28
4.0.3 Vergleich mit FDTD	30
4.1 Fazit	31
5 Teilausleuchtung	33
5.1 Modellbildung der Teilausleuchtung	33
5.1.1 Teilausleuchtung mit Fouriermethode	35
5.2 Teilausleuchtung mit Feldbildsynthese	37
5.3 Symmetrie der Struktur	39
5.4 Verifikationsmethoden der Teilausleuchtung	41
5.5 Vergleich und Verifikation mit den vorgestellten Methoden	42
5.6 Fazit	51
6 Raytracing	52
6.1 Grundverfahren der ebenen Welle	52
6.2 Vergleich der Testimplementierung und der Cornellbox	56
6.3 Einbau des ebenen Wellenkonzeptes in pbrt-v3	57
6.4 Fazit	63
7 Zusammenfassung und Ausblick	64
7.1 Ausblick	65

8 Literaturverzeichnis und weitere Quellen	66
9 Appendix A: Einstellung der FDTD Simulation	69
10 Appendix B: Formulierung der allgemeinen Reihenlösungen für die 2D-RCWA	72
11 Appendix C: Formulierung der Gaußschen Grundmode als Eingangsbeleuchtung	74
12 Appendix D: Die C++ Implementierung der RCWA Samplingfunktion in pbrt-v3	76

1 Einführung

Seit dem Ursprung der Halbleitertechnik um 1950 existiert die Motivation, eine stetige Verkleinerung von Transistoren auf Computerchips zu realisieren. So können mehr Bauelemente auf der gleichen Chipfläche angeordnet und mehr Berechnungsoperationen in gleicher Zeit durchgeführt werden. Möglich geworden ist dies durch geeignete Herstellungsverfahren unter dem Oberbegriff der Planartechnik. Dank einzelner Verfahren wie der EUV-Belichtung, Molekularstrahlepitaxie und Ionenimplantation sind so Strukturen mit Einzelementen in der Größenordnung der Wellenlänge des sichtbaren Lichtes (400nm - 700nm) und darunter formbar. Dies eröffnete neuartige Konzepte, diese Produktionsmöglichkeiten auch in der Herstellung von Optiken unter dem Oberbegriff der diffraktiven Optiken (Bauelemente mit Oberflächenstrukturierungen in der Größenordnung des sichtbaren Wellenlängenbereichs) einzusetzen. Hieraus ergeben sich verschiedenste Vorteile. Beispielsweise kann die Displayfokussierung mit Hilfe einer konventionellen Linse durch einen dünnen Mikrolinsenfilm direkt auf dem Display platzsparend ersetzt werden. Ferner treten bei Strukturierungen in der genannten Größenordnung der Wellenlänge Beugungseffekte des Lichtes auf. Diese können durch geschickte Auslegung beispielsweise zur Strahlfokussierung oder -abbildung genutzt werden. Einige dieser Effekte sind mit herkömmlichen Strahlengangsoptiken nicht erzeugbar sowie durch herkömmliche Optiksimulationen wie Raytracing (Lichtstrahlen werden als einzelne Geraden im Raum betrachtet, die geometrische Körper einer Szenerie schneiden) nicht simulierbar.

1.1 Motivation

In den letzten drei Jahren sind zur Klasse der diffraktiven Bauelemente noch die Gruppe der Metalinsen hinzugekommen, da das zugehörige Nanoimprintverfahren zur Herstellung geeignet optimiert worden ist. Ihre Auslegung ist eine der Motivationen dieser Arbeit. Bei einer Metalinse sind Einheitsgeometrien wie Zylinder oder Rechtecke aus Oxiden in periodischen Zellen auf einem Glaswafer aufgebracht. Über Parameter wie der Höhe eines Zylinders oder dem Rotationswinkel eines Blockes induziert so beispielsweise jeder Zellentyp eine zur Höhe proportionale Phasenverschiebung einer auf die Linse einfallenden ebenen Welle. Da die Zellen in ihren Abmessungen kleiner als die der einfallenden Wellenlänge sind, können, falls ersucht, Beugungseffekte umgangen werden. Mit einer geeigneten Phasenfunktion, die jeder Zelle auf der Linse die zugehörige Einheitsgeometrie zuweist, kann beispielsweise im einfachsten Fall so eine Linse hergestellt werden, die mit hoher numerischer Apertur auf einen Brennpunkt fokussiert [6]. Die Möglichkeit der beliebigen Phaseninduktion an diskreten Punkten mit Abständen kleiner als der Wellenlänge eröffnet völlig neue Möglichkeiten in Bezug auf optische Abbildungen wie der Bildeinkopplung in der Augmented Reality [7]. Das Kernproblem bei der Auslegung dieses Linsentyps ist jedoch die Simulation zur Verifikation und Testung der Funktionalität des Bauelementes. Aufgrund der Größenordnung der Einzelemente dieses Linsentyps sind hierfür Simulationsverfahren mit feldtheoretischem Ansatz wie beispielsweise der Finite Difference Time Domain (FDTD) nötig. Problematisch ist hierbei, dass bei diesem Simulationstyp der Rechenaufwand erheblich mit der Bauelementgröße skaliert. Konkret sind so beispielsweise in [6] lediglich Simulationen der Beleuchtung einer Metalinse bis zu einem Durchmesser von $24\mu\text{m}$ möglich. Somit sind rein konzeptionelle Untersuchungen von skalierungsunabhängigen optischen Abbildungen dieses Linsentyps wie der Fokussierung auf einen Brennpunkt möglich. Komplexere optische Abbildungen bei verschiedenen inhomogenen Eingangsbeleuchtungen, die aufgrund der beliebig ortsabhängigen

Phaseninduktion realisiert werden könnten, sind jedoch nicht konzeptionell herab skalierbar. Somit stehen auch bisher nur unzureichend Möglichkeiten für die virtuelle Auslegung, Verifikation und Testung solcher neuen Metalinsenkonzepte oder anderer diffraktiver Baugruppen zur Verfügung.

Eine weitere Motivation dieser Arbeit ist der gegenwärtige Entwicklungsstand bezüglich periodischer planartechnisch hergestellter optischer Bauelemente wie Folierungen oder Gitter. In den letzten Jahren sind in der Entwicklung des Raytracing Fortschritte derart gemacht worden, dass inzwischen Bilder von detailreichen Szenerien berechnet werden können, worin der Unterschied zu einer Fotografie der realen Szenerie nicht mehr erkennbar ist. Für die Simulation von herkömmlichen Glasoptiken wird das Raytracing schon eingesetzt, um die optische Interaktion des Bauteils mit einer Szenerie zu visualisieren. Bisher ist es für die genannten periodischen Bauelemente zwar möglich, beispielsweise mit einem rigorosen Feldansatz wie dem der Rigorous Coupled Wave Analysis (RCWA) Parameter wie Absorptionsgrade oder die Abhängigkeit der Lichtintensität über den Raumwinkel zu analysieren. Es existiert jedoch die Überlegung, ob es möglich wäre, die mikroskopische Simulation von periodischen optischen Strukturen beispielsweise durch die RCWA mit dem Raytracing zu vereinen.

Abbildung(1) zeigt diese prinzipielle Überlegung.



Abbildung 1: Fotografie als eine der Motivationen dieser Arbeit: Wie interagiert ein diffraktives optisches Bauelement, hier exemplarisch von einer CD repräsentiert, mit seiner Umgebung?

Es sei ein beliebiges diffraktives Element, hier exemplarisch eine CD zur Verifikation, nebst weiteren nicht diffraktiven Elementen und einer Beleuchtung in einer Box positioniert. Die Idee ist, mit einer Simulationssoftware diese gesamte Szenerie bezüglich des Kamerapunktes inklusive der wellenoptischen Effekte der CD zu visualisieren. So sollen die sichtbaren Licht- und Farbartefakte, hier vor allem die Farbschemen, noch vor der eigentlichen Herstellung erkennbar sei. So sind Informationen zu parasitären Effekten und der gesamten optischen Interaktion, die aus den Ergebnissen reiner Feldsimulatoren nicht hervorgehen, direkt visuell evident.

Die Motivation dieser Idee ist durch die eingangs erwähnten Herstellungsverfahren der Planartechnik gegeben. In diesen Verfahren stellen üblicherweise die Maskierungen bzw. im Nanoimprintverfahren der SiO_2 -Stempel als Einzelanfertigungen die teuersten Komponenten der Herstellung dar. Die Motivation der Abwicklung von Auslegung, Verifikation und Tes-

tung der optischen Bauelemente im virtuellen Bereich noch vor der Herstellung der ersten Optik ist somit ersichtlich.

1.2 Zielsetzung

Die Zielsetzung dieser Arbeit ist erstens, die bestehende Rigorous Coupled Wave Analysis (RCWA) zu erweitern, um Optimierungen bezüglich der Konvergenz und Rechengeschwindigkeit bei der Simulation von 2D-periodischen Zellen zu erlangen.

Zweitens soll die Modellbetrachtung der Teilausleuchtung, bei der eine beliebige inhomogene Eingangsverteilung eine räumlich begrenzt periodische Struktur beleuchtet, um ein neues Konzept erweitert werden, in welchem die maximale laterale Ausdehnung der zu simulierenden Struktur im Millimeterbereich liegt und somit der realen Größe des Bauelementes entspricht.

Drittens soll die Möglichkeit einer Integration der RCWA in das Prinzip des Raytracing gefunden werden, um so beliebig strukturierte periodische 2D-Bauelemente unter Berücksichtigung einer vollständigen Wellensimulation in einer virtuellen Szenerie photorealistisch zu visualisieren.

Alle entwickelten Methoden gilt es durch Referenzverfahren zu bestätigen und mit der Perspektive des Einsatzes dieser Methoden in der Auslegung von diffraktiven Optiken zu verifizieren.

1.3 Lösungsansatz

Um die Zielsetzung dieser Arbeit zu erreichen, ist folgendes Vorgehen festgelegt.

Kapitel 2: Stand der Technik

Es gilt den gegenwärtigen Stand der Technik bezüglich der Auslegung von diffraktiven Optiken und der hierfür verfügbaren Verfahren zu ermitteln. So sollen konkrete Ansatzpunkte für die eigene Entwicklungsarbeit sowie Defizite zwischen dem gegenwärtigen Forschungsstand und der hier definierten Zielsetzung gefunden werden.

Kapitel 3: Methodik

Basierend auf den ermittelten Forschungsdefiziten und relevanten Vorarbeiten ist eine genaue Methodik zu erstellen, mit der das Ziel dieser Arbeit erreicht werden kann. Ein besonderer Fokus liegt hierbei auf einer geeigneten Methodik zur Verifikation der neu entwickelten Verfahren, um diese in zukünftigen Anwendungen zuverlässig einsetzen zu können.

Kapitel 4: RCWA Erweiterung

In diesem Kapitel sollen die gesuchten Optimierungen des bestehenden 3D-RCWA Paketes implementiert und verifiziert werden.

Kapitel 5: Teilausleuchtung

Hier gilt es die Modellbetrachtung der Teilausleuchtung mit einem neuen Verfahren zu erweitern, sodass diffraktive Optiken in ihrer tatsächlichen Baugröße wellenoptisch simuliert werden können.

Kapitel 6: RCWA-Raytracing Kombination

Es ist die Erweiterung eines dem aktuellen Stand der Technik entsprechenden Raytracers (pbrt-v3 [25]) zu entwickeln, sodass hierin diffraktive Strukturen in einer Szenerie mit simuliert werden.

Kapitel 7: Validierung der Arbeit

Zuletzt muss rekapituliert werden, ob die Zielsetzung dieser Arbeit erreicht worden ist, ob die durchgeführten Verifikationen eine zukünftige Anwendung zulassen und welche weiterführenden Optimierungen oder Methoden motiviert werden.

1.4 Vorbemerkung

An dieser Stelle sei kurz darauf hingewiesen, dass in dieser Arbeit eine einheitliche Nomenklatur bezüglich optischer Strahlen verwendet wird, um die Verfahren klar abzugrenzen. Im Folgenden bezeichnen Wellen ebene homogene Wellen, die durch einen kartesischen Wellenvektor und eine komplexe Amplitude voll definiert sind. Im Bereich des rigorosen Feldansatzes besteht das Ausgangsfeld aus mehreren solcher ebenen homogenen Wellen, die hier auch als Beugungsordnungen bezeichnet werden. Strahlen beschreiben in dieser Arbeit Feldverteilungen in der Art beispielsweise eines Lichtkegels oder einer Gaußschen Grundmode, welche eine endliche und inhomogene Amplitudenverteilung und eine inhomogene Phasenfront aufweisen. Rays bezeichnen in dieser Arbeit Geraden im Raum und beziehen sich auf das Raytracing. Sie sind durch einen Basis- sowie einen Richtungsvektor definiert und ihnen kann ein Farbwert, eine Intensität oder ein Wellenlängenspektrum der Leuchtdichte zugeordnet werden.

2 Grundlagen der wellenoptischen Simulation

Mit der im Vorkapitel definierten Zielsetzung sowie der einleitenden Motivation gilt es nun zu untersuchen, wie bis zum heutigen Stand in Forschung und Entwicklung von diffraktiven Optiken vorgegangen wird und welche Forschungsdefizite hierbei zur eigenen Zielsetzung bestehen. Es sind schon bestehende Methoden zur Simulation diffraktiver Strukturen zu ermitteln, welche dann im Verlauf dieser Arbeit um eigene Erweiterungen zu ergänzen sind. In [23] wird ein erster Überblick über diese Situation gegeben. Als weit verbreitete Berechnungsmethode wird die Finite Difference Time Domain (FDTD) seit Jahrzehnten in der Hochfrequenztechnik, Optik und Photonik eingesetzt. Aufgrund des daraus folgenden Entwicklungsstandes eignet sie sich somit auch zur Verifikation neuer Methoden dieser Teilgebiete. Aufgrund ihres iterativen Verfahrens skalieren bei der FDTD jedoch Simulationszeit und Rechenaufwand mit der vierten bzw. fünften Potenz mit den Abmessungen des zu simulierenden Bauteils. Die Grenzen dieses Verfahrens liegen beispielsweise bei der Simulation von Metalinsen bei $24\mu\text{m}$ Durchmesser der Linse [6]. Daraus folgend ist die Simulation von im Millimeterbereich ausgedehnten wellenoptischen Strukturen mit der FDTD nicht möglich, was die Verifikation nur auf konzeptionelle Untersuchungen beschränkt. Alternativ hat sich seit einem Jahrzehnt in den Forschungsgebieten der Photonik und Optik die bereits erwähnte RCWA etabliert. Sie folgt einem rigorosen Lösungsansatz, in welchem gemäß der Fouriertheorie allgemeine Lösungen der Maxwellgleichungen im Voraus definiert werden. Über die Lösung eines Eigenwertproblems aus diesen Differentialgleichungen und Ausnutzung der Stetigkeit von elektromagnetischen Feldern an Grenzflächen werden diese allgemeinen Lösungen mit den zugehörigen Koeffizienten bestückt. Wie schon im Eingangskapitel erwähnt, existiert mit Hilfe der Fourier Modal Method [30], in dieser Arbeit auch als Fouriermethode bezeichnet, die Möglichkeit, eine räumlich begrenzte Struktur bei inhomogener Teilausleuchtung zu simulieren. Die lateralen Grenzen der zu simulierenden Struktur liegen hierbei in der Größenordnung von $100\mu\text{m}$, sodass ebenfalls eine Optimierung dieses Maximums für die Teilausleuchtung geeignet erscheint.

Parallel ist als Zielsetzung dieser Arbeit definiert, periodische diffraktive Strukturen in die Visualisierung eines Raytracers zu integrieren. Als aktueller Stand der Technik bezüglich des Raytracings gilt allgemein Physically Based Rendering Technique (pbrt-v3 [25]). Für die Integration einer diffraktiven Struktur existieren schon veröffentlichte Ansätze [24] [21] [38]. Hierbei werden jedoch anstatt einer wellenoptischen Simulation Ortsfunktionen für eine Oberfläche definiert und anschließend mit Hilfe der optischen Gittergleichung rudimentäre Beugungseffekte simuliert. Die Resultate dieser Ansätze genügen jedoch den Ansprüchen an die photorealistische Visualisierung dieser Strukturen im Raytracing nicht.

Im folgenden Kapitel werden vier der Kernverfahren aus Hochfrequenztechnik, Photonik und Optik jeweils bis zu ihrem heutigen Entwicklungsstand vorgestellt. Ferner werden jeweils bestehende Anwendungsszenarien aus bestehenden Veröffentlichungen gezeigt und parallel dazu die Defizite der verwendeten Verfahren bezüglich des Ziels dieser Arbeit identifiziert. Zum Abschluss des Kapitels werden diese Defizite noch einmal in Kurzfassung rekapituliert, um darauf basierend im Folgekapitel 3 eine geeignete Methodik für deren Behebung zu erstellen.

2.1 Optiken mit reiner Lichtbrechung

Zur Einführung in die Funktion und die Auslegung von optischen Bauelementen sei deren einfachste Baugruppe, die der optischen Linsen, betrachtet. Grundsätzlich sind diese aus

einem optischen Glas oder Polymer hergestellt. Schon seit über einem Jahrhundert wird hierbei das Material grob in Form gesägt und dann die endgültige Oberflächenform in mehreren Arbeitsschritten mit immer feinerem Korn geschliffen. Aufgrund dieser Bearbeitung sind optische Linsen üblicherweise rotationssymmetrisch und lassen sich über einen konstanten Radius beschreiben. Linsen dieser Art bieten die Möglichkeit, einfallendes Licht auf einen Brennpunkt zu fokussieren oder den Strahlengang umgekehrt aufzuweiten. Mit diesem Prinzip lassen sich so beispielsweise Teleskope oder Mikroskope herstellen. In beiden exemplarischen Anwendungen tritt jedoch beispielsweise der parasitäre Effekt der sphärischen Aberration auf. Aufgrund des Abstandes zur optischen Achse und der daraus folgenden ungleichen Schnittweite werden einfallende parallele Strahlen nur unscharf auf den gemeinsamen Brennpunkt gelenkt. Ein erster Ansatz, dieses Problem zu beheben, sind Asphären. Sie sind ebenfalls rotationssymmetrisch konstruiert, weisen jedoch an jedem Punkt ihrer Oberfläche einen unterschiedlichen Radius auf. Dennoch sind auch hier der möglichen optischen Abbildung aufgrund der Rotationssymmetrie Grenzen gesetzt. An diesem Punkt stellen Freiformlinsen den gegenwärtigen Entwicklungsstand dar. Hierbei folgt die Oberfläche dieses Linsentyps keiner Art von Symmetrie, sondern ist beliebig ausgelegt und im Regelfall nur durch mehrdimensionale Polynomgitter (Splines bzw. Nurbs) im Raum beschreibbar. Da jedoch auch dieser Linsentyp physikalisch nur dem Brechungsgesetz folgt, können alle der in diesem Abschnitt betrachteten Linsentypen durch Verfahren wie dem Raytracing vollständig simuliert werden. Beispielsweise ist es in dem Raytracingpaket Physically Based Rendering Technique [25] möglich, eine solche durch Splines beschriebene Freiform einzulesen und in einer Anwendungsszenerie mitsamt mehreren Lichtquellen und anderen Objekten photorealistisch zu visualisieren. Somit ist das Auslegungsziel dieser Arbeit für die Obergruppe der konventionellen Linsen schon erreicht worden.

2.2 FDTD

Als eine der Kernmethoden zur Simulation von elektromagnetischen Feldern an mikro- und nanostrukturierten Bauelementen gilt die Finite Difference Time Domain (FDTD). Ihr Grundprinzip ist in Abbildung(2) veranschaulicht.

In der FDTD wird der Simulationsraum durch ein Meshgitter, dem sogenannten Yee-Grid, diskretisiert. Bauelemente und Schichten werden dahingehend definiert, dass für jeden Diskretisierungspunkt ein Brechungsindex bzw. eine Permittivität angegeben wird. Ferner kann in einer beliebig positionierten Fläche eine explizite Verteilung als Eingangsfeld definiert werden. Nach Abschluss der Simulation lässt sich nun in der beispielhaft gezeigten Monitorebene das resultierende transmittierte Feld anzeigen. Somit kann mit dieser Simulationsmethode eine inhomogene Eingangsbeleuchtung auf ein beliebiges Bauelement betrachtet werden. Den Kern der FDTD bilden die Maxwellgleichungen. Sie sind an dieser Stelle auch in Bezug auf die weiteren im Laufe dieses Kapitels vorgestellten Verfahren in Gleichung(1) in differenzierlicher Schreibweise angegeben. Hierbei wird sowohl von der Abwesenheit von Stromdichten ausgingen als auch dass alle Materialeigenschaften durch die komplexe Permittivität $\epsilon = \epsilon_0 n^2$ bzw. Permeabilität vollständig repräsentiert werden. Ferner sei angemerkt, dass diese beiden Parameter insbesondere bei den Materialien, wie sie in dieser Arbeit zum Einsatz kommen, auch frequenzabhängig (Materialdispersion) sein können:

$$\vec{H} = \left(\frac{j}{\omega \mu} \right) \vec{\nabla} \times \vec{E} \quad \vec{E} = \left(\frac{-j}{\omega \epsilon_0 n^2} \right) \vec{\nabla} \times \vec{H} \quad (1)$$

Im Kontext dieser Arbeit werden nur Materialien mit einer Permeabilität von 1 betrachtet. Ferner sei hier die komplexe Einheit mit j bezeichnet. Der Nabla Operator $\vec{\nabla}$ wird über die

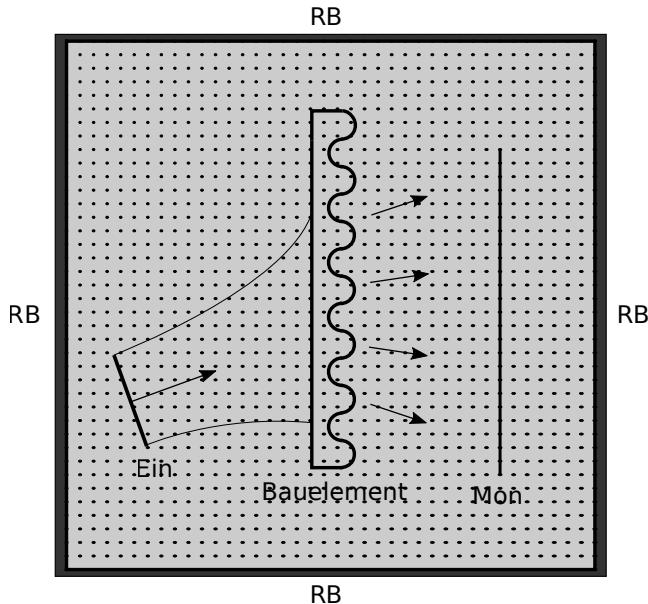


Abbildung 2: Grundkonzept der FDTD-Simulation: Der Simulationsbereich, eingegrenzt von Randbedingungen RB, wird durch ein Raumgitter diskretisiert, sodass anschließend die Maxwellgleichungen iterativ an jedem Knoten gelöst werden.

Rotation in Form des Kreuzproduktes auf das elektrische \vec{E} und das magnetische \vec{H} Feld angewandt. Der Parameter ω repräsentiert hierbei als Kreisfrequenz die Zeitabhängigkeit der periodischen Felder. Im FDTD Algorithmus werden diese beiden Gleichungen über die Formulierung einer Taylorreihe in Raum und Zeit diskretisiert. Anschließend werden diese Gleichungen solange iterativ für jeden Gitterpunkt berechnet, bis der Einschwingvorgang abgeklungen und die Abweichung zwischen den Diskretisierungspunkten einen Schwellenwert unterschritten hat. Für weiterführende Betrachtungen zur genauen Herleitung und weiteren Modalitäten der FDTD sei an dieser Stelle auf [23] verwiesen.

Da die FDTD seit Jahrzehnten in einem breiten Feld von Anwendungen der Hochfrequenztechnik und Photonik entwickelt und eingesetzt wird, existieren bereits einige kommerzielle sowie frei verfügbare Implementierungen dieses Algorithmus. Mittels der vorgestellten Methoden in Veröffentlichungen wie [8] und [15] können bei der Verwendung von beliebigen Materialien, Bauformen und Propagationsrichtungen der Felder konvergente Lösungen erzielt werden. Aus diesen Gründen eignet sich die FDTD besonders, andere neue Methoden der Optik und Photonik zu verifizieren.

In Bezug auf die Anwendung der FDTD in dieser Arbeit ist jedoch die relative Systemgröße dieses Verfahrens problematisch. Grundsätzlich skaliert bei der FDTD die Rechenzeit $\sim x^4$ und der Speicherbedarf $\sim x^3$ mit x als der Kantenlänge des Simulationsbereiches. Laut [23] ist dies das Hauptdefizit der FDTD bezüglich der Simulation von diffraktiven Optiken, da im besten Fall die Größe des Simulationsraumes den zweistelligen Mikrometerbereich nicht übersteigt. Beispielsweise bei der Auslegung von Metalinsen in [6] sind mit dem maximal möglichen Durchmesser der Linse von $24\mu m$ in der FDTD nur konzeptionelle Untersuchungen bei einer ebenen Welle als Feldeingang und der Fokussierung auf einen Brennpunkt als optische Abbildung der Linse möglich.

2.3 RCWA

Alternativ zu iterativen Verfahren wie der FDTD hat sich die Rigorous Coupled Wave Analysis (RCWA) mit ihrem rigorosen Lösungsansatz in den Entwicklungsgebieten der Photonik und Nanooptik etabliert. Wie im Einleitungskapitel erwähnt dient sie ferner als Ausgangspunkt dieser Arbeit. Die Abbildung(3) veranschaulicht das Grundprinzip der RCWA.

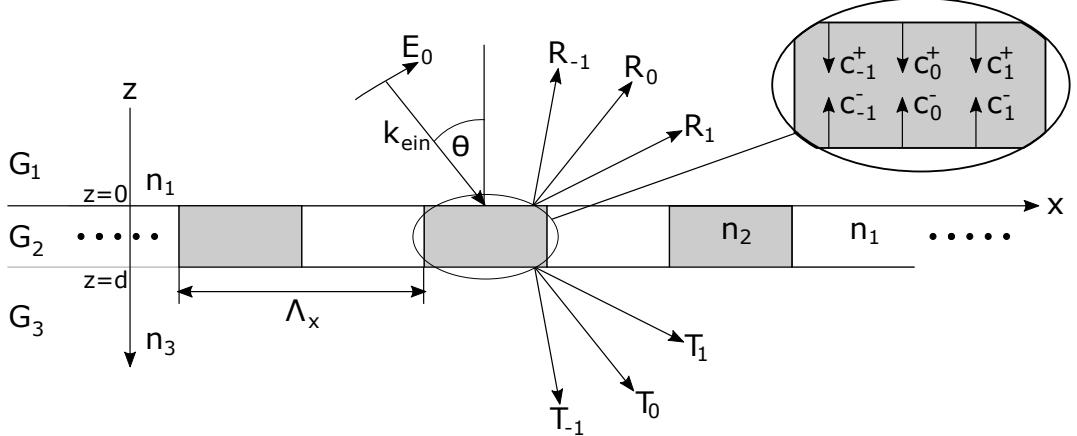


Abbildung 3: Modellbeschreibung der RCWA: Eine ebene homogene Welle trifft auf eine periodische Struktur und erzeugt Beugungsordnungen als reflektierte bzw. transmittierte ebene Wellen

In der Modellbeschreibung der RCWA sei eine in x-Richtung periodisch strukturierte Schicht G_2 gegeben, wobei sich die einzelnen Zellen entlang der x-Achse unendlich oft mit der Periode Λ_x wiederholen. In G_2 werden allgemeine Fourierreihen formuliert, die dann über die Lösung eines Eigenwertproblems und dem Gleichsetzen der Felder an Grenzflächen mit eindeutigen Koeffizienten bestückt werden und das Feld als Überlagerung ebener Wellen der Amplituden c_i im gesamten Simulationsgebiet angeben. Als Eingang dieser Simulation dient eine ebene homogene Welle beliebigen Eingangswinkel (azimutal und polar), einer Wellenlänge sowie beliebiger Polarisation, welche durch die Winkel ψ und δ beschrieben werden. Die Welle trifft auf eine periodische Zelle, die in allen 3 Raumrichtungen in Bezug auf ihre inneren Geometrien und die verwendeten Materialien beliebig definierbar ist. Die Grundannahme der RCWA ist hierbei, dass sich die betrachtete Zelle unendlich oft lateral periodisch fortsetzt. Eine detaillierte Ansicht dieses Modells inklusive aller Eingangsparameter ist in Abbildung(32) in Appendix 10 gezeigt.

Das Grundprinzip der RCWA wurde erstmals in [19] aufgestellt. In den folgenden Jahren ist diese Methode immer weiter erweitert worden, beispielsweise durch einen Mehrschichtenansatz [18], um eine Variation der Struktur in z-Richtung zu berücksichtigen. Insbesondere relevant für diese Arbeit sind die Erweiterung zum 2D-periodischen Fall [9] sowie die Behebung eines Konvergenzproblems in [10]. Ferner existiert für den 2D-periodischen Fall auch die Erweiterung zur Normalenvektortransformation (NVT), wie sie beispielsweise in [33] vorgestellt wird. Mit diesem lässt sich die Konvergenz der RCWA erheblich optimieren. Sie bietet die Möglichkeit, die Modifikationen zur Behebung des Konvergenzproblems in [10] auf die beiden periodischen Richtungen ideal anzupassen. Der Vorteil der RCWA in Bezug auf die Simulation diffraktiver Elemente ist, dass der Rechenaufwand im Vergleich zu anderen iterativen Verfahren wie der FDTD oder der Beam Propagation Method (BPM) erheblich

kleiner ausfällt, obwohl sie keinem Näherungsverfahren entspricht, sondern einen vollständig gelösten Feldansatz repräsentiert. Die Systemgröße erleichtert gerade Simulationen über Eingangsparameterbereiche in einer feinen Diskretisierung, ferner gibt die Betrachtung der Beugungsordnungen über einen Parameterbereich ein Verständnis über die Funktion der simulierten Struktur.

In einigen Anwendungen wird die RCWA auch direkt zur Simulation und Verifikation von periodischen Strukturen der eingangs genannten Größenordnung verwendet. Meist handelt es sich hierbei um dünn beschichtete periodische Folien oder Halbleiterelemente. Beispielsweise wird in [2] die RCWA angewandt, um in der Chipherstellung aufgebrachte periodische Kontrollstrukturen zu simulieren. Hier ist das Ziel, durch Beleuchtung und spektrale Vermessung mit Hilfe der Ausgangsordnungen im Herstellungsprozess direkt Informationen über das Aspektverhältnis oder die Dicke der aufgebrachten Strukturen zu bekommen. Mit Hilfe der RCWA sind hier die Ausgangsfelder bzw. -ordnungen schon bekannt und es kann so auf die tatsächliche Struktur aus der Messung innerhalb des Produktionsschrittes rückgeschlossen werden. In [1] wird die RCWA angewandt, um sowohl die Belichtung einer Struktur in der Lithografie zu untersuchen als auch Feldsimulationen der dadurch hergestellten Strukturen inklusive eventuell auftretender Defekte zu tätigen. In [13] und in [27] werden planarisch Polarisatoren hergestellt, die elliptisch polarisiertes Licht erzeugen bzw. manipulieren sollen. Diese Oberflächen sind über eine Fläche von bis zu mehreren Zentimetern periodisch strukturiert und die Einzelzellen werden mit Hilfe einer RCWA Simulation ausgelegt und verifiziert. Ferner wird die RCWA vorzugsweise in der Plasmonik eingesetzt wie beispielsweise in [22], um die optischen Eigenschaften von planarisch mit Gold strukturierten Nanooberflächen zu untersuchen, an welchen plasmonische Effekte vorrangig auftreten. Problematisch an dieser Stelle ist, dass all diese Betrachtungen der Annahme folgen, dass eine unendlich ausgedehnte ebene homogene Welle auf die sich unendlich mit gleichartigen Zellen periodisch fortsetzende Struktur fällt. Realistischere Szenarien mit einer inhomogenen Beleuchtung auf eine begrenzte Struktur sind hier nicht möglich, wären aber beispielsweise bei der Verwendung in der Lithografie sehr hilfreich.

2.3.1 Normalenvektortransformation

Mit dem Fokus dieser Arbeit sei an dieser Stelle die bereits erwähnte Normalenvektortransformation (NVT) näher betrachtet. Zur Erläuterung ihrer Modalität sei deshalb zuerst die Kernformulierung der RCWA betrachtet, welche sich auch hier von den Maxwellgleichungen(1) in differenzierlicher Schreibweise ableitet. Allgemein werden für die drei genannten Gebiete zu Beginn allgemeine Reihenlösungen erstellt, welche dann durch Lösung eines Eigenwertproblems sowie dem Gleichsetzen der Felder an den Grenzflächen vollständig gelöst werden. Die genaue mathematische Formulierung ist in [34] hergeleitet und erläutert. In Appendix 10 sind der Vollständigkeit halber die allgemeinen Lösungen des Systems sowie die nötigen Formulierungen der Wellenvektoren und ihrer Komponenten aufgeführt. Durch Einsetzen der allgemeinen Felddarstellungen aus Appendix 10 in die Maxwellgleichungen(1) werden die folgenden vier Gleichungen in Matrixschreibweise formuliert:

$$\begin{bmatrix} \partial \mathbf{S}_y / \partial(k_0 z) \\ \partial \mathbf{S}_x / \partial(k_0 z) \\ \partial \mathbf{U}_y / \partial(k_0 z) \\ \partial \mathbf{U}_x / \partial(k_0 z) \end{bmatrix} = \begin{bmatrix} 0 & 0 & \mathbf{K}_y \mathbf{E}^{-1} \mathbf{K}_x & \mathbf{I} - \mathbf{K}_y \mathbf{E}^{-1} \mathbf{K}_y \\ 0 & 0 & \mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_x - \mathbf{I} & -\mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_y \\ \mathbf{K}_y \mathbf{K}_x & \mathbf{E} - \mathbf{K}_y^2 & 0 & 0 \\ \mathbf{K}_x^2 - \mathbf{E} & -\mathbf{K}_x \mathbf{K}_y & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \\ \mathbf{U}_y \\ \mathbf{U}_x \end{bmatrix} \quad (2)$$

Die Vektoren \mathbf{S} und \mathbf{U} repräsentieren die Amplituden der Fourierreihen der E- bzw. H-Felddarstellungen im Gitterbereich. Die Diagonalmatrizen \mathbf{K}_x bzw. \mathbf{K}_y enthalten die Wellenvektorkomponenten k_x bzw. k_y . Die Matrix \mathbf{E} wird mit den Faktoren ϵ der Fourierreihe für die Permittivität im Gitterbereich bestückt. Die Definition der Amplituden und Wellenvektorkomponenten ist in Appendix 10 aufgeführt. Ferner müssen die Koeffizienten k_x , k_y und ϵ im 2D-periodischen Gitterfall gemäß einer Umordnungsvorschrift in die zugehörigen Matrizen eingesetzt werden. Für die genaue Modalität dieser Vorschrift sei auf [34] verwiesen.

Durch erneutes Einsetzen der Matrixgleichungen(2) ineinander lässt sich beispielsweise für die Amplituden des elektrischen Feldes ein Differenzialgleichungssystem zweiter Ordnung angeben,

$$\begin{bmatrix} \partial^2 \mathbf{S}_y / \partial(k_0 z)^2 \\ \partial^2 \mathbf{S}_x / \partial(k_0 z)^2 \end{bmatrix} = \begin{bmatrix} \mathbf{K}_x^2 + \mathbf{DE} & \mathbf{K}_y [\mathbf{E}^{-1} \mathbf{K}_x \mathbf{E} - \mathbf{K}_x] \\ \mathbf{K}_x [\mathbf{E}^{-1} \mathbf{K}_y \mathbf{E} - \mathbf{K}_y] & \mathbf{K}_y^2 + \mathbf{BE} \end{bmatrix} \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \end{bmatrix} \quad (3)$$

wobei die Matrizen \mathbf{D} und \mathbf{B} durch $\mathbf{D} = \mathbf{K}_y \mathbf{E}^{-1} \mathbf{K}_y - \mathbf{I}$ und $\mathbf{B} = \mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_x - \mathbf{I}$ definiert sind. Diese beiden Matrixsysteme bilden die Grundlagen zur folgenden Herleitung der NVT. Eine erste Vorstellung und die Anwendung des Prinzips der NVT ist schon in [33] veröffentlicht worden. Zur Veranschaulichung sei an dieser Stelle die Beispielstruktur in Abbildung(4) mit einer runden Struktur, hier einem Metallzyylinder, vorgestellt.

