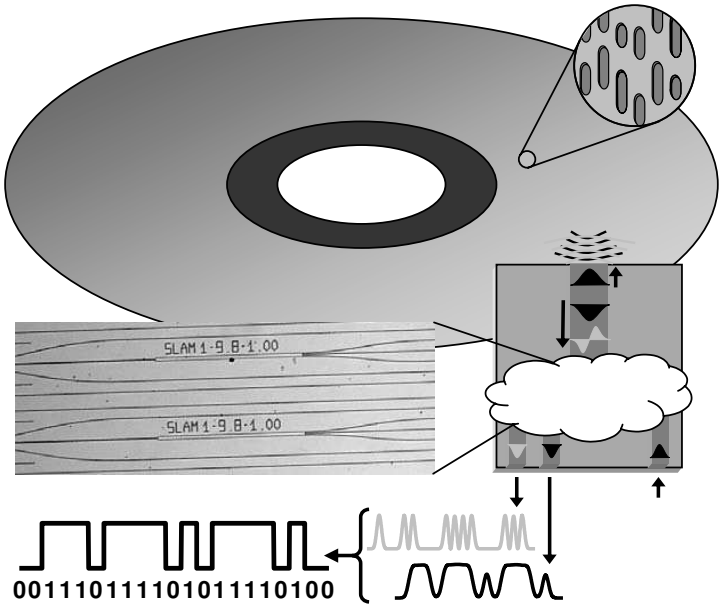
UNIVERSITEIT
GENT

Universiteit Gent
Faculteit Toegepaste Wetenschappen
Vakgroep Informatietechnologie

Performantie-analyse van een uitleessysteem voor optische schijven gebaseerd op een multimodale golfgeleider

Performance analysis of a multimode waveguide based optical disc readout system

Frederik Fransoo



INTEC

Proefschrift tot het behalen van de graad van
Doctor in de Toegepaste Wetenschappen
Academiejaar 2004-2005

Promotor:

Prof. dr. ir. R. Baets	Universiteit Gent, INTEC

Examencommissie:

Prof. dr. ir. R. Verhoeven, voorzitter	Universiteit Gent, INTEC
Prof. dr. ir. D. Van Thourhout, secretaris	Universiteit Gent, INTEC
Prof. dr. ir. J.J.M. Braat	Technische Universiteit Delft
Prof. dr. E. Stijns	Vrije Universiteit Brussel
Prof. dr. ir. P. De Visschere	Universiteit Gent, ELIS
Prof. dr. ir. P. Bienstman	Universiteit Gent, INTEC
Dr. ir. W. Bogaerts	Universiteit Gent, INTEC

Universiteit Gent
Faculteit Toegepaste Wetenschappen

Vakgroep Informatietechnologie (INTEC)
Sint-Pietersnieuwstraat 41
9000 Gent
BELGIUM

Tel.: +32-9-264.33.19
Fax: +32-9-264.35.93

Part of this work was carried out in the framework of the European project SLAM (IST-2000-26479).

voor mijn ouders

voor Tilly

Contents - Inhoudsopgave

Nederlandstalige tekst

Voorwoord - Preface

Dit proefschrift is tot stand gekomen door de inspanningen en steun van heel wat mensen. Ik wil dit voorwoord dan ook gebruiken om hen te bedanken. In de eerste plaats zou ik mijn promotor Prof. Roel Baets willen bedanken om mij de mogelijkheid te bieden mijn onderzoekswerk te verrichten. Zijn suggesties en ideeën waren altijd erg waardevol en hebben mij vaak vooruitgeholpen als ik ergens vast zat in mijn onderzoek. Verder wil ik ook Prof. Dries Van Thourhout bedanken. Van hem komt oorspronkelijk het idee om een multimodale golfgeleider te gebruiken voor het uitlezen van een optische schijf. Ook was hij in de latere jaren van mijn doctoraat een grote steun. Zijn opmerkingen en suggesties openden vaak een deur naar een nieuw onderzoekspad.

Dit alles was ook niet mogelijk zonder de steun van andere mensen in de vakgroep. Hierbij denk ik niet alleen aan vakgroepvoorzitter Prof. Lagasse, maar ook aan de mensen van de financiële dienst, die altijd klaar stonden voor hulp bij het indienen van onkostennota's, Kristien die zorgde dat het computerpark bleef functioneren, Hendrik en Luc die bereidwillig raad en hulp boden bij het opbouwen van de meetopstelling, en ten slotte natuurlijk de mensen van de technologie, Yu, Peter, Liesbeth, en Steven die ervoor zorgden dat mijn ontwerpen, werkelijk gefabriceerde componenten werden.

Ook kijk ik graag terug naar het jaar waarin ik de thesis van Annelies en Wouter heb mogen begeleiden. Niet alleen heb ik een aantal van hun resultaten in dit proefschrift kunnen verwerken, maar bovendien heb ik ook zelf veel bijgeleerd uit al hun vragen en opmerkingen die ze me stelden.

Verder ook bedankt aan al mijn collega's, in het bijzonder die van 'de 39'. Ons gebouw was misschien niet wat je zou noemen een hypermoderne kantoorfaciliteit, maar compenseerde dit overvloedig met een eigen apart karakter. De sfeer onder de collega's en zeker de legen-

darische koffiepauzes, waarin we dagelijks 'vergaderden' over allerlei politiek correcte en politiek minder correcte onderwerpen, zullen mij nog eeuwig bijblijven. Er zijn te veel namen om op te noemen, maar toch wil ik graag de collega's die met mij jarenlang ruimte 2.5 hebben gedeeld, Bart, Wim B., Peter, Wim W. en Hannes, vermelden. En misschien, als ik er toch nog iemand wil uitpikken, Lieven, die met zijn laaiend enthousiasme ervoor zorgde dat we het woord 'saaie werkplek' uit onze woordenschat mochten schrappen.

Niet alleen in Gent, maar ook in Delft heb ik metingen verricht om het concept van de multimodale golfgeleider te testen, ook heb ik er erg veel geleerd uit de vele gesprekken over de meer theoretische aspecten. Hartelijk dank daarom aan Haifeng, Jacco, Jos, Sami en Silvania, die mij elk op hun manier hebben bijgestaan bij mijn regelmatige bezoeken aan de TUDelft.

Of course, I also thank the other partners from the SLAM project. The sometimes long discussions on the regular meetings were an excellent opportunity to learn form each other's experience. From all the people who helped the project to a successful end, I would like to thank specifically Prof. Joseph Braat, who coordinated the whole project.

Tenslotte, last but not least, ook nog duizendmaal dank aan het thuisfront. Aan mijn ouders, die mij jarenlang gesteund hebben op alle mogelijke vlakken, zowel tijdens mijn studies als nu voor het voltooien van dit doctoraat. En ook aan mijn allerliefste schat, die altijd voor me klaarstond en me het laatste halfjaar nog eens extra heeft gesteund om dit werk tot een goed einde te brengen.

Frederik Fransoo

Gent, 15 december 2004

Nederlandse samenvatting - Dutch summary

1. Inleiding

1.1 Gegevensopslag

Gegevensopslag is al belangrijk sinds de uitvinding van het schrift, van spijkerschrift op kleitabletten tot stapels boeken gedrukt met de modernste drukpersen. Met de opkomst van de computer zijn de noden voor snelle opslag van enorme hoeveelheden gegevens echter sterk veranderd.

Vaak wordt een onderscheid gemaakt tussen geheugen en 'echte' gegevensopslag. Geheugen is een manier van gegevensopslag die vaak snel maar volatiel is. De 'echte' gegevensopslag is langzamer, krijgt informatie uit het geheugen en moet die voor langere tijd bewaren.

Deze laatste kan weer verder worden onderverdeeld in categorieën. Zo zijn er herschrijfbare technieken en opslagmethodes die eens geschreven, nooit meer veranderd kunnen worden. Dit laatste kan belangrijk zijn vanwege veiligheidsredenen. Een andere onderverdeling is verplaatsbaarheid. Zo is het veel gemakkelijker een optische schijf uit te wisselen met een andere persoon, dan de interne harde schijf te moeten uitbreken en herinstalleren. Er bestaan verschillende methodes van verplaatsbare gegevensopslagsystemen. Elk hebben ze hun eigen sterkte- en zwaktepunten. Belangrijk zijn: de totale gegevenscapaciteit, de snelheid waarmee die kan worden uitgelezen, de toegangssnelheid, de stabiliteit op lange termijn en ten slotte de prijs.

Een eerste methode is de ponskaart. Deze bestaat uit een papieren rol met kleine gaatjes in. Hoewel deze methode vroeger vaak werd gebruikt, is ze nu in onbruik geraakt. De magneetband, bestaat een lange strook met daarop kleine ferromagnetische deeltjes. Het voornaam-

ste nadeel van deze techniek is de lage toegangssnelheid, aangezien de band sequentieel moet worden uitgelezen. De verplaatsbare magneetschijf, is vooral bekend onder de vorm van de diskette, maar er bestaan ondertussen ook al varianten met een erg hoge opslagcapaciteit en uitleessnelheid. Flitsgeheugen[1] is een andere techniek en slaat bits op door elektronen op te sluiten in een transistor. Deze techniek heeft vooral de prijs als grote nadeel.

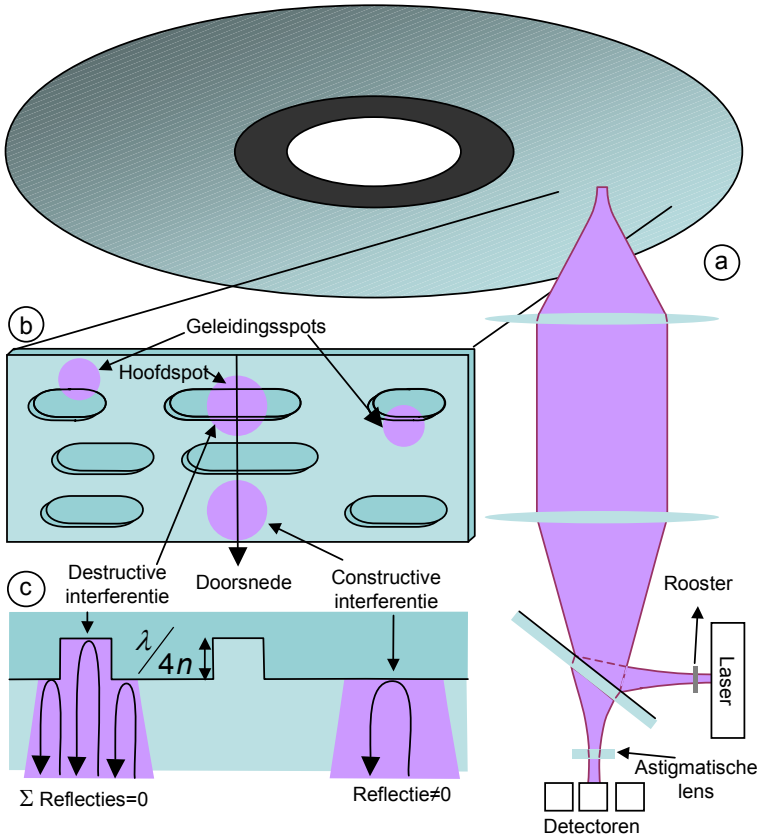
Tenslotte is er optische data opslag, zoals de CD[2]. Deze gebruikt een lichtbundel om bits te schrijven en uit te lezen. De belangrijkste voordelen ten opzichte van de directe concurrenten zijn: de lage prijs van de optische schijven zelf en de stabiliteit op langere termijn. Ook in de toekomst zullen optische schijven een belangrijke speler blijven in de distributie en archivering van tekst tot video.

1.2 Gegevensopslag op optische schijven

Hoewel het eerste voorbeeld van gegevensopslag op optische schijven al dateert van 1972, werd de techniek pas echt populair met de uitvinding van de CD in het begin van de jaren tachtig en de DVD[3] halfweg de jaren negentig. Welk formaat de derde generatie zal vormen is op dit moment nog onbeslist. De belangrijkste kandidaten zijn HD-DVD[4] en BD[5]. De eerstgenoemde heeft betere compatibiliteit met de DVD, de laatstgenoemde heeft een hogere gegevenscapaciteit.

De fysische principes achter al deze optische schijven zijn gelijk. Elke nieuwe generatie heeft echter zoals aangeduid in tabel 1, telkens een hogere gegevenscapaciteit en een hogere uitleessnelheid. Fig. 1 toont de draaiende optische schijf en de uitleeskop. Een laser belicht een door de diffractiewetten beperkte lichtvlek met diameter λ/NA op de schijf. Via een halfdoorlatende spiegel kan het gereflecteerde licht opgevangen worden op een detector. Fig. 1b en Fig. 1c tonen respectievelijk een boven- en een zijaanzicht van een detail van de schijf. Doordat de putjes op de schijf $\lambda/(NA)$ diep zijn, krijgt men destructieve interferentie als de lichtvlek boven een putje schijnt, en constructieve interferentie als er geen putje is. Zoals we in hoofdstuk 3 uitvoerig zullen beschrijven kan een diffractie-gelimiteerd systeem geen details onderscheiden met een spatiale periode groter dan $\lambda/(2NA)$.

[1] In het Engels: flash memory
[2] In het Engels: compact disc
[3] In het Engels: digital versatile disc
[4] In het Engels: high definition DVD
[5] In het Engels: blu-ray disc



Figuur 1: A: Een laser belicht een lichtvlek op een draaiende optische schijf. Door een halfdoorlatende spiegel wordt het gereflecteerde licht door een detector opgevangen. B: De putjes op de schijf zijn $\lambda/(4n)$ diep. C: De lichtvlek bedekt de helft van de putjes. Afhankelijk van de bitsequentie genereert dit destructieve en constructieve interferentie aan de detector.

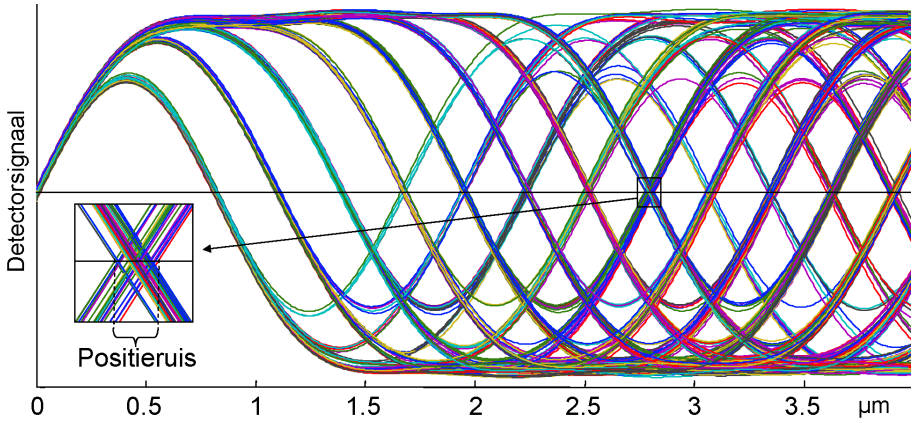
De reflectie van de lichtvlek op de schijf wordt niet alleen opgevangen om de bits te detecteren, maar wordt ook gebruikt om de lichtvlek nauwkeurig op het bitspoor te houden en om het brandpunt van de objectieflens op het schijfoppervlak te houden. Op de huidige optische schijven zijn de bits gecodeerd door middel van begrensde-looplengtecodes[6]. Dit wil zeggen dat de putjes op de schijf enkel voorkomen met specifieke lengtes. De kortst mogelijke putlengte is dan gelijk aan de helft van de laagste spatiale periode welke groter moet zijn dan $\lambda/(2NA)$. Tabel 1 geeft aan dat de marge hierop bij de CD een stuk

[6]In het Engels: run length limited codes (RLL)

	CD	DVD	BD
Golflengte (λ)	780 nm	650 nm	405 nm
Numerieke Apertuur (NA)	0,45	0,60	0,85
Afstand tussen twee sporen	1,6 µm	0,74 µm	0,32 µm
Kortste lengte van een putje	3T=0,84 µm	3T=0,40 µm	2T=160/149/138 nm
Netto gegevenscapaciteit per laag	650 MByte	4, 7 GByte	23,3/25/27 GByte
Netto uitleessnelheid ($1\times$)	1,47 Mbps	11 Mbps	36 Mbps

Tabel 1: Drie generaties optische schijven: de compact disc (CD), de digital versatile disc (DVD) en de blu-ray disc (BD).

groter is dan bij de BD. De overgang van een 'nul' naar een 'één' op de schijf wordt teruggevonden door te kijken wanneer het detectorsignaal een bepaalde drempelwaarde doorkruist. De fout op het meten van dit tijdstip wordt positieruis genoemd, en kan visueel voorgesteld worden door de signalen van een hele reeks bitpatronen bovenop elkaar te tekenen, zoals in Fig. 2.



Figuur 2: Oog diagram van het detector signaal in een CD. De detector signalen van alle mogelijke bitsequenties zijn boven elkaar getekend. De variatie in de x-coördinaat van het punt waar de signalen een zekere drempelwaarde overschrijdt geeft de 'positieruis'.

Hoe de technologie van optische gegevensopslag zal evolueren na de BD is nog niet duidelijk. Vanuit historisch standpunt bekeken lijkt het echter onwaarschijnlijk dat de vraag naar een hogere gegevenscapaciteit en een hogere uitleessnelheid nu zou stoppen. Bovendien is er ook een tendens naar miniaturisatie, dit wil zeggen dat dezelfde gegevenscapaciteit op een nog kleinere schijf zou worden opgeslagen.

Daarom worden verschillende technologieën onderzocht om de gegevenscapaciteit per oppervlakte te vergroten. Een eerste mogelijkheid is de numerieke apertuur te verhogen door te werken in het nabije veld, of door een immersielens te gebruiken. Een andere techniek is gebruik te maken van niet lineaire eigenschappen van de schijf zodat enkel het centrale gedeelte van de lichtvlek een signaal geeft op de detector. Een derde techniek is volumetrische gegevensopslag, door ofwel met meerdere lagen te werken of ofwel via holografische technieken. Een andere methode bestaat erin meerdere grijsniveaus te gebruiken, en hierdoor meerdere bits op te slaan op dezelfde ruimte. Verder is het ook mogelijk om door een betere signaalverwerking op de verkregen detectorsignalen, de invloed van nabijgelegen bits te compenseren. Hierdoor kunnen de bits dichter bij elkaar geplaatst worden.

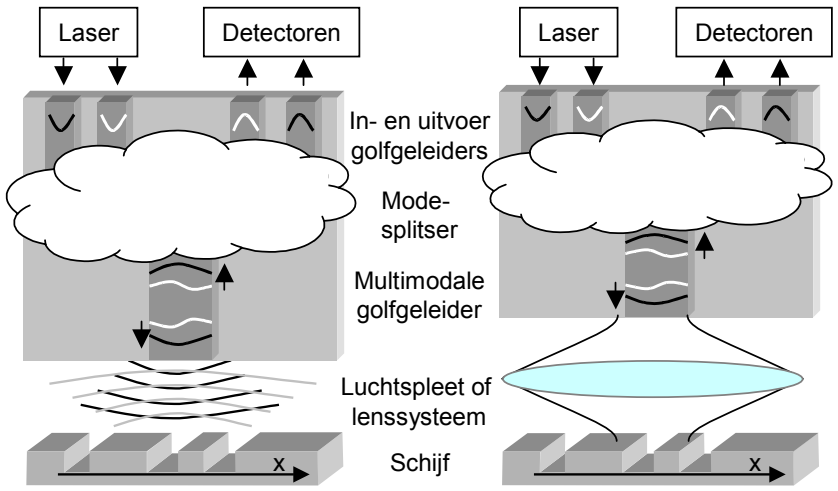
Nog een andere methode is om informatie te halen uit zowel de amplitude als uit de fase van het licht, dat in reflectie van de schijf wordt opgevangen. In deze thesis stellen we een methode voor om dit te doen met een multimodale golfgeleider.

1.3 De golfgeleider-detectiemethode

In dit doctoraat onderzochten we hoe de gegevensdichtheid op een optische schijf kan verbeterd worden door middel van een multimodale golfgeleider als uitleeskop. Fig. 3 geeft een schematisch overzicht van het voorgestelde golfgeleider-detectiesysteem. Een laser belicht via één of meerdere modes een kleine lichtvlek op de schijf. Het gereflecteerde licht wordt door dezelfde golfgeleidermodes terug opgepikt en via een fotonisch geïntegreerd circuit naar de detectoren geleid. De filosofie achter deze methode bestaat er in om informatie te halen uit zowel de amplitude als uit de fase van het gereflecteerde licht. Dit in tegenstelling tot een gewone detector, die enkel het vermogen opmeet. Een tweede motief is miniaturisatie; in principe zouden alle optische functies op een enkel optisch schijfje kunnen worden geïntegreerd, wat de robuustheid en kostprijs ten goede zou komen. Om de signalen van de verschillende modes te combineren werd een eenvoudige parameterextractiemethode ontworpen. Simulaties toonden aan dat door meerdere modes te gebruiken het aantal bitfouten kan worden gereduceerd.

1.4 Overzicht van dit werk

Dit doctoraat is onderverdeeld in verschillende hoofdstukken. In deze introductie staat een algemeen overzicht van gegevensopslag en wor-



Figuur 3: Schematisch overzicht van het golfgeleider-detectiesysteem. De figuur toont de versie in het nabije-veld aan de linkerkant en de versie in het verre veld aan de rechterkant. In beide gevallen wordt de schijf belicht met een of meerdere golfgeleidermodes. Dezelfde golfgeleidermodes pikken het gereflecteerde veld terug op. Een fotonisch geïntegreerd circuit verbindt de individuele modes met de laser en detectoren via in- en uitvoer golfgeleiders.

den de werkingsprincipes van de voorgestelde golfgeleider-detectiemethode uitgelegd. In hoofdstuk 2 zullen de principes van optische resolutie worden beschreven. Een uitvoerige analyse van de scannende microscoop komt in hoofdstuk 3 aan bod. In hoofdstuk 4 wordt aan de hand van een eenvoudige parameter-extractiemethode uitgelegd hoe het aantal bitfouten kan worden verminderd door de signalen van twee verschillende modes te combineren. De werking en het ontwerp van een fotonisch geïntegreerd circuit worden toegelicht in hoofdstuk 5. De meetresultaten zijn in hoofdstuk 6 gebundeld. In het laatste hoofdstuk maken we tenslotte een aantal conclusies.