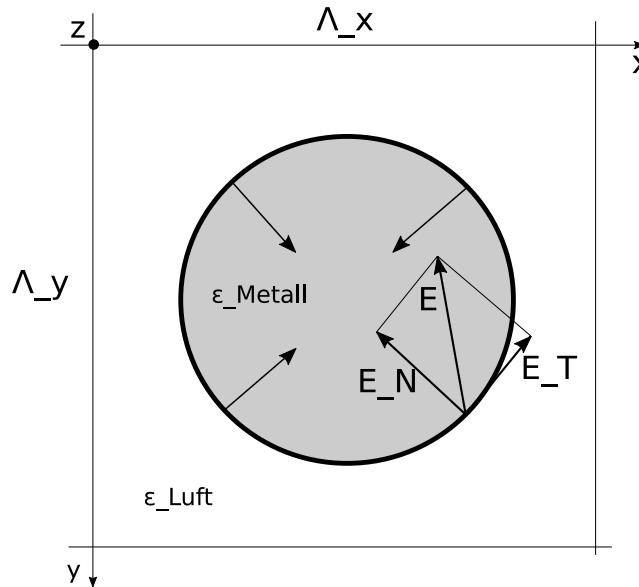


Abbildung 4: Anordnung zur Erläuterung der Normalenvektortransformation: eine luftgefüllte 2D-periodische Zelle enthält einen Metallzyylinder

Wie in Abbildung(3) bzw. in Abbildung(32) in Appendix 10 gezeigt, wird in der Permittivitätsmatrix der RCWA der Metallzyylinder durch Mikrozellen mit jeweils homogenem Brechungsindex angenähert. In den ersten Veröffentlichungen zur RCWA wie [26] ist hierbei das Problem ausgemacht worden, dass im eindimensionalen Gitterfall bei einer TM-Eingangspolarisation eine erheblich langsamere Konvergenz gegenüber der TE-Polarisation auftritt. In [10] ist eine Erläuterung dieses Problems präsentiert. Diese beinhaltet, dass die langsame Konvergenz durch die Synthese der Fourierdarstellung der Permittivität mit den

Fourierreihen der Feldlösung verursacht wird. Zur Lösung wird folgender empirischer Lösungsansatz vorgestellt:

Es sei im ersten Schritt eine eindimensional periodische Zelle betrachtet, in welcher somit die Permittivität innerhalb der gesamten Zelle durch die Funktion $\epsilon(x)$ vollständig beschrieben ist. Es werde nun der allgemeine Term der elektrischen Flussdichte $D_x = \epsilon_0 \epsilon(x) E_x$ betrachtet. Allgemein wird in der RCWA durch die Diskretisierung der Mikrozellen in Abbildung(32) für $\epsilon(x)$ eine Rechteckstruktur definiert. Wie bei der Angabe der allgemeinen Feldlösungen wird auch diese durch eine Fourierreihe angenähert. In der Herleitung der Reihendarstellungen der RCWA in [34] ist ersichtlich, dass eine Fourierzerlegung von $\epsilon(x)$ aufgrund der äquidistanten Terme zu den gleichen Basisvektoren im Fourierraum führt. Aufgrund der senkrechten Flanken der Rechteckstruktur enthält diese Entwicklung Überschwinger in den Koeffizienten hoher Ordnungen ("Gibbssches Ringing"), welche sich in der weiteren Multiplikation mit den Feldkoeffizienten auch auf das Feld fortsetzen. Der Ansatz aus [10] beinhaltet, dass die Multiplikation zweier Fourierreihen in zwei Fälle zu unterteilen ist:

Bei der Multiplikation zweier Fourierreihen ohne gemeinsame Sprungstellen kann die Multiplikation gemäß der Faltung geschrieben werden als:

$$\text{Laurent Umordnung: } D_n = \epsilon_0 \sum_{m=-\infty}^{\infty} \epsilon_{n-m} S_m \quad (4)$$

Die Faltung mit dem Verschiebungsexponenten $n - m$ kann im mehrdimensionalen Fall als $\epsilon_{n-m} \Rightarrow [[\epsilon]] = E$ realisiert werden, wobei die doppelte Klammer als allgemeine Konvention die Darstellung einer eindimensionalen Reihe von Koeffizienten als Toeplitzmatrix mit Indexvorschrift $(i, j) = i - j$ für jedes Element beschreibt. So ergibt sich der Vektor aus der Matrixmultiplikation $\mathbf{D} = \mathbf{E}\mathbf{S}$ mit dem Vektor \mathbf{S} .

Haben beide Reihen jedoch gemeinsame Sprungstellen an den gleichen Orten x , so gilt:

$$\text{Li Umordnung: } D_n = \epsilon_0 \sum_{m=-\infty}^{\infty} \left[\left[\frac{1}{\epsilon} \right] \right]_{n,m}^{-1} S_m \quad (5)$$

Dieser Ansatz behebt das Konvergenzproblem von eindimensionalen Gittern durch die Entkopplung der beiden Fourierreihen. Die Verkürzung der Rechenzeit und Optimierung der Konvergenz spielt jedoch im zweidimensionalen Gitterfall in dieser Implementierung eine erhebliche Rolle, insbesondere in Bezug auf die Anwendungen in Kapitel 5 und 6, da dieses Problem weiterhin eine langsamere Konvergenz bei 2D-periodischen Gittern verursacht. Hier besteht die Problematik insofern, dass beispielsweise bei der Struktur in Abbildung(4) die beiden vorgestellten Regelungen aufgrund der Rundheit nicht ideal angewandt werden können. An dieser Stelle kommt die NVT zum Einsatz. Die Grundidee hierbei ist, dass in jedem Zellenpunkt die Art der Fourierreihenmultiplikation ideal angepasst wird. An jedem diskretisierten Zellenpunkt des Gitters wird der elektrische Feldvektor so in zwei Komponenten zerlegt, dass eine Komponente exakt normal zu der jeweiligen Grenzfläche steht und die andere Feldkomponente tangential zur Grenzfläche liegt. In diesem Fall passen die Vorschriften, welche Multiplikation für welche Feldkomponente anzuwenden ist, exakt.

2.4 Teilausleuchtung

Die Einschränkung der RCWA bezüglich des homogenen Feldeingangs sowie die Annahme der unendlichen periodischen Wiederholung der betrachteten Zelle führt auf das eingangs erwähnte Konzept der Teilausleuchtung. Grundsätzlich wird der Oberbegriff der Teilausleuchtung in dieser Arbeit für das Problem definiert, im einfachsten Fall die realitätsnähere

Anordnung einer räumlich begrenzten periodischen Struktur bei einer ebenfalls räumlich begrenzten Eingangsbeleuchtung zu simulieren. Zur Entwicklung der Teilausleuchtung auf Basis der RCWA sind schon einige Arbeiten und kommerzielle Programme verfügbar. Maßgeblich zählt hierzu die Fourier Modal Method [30], in dieser Arbeit auch als Fouriermethode bezeichnet, welche auch schon als eine der Grundmethoden in [23] vorgestellt worden ist. Die Grundidee dieser Erweiterung ist in Abbildung(5) gezeigt.

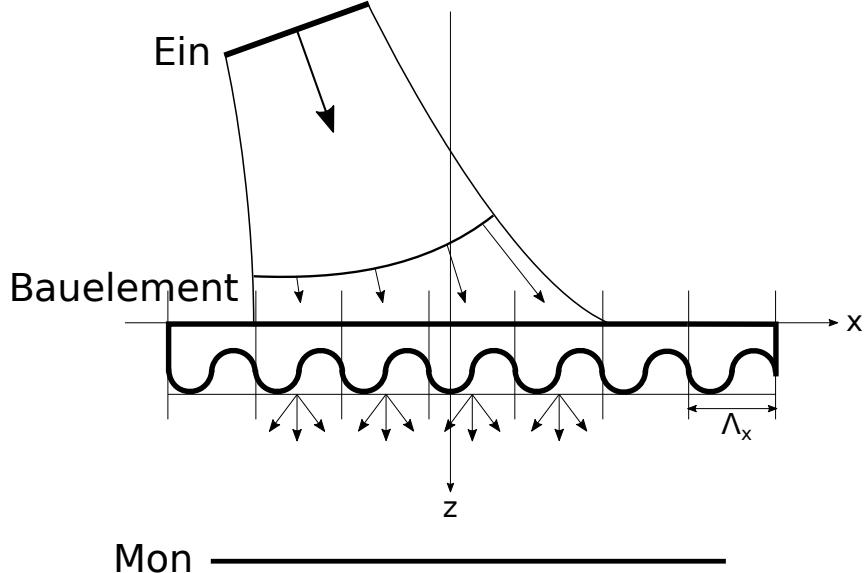


Abbildung 5: Das realitätsnähere Modell der Teilausleuchtung: eine beliebige inhomogene Feldverteilung fällt auf eine räumlich begrenzte periodische Struktur ein

Hierbei wird der Fall betrachtet, in welchem eine inhomogene Eingangsverteilung, in Abbildung(5) konzeptionell als Gaußsche Grundmode gezeigt, unter beliebigem Eingangswinkel auf eine zu simulierende Struktur einfällt. Diese wiederum besteht im einfachsten Falle aus gleichen periodischen Zellen, die sich nur begrenzt oft periodisch lateral in x- und y-Richtung fortsetzen. Nun wird das Eingangsfeld gemäß der gegebenen Verteilung an der Oberfläche der Struktur in einzelne ebene homogene Wellen zerlegt. Dies stellt eine der Grundlagen der Fourieroptik [14] dar, wonach jedes beliebige propagierende Feld durch eine endliche Reihenüberlagerung von ebenen homogenen Wellen angenähert werden kann. Nun gilt es, diese ebenen homogenen Wellen, wie in Abbildung(5) gezeigt, als Eingangswellen verschiedener RCWA-Zellen zu betrachten. Anschließend sind die Ausgänge dieser Simulationen zu einem Ausgangsfeld zu koppeln und an die gewünschte Stelle, hier die MonitorenEbene, zu propagieren. In [30] ist diese Modellbetrachtung das bisher einzige Mal vorgestellt worden und die Verknüpfung geschieht hierbei über die ausgehenden Beugungsordnungen der Simulation. Mit dieser Methode sind Strukturen mit maximalen Abmessungen einer Größenordnung über der FDTD simulierbar, dennoch sind die Größenordnungen des Auslegungsziels dieser Arbeit hiermit nicht zu erreichen. Sie eignet sich dennoch, eigene Verfahren zur Teilausleuchtung zu verifizieren.

Dieses Verfahren nach [30] wird beispielsweise in der kommerziellen Software VirtualLab von LightTrans [11] angewandt. VirtualLab ist für ein breites Spektrum an Anwendungen konzipiert, welche teilweise auch der Zielsetzung dieser Arbeit entsprechen. Die Beispiele-

wendungen umfassen hier auch Metalinsen für HUD-Displays sowie die Simulation von Kontrolloptiken für Wafer. Letzteres Beispiel in [12] werde hier näher betrachtet. Aus der Beschreibung der zu diesem Beispiel verknüpften Veröffentlichungen wie [37], [36] und [40] geht hervor, das auch hier mit einer RCWA gearbeitet wird. Sie wird nach dem Prinzip von [30] in das Gesamtsystem integriert und die Felder vor und hinter dem Wafer, dessen Strukturen mit einer RCWA berechnet und in den gezeigten Virtual Lab Beispielen meist entlang der z-Achse angeordnet werden, mittels der Fouriermethode propagiert. Trotz dieser Möglichkeiten sind Systemsimulationen der in dieser Arbeit geforderten Größenordnung bis in den Millimeterbereich hinein mit VirtualLab nicht evident.

2.5 Raytracing

Eines der Hauptziele dieser Arbeit ist es, diffraktive Strukturen in der Szenerie eines Raytracers mit zu berücksichtigen. So könnte über das vom Raytracingalgorithmus berechnete Bild ein erhöhter Grad an Informationen über die Auswirkung verschiedener Belichtungen, Orientierungen des Bauelementes oder Verzerrungen und Farbartefakte gewonnen werden. Die Abbildung(6) zeigt das grundsätzliche Konzept des Raytracings, welches hier einmal kurz erläutert wird, um die relevanten Vorarbeiten in bisherigen Veröffentlichungen in Zusammenhang zu bringen:

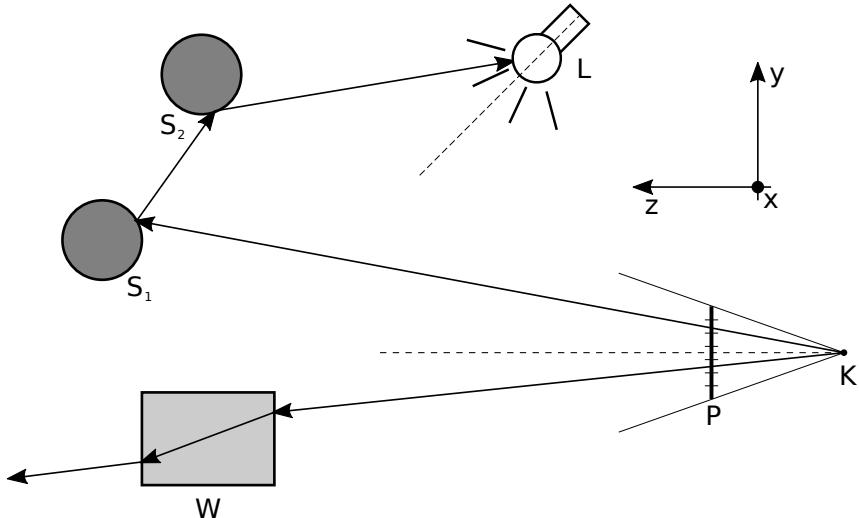


Abbildung 6: Die Grundidee des Raytracings besteht in der geometrischen Verfolgung von Rays durch eine Szenerie

Es sei eine Szenerie in einem kartesischen Koordinatensystem definiert. Zu Beginn sei der Kamerapunkt K betrachtet. Für diesen sind ein Ort sowie eine Blickrichtung, verdeutlicht durch die gestrichelte Linie, sowie ein Öffnungswinkel definiert. In einem Abstand sei eine Fläche normal zur Blickrichtung gesetzt, welche durch quadratische Pixel P unterteilt ist. Nun wird eine Gerade (Ray) mit ihrem Ursprung in K sowie einem Richtungsvektor definiert, sodass diese Gerade sowohl den Kamerapunkt K als auch den Pixel P schneidet. Nun sind in der weiteren Szenerie verschiedene geometrische Objekte definiert. Im einfachsten Fall können hierfür Kugeln S gesetzt sein, welche durch ihren Zentrumspunkt und einen Radius vollständig definiert sind. Nun wird für jedes in der Szenerie vorhandene Objekt berechnet, ob ein Schnittpunkt mit der momentan betrachteten Geraden existiert. Ist dies für mehrere

Objekte der Fall, so wird das Objekt mit dem kleinsten positiven Abstand zum Ursprung der Gerade betrachtet. In Abbildung(6) ist dies für die erste Kugel S_1 der Fall. Im nächsten Schritt gilt es zu ermitteln, wie dieser Ray mit der Kugel S_1 interagiert. Es sei an dieser Stelle exemplarisch definiert, dass die Kugel S_1 eine perfekt spiegelnde und glatte Oberfläche aufweist, sodass die Interaktion des Rays direkt ersichtlich ist. Wie allgemein von Spiegeln bekannt, wird hierbei einfallendes Licht nicht verfärbt und im exakt gleichen Winkel reflektiert. Somit ist im Algorithmus ein neuer Ray zu erstellen, welcher als Basispunkt den eben berechneten Schnittpunkt sowie einen Richtungsvektor mit gleichem Ausgangswinkel zur Oberfläche wie der vorherige Ray aufweist. Ferner wird dieser neue Ray mit einem Faktor 1 gewichtet, was hier der idealen Reflektion entspricht. Nun beginnt die Schnittpunktfindung erneut. In der gezeigten Szenerie wird nach der weiteren Reflektion an einer Kugel S_2 dann die Lichtquelle L erreicht. Diese emittiert gemäß einer raumwinkelabhängigen Verteilung eine Farbe beziehungsweise eine Intensität bei einer zugehörigen Wellenlänge. Mit dieser wird der Faktor des eintreffenden Rays multipliziert und die Strahlverfolgung abgebrochen. Der multiplizierte Wert entspricht nun der Farbe bzw. Intensität, die im zugehörigen Pixel auf der Kamerafläche zu sehen ist.

Nach diesem Grundprinzip werden Millionen von Rays im Kamerapunkt erzeugt und durch die Szenerie verfolgt. In Abbildung(6) ist ferner noch die Situation einer Transmission in ein optisch dichteres Medium gezeigt. Der zweite Ray trifft hierbei auf einen Würfel aus beispielsweise optischem Glas, bei welchem gemäß einer Gleichverteilung am Schnittpunkt entschieden wurde, dass dieser Ray in den Würfel hinein propagiert. Aufgrund des Übergangs in ein optisch dichteres Medium wird der Ray vom Lot weggebrochen. Zu sehen ist, dass dieser Ray nach Durchlauf des Würfels keinen Schnittpunkt mehr mit der restlichen Szenerie aufweist. In diesem Fall wird die Strahlverfolgung abgebrochen und der zugehörige Kamerapixel erhält den Faktor 0. Dazu sei hierbei vermerkt, dass, wie in Abbildung(6) gezeigt, entgegen dem physikalischen Grundprinzip Rays in der Kamera starten und bis zu den Lichtquellen hin verfolgt werden. Dieses Verfahren wird allgemein als “backward ray tracing” bezeichnet und wird in allen gängigen Raytracingalgorithmen angewandt. Diese Richtungsdefinition erweist sich im Allgemeinen als erheblich effizienter im Vergleich zum “forward ray tracing”, bei dem physikalisch korrekt die Rays in den Lichtquellen kreiert und durch die Szenerie hindurch bis zum Kamerapunkt hin verfolgt werden. Der Effizienzgewinn liegt maßgeblich daran, dass im “backward ray tracing” alle betrachteten Rays den geometrisch sehr kleinen Kamerapunkt schneiden und so zum Bild beitragen. Ferner zielen viele Optimierungen der Raytracingverfahren darauf ab, die Strahlverfolgung nach einigen Iterationen abhängig von der ermittelten Lichtintensität abzubrechen. Im “forward ray tracing” wird so ein erheblicher Anteil Rays mit hohem Beitrag weiterverfolgt, jedoch schneidet nur ein kleiner Teil den aufgrund seiner Größe schwer zu “treffenden” Kamerapunkt.

Um Szenerien zu visualisieren, bei denen das berechnete Bild mit der Fotografie der Szene vergleichbar ist, bedarf es einiger weiterer Methoden. Diese beziehen sich beispielsweise auf eine möglichst schnelle Schnittpunktfindung oder die angewandten Raumverteilungen, mit welchen die Richtungen der Rays innerhalb eines Pixels definiert werden (Kamerasampling). Diese sind, auch in Bezug auf die eingesetzten Materialien in der Szene, größtenteils schon entwickelt worden. Für tiefgreifendere Betrachtungen diesbezüglich sei deshalb hier auf Physically Based Rendering Technique [25] verwiesen. Hier ist eine vom Anwender über Skripte steuerbare Implementierung eines Raytracers entwickelt und parallel ein Buch veröffentlicht worden, in welchem alle angewandten Konzepte von Schnittpunktfindung, Sampling und weiteren von Grund auf hergeleitet und die jeweils zugehörige Implementierung kohärent

erläutert werden. Dieses Softwarepaket wird im Arbeitsgebiet des Raytracings als eine der fortschrittlichsten Systeme gesehen [24] und schon in verschiedenen Bereichen wie Computer Generated Imagery (CGI) oder auch für Akustiksimulationen eingesetzt. Dennoch sind in diesem Paket keine diffraktiven Strukturen visualisierbar, deren Einzelemente in der Größenordnung der Wellenlänge oder darunter liegen. Somit würde das Ziel dieser Arbeit erreicht, wenn eine Implementierung existiere, die es ermöglicht, zusätzlich in pbrt solche Strukturen mit simulieren zu können. Diese hätte den Vorteil, dass nebenbei die Vorteile von pbrt in Bezug auf Geschwindigkeit, Schnittpunktfindung und der dort definierten Materialien ohne nötige eigene Implementierung genutzt werden könnten.

Gesucht ist deshalb, ob an dieser Stelle schon andere Autoren Vorarbeit geleistet haben. Hierzu seien [24], [21] und [38] genannt, wo versucht wird, diffraktive Strukturen im Visualisierungsprozess zu berücksichtigen. Hierfür beschreiben die Autoren die Oberfläche als Ortsfunktion und setzen diese mit Hilfe einer Fouriertransformation in eine Reflektanzfunktion ein. Dies entspricht jedoch zum einen nicht einer vollständigen Feldsimulation, zum anderen sei beispielsweise das errechnete Bild einer CD in Fig.1 von [24] oder Fig. 6 in [38] betrachtet. Hier lässt sich direkt erkennen, dass ein erheblicher Teil der Farbartefakte, wie man sie vom Betrachten einer CD in Abbildung 8 kennt, in diesem Modell nicht zu sehen sind. Daraus folgt, dass mit der Formulierung aus [38] aufwendigere Strukturen wie Goldgitter oder Metalinsen nur ungenügend simuliert werden können. Eine CD stellt eine vergleichsweise einfache Struktur dar, deren Farbspektren größtenteils nur durch die Periode und die Dicke der Metallstrukturierung auf der Oberfläche erzeugt werden. Dies erklärt sich dadurch, dass unterschiedlich beschriebene CDs die exakt selben Farbartefakte verursachen. Es ist somit als unwahrscheinlich anzusehen, dass mit der in [38] vorgestellten Methodik komplexere Goldgitter oder diffraktive Linsensysteme im Raytracing korrekt dargestellt werden. Hierzu ähnlich wird auch in [24] vorgegangen. In [24] wird angeführt, dass grundsätzlich eine solche vollständige Feldsimulation innerhalb des Raytracingalgorithmus einerseits aufgrund des zu großen Rechen- und Zeitaufwandes nicht möglich sei. Andererseits gäbe es keine Möglichkeit, Rays in der Strahlverfolgung mit den Ein- und Ausgangsfeldern der Feldsimulation zu verknüpfen. Dies stellt somit ein Defizit für das Ziel dieser Arbeit dar.

2.5.1 Grundkonzepte des Raytracing

An dieser Stelle seien die Konzepte vorgestellt, die für die Berücksichtigung der RCWA im Rahmen dieser Arbeit vonnöten und gleichzeitig schon existent sind. Die Kernelemente des Raytracings bilden die Berechnung der resultierenden Ausgangsrichtung des Rays am Schnittpunkt sowie der Gewichtungsfaktor der resultierenden Richtungskombination. Welcher Faktor hierbei verwendet wird, hängt von der Farbdefinition des Raytracers ab. Im einfachsten Fall ist jedem Ray eine Rot-Grün-Blau (RGB) Farbwert zugeordnet, der iterativ an jedem Schnittpunkt mit einem dreiwertigen Skalierungsfaktor multipliziert wird. In anderen Algorithmen wird jedem Ray nur ein skalarer Faktor bei einer gegebenen Wellenlänge oder ein komplettes Spektrum zugeordnet. Ein intuitiver Zusammenhang für diesen Faktor an Spiegeln wurde in der Beschreibung von Abbildung(6) erläutert.

Um diesen Zusammenhang physikalisch für beliebige Materialien zu definieren, wird üblicherweise mit dem Konzept der bidirektionalen Verteilungsfunktion (bidirectional distribution function BxDF [25]) eingesetzt. Das x ist an dieser Stelle ein Platzhalter für den Fall, dass ein Reflektions- (BRDF), Transmissions- (BTDF), Streu- (BSDF) oder Reflektionsszenario unter Berücksichtigung von Streuprozessen in den Materialschichten unterhalb der Oberfläche (BSSRDF) betrachtet wird. Zur Einführung sei an dieser Stelle nur die BRDF betrachtet. Die Abbildung(7) soll hierbei unterstützen.

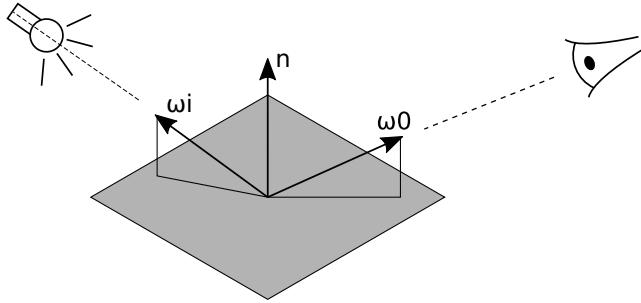


Abbildung 7: Das Grundkonzept der BRDF

Es wurde der Schnittpunkt des aktuellen Rays aus der Kamera mit der aktuell betrachteten Oberfläche berechnet. Die Richtung ω_o dieses Rays ist hierbei umgekehrt in einem zur Oberfläche normalen Koordinatensystem eingezeichnet. Nun gilt es herauszufinden, welche Richtung der weiterpropagierende Ray habe, dessen Richtung hier mit ω_i bezeichnet sei. Ferner ist es physikalisch relevant, welche Leuchtdichte am Schnittpunkt in die Richtung der Kamera ω_o abgestrahlt wird. Beschreibe $L_o(p, \omega_o)$ die Leuchtdichte des Rays der Richtung ω_o am aktuellen Schnittpunkt p und analog $L_i(p, \omega_i)$ selbige für die Richtung ω_i , so lässt sich nach [25] hierfür die gesuchte Leuchtdichte $L_o(p, \omega_o)$ formulieren:

$$L_o(p, \omega_o) = \int_{S^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) |\cos\theta_i| d\omega_i \quad (6)$$

Hierbei beschreibt S^2 die Integration über die Oberfläche der mathematischen Einheitskugel und $f(p, \omega_o, \omega_i)$ die gesuchte BRDF nach

$$f_r(p, \omega_o, \omega_i) = \frac{dL_o(p, \omega_o)}{L_i(p, \omega_i) \cos\theta_i d\omega_i} \quad (7)$$

Einer der Hauptaufgaben des Raytracingalgorithmus ist, genau dieses Integral zu berechnen. Im einfachsten Fall wie dem ideal verspiegelten Körper ist nur eine Ausgangsrichtung ω_i bei der gegebenen Richtung ω_o vorhanden; gleiches gilt auch für einfache Modellierungen von Metall, Glas oder Wasser. Im allgemeinen Fall erzeugen jedoch alle Raumrichtungen im Integral einen Beitrag, diese Oberflächen werden allgemein als diffus bezeichnet. Theoretisch könnten nun an jedem Schnittpunkt mit einer feinen Diskretisierung der Raumrichtungen Rays in diese losgeschickt und deren Pfadgewichtungen weiter berechnet werden. Hierbei steigt jedoch der Rechenaufwand lawinenartig mit jedem weiteren Schnittpunkt an. An dieser Stelle kommt die Monte-Carlo Theorie zur Anwendung. Sie ist besonders dann geeignet, wenn Integrale der Form von Gleichung(6), also mit hoher Dimensionalität, numerisch abzuschätzen sind. Ihr Grundprinzip ist, die Ausgangsrichtung gemäß einer Wahrscheinlichkeitsdichtefunktion auszuwählen und anschließend die resultierende Leuchtdichte der folgenden Raykombination mit dem Integral dieser Dichtefunktion zu gewichten.

2.6 Fazit

In der Überleitung zum folgenden Kapitel der Methodik sollen an dieser Stelle noch einmal kurz die ermittelten Forschungsdefizite in Bezug auf das Ziel dieser Arbeit aufgelistet werden:

Räumliche Maximalgröße des diffraktiven Bauelements

In den bisher erläuterten Verfahren zur wellenoptischen Simulation kann die maximal mögliche laterale Ausdehnung diffraktiver Bauelemente oder Systeme $100 \mu m$ nicht überschreiten.

Raytracing von diffraktiven Elementen

Aufgrund der wellenoptischen Effekte existiert bisher noch keine ausreichende Möglichkeit, diffraktive Bauelemente im Raytracing zu berücksichtigen und so virtuell die optische Interaktion mit dem späteren Einsatzgebiet einschätzen zu können.

Simulation neuartiger Bauelemente wie Metalinsen

Aufgrund der maximal möglichen Ausdehnung in der Simulation können Bauelemente wie Metalinsen bisher nur für rudimentäre Abbildungen wie einer Fokussierung auf einen Brennpunkt konzeptionell simuliert werden.

Handhabbarkeit der wellenoptischen Simulationen

Sowohl in Bezug auf die Abmessungen der Bauelemente als auch auf Simulationen über Parametermengen wie einem Wellenlängenspektrum sind die bisherigen Methoden sehr rechenintensiv und skalieren in Speicherbedarf und Rechenzeit zum Teil mit der vierten oder fünften Potenz, wofür Simulation auf Großgruppenrechnern im Stundenbereich nötig sind.

Das Ziel ist nun, eine geeignete Methodik zu formulieren, um diese Defizite zu lösen.

3 Methodik

Im nun folgenden Kapitel wird die Methodik vorgestellt, mit welcher das Ziel dieser Arbeit erreicht und die ermittelten Forschungsdefizite behoben werden sollen. Im ersten Schritt soll das bestehende 3D-RCWA Paket um eine 2D-Normalenvektortransformation (NVT) erweitert werden, um die Konvergenz und Rechengeschwindigkeit für 2D-periodische Zellen maßgeblich zu beschleunigen. Dies gilt insbesondere für den Aspekt, RCWA-Simulationen über Parameterbereiche der Eingangswinkel und -wellenlängen in der Teilausleuchtung und in der Raytracing-Weiterleitung zu nutzen. Für die Modellbetrachtung der Teilausleuchtung soll mit dem Ziel der Maximierung der lateralen Ausdehnung der zu simulierenden Struktur das Konzept der Feldbildsynthese entwickelt werden. Parallel hierzu soll die bereits existierende Fouriermethode nach [33] als Referenzmodell implementiert werden. Die Zielsetzung der Berücksichtigung diffraktiver Optiken im Raytracing soll durch die Möglichkeit einer RCWA Integration gegeben sein. Da hierfür auf keine bereits bestehenden Methoden zurückgegriffen werden kann, ist das Konzept, im Schritt einen vereinfachten Ansatz in einen selbst implementierten Raytracer testweise einzubauen. Zeigt die Gegenüberstellung dieses Ansatzes mit der Fotografie einer Szenerie vergleichbare Ergebnisse, soll ein ähnliches Konzept als Erweiterung für einen dem aktuellen Stand der Technik entsprechenden Raytracingalgorithmus (pbryt-v3 [25]) aufgesetzt werden. Der erhoffte Mehrwert ist hierbei, alle Vorteile dieser bestehenden Implementierung in Bezug auf Rechengeschwindigkeit und möglichem Detailgrad der betrachteten Szenerie zu nutzen.

Für jeden dieser Arbeitsschritte gilt es, die entwickelten Methoden durch Vergleiche mit Referenzsimulationen, Modellbeispielen und Veröffentlichungen zu verifizieren. Maßgeblich zählen hierzu die Vergleiche anhand von Modellstrukturen in Simulationen von FDTD und Fouriermethode. Entsprechen die Resultate vorher festgelegten Ansprüchen an Konvergenz und Abweichung, gilt das Ziel dieser Arbeit als erreicht.

3.1 Konzept

Die folgende Abbildung(8) zeigt eine Einordnung der für diese Arbeit relevanten Vorarbeiten, die grundlegende Methodik zum Erreichen des Ziels dieser Arbeit sowie die motivierenden Problemstellungen.

Grundsätzlich basieren die Ansätze dieser Arbeit auf einer im Vorfeld implementierten dreidimensionalen Zellensoftware, der Rigorous Coupled Wave Analysis (RCWA). Mit dieser Implementierung ist es möglich, eine dreidimensionale periodische Zelle feldtheoretisch zu simulieren, in der beliebige Materialtypen und Geometrien definierbar sind. Als Eingang dient eine ebene homogene Welle mit polarem, azimuthalem und zwei Polarisationswinkeln. Als Resultat können die Felder in beliebig orientierten Ebenen in den Gebieten über und unter der Struktur sowie innerhalb der Struktur angegeben werden. Ferner stehen die ausgehenden Beugungsordnungen mitsamt ihrer Ausgangswinkel und den komplexen Amplituden bzw. Beugungseffizienzen zur Verfügung. Nach Abschluss der Verifikation dieses Paketes in [34] motivierten [6] und [7] mit den Einheitszellen der dort konstruierten Metalinsen die Verwendung der RCWA zu deren Auslegung. Aus der Fragestellung, wie die RCWA hierfür verwendet werden könnte, ist ein erster Ansatz für die Feldbildsynthese in [34] entwickelt worden. Dieser beschränkt hierbei jedoch nur auf einen rudimentären Aufbau, in welchem die Metalinse eine ebene homogene Welle auf einen Brennpunkt fokussiert.

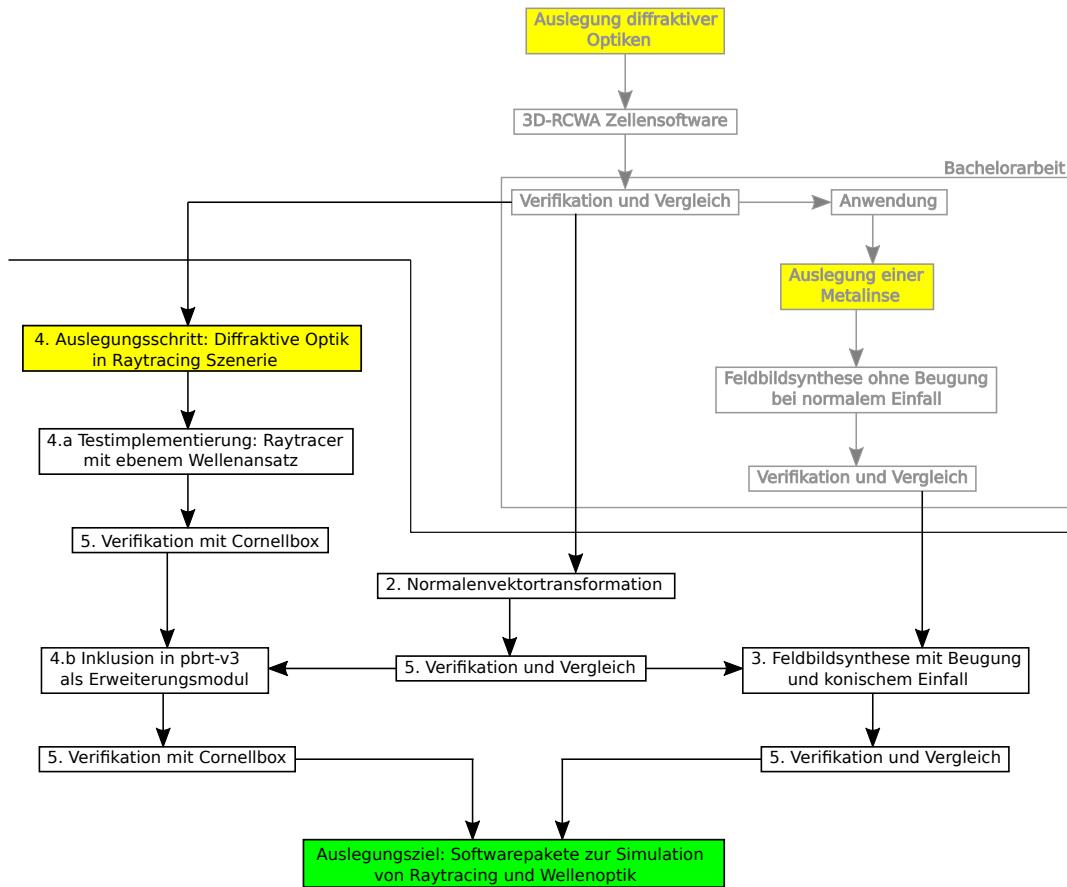


Abbildung 8: Einordnung der Motivationen (gelbe Flächen) und der Herangehensweise zum Erreichen des Ziels (grün) dieser Arbeit; die hellgrauen Module wurden schon im Vorfeld dieser Masterarbeit behandelt

3.2 Normalenvektortransformation

Im Zuge der im Folgeabschnitt vorgestellten Teilausleuchtung und der Mitnahme der RCWA-Simulation ins Raytracing ist direkt abzusehen, dass viele Simulationen einer Struktur durchzuführen sind, da hier die RCWA-Ergebnisse über Parametermengen wie Eingangswinkel und Wellenlängenspektren nötig sind. Dies geht einher mit dem Gesamtziel dieser Arbeit, die Feldsimulationen handhabbarer zu realisieren. Besonders motiviert das Problem, dass die Simulationszeiten bei 2D-periodischen Strukturen mit einem Faktor größer als $\Omega(n^2)$ steigen, wobei n die Anzahl der verwendeten Beugungsordnungen beschreibt. Gleichzeitig wurden auch Strukturen mit scharfkantigen Metallelementen simuliert, die auf der gegebenen Rechnerkonfiguration den Ansprüchen an die Konvergenz nicht genüge taten. Aus diesem Grund soll im ersten Teil dieser Arbeit der RCWA noch die Erweiterung einer NVT angefügt werden. Auch die Rundheit beziehungsweise Flankenwinkel der Einheitsgeometrien der Metallinsen aus [6] und [7] motivieren diesen Ansatz. Diese Idee basiert maßgeblich auf [33], die Formulierung muss jedoch den Grundsätzen der bisherigen Formulierung der RCWA in [34] folgen. Die Verifikation der NVT soll sowohl mit der ursprünglichen RCWA durchgeführt als

auch mit einer FDTD verglichen werden. Von der NVT werden maßgebliche Optimierungen der Rechengeschwindigkeit und Systemgröße bei gegebenem Konvergenzanspruch erwartet.