1.5 Publicaties

Het werk uitgevoerd in het kader van dit proefschrift leidde tot een aantal publicaties:

1. F.W.J. Fransoo, D. Van Thourhout, and R. Baets, Performance analysis of a multi-mode waveguide based optical disc readout system, Applied Optics, 34(17): 3480–3488, Jun. 2004.

2. S. Musa, S.F. Pereira, J.J.M. Braat, F.W.J. Fransoo, D. Van Thourhout, and R. Baets, Sub-wavelength detection with waveguides in optical recording, *SPIE proceedings*, 5380: 723–728, Sep. 2004.

3. S. Musa, S.F. Pereira, J.J.M. Braat, F.W.J. Fransoo, D. Van Thourhout, and R. Baets, Sub-wavelength detection with waveguides in optical recording, *Optical Data Storage Topical Meeting 2004*, Monterey , United States: 156–158, Apr. 2004.

4. F.W.J. Fransoo, D. Van Thourhout, L. Van Landschoot, S. Verstuyft, and R. Baets, A multimodal waveguide for enhanced performance in optical disc read-out, *EOS Topical Meeting in Advanced Imaging Techniques 2003*, Delft, The Netherlands: 115–117, Oct. 2003.

5. F.W.J. Fransoo, D. Van Thourhout, L. Van Landschoot, A. Verbiest, W.F.H. Van Parys, P. Van Daele, and R. Baets, A method for detecting sub-wavelength features by means of a multimode waveguide and a mode splitting photonic IC, *LEOS Annual 2003*, Glasgow, United Kingdom: 750–751, Nov. 2002.

6. F.W.J. Fransoo, D. Van Thourhout, L. Van Landschoot, A. Verbiest, W.F.H. Van Parys, P. Van Daele, and R. Baets, Using a multimodal waveguide for enhanced resolution in optical data storage, *FTW-symposium 2002*, Gent, Belgium: p 25, 2002.

7. F.W.J. Fransoo, P. Bienstman, and R. Baets, Photonic IC for optical detection with sub wavelength resolution, *IEEE/LEOS Benelux Chapter 2001*, Brussels, Belgium, pp. 205-208, Dec. 2001.

2. Optische resolutie

2.1 Inleiding

Traditioneel wordt de resolutie van een optisch systeem gedefinieerd als de kleinste afstand tussen twee punten, waarbij de afbeeldingen van die punten door een optische systeem nog onderscheidbaar zijn. Het meest bekende criterium is het Rayleigh-criterium:

> *Twee puntbronnen zijn net geresolveerd als het centrale maximum van de intensiteit van het diffractiepatroon van één van beide puntbronnen samenvalt met het eerste minimum van het diffractiepatroon van het andere punt.*

Voor twee totaal incoherente puntbronnen en een circulaire aberratievrije lens met numeriek apertuur NA, resulteert dit in een 'resolutie' van $0{,}601NA/\lambda$. Deze definitie is heel handig in de sterrenkunde, waar de sterren als volkomen incoherente bronnen beschouwd kunnen worden. Ze heeft in andere toepassingen echter maar een beperkt nut aangezien ze enkel naar twee geïsoleerde punten kijkt en geen rekening houdt met coherentie of ruis.

2.2 Puntspreidingsfunctie en optische overdrachtsfunctie

Een andere manier om te meten hoe goed een optisch systeem kleine details kan onderscheiden, is gebruik maken van de puntspreidingsfunctie (PSF[7]) en de optische overdrachtsfunctie (OTF[8]). In het vervolg van dit hoofdstuk maken we hierbij twee veronderstellingen: lineariteit en verschuivingsinvariantie.

Lineariteit en verschuivingsinvariantie Beschouw een optisch systeem dat een puntbron met amplitude A op positie (x_0, y_0) afbeeldt op een lichtvlek in een ander vlak met coördinaten x_1 en x_2. Bij een lineair systeem is de amplitude van deze lichtvlek evenredig met A. Als het systeem ook nog eens verschuivingsinvariant is, kan deze worden geschreven als een functie van het verschil tussen de coördinaten in het beeldvlak en de coördinaten van de puntbron: $K(x_1 - x_0, y_1 - y_0)$.

[7] In het Engels: point spread function
[8] In het Engels: optical transfer function

Puntspreidingsfunctie Bij een lineair systeem kan men de afbeelding van twee punten vinden door de afzonderlijke afbeeldingen van elk punt samen te voegen. Hoe dit samenvoegen moet gebeuren, hangt echter af van de onderlinge coherentie. Bij coherent licht mag men de complexe amplitudes optellen en bij incoherent licht mag men de intensiteiten optellen. Als men de amplitude van het licht in het voorwerpsvlak en het beeldvlak definieert als respectievelijk $\psi_1(x_1, y_1)$ en $\psi_1(x_1, y_1)$, krijgt men voor coherent en incoherent licht:

$$\psi_1(x_1, y_1) = \iint psf_{coh}(x_1 - x_0, y_1 - y_0)\, \psi_0(x_0, y_0)\, \mathrm{d}x_0\, \mathrm{d}y_0 \qquad (1)$$
$$\text{met } psf_{coh}(x, y) = K(x, y)$$
$$|\psi_1(x_1, y_1)|^2 = \iint psf_{incoh}(x_1 - x_0, y_1 - y_0)\, |\psi_0(x_0, y_0)|^2\, \mathrm{d}x_0\, \mathrm{d}y_0 \qquad (2)$$
$$\text{met } psf_{incoh}(x, y) = |K(x, y)|^2 .$$

Dit zijn eigenlijk twee uiterste gevallen voor het algemene geval van partieel coherent licht. Hiervoor moeten we eerst de zogenaamde zichtbaardheidsfunctie definiëren. Als $\psi_0(t)$ en $\psi_0'(t)$ het complexe optische veld voorstellen in twee punten, dan is de zichtbaardheidsfunctie $J(x_0, y_0; x_0', y_0')$ gelijk aan het gemiddelde over de tijd van het product van $\psi_0(t)$ en $\psi_0'(t)$. Bij partieel coherent licht krijgt men dan:

$$J_1(x_1, y_1; x_1', y_1') = \iiiint PSF(x_1 - x_0, y_1 - y_0; x_1' - x_0', y_1' - y_0') \times J_0(x_0, y_0; x_0', y_0')\, \mathrm{d}x_0\, \mathrm{d}y_0\, \mathrm{d}x_0'\, \mathrm{d}y_0' \qquad (3)$$
$$\text{met } PSF(x, y; x', y') = K(x, y)\, K^*(x', y') .$$

In de meeste optische scannende microscopen werkt men met coherent licht. Toch zullen we in het volgende hoofdstuk terugkomen op de vergelijkingen voor partieel coherent licht. De reden is dat de reflectie aan de schijf lineair is volgens de amplitude, en het elektrisch signaal van de detector lineair volgens de intensiteit van het licht.

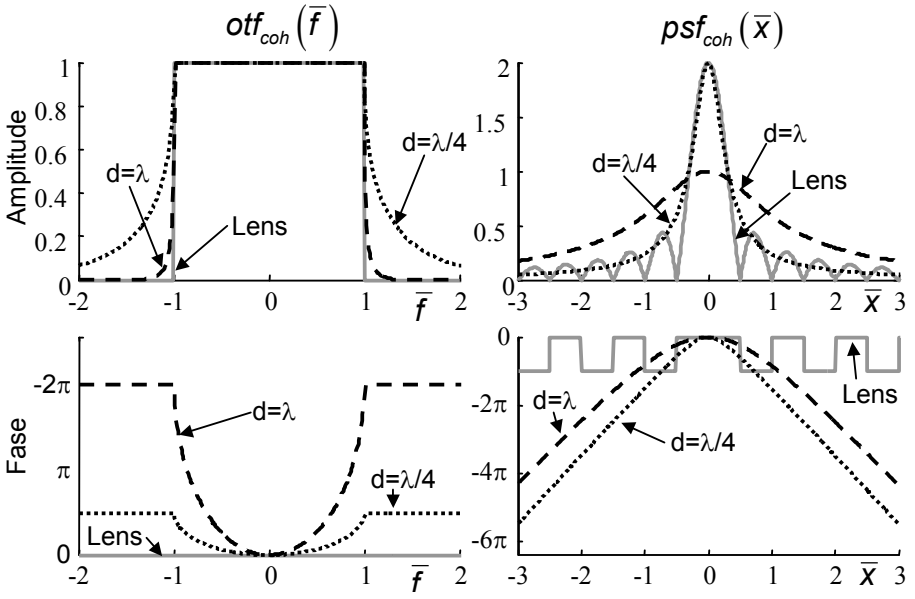
Optische overdrachtsfunctie De optische overdrachtsfunctie (OTF), geeft het frequentie-antwoord van een optische systeem. Voor coherent en incoherent licht is dit de Fourier transformatie van de puntspreidingsfunctie. Voor partieel coherent licht gebruiken we een iets

verschillende operator aangegeven met een tilde: $OTF = \widetilde{PSF}$ of voluit

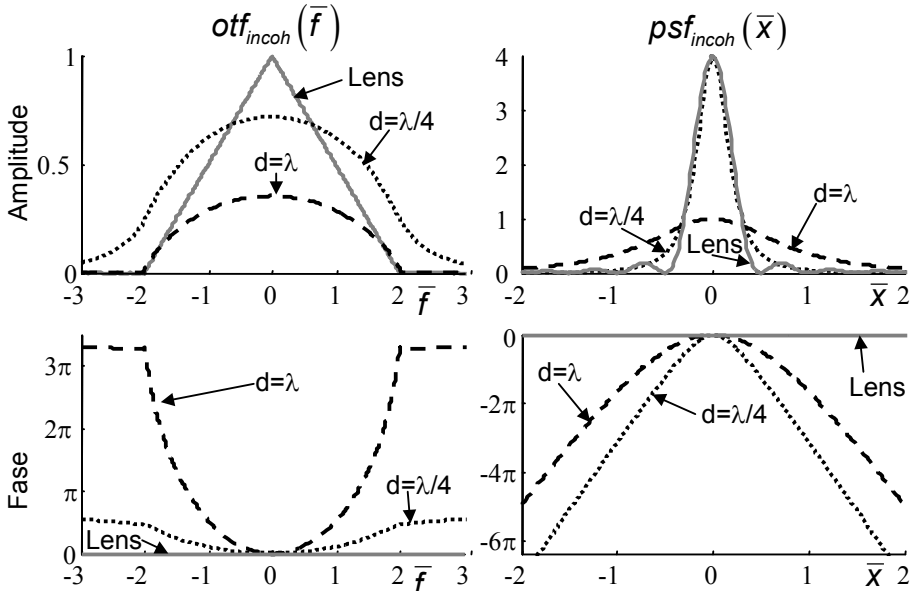
$$OTF(f,g;f',g') = \iint PSF(x,y;x',y') \times e^{-2\pi j(fx+gy+f'x'+g'y')}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y'. \quad (4)$$

2.3 Voorbeelden

Voor een aberratievrije lens of voor transmissie in de vrije ruimte kan men analytische uitdrukkingen berekenen voor de PSF en de OTF. Fig. 4 en Fig. 5 tonen respectievelijk de amplitude en de fase van de puntspreidingsfunctie en de optische overdrachtsfunctie voor coherent en incoherent licht. De grafieken tonen dat hoewel afbeelding in het nabije veld op een afstand van een golflengte een breder spectrum aan spatiale frequenties doorlaat dan in het verre veld, de puntspreidingsfunctie toch breder is. De reden is dat het faseverband van de verschillende frequenties door elkaar gehaald wordt.



Figuur 4: De amplitude en de fase van de coherente otf en psf langs de X-as. In volle grijze lijn voor een rechthoekige lens, in zwarte stippellijn voor het nabije veld met $z = \lambda/4$, in zwarte gebroken lijn voor het nabije veld met $z = \lambda$. Op de horizontale as staan genormeerde grootheden: $\overline{x} = xNA/\lambda$ en $\overline{f} = f\lambda/NA$ voor het lens systeem. $\overline{x} = x/\lambda$ en $\overline{f} = f\lambda$ voor in het nabije veld.

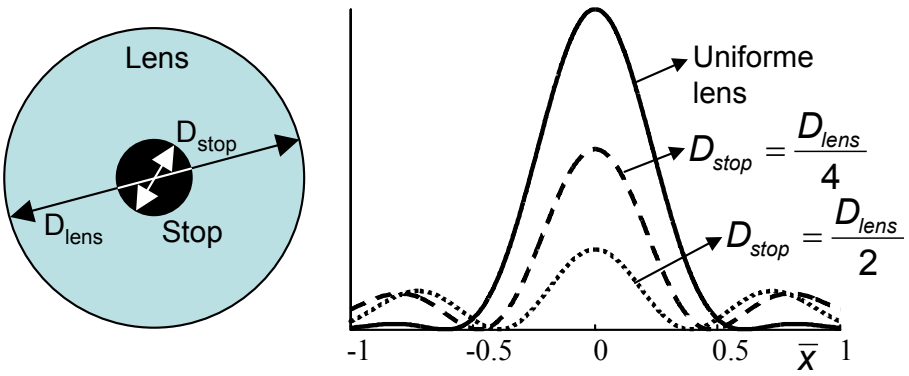


Figuur 5: De amplitude en de fase van de incoherente otf en psf langs de X-as. In volle grijze lijn voor een rechthoekige lens, in zwarte stippellijn voor het nabije veld met $z = \lambda/4$, in zwarte gebroken lijn voor het nabije veld met $z = \lambda$. Op de horizontale as staan genormeerde grootheden: $\overline{x} = xNA/\lambda$ en $\overline{f} = f\lambda/NA$ voor het lens systeem. $\overline{x} = x/\lambda$ en $\overline{f} = f\lambda$ voor in het nabije veld.

2.4 Resolutieversterking

In principe zijn er twee methodes om de resolutie van een lenssysteem te verbeteren. De eerste methode is door filters te plaatsen of door verschillende bundels te laten interfereren. Zoals Fig. 6 aangeeft kan dit bijvoorbeeld door het centrale gedeelte van de lens af te dekken. De breedte van de puntspreidingsfunctie daalt, en hiermee ook de resolutie zoals bepaald door het Rayleigh-criterium. Het nadeel is dat een gedeelte van het licht verloren gaat en de zijlobben relatief sterker worden. Over het algemeen kan men bepaalde aspecten van de resolutie verbeteren, maar gaat dit steeds ten koste van andere aspecten.

Een tweede manier bestaat erin gebruik te maken van de voorkennis van de puntspreidingsfunctie of bepaalde parameters van het optische veld in het voorwerpsvlak. Het is mogelijk hiermee details kleiner dan de diffractie-limiet te onderscheiden, maar enkel indien ruis klein is ten opzichte van het signaal.

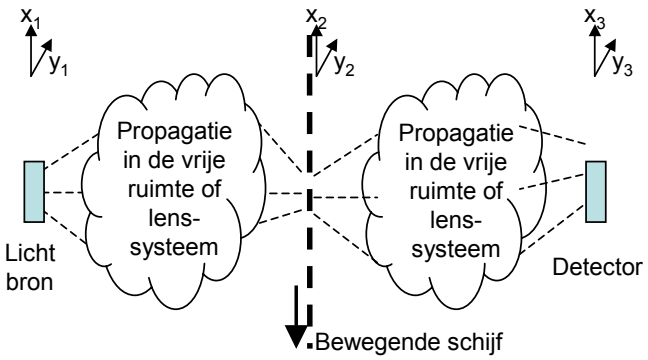


Figuur 6: Door het centrale gedeelde van een lens te bedekken, kan de puntspreidingsfunctie gewijzigd worden. De volle lijn geeft de puntspreidingsfunctie voor een lens zonder centrale filter. De gebroken lijn en de stippellijn tonen de gewijzigde puntspreidingsfunctie.

3. Analyse van de optische scannende microscoop

3.1 De optische scannende microscoop

Een scannende optische microscoop bestaat uit een lichtbron, een object, en een detector. De lichtbron belicht een (klein) gedeelte van het voorwerp waar het gereflecteerd wordt. Een detector pikt (een gedeelte van) deze reflectie weer op en zet dit om in een elektrisch signaal. Door het voorwerp te bewegen (te scannen), kan men sequentieel het volledige voorwerp bestrijken. Voorbeelden van een optische scannende microscoop zijn het DVD detectiesysteem, de confocale microscoop, en de in dit doctoraat voorgestelde golfgeleider-detectiemethode. Deze drie methodes zullen in dit hoofdstuk uitvoerig besproken en vergeleken worden.



Figuur 7: De lichtweg in een optische scanning microscoop. Een lichtbron wordt door de vrije ruimte of via een lens systeem naar de schijf geleid waar het een kleine lichtvlek vormt. Het gereflecteerde licht wordt terug opgevangen door een detector. Voor de duidelijkheid werd bovenstaande figuur als een transmissiesysteem getekend. Door de schijf te verschuiven, passeren verschillende bitpatronen in de lichtvlek. Het signaal van de detector verandert overeenkomstig met deze bitpatronen.

Voor een eenvoudige beschrijving van de optische eigenschappen, hebben we een aantal benaderingen ingevoerd. Vooreerst veronderstellen we dat er geen meervoudige reflecties zijn, er is dus enkel een eenmalige reflectie aan het voorwerp. Twee andere benaderingen zijn met elkaar verbonden: scalaire diffractie en polarisatie. Voor een rigoureuze beschrijving van de lenzen, de schijf en de golfgeleiders moet het licht beschreven worden als een vector. Wij benaderen het licht door een scalaire waarde, bovendien veronderstellen we dat de polarisatie

gedurende de lichtweg niet wordt gewijzigd. Een laatste benadering is dat we de reflectie van het voorwerp beschouwen als een complexe amplitude transmissie functie: $r(x, y)$. Onder deze voorwaarden kan het elektrisch signaal uit de detector geschreven worden als:

$$I(x_s) = \iiiint R\big(x - x_s, y - y_s; x' - x_s, y' - y_s\big) \times PSF\big(x, y; x', y'\big)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y' \tag{5}$$

$$= \iiiint \widetilde{R}\big(f, g; f', g'\big) \exp\big\{-2\pi j\left[\left(f - f'\right)x_s + \left(g - g'\right)y_s\right]\big\} \times OTF\big(-f, -g; -f', -g'\big)\,\mathrm{d}f\,\mathrm{d}g\,\mathrm{d}f'\,\mathrm{d}g'. \tag{6}$$

In deze vergelijkingen is $R(x, y; x', y') = r(x, y)\, r^*(x', y')$, de partieel coherente transmissiefunctie van de schijf. x_s en y_s duiden de translatie van het voorwerp aan. De tilde geeft een soort Fourier-transform aan en is beschreven in appendix A.

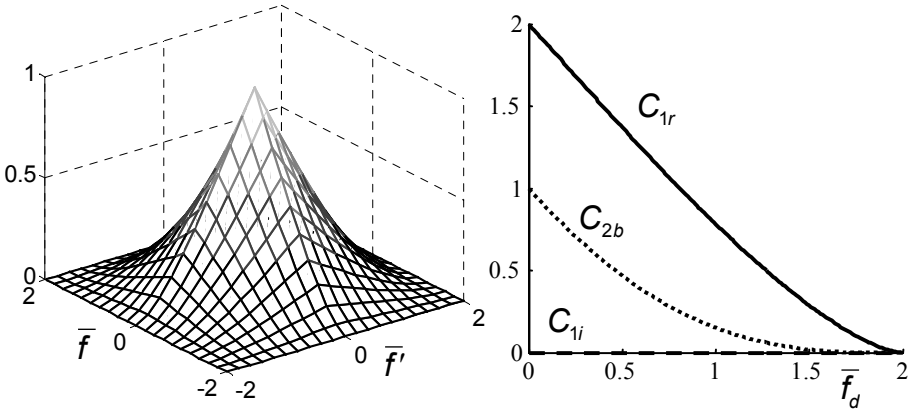
Het is niet zo eenvoudig om de partieel coherente overdrachtsfunctie $OTF(f, g; f', g')$ te interpreteren. Een beschrijving die toch enig inzicht geeft, is kijken naar het signaal van een optisch systeem voor een schijf bestaande uit een één-dimensionaal rooster bovenop een vaste reflectiecoëfficiënt.

Uit de wiskundige beschrijving volgt dan dat het detectorsignaal uit vier componenten zal bestaan. Een eerste is een constant signaal dat overeenkomt met de gemiddelde reflectie van de schijf. De twee volgende componenten, $\mathcal{C}_{1r}$ en $\mathcal{C}_{1i}$, geven elk een signaal met dezelfde frequentie als het rooster op de schijf. $\mathcal{C}_{1r}$ is evenredig met het reëel deel van de relatieve reflectie van het rooster, en $\mathcal{C}_{1i}$ is evenredig met het imaginaire deel daarvan. Tenslotte is er nog een vierde component, $\mathcal{C}_{2b}$, die een signaal geeft met de dubbele frequentie van het rooster en bovendien evenredig is met het kwadraat van de sterkte van het rooster. Bij zwakke roosters is deze component dus te verwaarlozen. Deze component geeft ook een idee in welke mate de signalen van de roosters met een verschillende periode zullen worden gemengd.

3.2 Berekening van de PSF en OTF

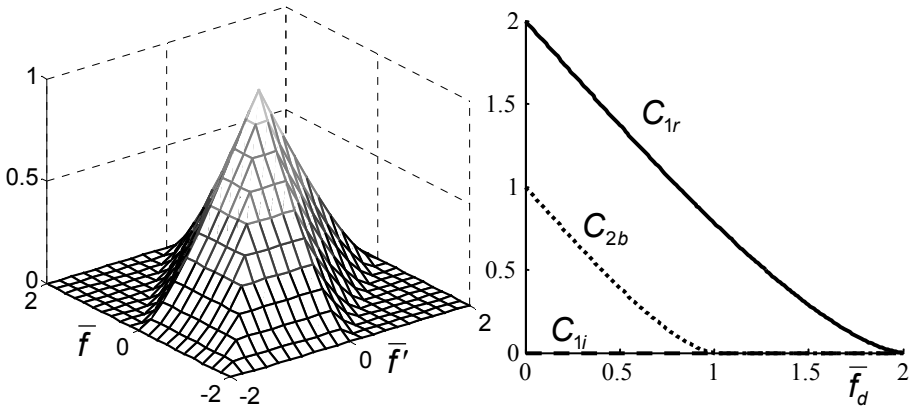
Door de optische lichtweg van belichtingsbron tot detector te volgen, kan de partieel coherente optische overdrachtsfunctie berekend worden. Als we veronderstellen dat de lenzen aberratie-vrij en circulair zijn en de golfgeleidermodes heel klein, dan blijken de golfgeleiderdetectiemethode en de confocale microscoop identiek. Fig. 8 toont de

optische overdrachtsfunctie en de doorsneden die de invloed op een één-dimensionaal rooster weergeven.



Figuur 8: De genormaliseerde partieel coherente OTF van het golfgeleider-detectiesysteem langs de X-as voor circulaire aberratievrije lenzen en heel kleine golfgeleider modes.

De eigenschappen van het DVD-uitleessysteem zijn echter wel verschillend, zoals weergegeven op Fig. 9: de component van het detectorsignaal met dezelfde frequentie als het rooster is weliswaar dezelfde, maar de tweede orde component, $\mathcal{C}_{2b}$ laat de hogere frequenties niet door.



Figuur 9: De genormaliseerde partieel coherente OTF van het uitgebreide-detectorsysteem langs de X-as voor circulaire aberratievrije lenzen en een heel kleine belichtingsbron.

3.3 Uitbreidingen en benaderingen

Om analytische uitdrukking te vinden voor de optische overdrachtsfunctie werden een aantal verdere benaderingen gedaan. De eerste is de dimensies met één reduceren. Door de gepaste benaderingen te maken kunnen we alle vergelijkingen onafhankelijk maken van de y. Verder hebben we ook de golfgeleidermodes benaderd door analytische functies. Zo kan de nulde orde mode goed benaderd worden door een Gaussische functie:

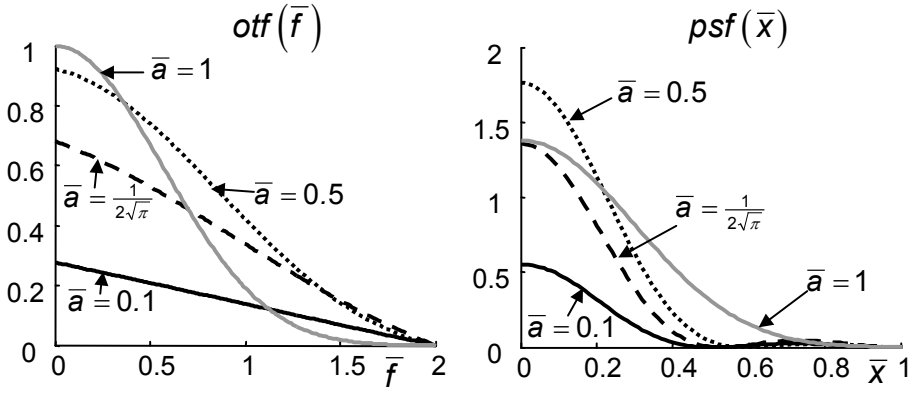
$$m^{0,air}(x) \approx \sqrt{\frac{\sqrt{2}}{a}}\,\mathcal{G}\Big(\frac{x}{a}\Big) \qquad \widehat{m^{0,air}}(f) \approx \sqrt{a\sqrt{2}}\,\mathcal{G}(af)\,. \tag{7}$$

De eerste en de tweede orde kunnen op een gelijkaardige manier door analytische functies worden benaderd. Om de notatie wat te vereenvoudigen zullen we in de rest van dit hoofdstuk werken met genormaliseerde parameters: $\overline{x} = xNA/\lambda$, $\overline{a} = aNA/\lambda$, en $\overline{f} = f\lambda/NA$.

In het kader van een samenwerking met de Technische Universiteit Delft werden ook twee uitbreidingen op de golfgeleider-detectiemethode onderzocht. De eerste uitbreiding is in plaats van één enkele multimodale golfgeleider, twee afzonderlijke monomodale golfgeleiders te gebruiken, waarvan de onderlinge fase van de modes door een PIC gecontroleerd kan worden. Een andere uitbreiding bestaat erin de golfgeleider lichtjes schuin te plaatsen onder een hoek θ. Hierdoor ontstaat er een soort asymmetrie die de eigenschappen van de golfgeleiderdetectiemethode grondig wijzigt.

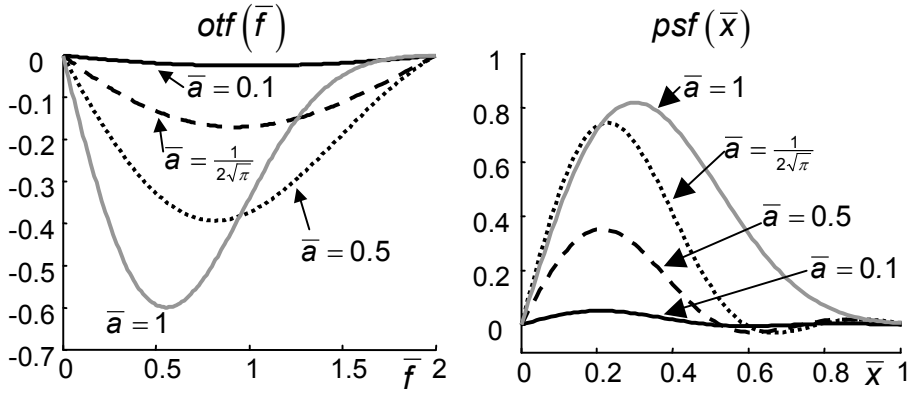
3.4 Golfgeleider-detectiemethode

Nulde orde mode Als de nulde orde mode gebruikt wordt zowel voor de belichting als de detectie krijgt men een optische overdrachtsfunctie die gelijkaardig is aan deze van een confocale microscoop. Op Fig. 10 staan de *otf* en de *psf* uitgezet voor verschillende waarden van de genormaliseerde breedte van de golfgeleider. De figuur geeft al aan dat hier een compromis zal moeten gemaakt worden. Voor een brede golfgeleider worden de hogere spatiale frequenties niet doorgelaten, en krijgen we een bredere puntspreidingsfunctie. Voor smalle golfgeleiders krijgt de *otf* een driehoekig profiel, maar daalt de totale doorvoer aan lichtvermogen. Door maximalisatie van de transmissie bij hogere frequenties vindt men een compromis bij $\overline{a} = 1/(2\sqrt{\pi})$.



Figuur 10: De $otf^{0,0}(\overline{f})$ en $psf^{0,0}(\overline{x})$ voor enkele waarden van de genormaliseerde golfgeleiderbreedte, $\overline{a}$. Voor de duidelijkheid is slechts de helft van de grafiek getoond. Zowel de otf als de psf zijn symmetrisch.

Eerste orde mode Als men voor de belichting de nulde orde mode gebruikt, maar voor de detectie de eerste orde mode krijgt men een totaal ander beeld. Zoals weergegeven op Fig. 11 lopen zowel de otf als de psf door de oorsprong. Voor de otf wil dit zeggen dat, in vergelijking met de detectie door middel van de nulde orde mode, de lagere spatiale frequenties gedempt worden. De hogere spatiale frequenties krijgen hierdoor een groter relatief gewicht.



Figuur 11: De $otf^{0,1}(f)$ en $psf^{0,1}(x)$ voor enkele waarden van de genormaliseerde golfgeleiderbreedte, $\overline{a}$. Beide functies zijn antisymmetrisch. De otf is zuiver imaginair, en de psf is zuiver reëel.

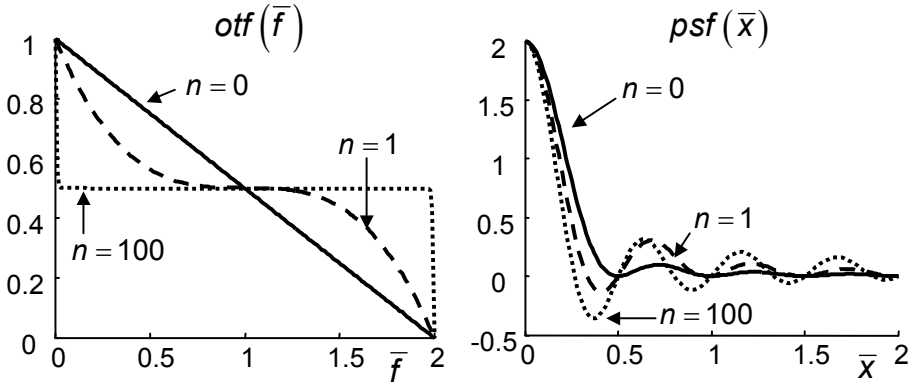
Hogere orde mode In principe is het ook mogelijk meerdere modes te combineren. Door een geschikte keuze van symmetrische hogere orde modes kan men voor de otf een vlak profiel benaderen. Zoals weergegeven in Fig. 12 heeft dit tot gevolg dat we een smallere puntspreidingsfunctie bekomen. Deze methode heeft echter twee nadelen. De eerste is dat net zoals in Fig. 6 de zijlobben een groter gewicht krijgen. Een tweede nadeel is dat het totale lichtvermogen sterk wordt gedempt. Wegens de normering is dit echter niet zichtbaar op de figuur.

Als $n+1$ symmetrische modes gecombineerd worden, kan men een otf maken met de vorm:

$$otf(f) \propto 1 + (1 - |f|\,\lambda/NA)^{2n+1}\,. \tag{8}$$

Voor $n = 0$ en $n = \infty$ is ook de psf een eenvoudige uitdrukking:

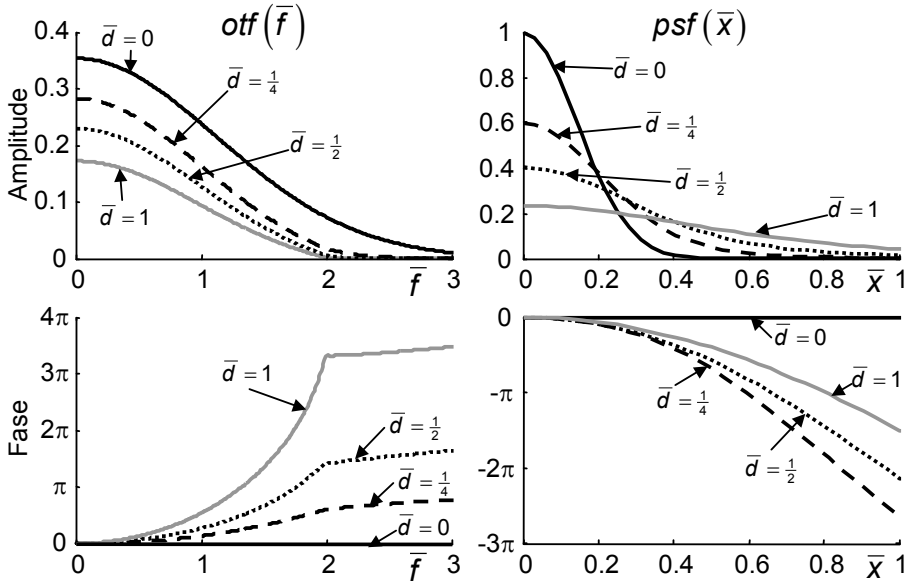
$$psf^{n=0}(x) \propto \mathrm{sinc}^2(2\pi NAx/\lambda) \qquad psf^{n=\infty}(x) \propto \mathrm{sinc}(4\pi NAx/\lambda)\,. \tag{9}$$



Figuur 12: De otf en psf voor detectie met een combinatie van $n+1$ even hogere orde modes, voor $n = 0$, $n = 1$ en $n = 100$. Voor de duidelijkheid werden de otf en psf genormaliseerd. Ze zijn beide symmetrisch en werden daarom enkel voor positieve $\overline{x}$ en $\overline{f}$ getekend.

Nabije veld In plaats van een lenssysteem te gebruiken kan de golfgeleider ook rechtstreeks boven de schijf geplaatst worden. In dat geval hangen de eigenschappen sterk af van de dikte van de luchtspleet tussen de golfgeleider en de schijf. Fig. 13 toont de otf en de psf in functie van deze spleetdikte. Bij een grotere luchtspleet worden ook de faseverschillen tussen de verschillende frequenties belangrijk. Hoewel dit

in principe de informatie-inhoud die kan worden overgedragen niet vermindert, wordt het uiteindelijke signaal wel moeilijker om te analyseren. Bemerk ook dat zelfs als er geen luchtspleet is, de resolutie niet oneindig is. Dit komt door de eindige afmetingen van de golfgeleider zelf.



Figuur 13: Amplitude en fase van de $otf^{0,0}(f)$ en $psf^{0,0}(x)$ voor waarneming in het nabije veld, voor $\overline{a} = \frac{1}{2\sqrt{\pi}}$. Voor de duidelijkheid werd enkel de helft van de grafiek getekend. De otf en de psf zijn beide symmetrisch.

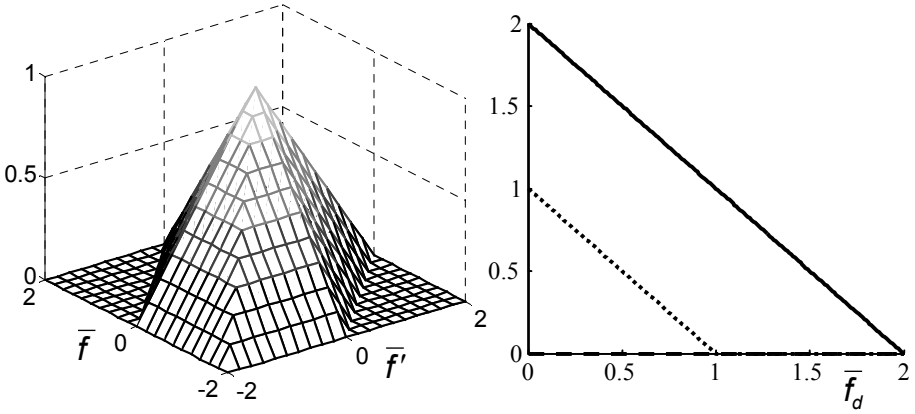
3.5 Uitgebreide-detectormethode

Voor de conventionele DVD detectiemethode, ook wel de uitgebreide-detectormethode genoemd, moeten de partieel coherente PSF en OTF gebruikt worden. Op Fig. 14 en Fig. 15 geven we een beeld van de OTF voor twee detectiesignalen.

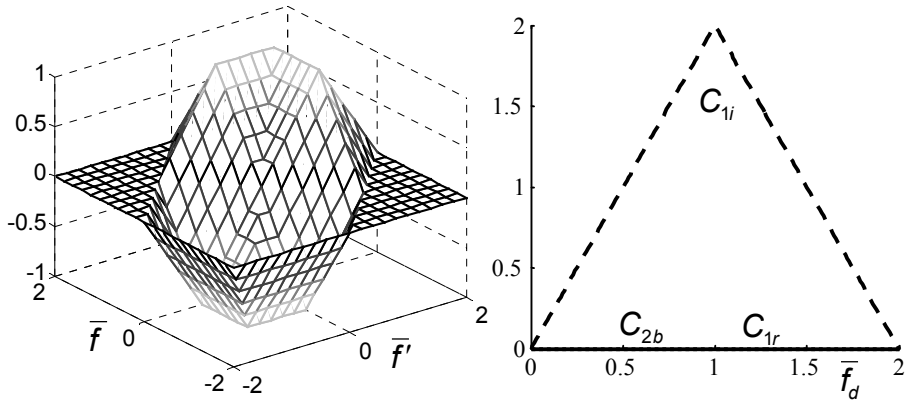
De eerste is de centrale-apertuurdetectie. Bij deze methode wordt al het lichtvermogen dat door de objectieflens opgepikt wordt, opgevangen op een grote detector.

Bij de tweede methode wordt de detector in twee helften gesplitst. Door het vermogen in de twee detectorhelften van elkaar af te trekken, krijgen we een 'duw-treksignaal'[9]. In sommige CD-systemen, wordt

[9] In het Engels: push-pull signal



Figuur 14: Aan de linkerkant: de genormaliseerde partieel coherente OTF voor het centrale apertuur signaal in een uitgebreide-detectorsysteem voor $\overline{a} \ll 1$. Aan de rechterzijde: de voornaamste doorsneden $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, en $\mathcal{C}_{2b}$.



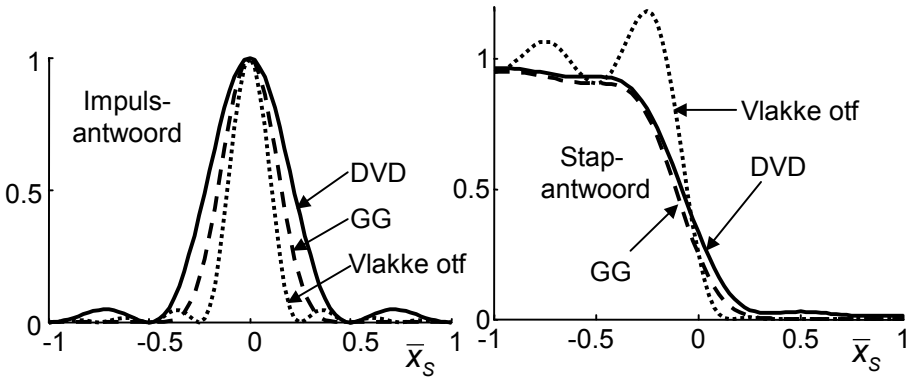
Figuur 15: Aan de linkerkant: de genormaliseerde partieel coherente OTF van het verschilsignaal van een gesplitste detector in een uitgebreide-detectorsysteem voor $\overline{a} \ll 1$. Aan de rechterzijde: de voornaamste doorsneden $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, en $\mathcal{C}_{2b}$.

dit ondermeer gebruikt om een bitspoor op de schijf nauwkeurig te volgen. Hoewel deze methode sterk gelijkt op de eerste orde mode detectie in de golfgeleider-detectiemethode, is ze toch sterk verschillend. Een opvallend verschil kunnen we zien op de grafiek op de rechterzijde in Fig. 15. De component $\mathcal{C}_{1r}$ is nul. Dit wil zeggen dat men enkel een signaal opvangt als de effectieve modulatiediepte een imaginair ge-

deelte bevat. Dit is een van de redenen waarom de diepte van de putjes in een CD een klein beetje groter of kleiner zijn dan $\lambda/(4n)$.

3.6 Confocale detectiemethode

Voor de confocale microscoop is het niet mogelijk een gesloten uitdrukking te vormen voor de OTF en de PSF. Men kan echter aantonen dat voor een heel kleine detectie-apertuur de confocale detectiemethode identiek is aan de golfgeleider-detectiemethode bij nulde orde mode detectie. En in het andere extreme geval, nl. bij een heel grote detectie-apertuur, is ze gelijk aan de uitgebreide-detectormethode.



Figuur 16: Een vergelijking van het signaal voor een kleine puls en voor een stapfunctie. De volle lijn toont het DVD detectie systeem, de gebroken lijn is de nulde orde detectie in het golfgeleider-detectiesysteem, en de stippellijn toont de curves voor een vlakke OTF.

3.7 Conclusies

In dit hoofdstuk werden de optische eigenschappen van de voorgestelde golfgeleider-detectiemethode vergeleken met deze van de DVD-detectiemethode en de confocale microscoop. Het is echter moeilijk om op basis van de optische overdrachtsfunctie ondubbelzinnig te bepalen welk systeem het beste is. Hiervoor zijn twee methodes gangbaar.

Een eerste methode is te kijken hoe goed het uiteindelijke signaal de bitpatronen op de schijf volgt. Dit is echter niet altijd even duidelijk. Fig. 16 geeft het impuls- en het stapantwoord weer van een DVD-detectiesysteem, van de nulde orde mode detectie met golfgeleider-detectiemethode en van een optisch systeem met een uniforme OTF,

zoals in Vgl. (8). De figuur geeft duidelijk aan dat het impulsantwoord van de DVD-detectiemethode breder is dan deze van het golfgeleider-detectiemethode, of deze van het systeem met een uniforme OTF. Een ander beeld krijgt men echter bij het stapantwoord: het systeem met een uniforme OTF geeft wel een steilere curve, maar dit ten koste van veel sterke franjes. Bovendien zorgt de grote asymmetrie tussen de overgang van 'nul' naar 'één' en omgekeerd ervoor dat het moeilijker wordt de juiste positie van deze overgang te bepalen.

Een tweede methode is gebaseerd op de Shannon limiet, die zegt dat de informatie-inhoud die binnen een bepaalde bandbreedte kan worden doorgestuurd evenredig is met $\log_2(1 + \mathrm{SNR})$. Hierbij is SNR de verhouding van het signaal tot de ruis. Deze wordt beïnvloed door de vorm van de otf maar ook door nog veel andere factoren. In het volgende hoofdstuk zullen we aantonen hoe de golfgeleider-detectiemethode door het combineren van de signalen van de nulde en de eerste orde mode, meer kleinere bits op de schijf zal kunnen reconstrueren dan het DVD-uitleessysteem of de confocale microscoop.