3.3 Teilausleuchtung

Wie bereits erwähnt ist im Vorfeld das Konzept der Feldbildsynthese in einem einfachen Ansatz entwickelt worden. Dieses ist für eine rudimentäre optische Abbildung der Metalinse schon kurz in [34] vorgestellt und verifiziert worden. Sie soll grundsätzlich allerdings für beliebige periodische Strukturen, Anordnungen und Eingangsfeldverteilungen implementiert werden. Die Möglichkeiten der Metalinsen, auch komplexere Abbildungen oder Beleuchtungs-szenarien zu behandeln, motivieren diese Entwicklung. In [30] ist diese Modellbetrachtung mit einem anderen Ansatz das bisher einzige Mal vorgestellt worden. Dieser Ansatz beruht auf der Verknüpfung über die ausgehenden Beugungsordnungen der RCWA Simulation. In Kapitel 2 sind schon die Grenzen dieses Verfahrens gezeigt. Das Erreichen des Ziels dieser Arbeit wird von der erwähnten und hierfür entwickelten Feldbildsynthese erhofft, jedoch gilt es die Methode aus [30] als Referenz parallel zum eigenen Verfahren der Feldbildsynthese zu implementieren, um mit ihr Vergleiche und Verifikationen durchzuführen.

3.4 Raytracing

Der letzte Schritt zum Auslegungsziel aus Abbildung 8 besteht in der Integration der RCWA beziehungsweise der Teilausleuchtung in das Raytracing. Aus der Ausarbeitung der RCWA sowie der Betrachtung dieses Problems besteht die Idee, im einfachsten Fall die Rays und die Beugungsordnungen der RCWA miteinander zu verknüpfen. Diese Idee wurde bisher noch in keiner Veröffentlichung genannt und soll deshalb im ersten Schritt in einem rudimentären Raytracingalgorithmus getestet werden, da ihre Funktionalität nicht garantiert ist. Hierfür eignet sich smallpt [31], wo als geometrische Objekte Kugeln sowie matte und ideal reflektierende Materialien definiert sind. Diese einfache Implementierung gilt es um einige Geometrien zu erweitern und zusätzlich Samplingfunktionen zu implementieren, in welchen eine RCWA-Oberfläche berücksichtigt ist. Eine der Grundideen hierbei ist ferner, die Schnelligkeit der RCWA-Simulation zu nutzen und im Vorfeld eine periodische Struktur für einen kompletten Eingangswinkel- und -wellenlängenbereich zu simulieren, diesen einzulesen und dann zur Laufzeit des Raytracers lediglich auf diese Simulationsdaten zuzugreifen, anstatt an jedem Schnittpunkt eine vollständige Feldsimulation durchzuführen. In den Bemerkungen zum Raytracing in Kapitel 2 sind das “backward ray tracing” und die zugehörigen Hintergründe genannt. Diese Methode erhöht die Effizienz des Raytracingalgorithmus maßgeblich, stellt jedoch die umgekehrte Reihenfolge zur RCWA-Simulation dar. Eine Idee zur Lösung dieses Problems sollte ebenfalls in diesem Ansatz getestet werden. Die Verifikation dieses Verfahrens orientiert sich an CGI üblichen Verfahren, eine Cornellbox zu bauen und dann Fotografie und errechnetes Bild miteinander zu vergleichen. Diese Idee stammt von der Cornell-University [5], wo ebenfalls solche Boxen gebaut wurden, um die eigenen bildgebenden Verfahren zu bestätigen. Sie ist insofern hilfreich, als dass in ihrer Bauform nur die in der Decke vorhandene lambertsche Lichtquelle im Code zu definieren sowie die Berandung der Szenerie durch Wände mit homogener Farbgebung im Code vergleichsweise einfach zu berücksichtigen ist. Auch sie wird in Kapitel 6 vorgestellt. Erst wenn mit dieser Testimplementierung der rudimentäre Ansatz als anwendbar betrachtet werden kann, sollen erweiterte Konzepte in pbtr eingebaut werden, sodass das Ziel dieser Arbeit zur Integration der RCWA in Raytracing erreicht werden kann. Der erhoffte Mehrgewinn ist hierbei, alle Vorteile von pbtr-v3, als Raytracer auf dem gegenwärtigen Stand der Technik, in Bezug auf

Rechengeschwindigkeit und Detailgrad der Szenerie nutzen zu können.

3.5 Verifikation und Vergleich

Ein besonderes Augenmerk liegt in dieser Arbeit auf der Verifikation der neu entwickelten Methoden. Das Erreichen des Arbeitsziels ist erst dann garantiert, wenn vorher gesetzte Ansprüche an die eigenen Methoden in Bezug auf Konvergenz und Abweichung vollständig erfüllt werden. Diese sind jedoch vorher eindeutig anzugeben. Im Idealfall ist hierfür eine Beispielstruktur zu definieren, für welche eine ideale analytische Lösung beispielsweise des Feldes in einer betrachteten Ebene existiert. In diesem Fall muss das eigene Verfahren mit stetig verfeinerter Diskretisierung des Simulationsraumes eindeutig ohne Restabweichung gegen diese Lösung konvergieren. Ist für eine gegebene Struktur keine analytische Lösung formulierbar, gilt es ein Simulationsverfahren heranzuziehen, welches schon durch genügend Veröffentlichungen als konvergent und fehlerfrei bestätigt ist und wobei somit auch dessen Handhabung, um diese Konvergenz zu garantieren, bekannt ist. Hierbei kann diese Lösung als Konvergenzziel zur Rate gezogen werden, gegen welches die eigene Methode ebenfalls bei stetig feinerer Diskretisierung konvergieren muss. Ferner müssen alle Artefakte in verglichenen Feldbildern eindeutig erklärbar sein. Problematisch ist an dieser Stelle, dass unter Umständen keine direkte Gegenüberstellung mit einer anderen Simulation in der Größenordnung der hier verwendeten Strukturen möglich ist. In diesem Fall ist ein Stufenprogramm mit Teilverifikationen zu erstellen, die in ihrer Gesamtheit als Verifikation ausreichen und welche einzeln vollständig erfüllt werden müssen. Hierauf wird in der Verifikation von Kapitel 5 näher eingegangen.

Zusätzlich wäre hier auch der Vergleich mit Messergebnissen eines realen Aufbaus ideal. Konzeptionell könnte beispielsweise ein Laser, welcher eine Gaußsche Grundmode emittiert, in schiefem Winkel zu einer Oberflächenstruktur oder einem Gitter der gewünschten Größenordnung positioniert werden. Dann könnte das Ausgangsfeld mit Hilfe eines Spektrometers gemessen und das Ergebnis von Messung und Simulation direkt verglichen werden. Solche Strukturen sind prinzipiell verfügbar, jedoch sind von Seiten der Hersteller nicht genügend Informationen zur Geometrie der Einzelemente bekannt, um sie explizit in der RCWA nachbilden zu können. Diese Geometrie repräsentiert Entwicklungsarbeit und Forschung von Seiten der Entwickler, damit eine gegebene Struktur beispielsweise die gewünschten Beugungseffekte in verschiedenen Wellenlängenbereichen erzielt. Sie wird somit nur ungern öffentlich preisgegeben. Abbildung(1) bietet hierfür jedoch mit der CD als diffraktives Element eine Möglichkeit. Dennoch gilt auch hier der Anspruch an die Konvergenz. Gleichzeitig bietet dies den Vorteil, rückwirkend auch nochmals die RCWA und Teilausleuchtung zu verifizieren, da auf die Simulationsergebnisse dieser Implementierungen im Raytracer zugegriffen wird. Als Kernelement der Verifikation dient in dieser Arbeit die FDTD, da sie schon seit einigen Jahrzehnten besteht und somit schon in hohem Grade untersucht und weiterentwickelt wurde. Sie wird an vielen Stellen im Bereich der Nanotechnologie, Hochfrequenztechnik und Plasmonik eingesetzt bzw. in zugehörigen Veröffentlichungen zur Verifikation genutzt. Sie ist ferner ebenfalls schon zur Verifikation der reinen 3D-RCWA-Zellensoftware aus [34] eingesetzt worden. Aufgrund der in Kapitel 2 zur FDTD gezeigten Skalierung des Berechnungssystems ist es essentiell, genügend Informationen über die richtige Handhabung der FDTD, insbesondere in Bezug auf die Konvergenz, zu erlangen, jedoch auch, um die Richtigkeit der als Referenz herangezogenen FDTD-Lösung zu garantieren. Zu diesen Themen existiert schon eine erhebliche Bandbreite an Veröffentlichungen, die hierbei zur Rate gezogen werden. Wichtig sind hierfür insbesondere [16], [17] und [8]. Aus diesen Arbeiten werden Vorschriften

und Vorgehen ermittelt, mit welchen die Simulationen in dieser Arbeit zu tatigen und welche in Appendix 9 aufgefuhrt sind.

3.6 Bemerkung zur Implementierung

An dieser Stelle sei noch eine kurze Bemerkung zur Implementierung der Software und des hier verwendeten Computersystems angefuhrt, um dem Leser einen groben Uberblick uber Handhabung sowie Rechenzeit und Speicherbedarf der vorgestellten Methoden zu geben. Das 3D-RCWA-Paket ist in der Sprache Python mit den Standardbibliotheken Numpy und Scipy implementiert. Es ist grundsatzlich objektorientiert aufgebaut, sodass beispielsweise zu Beginn eine leeres Schichtobjekt erzeugt wird, welches dann nacheinander mit Hilfe von Geometriefunktionen und Materialdaten bestuckt wird. Anschlieend konnen automatisierte Konvergenz- oder Simulationsfunktionen gestartet werden. Das Teilausleuchtungspaket ist auf ahnliche Weise implementiert worden und zu handhaben. Die vorsimulierten RCWA-Daten uber beispielsweise einen Eingangswinkelbereich werden hier im Format .dat eingelesen und ein Schichtobjekt gemss Symmetrie und geometrischer Transformationen bestuckt. Anschlieend kann eine beliebige Eingangsbeleuchtung definiert und durch die Halbrume uber und unter der Struktur propagiert werden. Der Raytracingversuch ist ebenso wie pbry-v3 und die zugehorigen Erweiterungen dieser Arbeit in C++ geschrieben. Alle drei Softwarepakete sind durch Skripte ansteuerbar, um so die Verwendung fur einen Anwender ohne Kenntnis der Programme und ihrer Funktionalitat zu erleichtern. Alle Implementierungen sind zusatzlich mit Multithreading versehen, sodass bei den in dieser Arbeit vorgestellten Simulationen die CPU-Auslastung mindestens 80% betragt. Dennoch kann das Multithreading uber die Anwenderskripte gesteuert werden, falls eine Simulation als Hintergrundprozess gedacht ist. Alle Simulationen sind auf einem Computer mit 16GB Arbeitsspeicher und Intel Core i7-3770 mit 3.4GHz erstellt worden. Auf diese Konfiguration beziehen sich auch alle Vergleiche dieser Arbeit bezuglich Rechenzeit und Speicherbedarf. Dementsprechend sind bei Verwendung der hier vorgestellten Implementierungen auf moderneren bzw. groeren Rechensystemen die Mglichkeiten in Bezug auf Bauelementgroe und Rechenzeit erweitert.

4 RCWA - Normalenvektortransformation

Im Kontext dieser Arbeit liegt der Fokus der Simulationsmethoden und der zugehörigen Implementierung auf der Optimierung der Rechengeschwindigkeit und Konvergenz. Konkret bedeutet dies für die RCWA, möglichst wenig Beugungsordnungen in der Simulation bei gleichem Konvergenzgrad verwenden zu müssen. Gerade in Bezug auf die spätere Verwendung in der Teilausleuchtung in Kapitel 5 sowie im Raytracing in Kapitel 6 sollen so größere Anzahlen an Simulationen über Eingangsparameterbereiche auf der gegebenen Rechnerkonfiguration handhabbar sein. In dieser Arbeit werden vorrangig zweidimensional periodische Zellen betrachtet, wo sich diese Gegebenheit eklatant in Bezug auf die Rechengeschwindigkeit auswirkt. Auf der Basis des bestehenden 3D-RCWA Paketes, als Vorarbeit und Grundlage dieser Arbeit aus [34], gilt die Normalenvektortransformation (NVT) nach [33] als maßgebliche Optimierung. Ihre Grundidee ist, die im Abschnitt der RCWA in Kapitel 2 vorgestellten Methoden zur Fourierreihenmultiplikation durch eine geeignete Normalenvektortransformation ideal anzupassen. Die Formulierung der NVT in dieser Arbeit orientiert sich an der Idee in [33], jedoch wird sie auf Basis der bestehenden Implementierung in [34] andersartig formuliert. Aufgrund dieser abweichenden Formulierung ist hierfür eine explizite Verifikation der NVT-Erweiterung vonnöten. Wie im Kapitel 3 schon ausgemacht soll hierfür einerseits die bestehende RCWA-Verifikation selbst die optimierte Konvergenz der NVT bestätigen. Hierbei wird auch ein Fall betrachtet, bei welchem entgegen von Veröffentlichungen zur RCWA diese ohne die NVT Erweiterung nicht konvergent ist. Andererseits werden die Simulationsergebnisse einer Modellstruktur in der RCWA mit NVT Erweiterung sowie einer parallelen FDTD Simulation über die explizite Angabe des Feldes in einer Schnittebene verglichen. Hieraus soll auch eine Abschätzung für die spätere Anwendung dieser Feldbilder getätigt werden.

4.0.1 Formulierung

Angesetzt wird an dieser Stelle bei der elektrischen Verschiebungsdichte \vec{D} , welche sich aus der Multiplikation von elektrischem Feld \vec{E} und Permittivität ϵ ergibt. In Abbildung(4) in Kapitel 2 ist die Zerlegung des elektrischen Feldvektors in eine zur Grenzfläche normale \vec{E}_N und eine tangentiale Komponente \vec{E}_T gezeigt. Diese Zerlegung wird direkt in die Verschiebungsdichte eingesetzt.

$$\vec{D} = \epsilon \vec{E} = \epsilon(\vec{E}_N + \vec{E}_T) = \epsilon \vec{N}(\vec{N} \bullet \vec{E}) + \epsilon(\vec{E} - \vec{N}(\vec{N} \bullet \vec{E})) \quad (8)$$

Das Normalenvektorfeld \vec{N} ist hierbei als Vektor mit drei kartesischen Komponenten für jeden diskretisierten Zellenpunkt definiert. In Abbildung(9) ist dieses Feld exemplarisch für eine Beispielstruktur gezeigt.

An Punkten auf einer Materialgrenzfläche müssen die Normalenvektoren normal zu dieser orientiert sein. In Gebieten zwischen den Grenzflächen haben die Richtungen der Normalenvektoren stetig zu verlaufen. Bei der Verwendung von Einheitsgeometrien wie Zylindern oder Blöcken kann dieses Vektorfeld intuitiv vordefiniert und Skalierung und Drehung auf die Struktur angepasst werden. Bei komplexeren Geometrien sind Verfahren wie die Schwarz-Cristoffel Transformation [3] zu verwenden, die jedoch an dieser Stelle nicht näher betrachtet werden.

Werden nun elektrisches Feld und Permittivität nach Fourierreihen entwickelt, ist in dieser Schreibweise direkt ersichtlich, dass im ersten normalen Term die Li-Regel (5) und im zweiten Term die Laurent-Regel (4) bei der Multiplikation der Reihen angewandt werden. Nun wird, wie in Abbildung(9) gezeigt, an jedem Punkt der Zelle, insbesondere an den

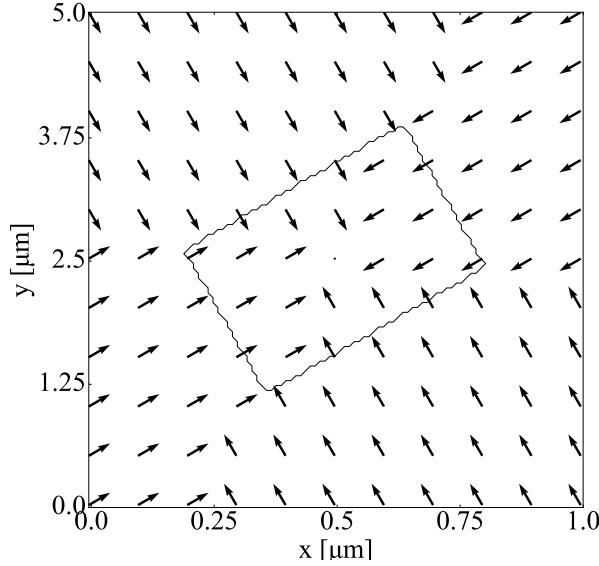


Abbildung 9: Verifikationsbeispiel: Goldelement mitsamt der Normalenvektoren in einer luftgefüllten Zelle

Materialgrenzflächen, der Normalenvektor angepasst. So kann an jedem Punkt ein Teil der Verschiebungsdichte jeweils exakt nach Li und Laurent multipliziert werden.

Im Übergang zur Matrixschreibweise der RCWA gelangt man zu:

$$\begin{aligned} \begin{bmatrix} \mathbf{D}_x \\ \mathbf{D}_y \\ \mathbf{D}_z \end{bmatrix} &= \left[\begin{bmatrix} 1 \\ \epsilon \end{bmatrix} \right]^{-1} [[\vec{\mathbf{N}} \vec{\mathbf{N}}]] \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} + [[\epsilon]] \left(\begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} - [[\vec{\mathbf{N}} \vec{\mathbf{N}}]] \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} \right) \quad (9) \\ &= [[\epsilon]] \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} + \epsilon_\Delta \begin{pmatrix} \mathbf{N}_x \mathbf{N}_x & \mathbf{N}_x \mathbf{N}_y & \mathbf{N}_x \mathbf{N}_z \\ \mathbf{N}_y \mathbf{N}_x & \mathbf{N}_y \mathbf{N}_y & \mathbf{N}_y \mathbf{N}_z \\ \mathbf{N}_z \mathbf{N}_x & \mathbf{N}_z \mathbf{N}_y & \mathbf{N}_z \mathbf{N}_z \end{pmatrix} \begin{bmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{bmatrix} \quad \text{mit } \epsilon_\Delta = \left(\left[\begin{bmatrix} 1 \\ \epsilon \end{bmatrix} \right]^{-1} - [[\epsilon]] \right) \end{aligned}$$

Da sich in dem hier vorgestellten Fall die Normalenvektortransformation nur auf eine RCWA-Schicht beschränkt, in welcher der Brechungsindex in z-Richtung invariant ist, ergibt sich die Komponente $\mathbf{N}_z = 0$. Die z-Komponente \mathbf{D}_z wird für das Matrixsystem(2) der RCWA nicht benötigt. Die beiden nötigen Gleichungen lauten somit:

$$\mathbf{D}_x = [[\epsilon]] \mathbf{E}_x + \epsilon_\Delta (\mathbf{N}_x \mathbf{N}_x \mathbf{E}_x + \mathbf{N}_x \mathbf{N}_y \mathbf{E}_y) \quad (10)$$

$$\mathbf{D}_y = [[\epsilon]] \mathbf{E}_y + \epsilon_\Delta (\mathbf{N}_y \mathbf{N}_x \mathbf{E}_x + \mathbf{N}_y \mathbf{N}_y \mathbf{E}_y)$$

Diese Formulierungen werden nun auf die letzten beiden Gleichungen in(2), in welchen die Verschiebungsdichte identifiziert wird, angewandt und somit kann formuliert werden:

$$\frac{\Delta U_y}{\Delta z} = \frac{k_y}{k_0} k_x S_y + k_0 \left([[\epsilon]] \mathbf{S}_x + \epsilon_\Delta [[\mathbf{N}_x^2]] \mathbf{S}_x + \epsilon_\Delta [[\mathbf{N}_x \mathbf{N}_y]] \mathbf{S}_x \right) - \frac{k_y}{k_0} k_y S_x \quad (11)$$

$$\frac{\Delta U_x}{\Delta z} = \frac{k_x}{k_0} k_x S_y - k_0 \left([[\epsilon]] \mathbf{S}_y + \epsilon_\Delta [[\mathbf{N}_y \mathbf{N}_x]] \mathbf{S}_x + \epsilon_\Delta [[\mathbf{N}_y^2]] \mathbf{S}_y \right) - \frac{k_x}{k_0} k_y S_x$$

Für die Definition der Wellenvektorkomponenten und der zugehörigen Matrizen sei an dieser Stelle ebenfalls auf Appendix 10 verwiesen. Umgeschrieben in Matrixform synthetisieren so die unteren beiden Reihen in Gleichung(2) zu

$$\begin{bmatrix} \partial \mathbf{U}_y / \partial(k_0 z) \\ \partial \mathbf{U}_x / \partial(k_0 z) \end{bmatrix} = \begin{bmatrix} \mathbf{K}_y \mathbf{K}_x + \epsilon_\Delta [[\mathbf{N}_x \mathbf{N}_y]] & \mathbf{E} - \mathbf{K}_y^2 + \epsilon_\Delta [[\mathbf{N}_x^2]] \\ \mathbf{K}_x^2 - \mathbf{E} - \epsilon_\Delta [[\mathbf{N}_y^2]] & -\mathbf{K}_x \mathbf{K}_y - \epsilon_\Delta [[\mathbf{N}_y \mathbf{N}_x]] \end{bmatrix} \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \end{bmatrix} \quad (12)$$

Analog zur ursprünglichen Formulierung in Gleichung(3) werden diese Gleichungen ineinander zu einem Differentialsystem zweiter Ordnung eingesetzt, sodass das Differentialgleichungssystem zweiter Ordnung nun stattdessen folgendermaßen anzugeben ist.

$$\begin{bmatrix} \partial^2 \mathbf{S}_y / \partial(k_0 z)^2 \\ \partial^2 \mathbf{S}_x / \partial(k_0 z)^2 \end{bmatrix} = \begin{bmatrix} \mathbf{K}_x^2 + \mathbf{K}_y \mathbf{E}^{-1} \mathbf{K}_x \epsilon_\Delta [[\mathbf{N}_x \mathbf{N}_y]] + \mathbf{D}(\mathbf{E} + \epsilon_\Delta [[\mathbf{N}_y^2]]) \\ \mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_y (\mathbf{E} + \epsilon_\Delta [[\mathbf{N}_y^2]]) - \mathbf{K}_y \mathbf{K}_x + \mathbf{B} \epsilon_\Delta [[\mathbf{N}_x \mathbf{N}_y]] \\ \mathbf{K}_y \mathbf{E}^{-1} \mathbf{K}_x (\mathbf{E} + \epsilon_\Delta [[\mathbf{N}_x^2]]) - \mathbf{K}_x \mathbf{K}_y + \mathbf{D} \epsilon_\Delta [[\mathbf{N}_y \mathbf{N}_x]] \\ \mathbf{K}_y^2 + \mathbf{K}_x \mathbf{E}^{-1} \mathbf{K}_y \epsilon_\Delta [[\mathbf{N}_x \mathbf{N}_y]] + \mathbf{B}(\mathbf{E} + \epsilon_\Delta [[\mathbf{N}_x^2]]) \end{bmatrix} \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \end{bmatrix} \quad (13)$$

Dieses System wird nun analog zur 2D-Standardformulierung als Eigenwertproblem gelöst. Die nun erhaltenen Gleichungen werden in(12) eingesetzt, um auch die Lösungen der magnetischen Feldamplituden \mathbf{U} zu erhalten. Alles weitere erfolgt wie in der ursprünglichen RCWA in [34].

4.0.2 Konvergenz und Verifikation

Da in [34] die RCWA auch im 2D-Fall in allen Varianten verifiziert und verglichen worden ist, kann diese zu Vergleich und Verifikation der neuen NVT-RCWA verwendet werden. Ferner wird an dieser Stelle ein Fall betrachtet, der in den Veröffentlichungen zur RCWA und insbesondere zur Normalenvektortransformation nicht zu finden ist. Üblicherweise wird in den im Kapitel 2 vorgestellten Veröffentlichungen eine rudimentäre Geometrie mit rein realem Brechungsindex simuliert. Die Ergebnisse der dort aufgeführten NVT-Formulierung sind jedoch schon mit dem ursprünglichen 3D-RCWA-Paket ohne NVT nachbildbar. Aus diesem Grund wird hier eine periodische Zelle mit scharfkantigem Metallelement und somit hohem imaginären Anteil des Brechungsindexes zur Simulation ausgewählt, ein Fall, wie er auch in den späteren Anwendungen von Teilausleuchtung und Raytracing zu sehen sein wird. Die Simulation einer solchen Struktur mit der ursprünglichen RCWA Formulierung konvergiert auf Desktop-PCs zum Teil nicht, weshalb sich von der NVT-Formulierung ein erheblicher Mehrgewinn erhofft wird.

Die hier betrachtete Beispielstruktur ist schon in Abbildung(9) gezeigt. Die Gitterperiode sei in x- und y-Richtung jeweils $1\mu\text{m}$ weit. Die Zelle ist in Draufsicht in Abbildung(9) mit dem Normalenvektorfeld gezeigt. Das gedrehte Rechteck (Drehwinkel 30°) aus Gold sei seitlich umgeben von Luft und $1\mu\text{m}$ dick. Ferner liegt über und unter der gezeigten Zelle noch eine 100nm dicke Luftsicht. Der obere und untere Halbraum sei mit Luft gefüllt. Eine ebene homogene Welle mit einer Wellenlänge von 550nm und einem linearen Polarisationswinkel von 45° trifft unter einem azimuthalen Winkel von 50° und einem polaren Winkel von 40° auf die gezeigte Struktur.

In Abbildung(10) sind zu einer ersten Aussage zur Konvergenz dieser neuen Formulierung die inneren Beugungseffizienzen über die Anzahl der verwendeten Beugungsordnungen aufgetragen. Wie sich die Beugungseffizienzen aus den berechneten Feldamplituden ableiten ist in Appendix 10 definiert. Die Angabe der Beugungsordnungszahl auf der x-Achse beschreibt

hierbei die Gesamtzahl der Ordnungen (negative + nullte + positive) einer Achse. Für das Simulationsgebiet sind in jeder Achse die gleiche Anzahl an Beugungsordnungen definiert. Die insgesamt auftretenden bzw. verwendeten Ordnungen ergeben sich aus der Multiplikation beider Anzahlen, in diesem Fall also dem Quadrat der angegebenen Zahl auf der x-Achse.

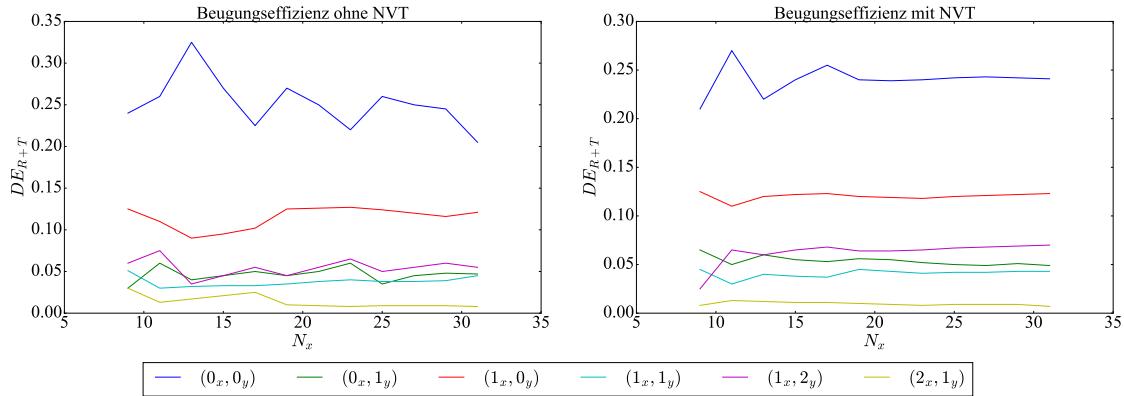


Abbildung 10: Beugungseffizienzen der inneren positiven Ordnungen über die Gesamtzahl der verwendeten Beugungsordnungen jeweils einer Achse. Links die Standard 3D-RCWA, rechts die 3D-RCWA inklusive NVT-Erweiterung

Die Betrachtung der Beugungseffizienzen als Konvergenzaussage ist eine gängige Methode in Veröffentlichungen zur RCWA wie beispielsweise in [32]. Die Konvergenz gilt hierbei als erreicht, wenn sich der Kurvenverlauf mit zunehmender Anzahl an verwendeten Beugungsordnungen nicht mehr ändert.

Hierbei ist im linken Teilbild von Abbildung(10) die langsamere Konvergenz der Standard RCWA ohne NVT gut erkennbar. Die 31 Beugungsordnungen einer Achse repräsentieren hierbei den letzten ermittelten Simulationswert. Werden mehr Ordnungen in der Simulation verwendet, so passen die Hauptkoeffizientenarrays, für welche beispielsweise ein Eigenwertproblem zu lösen ist, nicht mehr komplett in den Arbeitsspeicher der gegebenen Rechnerkonfiguration. In diesem Fall muss das Array gestückelt und jeweils kontinuierlich Teile des Arrays in den Arbeitsspeicher hinein und wieder heraus geladen werden. Hierbei steigt die Simulationszeit nahezu exponentiell an, da dies eine Variante des Cache-Miss Problems ist. Auf Gegebenheiten dieser Art bezieht sich auch wiederum eins der Ziele dieser Arbeit, Simulationen dieser Form "handhabbar" zu machen, was im gezeigten Fall verfehlt wird. Im gegebenen Bereich ist aufgrund der mit Erhöhung der Ordnungszahl nicht geringer werdenden Schwankungen direkt ersichtlich, dass diese Implementierung den Ansprüchen an Konvergenz und Abweichung nicht genügt.

Jedoch kann im rechten Teilbild mit der Anwendung der NVT eine erheblich schnellere Konvergenz ausgemacht werden. Als Konvergenzschwelle kann hier eine Ordnungszahl von 19 definiert werden. Ab dieser Ordnungszahl resultiert eine maximale absolute Abweichung in der Ordnung $(0,1)$ (grüne Kurve) mit 0.005. Simulationsergebnisse mit Ordnungszahlen größer oder gleich 19 können somit als genügend und verwendbar betrachtet werden. Die dennoch existierenden Schwankungen der Kurven ab diesem Punkt werden durch die Abbildung der schrägen Flanken durch diskretisierte rechteckige Blöcke verursacht.

Allgemein ist der Speicherbedarf der Hauptkoeffizientenmatrizen der RCWA proportional zur Gesamtzahl an Ordnungen in der Simulation in der vierfachen Potenz. Sei für einen kurzen Moment angenommen, das Ergebnis der alten Formulierung sei bei 31 Ordnungen

konvergent, was sichtlich nicht der Fall ist, reduziert sich so der Speicherbedarf bei 19 verwendeten Ordnungen auf 14 % des Bedarfes bei 31 Ordnungen. Bezogen auf die Rechenzeit ist das Ergebnis noch gravierender, selbige reduziert sich hier auf 10 %.

4.0.3 Vergleich mit FDTD

Zur Verifikation wird hier die Anordnung aus Abbildung(9) in einer FDTD-Simulation nachgebildet. Hierbei sind zur Festlegung der Simulationseinstellungen die Methoden aus Appendix 9 angewandt. Somit kann das hier präsentierte FDTD Simulationsergebnis als den für diese Arbeit allgemein definierten Ansprüchen an Konvergenz und Abweichung genügend betrachtet und als Referenz herangezogen werden. Die FDTD-Zelle hat hierbei die gleichen lateralen Abmessungen wie die RCWA Zelle, was aufgrund der Luftbereiche um die Goldzelle herum kein Konvergenzproblem bezüglich der FDTD-Randbedingungen darstellt. Ferner ist die FDTD-Zelle $4\mu\text{m}$ dick, wobei die Quellenfläche, welche die ebene homogene Welle emittiert, $1\mu\text{m}$ über der Goldzelle positioniert ist. Im ersten Schritt werden die Felder in einigen 100nm Abstand zum Goldplättchen über Monitorebenen in der FDTD ausgegeben und mit den entsprechenden Ergebnissen der RCWA verglichen. Die Beträge der x-Komponente des elektrischen und des magnetischen Feldvektors in einer Ebene 500nm über dem Goldplättchen sind in Abbildung(11) gezeigt.

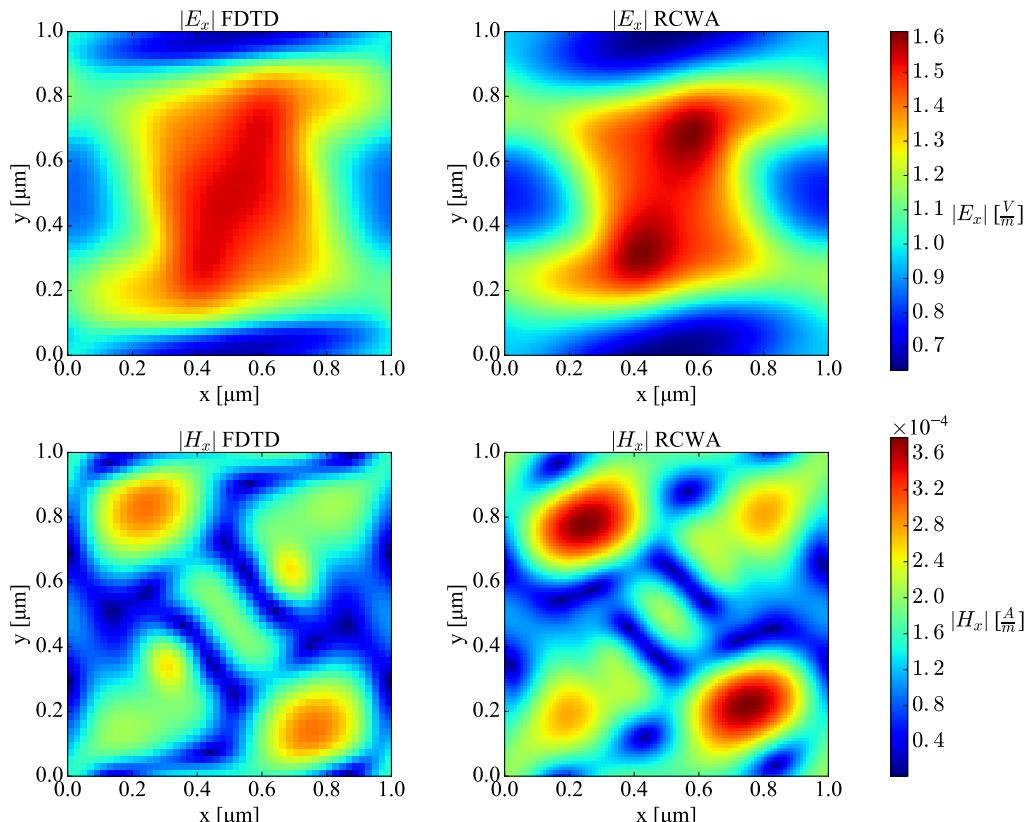


Abbildung 11: Betrag der x-Komponente des elektrischen (links) und magnetischen (rechts) Feldvektors in einer xy-Ebene 500nm über dem Goldplättchen. Links simuliert mit der FDTD, rechts mit der 3D-RCWA inklusive NVT

Die Wahl der Eingangswellenlänge ist an dieser Stelle auf 550nm gefallen, da bei der Simulation mit Permittivitäten in der Nähe einer Absorptionskante des Goldes stärkere Abweichungen als mit Permittivitäten fern einer der Absorptionskanten erwartet werden. So sind die Aussagen dieser Betrachtung zu Konvergenz und Abweichung triftiger.

Die RCWA ist mit den im Teilabschnitt zuvor benannten 19 Ordnungen pro Achse angesetzt, wobei die Konvergenz der RCWA bei dieser Ordnungszahl nach Abbildung (10) bereits gezeigt ist. In dieser Ansicht sind beide Felddarstellungen als vergleichbar und den Ansprüchen an die Abweichung als genügend zu betrachten. Lediglich in der Darstellung des E-Feldes der RCWA-Simulation sind im mittleren roten Bereich zwei dunkelrote Maxima statt einem durchgängigen wie in der FDTD zu erkennen. Die maximale absolute Abweichung von $0.03 \frac{V}{m}$ der gesamten Darstellung kann an diesen beiden Maximalpunkten ausgemacht werden. In der Darstellung des H-Feldes liegen die beiden Maximalwerte der RCWA ebenfalls über denen der FDTD, hierbei mit einer absoluten Abweichung von $3 * 10^{-5} \frac{V}{m}$.

Der Rechenzeit der RCWA Simulation von 12 Sekunden bei 19 Ordnungen pro Achse steht eine Simulationszeit der FDTD von 20 Minuten gegenüber. Der Mehrgewinn der RCWA für diese Anwendung ist somit hier direkt ersichtlich. Im Bezug zum Ziel dieser Arbeit, der Handhabbarkeit von Simulationen zur Durchführung von Berechnungen über ganze Eingangsparameterbereiche zeigt sich die RCWA insbesondere mit dieser NVT-Formulierung als geeignet.

In Abbildung(12) sind die gleichen Felddarstellungen für die Mitte des Goldplättchen gezeigt. Hierbei soll das Verhalten an den Metallkanten des Goldelementes untersucht werden, insbesondere da bei der festgelegten Wellenlänge die Permittivität des Goldes nahe einer der Absorptionskanten liegt.

Es ist zu erkennen, dass das Aufschwingen des Feldes an der Metallkante bei sofortigem Abklingen im Material in beiden Fällen gegeben ist. Der einzige erkennbare Unterschied liegt hierbei, dass dieser Effekt bei der Betrachtung des elektrischen Feldes an den Ecken des Goldplättchens in der FDTD zentrierter und um einen Absolutbetrag von $0.12 \frac{V}{m}$ höher zu sehen ist als in der RCWA. Dies ist auf die Darstellung des Feldes in der RCWA als Überlagerung von ebenen Wellen zurückzuführen, deren Amplituden nicht auf einen Punkt konzentriert werden können. Dennoch liegen alle Wertebereiche und Feldartefakte auf den gleichen Positionen.

An dieser Stelle ist das Ziel vor allem, die Lösungen in einigem Abstand von der eigentlichen Struktur zu berechnen, um sie in der Teilausleuchtung oder dem Raytracing durch einen homogenen Raum weiter zu propagieren. Somit fällt dieser Effekt in den folgenden Anwendungen nur wenig ins Gewicht.

4.1 Fazit

Mit der Erweiterung der Normalenvektortransformation im 2D-periodischen Gitterfall ist für das ursprüngliche 3D-RCWA Paket ein geeignetes Optimierungsverfahren gefunden worden. Mit dieser Implementierung es nun als verifiziert geltend möglich, in der gegebenen Rechenkonfiguration eine metallische Struktur zu simulieren, für die mit der vorherigen Implementierung der RCWA die Ansprüche an Konvergenz und Abweichung nicht erreicht worden sind. Bezogen auf die Auslegung von Metalinsen oder Folierungen als Motivation dieser Arbeit ist es nun realisierbar, beispielsweise die Einheitszellen der Metalinsen über die Parameterbereiche von Eingangswinkel und -wellenlänge in weniger als vier Stunden innerhalb der Konvergenzanforderungen zu simulieren. Mit der vorherigen Implementierung ohne NVT sind für die Gesamtzahl von 6400 Simulationen knapp 28 Stunden benötigt worden.

Der gezeigte Vergleich mit der FDTD-Simulation verifiziert ferner die Simulationsergebnisse der Felddarstellungen in den Bereichen über und unter der Struktur für die Anwendung im folgenden Kapitel.

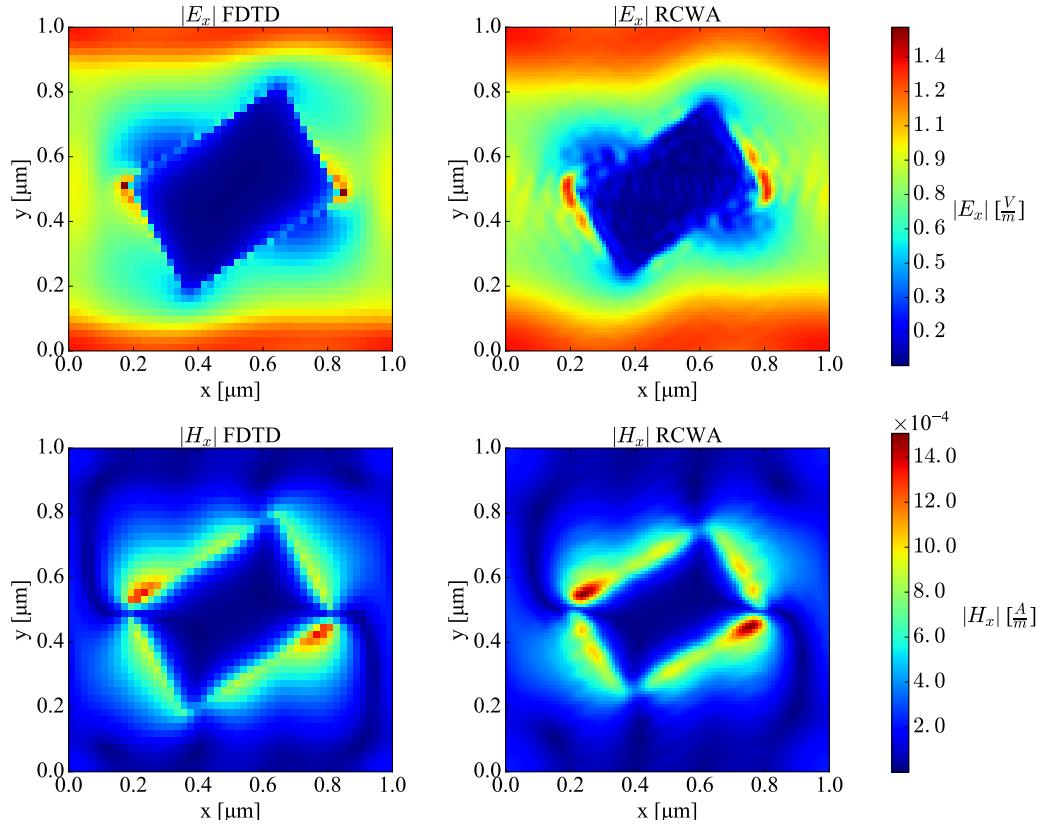


Abbildung 12: Betrag der x-Komponente des elektrischen (oben) und magnetischen (unten) Feldvektors in einer xy-Schnittebene genau in der Mitte des Goldplättchen. Links simuliert mit der FDTD, rechts mit der 3D-RCWA inklusive NVT

5 Teilausleuchtung

Dieses Kapitel widmet sich der Modellbetrachtung der Teilausleuchtung. Auf Grundlage der RCWA gilt hier die erweiterte Betrachtung einer inhomogenen Eingangsbeleuchtung auf eine periodische begrenzt ausgedehnte Struktur. Es werden zwei Verfahren vorgestellt, die später auch verifiziert und durch das gesamte Thema der Teilausleuchtung begleiten werden. Das erste Verfahren wird hier als Fouriermethode bezeichnet. Sie wurde in Kapitel 2 kurz vorgestellt und orientiert sich maßgeblich an der Arbeit in [30]. Die Fouriermethode wird in dieser Arbeit noch um einige Aspekte erweitert und an manchen Stellen andersartig formuliert, dient hier jedoch maßgeblich für Vergleich und Verifikation der eigenen Methode. Gleichzeitig ist in der Implementierung dieser Methode gemäß [30] ein Fehler entdeckt worden, dessen genaue Modalitäten sowie dessen Behebung in diesem Kapitel aufgezeigt werden. Das zweite Verfahren, im weiteren Verlauf als Feldbildsynthese bezeichnet, ist selbst aus der Problemstellung der Metalinsen heraus entwickelt worden. Ein Kernelement dieses Kapitels ist die Verifikation des neuen Verfahrens der Feldbildsynthese, wobei der Fokus insbesondere auf die Nahfeldbetrachtung im Vergleich zur FDTD sowie der Ermittlung der maximal möglichen lateralen Ausdehnung der zu simulierenden Struktur liegt. Letzteres ist gleichzeitig eines der Kernziele dieser Arbeit.

5.1 Modellbildung der Teilausleuchtung

Das grundlegende Konzept der Teilausleuchtung in Abbildung(13) gezeigt.

Im einfachsten Fall trifft in dieser Anordnung eine ebene homogene Welle auf die Struktur. Für diese Welle sei ein strahleigenes Koordinatensystem $Ox'y'z'$ definiert, in welchem die Welle nur in z' -Richtung mit $\exp(-jk_0n_1z')$, mit dem Brechungsindex n_1 im Halbraum über der Struktur, propagiert. Die rein lineare Polarisation ist in der $x'y'$ -Ebene durch den Winkel ψ definiert (hier könnte der Vollständigkeit halber noch der Polarisationswinkel δ wie in der RCWA-Zellensimulation hinzugenommen werden, um elliptische Polarisationen zu definieren; dieser Fall wird hier jedoch nicht betrachtet). Die Amplitude der einfallenden TEM-Welle, mit einem E-Feldvektor in der $x'y'$ -Ebene, sei über diese Ebene hinweg inhomogen. Beispielsweise entspreche sie einer Gaußverteilung mit dem Zentrum im Koordinatenursprung von $Ox'y'z'$. Der Übergang von $Ox'y'z'$ zum globalen Koordinatensystem $Oxyz$ der Struktur wird durch die Einfallsinkel θ und ψ dargestellt. Die Struktur selbst sei in ihrem Gesamtmaßen endlich in der xy -Ebene mit ihrer Oberfläche bei $z = 0$ angeordnet. Für das gesamte Feld wird eine Kantenlänge von L_x bzw. L_y definiert. In diesem Feld sind die periodischen Zellen der Anzahl Z_x bzw. Z_y bis zu einer gewissen Ausdehnung angeordnet. Die Zellennummern werden durch die Laufvariablen map_x bzw. map_y iteriert. Ferner wird in Abbildung(13) eine geometrische Form in der Ebene $z = 0$ definiert. Diese wird beispielhaft durch den äußeren angegraute Kreis dargestellt. Hier gilt, dass das Innere des Kreises mit den periodischen Zellen gefüllt sei und das Gebiet außerhalb ein perfekt absorbierendes Material sei. Feldanteile der Eingangsbeleuchtung, die auf dieses absorbierende Material treffen, werden so in der Implementierung auf Null gesetzt.

An dieser Stelle sei angemerkt, dass sich die folgenden Formulierungen des elektrischen Feldes in der Form $\tilde{\mathbf{E}}$ immer auf einen komplexen Feldvektor mit drei kartesischen Feldkomponenten beziehen.

Das Ziel ist nun, das elektrische Feld im gesamten Halbraum über und unter der Struktur zu berechnen. Im ersten Schritt wird hierzu das einfallende Feld der Eingangswelle auf der Struktur für jeden Diskretisierungspunkt (x,y) bei $z = 0$ berechnet. Der Zentrumspunkt der

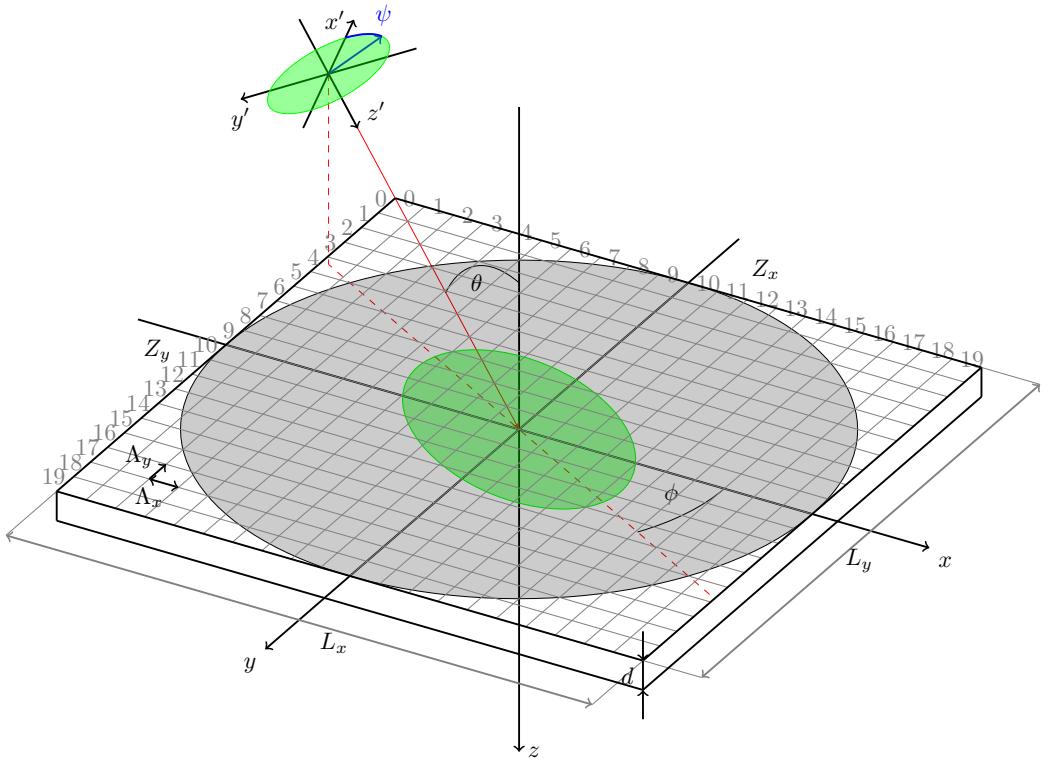


Abbildung 13: Vereinfachte Darstellung der Anordnung in der Teilausleuchtung: eine ebene homogene Welle mit begrenzter Amplitudenform (grün) fällt in schiefem Winkel auf eine räumlich begrenzte Zellenstruktur der Dicke d ein. Die periodischen Zellen mit Perioden Λ_x bzw. Λ_y werden durch die grauen Gitterlinien verdeutlicht.

Eingangswelle auf der Struktur ist beliebig. In Abbildung(13) entspricht er zur besseren Er-sichtlichkeit dem Koordinatenursprung der Struktur. Das Eingangsfeld bei $z = 0$ wird im Folgenden mit $\tilde{\mathbf{E}}_{ein}(x, y)$ bezeichnet. Der in Abbildung(13) gezeigte Fall stellt eine idealisierte Eingangsbeleuchtung dar, bei der aufgrund der ebenen Phasenfront die Eingangs- und Polarisationswinkel an jedem Punkt (x, y) der Fläche $z = 0$ gleich sind. Lediglich Feldbetrag und Phase müssen an jedem Punkt mit Hilfe von Koordinatentransformationen individuell berechnet werden; dies wird im Verlauf dieses Kapitels auch exemplarisch gezeigt. Diese Art der Eingangsbeleuchtung wird im Folgenden als Teilausleuchtung ohne Beugung bezeichnet. In dieser Anordnung und den Verifikationsbeispielen werden jedoch auch Eingangsfelder betrachtet, bei welchen auch die Winkel an jedem Punkt (x, y) unterschiedlich sind. Ein Beispiel hierzu ist die Gaußsche Grundmode (Appendix 11) mit der Gleichung im strahleigenen Koordinatensystem $O'x'y'z'$ aus [30]. Diese Strahlform wird auch in einigen folgenden Verifikationsbeispielen als Eingangsbeleuchtung verwendet. In dieser Eingangsbeleuchtung existiert eine ortsabhängige Krümmung der Phasenfront. Die Strahlform weitet sich in Propagationsrichtung aus. Deshalb wird im folgenden immer von einer Teilausleuchtung mit Beugung gesprochen, wenn die Eingangswinkel an der Ebene (x, y) sich über die Fläche hinweg unterscheiden.

5.1.1 Teilausleuchtung mit Fouriermethode

Die Formulierung der Fouriermethode basiert maßgeblich auf [30]. Die Grundidee dieser Methode ist es, beliebige räumlich inhomogene Feldverteilungen durch eine Fourierreihe darzustellen. Wie im eindimensionalen Fall, in welchem eine beliebige Funktion durch eine endliche Fourierreihe angenähert werden kann, ist dies auch im dreidimensionalen Fall möglich. Das besondere hierbei ist, dass jeder einzelne skalierte Exponentialterm einer ebenen homogenen Welle entspricht. Die Überlagerung dieser ebenen Wellen durch die Reihensumination nähert so die tatsächliche, inhomogene Feldverteilung an. Der Nutzen dieser Formulierung ist nun, dass gleichzeitig eine ebene homogene Welle dem Feldeingang einer RCWA entspricht. So mit kann jede einfallende homogene Welle, abhängig von ihren Winkeln und skaliert durch Eingangsphase und -amplitude, mit einer RCWA-Simulation verknüpft werden. Das Ausgangsfeld wird nun auf exakt gleiche Weise als Superposition formuliert, nur dass die ebenen Wellen den ausgehenden Beugungsordnungen der RCWA entsprechen.

Zu Beginn wird die Simulationsebene (x, y) bei $z = 0$ diskretisiert und daraus die nötigen Richtungsvektoren dieser Ebene für die Fouriertransformation abgeleitet:

$$\begin{aligned} k_{xn} &= n \frac{2\pi}{L_x} \quad \text{mit} \quad n = -M_x // 2 \dots M_x // 2 \\ k_{ym} &= m \frac{2\pi}{L_y} \quad \text{mit} \quad m = -M_y // 2 \dots M_y // 2 \end{aligned} \quad (14)$$

wobei M_x bzw. M_y die Gesamtzahl der Diskretisierungspunkte der jeweiligen Achse beschreiben. Der Operator $//$ definiert hierbei die Ganzahldivision. Die Vektorkomponente k_z wird hier aufgrund der Eingangsrichtung umgekehrt nach

$$k_{zmn} = \begin{cases} \sqrt{(k_0 n_1)^2 - k_{xn}^2 - k_{ym}^2} & \text{für } k_0 n_1 > \sqrt{k_{xn}^2 + k_{ym}^2} \\ 0 & \text{für } k_0 n_1 < \sqrt{k_{xn}^2 + k_{ym}^2} \end{cases} \quad (15)$$

formuliert. Vektorkomponenten von $k_z = 0$ entsprechen hierbei nicht ausbreitungsfähigen Wellen der Fourierreihe und die zugehörigen Eingangsfeldkomponenten werden in der Implementierung zu Null gesetzt. Nun wird das schon vorgestellte Eingangsfeld $\tilde{\mathbf{E}}_{ein}(x, y)$ über die in Abbildung(13) gezeigte Fläche äquidistant diskretisiert mit den Gesamtzahlen der Diskretisierungspunkte M_x bzw. M_y zu $\tilde{\mathbf{E}}_{ein}(x_i, y_j)$. Anschließend wird in den räumlichen Fourierbereich transformiert:

$$\vec{\mathcal{F}}_{ein}(k_{xn}, k_{ym}) = \frac{1}{M_x M_y} \sum_{i=-M_x//2}^{M_x//2} \sum_{j=-M_y//2}^{M_y//2} \tilde{\mathbf{E}}_{ein}(x_i, y_j) \exp(1j(k_{xn} x_i + k_{ym} y_j)) \quad (16)$$

$1j$ entspricht hierbei der komplexen Einheit. Jede Wellenvektorkombination entspricht geometrisch einer Richtung, welche sich durch jeweils einen azimuthalen, polaren und Polarisationswinkel darstellen lässt. Unter Berücksichtigung der Richtung des Eingangsfeldes berechnen sich die ersten beiden Winkel nach:

$$\theta_{m,n} = \cos^{-1}\left(\frac{k_{zmn}}{k_0 n_1}\right) \quad \text{und} \quad \phi_{m,n} = \tan^{-1}\left(\frac{k_{ym}}{k_{xn}}\right) \quad (17)$$

Ferner muss an dieser Stelle auch der lineare Polarisationswinkel für jede ebene Welle der Fourierreihe individuell berechnet werden. Dessen etwas aufwendigere Berechnung ist in Appendix 11 ausformuliert. Im einfachen Fall einer inhomogenen Eingangsbeleuchtung ohne

Beugung muss von diesen drei Eingangswinkeln jeweils immer nur der Eingangswinkel der in Abbildung(13) gezeigten z'-Achse berücksichtigt werden.

Zum Verständnis werde an dieser Stelle für einen kurzen Moment angenommen, die Struktur in Abbildung(13) sei eine perfekt reflektierende Spiegelfläche. Es wird nun versucht, das reflektierte Feld im Halbraum über der Struktur zu definieren. Das reflektierte Ausgangsfeld im Fourierraum bei $z = 0$ kann direkt angegeben werden zu $\tilde{\mathbf{E}}_{RSF}(k_{xn}, k_{ym}) = \tilde{\mathbf{1}} * \tilde{\mathcal{F}}_{ein}(k_{xn}, k_{ym})$, da bei einem Spiegel das Eingangsfeld lediglich um einen Phasensprung von π gedreht wird. $\tilde{\mathbf{E}}_{RSF}$ bezeichnet hierbei das Ausgangsfeld im Fourierraum und der Vektor $\tilde{\mathbf{1}}$ die Faktorisierung durch den ideal reflektierenden Spiegel. Das Ausgangsfeld bei einer Koordinate $z < 0$ lässt sich nun mit einem Propagationsterm in z -Richtung im Fourierraum und anschließender Rücktransformation berechnen.

$$\tilde{\mathbf{E}}_{RS}(x, y, z) = \sum_{n=-M_x//2}^{M_x//2} \sum_{m=-M_y//2}^{M_y//2} [\tilde{\mathbf{E}}_{RSF}(k_{xn}, k_{ym}) \exp(jk_{zmn}z)] \exp(-1j(k_{xn}x + k_{ym}y)) \quad (18)$$

Die Propagationsvektorkomponente ergibt sich standardmäßig zu

$$k_{zmn} = \begin{cases} \sqrt{(k_0 n_1)^2 + k_{xn}^2 + k_{ym}^2} & \text{für } k_0 n_1 > \sqrt{k_{xn}^2 + k_{ym}^2} \\ -j \sqrt{k_{xn}^2 + k_{ym}^2 - (k_0 n_1)^2} & \text{für } k_0 n_1 < \sqrt{k_{xn}^2 + k_{ym}^2} \end{cases} \quad (19)$$

Die komplexe Einheit zeigt an dieser Stelle ebene Wellen mit zugehöriger Raumrichtung an, die nicht ausbreitungsfähig sind und direkt mit einem reellen Exponentialterm abklingen. Diese Spiegelkonfiguration wird zu einem späteren Zeitpunkt in der Verifikation angewandt, um die Koordinatentransformationen, Strahlformulierungen und Propagationen der zugehörigen Implementierung zu testen.

Nun wird der Effekt einer tatsächlichen periodischen, im einfachsten Fall erst eindimensionalen, Gitterstruktur mit einbezogen. Für eine gegebene Richtung (k_{xn}, k_{ym}) , zu der nach den obigen Formeln der zugehörige azimuthale, polare und Polarisationswinkel bekannt ist, wird eine RCWA-Simulation durchgeführt. Als Ergebnis stehen nun für jede Richtung (k_{xn}, k_{ym}) genau N_x komplexe Ausgangsamplitudenvektoren $\tilde{\mathbf{R}}_i$ für den oberen bzw. nochmals N_x komplexe Ausgangsamplitudenvektoren $\tilde{\mathbf{T}}_i$ für den unteren Halbraum zur Verfügung. Hierbei bezeichnet N_x die Gesamtzahl an Beugungsordnungen der RCWA-Simulation. Die zugehörigen Wellenvektoren der RCWA k_{xi} und k_y bestimmen die Richtungen der ausgehenden ebenen homogenen Wellen (Beugungsordnungen). Nun werde nochmals Gleichung(18) betrachtet. Statt mit dem konstanten Faktor 1 wie im Spiegelfall werden nun an jede Richtung (k_{xn}, k_{ym}) die zugehörigen reflektiven Ordnungsamplituden gekoppelt. An dieser Stelle unterscheiden sich Formulierung und Vorgehen maßgeblich von der Veröffentlichung dieser Methode in [30]. Die hier gezeigte Reihenfolge und Ausführung der Kopplung geschieht in anderer Reihenfolge und über eine Faltung. In [30] werden im Ergebnis nur die einzelnen Ordnungen bei einem genau zentralen Einfall des Gaußstrahls betrachtet. Wird jedoch die Formulierung in [30] angewandt, so sind zwar die ausgehenden Strahlformen richtig, liegen jedoch immer zentral im Mittelpunkt, was bei der Betrachtung eines schrägen Einfalls in einem Abstand von der Struktur intuitiv nicht korrekt ist. Diese Gegebenheiten ergeben sich aus den Testungen und Fehlerkorrekturen mit Hilfe der Methoden im Abschnitt Verifikation und Vergleich diesen Kapitels.

Der Koppelterm wird hier definiert nach:

$$\tilde{\mathbf{R}}_{zi} = IFFT2 \left(\vec{\mathcal{F}}_{ein}(k_{xn}, k_{ym}) * \tilde{\mathbf{R}}_i(k_{xn}, k_{ym}) \exp(1jk_{1zmniz}) \right) \quad (20)$$

Die Fourierrücktransformation an dieser Stelle geschieht über die Indizes n und m auf gleiche Weise wie beispielsweise in Gleichung(18), wobei sich in diesem Fall der Propagationsterm k_{1zmn} nach

$$k_{1zmn} = \begin{cases} \sqrt{(k_0 n_1)^2 + k_{xmn}^2 + k_{ymn}^2} & \text{für } k_0 n_1 > \sqrt{k_{xmn}^2 + k_{ymn}^2} \\ -j \sqrt{k_{xmn}^2 + k_{ymn}^2 - (k_0 n_1)^2} & \text{für } k_0 n_1 < \sqrt{k_{xmn}^2 + k_{ymn}^2} \end{cases} \quad (21)$$

ergibt. k_{xmn} und k_{ymn} beschreiben hierbei jeweils die ausgehenden Wellenvektoren k_x und k_y der RCWA für die jeweilige Eingangskombination (k_{xn}, k_{ym}) . Das Gesamtfeld ergibt sich nun über die Summation über die Ordnungen unter Berücksichtigung der Beugungsordnungs-wellenvektoren

$$\tilde{\mathbf{E}}_R(x, y, z) = \sum_{i=-N_x//2}^{N_x//2} \tilde{\mathbf{R}}_{zi} \exp(-1j2\pi(-jx/\Lambda_x)) \quad (22)$$

Analog lässt sich für das transmittierte Feld im Halbraum unter der Struktur definieren:

$$\tilde{\mathbf{T}}_{zi} = IFFT2 \left(\vec{\mathcal{F}}_{ein}(k_{xn}, k_{ym}) * \tilde{\mathbf{T}}_i(k_{xn}, k_{ym}) \exp(-1jk_{3zmn}(z-d)) \right) \quad (23)$$

Hierbei berechnet sich k_{3zmn} nach Gleichung(21), wobei hier n_1 durch n_3 zu ersetzen ist.

$$\tilde{\mathbf{E}}_T(x, y, z) = \sum_{i=-N_x//2}^{N_x//2} \tilde{\mathbf{T}}_{zi} \exp(-1j2\pi(-jx/\Lambda_x)) \quad (24)$$

Im 2D-Gitterfall fließen generell die Amplituden $\vec{\mathcal{R}}_{j,i}$ sowie ebenfalls die zugehörigen Wellenvektorkomponenten über den Index j mit der Anzahl N_y Ordnungen in diese Formulierungen auf gleiche Weise ein.

5.2 Teilausleuchtung mit Feldbildsynthese

In diesem Abschnitt wird ein alternatives Verfahren zur gezeigten Fouriermethode vorgestellt. Dieses Verfahren ist, wie eingangs schon erläutert, aus der Problemstellung einer Metallinse heraus entwickelt worden. Der Mehrgewinn dieses neuen Verfahrens liegt insbesondere in der Gesamtgröße der Hauptkoeffizientenmatrix, welche mindestens um den Faktor $N_x N_y$ kleiner ist als die der Fouriermethode. Insbesondere für die Zielsetzung der Feldsimulation von diffraktiven Elementen in der Größenordnung von mehreren Millimetern ist dieses daher geeignet.

Begonnen wird wieder mit dem Eingangsfeld $\tilde{\mathbf{E}}_{ein}(x, y)$ an der Oberfläche der Struktur bei $z = 0$. Hier wird das Eingangsfeld jedoch äquidistant auf die Zellen zur Zellenanzahl Z_x bzw. Z_y einer Achse zu $\tilde{\mathbf{E}}_{ein}(x_n, y_m)$ diskretisiert, wobei jeder diskretisierte Feldpunkt in der Mitte der jeweiligen Zelle liegt. Ferner wird für jede Zelle am selben diskretisierten Ort eine Eingangswinkelkombination sowie zugehörige Amplitude und Phase berechnet. Diese orientiert sich jedoch nicht wie bei der Fouriermethode an den Raumvektoren k_x und k_y der Fouriertransformation, sondern an den Normalenvektoren der Phasenfläche des aktuell betrachteten Punktes. Im einfachsten Fall einer einfallenden Amplitudenverteilung ohne Beugung ist diese

Eingangswinkelkombination für alle drei Eingangswinkel für jede Zelle gleich. Sie entspricht dem globalen azimuthalen, polaren und Polarisationseingangswinkel.

Im Falle eines einfallenden Strahls mit Beugung wie beispielsweise der vorgestellten Gaußmode (Appendix 11) gestaltet sich die Berechnung der Eingangswinkelfelder aufwendiger. Im strahleigenen Koordinatensystem $Ox'y'z'$ wird für jeden diskretisierten Punkt der Vektor $(x'_n, y'_m, \sqrt{R_{xz}^2 - (x'^2_n + y'^2_m)} sgn(z'_n))$ als Normalenvektor der differentiellen Phasenfläche dieses Punktes formuliert und anschließend über eine Kugelkoordinatentransformation mit ϕ_e und θ_e in das globale Koordinatensystem $Oxyz$ überführt. Dort wird mit Hilfe des Skalarproduktes der resultierende Vektor jeweils auf die Einheitsvektoren der x-, y- und z-Achse zu den Komponenten k, p und t projiziert. Anschließend ergeben sich die Eingangswinkelkarten zu

$$\theta_{m,n} = \cos^{-1}(t_{m,n}) \quad \text{und} \quad \phi_{m,n} = \tan^{-1}\left(\frac{p_{m,n}}{k_{m,n}}\right) \quad (25)$$

Der Polarisationswinkel $\psi_{m,n}$ eines jeden Diskretisierungspunktes ergibt sich ebenfalls nach Formel(57) bzw. (58) in Appendix(11), nur dass in dieser Methode der Subwellenvektor $\hat{\mathbf{k}}_{xn,ym}$ über die eben berechneten Eingangswinkelfelder definiert wird.

$$\hat{\mathbf{k}}_{xn,ym} = (\sin(\theta_{m,n})\cos(\phi_{m,n}), \sin(\theta_{m,n})\sin(\phi_{m,n}), \cos(\theta_{m,n})) \quad (26)$$

Im nächsten Schritt wird ein Parameterdurchlauf der mit dieser Vorschrift berechneten und somit benötigten Eingangswinkel im 3D-RCWA-Paket inklusive der NVT aus Kapitel4 durchgeführt. Hierbei wird zusätzlich zur Ausgabe der Beugungseffizienzen oder der Amplituden für die Fouriermethode auch das Feldbild einer kartesischen Komponente des E oder H Feldes in einer Schnittebene, welche vollständig innerhalb der RCWA-Struktur liegt, ausgegeben. Es ist somit möglich, beispielsweise über bzw. unter den eigentlichen Elementen eine 100nm dicke homogene Materialschicht in der RCWA zu definieren. Der Brechungsindex dieser Schicht muss mit dem Index des jeweiligen Halbraums in der Teilausleuchtung übereinstimmen aufgrund der homogenen Freiraumpropagatoren dieser beiden Methoden. In den meisten Anwendungen ist diese Schicht mit Luft gefüllt. Das Feld wird nun in einer xy-Ebene bei einer z-Koordinate innerhalb dieser beiden Schichten zu $\tilde{\mathbf{R}}_{i,j}(\theta, \phi, \psi)$ bzw. analog $\tilde{\mathbf{T}}$ berechnet und abgespeichert. Hierbei iterieren i bzw. j die Anzahl an Ausgabepunkten p_x bzw. p_y pro Achse. Die Winkel θ, ϕ, ψ geben die zugehörigen Eingangswinkel dieser Amplitude an. Es ist zu beachten, dass vor der Verwendung des reflektiven Feldbildes innerhalb der 3D-RCWA Simulation die einfallende ebene Welle $\tilde{\mathbf{E}}_{ein,i,j}$ von dem berechneten Feld nach $\tilde{\mathbf{R}}_{i,j} = \tilde{\mathbf{R}}_{RCWA,i,j} - \tilde{\mathbf{E}}_{ein,i,j}$ subtrahiert wird. Andernfalls würde in der Feldbildsynthese das Eingangsfeld im oberen Halbraum der Teilausleuchtung mit propagiert werden. Anschließend werden die RCWA-Feldbilder in Abhängigkeit von den berechneten Eingangswinkeln über die Fläche $z = 0$ hinweg ihren Zellen zugewiesen. Das resultierende Feld wird nun mit der Eingangsbeleuchtung im Reflektionsfall $\tilde{\mathbf{E}}_{R0}$ bei $z = 0$ formuliert.

$$\begin{aligned} \tilde{\mathbf{E}}_{R0}(x_q, y_p, z = 0) &= \tilde{\mathbf{R}}_{i,j}(\theta_{m,n}, \phi_{m,n}, \psi_{m,n}) * \tilde{\mathbf{E}}_{ein}(n, m) \\ \text{mit } m &= p//spc_y, \quad n = q//spc_x, \quad j = p\%spc_y, \quad i = q\%spc_x \\ \text{und } p &= -Z_y spc_y // 2 ... Z_y spc_y // 2, \quad q = -Z_x spc_x // 2 ... Z_x spc_x // 2 \end{aligned} \quad (27)$$

Die Parameter spc_x bzw. spc_y beschreiben hierbei die in der Methode berücksichtigten Anzahlen der Punkte der eingelesenen Feldbilder, der Operator $\%$ entspricht der Modulusoperation.

Um nun das Ausgangsfeld an einem beliebigen Punkt im oberen Halbraum zu berechnen, wird mit Hilfe der Fouriertransformation und dem Propagationsterm formuliert:

$$\tilde{\mathbf{E}}_{Rz}(x_q, y_p, z) = \text{IFFT2}((\text{FFT2}(\tilde{\mathbf{E}}_{R0}) * \exp(1jk_{1z}z))) \quad (28)$$

In dieser Hin- und Rücktransformation sei derweil der Normierungsfaktor der Fouriertransformation enthalten. Der Propagationsterm k_{1z} ergibt sich zu:

$$k_{1z} = \begin{cases} \sqrt{(k_0 n_1)^2 + k_{xpq}^2 + k_{ypq}^2} & \text{für } k_0 n_1 > \sqrt{k_{xpq}^2 + k_{ypq}^2} \\ -j \sqrt{k_{xpq}^2 + k_{ypq}^2 - (k_0 n_1)^2} & \text{für } k_0 n_1 < \sqrt{k_{xpq}^2 + k_{ypq}^2} \end{cases} \quad (29)$$

Wichtig an dieser Stelle ist der Umgang mit den Linien bei den Koordinaten $x = 0$ bzw. $y = 0$. Eine unsymmetrische 2D-Fouriertransformation bei Verschiebung des Spektrums um die zentrale Null herum erhält die Symmetrie der Eingangsstruktur. Somit sind die jeweiligen Nulllinien nur mit dem Rand einer Zelle zu belegen, um die Feldabtastpunkte symmetrisch um die x- bzw. y-Achse herum anzuordnen. Analog gilt für den unteren Halbraum:

$$\tilde{\mathbf{E}}_{T0}(x_q, y_p, z = d) = \tilde{\mathbf{T}}_{i,j}(\theta_{m,n}, \phi_{m,n}, \psi_{m,n}) * \tilde{\mathbf{E}}_{ein}(n, m) \quad (30)$$

$$\tilde{\mathbf{E}}_{Tz}(x_q, y_p, z) = \text{IFFT2}((\text{FFT2}(\tilde{\mathbf{E}}_{T0}) * \exp(-1jk_{3z}(z - d)))) \quad (31)$$

wobei sich in diesem Fall der Propagationsterm k_{3z} nach

$$k_{3z} = \begin{cases} \sqrt{(k_0 n_3)^2 + k_{xpq}^2 + k_{ypq}^2} & \text{für } k_0 n_3 > \sqrt{k_{xpq}^2 + k_{ypq}^2} \\ -j \sqrt{k_{xpq}^2 + k_{ypq}^2 - (k_0 n_3)^2} & \text{für } k_0 n_3 < \sqrt{k_{xpq}^2 + k_{ypq}^2} \end{cases} \quad (32)$$

formuliert. Dieses Verfahren ist, wie direkt ersichtlich, nicht physikalisch trivial, da Feldbilder mit unterschiedlicher Anregung aneinander gesetzt werden und so Sprünge der Amplituden an den Zellenrändern entstehen. Die Anwendung dieser Methode in der Simulation von Metalinsen [34] bei senkrechtem Einfall zeigt jedoch eindeutig die korrekte Funktion dieser Methode. Eine der ersten Annahmen für diese Methode ist, dass sie nur konvergent sei, falls die Beugung der Eingangsbeleuchtung und somit die Variation der Eingangsphase gering ist, da dann die Übergänge zwischen den Zellen besonders im Bezug auf die Phase stetiger sind. Wann genau dieses Modell konvergiert und anwendbar ist, wird deshalb in den folgenden Teilkapiteln der Verifikation genauer untersucht.

5.3 Symmetrie der Struktur

An dieser Stelle wird kurz die Ausnutzung der Symmetrie bezüglich des azimuthalen Eingangswinkels ϕ_e von Strukturen vorgestellt, wie sie in vielen der simulierten Strukturen wie beispielsweise der Einheitszellen von Metalinsen vorherrscht und in beiden Methoden der Teilausleuchtung sowie im folgenden Kapitel 6 angewandt wird. Abbildung(14) zeigt die Symmetrievorschrift einer 2D-periodischen RCWA-Struktur mit 45° Symmetrie, wie sie beispielsweise in der 2D-Goldstruktur im folgenden Verifikationsabschnitt ausgenutzt wird. Das rechteckige Kästchen entspricht einem Block gleicher Kantenlängen, der mittig in einer 2D periodischen Zelle mit $\Lambda_x = \Lambda_y$ angeordnet ist. Die kleine kreis- bzw. rechteckförmige Markierung auf dem Kästchen zeigt die nötige Drehung bzw. Transponierung des Kästchens an.