4. Extractie van het bitpatroon

4.1 Inleiding

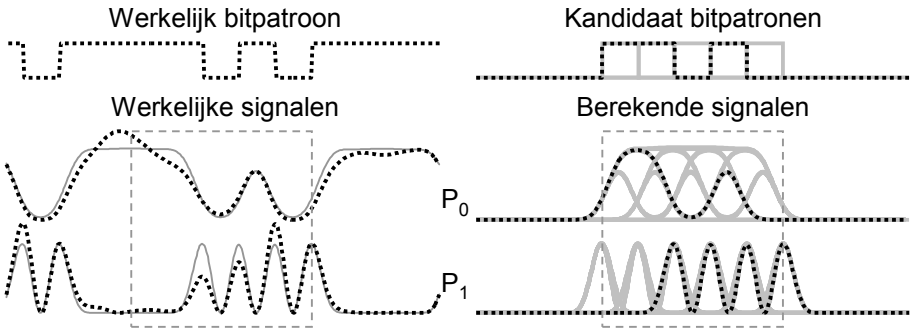
In een conventioneel CD- of DVD-uitleessysteem worden de 'nullen' en 'enen' van de bitsequentie uit het detectorsignaal gehaald door te kijken wanneer het signaal een zekere drempelwaarde kruist. Deze punten markeren een verandering van een 'één' naar een 'nul' of omgekeerd. Deze methode is snel en betrouwbaar maar gooit een stuk van de informatie uit de vorm van het signaal weg. Met andere methodes zoals door met meerdere grijsniveaus te werken [1, 2], kan de gegevenscapaciteit verhoogd worden. In dit hoofdstuk zullen we een systeem evalueren op basis van de fractie aan bitfouten. Dit is het aantal keer dat het detectie systeem de verkeerde bit aangeeft, gedeeld door het totaal aantal gedetecteerde bits.

Golfgeleider-detectiesysteem De resultaten in hoofdstuk drie toonden dat de optische overdrachtsfunctie van een confocale microscoop gelijk is aan deze van de golfgeleider-detectiemethode met belichting en detecting via de nulde orde mode. Bovendien hebben metingen aangetoond dat het uitlezen van een DVD met een confocale microscoop of met een klassiek DVD-uitleessysteem, tot dezelfde positieruis en fractie aan bit fouten leidden. Het idee is om naast de nulde orde ook met hogere orde modes te detecteren. Deze leveren elk een elektrisch signaal P_j. In dit onderzoek hebben we ons toegespitst op een combinatie van de nulde en eerste orde mode.

Ruis In een werkelijk uitgevoerd systeem is het gedetecteerde signaal verstoord door verschillende bronnen. Deze verstoring kan veroorzaakt zijn door intersymboolinterferentie (ISI) binnenin een bitspoor, overspraak tussen twee verschillende sporen, of door allerhande soorten ruis. Deze laatste bestaan o.a. uit detectorruis, ruis op de schijf zelf, parasitaire reflecties en overspraak in het fotonisch geïntegreerd circuit. Deze vormen van verstoring werden in onze simulaties ingebouwd door ruis, ISI en overspraak toe te voegen aan de bitpatronen.

4.2 Parameter-extractiemethode

Fig. 17 toont het algemene idee achter de parameter-extractiemethode. Veronderstel een werkelijk bitpatroon. Daarvan kunnen we het detectorsignaal van de verschillende modes berekenen als $P_j(t)$ inclusief



Figuur 17: Schematisch overzicht van de parameter extractie methode. In de linkerkolom staan de gedetecteerde signalen P_j: het theoretische signaal in het grijs en de werkelijke signalen verstoord door ISI, overspraak en ruis in het zwart. Voor de duidelijkheid werden enkel $P_0(t)$ en $P_1(t)$ getekend. In de rechter kolom: uit de k kandidaat bit patronen worden de theoretische signalen $C_j^k(t)$ berekend. Het idee is dat bitpatroon te vinden, waarvoor $C_j^k(t)$ het best gelijkt op de gemeten signalen $P_j(t)$.

overspraak of ruis. Voor de waarnemer zijn enkel deze detectorsignalen gekend. Voor een reeks van n testpatronen kunnen we ook de theoretische detectorsignalen berekenen, die we noteren als $C_j^k(t)$. Als we nu kijken welke van deze signalen het minst verschilt met het gemeten signaal, hebben we een aanwijzing voor het onderliggende werkelijke bitpatroon:

$$S_j^k = \int_{t_l}^{t_u} \Big(P_j(t) - C_j^k(t)\Big)^2 \,\mathrm{d}t. \tag{10}$$

Van zodra er ruis of overspraak is, kan het echter gebeuren dat we hiermee verkeerd gokken. Bovendien kan het gebeuren dat de beste keuze voor de nulde orde en deze voor de eerste orde niet dezelfde zijn. In deze gevallen nemen we een lineaire combinatie van de verschilintegralen van de betreffende modes:

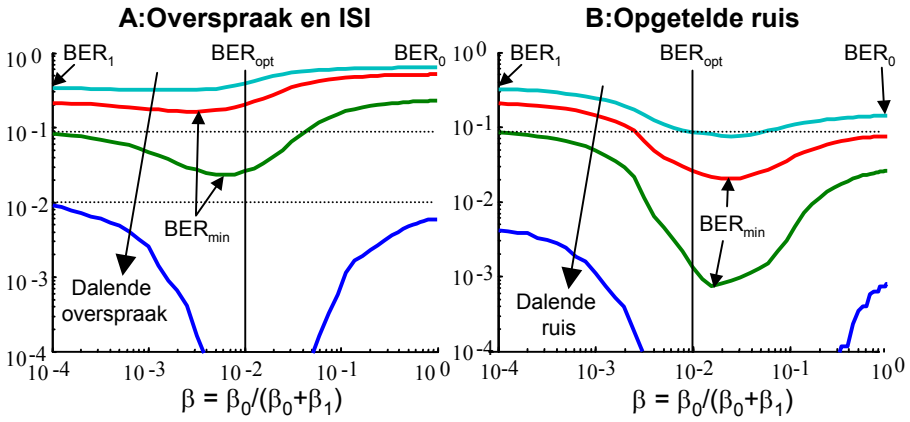
$$S_T^k = \sum_{j=0}^{N-1} \beta_j S_j^k. \tag{11}$$

De kleinste waarde van deze lineaire combinatie geeft ons een gok die een grote kans heeft om naar het werkelijke bitpatroon te leiden. Door stap voor stap telkens nieuwe kandidaat patronen voorop te stellen kunnen alle bits van de schijf teruggevonden worden. Het algoritme kan worden uitgebreid door tijdens het gokken gebruik te maken van de kennis van de bits die al teruggevonden zijn.

4.3 Implementatie

Het algoritme werd in MATLAB[10] geïmplementeerd. Eerst worden de puntspreidingsfunctie en de optische overdrachtsfunctie berekend door middel van snelle Fouriertransformaties[11]. In een tweede stap worden de signalen van de kandidaat patronen berekend. In de derde stap wordt het gemeten signaal van het werkelijke bitpatroon aangemaakt. Dit gebeurt onder verschillende omstandigheden van ruis en overspraak. In de vierde stap worden de verschilintegralen berekend, en tenslotte worden in de vijfde stap lineaire combinaties gemaakt, waarvan de beste gok voor het werkelijke bitpatroon wordt afgeleid.

De derde, vierde en vijfde stap worden verschillende keer herhaald. Op basis daarvan kan men een statistiek opmaken die aangeeft hoe vaak er een foute gok optreedt. Dit geeft de fractie aan bit fouten[12]. De fractie aan bitfouten die men bekomt door enkel de nulde orde te gebruiken, is weergegeven met BER_0. Als men enkel de eerste orde gebruikt, krijgt men BER_1.



Figuur 18: Fractie aan bitfouten als een functie van β. Het linker uiteinde geeft BER_1 en het rechter uiteinde geeft BER_0. Op de linker grafiek staat de fractie aan bitfouten voor verschillende hoeveelheden opgetelde ruis en een bitlengte van $0.15\lambda/NA$. En op de rechterzijde bij intersymboolinterferentie, verschillende hoeveelheden aan overspraak, en een bitlengte van $0.25\lambda/NA$.

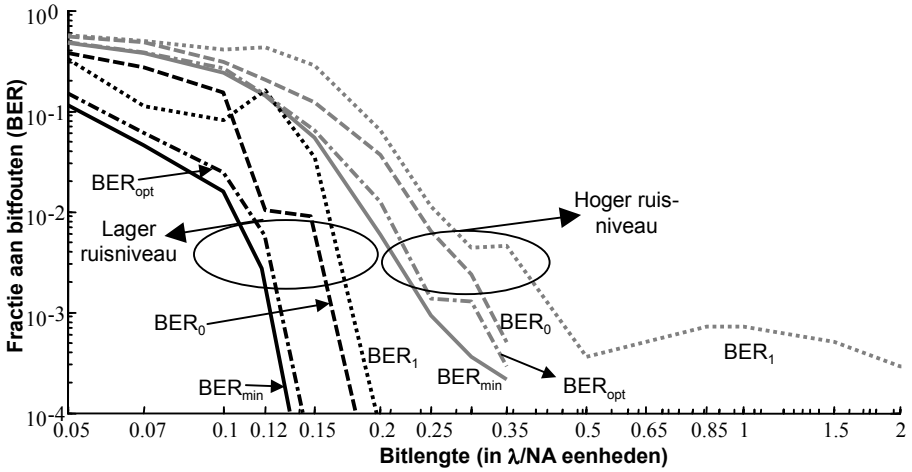
[10] MATLAB (MATrix LABoratory) is een numeriek softwarepakket van Mathworks Inc. (http://www.mathworks.com/).

[11] In het Engels: fast Fourier transforms (FFT)

[12] In het Engels: bit error rate (BER)

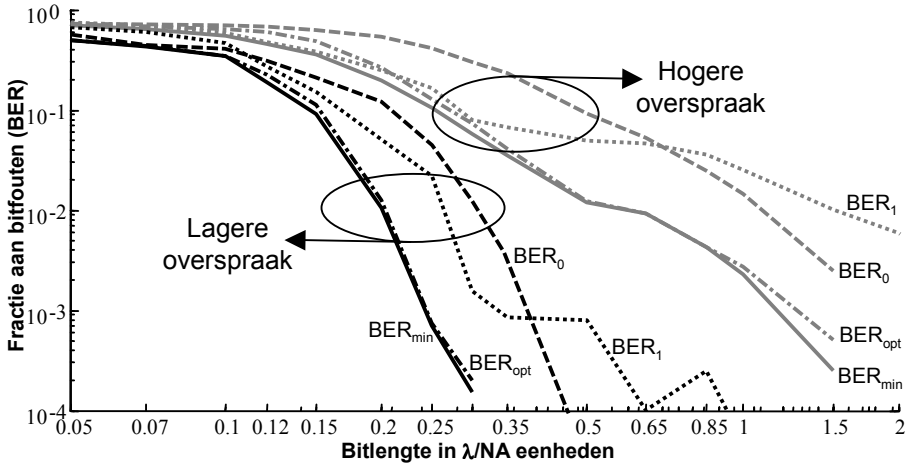
4.4 Resultaten

Fig. 18 toont de fractie aan bitfouten als een functie van de coëfficiënten in de lineaire combinatie van Vgl. (11). De figuur toont twee verschillende configuraties van bitlengte en ruis. In sommige gevallen is BER_0 groter dan BER_1 in andere gevallen is het omgekeerde het geval. In de meeste gevallen is de fractie aan bitfouten bij een tussenliggende lineaire combinatie echter nog een stuk lager. Wij vonden bij onze configuratie een optimale waarde van $\beta_0/\left(\beta_0+\beta_1\right)=\beta$ bij 10^{-2}. Deze keuze hangt echter sterk af van de parameters van de golfgeleider-detectiemethode zelf. In het ideale geval zouden we de simulaties moeten uitvoeren op heel veel bits. Wij hebben onze simulaties echter beperkt tot 10^4 bitpatronen.



Figuur 19: Fractie aan bitfouten als een functie van de bitlengte bij opgetelde witte ruis. De grijze lijnen tonen de fractie aan bitfouten voor een schijf met een grote hoeveelheid ruis, de zwarte lijnen tonen de situatie met een kleinere hoeveelheid ruis.

Fig. 19 toont de fractie aan bitfouten in functie van de bitlengte voor twee verschillende niveaus van opgetelde ruis. Over het algemeen leidt de nulde orde mode tot minder bitfouten dan de eerste orde mode. Een combinatie van beide zorgt echter voor nog minder bitfouten. Op Fig. 20 staat een gelijkaardige grafiek, maar voor overspraak en intersymboolinterferentie. Bij grote bits overweegt de overspraak en is de eerste orde mode gunstiger, voor kleinere bits is de intersymboolinterferentie het belangrijkst en is BER_0 lager. Ook hier weer geeft een



Figuur 20: Fractie aan bitfouten als een functie van de bitlengte, bij overspraak. De zwarte lijnen tonen de fractie aan bitfouten voor een schijf met sterke overspraak, de grijze lijnen bij zwakkere overspraak. De intersymboolinterferentie is in beide gevallen gelijk.

combinatie van beide modes echter nog een daling van het aantal bit fouten.

4.5 Conclusies

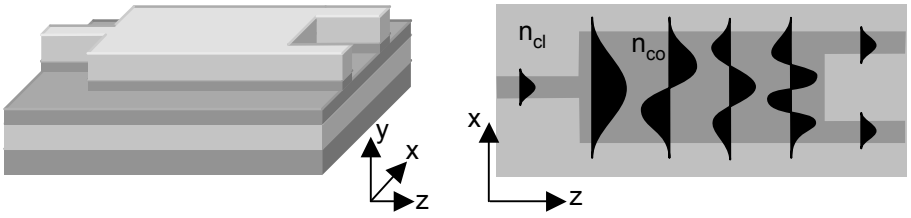
In dit hoofdstuk bespraken we een eenvoudige parameter-extractiemethode om de bitpatronen op de optische schijf te achterhalen op basis van de detectorsignalen van de nulde en de eerste orde mode. Het nulde orde mode signaal is gelijk aan dat van een confocale microscoop. Deze geeft op zijn beurt dezelfde resultaten als een klassiek oppiksysteem. Het nieuwe aan de golfgeleider-detectiemethode is dat we ook een eerste orde mode signaal oppikken. Een combinatie van beide signalen samen leidt bijna altijd tot een daling van het aantal bitfouten.

5. Fotonisch geïntegreerd circuit

Een fotonisch geïntegreerd circuit (PIC)[13] is een schijfje waarop een aantal optische functies uitgevoerd worden op een klein halfgeleiderschijfje. In dit hoofdstuk bespreken we een PIC die verschillende modes kan afsplitsen. Dit kan door middel van een multimode-interferentiekoppelaar, beschreven in sectie 5.2, of met een directionele koppelaar, beschreven in sectie 5.3. Verder wordt ook de werking van een optische interferometer beschreven, een component die het faseverschil tussen twee golfgeleiders kan meten. Tenslotte bespreken we het ontwerp en de vezelkoppeling

5.1 Mode splitsers gebaseerd op MMI's

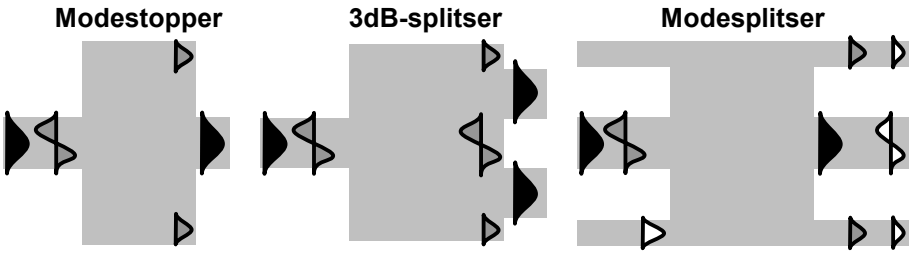
Een methode om de modes in een multimodale golfgeleider te controleren werd voorgesteld door Leuthold [3]. Het basisprincipe is gebaseerd op de verschillende afbeeldingseigenschappen van de symmetrische en antisymmetrische modes in een multimode-interferentiekoppelaar (MMI)[14]. Fig. 21 toont schematisch de werking van deze component. Voor de duidelijkheid werd de drie-dimensionale structuur vereenvoudigt tot twee dimensies.



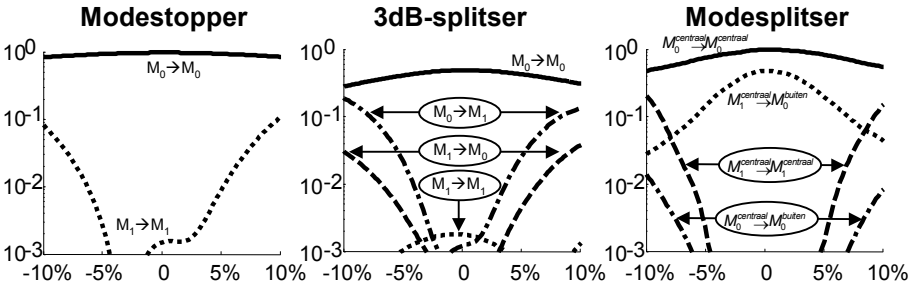
Figuur 21: Een 1×2 MMI en de tweedimensionale voorstelling. Een of meerdere ingangsgolfgeleiders exciteren de modes van een brede multimodale sectie. Door een verschil in propagatieconstante tussen deze modes verkrijgt men aan de andere zijde een interferentiepatroon. Bij een juist gekozen lengte krijgt men een of meerdere afbeeldingen van het ingangsveld.

Wij gebruikten in ons ontwerp drie soorten componenten, schematisch voorgesteld in Fig. 22. Voor elk van deze componenten proberen we ongewenste overspraak te vermijden. Zo willen we bijvoorbeeld niet dat de modestopper de eerste orde mode voor een klein gedeelte

[13] In het Engels: photonic integrated circuit
[14] In het Engels: multimode interference coupler



Figuur 22: Een schematische voorstelling van drie componenten in ons ontwerp (niet op schaal getekend). De modestopper laat enkel de nulde orde mode door en stopt de eerste orde mode. De 3dB-splitser maakt twee gelijke afbeeldingen van de nulde orde mode, en stopt de eerste orde mode. De modesplitser stuurt de nulde orde mode van de centrale ingangsgolfgeleider naar de centrale uitgangsgolfgeleider, de eerste orde mode van de centrale ingangsgolfgeleider wordt verdeeld over de buitenste uitgangsgolfgeleiders als een nulde orde mode. Een excitatie van de buitenste ingangsgolfgeleiders geeft een eerste orde mode in het midden en twee nulde orde modes aan de buitenkant.



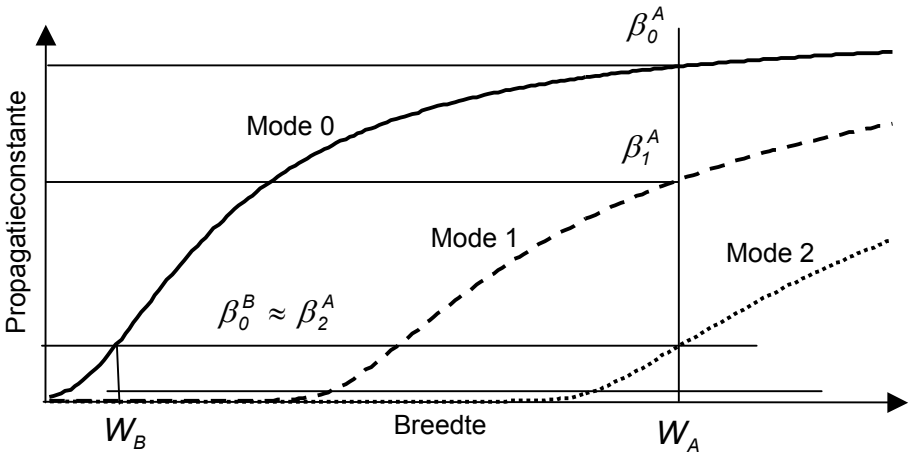
Figuur 23: Toleranties op de lengte van de MMI's zoals ontworpen voor masker 2. De figuur geeft voor elk van de drie componenten de vermogenstransmissie weer in functie van de afwijking op de lengte. Ook de voornaamste mogelijkheden van overspraak staan aangeduid. Hoe sterker de werkelijke lengte afwijkt van de ontworpen lengte, hoe meer overspraak, en hoe zwakker het gewenste signaal.

zou doorlaten. Theoretisch kunnen we deze overspraak heel laag houden, maar in de praktijk zijn de dimensies en de brekingsindices van de gefabriceerde componenten nooit identiek aan deze van het ontwerp. Daarom werd in het ontwerp ook gekeken naar de toleranties voor kleine afwijkingen. Fig. 23 toont het effect bij een afwijking op de lengte. Dezelfde grafieken kan men ook gebruiken voor de afwijkingen op de breedte, aangezien men uit de theorie van de MMI kan bewijzen

dat een relatieve afwijking op de breedte hetzelfde effect heeft als een dubbel zo grote relatieve afwijking op de lengte.

5.2 Mode splitsers gebaseerd op directionele koppelaars

In de vorige sectie toonden we dat het met een MMI mogelijk is de even en de oneven modes op te splitsen. Een andere techniek om modes op te splitsen is de directionele koppelaar. De directionele koppelaar bestaat uit twee golfgeleiders, die elkaar naderen, een eind parallel lopen en zich dan weer verwijderen. In de koppelsectie koppelt het licht van de ene naar de andere golfgeleider. Dit proces hangt af van de overlap tussen de twee golfgeleidermodes, hun propagatieconstante en de lengte van de koppelsectie.