In der RCWA Simulation ist so nur ein Parametersweep bezüglich des Eingangswinkels ϕ_e

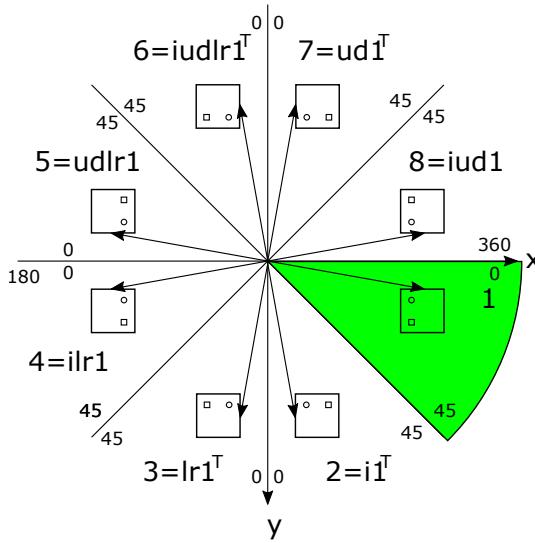


Abbildung 14: 45° Symmetrie einer 2D-periodischen RCWA-Struktur bezüglich des azimuthalen Eingangswinkels ϕ_e ; lediglich der grüne Bereich ist tatsächlich in der RCWA zu simulieren, alle weiteren Abschnitte werden aus den gezeigten Symmetrievorschriften synthetisiert.

von 0° bis 45° durchzuführen. Die restlichen azimuthalen Eingangswinkel berechnen sich aus diesen Simulationsergebnissen nach den gezeigten Vorschriften aus dem grünen Hauptbereich mit der Markierung 1. Die Markierung T bedeutet, dass in diesem Eingangswinkelbereich die zugehörigen RCWA-Feldbilder bzw. die Matrix der Amplituden in der Fouriermethode in der Zuweisung zum Bildarray zu transponieren sind. Analog gilt für lr (left-right) und ud (up-down), dass die Feldbilder bzw. die Matrix bezüglich der ersten bzw. zweiten Achse umzukehren sind. Um die Winkelreihenfolge einzuhalten, besteht in jedem zweiten Abschnitt die Markierung i (invers), wo die Reihenfolge des Winkelbereichs umzukehren ist. In der Python Implementierung des Teilausleuchtungspaketes wird so beispielsweise der Abschnitt 6 von Abbildung(14) mit `E_refl_sample[theta_mask, phi_xy_inv_trans, psi_mask, ::-1, ::-1].transpose(0,1,3,2)` berücksichtigt, wobei theta_mask bzw. psi_mask die Maskierindizes des polaren bzw. Polarisationswinkelfeldes beschreiben. `phi_xy_inv_trans` bezeichnet hierbei die Indexmaske, die für den in Abbildung(14) gezeigten Abschnitt 6 gültig ist. “`::-1`” entspricht der Invertierung der Elementereihenfolge der jeweiligen Achse, hier somit der Operation ud bzw. lr. Einer der Vorteile dieser Methode ist, dass die gesamte Zuweisung und Umordnung der Datenarrays über die weitere Arrays mit den jeweiligen Indizes bzw. Datenschlüsseln geschieht. So müssen die eigentlichen Datenwerte erst bei der Erstellung des Hauptarrays geladen werden. Neben der Lese und Speicherzeit, die auf Karten (Maps) mit Datenschlüsseln um einen Faktor größer 100 schneller ist als Lese- oder Schreiboperationen der tatsächlichen Daten, ist es so auch möglich, bei besonders großen Strukturen, für die Daten der gesamten in der RCWA simulierten Parameterräume nicht in den Arbeitsspeicher der gegebenen Rechnerkonfiguration passen würden, dennoch nur die benötigten Daten nach dem Mapping zu laden. Dies erhöht die maximal mögliche laterale Größe der zu simulierenden Struktur erheblich.

5.4 Verifikationsmethoden der Teilausleuchtung

Die beiden in den Teilkapiteln zuvor präsentierten Simulationsmethoden sind besonders für Feldsimulationen diffraktiver Strukturen der geforderten Größenordnung dieser Arbeit konzipiert. Dieser Mehrgewinn ist jedoch auch gleichzeitig ein Problem für die Verifikation dieser Verfahren. Besonders die Methode der Feldbildsynthese ist hier neu entwickelt und muss somit, bevor sie verlässlich eingesetzt werden kann, erst noch verifiziert werden. Typischerweise wird bei dieser Art von Verifikation idealerweise eine Modellstruktur exakt in einer Referenzsimulation basierend auf einem anderen Verfahren aufgesetzt und anschließend beispielsweise die Feldplots beider Simulationen miteinander verglichen. Wie auch schon in der Verifizierung der reinen RCWA würde, wie im Kapitel Methodik schon erläutert, eine FDTD-Simulation am geeignetsten verwendet. Die Größenordnung der Zielstrukturen für die beiden Methoden der Teilausleuchtung liegt jedoch in einem Bereich, welcher mit einer FDTD bei weitem nicht mehr simulierbar ist. Aus diesem Grund wurden mehrere Stufen an Teilverifikationen definiert. Falls alle Ansprüche der einzelnen Stufen für sich erreicht werden, gilt insbesondere die neue Methode der Feldbildsynthese als verifiziert und einsetzbar. Diese Stufen sind:

Spiegelkriterium

In diesem Fall wird zur Berechnung des reflektierten Feldes angenommen, die Struktur sei ein perfekt leitender Spiegel. Für das transmittierte Feld sei die Struktur Luft. So können durch Betrachtung des Ausgangsfeldes die Propagatoren und die Winkelberechnung etc. der Implementierung überprüft werden, da die analytischen Lösungen beispielsweise des Ausgangsstrahls an dem jeweiligen Ort bekannt sind sowie erwartete Symmetrien erfüllt werden müssen.

FDTD-Nahfeld

Eine Beispielstruktur wird in der für die FDTD maximal möglichen Größe bei der gegebenen Rechnerkonfiguration aufgesetzt. Als Referenz wird dieselbe Struktur in beiden Methoden der Teilausleuchtung simuliert und anschließend Feldplots in einer Ebene nah über bzw. unter der Struktur direkt miteinander verglichen. Die Abweichung zwischen den drei so erhaltenen Darstellungen soll den Konvergenzansprüchen aus dem Kapitel 3 genügen. Vorausgehend ist hier eine Untersuchung der nötigen Diskretisierung von Raum und Zeit sowie der zu setzenden Randbedingungen in der FDTD notwendig, um die abgeschlossene Konvergenz der als Referenz gesetzten FDTD-Lösung zu garantieren.

FDTD-Fernfeld

Das Ausgangsfeld derselben Struktur aus der Stufe zuvor wird in einem Abstand von einer Größenordnung über dem Nahfeld unter Zuhilfenahme der Fernfeldbetrachtung der FDTD mit der Teilausleuchtung verglichen.

Ordnungsposition

Der Fourieroptik [35], welcher auch die hier gezeigte Referenzmethode zu Grunde liegt, beinhaltet als fundamentale Aussage, dass jede beliebige Feld- oder Strahlform durch eine Reihe ebener homogener Wellen im Raum dargestellt werden kann. Ferner kann in ihr die Propagation eines Strahls durch ein Bauelement über die Faltung beschrieben werden. In der Fouriertransformation wird allgemein eine Faltung in eine Multiplikation überführt. Ferner kann ein Gitter im Fourieraum allgemein als Dirackamm (ebene homogene Wellen im Raum mit diskreten Abständen der Wellenvektorkomponenten) dargestellt werden. Somit kann grundsätzlich bei der Propagation eines Gaußstrahls

auf ein Gitter erwartet werden, dass sich das Ausgangsfeld aus verzerrten Gaußverteilungen zentriert um die Ordnungspositionen nahe der nullten Ordnung des Gitters zusammensetzt, da die Gaußfunktion gegenüber der Fouriertransformation invariant ist. Diese Positionen lassen sich über die Wellenvektoren der Gitter, hier exemplarisch für den oberen reflektiven Halbraum nach

$$k_x = k_0 \left(n_1 \cos \phi_e \sin \theta_e - n \frac{\lambda}{\Lambda_x} \right) \quad k_y = k_0 \left(n_1 \sin \phi_e \sin \theta_e - m \frac{\lambda}{\Lambda_y} \right) \quad (33)$$

$$k_z = \begin{cases} \sqrt{(k_0 n_1)^2 - k_x^2 - k_y^2} & \text{für } k_0 n_1 > \sqrt{k_x^2 + k_y^2} \\ 0 & \text{sonst} \end{cases}$$

$$x = \frac{k_x}{k_z} z + x_{zentr} \quad \text{und} \quad y = \frac{k_y}{k_z} z + y_{zentr}$$

berechnen, wobei m bzw. n die Beugungsordnungen des Gitters indizieren und der Punkt (x_{zentr}, y_{zentr}) den Auftreffpunkt des Hauptwellenvektors auf die xy-Ebene bei $z = 0$ beschreibt. Sie werden mit dem tatsächlichen Ausgangsfeld bei einer Struktur in der Zielgrößenordnung verglichen.

Konvergenz Feldbildsynthese zu Fouriermethode

Hier soll untersucht werden, inwieweit beide Methoden im Fernfeld bei besonders flachen Eingangswinkeln oder starker Eingangsbeugung gegeneinander konvergieren.

Das Aufsetzen der FDTD-Simulationen und die Einstellung der zugehörigen Parameter orientiert sich hierbei ebenfalls am Leitfaden in Appendix 9.

5.5 Vergleich und Verifikation mit den vorgestellten Methoden

Nach der Vorstellung der beiden Methoden zur Berechnung der Teilausleuchtung wird nun das Stufenprogramm zur Verifikation von Feldbildsynthese und Fouriermethode anhand von Beispielstrukturen angewandt. Begonnen wird mit dem Spiegelkriterium. Hierfür ist in Abbildung(15) ein Spiegelfläche der Kantenlänge $40\mu m \times 40\mu m = (L_x \times L_y)$ definiert.

In der Feldbildsynthese ist exemplarisch ein Gaußstrahl mit einem Taillenradius von $2\mu m$ und einem radialen Taillenabstand von $10\mu m$ zum Mittelpunkt der Fläche $z = 0$ als Eingangsbeleuchtung definiert. Alle drei Eingangswinkel des Strahls sind auf 45 Grad gesetzt. Um die Eingangsbeleuchtung mit und ohne Beugung zu untersuchen ist in der Fouriermethode dagegen eine Teilausleuchtung ohne Beugung definiert, somit eine Eingangswelle mit ebener Phasenfront, jedoch der Amplitudenverteilung einer Gaußfunktion. Diese weist ebenfalls einen Taillenradius von $2\mu m$ auf. Die in Abbildung(15) betrachtete xy-Fläche liegt bei z-Koordinaten von $z = -\sqrt{2} \cdot 5\mu m$ bzw. $z = -\sqrt{2} \cdot 10\mu m$, da so die Zentren der Strahlformen genau bei $(5\mu m, 5\mu m)$ bzw. $(10\mu m, 10\mu m)$ in der Ebene angeordnet sein müssen.

Die gezeigten Darstellungen erfüllen diese Anforderungen, insbesondere da die Positionen der Strahlen übereinstimmen und die Strahlformen bis auf eine geringe Abflachung erhalten bleiben, sodass die Formulierung der Propagatoren als richtig gesehen werden kann. In beiden Bildern stellt die Strahlform im linken oberen Bereich die Eingangsbeleuchtung dar. Im Bild der Feldbildsynthese ist die Aufweitung des Gaußstrahls bei der Vorwärtspropagation zu erkennen.

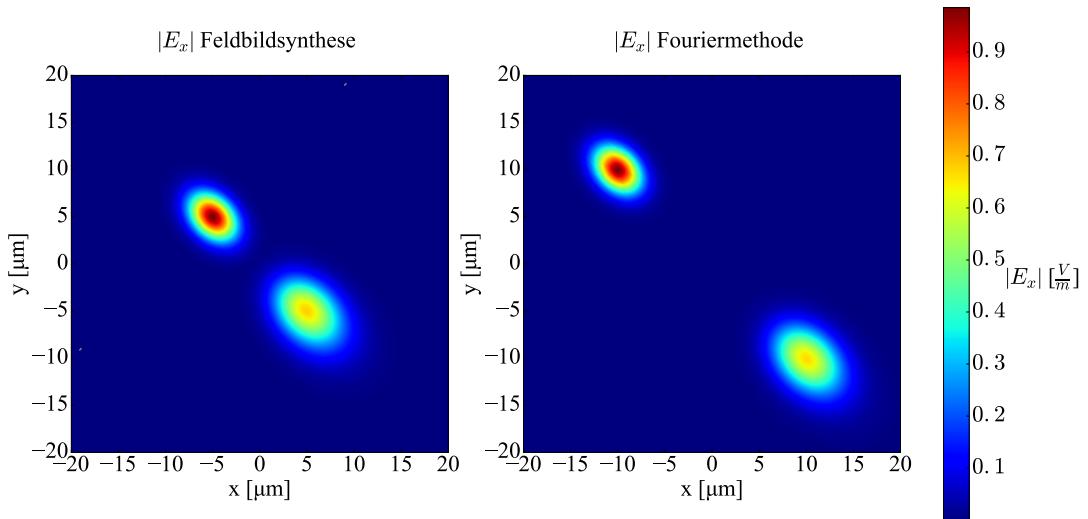


Abbildung 15: Spiegeltest zur Überprüfung der Implementierung: links Feldbildsynthese mit einfallendem Gaußstrahl, rechts Fouriermethode mit einfallender ebener Welle. Der Eingangsstrahl ist jeweils im linken oberen Quadranten angeordnet.

In der nächsten Stufe wird zuerst das Nah- und anschließend das Fernfeld anhand von zwei Beispielstrukturen in der FDTD verglichen. Begonnen wird mit einer Struktur eindimensionaler Periodizität. Die periodische Zelle sowie die gesamte Struktur sind in Abbildung(16) gezeigt.

Das Modellbeispiel ist wie folgt aufgebaut: In einem kartesischen Koordinatensystem ist in der xy-Ebene bei $z = 0$ eine Fläche der Kantenlängen $19.2\mu\text{m} \times 19.2\mu\text{m}$ angeordnet. Diese Fläche enthält 12×12 Zellen der Kantenlänge $1.6\mu\text{m} \times 1.6\mu\text{m}$; eine dieser Zellen ist auf der rechten oberen Seite von Abbildung(16) gezeigt. Der blaue Körper im Bild entspricht Polycarbonat mit einem wellenlängenunabhängigen Brechungsindex von 1.5. Die grauen Schichten repräsentieren Aluminium, wobei der komplexe Brechungsindex von Referenz [20] bezogen ist. Die Zellen wiederholen sich über die gesamte Struktur hinweg entlang der x-Achse mit der Periode von $1.6\mu\text{m}$. Die Geometrie einer periodischen Zelle entlang der x-Achse ist über die grauen Gitterlinien in Abbildung(16) unten links gezeigt, wobei die Schrittweite der Bezeichnungslinien in x und z-Richtung jeweils 100nm beträgt. Die Dicke der Polycarbonatschicht bis zur Unterkante der ersten Aluminiumschicht beträgt in diesem Modell $1\mu\text{m}$ (dies entspricht nicht der realen Dicke der Carbonatschicht der CD, stellt aber genügend Dicke für ihre Berücksichtigung in der Simulation dar). Entlang der y-Achse ist die Gesamtstruktur invariant (1D-Periodizität). Die jeweils äußerste Zellenlinie ist hierbei nicht mit Zellen, sondern mit Luft belegt, um genügend Abstand zwischen dem metallischen Material und den Randflächen der FDTD-Zelle zu garantieren. Über und unter der gezeigten Struktur befindet sich jeweils noch eine 100nm dicke Luftsicht; ferner sind die Halbräume über und unter der Struktur mit Luft gefüllt. Als Eingangsbeleuchtung dient hier ein Gaußstrahl (Appendix 11) mit einem Strahlradius von $4\mu\text{m}$ in der Strahlaille bei einer Wellenlänge von 400nm. Die Strahlaille befindet sich $5\mu\text{m}$ von der Quellenfläche entfernt ($z = -1\mu\text{m}$) und die Propagationsrichtung entspricht der z-Achse, es wird somit ein normaler Einfall (azimuthaler und

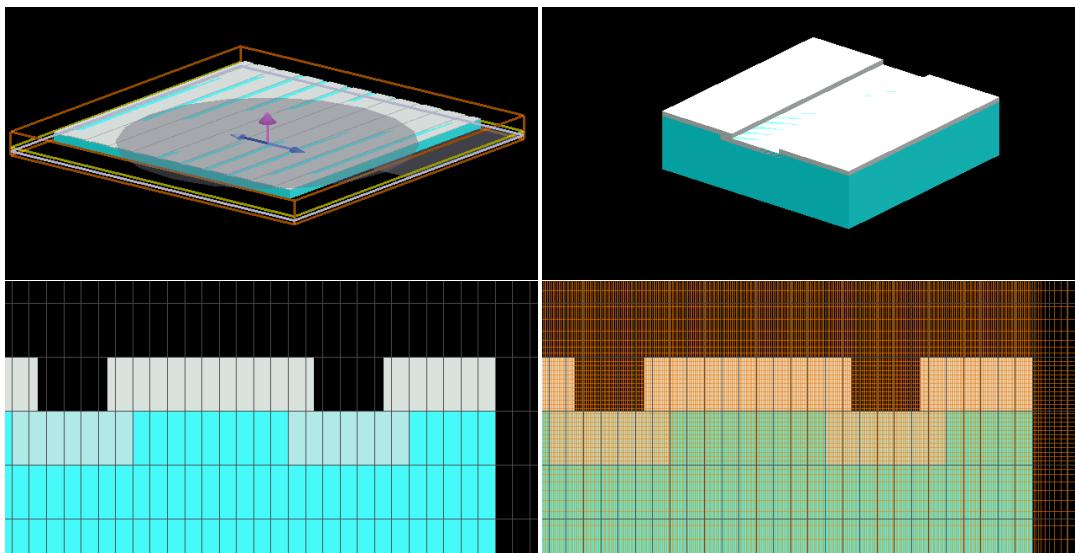


Abbildung 16: Oben links: Perspektivische Ansicht der gesamten 1D-Verifikationsstruktur in negativer z-Richtung - der pinke Pfeil repräsentiert die z-Achse und gleichzeitig die Einfallsrichtung des Gaußstrahls; die blauen Pfeile stellen die Linie der linearen Eingangspolarisation dar, oben rechts: Ansicht einer einzelnen periodischen Zelle, unten links: Ansicht der xz-Ebene für die exakten Abmessungen, unten rechts: durchgehend feine Meshauflösung der Metallschicht.

polarer Winkel der Hauptachse sind 0°). Die Quellenfläche ist bei $z = -1\mu\text{m}$ angeordnet. Die folgend gezeigten FDTD Simulationsergebnisse sind gemäß Appendix 9 vollzogen worden und können deshalb, wie in Kapitel 3 und im vorherigen Teilabschnitt beschrieben, als konvergent erachtet werden. Die Polarisation des elektrischen Feldvektors in der Strahltaille ist linear mit einem Polarisationswinkel ψ von 0° , um die Symmetrie der Struktur in der FDTD ausnutzen zu können. Der tatsächlich simulierte Bereich ist in der perspektivischen Ansicht der Gesamtstruktur in Abbildung(16) durch den angegraute Quadranten gezeigt. Problematisch bei dieser Struktur ist die durchgängige Metallfläche, die in der ‘Conformal Mesh Technology’ (CMT) von Lumerical sehr fein aufgelöst werden muss; siehe hierzu das Meshbild in Abbildung(16). Dies limitiert die maximale laterale Ausdehnung der Beispielstruktur auf die vorgestellten Maße. Die Berechnungszeit dieser Anordnung liegt unter Konvergenz bei 1.5 Stunden und der Berechnungsaufwand bei 7GB bei der genannten Rechenkonfiguration dieser Arbeit. Hiermit ist auch schon die Größenordnung und das Problem des direkten Vergleichs einer größeren Struktur ersichtlich.

Das Modellbeispiel wird nun im Teilausleuchtungspaket aufgesetzt. Hierbei kann wie in der FDTD die 90° Symmetrie der Struktur ausgenutzt werden. Vor dem eigentlichen Vergleich mit der FDTD ist für die beiden Methoden des Teilausleuchtungspakets die Konvergenz zu zeigen. Dies ist exemplarisch in Abbildung(17) für die Feldbildsynthese gezeigt. Sie zeigt eindeutig die gewünschte Konvergenz des Verfahrens. Das Ziel ist, dass sich ab einer gewissen Anzahl an Diskretisierungspunkten pro Zelle (DpZ) die Maximalwerte und Feldartefakte nicht mehr ändern. Diese Anzahl kann dementsprechend als Konvergenzschwelle festgelegt werden. Aus dieser Betrachtung folgend können 10 DpZ als konvergentes Ergebnis für den Vergleich festgesetzt werden. Exemplarisch sind hier auch einmal die Berücksichti-

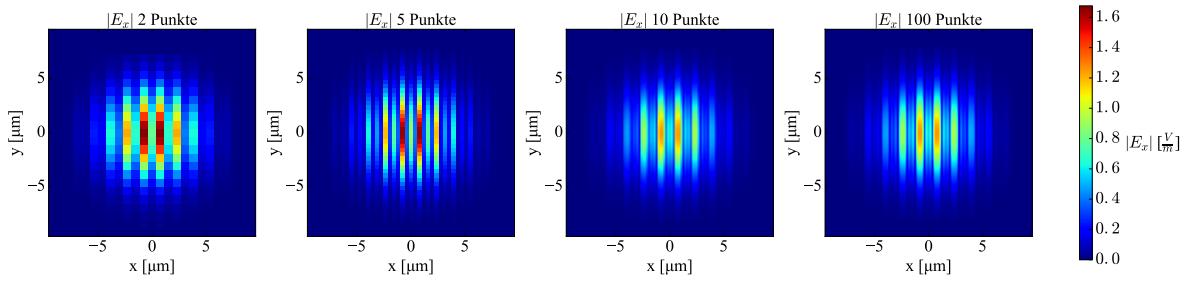


Abbildung 17: Der Absolutbetrag der x-Komponente des elektrischen Feldes in einer xy-Ebene bei $z = -0.5\mu\text{m}$, somit über der Struktur (Reflektion), bei vier verschiedenen Diskretisierungsstufen: 2 Diskretisierungspunkte pro Zelle (DpZ), 5 DpZ, 10 DpZ und 100 DpZ

gung aller 100 Bildpunkte aus der RCWA gezeigt, die das Erreichen des Konvergenzpunktes bei 10 DpZ nochmals verdeutlichen.

Für die Fouriermethode wird auf gleiche Weise für den folgenden Vergleich dieser Beispielstruktur das Verwenden der innersten 21 Beugungsordnungen bei 10 DpZ festgelegt. Analog werden sämtliche Konvergenzuntersuchungen der vorgestellten Ergebnisse von Feldbildsynthese und Fouriermethode in diesem Kapitel vor dem eigentlichen Vergleich durchgeführt, sodass alle folgend gezeigten Ergebnisse als konvergent betrachtet werden können. Aus Platzgründen ist die Konvergenzuntersuchung jedoch nur in diesem Fall exemplarisch dargestellt. In Abbildung(18) steht somit das konvergente Simulationsergebnis von FDTD, Feldbildsynthese und Fouriermethode zur Verfügung, um nun die eigentlichen Verifikationsaussagen treffen zu können.

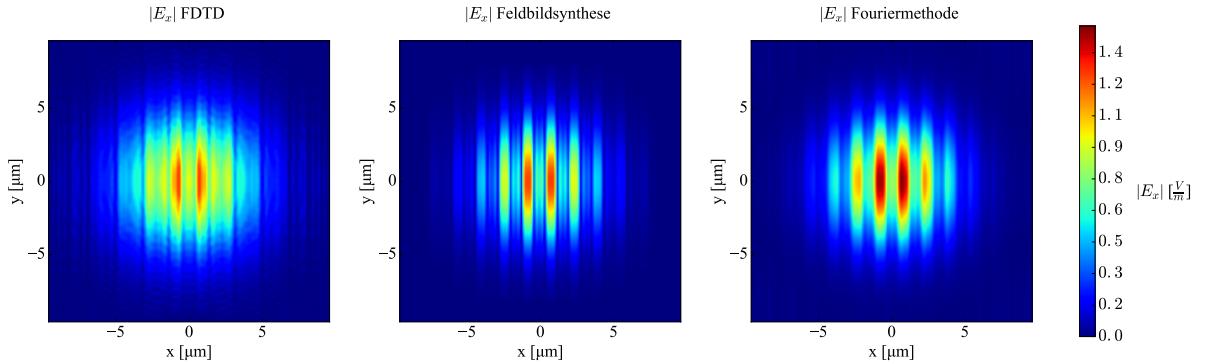


Abbildung 18: Gleiche Modalität wie in Abbildung(17): links Lumerical-FDTD, mittig Feldbildsynthese, rechts Fouriermethode

Im Vergleich zwischen der Feldbildsynthese und der FDTD zeigt sich, dass die Maximalwerte beider Verfahren übereinstimmen. Lediglich die Maximalwerte der Fouriermethode weichen um +10% ab, was auch durch Erhöhung der DpZ nicht zu beheben ist. Die Abweichung dieser Methode spielt hier jedoch eine untergeordnete Rolle, da der Fokus auf der Verifikation der eigenen Methode mit der FDTD liegt. Ferner sind auch die Formen der Feldartefakte nahezu gleich, bei der FDTD sind die Maximalwerte mehr auf die Metallkanten des Aluminiums konzentriert. Dieser Effekt wird ebenfalls im Vergleich der NVT im vorigen Kapitel

4 beobachtet und erläutert. Die Feldartefakte sind in der Feldbildsynthese wie auch in der Fouriermethode gleichartig kontinuierlich, was sich analog auf die Fourierreihenformulierung der RCWA rückzuführen lässt. Die Feldbildsynthese genügt somit dem definierten Anspruch an die Abweichung im Vergleich, die Fouriermethode dagegen nicht.

Nun wird noch ein zweites Beispiel zur Verifikation der Teilausleuchtung herangezogen. Auch dieses wird an weiteren Stellen in dieser Arbeit aufgegriffen und deshalb hier detailliert mit Hilfe von Abbildung(19) vorgestellt.

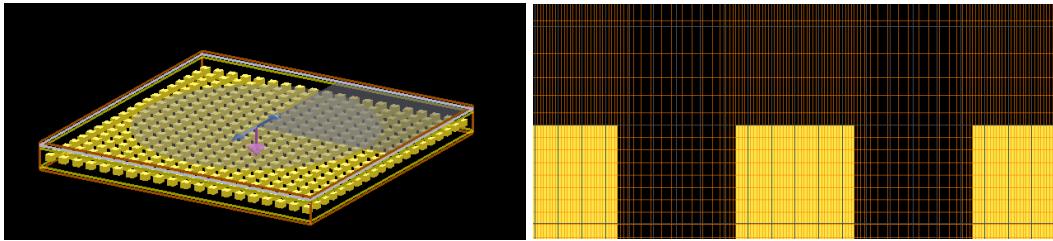


Abbildung 19: Links: Perspektivische Draufsicht auf die 2D-Verifikationsstruktur in Lumerical, Polarisation und Einfallsrichtung stimmen mit dem vorherigen Beispiel überein, ebenso wird hier die 90° Symmetrie der Struktur in der FDTD ausgenutzt, rechts: grobere Meshauflösung der Luftbereiche zwischen den Goldzellen, definiert durch Lumerical-CMT

In diesem Modell wird eine Struktur mit 2D-periodischen Zellen untersucht; somit kann auch hier direkt die Normalenvektortransformation aus Kapitel 4 angewandt werden. Eine periodische Zelle hat die Abmessungen $1\mu\text{m} \times 1\mu\text{m}$ und setzt sich sowohl in x- als auch in y-Richtung periodisch über die Struktur hinweg fort. Die Zellenoberkante liegt bei $z = 0\mu\text{m}$ und jede Zelle hat eine Dicke von $1.2\mu\text{m}$. Entlang der z-Achse sind zuerst eine 100nm dicke Luftsicht, dann ein $1\mu\text{m}$ dickes Goldquadrat mit einer xy-Kantenlänge von 500nm x 500nm mittig in Luft positioniert und anschließend nochmals eine 100nm dicke Luftsicht gesetzt. Die Gesamtstruktur hat eine Abmessung von $20\mu\text{m} \times 20\mu\text{m}$, in welcher die Zellen in beiden Achsen periodisch angeordnet sind. Als Eingangsverteilung dient derselbe Gaußstrahl wie im Beispiel zuvor, diesmal allerdings bei einer Wellenlänge von 660nm. Die Quellenfläche ist hierbei auf $z = -1.5\mu\text{m}$ gesetzt. Der gesamte FDTD-Simulationsbereich hat hierbei eine Größe von $20\mu\text{m} \times 20\mu\text{m} \times 4.5\mu\text{m}$. Die lateralen Abmessungen der Struktur können somit größer als im Beispiel zuvor definiert werden, da mit der CMT-Technologie von Lumerical die Luftbereiche zwischen den Goldelementen im kontinuierlichen Übergang grober diskretisiert werden. Sonstige FDTD-Einstellungen beispielsweise bezüglich der Symmetrie sind aus dem vorherigen Beispiel der CD-Zellenstruktur übernommen.

In der RCWA-Zellensimulation und der Teilausleuchtung kann hierbei die 45° Symmetrie der Struktur, wie sie im Teilabschnitt Symmetrie dieses Kapitels präsentiert ist, ausgenutzt werden. Die RCWA-Zellensimulation ist mit 21 Beugungsordnungen pro Achse bei einer Eingangswinkelschrittweite von 2° angesetzt. Die konvergenten Ergebnisse des elektrischen Feldes aller drei Methoden an einer reflektiven Position bei $z = -1\mu\text{m}$ sind in Abbildung(20) gezeigt.

Die Feldbildsynthese enthält hierbei ebenfalls 10 DpZ, die Fouriermethode 20 DpZ, wobei in dieser nur die innersten 11 von 21 Beugungsordnungen aus der RCWA berücksichtigt sind. Auch hier stimmen die Maximalwerte und Feldformen von FDTD und Feldbildsynthese überein. Bei der Feldbildsynthese ist das Feld zwischen den gaußförmigen Maxima im

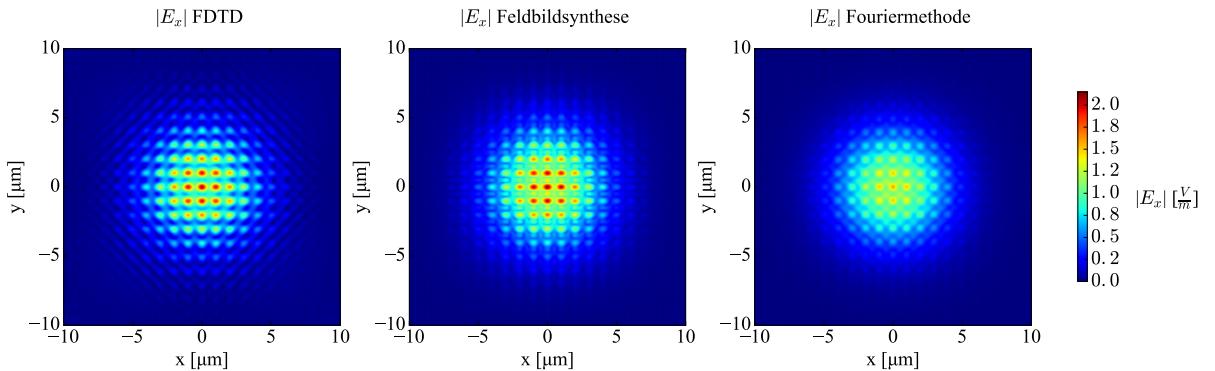


Abbildung 20: Vergleich des Absolutbetrages der x-Komponente des elektrischen Feldes in einer xy-Ebene bei $z = -1 \mu\text{m}$ über der Struktur (Reflektion): links FDTD, mittig Feldbildsynthese, rechts Fouriermethode

inneren des Zellenbereichs etwas höher und mit Nebenmaxima versehen, was sich wie in den vorherigen Vergleichen auf die Darstellung des Feldes durch Überlagerung ebener Wellen rückführen lässt. Die Fouriermethode zeigt zwar konzeptionell ähnliche Feldverläufe, weicht jedoch diesmal mit ihren Maximalwerten auf bis zu -25% zu FDTD und Feldbildsynthese ab. Somit ist auch hier der Anspruch an die gewünschte Fehlertoleranz im Vergleich verfehlt. Analog sind die konvergenten Ergebnisse an einer Position hinter der Goldstruktur in Abbildung(21) gezeigt.

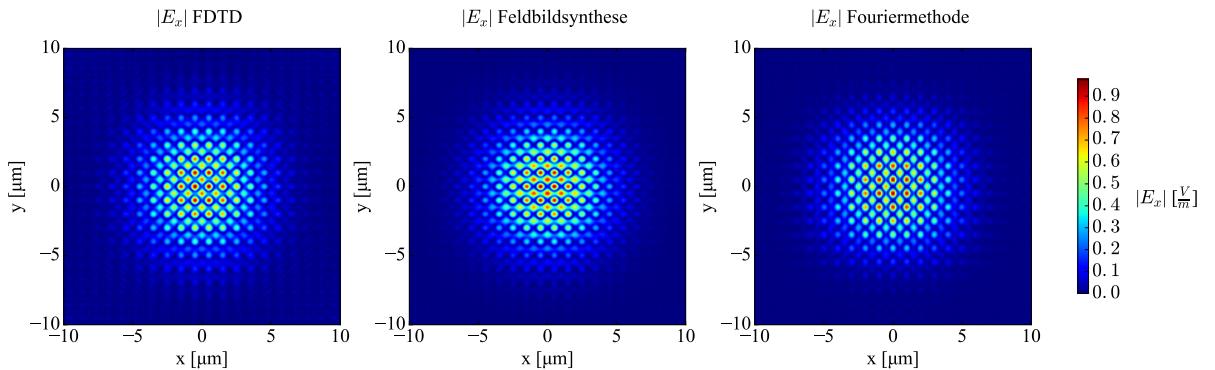


Abbildung 21: Vergleich des Absolutbetrages der x-Komponente des elektrischen Feldes in einer xy-Ebene bei $z = 2 \mu\text{m}$, somit 900nm unter der Struktur (Transmission): links FDTD, mittig Feldbildsynthese, rechts Fouriermethode

In dieser Vergleichsbetrachtung des elektrischen Feldes sind alle drei Methoden sowohl in Bezug auf die Maximalwerte als auch die Feldartefakte vergleichbar. Die an dieser Stelle wiederum vorhandene Konvergenz der Maximalwerte in der Fouriermethode liegt an der fehlenden Eingangsbeleuchtung im unteren Halbraum. In den bisherigen Betrachtungen im oberen reflektiven Halbraum ist zusätzlich zum reflektierten Ausgangsfeld auch die Eingangsbeleuchtung im Feld vorhanden. Die Fouriermethode hat somit Schwierigkeiten, das reflektierte Feld und den auf die Struktur einfallenden Gaußstrahl korrekt zu überlagern. Aufgrund der minimalen bis nicht vorhandenen Abweichung der konvergenten Ergebnisse von

Feldbildsynthese und FDTD im Nahfeld kann dieser Verifikationsschritt somit als erfüllt betrachtet werden.

Im nächsten Verifikationsschritt wird das Fernfeld der FDTD mit dem Ausgangsfeld der Feldbildsynthese in Abbildung(22) für die zweite Beispielstruktur verglichen.

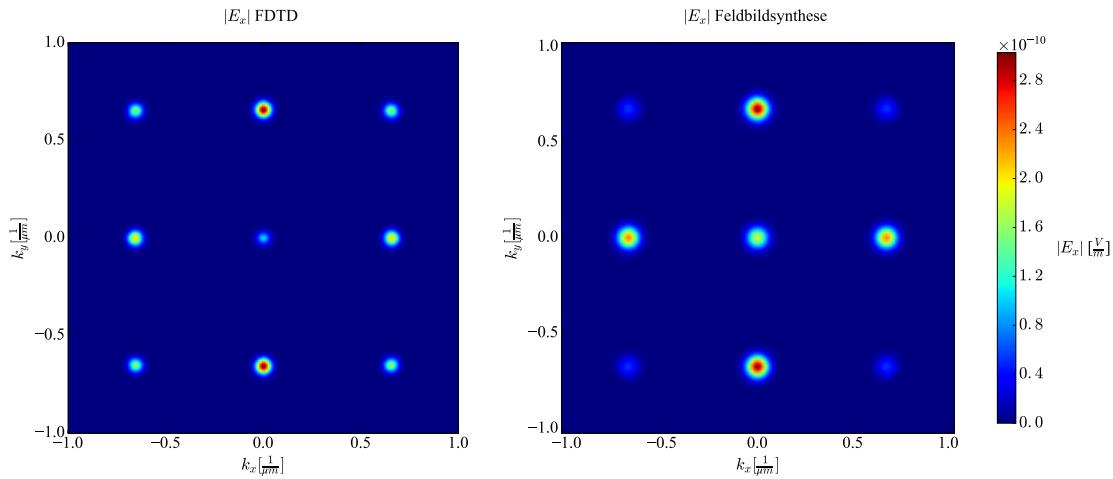


Abbildung 22: Vergleich des Absolutbetrages der x-Komponente des elektrischen Feldes in der Fernfeldnäherung in Transmission der 2D-Goldstruktur: links mit der FDTD, rechts mit der Feldbildsynthese

Die Fernfeldnäherung in Lumerical wird durch eine zweidimensionale Fouriertransformation realisiert, was dem theoretischen Grundsatz der Fourieroptik entspricht, dass der Übergang von Nah- zu Fernfeld von einer Fouriertransformation angenähert werden kann. Auf diese Weise wird auch das Fernfeld in der Feldbildsynthese angegeben. Die gezeigten Amplitudenbeträge sind hierbei mit dem Normierungsfaktor der Fouriertransformation gewichtet.

In Abbildung(22) sind die diskretisierten Beugungsordnungen in x und y-Richtung zu erkennen. Aufgrund des normalen Einfalls befindet sich die Beugungsordnung $(0,0)$ genau mittig. Ein positiver Index einer Beugungsordnung korrespondiert zu einer negativen Vektorkomponente, sodass beispielsweise die Ordnung $(1,1)$ links oben zu finden ist. Auch in diesem Vergleich stimmen die Maxima überein, es ist eine leichte Aufweitung der gaußförmigen Feldartefakte in der Feldbildsynthese gegenüber der FDTD zu erkennen. Auch die Grundüberlegung der Fourieroptik, dass in dieser Darstellung Gaußfunktionen auf den Beugungsordnungen positioniert zu sehen sein sollten, wird von beiden Methoden erfüllt. Dieser Zusammenhang wird im hierauf folgenden Verifikationsschritt nochmals genauer betrachtet. Aus Vergleichsgründen ist für die Feldbildsynthese die Anzahl der DpZ aus der vorherigen Abbildung(21) beibehalten.

Die Fernfeldbetrachtung der Fouriermethode wird an dieser Stelle nicht in den Vergleich einbezogen. Im Fourieraum dieser Methode existieren nur die einzelnen Beugungsordnungen der Methode, deren Orte analytisch über die Wellenvektoren festgelegt sind. Der aussagekräftigere Vergleich liegt hier jedoch eher bei den Methoden FDTD und Feldbildsynthese, in welcher die sichtbaren Ausgangsordnungen des Fernfeldes rein über die Felddarstellungen, die keiner Darstellung über die analytischen Beugungsordnungen entsprechen, erzeugt wer-

den. Aufgrund der vergleichbaren Ergebnisse kann auch dieser Verifikationsschritt kann als gültig abgeschlossen betrachtet werden.

Im vorletzten Schritt dieses Vergleichskapitels soll nun die Konvergenz der beiden Methoden der Teilausleuchtung einer erheblich größeren Struktur unter flachem Eingangswinkel untersucht werden. Der zugehörige Vergleich ist in Abbildung(23) gezeigt.

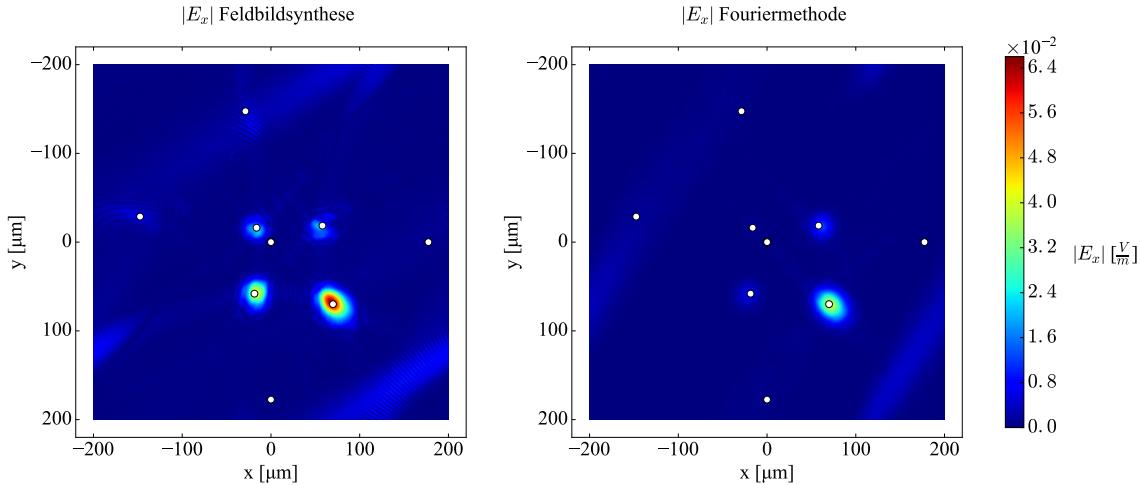


Abbildung 23: Der Absolutbetrag der x-Komponente des elektrischen Feldes bei $z = 100\mu\text{m}$ (Transmission) unter flachem Einfall (polarer Winkel von 45 Grad) auf die 2D-Goldstruktur. Die analytischen Ordnungspositionen sind durch die weißen Punkte dargestellt: links Feldbildsynthese, rechts mit Fouriermethode

Als analytische Überprüfungsmittel dienen an dieser Stelle die nach Gleichung(33) vorberechneten Ordnungspositionen. Die hier gezeigte Struktur entspricht den Goldzellen aus Abbildung(19), diesmal jedoch mit einer Kantenlänge L_x bzw. L_y von $400\mu\text{m}$. Als Feldeingang dient ein Gaußstrahl mit $4\mu\text{m}$ Taillenradius, wobei selbige $100\mu\text{m}$ radial von der Struktur entfernt liegt. Alle drei Eingangswinkel liegen bei 45° , wobei der Gaußstrahl genau mittig einfällt. Für die gezeigten konvergenten Ergebnisse werden in der Feldbildsynthese 20 DpZ, in der Fouriermethode 15 DpZ bei Berücksichtigung der innersten 5 Ordnung pro Achse von den jeweils insgesamt 21 verfügbaren Ordnungen aus der RCWA gesetzt.

In diesem Vergleich ist die Methodik der Fouriertransformationen gut zu erkennen. Für sie gilt die Annahme, dass sich der transformierte Ausschnitt einer Funktion oder eines Bildes unendlich oft periodisch fortsetzt. Dies ist in beiden Feldbildern ersichtlich; die leichten äußeren Feldartefakte werden auf einer Seite abgeschnitten und auf der anderen Seite aufgrund dieser periodischen Formulierung fortgesetzt. Diese Charakteristik ist jedoch grundsätzlich inkorrekt, hierauf wird jedoch in Abbildung(24) noch einmal Bezug genommen. An dieser Stelle ist eine Abweichung der Fouriermethode von der Feldbildsynthese evident. Die Orte der ausgehenden Ordnungen stimmen zwar überein, jedoch treten gerade die inneren Ordnungen hier abgeschwächt auf. Ferner ist eine um $\sim 50\%$ niedrigere Maximalintensität zu erkennen. Dieses Problem wird auch nicht durch die Berücksichtigung nur der inneren Ordnungen verursacht. Sowohl die Erhöhung der DpZ als auch die Erhöhung der verwendeten Ordnungen beheben diese Abweichungen nicht. Jedoch ist auch in einigen der oberen Vergleiche öfters die Abweichung der Fouriermethode sowohl von der FDTD als auch von der

Feldbildsynthese evident. Der Hauptfokus dieser Arbeit liegt allerdings auf der Verifikation der eigenen Feldbildsynthese, weshalb diese Problematik das Erreichen des Ziels dieser Arbeit nicht beeinträchtigt. Gleichzeitig stellt die gezeigte Anordnung die maximal simulierbare laterale Ausdehnung der Struktur von $400 \mu\text{m}$ unter Konvergenz für die Fouriermethode bei der gegebenen Rechnerkonfiguration dar. Größere Ausdehnungen führen zu einer Divergenz des Ergebnisses.

Nun kann die Feldbildsynthese auf ihre maximale Ausdehnung hin untersucht werden. Das Ergebnis dieser Maximumsfindung ist in Abbildung(24) gezeigt.

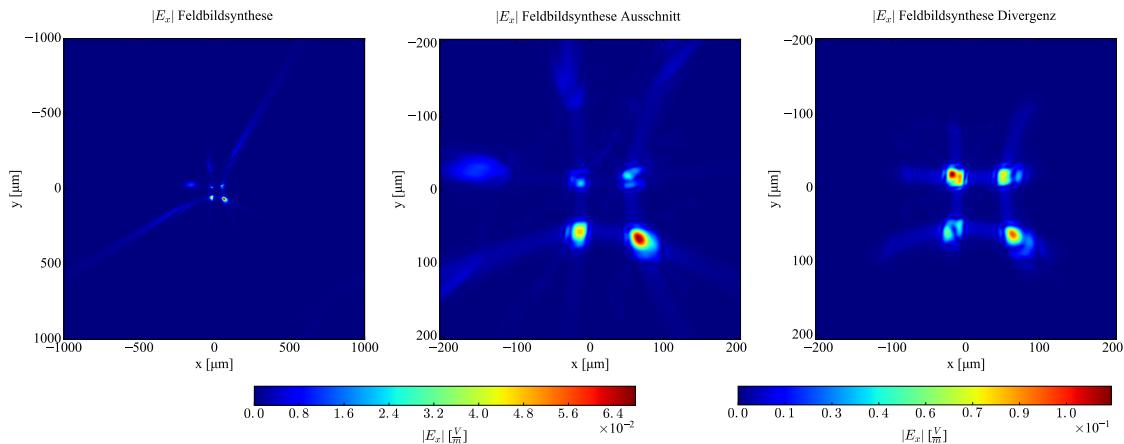


Abbildung 24: Der Absolutbetrag der x-Komponente des elektrischen Feldes: links die maximal mögliche Ausdehnung von 2mm, mittig der Ausschnitt der maximalen Ausdehnung zum Vergleich mit Abbildung(23), rechts das divergente Ergebnis bei 3mm Kantenlänge L_x im gleichen Ausschnitt.

Der Abstand der betrachteten Fläche von der Struktur sowie die Eingangsbeleuchtung entsprechen den Angaben zu Abbildung(23). Lediglich die laterale Ausdehnung L_x bzw. L_y wird schrittweise erhöht. Das konvergente gesuchte Ergebnis ist ja nun bereits aus Abbildung(23) bekannt und muss weiterhin erhalten bleiben. Somit ist durch Minimierung der verwendeten DpZ die maximale Ausdehnung zu finden.

Die maximale Kantenlänge der Struktur beträgt 2mm bei einer Verwendung von 5 DpZ und stellt somit das absolute Maximum an lateraler Ausdehnung dieser Methode bei der gegebenen Rechnerkonfiguration dar. Ab diesem Maximum beginnt bei Reduktion der DpZ eine Verschiebung der Energie zwischen den Ordnungen als erstes Anzeichen von Divergenz. Diese ist für eine 3mm Struktur bei 3 DpZ gut zu erkennen. Dennoch zeigt diese Analyse den maßgeblichen Vorteil der Feldbildsynthese dahingehend, dass die gezeigte maximale Ausdehnung nahezu eine Potenz über der maximal möglichen Ausdehnung der Fouriermethode aus Abbildung(23) liegt. Nur mit dieser Methode ist die Konvergenz bei 2mm Kantenlänge unter der Verwendung von 5 DpZ, besonders in Bezug auf einen Feldeingang unter Beugung, zur Struktur relativ kleine Ausleuchtung und flachem Eingangswinkel garantiert. Dennoch ist dies eine numerisch aufwendige Struktur aufgrund des scharfkantigen Übergangs von Luft zu Metall, somit müsste beispielsweise bei der Simulation von nicht metallischen Metalinsen eine größere laterale Abmessung unter Konvergenz möglich sein.

An dieser Stelle sei nochmals Rückbezug zu den periodischen Artefakten aus Abbildung(23) gegeben. Diese treten in Abbildung(24) nicht auf. Stattdessen sind langgezogene Artefakte ersichtlich. Diese werden in Abbildung(23) aufgrund ihres sehr flachen Ausgangswinkels bei einer zu kleinen lateralen Abmessung der Fourierfläche fälschlicherweise am Rand abgeschnitten und auf der gegenüberliegenden Seite fortgesetzt.

5.6 Fazit

In diesem Kapitel ist das neue Verfahren der Feldbildsynthese zur Simulation der Teilausleuchtung vorgestellt worden. Das zugehörige Stufenprogramm zur Verifikation kann als abgeschlossen und den Ansprüchen an Konvergenz und Abweichung genügend betrachtet werden. Die Feldbildsynthese erfüllt die einzelnen Stufen vollständig, insbesondere in den im Vorhinein als kritisch betrachteten Punkten des Fernfeldes und der Eingangsbeleuchtung mit Beugung unter flachem Eingangswinkel. Mit der gegebenen Abschätzung zur minimalen Anzahl an Diskretisierungspunkten kann diese neue Methode somit zuverlässig für die gesuchte wellenoptische Simulation der gewünschten Strukturen eingesetzt werden. Es ist eine maximal mögliche laterale Abmessung von 2mm bei einer numerisch aufwändigen Metallstruktur ausgemacht worden. Diese Abmessung liegt im Bereich der Zielsetzung dieser Arbeit und eine Größenordnung über dem zur Verifikation implementierten Verfahren der Fouriermethode nach [30]. Bei der Fouriermethode ist ferner eine fehlerbehaftete Formulierung in [30] ausgemacht und behoben worden, sodass insbesondere die Positionen der Feldartefakte korrekt dargestellt werden. Die Rechenzeit der Feldbildsynthese motiviert auch für verschiedenste Anwendungen Simulationen über Parameterbereiche der Eingangswinkel oder Wellenlängen. Ferner ermöglicht diese Methode bezüglich der Motivation dieser Arbeit erstmals, optische Bauelemente wie beispielsweise Metalinsen in ihrer realen Baugröße wellenoptisch zu simulieren. Die Feldbildsynthese bietet so die Möglichkeit, diffraktive Optiken mit inhomogener Eingangsbeleuchtung und beliebiger optischer Abbildung anstatt einer reinen Fokussierung auf einen Brennpunkt simulativ auszulegen und zu verifizieren.

6 Raytracing

Im folgenden Kapitel wird nun das letzte der drei Hauptkonzepte zum Erreichen des Auslegungsziels dieser Arbeit vorgestellt. In Kapitel 3 wurde schon das grundlegende Verfahren des Raytracings sowie seine bisherigen Anwendungen erläutert. Wie eingangs erwähnt, werden vor Ausführung des Visualisierungsprozesses Simulationen in RCWA und Teilausleuchtung über Parameterräume wie Eingangswinkel und Wellenlängen durchgeführt und in Dateien abgespeichert. Diese werden dann im Raytracingalgorithmus lediglich eingelesen, sodass zur Laufzeit nur auf diese Simulationswerte zugegriffen werden muss. Dieser Ansatz unterscheidet sich grundlegend von den bisherigen Arbeiten zu diesem Thema aus Kapitel 2. In diesen Arbeiten wird unter anderem eher versucht wird, die wellenoptische Effekte direkt zur Laufzeit des Raytracers zu berechnen. Für den eigenen Ansatz müssen jedoch geeignete Schnittstellen und Samplingfunktionen entwickelt werden. Nach der Vorstellung der Grundkonzepte werden diese hergeleitet und die gesamte Implementierung abschließend anhand eines Beispielaufbaus vorgestellt. Anschließend wird die Entwicklung der Implementierung für pbtr-v3 vorgestellt. Der maßgeblich ersuchte Vorteil an dieser Stelle ist, alle Möglichkeiten von pbtr-v3 in Bezug auf Rechengeschwindigkeit und Detailgrad der Szenerie mit nutzen zu können. Auch diese Erweiterung wird exemplarisch mit einer realen Szenerie verglichen.

6.1 Grundverfahren der ebenen Welle

Im folgenden Teilabschnitt wird die in C++ implementierte Testsoftware vorgestellt. Sie orientiert sich an der Implementierung von smallpt in [31]. In einer Hauptfunktion, welche für jeden Pixel der Kamera aufgerufen wird, ist ein Kamerapunkt mit Sichtrichtung im Raum definiert. Für jedes Sample wird vom Kamerapunkt aus ein Ray der zugehörigen Klasse erzeugt. Diese Klasse beinhaltet zwei Vektoren, einen Basispunkt und einen Richtungsvektor, sodass ein Ray geometrisch eine Gerade beschreibt. Die Richtung des ausgehenden Kamerarays wird gemäß einer Gleichverteilung erstellt. Hierbei muss eine Dreiecksfilterung der Gleichverteilung für die Richtungsvektoren durchgeführt werden, da, wie in [29] formuliert, eine Einheitskugel auf eine Einheitskreisfläche abgebildet wird. Ab diesem Punkt unterscheidet sich die Implementierung von den meisten Raytracingalgorithmen, da mit Berücksichtigung des Einbaus der RCWA hier spektrale Werte der Szenerie statt RGB-Farbwerthen verwendet werden. Hierbei wird so vorgegangen, dass für jedes Sample ein Spektrum erstellt wird. Dieses enthält beispielsweise 31 spektrale Samples äquidistant im sichtbaren Wellenlängenbereich von 380nm bis 780nm verteilt. Für jedes spektrale Sample bei zugehöriger Wellenlänge wird die Hauptfunktion “radiance” dieser Implementierung aufgerufen; die genauen Modalitäten dieser Funktion werden im folgenden Absatz vorgestellt. Grundsätzlich wird ihr an dieser Stelle die Wellenlänge des aktuellen spektralen Samples sowie der aktuelle Kameraray übergeben. Nachdem diese Funktion für jedes spektrale Sample ausgeführt worden ist, muss das gesamte Spektrum mit den genormten Werten des CIE-Farbsystems in einen XYZ bzw. RGB-Farbwert konvertiert und gegebenenfalls bezüglich eines Schwellenwertes abgeschnitten werden (“clamping”). Da somit ein Wellenlängenspektrum der Lichtintensität einer Pixelfarbe zugeordnet ist, muss hierbei gelten, dass jedes spektrale Sample eines Pixelsamples exakt mit demselben Kameraray (gleiche Richtung und gleicher Ursprung) startet. Diese Farbkonvertierung basiert auf der allgemeinen Farbtheorie, wie sie beispielsweise in [28] tiefgreifender erläutert wird.

Nun werde die genannte “radiance” Funktion näher betrachtet. Auch sie orientiert sich an [31]. In ihr wird zu Beginn der erste Schnittpunkt des Kamerastrahls mit den geometri-

schen Objekten der Szenerie berechnet. Die geometrischen Objekte sind über eine Oberklasse verknüpft als eigene Klassen definiert. Am Beispiel einer Oberfläche gezeigt sind für jede Instanz geometrische Parameter wie der Ursprung und Kantenlängen sowie der Materialtyp gesetzt. Für jede Geometrieklasse sind ferner die eigenen Methoden “intersection” zur Schnittpunktfindung mit einer Gerade, hier dem Ray, sowie “get-normal” zur Rückgabe des Normalenvektors der Oberfläche am aktuellen Schnittpunkt. In der radiance Funktion wird nun eine Liste aller gesetzten Geometrien iteriert und der Schnittpunkt des aktuellen Rays zu jeder Geometrie berechnet. Nach diesem Durchlauf, falls mit mindestens einer Geometrie ein Schnittpunkt gefunden wurde, gilt das Objekt mit der kleinsten Distanz zum Schnittpunkt als das aktuelle Objekt. Bei der Schnittpunktberechnung zu jedem Objekt muss zu Beginn eine Transformation des Rays ins lokale Koordinatensystem dieses Objektes vollzogen werden. Jedes geometrische Objekt wird zuerst im kartesischen Ursprung definiert und anschließend mit Hilfe von Translation und Rotation an seinen Zielpunkt bewegt. Diese Transformation wird dann für die Schnittpunktfindung abgespeichert.

Im nächsten Schritt wird ein lokales Koordinatensystem des aktuellen Schnittpunktes bezüglich des Normalenvektors und der zugehörigen Oberfläche erstellt. In der testweisen Implementierung dieses Teilkapitels sind lediglich die Materialklassen Diffraktiv (matte Oberflächen), Specular (ideale Verspiegelung), Refractive (Glaskörper) und RCWA verfügbar, wobei sich die Implementierung der ersten drei Materialklassen mitsamt ihrer Samplingfunktion an smallpt [31] orientiert. Grundsätzlich wird am Schnittpunkt ein Faktor ermittelt, welcher mit dem vorherigen, zwischengespeicherten Faktor für die aktuell verwendete Wellenlänge multipliziert wird.

Im Folgenden wird nun allerdings der Fokus auf die neue Methodik am Schnittpunkt mit einem RCWA Material gelegt. Zu Beginn wird am aktuellen Schnittpunkt ein Koordinatensystem in Korrelation zur Oberfläche, dem Normalenvektor und der Orientierung der Struktur an dieser Stelle definiert. Beispielsweise muss beim Modellbeispiel der CD die rotationssymmetrische Anordnung der periodischen Zellen mit berücksichtigt werden. Dieses Koordinatensystem entspricht dann exakt dem der RCWA aus Abbildung(32). Das Ziel ist nun, den Gewichtungsfaktor und die Ausgangsrichtung des weiterzuverfolgenden Strahls aus den zuvor berechneten RCWA-Simulationswerten heraus zu ermitteln. An dieser Stelle besteht jedoch das Hauptproblem dieser Zusammenführung.

Aufgrund der Samplingeffizienz werden, wie im Teilabschnitt Raytracing des Kapitels 3 schon erläutert, im Raytracing die Raypfade in umgekehrter physikalischer Richtung, somit beginnend am Kamerapunkt und endend in einer Lichtquelle, verfolgt. Die RCWA hingegen stellt das Lösungssystem ihrer Differentialgleichungen um die einfallende ebene Welle herum auf; somit sind die Auswahl der resultierenden Beugungseffizienzen (siehe Appendix 10) und der Ausgangsrichtungen komplett abhängig von den Eingangswinkeln der einfallenden Welle. Dies entspricht somit der umgekehrten Richtung der Strahlberechnung im Raytracer.

Abbildung(25) zeigt einen ersten Ansatz zur Lösung dieses Problems.

Der rote Vektor entspricht dem ankommenden Ray im Raytracer und gleichzeitig einer ausgehenden Beugungsordnung der RCWA im lokalen Koordinatensystem des aktuellen Objektes, welches durch Transformationen in das der RCWA überführt worden ist. In diesem Koordinatensystem werden azimuthaler ϕ_a und polarer θ_a Winkel des roten Rays berechnet. Ferner wurde dem Ray in der Erzeugung im Kamerapunkt während der Iteration des Pixelspektrums eine Wellenlänge zugewiesen. Nun werde für diesen Schnittpunkt nach einer gleichverteilten Zufallsfunktion aus der Menge der in der RCWA verwendeten Beugungsordnungen eine ausgewählt und dem Strahl zugewiesen. Mit den beiden Ausgangswinkeln sowie der gegebenen

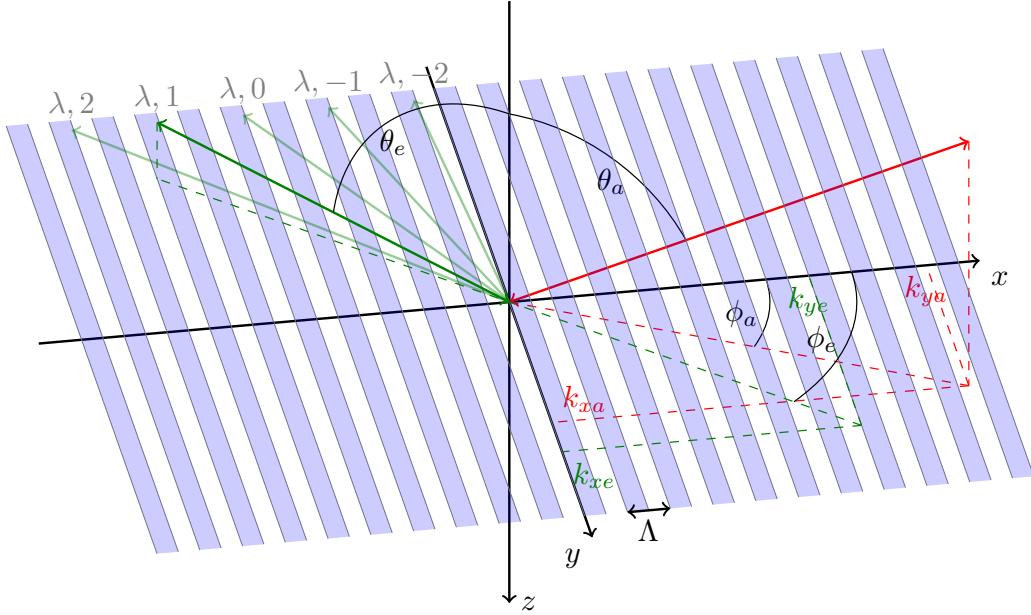


Abbildung 25: Konzept zur Behebung des Richtungsproblems: hier ist zur Übersichtlichkeit eine rein reflektierende Struktur gezeigt

Wellenlänge λ und dieser Beugungsordnung n_{res} ist nun eine Raykombination von Ausgangs-(rot) zu Eingangsray (grün) dank der Stetigkeit an elektromagnetischen Grenzflächen gegeben. Hierzu werden die Wellenvektoren der RCWA im Reflektionsfall formuliert:

$$k_{xa} = \cos(\phi_a) \sin(\theta_a) \frac{2\pi}{\lambda} n_1 \quad \text{und} \quad k_{ya} = \sin(\phi_a) \sin(\theta_a) \frac{2\pi}{\lambda} n_1 \quad (34)$$

$$k_{xe} = k_{xa} + n_{res} \frac{2\pi}{\Lambda} \quad \text{und} \quad k_{ye} = k_{ya}$$

Da hier nur eine eindimensionale Periodizität in x-Richtung betrachtet wird, ist die y-Komponente des Wellenvektors invariant. Auf diese Weise wurde nun die zugehörige Eingangswelle im RCWA-System definiert. Die Floquettheorie nach [4] garantiert hierbei, dass die zugehörige Lösung über die Periodizität eindeutig ist. Abbildung(25) zeigt nochmals dieses Prinzip. Die grünen Pfeile sind die verschiedenen Eingänge der RCWA bei der gegebenen Wellenlänge, wenn der aktuelle Ray in der RCWA als ausgehende $n_{res} = -2, -1, \dots$ Beugungsordnung betrachtet wird. Über die Formulierung der Wellenvektorkomponenten lassen sich nun die zugehörigen Eingangswinkel berechnen, wobei die Projektion der Richtungen und Eingangswinkel, wie in Abbildung(25) gezeigt, zu beachten sind:

$$\phi_e = \tan^{-1} \left(\frac{k_{ye}}{k_{xe}} \right) \quad \text{und} \quad (35)$$

$$\theta_e = \begin{cases} \sin^{-1}\left(\frac{k_{ye}}{k_0 n_1 \sin(\phi_e)}\right), & \text{falls } |k_{ye}| > 0 \\ \sin^{-1}\left(\frac{k_{xe}}{k_0 n_1 \cos(\phi_e)}\right), & \text{falls } k_{ye} = 0 \text{ und } |k_{xe}| > 0 \\ \pi & \text{sonst} \end{cases}$$

Der Eingangswinkel $\theta_e = \pi$ beschreibt hierbei einen nicht ausbreitungsfähigen Weg ohne Propagationsterm von Eingangs- zu Ausgangsray. Nun sind die Eingangswinkel der einfallenden Welle, somit also des weiterzuverfolgenden Rays, wofür direkt der ausgehende Richtungsvektor des Rays berechnet wird, sowie der zugehörige Wert aus der RCWA Simulation bekannt. Die Winkel sind hierbei aus einem stetigen Zahlenbereich berechnet, jedoch sind aus der RCWA-Simulation nur eine endliche Menge mit diskretisierten Eingangswinkeln berechnet worden. Nach Ausnutzung der Symmetrie nach dem Prinzip aus Kapitel 5 wird anschließend eine bilineare Interpolation nach [39] für den gegebenen Eingangswinkel durchgeführt, da nur diskrete Eingangswinkel und die zugehörigen Simulationen vorhanden sind.

$$DE_{res} = N \frac{1}{(\phi_{eo} - \phi_{eu})(\theta_{eo} - \theta_{eu})} [\theta_{eo} - \theta_e \quad \theta_e - \theta_{eu}] * \begin{bmatrix} DE_{\theta_{eu}, \phi_{eu}, n_s} & DE_{\theta_{eu}, \phi_{eo}, n_s} \\ DE_{\theta_{eo}, \phi_{eu}, n_s} & DE_{\theta_{eo}, \phi_{eo}, n_s} \end{bmatrix} [\phi_{eo} - \phi_e \quad \phi_e - \phi_{eu}] \quad (36)$$

Hierbei beschreiben die Winkel θ_{eu} bzw. θ_{eo} den in der Simulation vorhandenen nächsten unteren bzw. oberen Eingangswinkel sowie Terme der Form $DE_{\theta_{eu}, \phi_{eu}, n_s}$ die zugehörigen Beugungseffizienzen. Der Faktor N entspricht der Normierung dieser Zufallsentscheidung nach dem Monte Carlo Prinzip. Im einfachsten hier gezeigten Fall, in welchem die Ordnungen gemäß einer Gleichverteilung ausgewählt werden, gilt somit $N = N_{ges}$, wobei N_{ges} der Gesamtzahl der in der zufälligen Ordnungswahl berücksichtigten Ordnungen entspricht. Aus Effizienzgründen können hierbei auch beispielsweise nur die inneren 11 Ordnungen um die nullte Ordnung angeordnet berücksichtigt werden, obwohl in der RCWA-Simulation mit 61 Ordnungen simuliert worden ist. Theoretisch könnte an dieser Stelle weiterhin statt der bilinearen Interpolation auch eine bicubische durchgeführt werden; Vergleichstests ergeben hierfür jedoch keinen Mehrgewinn.

Im letzten Schritt wird in der "radiance"-Funktion noch eine zusätzliche "Russian Roulette" Abschätzung für die RCWA Materialien formuliert, die ab einer im Steuerskript angebbaren minimalen Anzahl an bisherigen Schnittpunkten (bspw. 5 Schnitte) aktiviert wird. Sie orientiert sich an [25], nur dass hier die nötigen Faktoren aus der spektralen Betrachtung synthetisiert werden müssen. Das Ziel der Russian-Roulette Abschätzung ist, Pfade mit geringerer Intensität früher als Pfade mit höherer Intensität abzubrechen und so die Effizienz der Strahlverfolgung zu steigern. Sie wird hier für die spektrale Implementierung gemäß

$$f = \begin{cases} \frac{f - q * e}{1 - q}, & \text{falls } \omega > q \\ e & \text{sonst} \end{cases} \quad (37)$$

formuliert. Hierbei entspricht e dem Lichtemissionsfaktor des aktuellen Objekts, ω einer in dieser Funktion erzeugten gleichverteilten Zufallszahl zwischen 0 und 1 sowie q einer vorher definierten Abbruchwahrscheinlichkeitsschwelle. f sei hier der zurückgegebene, über die Pfadschnittpunkte hinweg iterativ multiplizierte Spektralwert von DE_{res} .

6.2 Vergleich der Testimplementierung und der Cornellbox

Um die vorgestellte Implementierung zu verifizieren und zu beurteilen, ob das zugehörige Konzept zur Integration der RCWA in das Raytracingverfahren funktioniert, bedarf es einer Vergleichsszenerie. An der Cornell-University entstand die Idee, zur Verifikation der eigenen bildgebenden Verfahren eine Box zu konstruieren, in diese hinein zu fotografieren und die Fotografie mit dem im Raytracing errechneten Bild zu vergleichen. Diese Box sollte möglichst einfach im Raytracer zu realisieren sein, um Fehlerquellen besser eingrenzen zu können. Beispielsweise werde eine neue Samplingfunktion für ein im Raytracer bisher nicht definiertes Material entwickelt und sei nun zu testen. Ein Objekt dieses Materials werde so in der Box positioniert ist und die zugehörige Fotografie aufgenommen. Falls dann beispielsweise die Wände im errechneten Bild die gleiche Farbe und Helligkeit wie in der Fotografie aufweisen, jedoch im Bereich des neuen Objekts andere Farbartefakte als im errechneten Bild zu sehen sind, kann der Fehler eindeutig auf die neu entworfene Samplingfunktion eingegrenzt werden. Die Box besteht im einfachsten Fall aus drei Wänden sowie Deckel und Boden, wobei diese aus einem Werkstoff mit möglichst glatter, aber dennoch diffuser Oberfläche gefertigt sein sollten. So können die Wände im Raytracer als homogene ebene Oberflächen mit nur einem einzigen Farb- bzw. Reflektionswert definiert werden. In diesem Aufbau sind die Platten aus geschäumtem Simopor gefertigt und in mehreren Lackiergängen mit Sprühfarbe benetzt worden. Die Decke enthält ein kreisrundes Loch mit einer möglichst Lambertschen Lichtquelle, der am einfachsten in der Implementierung zu definierenden Strahlquelle, da sie mit der gleichen Leistungsdichte in alle Raumrichtungen abstrahlt. Hierfür ist ein LED-Array direkt auf das Loch aufgebracht und für konstante Helligkeit mit einer stromstabilen Versorgungsschaltung versehen. In der Box werden nun die zu untersuchenden Objekte positioniert. Hier sollte ein Objekt mit diffraktiver Strukturierung vorhanden sein, um die RCWA zu testen. Aus den in Kapitel 2 und 3 genannten Gründen wird hier für den ersten Vergleich auf eine CD zurückgegriffen, zu deren Strukturierung genügend Information auffindbar ist, um sie in der RCWA nachzubilden. Um auch verschiedene Farbschemen nach mehrmaligen Reflexionen sowie verschiedene Lichteinfälle zu berücksichtigen, wurden zusätzlich noch eine Glaskugel und ein -kubus in der Szenerie platziert. Zum Nachbau der gesamten Szenerie im Raytracer müssen alle Positionierungen der Kamera und dieser drei Objekte sowie die Abmessungen der Box und Quelle selbst exakt vermessen werden, da schon kleine Verschiebungen direkt andere Farbschemen und Reflexionen erzeugen.

In Abbildung(26) ist ein Vergleich der mit diesem Verfahren visualisierten Szenerie und der Fotografie der Cornellbox gezeigt.

Mit diesem Vergleich soll abgeschätzt werden, ob die vorgestellte Idee, die Rays im Bereich der RCWA direkt als ebene homogene Wellen zu betrachten und die im vorherigen Teilkapitel präsentierte Samplingfunktion zu entwerfen, praktikabel ist. Aus Abbildung(26) geht hervor, dass diese Zielsetzung erfüllt wird. Betrachtet man alle vorhandenen Farbschemen im Bild, auch diejenigen, die erst durch Mehrfachreflektion erzeugt werden, ist ersichtlich, dass in beiden Bildern sowohl die Reihenfolge der Farben als auch deren Form und die Gradienten der Farben zueinander übereinstimmen. Selbst Streuungen wie im linken oberen Eck sind in der exakt gleichen Form und Farbe existent. Das Hauptproblem bei diesem Visualisierungsprozess ist die Verdunkelung des Bildes. Es wird angenommen, dass hierbei ein Fehler in der Normierung des Bildes sowie der RCWA-Werte auftritt. Aus diesem Grund fehlt im berechneten Bild eine Größenordnung an Grundhelligkeit; gleichzeitig sind die Farbschemen der RCWA um eine Größenordnung heller als die Umgebung. Dieser Fehler konnte bisher noch nicht ausgemacht und behoben werden. Dennoch zeigt dieser Raytracingversuch, dass



Abbildung 26: Vergleich des ersten Raytracingversuchs: links die Fotografie, rechts das errechnete Bild im eigenen Raytracer

die vorgestellte Grundidee zur Verknüpfung von RCWA und Raytracing funktioniert. Problematisch ist bei dieser Testimplementierung die hohe Rechenzeit sowie die beschränkten Möglichkeiten in Bezug auf die Detaillierung der Szenerie.

6.3 Einbau des ebenen Wellenkonzeptes in pbrt-v3

Wie schon zuvor erläutert, dient die Implementierung im vorhergehenden Teilkapitel lediglich als Machbarkeitsstudie zum Einbau des ebenen Wellenkonzeptes in das Raytracing. Ein Problem dieser Implementierung ist die Rechengeschwindigkeit und Rechenaufwand des Raytracingalgorithmus. Das errechnete Bild in Abbildung(26) benötigt unter Speicheroptimierung und Multithreading im Code bei der in dieser Arbeit verwendeten Rechnerkonfiguration eine Berechnungszeit von 4 Stunden. Ferner sind in dieser Implementierung nur rudimentäre Geometrien und Lichtquellen sowie nur die vier gegebenen Materialien definierbar. Da das vorgestellte Prinzip jedoch funktioniert, gilt als nächste Zielsetzung, für die Realisierung einen der aktuell fortschrittlichsten Raytracer zu verwenden und die Methoden der Wellenoptik so zu implementieren, dass sie als Erweiterung hinzugefügt werden können. So können alle Vorteile in Bezug auf Rechengeschwindigkeit, Materialien und Komplexität der Szenerie genutzt werden.