Figuur 24: De propagatie constante als een functie van de golfgeleiderbreedte voor verschillende modes. De nulde orde mode in volle lijn, de eerste orde mode in gebroken lijn en de tweede orde mode in stippellijn. Voor een goede koppeling, moeten de breedtes van de golfgeleiders zo gekozen worden dat de propagatieconstante van de nulde orde mode in golfgeleider B, β_0^B, gelijk is aan de propagatieconstante van de tweede orde mode in golfgeleider A, β_2^A.

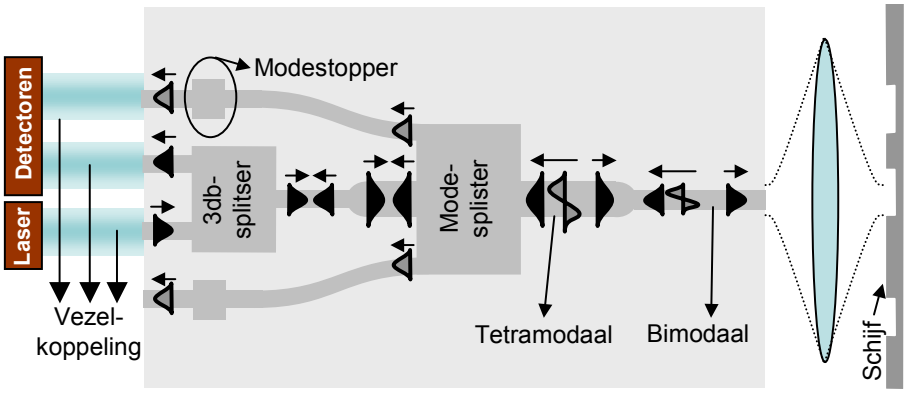
Fig. 24 toont de propagatieconstante van een type golfgeleider in functie van de breedte. Een significante overkoppeling vindt pas plaats als de propagatieconstantes van twee golfgeleidermodes bijna gelijk zijn. Door de breedte van golfgeleider A en B zó te kiezen dat de propagatieconstante van de tweede orde van golfgeleider A, β_2^A gelijk is aan de propagatieconstante van de nulde orde van golfgeleider B, β_0^A, kunnen we de tweede orde mode van golfgeleider A afsplitsen.

5.3 Optische interferometer

In samenwerking met de Technische Universiteit Delft werd ook een optische interferometer ontworpen en gemaakt. Deze component bestaat uit een cascade van 3dB-splitsers waarmee we het faseverschil tussen twee golfgeleidermodes kunnen opmeten. Met deze techniek kan een uitleeskop met twee nabijgelegen monomodale golfgeleiders een alternatieve implementatie van een uitleeskop met één enkele bimodale golfgeleider vormen.

5.4 Ontwerp

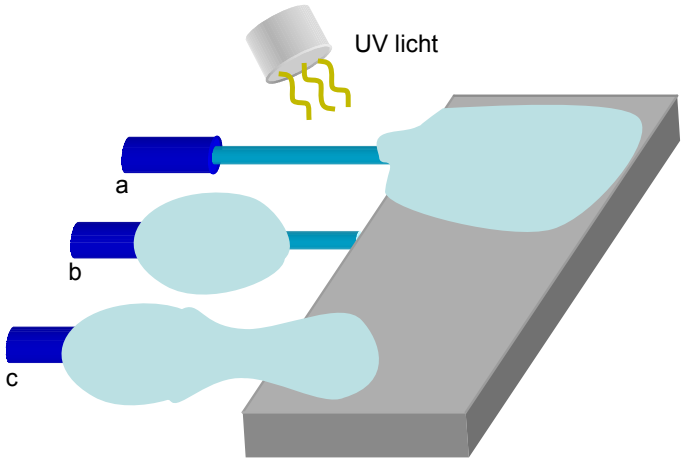
In de loop van het doctoraat werden twee maskers ontworpen, waarop verschillende componenten werden geplaatst. De componenten op het eerste masker bestaan uit twee modesplitsers gekoppeld door een bimodale golfgeleider en een aantal ingangs- en uitgangsgolfgeleiders. De componenten op dit eerste masker zijn zo ontworpen dat de overspraak in de modesplitser kan worden opgemeten door het vermogen in de verschillende uitgangsgolfgeleiders op te meten. Het tweede masker is een complexere uitbreiding op het eerste masker en laat ook toe om vezels te koppelen aan de componenten. Hiermee is het mogelijk om tegelijkertijd te belichten en te detecteren.



Figuur 25: Een schematisch overzicht van de optische lichtweg in de componenten van masker 2, in de uiteindelijke implementatie om optische schijven uit te lezen.

Fig. 25 toont een schematisch overzicht van hoe deze component uiteindelijk zal gebruikt worden voor het uitlezen van een optische schijf. Een laser koppelt via een vezel en één van de ingangsgolfgelei-

ders licht in het PIC en vormt uiteindelijk een lichtvlek op de schijf. Het gereflecteerde licht wordt in de nulde en de eerste mode weer opgepikt en gaat via twee andere golfgeleiders naar de detectoren. Op het masker zelf staan naast deze componenten ook nog een aantal extra componenten om het fabricageproces te controleren en om de overspraak in de modesplitsers te kunnen opmeten. Bovendien staat er ook een component op die ontworpen is voor individuele koppeling met drie afzonderlijke vezels.



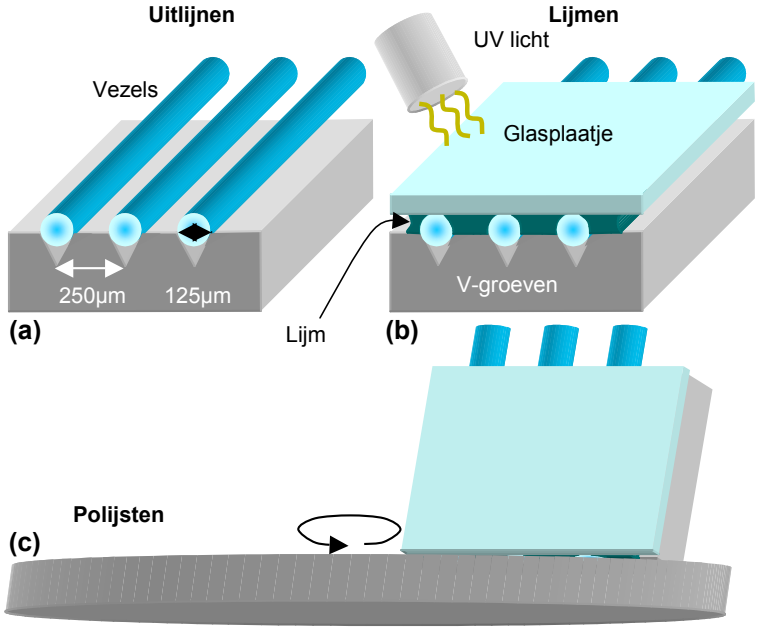
Figuur 26: Versterken van de vezelverbinding door middel van extra lijmdruppels.

5.5 Vezelkoppeling

Voor het koppelen van de golfgeleiderplaatjes met optische vezels zijn twee technieken gebruikt. Bij de eerste techniek worden de vezels individueel aan het plaatje gelijmd. Het probleem hierbij is dat door de sterke cohesiekrachten de lijm voor het grootste gedeelte naar het plaatje wordt getrokken. De uiteindelijke koppeling blijkt heel breekbaar te zijn. Door eerst een druppel een paar millimeter van het plaatje weg uit te harden via UV-licht wordt er een soort brug gemaakt waarin de volgende druppel blijft hangen. Als ook deze wordt uitgehard, krijgt men een stevige koppeling, zoals getoond op Fig. 26.

Een tweede techniek is weergegeven op Fig. 27 en bestaat erin de vezels eerst in een rij van V-vormige groeven te kleven. Met een gelijmd glasplaatje worden deze vezels stevig vastgezet. Het eindfacet

wordt vervolgens mooi glad gepolijst, waarna de vezelrij als geheel aan het golfgeleiderplaatje kan worden gelijmd. Het voordeel van deze tweede techniek is dat de vezels veel dichter bij elkaar kunnen worden geplaatst. Hierdoor kunnen ook de golfgeleidercomponenten kleiner gemaakt worden en passen er meer op hetzelfde plaatje.



Figuur 27: Vezeluitlijning in V-groeven. Doorheen een glasplaatje dat op de vezels in de groeven drukt, wordt door middel van UV-licht de uitlijning permanent gemaakt. Op het einde worden de facetten gepolijst om een zuiver en vlak oppervlak te verkrijgen.

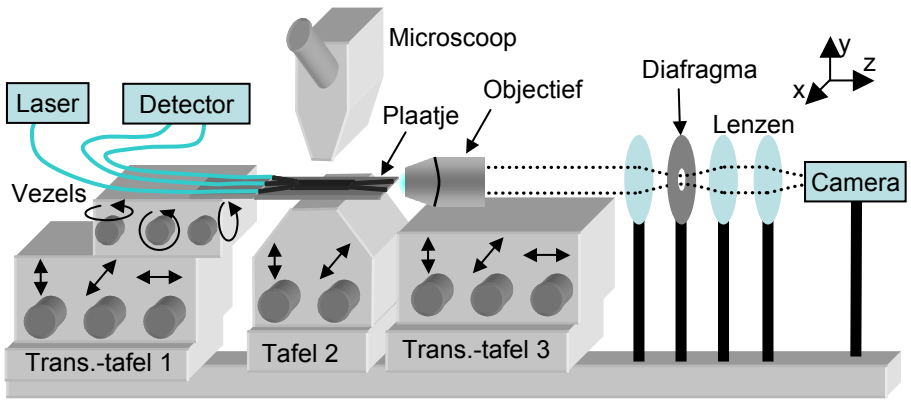
5.6 Toekomstige uitbreidingen

De huidige componenten kunnen op twee manieren worden uitgebreid. Een eerste uitbreiding is om naast de modesplitsers ook de detectoren en de laser op hetzelfde halfgeleiderschijfje te integreren. Hierdoor zou men een goedkope, robuuste, en compacte uitleeskop kunnen bekomen. Een tweede uitbreiding is om de detectiegolfgeleider zowel horizontaal als verticaal bimodaal te maken. Op die manier zou men vier in plaats van twee signalen krijgen. Met een combinatie van deze signalen zou het aantal bitfouten verder verlaagd kunnen worden en eventueel ook de focusering en de spoorvolging kunnen worden geïntegreerd.

6. Metingen

6.1 Metingen op de afzonderlijke golfgeleiderplaatjes

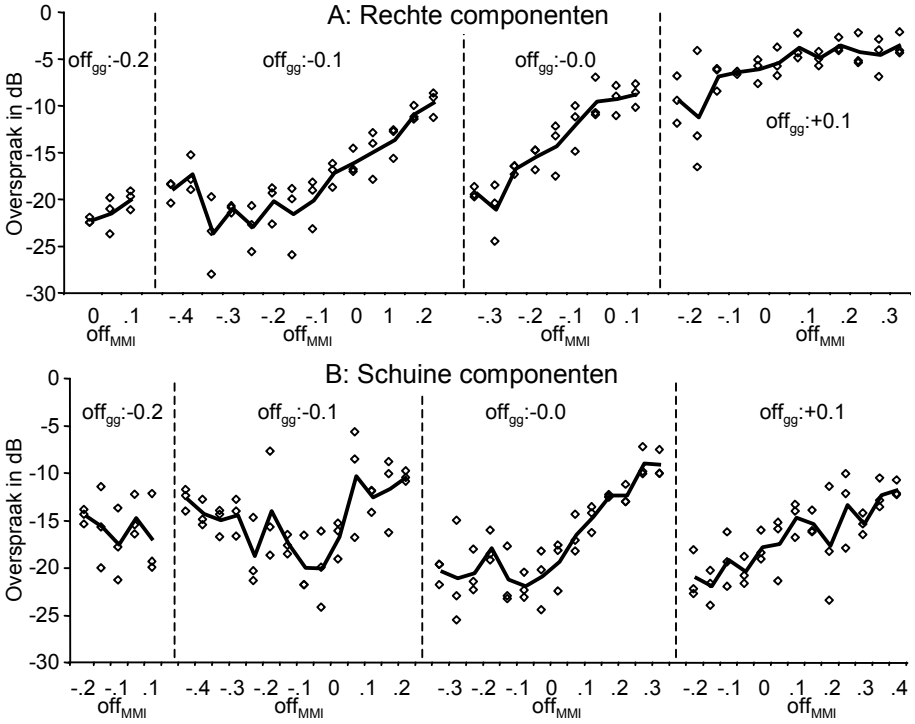
Fig. 28 toont een schematisch overzicht van de meetopstelling in de vakgroep INTEC aan de Universiteit Gent. De opstelling bestaat uit drie delen: via de eerste translatietafel wordt het licht ingekoppeld in de golfgeleiders, die op de tweede translatietafel liggen. Via het uitkoppelobjectief op de derde translatietafel wordt het licht verder geleid naar een detector. Een diafragma laat toe om één welbepaalde uitgangsgolfgeleider te selecteren.



Figuur 28: Meetopstelling voor het testen van de componenten, zoals gebruikt voor het uitlijnen van de vezelrij met het golfgeleiderplaatje. Een vezelrij (op translatietafel 1) wordt uitgelijnd met het golfgeleiderplaatje (op translatietafel 1). Het licht wordt uitgekoppeld door middel van een objectieflens (translatietafel 3). Daarna wordt het licht verder geleid door een diafragma naar de detector.

Zoals vermeld in hoofdstuk 5 werden de maskers zo ontworpen dat de overspraak van de modesplitser kan worden opgemeten. Omdat het bijna onvermijdelijk is dat bij de fabricage de componenten iets te breed of iets te smal zijn, werden op het masker verschillende maal dezelfde component geplaatst met telkens een licht verschillende breedte voor de MMI, of voor de golfgeleiders. Het idee is door deze variatie in het ontwerp op voorhand te compenseren voor eventuele de fabricagefouten. Hiermee zou minstens een aantal van de componenten de juiste breedte moeten hebben. Fig. 29 toont hoe de overspraak in de modesplitsers op masker 1, varieert volgens de variatie op de breedte van de golfgeleiders, off_{gg}, en deze van de MMI's, off_{mmi}. De algemene trend

is dat de kleinste overspraak gevonden wordt bij de smalste MMI's. Hieruit kunnen we besluiten dat de gefabriceerde componenten iets breder zijn dan ontworpen. Niet alle componenten zijn even goed, maar er zitten wel componenten bij waar de overspraak lager is dan -22 dB. Metingen op de componenten van het tweede masker waren iets minder succesvol. De beste component had een overspraak van -15 dB.



Figuur 29: Overspraak in de buitenste uitgangsgolfgeleiders van de modesplitsers van masker 2. Op de bovenste grafiek staan de meetresultaten van de rechte componenten, op de onderste grafiek deze van de schuine componenten. De vierkantjes tonen de overspraak opgemeten aan één van de ingangsgolfgeleiders. De volle lijn geeft het gemiddelde van deze verschillende metingen. Sommige componenten waren beschadigd en konden niet worden opgemeten.

6.2 Het resolveren van kleine details

Nadat de componenten uitvoerig afzonderlijk getest waren zijn de beste gebruikt om te testen in welke mate kleine details op een masker wor-

den opgepikt door de nulde en de eerste orde mode. Hiervoor werd een nieuwe opstelling gebouwd gelijkaardig aan deze van Fig. 28. Het grootste verschil is dat bovenop translatietafel 1 een extra translatietafel is geplaatst waarmee een masker kan worden gepositioneerd. Op dit masker staan een aantal teststructuren, waaronder fijne lijntjes, een aantal bitpatronen en een reeks roosters met verschillende periodes. De afstand van dit masker tot het golfgeleiderplaatje wordt gecontroleerd door middel van een capacitieve sensor.

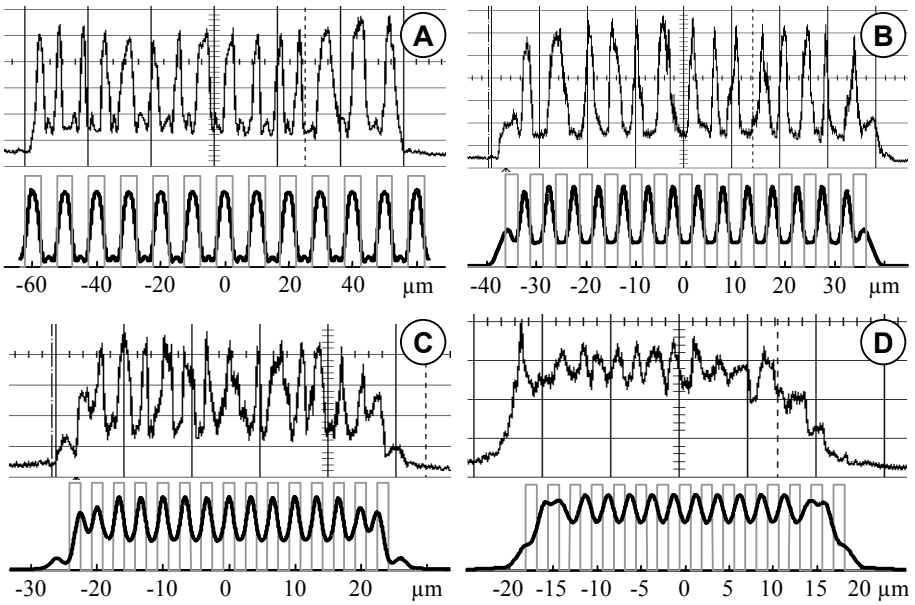
Fig. 30 toont de simulaties en de metingen van het licht dat door de nulde orde wordt opgepikt als een rooster langs de golfgeleider wordt geschoven. De simulaties en de metingen konden in goede overeenkomst worden gebracht door te veronderstellen dat de luchtspleet tussen het masker en de golfgeleider een 20-tal micrometer bedraagt. De reden waarom dit een stuk meer is dan de geplande 1 µm, is wellicht dat het erg moeilijk is er voor te zorgen dat het masker en het golfgeleiderplaatje 100% parallel lopen.

6.3 Reflectiemetingen

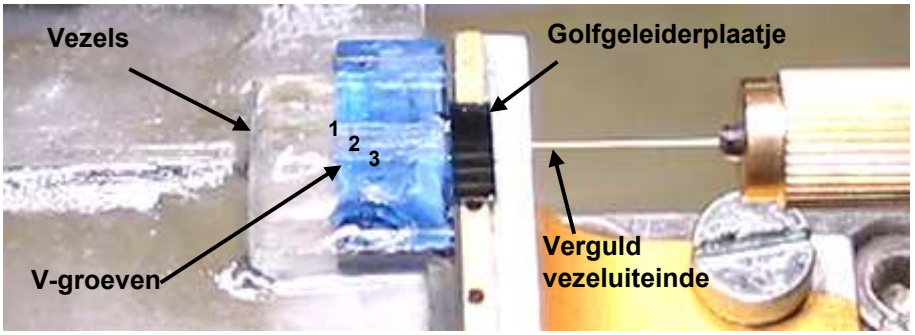
Met de beste component van het tweede masker werd een reflectiemeting gedaan. Omdat het masker met de roosters te veel oppervlakteruwheid had, was dit niet bruikbaar in reflectie. In plaats daarvan hebben we een met goud opgedampt vezeluiteinde gebruikt. Fig. 31 toont links een vezelrij, rechts het vezeluiteinde en daartussen de golfgeleidercomponenten. Uit de metingen bleek dat de reflectie aan het goudfacet goed meetbaar was in de nulde orde mode. Voor conclusies met betrekking tot de resolutie was het signaal echter te zwak in vergelijking met de parasitaire reflecties in de optische lichtweg.

6.4 Conclusies

De metingen op de afzonderlijke componenten toonden dat het mogelijk is de eerste orde af te splitsen met een overspraak lager dan -22 dB. De metingen om de resolutie van de golfgeleider-detectiemethode te quantificeren waren echter minder succesvol, bovendien bleek dat de verliezen aan lichtvermogen vrij groot waren.



Figuur 30: Detectie met de nulde orde mode van extern belichte roosters. De figuur toont de metingen bovenaan en de simulaties onderaan. De roosterperiode in de vier figuren is respectievelijk: 10λ(A), 5λ(B), $3{,}3\lambda$(C), $2{,}5\lambda$(D). Omdat het rooster manueel geschoven werd, zijn de x-coördinaten enkel kwalitatief correct. Met de simulaties werden de meetresultaten benaderd door een luchtspleet van 20λ tussen het rooster en het golfgeleiderplaatje te veronderstellen.



Figuur 31: Detail van de meetopstelling voor reflectiemetingen met de componenten van masker 2. Vezels in V-groeven worden uitgelijnd met de in- en uitgangsgolfgeleiders van de component op het plaatje. Een met goud opgedampt vezeluiteinde wordt uitgelijnd met het golfgeleiderfacet aan de andere kant van de component. Vezel 1 zorgt voor de belichting met de nulde orde mode; vezel 2 en vezel 3 detecteren respectievelijk de nulde orde en de eerste orde in reflectie.

7. Conclusies

In dit doctoraat werden de resultaten gebracht die het potentieel aantoonden van een multimodale golfgeleider als een uitleeskop voor optische schijven. Met het zogenoemde golfgeleider-detectiesysteem wordt het gereflecteerde licht van de schijf via twee afzonderlijke modes opgevangen. De optische eigenschappen van dit systeem werden grondig geanalyseerd en vergeleken met deze van een conventionele DVD-uitleesmethode en de confocale microscoop. De simulaties toonden aan dat door de signalen van twee modes te combineren het aantal bitfouten kan worden gereduceerd. We toonden ook de analyse en het ontwerp van een fotonisch geïntegreerd circuit dat de nulde en de eerste orde mode opsplitst.

7.1 Minpunten

Lichtvermogen De golfgeleider-detectiemethode houdt minder lichtvermogen over dan een conventionele DVD methode. Dit verlies is voor een stuk intrinsiek tengevolge van de noodzakelijk kleine afmetingen van de golfgeleider-detectiemethode, anderzijds zijn er ook verliezen in het PIC en aan de interfaces tussen de vezel en halfgeleider.

Overspraak Een tweede minpunt is een groter aantal mogelijke bronnen van overspraak. We hebben weliswaar experimenteel aangetoond dat de overspraak binnen in het PIC lager kan zijn dan -22 dB. De vraag blijft echter of dit voldoende is. Bovendien kunnen parasitaire reflecties aan de facetten ook nog voor extra overspraak zorgen tussen de belichtingsmode en de detectiemode.

Compatibiliteit In tegenstelling tot harde schijven moet de technologie van optische schijven internationaal gestandaardiseerd worden. Onze methode is op een aantal vlakken erg verschillend van de huidige systemen wat een mogelijke implementatie sterk zou kunnen afremmen.

Snelheid Naast capaciteit is ook de uitleessnelheid erg belangrijk. In vergelijking met de signaalverwerking in een DVD-uitleessysteem zijn de verschilintegralen van de parameter-extractiemethode relatief rekenintensief. Het zal afhangen van de rekenkracht van de toekomstige computerhardware of dit een belangrijk probleem vormt.

7.2 Potentieel en toekomstige perspectieven

Resolutie Met de golfgeleider-detectiemethode is het mogelijk zowel fase- als amplitude-informatie uit het gereflecteerde licht van de schijf op te vangen. In hoofdstuk 4 toonden we aan de hand van een eenvoudige parameter-extractiemethode aan dat we hiermee het aantal bitfouten kunnen verkleinen, wat de deur opent naar kleinere bitlengtes. De simulaties toonden ook dat het alleen nut heeft in het nabije veld te werken als de luchtspleet kleiner is dan de golflengte.

Compatibiliteit met meerlaagstechnieken Vanwege haar confocale aard heeft de golfgeleider-detectiemethode een goede diepteresolutie. Dit betekent dat enkel licht dat gereflecteerd wordt aan de laag die zich op brandpuntsafstand bevindt, zal worden opgevangen door de detector.

Integratie en miniaturisatie In principe moet het mogelijk zijn om naast de modesplitser ook andere optische functies op het PIC te integreren. Dit komt niet alleen de kostprijs bij massafabricage ten goede, maar kan ook leiden tot erg kleine uitleeskoppen.

Parallellisatie De integratie van de optische uitleeskop op één enkel golfgeleiderplaatje opent ook de deur naar parallel uitlezen en schrijven, wat zonder veel extra kosten tot hogere uitleessnelheden kan leiden.

Uitbreiding naar drie dimensies In dit doctoraat lag de nadruk op het verhogen van de gegevensdichtheid binnen één spoor. Een volgende stap zou erin bestaan een golfgeleider te gebruiken die niet alleen bimodaal is in de richting van het spoor maar ook in de richting daar loodrecht op. Hiermee zou niet alleen de gegevensdichtheid nog verder kunnen worden verhoogd, maar zou ook het mechanisme om een bitspoor op de schijf nauwkeurig te volgen, kunnen worden geïntegreerd.

English text

Contents

Chapter 1

Introduction

In the first section we will give a general overview of data storage in general. In a second section we will give an outline of optical disc[1] data storage. Then we will explain the new waveguide detection method proposed in this thesis. This method uses a multimodal waveguide to retrieve the bit patterns on the disc by illuminating the disc and capturing the reflected light through different waveguide modes. Finally we will give the outline of the thesis and the publications published in this framework.

1.1 Data storage

1.1.1 Introduction

Starting with the invention of writing, data storage has been a very important tool. Signs carved in stone and scriptures drawn on papyrus or parchment were the early methods to store information for later use. Later on the printing press facilitated copying of information, and large libraries were filled with piles of books. With the advent of computers, which generated enormous amounts of information, data storage changed drastically. The focus shifted from long term storage to speed and capacity. To cope with the increasing need for fast and high ca-

[1] The correctness of the spellings 'disc' and 'disk' is not trivial. The spelling not only varies from continent to continent but also from usage to usage. The most common logic is that an 'optical disc' is written with a 'c', as the inventors Philips and Sony used the British spelling. A 'hard disk' on the other hand is written with a 'k' as it originates from the American company IBM.

pacity data storage, different technologies evolved that each provide solutions for different needs.

A commonly used division is primary and secondary storage. Primary storage is the place where information is stored for the first time, it is volatile, and often named with the term: *memory*. This type of data storage needs in the first place to be very fast, and is further subdivided into subcategories as CPU registers, cache memory, and ordinary random access memory (RAM). Historically the technology was vacuum tube-based, but since the sixties RAM is made on an electronic integrated circuit. The bits are stored as the state of a Flip-Flop (Dynamic RAM) or the charge in a tiny capacitor (Static RAM). The latter tends to be faster and is often used for CPU registers and cache memory, but is is more expensive as it requires more transistors.

Secondary storage, or off-line storage, gets the data from the primary storage, and holds this information for a longer time. This type of storage is slower, has a higher capacity and is in general non-volatile. Off-line storage can be further subdivided into classes based on different properties such as rewritability and removability. Rewritability defines whether a bit once written can ever be changed. Rewritability is very often a necessity, but can be a disadvantage for reasons of security. Immutable discs are written once and cannot maliciously or accidentally be changed. In principle all non-volatile storage systems are removable, but some more than others. Demounting, moving and remounting a fixed hard drive requires more effort than swapping optical discs between two optical drives. One also differentiates between removable storage where the read-out hardware is moved along with the storage medium as for a removable hard disk, and removable storage where the read-out hardware remains fixed as for optical discs.

1.1.2 Removable storage

Removable data storage as a commercial commodity is a growing market, which includes consumer products as digital home video storage, movie distribution, and data backup. Also for professional purposes removable data storage is important as a means of storing medical images, huge databases or large backup files. Compared to the consumer market long term stability of the stored data is far more important.

There are different competing technologies eager to take a share of this market. These technologies are benchmarked along different properties such as: The **capacity** is the total number of bytes that can be

stored and is mostly measured in gigabyte[2]. The **data rate** gives how fast a continuous stream of information can be read out and is commonly measured in megabits per second (Mbps). The **access time** is the average delay it takes to find and read a given bit, and is measured in milliseconds. **Stability** has no unambiguous measure, but is very important for several products, where long term storage is needed. Finally **the price** of the drive unit, and the price of the replaceable media also form an important factor. Throughout the last decades different technologies have made a huge evolution and have each claimed a place on the data storage market:

- The earliest example of removable computer storage is the punch card. It consists of a paper strip in which holes are punched, which can later be read out. The punch card as a means of storage is nearly extinct. The Florida voting ballot is a notable exception.
- Magnetic tape storage stores the data by switching the magnetization of small ferromagnetic particles on a long strip fitted into a 12 cm cartridge. Recent developments showed cartridges of one terabyte that can be read out at 35 Mb/s [4]. The main disadvantage over other storage types is that it can only be read out sequentially.
- Removable magnetic disks or disk drives still form an important share of the storage market, from simple floppy disks to removable hard drives containing a complete hard disk drive including read-out mechanics and electronics in a single case. Similar to the magnetic tape, the information is stored by switching the magnetization of small ferro-magnetic domains. The total data capacity of such a hard disk can reach several 100 gigabytes and can be read-out at several 100 megabit/s. The access time on the other hand is in the order of a few milliseconds.
- Flash memory is a type of EEPROM (Electrically Erasable Programmable 'Read-Only' Memory), which stores a bit by trapping electrons in a transistor. Flash memory is fast and noiseless, as there are no mechanical elements, but has a higher cost per megabyte than the other removable storage types. Typical data capacity is several gigabytes and the transfer speed is several megabits per second.

[2] In contrast to other computer hardware, data storage manufacturers use the official SI units. This means that 1 gigabyte equals 10^9 bytes, which is not equal to $2^{30} = 1024^3$.

- Optical data storage uses a light beam to write and read out bits, mostly on circular discs. It exists as well as a read-only, a write once, or as a fully rewritable version. In the next section this type of storage will be described more extensively. The specifications of the current optical data storage systems are summarized in table 1.1.

- In the future technologies as parallel detection with an atomic force microscope may push the frontiers of data density further up to $1\ \mathrm{Tb/in^2}$ [5, 6].

Optical data storage has some clear advantages over its direct competitors. The first is the price. Although the drive may cost a few hundred euros, the disc costs only a few tenths of a euro. Further, optical discs are not altered by electromagnetic interference and prove to be the best choice with respect to long term stability. Finally there are optical discs that can only be written once and not be altered afterwards, which is very important for security reasons in data that must be trusted on. Therefore it can be foreseen that optical disc data storage will dominate the storage market for content distribution and data archival. In other niches such as video recording or as a storage means in mobile devices it may also find a place among other technologies.

1.2 Optical disc data storage

1.2.1 Introduction

The first real example of optical disc data storage, is the Laserdisc, which was demonstrated by Philips and Music Corporation of America (MCA) in 1972 [7]. As an analog video recording means it provided higher quality than the VHS magnetic tapes. But the Laserdisc did not achieve the same success as the VHS tapes, partially because it was a read-only medium. Based on this first optical disc Philips developed a new type of disc: the compact disc (CD). The optical read-out mechanism was very similar tot that of the Laserdisc but the disc size was decreased to 12 cm, the data was recorded in a digital form. Moreover cheap semiconductor lasers made it possible to miniaturize the optical pick-up head, which reduced the cost of the drive. The wavelength λ is 780 nm. The CD became a success and so did its successor, the digital versatile disc (DVD), which was presented as a common standard in 1995, by a group of companies including Sony, Philips, Toshiba and

Matsushita [8]. This system has an operation wavelength of 650 nm. Which will be the successor in the third generation is still under debate. At the time of writing there are two competing formats both using blue light with a wavelength of 405/nm. The HD-DVD, which is backed by NEC and Toshiba, uses the same optics as the DVD. The blu-ray disc (BD) format on the other hand, is backed by Philips, Sony and a range of other companies, and use a very high numerical aperture lenses combined with blue light. The HD-DVD has the promise to be cheaper and has better backward-compatibility with the DVD. The blu-ray disc on the other hand has a higher data capacity and a broader support from industrial companies. Which format will eventually form the third generation optical disc drive depends however much more on intellectual property, marketing and money than on the technological differences.

1.2.2 Working principles

The working principles of the three generations of optical disc drives, CD [9], DVD [10], and BD [11] are identical. Each generation has however a higher capacity, and a faster read-out speed than the previous one.

Read-out Fig. 1.1a shows the rotating optical disc and the pick-up head. A laser illuminates a diffraction limited spot on the disc (FWHM $= \lambda/(2NA)$)[3]. Using a semi-transparent mirror the reflected light is guided to the detector. Fig. 1.1b and Fig. 1.1c show respectively a side and a top view of a detail on the disc. The disc consists of tracks with lands and pits forming a bit sequence. If the laser spot is focussed over a pit, the total reflection is zero as the pits are $\lambda/(4n)$ deep and the laser spot covers only half of the track. If the spot covers a land, the detector receives a reflection signal from the disc.

In chapter 3 we will describe extensively how well small details on the disc can be measured with this optical system. We will explain that because of the diffraction limit, a spatial period smaller than $\lambda/(2NA)$ can not be measured. This means that the track pitch should be larger than $\lambda/(2NA)$ and the minimum pit length larger than half this period, $\lambda/(4NA)$. This is the case for all three generations. For the compact disc system the margin is however much larger than for the blu-ray disc. As shown in table 1.1, the track pitch of the CD system is 1.8 times

[3] FWHM: full width at half maximum, λ: the wavelength, NA: the numerical aperture.

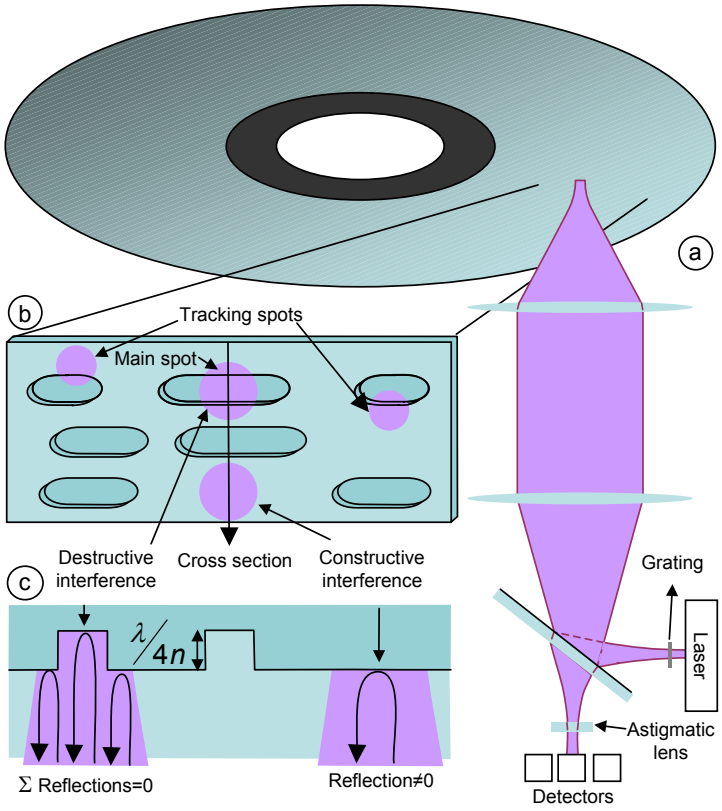


Figure 1.1: A: A laser illuminates on spot on the rotating optical disc. Through a semi-transparent mirror the reflected light is captured by a detector. B: The pits on the disc are $\lambda/(4n)$ deep. C: The spot covers half of the pits. Depending on the bit sequence this generates destructive and constructive at the detector side.

larger than the theoretical minimum distance. For the DVD system this distance is reduced to 1.37 times, and for the blu-ray disc, the track pitch is only 1.33 times larger than the diffraction limit.

The tolerance to tilt variations scales proportional to NA^3 and inversely proportional to the thickness of the cover layer. Therefore the distance of the pits to the irradiated surface has been decreased from 1.2 mm to 0.1 mm, as shown on Fig. 1.2. This means that scratches on the surface will have a bigger impact on the BD than on a CD. A 2 µm thick, scratch resistant coating on top of the cover layer and a better error correction scheme must compensate for this.

	CD	DVD	BD
Wavelength (λ)	780 nm	650 nm	405 nm
Numerical Aperture (NA)	0.45	0.60	0.85
$\lambda/(2NA)$	0.87 µm	0.54 µm	0.24 µm
Free working distance	1.2 mm	1.0 mm	$0.05 - 0.1$ mm
Cover layer	1.2 mm	0.6 mm	0.1 mm
Track pitch	1.6 µm	0.74 µm	0.32 µm
Shortest pit length	3T=0.84 µm	3T=0.40 µm	2T=160/149/138 nm
Modulation code efficiency	47%	50%	67%
ECC + format efficiency	60%	85%	82%
Total efficiency	28%	42%	55%
Surface containing data	86 cm^2	87.6 cm^2	87.6 cm^2
Netto data capacity per layer	650 MByte	4.7 GByte	23.3/25/27 GByte
Netto data transfer rate ($1\times$)	1.47 Mbps	11 Mbps	36 Mbps

Table 1.1: Three generations of optical discs: the compact disc (CD), the digital versatile disc (DVD), and the blu-ray disc (BD).

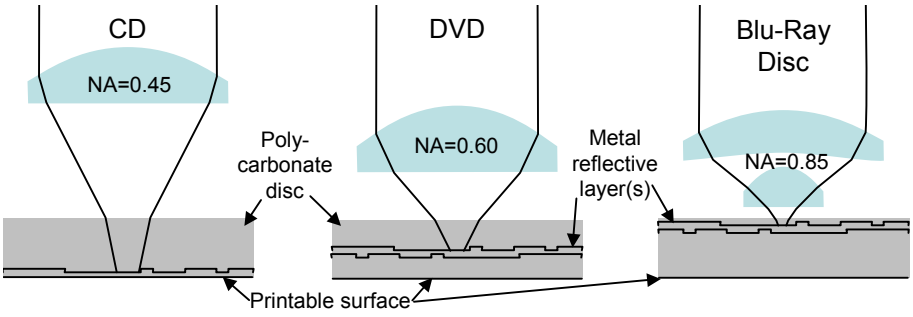


Figure 1.2: Three generations of optical discs: CD, DVD, and BD. The numerical aperture is increased from 0.45 over 0.6 to 0.85. The cover layer thickness is decreased from 1.2 mm over 0.6 mm to 0.1 mm. Remark that because of the decreased depth of field the DVD and Blu-ray disc can have multiple layers of bit patterns.

Focusing and tracking The depth of field is the range in which the object may move along the optical axis, without blurring the image. In the optical disc detection system it is approximately $2\lambda/NA^2$, which means that the flying height of the lens should be controlled to the submicrometer level. The same applies to the control of the lateral movement. The deviation of the laser spot along the track should be less than a tenth of a micron.

There are different techniques to do this. One of them, very often used for the compact disc, is shown on Fig. 1.1 [10]. A grating in the illumination path forms three spots on the disc. The central spot (the zeroth order diffraction at the grating) follows the tracks and returns

the bit signal. The two side spots (first order diffraction at the grating) are focussed on the left and on the right of the track at a different position (Fig. 1.1b). The three spots are detected by three separate detectors on a row. The signal from the outer detectors captures the side spots. The average difference in between these two signals gives a tracking signal. By keeping this signal zero, the main spot is kept on track, the reflection is captured on the central detector and results in the bit signal.

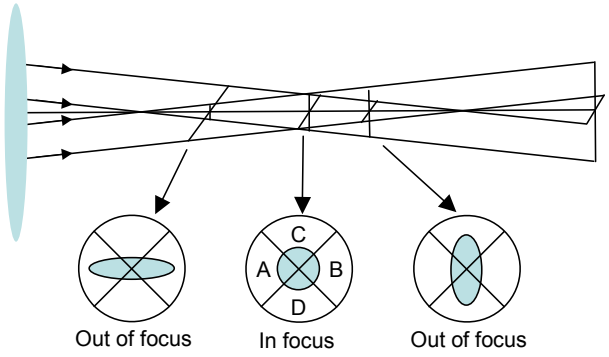


Figure 1.3: Focussing in a CD system, based on astigmatism of cylindrical lens. A cylindrical lens in the optical return path has a different focus along the vertical and the horizontal axes. If the objective lens is too far or too close from the disc, the image on the quadrant detector will be an ellipse. By adding and subtracting the quadrants $(A+C)-(B+D)$ a focussing signal is obtained.

For the focussing a supplementary astigmatic lens is placed in the optical return path. As shown on Fig. 1.3 the focus distance along the y-axis is different than along the x-axis. The central detector is split into four quadrants and is placed just in between the two foci. If the disc is in focus, a circular spot will be formed on the detector. Otherwise the spot will be elliptical. By adding and subtracting the power detected in the quadrants as: (A+C)-(B+D), a focussing signal is retrieved. The bit signal is retrieved by the total power (A+B+C+D), which is not influenced by the astigmatism.

Other techniques for tracking are the push-pull method, used in e.g. DVD-RAM and the differential phase detection used in e.g. DVD-RW.

Coding and bit detection The current optical discs are coded with run length limited (RLL) codes such as eight-to-fourteen modulation (EFM) for the compact disc, EFMplus for the DVD and a (1,7)RLL cou-

pled with partial response maximum likelihood (PRML) signal processing in the blu-ray disc [12, 13]. These mechanisms code a block of n data bits, within a block of more than n bits on the disc, to assure a minimum and maximum on the number of adjacent identical bits. This means that the shortest pit length, or the shortest space in between two pits, equals three bits for the CD and DVD and two bits for the BD. The time unit defined by the length of one bit divided by the speed of the disc, is named T. The shortest period in the detector signal will be twice the shortest pit length, i.e. $6T$ for the CD and DVD system and $4T$ for the BD system.

The bits on the disc are now retrieved from the detector signal by checking when this signal passes a certain threshold value. The method does not work when the minimal period is smaller than the diffraction limit: $6T < \lambda/(2NA)$. Even above the diffraction limit noise, crosstalk and inter-symbol interference will shift the time when the signal passes the threshold. This shift is called jitter. This jitter can be made clear by plotting the detector signals of all possible bit sequences on one graph, as shown on Fig. 1.4.

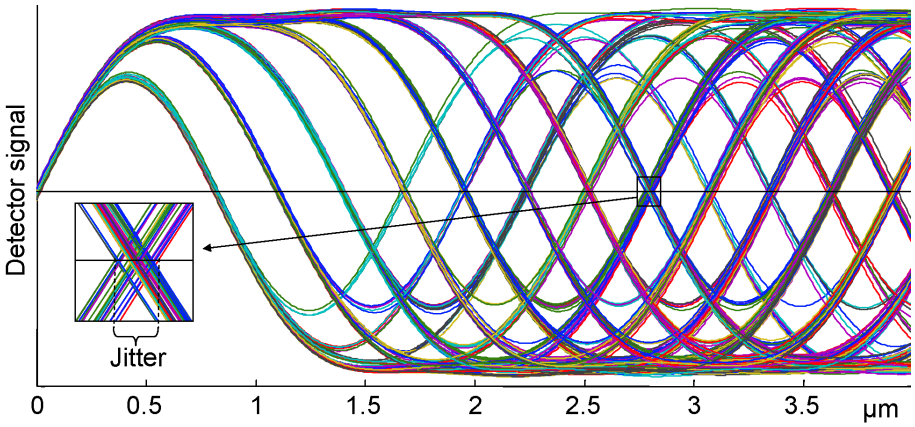


Figure 1.4: Eye patterns of the CD detector signal. The detector signals for all possible bit sequences are plotted on top of each other. The variation in the x-coordinate of the point where the signals cross a certain threshold level gives the jitter.