In Kapitel 3 ist hierfür pbrt-v3 ausgewählt worden. Die Erweiterung der bestehenden Implementierung von pbrt besteht hierbei aus zwei Kernelementen, zum Einen der neuen Materialklasse “RCWA_Material” und zum Anderen der Samplingklasse “RCWA_reflection”. Die Benennung der neu implementierten Klassen und Funktionen orientiert sich hierbei an der Konvention von pbrt. In der Materialklasse wird eine Instanz des RCWA-Materials definiert, welches mit einer Geometrie, im Falle der CD einer Kreisscheibe, verknüpft ist. Hier werden die nötigen Parameter der RCWA wie die Gitterperioden oder Brechungsindizes der umgebenden Halbräume gesetzt sowie die Datenarrays der RCWA-Simulation eingelesen und diesem Materialobjekt zugewiesen. Das Kernelement der Materialklasse ist die “Compute Scattering”-Funktion, welche in der globalen Pfadfunktion (Integrator) von pbrt nach Finden eines Schnittpunktes mit einer Geometrie aufgerufen wird. In ihr gibt es beispielsweise die Möglichkeit, auf vorher definierte Texturen am aktuellen Schnittpunkt zuzugreifen. Dies ist ein Beispiel für den Vorteil der Nutzung der bestehenden pbrt-Implementierung, da diese Texturklassen schon bereitgestellt sind und hier genutzt werden können, um die Rotation

des Koordinatensystems der RCWA durchzuführen. Am Beispiel der CD ist ersichtlich, dass das Gitterkoordinatensystem rotationssymmetrisch für jeden Punkt auf der Oberfläche der CD gedreht werden muss. Anschließend werden die RCWA-Parameter, der Rotationswinkel sowie die Zeiger auf die Datenarrays an die Samplingfunktion der Klasse “RCWA_reflection” übergeben.

Wie im Grundlagenkapitel 2 vorgestellt, ist das Hauptkonzept der Interaktion von Rays mit Materialien das der BxDF. Dieses Konzept ist auf gleiche Weise in pbrt implementiert, sodass nach Finden eines Schnittpunktes im Integrator eine der Samplingfunktionen des jeweiligen Materials aufgerufen wird. Hierbei wird grundlegend zwischen zwei Typen dieser Funktionen unterschieden. Zum Einen sind Funktionen definiert, bei denen der Beitrag aus allen Raumrichtungen der Halbkugel gemäß Gleichung(6) integriert wird. Diese Funktionen “f” werden hauptsächlich bei diffusen Materialien verwendet. Zum Anderen sind Funktionen definiert, bei denen diskrete Richtungen für die Rays vorliegen, welche gemäß eines Monte-Carlo-Systems ausgewählt werden. Diese Funktionen “Sample_f” werden für Materialien wie Glas oder Metall angewandt und sind auch für die RCWA einzusetzen, da hier, wie in der Testimplementierung des vorherigen Abschnittes schon gezeigt, die Richtung des neuen Rays diskret innerhalb der Methode zu berechnen ist. Auf diese neue Funktion “Sample_f” der Klasse “RCWA_reflection” wird nun näher eingegangen.

Ein grundlegender Unterschied zur Testimplementierung im Teilkapitel zuvor ist, dass in der spektralen Version von pbrt jedem Ray ein komplettes Spektrum zugewiesen ist. In der Testimplementierung enthält der Ray jedoch nur einen Intensitätswert bei einer Wellenlänge. Somit muss auch bei der Faktorisierung jedes Pfades an einem Schnittpunkt ein volles Gewichtungsspektrum ermittelt werden. Folgendermaßen muss an dieser Stelle eine neue Samplingfunktion entwickelt werden, da in dieser Methode ebenfalls die neue ausgehende Richtung des Rays zu berechnen ist. Diese ist wie schon gezeigt abhängig von der Wellenlänge. Das neue Vorgehen wird hierbei zur Übersichtlichkeit zuerst, wie im Falle der CD angewandt, im eindimensionalen Gitterfall mit reiner Reflexion hergeleitet. Zu Beginn der Funktion “Sample_f” wird ein Zufallssystem mit zwei Zufallsvariablen erstellt. Gleichverteilt wird eine Beugungsordnung n_{rand} sowie eine Wellenlänge λ_{rand} aus den zugehörigen Parameterräumen ausgewählt. Mit den bekannten Ausgangswinkeln des aktuellen Rays ϕ_a und θ_a kann so auf gleiche Weise wie in der vorigen Implementierung mit der Floquet-Bedingung und der Stetigkeit an den Grenzflächen ein Pfad erstellt werden, der im Folgenden als Zufallspfad bezeichnet sei. Abbildung(27) soll bei dieser Herleitung unterstützen.

Der rote Pfeil entspricht hierbei dem ankommenden Ray, der die aktuelle Geometrie schneidet. Der grüne Pfeil beschreibt die Richtung des neuen ausgehenden Rays, welche in Abhängigkeit der beiden Zufallsparameter berechnet ist. Diese Richtung stellt auch die weiterführende Propagation des Rays im Integrator dar. Nun sind die vier Winkel der beiden Richtungen ϕ_a , θ_a , ϕ_e und θ_e bekannt. Das diskrete Spektrum des aktuellen Rays bzw. des zugehörigen Schnittpunktes wird nun iteriert. Für jede Wellenlänge λ_i des Spektrums und den gegebenen Ein- bzw. Ausgangswinkeln wird die zugehörige Beugungsordnung bei dieser Wellenlänge berechnet.

$$N_{xi} = n_1 \frac{\Lambda_x}{\lambda_i} (\cos\phi_e \sin\theta_e - \cos\phi_a \sin\theta_a) \quad (38)$$

Mit dieser Beugungsordnung ließe sich nun die zugehörige Pfadgewichtung aus den RCWA-Daten als spektraler Wert dieser Wellenlänge zuordnen. Jedoch entspricht aufgrund der Stetigkeit der Winkelräume diese Beugungsordnung einer rationalen Zahl. Gesucht ist jedoch eine Ganzzahl, welche in der RCWA-Simulation vorhanden ist und wofür die Beugungsef-

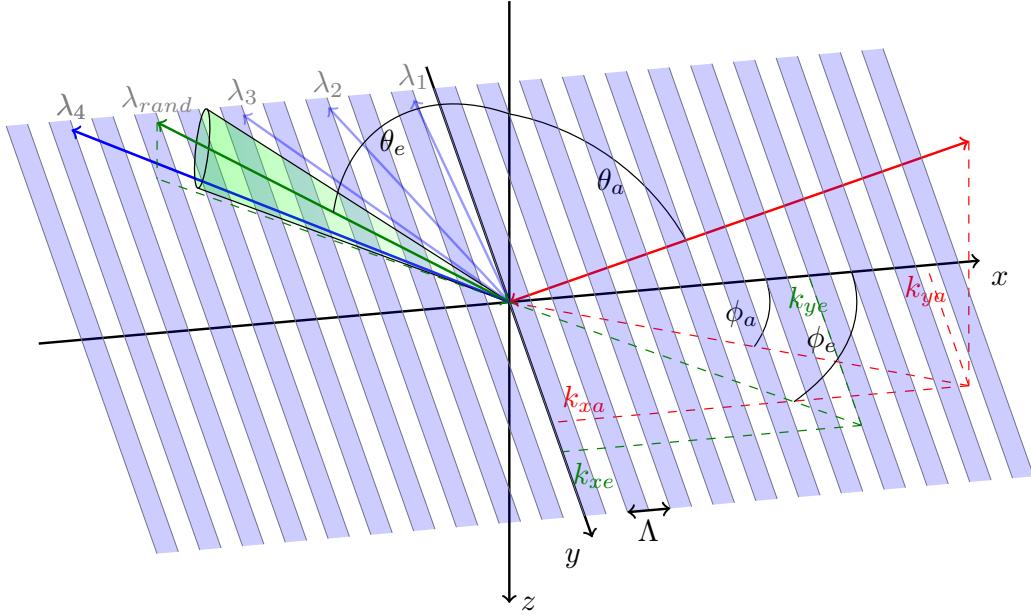


Abbildung 27: Samplingkonzept in pbrt: hierbei ist wieder nur eine rein reflektierende Struktur gezeigt. Die blauen Pfeile stellen hierbei die verschiedenen Ausgangsrays in Abhängigkeit von der Wellenlänge dar

fizienz bekannt ist. Nur falls die Differenz der ermittelten rationalen Zahl zu seiner durch Rundung zugehörigen Ganzzahl unterhalb eines vorher festgelegten Schwellenwertes Δ mit beispielsweise 0.05 liegt, ist dieser ermittelte Pfad als gültig betrachtet und die zugehörige Beugungseffizienz wird für diese Wellenlänge im Spektrum mit berücksichtigt. In Abbildung(27) ist dies durch den Kegel um die Zufallswellenlänge (grün) herum angezeigt. Beispielsweise hat das vierte Wellenlängensample (blau) eine Differenz zwischen rationaler zu Ganzzahl von weniger als Δ , somit liegt sie innerhalb des Kegels und wird im Spektrum mit berücksichtigt. Die Differenzen der anderen drei Wellenlängen sind größer als Δ , liegen also außerhalb des Kegels und werden so im Spektrum nicht berücksichtigt bzw. zu Null gesetzt. Zuletzt wird auch hier zur Ermittlung jedes Spektralwertes nochmals die Symmetriemethode aus Abschnitt 5.3 sowie die bilineare Interpolation nach Gleichung(36) für jedes Wellenlängensample durchgeführt.

An dieser Stelle sei auf die Verwendung der Integratoren in pbrt hingewiesen. Im Folgenden sind der “direct lighting” Integrator sowie das Pfadintegral zu sehen. Bei ersterem wird bei der Auswahl des zu verfolgenden Rays dessen Gewichtung bzw. ermittelte Faktorisierung beispielsweise durch die Beugungseffizienz an einer RCWA Oberfläche mit berücksichtigt. Schneidet der Ray beispielsweise eine matte Oberfläche wie die Wände der Cornellbox, so wird allgemein die ausgehende Rayrichtung gemäß einer stetigen räumlichen Wahrscheinlichkeitsverteilung ausgewählt. Somit erscheint die Oberfläche diffus. Beim direct lighting Integrator wird zusätzlich bei der zu ermittelnden Richtung noch die Gewichtung dieser Richtung bei der Auswahl einbezogen, sodass Richtungen mit kleinerer Gewichtung weniger oft ausgewählt bzw. die zugehörigen Rays abgebrochen werden. Aus diesem Grund erscheinen

im zugehörigen visualisierten Bild diffuse Oberflächen, welche nur durch eine Streubelichtung erhellt werden, als dunkel. Beim Pfadintegral dagegen wird die Gewichtung des Pfades bei der Auswahl der Rayrichtung nicht berücksichtigt. Der Nachteil hierbei ist jedoch eine höhere Rechenzeit, im unten gezeigten Fall benötigt die Implementierung 22 Minuten anstatt der 15 Minuten bei Verwendung des “direct lighting” Integrators.

In Abbildung(29) ist der Vergleich der errechneten Szenerie in pbrt mit beiden Integratoren und der Fotografie(28) der Cornellbox gezeigt.



Abbildung 28: Die Fotografie der Cornellbox

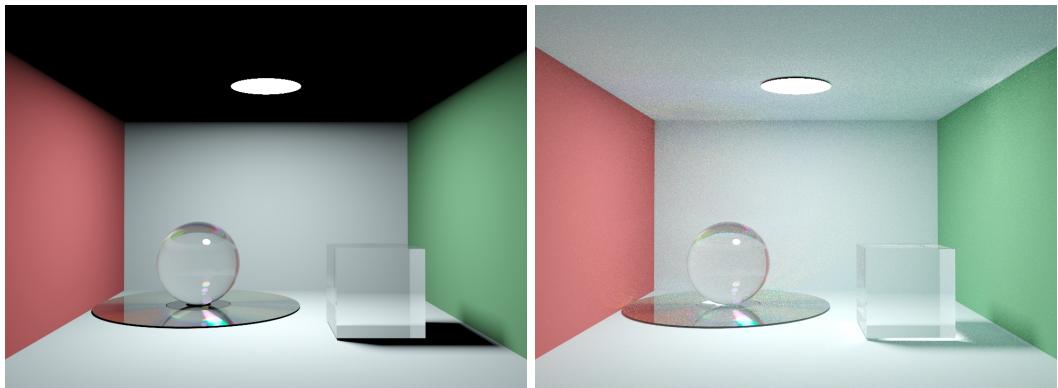


Abbildung 29: Das Berechnungsresultat der Cornellbox mit der pbrt-Implementierung: links der direct lighting Integrator, rechts das Pfadintegral

Die Vergleichsbilder aus Abbildung(29) sind mit 2045 Samples pro Pixel bei einer Auflösung von 978 zu 734 Bildpixeln berechnet worden. Im Vergleich zeigt sich bei beiden Integratoren, dass die grundlegenden Formen, Positionen und Helligkeiten der Farbartefakte auch hier übereinstimmen. Sie weisen auch innerhalb ihrer Form den gleichen Farbgradienten wie in der Fotografie auf. Das leichte Farbrauschen auf der CD-Oberfläche zeigt, dass die Überlagerung der RCWA-Strahlen die zu sehende Farbe der CD-Oberfläche erzeugt, da die gesamte Scheibe im Raytracingalgorithmus allein mit der vorgestellten Samplingfunktion berechnet wird. Die Konvergenz dieses Vorgangs ist auch mit Erhöhen der Samples pro Pixel zu beob-

achten. Im gezeigten Fall wird jedoch im Gegensatz zur vorigen testweisen Implementierung die Grundhelligkeit des Bildes erreicht und auch die Relation der Helligkeit der Farbartefakte zu den restlichen Gebieten des Bildes ist gleich. Ferner sind die Reflexionen und Bildelemente auf der CD, die gänzlich durch die RCWA-Funktion berechnet werden, mit der Fotografie übereinstimmend.

Die Integration einer Samplingfunktion mit stochastischer Auswahl von Eingangsparametern kann einen erhöhten Grad des Rauschens verursachen. Jedoch ist im rechten Bild von Abbildung(29) der gleiche Grad des Rauschens im Schatten des Glaswürfels zu sehen. Dessen Berechnung wiederum wird gänzlich durch pbtr eigene Algorithmen ohne Zugriff auf die RCWA Methoden gehandhabt, sodass die eigene Samplingfunktion kein stärkeres Rauschen induziert als die schon in pbtr implementierten Samplingfunktionen. Ferner sei an dieser Stelle auch angemerkt, dass sich die Rechengeschwindigkeit dieses Bildes in pbtr bei Entfernen der RCWA-Oberfläche nicht maßgeblich ändert. Somit fällt die RCWA-Samplingfunktion auch in diesem Aspekt nicht zurück. In Anbetracht des in Kapitel 2 und 3 genannten Standes von pbtr-v3 in Bezug auf Rechengeschwindigkeit und realistischer Abbildung können hier vorgestellten neuen Raytracingmethoden dieser Arbeit mit deren Entwicklungsstand einher gehen.

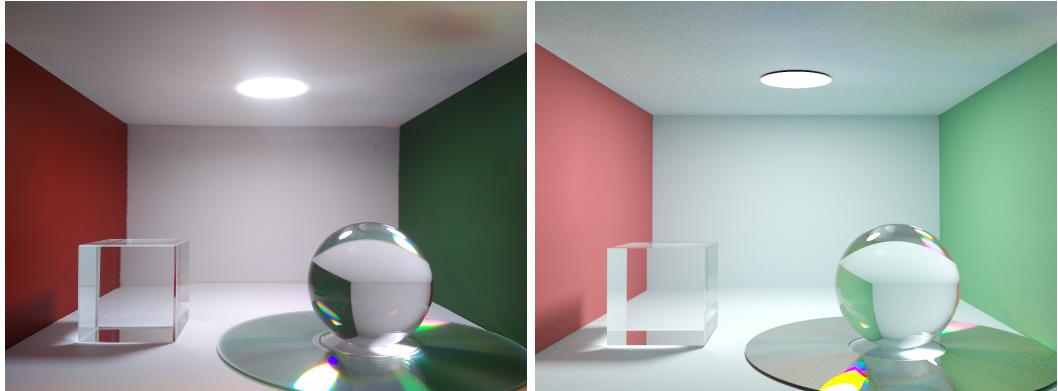


Abbildung 30: Das Berechnungsresultat der Cornellbox mit der pbtr-Implementierung in einer zweiten Anordnung: links die Fotografie der realen Szenerie, rechts das Ergebnis aus pbtr mit dem Pfadintegral

In Abbildung(30) sind die Elemente in der Box nochmals auf andere Weise angeordnet. Auch hier stimmen insbesondere die Positionen der Farbartefakte exakt überein. Dennoch zeigt diese Betrachtung die Schwierigkeiten bezüglich der korrekten Einstellung der Helligkeit der Lichtquelle. Es ist zu erkennen, dass im berechneten Bild eine leicht höhere Grundhelligkeit der Szenerie besteht. Gleichzeitig wird das Farbartefakt im vorderen linken Bereich der CD in der Fotografie weiß überstrahlt, wohingegen im berechneten Bild an dieser Stelle noch eine Gelbstufe zu erkennen ist. Diese Gegebenheit ist der verwendeten Kamera geschuldet, von der nicht ermittelt werden kann, inwieweit besonders helle Lichteinfälle auf dem Sensor durch die Kamerasoftware abgeschnitten oder skaliert werden. Dies führt an einigen Stellen dazu, dass besonders helle Farbschemen in leicht unterschiedlicher Helligkeit dargestellt sind.

Bis zu diesem Punkt wurden die Samplingfunktionen nur für ein eindimensionales Gitter mit rein reflektiertem Raypfad vorgestellt. Die endgültige Samplingfunktion, bei der sowohl

Reflektion als auch Transmission sowie zweidimensionale RCWA-Materialien berücksichtigt werden, ist zur Übersichtlichkeit in einer nicht optimierten Version in Appendix 12 gezeigt. Hier sind auch einige Erläuterungen innerhalb des Codes angefügt. Konkret bedeutet die Erweiterung zur zweidimensionalen Periodizität, dass für das Zufallssystem noch eine weitere Beugungsordnung bezüglich der y-Achse gleichverteilt ausgewählt wird. Ferner wird gleichermaßen die Symmetrieberechnung sowie die Ermittlung der resultierenden Beugungsordnung aus der Differenzbetrachtung mit Δ für die y-Ordnung durchgeführt. Auch gilt es allgemein bei Berücksichtigung der Transmission über eine Gleichverteilung in der Samplingfunktion zu entscheiden, ob eine reflektivte oder transmittierte Richtung auszuwählen ist. Hierbei ist, wie in 12 gezeigt, zusätzlich die Monte Carlo Normierung am Ende der Samplingfunktion anzupassen. Auch müssen Strukturen, die nicht wie die CD nur reflektieren, sodass keinerlei Beleuchtung von der gegenüberliegenden Seite möglich ist, von beiden Seiten der Struktur bezüglich Reflektion und Transmission simuliert und auch im Raytracer berücksichtigt werden. Dieses “see through” Prinzip ist in Abbildung(31) exemplarisch gezeigt.



Abbildung 31: Visualisierung der Cornellbox mit der 2D-periodischen Goldstruktur aus Kapitel 5 unter Verwendung des Pfadintegrals: Rechenzeit ebenfalls 22 Minuten bei 2048 Samples pro Pixel und einer Auflösung von 978 zu 734. Es sei insbesondere die goldene Farbgebung des Gitters angemerkt, die rein durch die Überlagerung der RCWA-Faktorisierung entsteht und nicht wie bei Oberflächen ohne Diffraktion explizit als Farbwert angegeben ist.

Die 2D-Goldstruktur der Teilausleuchtung wird in der Szenerie von beiden Seiten beleuchtet. Der Betrachter kann somit durch die Struktur hindurch sehen, gleichzeitig beleuchtet die Szenerie auf der Seite des Betrachters auch das Gebiet hinter der Goldstruktur. Jedoch muss hierzu die gleiche Struktur in umgekehrter z-Richtung nochmals für alle Parametermengen in der RCWA berechnet werden. Eine Vereinfachung des in Abbildung(31) Beispiels ist, dass diese Struktur auch symmetrisch bezüglich der z-Achse ist, sodass für die RCWA-Daten der anderen Goldseite die gleichen Daten verwendet werden können. Dennoch muss auch innerhalb des Raytracings die Orientierung und Rotation der Koordinatensysteme auf

der jeweiligen Seite des Gitters beachtet werden.

Diese Struktur ist hier der Vollständigkeit halber berechnet worden und soll die Möglichkeiten der Implementierung dieser Raytracing Erweiterung aufzeigen. Obwohl alle bisherigen Verifikationen der Samplingfunktionen und somit des Grundprinzips bis zu diesem Punkt erfolgreich waren und in der vergleichsweise trivialen Erweiterung zum 2D-Fall wenig Fehlermöglichkeiten bestehen, kann dieses Bild dennoch nicht verifiziert werden. Die zweidimensionale Goldstruktur ist eine reine Modellstruktur zur Verifikation in dieser Arbeit. Sie kann also nicht real in der Box fotografiert oder in sonstiger Weise bezüglich der photorealistischen Visualisierung verglichen werden.

6.4 Fazit

Im Vergleich der in diesem Kapitel gezeigten Ergebnisse mit bisherigen Arbeiten wie der Visualisierung einer CD in Fig.6 von [38] zeigt sich der Mehrgewinn des hier vorgestellten Verfahrens. Der Detailgrad der Farbartefakte und -positionen in Abbildung(29) ist maßgeblich optimiert im Vergleich zu den Ergebnissen in [38]. Mit der Perspektive, dass die CD hier lediglich ein Verifikationsmittel darstellt und die Strukturen, für welche dieses Verfahren eigentlich im Rahmen des Auslegungsziels dieser Arbeit konzipiert ist, erheblich komplexer aufgebaut sind, gilt es als nicht realistisch, mit dem in [38] und den weiteren in Kapitel 2 gezeigten bisherigen Verfahren diese Strukturen im gegebenen Anspruch zu simulieren. Da die hier gezeigten Ergebnisse im Rahmen der in Kapitel 3 definierten Ansprüche an Konvergenz und Abweichung genügen, könnte dies mit der vorgestellten Methode jedoch erreicht werden.

Es ist somit eine Methodik gefunden worden, mit welcher optische periodische Bauelemente in einem vollständigen wellenoptischen Ansatz photorealistisch visualisiert werden können. Mit der Implementierung dieser Methodik als Erweiterung für Physically Based Rendering Technique (pbrt-v3 [25]) kann diese optische Baugruppe in detaillierten Anwendungsszenarien simuliert und so ein erhöhter Grad an Information über die optische Interaktion der Bauelemente mit ihrer Umgebung gewonnen werden.

7 Zusammenfassung und Ausblick

Mit den planaren Herstellungsverfahren der Halbleitertechnik besteht seit einigen Jahren die Möglichkeit, Oberflächen in der Größenordnung der sichtbaren Wellenlängen und darunter zu strukturieren. Der Einsatz dieser Verfahren motiviert mit neuartigen Bauelementen, unter dem Oberbegriff der diffraktiven Optiken, beispielsweise Beugungseffekte in optischen Abbildungen zu nutzen. Die Herausforderungen liegen hierbei in der Auslegung dieser Optiken, da für die Ermittlung des Lichtverhaltens an Strukturierungen dieser Größenordnung rechenintensive Feldansätze nötig sind. Dies beschränkt die simulative Auslegung diffraktiver Optiken auf rein konzeptionelle Untersuchungen, in welchen innerhalb des Simulationsverfahrens das Bauteil aufgrund des Rechenaufwandes in den unteren Mikrometerbereich herab skaliert werden muss. So können diese Bauelemente bisher nur für einfache optische Abbildungen wie der Fokussierung einer ebenen Welle auf einen Brennpunkt virtuell verifiziert und getestet werden. Die Möglichkeit, das Bauelement in seiner realen Größe von mehreren Millimetern zu simulieren und so die eigentlichen Möglichkeiten auszunutzen, komplexere Abbildungen bei inhomogenen Eingangsbeleuchtungen zu realisieren, ist bisher nicht gegeben. Ferner motivieren die Verfahren der Planartechnik, in welchen sich die Herstellung von Prototypen sehr kostenintensiv gestaltet, die virtuelle Testung des Bauelementes in seinem späteren Einsatzgebiet noch vor Herstellung der ersten Optik. Hierfür ist bereits das Verfahren des Raytracing gegeben, mit welchem detaillierte Szenerien photorealistisch visualisiert werden können. Bis zum aktuellen Stand lässt dieses Verfahren jedoch die Berücksichtigung von diffraktiven Elementen aufgrund der nötigen Feldsimulation nicht zu.

Mit dieser Arbeit stehen nun drei Softwarepakete zur Verfügung, die zur Auslegung und Testung diffraktiver Optiken im virtuellen Bereich maßgeblich beitragen. Alle drei Pakete sind in dieser Arbeit verifiziert, um sie zuverlässig in den geforderten Anwendungsbereichen nutzen zu können. Im ersten Softwarepaket der Rigorous Coupled Wave Analysis (RCWA) ist nun die Erweiterung der Normalenvektortransformation eingesetzt, deren Grundidee jedoch schon von anderen Autoren veröffentlicht worden ist. Hierzu wird in dieser Arbeit eine Implementierung hergeleitet und der zugehörige maßgebliche Mehrgewinn im Kontext der Optimierung von Konvergenz- und Rechengeschwindigkeit aufgezeigt. Mit diesem erweiterten Verfahren sind nun konvergente Berechnungen über Eingangsparameterbereiche mit mehreren 1000 Simulationen im Stundenbereich möglich.

Zur Simulation der inhomogenen Eingangsbeleuchtung einer begrenzt periodischen Struktur ist die neuartige Methode der Feldbildsynthese entwickelt und verifiziert. Gerade im Vergleich zur parallel als Verifikation implementierten Fouriermethode nach [30], welche zum aktuellen Zeitpunkt als die einzige Methode für eine wellenoptische Simulation von Strukturen mit einer lateralen Größe von bis zu $100\mu\text{m}$ gesehen werden kann, zeigen sich hier eklatante Vorteile. Zum Einen reicht die gezeigte maximal mögliche Ausdehnung in den gewünschten Millimeterbereich hinein. Zum Anderen liegen die Ergebnisse im Vergleich zum Verifikationsverfahren der Finite Difference Time Domain (FDTD), insbesondere im für diese Methode eher kritischen Nahfeld, erheblich näher an deren Ergebnissen der FDTD als die der Fouriermethode. Ferner wird gleichzeitig ein Mehrgewinn in Bezug auf Rechengeschwindigkeit und Speicherbedarf ausgemacht. In Bezug auf die maximale Ausdehnung der Strukturen in der neuen Methode der Feldbildsynthese eröffnet diese neue Auslegungsmöglichkeiten für diffraktive Optiken. Sie kann bis zum gegenwärtigen Stand als einzige vollständig wellenoptische Simulation mit Feldansatz in dieser Größenordnung des Millimeterbereichs gesehen werden. Mit ihr sind nunmehr statt rein konzeptioneller Untersuchungen in Arbeitsgebieten wie Me-

talinsen, periodischen Metallstrukturen oder Gittern Simulationen dieser Bauelemente in ihrer realen Größe bei beliebiger Anordnung und verschiedenen inhomogenen Eingangsbeleuchtungen möglich.

Ferner ist in dieser Arbeit das Auslegungsziel bezüglich der Visualisierung diffraktiver Elemente mit Hilfe des Raytracing erreicht bzw. behoben. Gerade in Rückbezug auf das motivierende Bild in Abbildung(8), als auch im Vergleich zu bisherigen Arbeiten zum Raytracing diffraktiver Elemente in Fig.6 von [38] repräsentieren die Ergebnisse der Raytracingweiterung eine Lösung des motivierenden Problems. Die Implementierung dieser Arbeit kann als Zusatzpaket zu Physically Based Rendering Technique (pbrt-v3 [25]) installiert werden, sodass wie gewünscht alle bereits existierenden Möglichkeiten des Raytracings und die Vorteile gerade in Bezug auf die Rechengeschwindigkeit und möglichem Detaillierungsgrad der Szenerie von pbrt-v3 genutzt werden können.

Abschließend ist zu erwähnen, dass sich die in dieser Arbeit vorgestellten Rechenzeiten und in der Simulation maximal möglichen Ausdehnungen der diffraktiven Bauelemente auf ein Desktop Rechensystem beziehen. Dementsprechend sind bei der Verwendung der vorgestellten Simulationsverfahren auf moderneren Computersystemen die Möglichkeiten gerade in Bezug auf die maximalen lateralen Abmessungen des Bauelements erweitert.

7.1 Ausblick

Die vorgestellten Methoden sind noch in einigen Aspekten erweiterbar. Für das Softwarepaket der RCWA sind Modifikationen wie die Formulierung für anisotrope Materialien als auch die Implementierung einer Koordinatentransformation geplant. Mit letzterer soll die Variation der Struktur bezüglich der z-Richtung in einer einzigen RCWA-Schicht darstellbar sein und so der Multilayeransatz obsolet werden. Insbesondere von dieser Erweiterung wird ein zusätzlicher Konvergenz- und somit Geschwindigkeitsvorteil, gerade bezüglich der Simulation über Parameterräume für die Anwendung in Teilausleuchtung und Raytracing, erhofft. Die Teilausleuchtssoftware kann bis zum aktuellen Stand als abgeschlossen betrachtet werden. Sie ist für verschiedenste Anwendungen von der Auslegung von Head-Up-Display Systemen bis hin zur Sensorik von Halbleiterbauelementen einsetzbar. Für die vorgestellte Raytracingweiterung für pbrt-v3 sind Modifikationen bezüglich der Teilausleuchtung geplant. Das Ziel ist es hierbei, auf gleiche Weise wie im zweidimensionalen Goldgitterfall(31) Strukturen wie beispielsweise Metalinsen, diffraktive Linsen oder komplett optische Systeme bestehend aus diffraktiven Bauelementen nach dem “see-through” Prinzip im Raytracing zu visualisieren. Hierbei gilt das Konzept, mit Hilfe der Teilausleuchtung beide Seiten des System über die Parameterräume der Wellenlänge und der Eingangswinkel zu simulieren, sodass im Raytracing der physikalische Lichteinfall von beiden Seiten auf eine solche Optik berücksichtigt werden würde. Es existieren bereits Ideen, diese relativ große Menge an Simulationen in der Teilausleuchtung bezüglich der Rechengeschwindigkeit zu optimieren.

8 Literaturverzeichnis und weitere Quellen

Literatur

- [1] Andreas Erdmann. "Simulation photolithographischer Prozesse: Grundlagen, Anwendungspotential und Herausforderungen". In: *Franhofer IISB, Erlangen* ().
- [2] Lukas Bahrenberg u. a. "Spectroscopic reflectometry in the EUV for critical dimension metrology". In: *Metrology, Inspection, and Process Control for Microlithography XXIII*. Hrsg. von Ofer Adan und Vladimir A. Ukrainstev. SPIE, 24.02.2019 - 28.02.2019. ISBN: 9781510625655. DOI: 10.1117/12.2513173.
- [3] Wesley Pacheco Calixto u. a. "Electromagnetic Problems Solving by Conformal Mapping: A Mathematical Operator for Optimization". In: *Mathematical Problems in Engineering* 2010 (2010), S. 1–19. ISSN: 1024-123X. DOI: 10.1155/2010/742039.
- [4] Emine Cambel. "Seminararbeit: Floquet-Theorie für gewöhnliche Differentialgleichungen". In: *IGPM RWTH Aachen* (2013).
- [5] *History of the Cornell Box*. URL: <http://www.graphics.cornell.edu/online/box/history.html>.
- [6] M. Khorasaninejad u. a. "Polarization-Insensitive Metalenses at Visible Wavelengths". In: *Nano letters* 16.11 (2016), S. 7229–7234. DOI: 10.1021/acs.nanolett.6b03626.
- [7] Gun-Yeal Lee u. a. "Metasurface eyepiece for augmented reality". In: *Nature communications* 9.1 (2018), S. 4562. DOI: 10.1038/s41467-018-07011-5.
- [8] Antonino Calà Lesina u. a. "On the convergence and accuracy of the FDTD method for nanoplasmatics". In: *Optics express* 23.8 (2015), S. 10481–10497. DOI: 10.1364/OE.23.010481.
- [9] Lifeng Li. "Fourier modal method for crossed anisotropic gratings with arbitrary permittivity and permeability tensors". In: *Journal of Optics A: Pure and Applied Optics* 2003.5 (2003), S. 345–355. URL: stacks.iop.org/JOptA/5/345.
- [10] Lifeng Li. "Use of Fourier series in the analysis of discontinuous periodic structures". In: *Optical Society of America* 1996.13 (1996), S. 1870–1876.
- [11] LightTrans International UG. "Beyond Ray Tracing: Innovative Optical Simulations with Fast Physical Optics". In: *Free VirtualLab Fusion Seminar 2018* (2018).
- [12] LightTrans International UG. *Inclusion of gratings - LightTrans*. URL: www.lighttrans.com/applications/imaging-systems/inclusion-of-gratings.
- [13] Ming-Yi Lin u. a. "Design and fabrication of birefringent nano-grating structure for circularly polarized light emission". In: *Optics express* 22.7 (2014), S. 7388–7398. DOI: 10.1364/OE.22.007388.
- [14] Bert Hecht Lukas Novotny. "Principles of Nano-Optics: Propagation and focusing of optical fields". In: *Cambridge University Press* 2012 (2012).
- [15] Lumerical. *High-Performance Photonic Simulation Software - Lumerical*. URL: <https://www.lumerical.com/>.
- [16] Lumerical Support. *Convergence testing process for FDTD simulations*. URL: <https://support.lumerical.com/hc/en-us/articles/360034915833-Convergence-testing-process-for-FDTD-simulations>.

- [17] Lumerical Support. *Symmetric and anti-symmetric BCs in FDTD and MODE*. URL: <https://support.lumerical.com/hc/en-us/articles/360034382694-Symmetric-and-anti-symmetric-BCs-in-FDTD-and-MODE>.
- [18] M. G. Moharam, Drew A. Pommet, Eric B. Grann and T. K. Gaylord. “Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach”. In: *Optical Society of America* 1995.12 (1995), S. 1077–1086.
- [19] M. G. Moharam, Eric B. Grann, Drew A. Pommet and T. K. Gaylord. “Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings”. In: *Optical Society of America* 1995.12 (1995), S. 1068–1076.
- [20] M. N. Polyanskiy. *Refractive index database*. URL: <https://refractiveindex.info>.
- [21] Michal Radziszewski, Krzysztof Boryczko and Witold Alda. “An Improved Technique for Full Spectral Rendering”. In: () .
- [22] Elie Nadal u.a. “Optical Properties of Complex Plasmonic Materials Studied with Extended Effective Medium Theories Combined with Rigorous Coupled Wave Analysis”. In: *Materials (Basel, Switzerland)* 11.3 (2018). ISSN: 1996-1944. DOI: 10.3390/ma11030351.
- [23] Donald C. O’Shea u.a. *Diffractive optics: Design, fabrication, and test*. Second printing. Bd. Volume TT62. Tutorial Texts in Optical Engineering. Bellingham: SPIE Press, 2015. ISBN: 9780819451712.
- [24] Vitor F. Pamplona, Manuel M. Oliveira und Gladimir V. G. Baranowski. “Photorealistic models for pupil light reflex and iridal pattern deformation”. In: *ACM Transactions on Graphics* 28.4 (2009), S. 1–12. ISSN: 0730-0301. DOI: 10.1145/1559755.1559763.
- [25] Matt Pharr, Greg Humphreys und Wenzel Jakob. *Physically based rendering: From theory to implementation*. Third edition. Cambridge, MA: Morgan Kaufmann, 2017. ISBN: 9780128007099. URL: <https://www.pbrt.org>.
- [26] Philippe Lalanne and G. Michael Morris. “Highly improved convergence of the coupled-wave method for TM polarization”. In: *Optical Society of America* 1996.13 (1996), S. 779–784.
- [27] Pengfei Qiao u.a. “Theory and design of two-dimensional high-contrast-grating phased arrays”. In: *Optics express* 23.19 (2015), S. 24508–24524. DOI: 10.1364/OE.23.024508.
- [28] Abhay Sharma. *Understanding color management*. 2nd edition. The Wiley-IS&T series in imaging science and technology. Hoboken NJ: Wiley, 2018. ISBN: 9781119223689.
- [29] P. Shirley und R. K. Morley. *Realistic Ray Tracing, Second Edition*. Ak Peters Series. Taylor & Francis, 2003. ISBN: 9781568811987. URL: <https://books.google.ch/books?id=yw0tPMpCcY8C>.
- [30] Shun-Der Wu, Thomas K. Gaylord, Elias N. Glytsis und Yu-Ming Wu. “Three-dimensional converging–diverging Gaussian beam diffraction by a volume grating”. In: *Optical Society of America* 2005.22 (2005), S. 1293–1304.
- [31] *smallpt: Global Illumination in 99 lines of C++*. URL: <http://www.kevinbeason.com/smallpt/>.
- [32] Song Peng und G. Michael Morris. “Efficient implementation of rigorous coupled-wave analysis for surface-relief gratings”. In: *Optical Society of America* 1994.12 (1994).