To reduce the bit errors, the data bits are coded on the disc with an error-correction mechanism. For the CD and DVD system this is the cross-interleaved Reed-Solomon (CIRC) detection. For BD one uses a long distance code coupled with a burst indicator subcode (LDC/BIS). As the BD has a very thin cover layer, scratches or fingerprints might

influence the detector signal. Therefore the error correcting codes operate on a large block of bits.

1.2.3 Future of optical data storage

For future optical data storage beyond the BD system there are different open questions. The first question is "Why?". The driving force behind the CD was digital music, the driving force behind DVD movie distribution and BD is seen as the new format for high definition television (HDTV). At the moment there is no clear consumer product that would request a new generation of optical discs. From a historical point of view it is however very unlikely that at a given time there will not be any need for bigger data capacity on the same disc, nor any need for an increased data rate. Moreover there is a second incentive for further research, which is miniaturization. The current optical discs have a 12 cm diameter. For mobile devices one would like to shrink the discs as well as the disc drives such that it fits in e.g. a compact digital video camera. If one would like to store the same information content on these smaller discs, the data density needs to be increased. To increase this data density different techniques are investigated. We have grouped them together into a few categories:

- A first method is to go further on the path from CD over DVD and BD: increasing the numerical aperture and/or decreasing the wavelength. An increase in numerical aperture can either be done by working in the near-field [14] or by using solid immersion lenses [15]. With this last technique an immersion oil with a high refractive index is placed in between the objective lens and the disc. As the numerical aperture scales proportional with this refractive index, the resolution limit can be improved. With these techniques one can raise the numerical aperture above 1. The main disadvantages of these approaches is that removability of the disc is either lost or very difficult.

- Totally different are techniques in which the disc is adapted such that even with a larger spot smaller details can be read-out. One of these methods is the Magnetic AMplifying Magneto-Optical System (MAMMOS) [16]. Above a certain intensity threshold, the magnetic field in the storage layer is expanded over a larger area in the conversion layer. Because the threshold intensity is only reached in the small centre area of the optical spot, small features

can be read out. Another method is the super resolution near-field structure (super RENS) [17, 18]. In this layered structure the laser beam opens a small aperture in the topmost layer at the point of highest light intensity. Very small features in the layer below can then be read out.

- Another technique is volumetric storage. This can be done by storing the information onto different layers which can each be read out by focussing the objective lens correctly. This technique has already been implemented in dual-layer DVD, and will also be implemented in the BD. Care should be taken that the axial point spread function is sufficiently narrow to prevent interference in between the different layers [19, 20]. The third dimension can also be explored by holographic storage [21, 22]. With this technique the 'information planes' are stored by using a different reference beam.

- Storing more than one bit in a feature sized above the diffraction limit, can also increase the storage capacity many times. One technique to do so is detecting different 'gray-levels' instead of only detecting zero and one [1, 2, 23].

- By post-processing the retrieved signals, one can compensate the influence of neighbouring bits. Methods such as 'TWO Dimensional Optical Storage' (TWODOS) [24, 25, 26] or Matrix-Method Deconvolution (MMD) [27] do this on a two-dimensional scale.

Our method is to retrieve as well the phase as the amplitude of the light reflected from the disc. Combined with a parameter extraction method the bits can be spaced closer together for a given bit error rate. This method will be outlined in section 1.3

1.3 The waveguide detection method

In this thesis we investigated how the data density on optical disc can be improved by using a multimodal waveguide as a pick-up means. Fig. 1.5 gives a schematic overview of the detection system. A laser source excites one or more modes of a multimodal waveguide through a photonic integrated circuit. These modes are imaged onto the disc, either with a lens system in the far-field or by free space propagation in proximity of the disc. The light reflected from the disc is captured

by the modes of the same multimodal waveguide. The light captured by these modes is led via different output waveguides to the detectors. By inspecting the optical disc through two parallel channels, the bit patterns on the tracks can be reconstructed.

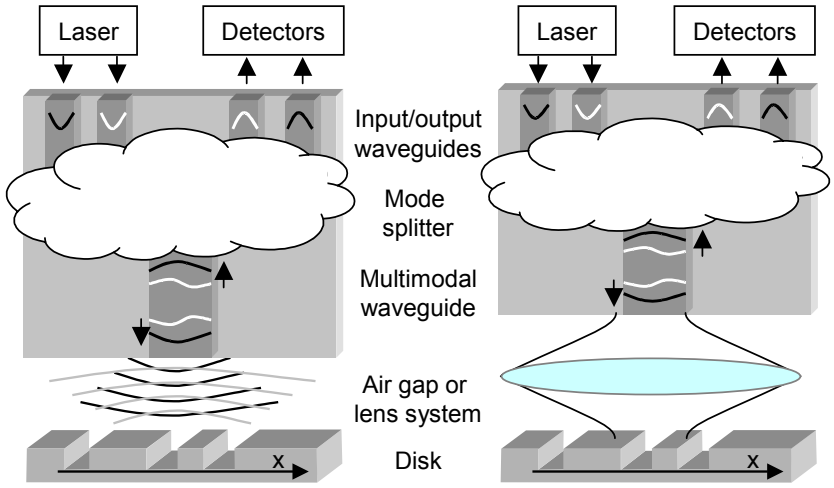


Figure 1.5: Schematic view of the waveguide detection system. The picture shows the near-field version on the left and the far-field version on the right. In both cases the disc is illuminated with one or more waveguide modes. The same waveguide modes capture the reflected light. A photonic integrated circuit connects the individual modes through input and output waveguides with the laser and detectors.

The rationale behind the approach is that by using a multimodal waveguide as a pick-up head, both the phase and the amplitude of the reflected light can be recovered. This in contrast to an ordinary detector, which detects only the intensity of the light, which is the square of the amplitude. This phase and amplitude information are recovered in parallel by analysis of the excitations of the different modes in the multimodal waveguide. This capability of detecting phase and amplitude has been described in similar experiments with multimode fibres [28, 29]. In this thesis we combine the multimodal waveguide with a photonic integrated circuit which gives a much stronger control over the individual modes.

A second rationale is miniaturization. In principle all optical functions, including lasers and detectors could be integrated in a single chip, with the potential to decrease the size and costs and improve the robustness of the component.

To combine the signals from the different modes a simple parameter extraction method has been designed. These simulations showed that reconstructing the bit pattern with both waveguide modes can lead to a lower bit error rate than would be obtained with the conventional pick-up heads.

1.4 Outline of this work

The analysis of the waveguide detection method as a means to increase the data density of optical discs has been described in different chapters. In this introduction we gave an overview of current data storage techniques and described the working principles of the proposed waveguide detection method. In chapter 2 the principles behind optical resolution in general will be outlined. Starting from the 19th century Rayleigh criterion to current concepts as model-based resolution. The resolution of the scanning microscope will be extensively analysed in chapter 3. This analysis includes the conventional optical disc readout, the confocal microscope and the waveguide detection method. The parameter extraction method to retrieve the bits on the disc will be explained in chapter 4. An analysis of the bit error rate as a function of the minimum bit length and noise influences will show how a combination of the different modes can lead to an improved performance. Chapter 5 discusses the principles and design of a photonic IC connecting the input and output waveguides with the individual waveguide modes. The measurement setup and experimental results are described in chapter 6. Chapter 7 concludes with a recapitulation of the properties of the waveguide detection method, the possible drawbacks, and the potential future applications.

1.5 Publications

The results obtained within this work have led to a number of publications.

International Journals

1. F.W.J. Fransoo, D. Van Thourhout, and R. Baets, Performance analysis of a multi-mode waveguide based optical disc readout system, Applied Optics, 43(17): 3480–3488, Jun. 2004.

International Conference Proceedings

2. S. Musa, S.F. Pereira, J.J.M. Braat, F.W.J. Fransoo, D. Van Thourhout, and R. Baets, Sub-wavelength detection with waveguides in optical recording, *SPIE proceedings*, 5380: 723–728, Sep. 2004.

3. S. Musa, S.F. Pereira, J.J.M. Braat, F.W.J. Fransoo, D. Van Thourhout, and R. Baets, Sub-wavelength detection with waveguides in optical recording, *Optical Data Storage Topical Meeting 2004*, Monterey , United States: 156–158, Apr. 2004.

4. F.W.J. Fransoo, D. Van Thourhout, L. Van Landschoot, S. Verstuyft, and R. Baets, A multimodal waveguide for enhanced performance in optical disc read-out, *EOS Topical Meeting in Advanced Imaging Techniques 2003*, Delft, The Netherlands: 115–117, Oct. 2003.

5. F.W.J. Fransoo, D. Van Thourhout, L. Van Landschoot, A. Verbiest, W.F.H. Van Parys, P. Van Daele, and R. Baets, A method for detecting sub-wavelength features by means of a multimode waveguide and a mode splitting photonic IC, *LEOS Annual 2003*, Glasgow, United Kingdom: 750–751, Nov. 2002.

National Conference Proceedings

6. F.W.J. Fransoo, D. Van Thourhout, L. Van Landschoot, A. Verbiest, W.F.H. Van Parys, P. Van Daele, and R. Baets, Using a multimodal waveguide for enhanced resolution in optical data storage, *FTW-symposium 2002*, Gent, Belgium: p 25, 2002.

7. F.W.J. Fransoo, P. Bienstman, and R. Baets, Photonic IC for optical detection with sub wavelength resolution, *IEEE/LEOS Benelux Chapter 2001*, Brussels, Belgium, pp. 205-208, Dec. 2001.

Chapter 2

Optical resolution

In imaging resolution has always been an important issue, but it is not unambiguously defined and has many interpretations. In this chapter we will give a brief overview of some concepts in resolution. We will first describe the properties of an optical system for coherent, incoherent, and partially coherent light. In the next section we will give a brief overview of some classical resolution enhancement techniques. Finally we will describe the influence of noise. The analysis of the resolution in a scanning microscope is a special case and will be dealt with in chapter 3.

2.1 Introduction

A traditional definition of the resolution of an optical system operating in the far-field is the smallest distance between two points, of which the image by this optical systems still shows two resolved points [30]. The most famous criterion is the Rayleigh criterion, which states:

> *Two point sources are just resolved if the central maximum of the intensity diffraction pattern produced by one point source coincides with the first zero of the intensity diffraction pattern produced by the other.*

For two totally incoherent light sources and a point spread function (spatial impulse response) related to an aberration-free circular lens with numerical aperture NA, this results in a 'resolution' of $\lambda/0.601NA$. Another classical definition is the Sparrow limit, which defines that two points are resolved if the second order derivative of the composite intensity distribution becomes zero at the centre between the images of

two point sources. This means the points are resolved as soon as there is a dip in the intensity in between the images of the two points, which gives $\lambda/0.47NA$. These two definitions are simple and have much practical use in astronomy, where the stars can be seen as mutually incoherent point sources. They have however a limited validity for other applications, as they only look to two isolated point sources in the far-field and disregard coherence and noise.

2.2 Point spread function and optical transfer function

Another way to measure how well small features can be resolved by an imaging system is the point spread function and the optical transfer function. Before analysing these functions we will firstly describe two assumptions: linearity and shift-invariance.

2.2.1 Linearity and shift-invariance

Assume an optical system as on Fig. 2.1, which images the optical field in the object plane on the left to the image plane on the right. In general a point source with amplitude A, $A\delta(x_0, y_0)$, will be imaged onto a larger spot $A \cdot K(x_0, y_0; x_1, y_1; A)$ in the object plane. This image of a point source will be called the transmission function. An imaging system is called linear if the transmission function does not depend on the amplitude of the point source: $K(x_0, y_0; x_1, y_1; A) = K(x_0, y_0; x_1, y_1)$. If the imaging systems is also shift-invariant this can be further simplified to: $K(x_0, y_0; x_1, y_1; A) = K(x_1 - x_0, y_1 - y_0)$. An imaging system is only shift-invariant if there is no magnification and if the origin is laying on the optical axis. An extension to imaging systems with a demagnification factor γ is however trivial, as will be shown in section 2.3.

This linearity and shift-invariance do not hold for all optical systems, and complicates a correct definition of optical resolution. Nonlinear effects are sometimes used to increase the resolution as for example in magneto-optical discs [16] or the super-RENS disc [17, 18]. Lenses with aberrations such as coma or astigmatism are examples of non-shift-invariant imaging systems. In the remainder of the text we will however focus on linear, shift-invariant imaging systems.

These imaging systems can be fully described by a point spread function and an optical transfer function. In the next paragraphs we will explain how these functions can be derived from the transmission

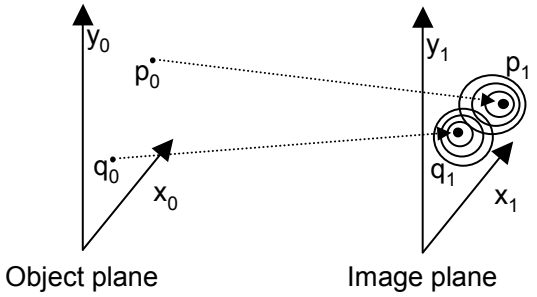


Figure 2.1: In a linear system a point source with a certain amplitude in the one plane (the object plane) is imaged onto a larger spot in an other plane (the image plane). When two point sources are imaged, the respective point spread functions can be combined. The nature of this combination depends however on the coherence of the original point sources.

function K. To keep the notations simple, the optical fields are written as a scalar value.

2.2.2 Point spread function

As shown on Fig. 2.1 each point source in the object plane gives a broader spot in the image plane. These spots can be combined to construct the complete image. The nature of this addition depends however on the coherence of the original point sources.

If all the point sources in the object plane have a fixed phase relationship, which is the case for fully monochromatic light, the light is called coherent and one can find the total image by adding up the complex field of the images of the individual points:

$$\psi_1(x_1, y_1) = \iint psf_{coh}(x_1 - x_0, y_1 - y_0)\, \psi_0(x_0, y_0)\, \mathrm{d}x_0\, \mathrm{d}y_0 \quad \text{with } psf_{coh}(x, y) = K(x, y)\,. \tag{2.1}$$

If the phase relationships in between the point sources vary randomly, which is the case for white light, the light is called fully incoherent and one should not add up the complex fields but the intensities:

$$|\psi_1(x_1, y_1)|^2 = \iint psf_{incoh}(x_1 - x_0, y_1 - y_0)\, |\psi_0(x_0, y_0)|^2\, \mathrm{d}x_0\, \mathrm{d}y_0 \quad \text{with } psf_{incoh}(x, y) = |K(x, y)|^2\,. \tag{2.2}$$

In most cases the optical spectrum is nor infinite nor infinitesimal but has a certain width. In this case light is said to be partially coherent. If we write the complex optical field in a point $P(x, y)$, as $\psi(x, y; t)$, and similarly as $\psi(x', y'; t)$ in the point $P'(x', y')$, the mutual intensity, also called visibility, is defined as $J(x, y; x', y') = \langle \psi(x, y; t)\, \psi^*(x', y', t) \rangle$ where the brackets indicate a time averaging [31, Chapter 10]. The fully coherent and the fully incoherent case are extremes of this partially coherent case.

Fully coherent light is monochromatic, which means that the optical field $\psi(x, y; t)$ can be written as $\psi(x, y) \exp(-j\omega t)$.

$$J\big(x, y; x', y'\big) = \big\langle \psi(x, y)\, \psi^*\big(x', y'\big) \big\rangle = \psi(x, y)\, \psi^*\big(x', y'\big)\,. \tag{2.3}$$

For the fully incoherent case, the phase difference between two point sources varies randomly in time. Which means that the time average is only non-zero for a product between twice the same point:

$$J\big(x, y; x', y'\big) = \psi(x, y)\, \psi^*\big(x', y'\big)\, \delta\big(x - x', y - y'\big) \tag{2.4}$$

$$= |\psi(x, y)|^2\, \delta\big(x - x', y - y'\big)\,. \tag{2.5}$$

For partially coherent light one can also define a point spread function, by adding up these mutual intensities.

$$J_1\big(x_1, y_1; x_1', y_1'\big) = \iiiint PSF\big(x_1 - x_0, y_1 - y_0; x_1' - x_0', y_1' - y_0'\big) \times J_0\big(x_0, y_0; x_0', y_0'\big)\, \mathrm{d}x_0\, \mathrm{d}y_0\, \mathrm{d}x_0'\, \mathrm{d}y_0' \tag{2.6}$$

$$\text{with } PSF\big(x, y; x', y'\big) = K(x, y)\, K^*\big(x', y'\big)$$

This can be understood as follows: We omit the y-coordinate, and assume two point sources in the object plane, each with a different time dependency: $\psi(x_0, y_0)\, f(t)$, and $\psi(x_0', y_0')\, g(t)$. These point sources are imaged onto two spots in the image which can respectively be written as:

$$K(x_1 - x_0, y_1 - y_0)\, \psi(x_0, y_0)\, f(t)\,, \tag{2.7}$$

$$\text{and } K\big(x_1' - x_0', y_1' - y_0'\big)\, \psi\big(x_0', y_0'\big)\, g(t)\,. \tag{2.8}$$

The mutual intensity in the image plane can then be written as:

$$\begin{aligned} J\big(x_1, y_1; x_1', y_1'\big) &= \psi(x_0, y_0)\, \psi^*\big(x_0', y_0\big)\, \langle f(t)\, g^*(t) \rangle \\ &\quad \times K(x_1 - x_0, y_1 - y_0)\, K^*\big(x_1' - x_0', y_1' - y_0'\big) \\ &= J\big(x_0, y_0; x_0', y_0'\big) \\ &\quad \times PSF\big(x_1 - x_0, y_1 - y_0; x_1' - x_0', y_1' - y_0'\big) \end{aligned} \tag{2.9}$$

2.2.3 Optical transfer function

For a linear shift-invariant optical system it is possible to define the frequency response of the optical system: the optical transfer function (OTF). For the fully coherent and the fully incoherent case this is the Fourier transform of the point spread function. For partially coherent light this is slightly different. To extend the definition to partially coherent light we introduce some definitions from Appendix A. With the definition of the Fourier transform as in section A.4, we find:

$$\widehat{\psi}(f) = \mathcal{F}[\psi(x)](f) = \int \psi(x) \exp(-2\pi j f x)\,\mathrm{d}x. \tag{2.10}$$

We also use a new operator defined in section A.5 which is very similar to the Fourier transform:

$$\widetilde{\Psi}(f,g;f',g') = \iiiint \exp\left[-2\pi j\,(fx+gy-f'x'-g'y')\right] \times \Psi(x,y;x',y')\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y'. \tag{2.11}$$

Performing the tilde-operator on left and right hand side of Eq. (2.6) one finds the partially coherent optical transfer function, $OTF = \widetilde{PSF}$.

$$\begin{aligned} \widetilde{J_1}(f,g;f',g') &= OTF(f,g;f',g')\,\widetilde{J_0}(f,g;f',g') \\ \text{with } OTF(f,g;f',g') &= \widehat{K}(f,g)\left[\widehat{K}(f',g')\right]^* \end{aligned} \tag{2.12}$$

In the case of full coherence Eq. (2.12) can be transformed to

$$\begin{aligned} \widehat{\psi_1}(f,g)\left[\widehat{\psi_1}(f',g')\right]^* &= \widehat{K}(f,g)\left[\widehat{K}(f',g')\right]^* \widehat{\psi_0}(f,g)\left[\widehat{\psi_0}(f',g')\right]^* \\ \text{or} \qquad \widehat{\psi_1}(f,g) &= otf_{coh}(f,g)\,\widehat{\psi_0}(f,g) \\ \text{with } otf_{coh}(f,g) &= \widehat{psf_{coh}}(f,g) \end{aligned} \tag{2.13}$$

In the case of complete incoherence one can write:

$$\begin{aligned} \widehat{|\psi_1|^2}(f,g) &= \mathcal{F}[K(x,y)\,K^*(x,y)]\,\widehat{|\psi_0|^2}(f,g) \\ \text{or} \qquad \widehat{I_1}(f,g) &= otf_{incoh}(f,g)\,\widehat{I_0}(f,g) \\ \text{with } otf_{incoh}(f,g) &= \widehat{psf_{incoh}}(f,g) = \widehat{K}(f,g) \otimes \widehat{K}(f,g)\,. \end{aligned} \tag{2.14}$$

This last equation shows that the relationship in between the coherent and the incoherent otf is a convolution, which is as expected.

While the width of the point spread function gives a measure for how well closely spaced objects can be discriminated, the optical transfer function gives a measure on how well spatial frequencies are preserved from the object to the image plane. The amplitude of the OTF shows how well a spatial frequency is transmitted. Many optical systems have a so called cut-off frequency, above which the OTF is zero. This means that features with a spatial frequency higher than this cut-off frequency are lost in the image plane. The phase of the OTF indicates how the different frequency components interfere. The coherent OTF of a defocussed lens has the same amplitude as that of a lens in focus, but the image is blurred because of the phase lag between the different spatial frequencies.

One could say that while the coherent transfer function operates linearly on the Fourier transform of the optical field (complex amplitude), the incoherent transfer function operates linearly on the Fourier transform of the optical intensity. For partially coherent light, the interpretation is more complex. We will return to this issue in section 3.1.5, where the description of the optical properties of a scanning microscope will be described by a partially coherent PSF and OTF. The reason is that the scanning microscope combines elements that operate linearly with the optical field (such as the reflection of the disc), and elements that operate linearly with the optical intensity (such as a detector).