- [33] Thomas Schuster, Johannes Ruoff, Norbert Kerwien, Stephan Rafler and Wolfgang Osten. “Normal vector method for convergence improvement using the RCWA for crossed gratings”. In: *Optical Society of America* 2007.24 (2007), S. 2880–2890.
- [34] Tobias Lange. “Modellbasierte Untersuchung zur Propagation von Gaussstrahlen durch periodisch dicke Gitter: Bachelorarbeit”. In: *Fraunhofer ILT* 2019 (2019).
- [35] TU Dortmund lsii. “Fourier Optics”. In: *lsii/cms* (). URL: www.mathematik.tu-dortmund.de/lsiii/cms/de/schriften.
- [36] Zongzhao Wang u. a. “Application of the semi-analytical Fourier transform to electro-magnetic modeling”. In: *Optics express* 27.11 (2019), S. 15335–15350. DOI: [10.1364/OE.27.015335](https://doi.org/10.1364/OE.27.015335).
- [37] Zongzhao Wang u. a. “Theory and algorithm of the homeomorphic Fourier transform for optical simulations”. In: *Optics express* 28.7 (2020), S. 10552–10571. DOI: [10.1364/OE.388022](https://doi.org/10.1364/OE.388022).
- [38] Sebastian Werner u. a. *Scratch iridescence: Wave-optical rendering of diffractive surface structure*. URL: <http://arxiv.org/pdf/1705.06086v1.pdf>.
- [39] Wikipedia, Hrsg. *Bilinear interpolation*. URL: https://en.wikipedia.org/w/index.php?title=Bilinear_interpolation&oldid=1000748657.
- [40] Liangxin Yang u. a. “Light shaping by freeform surface from a physical-optics point of view”. In: *Optics express* 28.11 (2020), S. 16202–16210. DOI: [10.1364/OE.392420](https://doi.org/10.1364/OE.392420).

9 Appendix A: Einstellung der FDTD Simulation

In diesem Abschnitt werden die Einstellungen der FDTD Simulationen aufgeführt, welche in den hier vorgestellten Simulationen angewandt worden sind. Als FDTD-Simulationstool ist in dieser Arbeit die kommerzielle Software Lumerical [15] ausgewählt. Der folgende Leitfaden orientiert sich unter anderem an [8], wo Konvergenzuntersuchungen mit dem Hintergrund der Plasmonik an einer Goldstruktur getätigt worden sind. Hierfür existiert eine ideale, analytische Lösung, welche mit Hilfe der Mie-Streuung formuliert werden kann. Ferner sind auch die Ergebnisse von [16] und [17], wo anhand der analytischen Mielösung für einen Golddraht das ideale Vorgehen zur Konvergenzuntersuchung erläutert wird, in diesen Leitfaden mit einbezogen.

Symmetrie

Sowohl in der RCWA-Zellensimulation als auch in der Teilausleuchtung werden Strukturen in lateraler Richtung durch eine ungerade Anzahl an Diskretisierungspunkten repräsentiert. Somit muss auch in der FDTD eine ungerade Anzahl an Diskretisierungspunkten in der xy-Ebene gesetzt werden. Die Diskretisierung in z-Richtung wird hiervon nicht beeinflusst. Ferner besteht bei der Verifikation der Teilausleuchtung die Idee, in der FDTD Symmetrien der Anordnung auszunutzen, um so die maximal mögliche laterale Ausdehnung der Beispielstruktur zu erhöhen. Die Beispielstrukturen in dieser Arbeit sind so aufgebaut, dass sie in vier Quadranten eingeteilt werden können. So muss beispielsweise bei einem normal und mittig einfallenden Gaußstrahl nur ein Quadrant der Anordnung simuliert werden. Die Lösung des Gesamtgebietes ergibt sich dann aus Drehung und Zusammensetzung dieser Quadrantenlösung. Auch hier gilt es besonders auf die symmetrische Lage mit dem Punkt des Ursprungs bezüglich der Quelle zu achten. Lumerical bietet die Möglichkeit bei symmetrischen Anordnungen, Grenzflächen an Symmetrielinien zu setzen. Wichtig ist hierbei, welche Art der Symmetrie an diesen Linien gesetzt werden muss. Hierbei werde auf [17] verwiesen. Die Unterscheidung zwischen symmetrischer und antisymmetrischer Grenzfläche orientiert sich an der Polarisation der Quelle. Das Symmetriekonzept kann nur angewandt werden, wenn die Quellenpolarisation tangential zu einer Symmetrielinie ist. Dies wird in den Beispielen dieser Arbeit derart angewandt, dass in allen Simulationen ein linearer Polarisationswinkel von 0° gesetzt wird und so die Grenzflächen parallel zur x-Achse (y-Randflächen) symmetrisch und dementsprechend die x-Randflächen antisymmetrisch festgelegt sind. Auch sollten zur Verifikation Strukturen gewählt werden, die im Yee-Grid der FDTD besonders akkurat abgebildet werden. Aus diesem Grund dienen beispielsweise die rechteckigen Goldplättchen der Verifikation im Kapitel der Teilausleuchtung 5.

FDTD-Fernfeld

Im Gegensatz zur Nahfeldbetrachtung, in der das Feld explizit in Ebenen, welche in der Simulationszelle stehen, betrachtet wird, soll hier die Fernfeldnäherung der FDTD genutzt werden. In Lumerical geschieht dies über die Überlagerung von propagierten ebenen Wellen aus dem Nahfeld.

Randbedingungen

Sowohl für die reine RCWA-Zellensimulation als auch für die Teilausleuchtung werden für die Seitenwände der FDTD-Zelle periodische Randbedingungen (PBC) gewählt. Diese Art der Randbedingungen basiert auf dem Bloch Theorem, welchem auch die Differentiallösungen der RCWA zugrunde liegen. In ihr gilt die Voraussetzung, dass

für die Feldlösung $\phi(x + \Lambda_x) = \phi(x)$, mit Λ_x als der Kantenlänge der betrachteten Zelle, gilt. Somit entspricht die Feldlösung auf dem linken Rand des Lösungsgebietes der Feldlösung auf dem gegenüberliegenden rechten Rand des Gebietes. Die Deckel- und Bodenflächen der FDTD-Zelle werden als stabilisierte Perfect Matched Layer (PML) festgelegt. Diese stellt eine Erweiterung des Lösungsraumes durch eine festgesetzte Anzahl an Schichten dar, welche durch ihr anisotropes Material im Idealfall das Feld an der Grenzfläche perfekt reflektionsfrei absorbieren. In Veröffentlichungen zu diesem Thema wurde eine Anzahl von 8 - 12 PML-Schichten empfohlen. Für alle Randbedingungen, insbesondere für die PML, sollte ein Mindestabstand von $\frac{\lambda_{min}}{2}$ von metallischen Elementen zu den Randbedingungen gelten. Deshalb sei bei allen Simulationen dieser Arbeit festgelegt, dass alle Quellen, Strukturen und Beobachtungspunkte um mindestens $\frac{\lambda}{2}$ von ihrer nächsten Randfläche entfernt seien. Der Raum dazwischen sei in jedem Fall homogen mit Luft gefüllt. Ausgenommen von dieser Betrachtung ist die Verwendung von Symmetrielinien, wo das Setzen der Grenzflächen schon im oberen Abschnitt Symmetrie beschrieben ist.

Diskretisierungsstufen von Raum und Zeit

Die Feinheit des Meshgrids und der Zeitschritte stellen hier die aufwendigste Einstellung dar, da der Rechenaufwand bei den vorgestellten Beispielstrukturen im Bereich von mehreren Gigabyte und Stunden an Simulationszeit liegt. Gerade da in dieser Arbeit eine Methode verifiziert werden soll, die für besonders große Strukturen konzipiert ist, sollte auch die FDTD-Struktur größtmöglich gesetzt sein. Als erste Regelung wird hierzu in [8] und [16] angeführt, dass die räumliche Schrittweite der Zellenpunkte in allen drei Raumachsen gleich sein sollte ($d_x = d_y = d_z$). Grundsätzlich gilt für FDTD-Simulationen die Abschätzung des Courant Limits aus [8] für die Raum- Δx und Zeitdiskretisierungsschritte Δt :

$$\min(\Delta x, \Delta y, \Delta z) \leq \frac{\lambda_{min}}{10n_{max}} \quad \text{und} \quad \Delta t \leq \frac{1}{v\sqrt{\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2}}} \quad (39)$$

Hierbei beschreibt n_{max} den betragsmäßig größten Brechungsindex aller in der Gesamtstruktur vorkommenden Materialien und λ_{min} die kleinste simulierte Wellenlänge in der gesamten Bandbreite des Eingangswellenpaketes. Mit der Vorschrift einer gleichen Diskretisierung in alle drei Raumrichtungen gilt nun:

$$\Delta t \leq \frac{\Delta x}{\sqrt{3}\frac{c_0}{\sqrt{\epsilon_\infty}}} \quad (40)$$

Lumerical verwendet bei der Erstellung des räumlichen Meshgitters eine eigene “Conformal Mesh Technologie” (CMT), deren genaues Verfahren nicht veröffentlicht wird. Es ist bekannt, dass beispielsweise in der Nähe von Metallkanten automatisch die räumliche Diskretisierung verfeinert wird. Durch schrittweises Erhöhen der Gesamtzahl an Diskretisierungspunkten einer kartesischen Achse von 6 bis hin zu maximal 34 Diskretisierungspunkten pro Arbeitswellenlänge ist die Konvergenz der Struktur zu testen. Die Simulationszeit ist in allen Simulationen auf 100fs gesetzt. Für diese Abschätzung gilt, dass die Simulationszeit mindestens der Durchlaufzeit (Lichtgeschwindigkeit des optisch dichtesten Mediums der Struktur mal Dicke der Anordnung) entsprechen muss. Ferner stoppt Lumerical automatisch, wenn die Gesamtenergie in der Simulation 1e-5 der Eingangsenergie erreicht.

Quelle

Wie schon zu den Randbedingungen erläutert, sollte die Quelle im Idealfall bezüglich ihrer lateralen Anordnung weitestgehend möglich von den seitlichen Randflächen der FDTD-Zelle entfernt sein. So sollte nahezu kein Feld mehr in den randnahen Gebieten angeregt werden. In der ebenen Wellenanregung der Verifikation für die NVT ist dies nicht zu realisieren. Die Startphase der Eingangswelle in der Quelle spielt hier keine Rolle, da in den folgenden Analysen immer der Absolutbetrag der Amplitude des Feldes betrachtet wird. Zu Beachten sind bei der Verwendung von Metallen im Simulationsgebiet, dass bei einer Eingangswellenlänge in der Nähe einer Absorptionskante des Metalls Abweichungen und Überschwinger eher auftreten als weiter entfernt. Bei der Verwendung von Gold empfiehlt sich so beispielsweise eine Arbeitswellenlänge von 500nm. Dieses Problem wird in [8] beschrieben; hier wurden Divergenzprobleme bei einer Schrittweite von 5nm beobachtet, die aufgrund der verwendeten Struktur und der vorhandenen analytischen Lösung zum Vergleich auf die Resonanzen des Goldes rückzuführen sind. Ferner entspricht die Anregung im Modell der RCWA und der Teilaustrahlung einer kontinuierlichen Welle. Diese “Continuous Wave” (CW) Anregung muss ebenfalls in der FDTD berücksichtigt werden. Die erste Anforderung ist hierbei die Gesamtzeit der Simulation so festzulegen, sodass alle eventuellen Einschwingvorgänge abgeklungen sind. In Lumerical wird diese CW-Anregung über einen möglichst langsam ansteigenden Puls realisiert; der Einschwingungspunkt wird automatisch beim Setzen der Wellenlänge mit berücksichtigt. Da hier auch keine wellenlängenverschiebende Materialien betrachtet werden, ist die Simulation mit einem breitbandigen Spektrum keine Fehlerquelle. Ferner sinkt mit der Pulslänge die Dispersion, sodass im Endeffekt weniger räumliche Diskretisierungspunkte pro Wellenlänge für die gleiche Konvergenz nötig sind. Die gesamte Definition von Pulslänge und zeitlicher Pulsposition wird automatisch über die Wellenlängenangabe in Lumerical vollzogen.

Beobachtungspunkte

In Lumerical können über geometrische Monitorobjekte Flächen definiert werden, in welchen das elektrische Feld bezüglich aller Vektorkomponenten nach der Berechnung explizit angezeigt werden kann. Die Monitore geben automatisch das Feld bei der gesetzten Mittenfrequenz des Eingangsspektrums an.

Generell wird von Lumerical empfohlen, zu Beginn der Simulation eine Divergenzüberprüfung durchzuführen. Hierzu sollten testweise alle Randbedingungen zu perfektem Metall (PMC) gesetzt werden. Besteht der Fehler weiterhin, ist der Zeitschritt d_t zu grob gewählt. Ferner ist häufig die PML-Randbedingung eine Divergenzursache, beispielsweise weil zu wenige PML-Schichten verwendet wurden.

10 Appendix B: Formulierung der allgemeinen Reihenlösungen für die 2D-RCWA

Im folgenden Abschnitt werden die im Zusammenhang mit dieser Arbeit verwendeten Reihenentwicklungen der zweidimensionalen RCWA aus [34] vorgestellt. Als Ausgangspunkt dient das in Abbildung(32) gezeigte Gitterproblem.

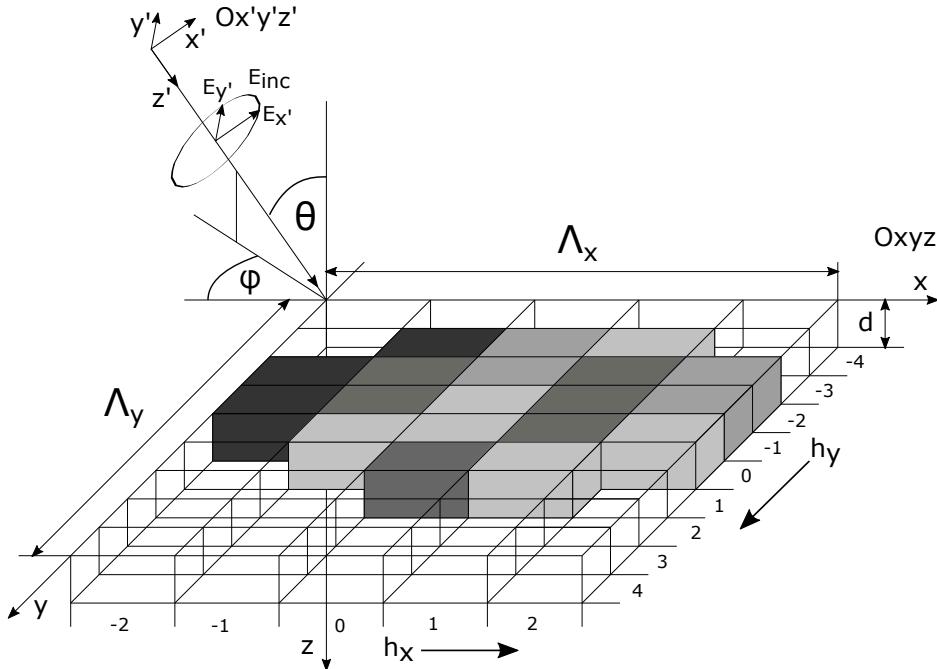


Abbildung 32: Allgemeiner Aufbau des Gitterproblems: eine ebene Welle fällt auf eine Zelle des periodischen Gitters ein. Die Grautöne jeder Mikrozelle repräsentieren unterschiedliche exemplarische Brechungsindizes, die innerhalb einer Mikrozelle an jedem Punkt konstant sind

Eine ebene homogene Welle der Vakuumwellenlänge λ trifft in einem homogenen Gebiet mit Brechungsindex n_1 unter einem azimuthalen ϕ und einem polaren Winkel θ auf ein periodisches Gitter aus isotroper Materialien. Das Gitter selbst sei durch eine einzelne Zelle beschreibbar, die durch Translation in x- und y-Richtung unendlich oft periodisch angeordnet ist. Diese Zelle habe eine Dicke d und eine Periode Λ_x bzw. Λ_y entlang der jeweiligen Koordinatenachse. Für jede Gitterachse stehen $2N_x - 1$ bzw. $2N_y - 1$ Punkte in der xy-Ebene bei $z = 0$ zur Verfügung. Hierbei bezeichnen N_x und N_y die Gesamtzahl der in dieser Simulation betrachteten Beugungsordnungen der jeweiligen Achse. Die Struktur der Zelle wird durch den Brechungsindex bzw. die Permittivität beschrieben, welcher für jeden der diskretisierten Punkte (Mikrozelle) anzugeben ist und innerhalb einer der durch Grautöne in Abbildung(32) gezeigten Mikrozellen konstant ist. Ferner ändert sich der Brechungsindex in dieser Betrachtung entlang der z-Achse über die Dicke d hinweg nicht. Das homogene Medium unterhalb der Struktur wird durch den Brechungsindex n_3 repräsentiert. Das elektrische Feld lässt sich in den Gebieten über und unter der Struktur als Fourierreihenentwicklung darstellen.

$$\mathbf{E}_1(x, y, z) = \mathbf{E}_{\text{inc}} + \sum_j \sum_i \mathbf{R}_{ji} \exp[-j(k_{xi}x + k_{yj}y - k_{1,zji}z)] \quad (41)$$

$$\mathbf{E}_3(x, y, z) = \sum_j \sum_i \mathbf{T}_{ji} \exp(-j[k_{xi}x + k_{yj}y + k_{3,zji}(z - d)]) \quad (42)$$

wobei \mathbf{R}_{ji} bzw. \mathbf{T}_{ji} die komplexen Amplitudenvektoren dieser ebenen Wellen repräsentieren. Die Wellenvektoren sind hierbei folgendermaßen festgelegt:

$$k_{xi} = k_0 n_1 \sin \theta \cos \phi - i \frac{2\pi}{\Lambda_x} \quad (43)$$

$$k_{yj} = k_0 n_1 \sin \theta \sin \phi - j \frac{2\pi}{\Lambda_y}$$

$$k_{med zji} = \begin{cases} +\sqrt{(k_0 n_{med})^2 - k_{xi}^2 - k_{yj}^2} & \text{für } k_0 n_{med} > \sqrt{k_{xi}^2 + k_{yj}^2} \\ -j\sqrt{k_{xi}^2 + k_{yj}^2 - (k_0 n_{med})^2} & \text{für } k_0 n_{med} < \sqrt{k_{xi}^2 + k_{yj}^2} \end{cases}$$

mit $\text{med} = 1$ bzw 3 für das jeweilige Gebiet. Die Indizes i bzw. j iterieren die Beugungsordnungen von $-n_x$ bis n_x bzw. $-n_y$ bis n_y mit $n_x = N_x // 2$ bzw. $n_y = N_y // 2$. // beschreibt hierbei eine Ganzzahldivision. Aufgrund der Stetigkeitsbedingung der tangentialen Feld- und somit auch der Wellenvektorkomponenten können im Gitterbereich das elektrische und magnetische Feld durch eine Fourierreihenentwicklung mit den gleichen Basen angegeben werden:

$$\mathbf{E}_g(x, y, z) = \sum_j \sum_i [S_{xji}(z) \hat{\mathbf{e}}_x + S_{yji}(z) \hat{\mathbf{e}}_y + S_{zji}(z) \hat{\mathbf{e}}_z] \exp[-1j(k_{xi}x + k_{yj}y)] \quad (44)$$

$$\mathbf{H}_g(x, y, z) = -j \left(\frac{\epsilon_0}{\mu_0} \right)^{1/2} \sum_j \sum_i [U_{xji}(z) \hat{\mathbf{e}}_x + U_{yji}(z) \hat{\mathbf{e}}_y + U_{zji}(z) \hat{\mathbf{e}}_z] \exp[-1j(k_{xi}x + k_{yj}y)]$$

Hierbei sind die Amplituden S und U skalar und die Vektoren $\hat{\mathbf{e}}$ die Einheitsvektoren der Koordinatenachsen. Der Vorfaktor des H-Feldes entspricht der Verknüpfung von E und H-Feld einer TEM-Welle.

Die Beugungseffizienzen ergeben sich nach Lösung des Gesamtsystems zu:

$$DE_{rji} = \left(|R_{s,ji}|^2 + \frac{1}{n_1^2} |R_{p,ji}|^2 \right) \text{Re} \left(\frac{k_{1,zji}}{k_0 n_1 \cos \theta} \right) \quad (45)$$

$$DE_{tji} = \left(|T_{s,ji}|^2 + \frac{1}{n_3^2} |T_{p,ji}|^2 \right) \text{Re} \left(\frac{k_{3,zji}}{k_0 n_1 \cos \theta} \right)$$

Sie beschreiben jeweils den Anteil der Leistungsdichte von Beugungsordnung (j, i) bezogen auf die gesamte Leistungsdichte der einfallenden Welle. Die Amplituden $R_{s,ji}$, $R_{p,ji}$, $T_{s,ji}$ und $T_{p,ji}$ ergeben sich aus der Transformation in ein Flächenkoordinatensystem (sp-System) normal zum Ausgangswellenvektor der betrachteten Ordnung. Hierzu liegen die jeweiligen Beugungsordnungen (j, i) projiziert auf die xy-Ebene in einem azimuthalen Winkel zur x-Achse:

$$\varphi_{j,i} = \tan^{-1}(k_{yj}/k_{xi}) \quad (46)$$

Mit der TEM-Wellennatur im isotropen und homogenen äußeren Halbraum liegen die Feldvektoren von E und H-Feld in einer Ebene senkrecht zur Ausbreitungsrichtung. Diese Ebene

lässt sich durch zwei Richtungsvektoren beschreiben, sodass für die Feldamplituden in diesem System nur zwei statt den drei kartesischen Koordinaten angegeben werden müssen:

$$\hat{\mathbf{e}}_s = \frac{\hat{\mathbf{e}}_z \times \mathbf{k}}{|\hat{\mathbf{e}}_z \times \mathbf{k}|} \quad \text{und} \quad \hat{\mathbf{e}}_p = \frac{\hat{\mathbf{e}}_s \times \mathbf{k}}{|\hat{\mathbf{e}}_s \times \mathbf{k}|} \quad (47)$$

Hierbei ist zu beachten, dass der \mathbf{k} -Vektor im oberen reflektiven Halbraum eine negative z -Komponente besitzt. In diesem Flächenkoordinatensystem lassen sich die Amplituden der jeweiligen Ordnung wie folgt darstellen:

$$R_{s,ji} = \mathbf{R}_{ji} \cdot \hat{\mathbf{e}}_s = \cos \varphi_{ji} R_{yji} - \sin \varphi_{ji} R_{xji} \quad (48)$$

$$R_{p,ji} = j n_1 \mathbf{R}_{ji} \cdot \hat{\mathbf{e}}_p = \frac{-j}{k_0} [\cos \varphi_{ji} (k_{1,zji} R_{xji} + k_{xji} R_{zji}) + \sin \varphi_{ji} (k_{yj} R_{zji} + k_{1,zji} R_{yji})]$$

Analog werden an der zweiten Grenzfläche die Amplituden angegeben.

$$T_{s,ji} = \cos \varphi_{ji} T_{yji} - \sin \varphi_{ji} T_{xji} \quad (49)$$

$$T_{p,ji} = (j/k_0) [\cos \varphi_{ji} (k_{3,zji} T_{xji} - k_{xi} T_{zji}) - \sin \varphi_{ji} (k_{yj} T_{zji} - k_{3,zji} T_{yji})]$$

11 Appendix C: Formulierung der Gaußschen Grundmode als Eingangsbeleuchtung

Die Gaußsche Grundmode nach [30] lautet im strahleigenen Koordinatensystem $Ox'y'z'$:

$$\mathbf{E}^{inc} = \left[\frac{w_{0x'}}{w_{x'}(z')} \right]^{1/2} \left[\frac{w_{0y'}}{w_{y'}(z')} \right]^{1/2} \exp \left\{ - \left[\left(\frac{x'}{w_{x'}(z')} \right)^2 + \left(\frac{y'}{w_{y'}(z')} \right)^2 \right] \right\} \quad (50)$$

$$\exp \left\{ -j \frac{1}{2} k \left[\frac{x'^2}{R_{x'}(z')} + \frac{y'^2}{R_{y'}(z')} \right] \right\} \exp(-jk_0 n_1 z') \hat{\mathbf{e}}$$

In der Strahltaille bei $z' = 0$ wird der Strahlradius $w_{0x'}$ bzw. $w_{0y'}$ definiert, wobei im Falle eines nicht zur z' -Achse rotationssymmetrischen Intensitätsprofils beide Strahlradien angegeben werden müssen. Der Strahlradius wird als der Abstand des Punktes bzw. der Kreislinie von der z' -Achse definiert, an welchem die Intensität nur noch $\frac{1}{e}$ der Intensität des zugehörigen Hauptachsenpunktes beträgt. Der Strahlradius $w_{x'}(z')$ am Ort z' , der Radius der Phasenfläche $R_{x'}(z')$ am Ort z' und die Rayleighdistanz $z_{0x'}$ lassen sich wie folgt angeben:

$$w_{x'}(z') = w_{0x'} \left[1 + \left(\frac{z'}{z_{0x'}} \right)^2 \right]^{1/2} \quad (51)$$

$$R_{x'}(z') = z' \left[1 + \left(\frac{z_{0x'}}{z'} \right)^2 \right] \quad (52)$$

$$z_{0x'} = \frac{\pi n_1}{\lambda_0} w_{0x'}^2 \quad (53)$$

Analog werden die drei Parameter für die y-Komponente definiert. Der Polarisationseinheitsvektor $\hat{\mathbf{e}}$ ergibt sich aus der Kugelrotation des Winkels ψ der linearen Polarisation vom Koordinatensystem $Ox'y'z'$ nach $Oxyz$:

$$\begin{aligned}\hat{\mathbf{e}} = e_x \hat{\mathbf{x}} + e_y \hat{\mathbf{y}} + e_z \hat{\mathbf{z}} &= (\cos\psi_e \cos\phi_e \cos\theta_e - \sin\psi_e \sin\phi_e) \hat{\mathbf{x}} + \\ &(\cos\psi_e \sin\phi_e \cos\theta_e + \sin\psi_e \cos\phi_e) \hat{\mathbf{y}} + (\cos\psi_e \sin\theta_e) \hat{\mathbf{z}}\end{aligned}\quad (54)$$

Zur Übersichtlichkeit seien hier einmal kurz der Wellenvektor der Hauptachse \mathbf{k} sowie der Subwellenvektor $\mathbf{k}_{xn,ym}$ angegeben.

$$\begin{aligned}\mathbf{k} = \sin\theta_e \cos\phi_e \hat{\mathbf{x}} + \sin\theta_e \sin\phi_e \hat{\mathbf{y}} + \cos\theta_e \hat{\mathbf{z}} \quad \text{bzw.} \quad \hat{\mathbf{k}} = \mathbf{k}/|\mathbf{k}| \\ \mathbf{k}_{xn,ym} = k_{xn} \hat{\mathbf{x}} + k_{ym} \hat{\mathbf{y}} + k_{zm} \hat{\mathbf{z}} \quad \text{bzw.} \quad \hat{\mathbf{k}}_{xn,ym} = \mathbf{k}_{xn,ym}/|\mathbf{k}_{xn,ym}|\end{aligned}\quad (55)$$

Ferner gilt es den Einheitsvektor der TM-Polarisation im Koordinatensystem $Oxyz$ anzugeben, die sich aus der Formulierung der \mathbf{e}_s und \mathbf{e}_p Vektoren aus Appendix 10 ergibt, da dieser Vektor exakt \mathbf{e}_p für die nullte Ordnung entspricht.

$$\begin{aligned}\hat{\mathbf{e}}_s = -\sin\phi_e \hat{\mathbf{x}} + \cos\phi_e \hat{\mathbf{y}} + 0 \cdot \hat{\mathbf{z}} \\ \mathbf{e}_p = \hat{\mathbf{e}}_s \times \hat{\mathbf{k}} \quad \text{bzw.} \quad \hat{\mathbf{e}}_p = \mathbf{e}_p/|\mathbf{e}_p|\end{aligned}\quad (56)$$

So wird nun der Einheitspolarisationsvektor des Subvektors definiert:

$$\hat{\mathbf{e}}_{xn,ym} = \frac{(\hat{\mathbf{k}} \times \hat{\mathbf{e}}) \times \hat{\mathbf{k}}_{xn,ym}}{|(\hat{\mathbf{k}} \times \hat{\mathbf{e}}) \times \hat{\mathbf{k}}_{xn,ym}|} \quad (57)$$

Zuletzt kann der Polarisationswinkel für jede Fourierordnungskombination angegeben werden:

$$\Psi_{e,xn,ym} = \cos^{-1}(\hat{\mathbf{e}}_{TM} \cdot \hat{\mathbf{e}}_{xn,ym}) \frac{(\hat{\mathbf{e}}_{TM} \times \hat{\mathbf{e}}_{xn,ym}) \cdot \mathbf{k}}{|\hat{\mathbf{e}}_{TM} \times \hat{\mathbf{e}}_{xn,ym}| |\mathbf{k}|} \quad (58)$$

12 Appendix D: Die C++ Implementierung der RCWA Samplingfunktion in pbrt-v3

```
116
117 Spectrum RCWA_reflection::Sample_f(const Vector3f &wo, Vector3f *wi,
118     const Point2f &u, Float *pdf,
119     BxDFType *sampledType) const {
120
121     //Initialisierung der Zufallsfunktion
122     static std::default_random_engine gen;
123     std::uniform_real_distribution<double> dis(0., 1.); //TODO: stattdessen u verwenden und transformieren
124
125     //Monte Carlo System
126     int wave_ind_rand = int(dis(gen) * Float(wavel_size)); //Monte Carlo → Zufallswellenlaenge des Systems
127     Float wavelength_rand = Float(wave_ind_rand) * 300. / Float(wavel_size) + 400.;
128     int n_x = (namount_x - 1) / 2;
129     int n_y = (namount_y - 1) / 2;
130     int n_rand_x = int(dis(gen) * Float(namount_x)) - n_x; //Monte Carlo → Zufallsordnung
131     int n_rand_y = int(dis(gen) * Float(namount_y)) - n_y; //Monte Carlo → Zufallsordnung
132     Float RT = 1.;
133     if (see_through == true) RT = dis(gen) > 0.5 ? 1. : -1.; //see = 1. -> Reflweg; see = -1. -> Transweg
134     //Transweg bedeutet bzgl. des Vorzeichens, der Strahl wechselt den Halbraum →
135
136     //Eingangsray Raytracer = Ausgangsordnung RCWA
137     //behandelte Vektoren wo und wi sind in einem lokalen Oberflaechensystem
138     //Rotation gibt Orientierung des RCWA-Systems bzgl. dieses Koordsys an
139     Transform rot = RotateZ(-Rot); //rotiere gegen den Uhrzeiger bei z aus Papierebene heraus
140     Vector3f wo_rot = rot(wo);
141     Float theta_out = acos(AbsCosTheta(wo_rot)); //bei ebener Oberflaeche in xy-Ebene
142     Float phi_out = atan2(wo_rot.y, wo_rot.x); //Ausgang der RCWA
143
144     Float n_med = wo_rot.z > 0. ? n1 : n3; //ob ueber oder unter der Struktur
145     //n1 und n3 richten sich hierbei nach der Halbraumbetrachtung
146
147     //Wellenvektorkomponenten
148     Float kmed = (2 * Pi / wavelength_rand) * n_med; //Wellenvektor im Medium
149     Float kxo = cos(phi_out) * sin(theta_out) * kmed;
150     Float kyo = sin(phi_out) * sin(theta_out) * kmed;
151     Float kxe = kxo + 2. * Float(n_rand_x) * Pi / LAMBDA_x;
152     Float kye = kyo + 2. * Float(n_rand_y) * Pi / LAMBDA_y;
```

```

153  Float phi_in = atan2(-kye, -kxe); //Eingangswinkel der RCWA
154
155  Float kz = sqrt(kmed*kmed - kxe*kxe - kye*kye); //kz hierbei      ↵
     Betragsskomponenten von kz
156
157  Float theta_in;
158  if (abs(kz / kmed) < 1) theta_in = acos(kz / kmed); //Weg nur existent    ↵
     falls ausbreitungsfähig
159  else {
160      *wi = Vector3f(0., 0., 0.);
161      return Spectrum(0.f);
162  }
163
164  Transform rotrot = RotateZ(Rot); //rotiere im Uhrzeiger bei z aus      ↵
     Papierebene heraus; entgegengesetzt zu oben
165  Vector3f wi_rot = Vector3f(sin(theta_in)*cos(phi_in), sin(theta_in)*sin      ↵
     (phi_in), cos(theta_in) * copysign(1., wo_rot.z) * RT);
166 //das Vz von wo und das RT-Kriterium bestimmen den resultierenden Halbraum
167 if (wo_rot.y * wi_rot.y > 0.) { //falls wo_rot und wi_rot gleiches y-Vz    ↵
     haben -> selbe y-Halbebene -> Abbruch
168     *wi = Vector3f(0., 0., 0.);
169     return Spectrum(0.f);
170 }
171 *wi = rotrot(wi_rot); //neue Richtung zur Weiterpropagation
172
173 phi_in = atan2(kye, kxe); //TODO: hier statt Neuberechnung diesen Winkel    ↵
     über Symmetriebetrachtung
174 auto ergebnis = symm(phi_in, phi_out); //Symmetriefunktion
175
176 float angle_x, angle_y, wavelength, N_x, N_y;
177 int N_x_int, N_y_int, wave_loop;
178 Spectrum r(0.f);
179
180 angle_x = cos(phi_in) * sin(theta_in) - cos(phi_out) * sin(theta_out);
181 angle_y = sin(phi_in) * sin(theta_in) - sin(phi_out) * sin(theta_out);
182
183 const float *DE = wo_rot.z > 0. ? (RT > 0. ? DER_1 : DET_1) : (RT > 0. ?    ↵
     DER_3 : DET_3); //1. oben oder unten 2. R oder T
184 float MC = see_through == true ? 2. * namount_x * namount_y : namount_x *    ↵
     namount_y; //Monte Carlo: Faktor 2 aufgrund der oberen    ↵
     gleichverteilten Wegentscheidung
185 //zur Logik stelle man sich die CD mit see_through = true vor -> Hälfte    ↵
     der Strahlen wird zu 0 gesetzt -> Faktor 2
186
187 for (int i = 0; i < 60; i = i + 1) { //siehe spectrum.h Zeile 282 und    ↵
     293 bzw 292 fuer den Wellenlängenbereich
188
189     wave_loop = int(i / (60. / float(wavel_size)) + 0.1); //RCWA-Index
190     wavelength = 400. + float(wave_loop) * 300. / float(wavel_size); //    ↵
     Schleifenwellenlänge
191
192     N_x = angle_x * n1 * LAMBDA_x / wavelength; //TODO: eleganter
193     N_x_int = int(copysign(abs(angle_x * n1 * LAMBDA_x / wavelength) +    ↵
     0.5, angle_x * n1 * LAMBDA_x / wavelength));
194
195     N_y = angle_y * n1 * LAMBDA_y / wavelength;

```

```

196     N_y_int = int(copysign(abs(angle_y * n1 * LAMBDA_y / wavelength) +      ↵
197     0.5, angle_y * n1 * LAMBDA_y / wavelength));
198     if ((abs(N_x - N_x_int) < 0.1) && (abs(N_y - N_y_int) < 0.1)) {      // ↵
199         TODO: eleganter
200         if (((N_x_int > -n_x-1) && (N_x_int < n_x+1)) && ((N_y_int > -      ↵
201             n_y-1) && (N_y_int < n_y+1))) {
202             //Anwenden der Symmetrie
203             N_x_int = N_x_int * int(ergebnis.flip_x) + n_x;
204             N_y_int = N_y_int * int(ergebnis.flip_y) + n_y;
205             if (ergebnis.trans > 0.) {
206                 int temp = N_x_int;
207                 N_x_int = N_y_int;
208                 N_y_int = temp;
209             }
210
211             r[i] = rcwa_distrib(wave_loop, theta_in, ergebnis.phi_in_sym,      ↵
212             N_y_int, N_x_int, DE) * MC;
213             //rcwa_distrib ist die bilineare Interpolationsfunktion mit      ↵
214             //Zugriff auf die RCWA-Daten DER bzw. DET
215         }
216     }
217     *pdf = 1.;
218     return r / AbsCosTheta(w0);
219 }
220

```

Eidesstattliche Versicherung (Affidavit)

Lange, Tobias

Name, Vorname
(Last name, first name)

209060

Matrikelnr.
(Enrollment number)

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit/Masterarbeit* mit dem folgenden Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present Bachelor's/Master's* thesis with the following title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution.

Titel der Bachelor-/Masterarbeit*:
(Title of the Bachelor's/ Master's* thesis):

Entwicklung und Verifikation von Algorithmen zur Vereinigung von Methoden der Wellenoptik und des Raytracing

*Nichtzutreffendes bitte streichen
(Please choose the appropriate)

Dortmund, 29.1.2021

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird gfls. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to €50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, section 63, subsection 5 of the North Rhine-Westphalia Higher Education Act (*Hochschulgesetz*).

The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:**

Dortmund, 29.1.2021

Ort, Datum
(Place, date)

Unterschrift
(Signature)

**Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.