2.3 Examples

In this section we will derive the point spread function and optical transfer function for coherent light for some typical systems: a simple lens and free space propagation. We will use these equations later on in chapter 3. The equations for the incoherent, or the partially coherent cases can easily be derived from the coherent case:

$$\begin{aligned} psf_{incoh}(x,y) &= |psf_{coh}(x,y)|^2 \\ otf_{incoh}(f,g) &= otf_{coh}(f,g) \otimes otf_{coh}(f,g) \\ PSF_{part.coh}(x,y;x',y') &= psf_{coh}(x,y)\, psf_{coh}{}^*(x',y') \\ OTF_{part.coh}(f,g;f',g') &= otf_{coh}(f,g)\, otf_{coh}{}^*(f',g') \end{aligned} \tag{2.15}$$

In the next chapter we will use 'psf' and 'otf' with lower case letters for the coherent case. 'PSF' and 'OTF' with capital letters will be used for the partially coherent functions.

2.3.1 Imaging by a thin lens

In general the design and use of lenses can be very complex. Especially for lenses with a high numerical aperture. We will not give a fully rigorous calculation but show the results for a thin lens in the paraxial approximation.

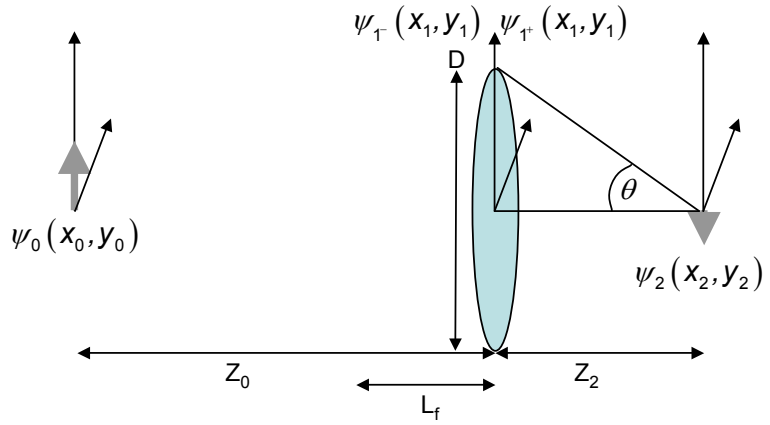


Figure 2.2: The imaging by a thin lens from the object plane (x_0, y_0), to the image plane (x_2, y_2), can be seen as a demagnification of the optical field in the object plane, ψ_0, with a factor $-z_0/z_2$ followed by a convolution with the point spread function of the lens, psf_{lens}.

A thin lens with focal length L_f and pupil function $p_{lens}(x, y)$ has a transmission function:

$$t_{lens}(x, y) = p_{lens}(x, y) \exp\left[j \frac{\pi}{L_f \lambda} \left(x^2 + y^2\right)\right]. \tag{2.16}$$

If the diameter of the optical field along the x- and y-axis, is much smaller than $\sqrt{\lambda z_0}$ and $\sqrt{\lambda z_2}$, we can apply the Fresnel approximation for diffraction. If we calculate the optical field as shown on Fig. 2.2, we

find:

$$\psi_{1^-}(x_1, y_1) = \frac{-\exp(-j\lambda z_0)}{j\lambda z_0} \exp\left[\frac{-j\pi}{\lambda z_0}\left(x_1^2 + y_1^2\right)\right] \times \iint \psi_0(x_0, y_0) \exp\left[\frac{j\pi}{\lambda z_0}\left(x_0^2 + y_0^2\right)\right] \times \exp\left[\frac{-j2\pi}{\lambda z_0}\left(x_0 x_1 + y_0 y_1\right)\right] \mathrm{d}x_0\, \mathrm{d}y_0 \quad (2.17)$$

$$\psi_{1^+}(x_1, y_1) = \psi_{1^-}(x_1, y_1)\, p_{lens}(x_1, y_1) \exp\left[j\frac{\pi}{L_f \lambda}\left(x_1^2 + y_1^2\right)\right] \quad (2.18)$$

$$\psi_2(x_2, y_2) = \frac{-\exp(-j\lambda z_2)}{j\lambda z_2} \exp\left[\frac{-j\pi}{\lambda z_2}\left(x_2^2 + y_2^2\right)\right] \times \iint \psi_{1^+}(x_1, y_1) \exp\left[\frac{-j\pi}{\lambda z_2}\left(x_1^2 + y_1^2\right)\right] \times \exp\left[\frac{j2\pi}{\lambda z_2}\left(x_1 x_2 + y_1 y_2\right)\right] \mathrm{d}x_1\, \mathrm{d}y_1 \quad (2.19)$$

Combining these equations one finds:

$$\psi_2(x_2, y_2) = \frac{\exp[-j\lambda\,(z_0 + z_2)]}{\lambda^2 z_0 z_2} \exp\left[\frac{-j\pi}{\lambda z_2}\left(x_2^2 + y_2^2\right)\right] \times \iint\iint \exp\left\{j\frac{2\pi}{\lambda}\left[\left(\frac{x_0}{z_0} + \frac{x_2}{z_2}\right) x_1 + \left(\frac{y_0}{z_0} + \frac{y_2}{z_2}\right) y_1\right]\right\} \times p_{lens}(x_1, y_1) = \exp\left[j\frac{\pi}{\lambda}\left(\frac{1}{L_F} - \frac{1}{z_0} - \frac{1}{z_2}\right)\left(x_1^2 + y_1^2\right)\right] \times \exp\left[\frac{-j\pi}{\lambda z_0}\left(x_0^2 + y_0^2\right)\right] \psi_0(x_0, y_0)\, \mathrm{d}x_0\, \mathrm{d}y_0\, \mathrm{d}x_1\, \mathrm{d}y_1 \quad (2.20)$$

If the object plane, the image plane and the lens are positioned such that $\frac{1}{L_f} = \frac{1}{z_0} + \frac{1}{z_2}$ and with the definition

$$psf_{lens}(x,y)\,\frac{1}{(\lambda z_2)^2}\widehat{p_{opt}}\left(\frac{x}{\lambda z_2},\frac{y}{\lambda z_2}\right), \tag{2.21}$$

Eq. (2.20) can be simplified to:

$$\begin{aligned}\psi_2(x_2,y_2) = &-\exp[-j\lambda\,(z_0+z_2)]\,\frac{z_2}{z_0}\exp\left[\frac{-j\pi}{\lambda z_2}\left(x_2^2+y_2^2\right)\right]\\ &\times\iint psf_{lens}\left[\frac{z_2x_0}{z_0}+x_2,\frac{z_2y_0}{z_0}+y_2\right]\\ &\times\exp\left[\frac{-j\pi}{\lambda z_0}\left(x_0^2+y_0^2\right)\right]\psi_0(x_0,y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0.\end{aligned} \tag{2.22}$$

The phase factors can be disregarded as we assumed that x_0 and y_0 are much smaller than $\sqrt{\lambda z_0}$ and $\sqrt{\lambda z_2}$. Which means the equation can further be simplified to:

$$\psi_2(x_2,y_2) = -\frac{z_0}{z_2}\psi_0\left(-\frac{z_0}{z_2}x_2,-\frac{z_0}{z_2}y_2\right)\otimes psf_{lens}(x_2,y_2)\,. \tag{2.23}$$

In this equation $-\frac{z_0}{z_2}$ is the linear demagnification factor, which we will write as γ.

Because of the paraxial approximations the above equations are only valid for lenses with a small diameter compared to z_0 and z_2. For larger lenses one needs to use the numerical aperture NA, defined as the refractive index of the medium multiplied by the sine of the angle θ, shown on Fig. 2.2. θ is the half of the angle at which the lens is seen from the centre of the image plane. For a refractive index n, and a lens diameter D this means: $NA = n\frac{D/2}{\sqrt{z_2^2+D/2}}$. In the rest of this text, NA or NA_d will be used to describe the numerical aperture at the disc side. The numerical aperture at the waveguide or detector, will be written as $NA_a = \frac{NA}{|\gamma|}$.

$psf_{lens}(x,y)$ is the coherent point spread function of a lens. For a square lens geometry this gives:

$$otf_{coh}(f,g) = \Pi\left(\frac{f\lambda}{2NA}\right)\Pi\left(\frac{g\lambda}{2NA}\right) \tag{2.24}$$

$$psf_{coh}(x,y) = \left(\frac{2NA}{\lambda}\right)^2\frac{\sin\left(\frac{2\pi NA}{\lambda}x\right)}{\left(\frac{2\pi NA}{\lambda}x\right)}\frac{\sin\left(\frac{2\pi NA}{\lambda}y\right)}{\left(\frac{2\pi NA}{\lambda}y\right)}. \tag{2.25}$$

Using $r = \sqrt{x^2 + y^2}$, and $h = \sqrt{f^2 + g^2}$ a circular lens geometry gives:

$$otf_{coh}(f,g) = \begin{cases} 1, & \text{if } h \leq NA \\ 0, & \text{if } h > NA \end{cases} \tag{2.26}$$

$$psf_{coh}(x,y) = \left(\frac{NA}{\lambda}\right)^2 \pi \frac{J_1\left(\frac{2\pi NA}{\lambda} r\right)}{\frac{2\pi NA}{\lambda} r}, \tag{2.27}$$

$$\text{with } J_1(x) \text{ the Bessel function of the first kind.} \tag{2.28}$$

The incoherent point spread function and optical transfer function can easily be derived from the equations for coherent light using Eq. (2.15). A square lens geometry results in:

$$otf_{incoh}(f,g) = \Pi\left(\frac{\lambda f}{2NA}\right)\left(1 - \frac{|\lambda f|}{NA}\right)\Pi\left(\frac{\lambda g}{2NA}\right)\left(1 - \frac{|\lambda g|}{NA}\right) \tag{2.29}$$

$$psf_{incoh}(x,y) = \left[\left(\frac{2NA}{\lambda}\right)^2 \frac{\sin\left(\frac{2\pi NA}{\lambda} x\right)}{\left(\frac{2\pi NA}{\lambda} x\right)} \frac{\sin\left(\frac{2\pi NA}{\lambda} y\right)}{\left(\frac{2\pi NA}{\lambda} y\right)}\right]^2. \tag{2.30}$$

Using $r = \sqrt{x^2 + y^2}$, and $h = \sqrt{f^2 + g^2}$ one finds for a circular lens geometry:

$$otf_{incoh}(f,g) = \begin{cases} \frac{2}{\pi} \arccos\left(\lambda \frac{h}{2NA}\right) - \lambda \frac{h}{2NA} \sqrt{1 - \lambda \frac{h}{2NA}}, & \text{if } h \leq 2\frac{NA}{\lambda} \\ 0, & \text{if } h > 2\frac{NA}{\lambda} \end{cases} \tag{2.31}$$

$$psf_{incoh}(x,y) = \left[\left(\frac{NA}{\lambda}\right)^2 \pi \frac{J_1\left(\frac{2\pi NA}{\lambda} r\right)}{\frac{2\pi NA}{\lambda} r}\right]^2. \tag{2.32}$$

In general the propagation between object and image plane can be seen as a simple demagnification followed by a convolution with a point spread function. In the Fourier space the convolution is translated to a multiplication.

2.3.2 Near-field propagation

In case of free space propagation it is also possible to define a point spread function. This can be done by using the Fresnel-Kirchoff diffraction integral [31, Chapter 8.4], which is founded on the Huygens-Fresnel principle saying that the propagation of a wavefront can be

found by substituting each point with a spherical wave. With $r = \sqrt{x^2 + y^2 + z^2}$ the point spread function can be written as:

$$psf_{coh}(x, y) = \frac{j}{\lambda r} \exp\left(-j\frac{2\pi}{\lambda} r\right) \frac{z}{r} \tag{2.33}$$

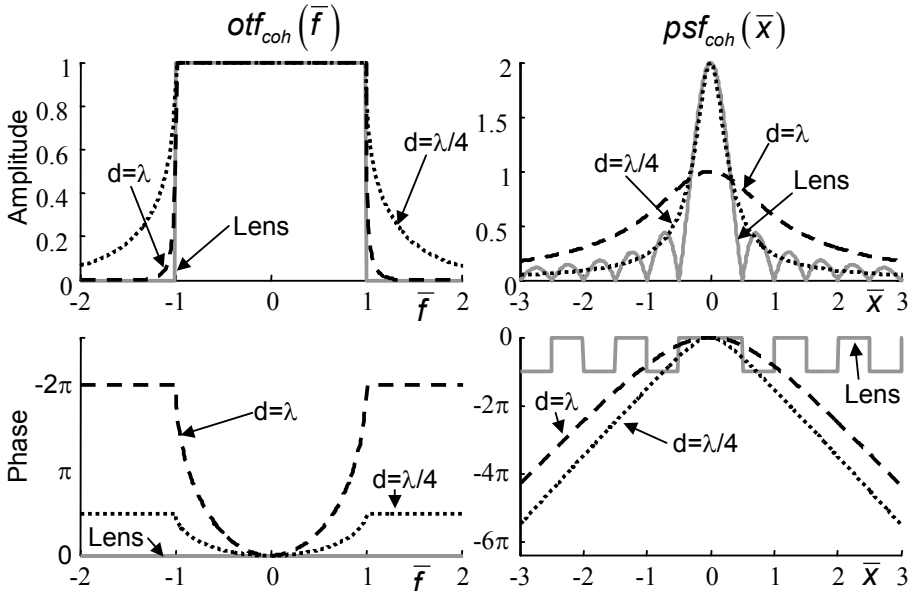


Figure 2.3: Amplitude and phase of the coherent otf and psf along the x-axis. In solid gray line for a rectangular lens, in dotted black line for near-field with $z = \lambda/4$, in dashed black line for near-field with $z = \lambda$. On the horizontal axis $\overline{x} = xNA/\lambda$ and $\overline{f} = f\lambda/NA$ for the lens system, and $\overline{x} = x/\lambda$ and $\overline{f} = f\lambda$ for the near-field.

Another approach is calculating the optical transfer function by expansion in plane waves as explained in appendix C.2. The optical transfer function can then be written as:

$$otf_{coh}(f, g) = \begin{cases} \exp\left(-j\frac{2\pi}{\lambda} z\sqrt{1 - \lambda^2 h^2}\right), & \text{if } \lambda h < 1 \\ \exp\left(-\frac{2\pi}{\lambda} z\sqrt{\lambda^2 h^2 - 1}\right), & \text{if } \lambda h > 1 \end{cases} \tag{2.34}$$
$$\text{with } h^2 = f^2 + g^2.$$

Both methods lead to the same results as can be checked by the identity $\widehat{psf}(x, y) = otf(f, g)$.

Fig. 2.3 and Fig. 2.4 compare the amplitude and the phase of respectively the coherent and incoherent psf and otf. Both graphs give

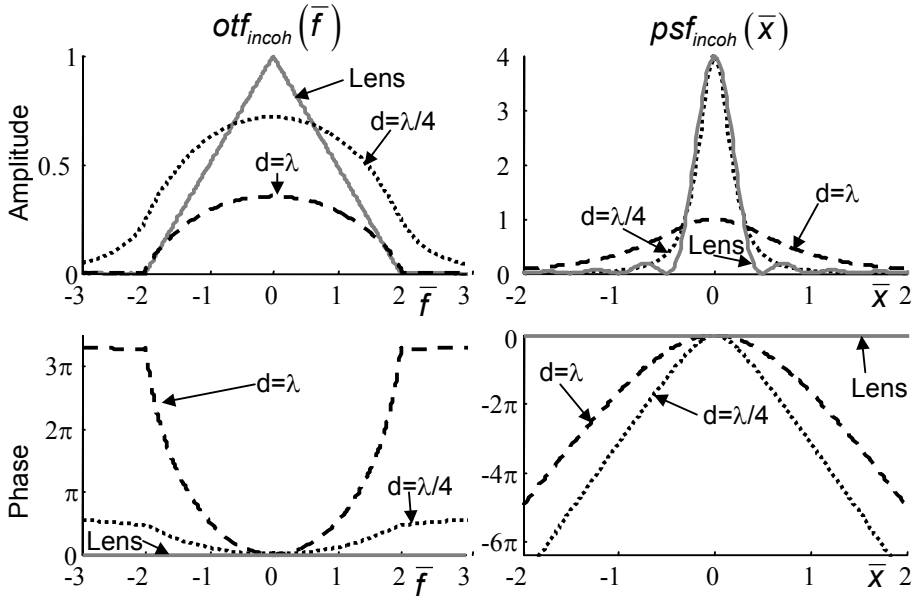


Figure 2.4: Amplitude and phase of the incoherent otf and psf along the x-axis. In solid gray line for a rectangular lens, in dotted black line for near-field with $z = \lambda/4$, in dashed black line for near-field with $z = \lambda$. On the horizontal axis $\overline{x} = xNA/\lambda$ and $\overline{f} = f\lambda/NA$ for the lens system, and $\overline{x} = x/\lambda$ and $\overline{f} = f\lambda$ for the near-field.

a comparison for a lens system and for two near-field configurations. The amplitude of the psf for a near-field gap of $\lambda/4$, and that of a lens system with $NA = 1$ are very similar. The graph for the otf shows however that the near-field configuration picks up a broader spectrum of spatial frequencies, at the expense of phase delay in between the frequencies. The graph shows that although the optical transfer function of the near-field propagation of λ transmits higher spatial frequencies, its point spread function can be broader than that of the imaging with a lens, e.g. for a gap of λ. This broad point spread function is due to the phase differences in between the spatial frequencies.

2.3.3 Imaging to Fourier plane

For later use we also present an example of a non-shift invariant imaging system. Assume the object plane is placed at the focus of the lens and the image plane just behind the lens. Starting again with Eq. (2.17) we find:

$$\begin{aligned}\psi_2(x,y) &= \psi_{1^-}(x,y)\, t_{lens}(x,y) \\ &= \frac{\exp(-j\lambda z_0)}{j\lambda z_0} \exp\left[j\pi \frac{x^2+y^2}{\lambda}\left(\frac{1}{L_f}-\frac{1}{z_0}\right)\right] \\ &\qquad \times p_{lens}(x,y)\, \widehat{\psi_0}\left(\frac{x}{\lambda z_0}, \frac{y}{\lambda z_0}\right) \end{aligned} \tag{2.35}$$

For $z_0 = L_f$ and disregarding the constant phase factor one finds:

$$\psi_2(x,y) = \frac{1}{\lambda z_0} p_{lens}(x,y)\, \widehat{\psi_0}\left(\frac{x}{\lambda z_0}, \frac{y}{\lambda z_0}\right) \tag{2.36}$$

2.4 Resolution enhancement

In principle there are two ways to enhance the resolution of a lens system. The first is optically, by changing the amplitude and/or phase transmission in the optical path, or by interfering different paths with each other. The second is by making use of the foreknowledge of the point spread function and/or the features in the object plane.

2.4.1 Optical resolution enhancements

An example to show a problem with the classical two point resolution limit is apodization. It is possible to improve the resolution in the Rayleigh sense by blocking parts of the lens transmission. Fig. 2.5 shows how the width of the point spread function of a lens can be reduced by blocking the central part of the lens. If $1/8$ of the surface is blocked, the Rayleigh and the Sparrow distance are reduced to $0.47NA/\lambda$ and $0.425NA/\lambda$. By blocking a quarter of the lens aperture this becomes: $0.39NA/\lambda$ and $0.375NA/\lambda$. This gain in 'resolution' has as drawback that the overall power decreases and that the relative amplitude of the side lobes increases. With a different stop aperture it is also possible to work the other way around and reduce the amplitude of the side lobes at the expense of a broader main lobe. In general apodization achieves an improvement of certain aspects at the expense of others.

With more complex filters on top of the lens aperture it is in principle possible to reduce the width of the main lobe to any given value at the expense of very high side lobes. Using interference in between two

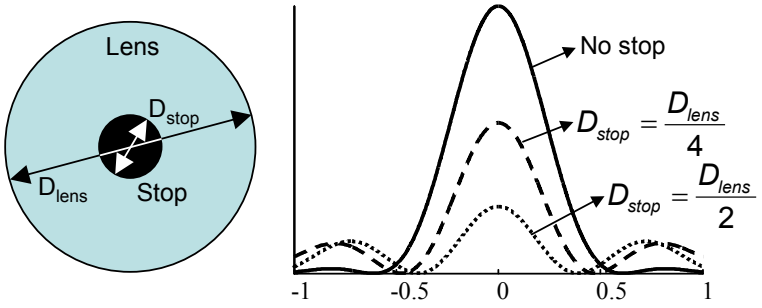


Figure 2.5: By blocking the central circle of a lens the point spread function can be altered. The solid line gives the point spread function for the lens without stop. The dashed and the dotted line give the apodized point spread function.

different point spread functions the effect of the side lobes can be reduced. The resolution increase is however still accompanied by a high overall loss in power throughput [32].

2.4.2 Model-based resolution

Till now we have looked to what extent the features in the object plane are distorted by imaging them into the image plane. If one could however know the exact shape of the point spread function, the original image could be reconstructed [33]. One method is based on the assumption that the object is finite in size, which assures that its spectrum is analytic. The frequencies within the otf-band can be reconstructed by dividing through the optical transfer function otf. The frequencies beyond the cut-off frequency can then be found by analytic continuation in the frequency domain. If one knows the object spectrum, one can find the object itself, which means in principle infinite resolution can be achieved. Unfortunately these methods are very sensitive to noise. Moreover most detection methods do not observe the phase of the optical field in the image plane. This means that in practice the resolution enhancements are not as big as one could expect.

In practice detected images are always disturbed by noise. Moreover the point spread function cannot be known with infinite accuracy. This means optical resolution is ultimately not limited by the Raleigh distance but by systematic and random errors. The effective resolution also depends on how much a priori knowledge one has on the object. A very recent statistical analysis of resolution and noise by Sharam [34] gives as a rule of thumb that for two equally bright point sources the

minimal resolvable distance scales with $\frac{1}{\sqrt[4]{\mathrm{SNR}}}$, with SNR the signal to noise ratio. This looks very similar to the information theory from Shannon [35], which says that there is an upper limit to the information that can be retrieved for a given bandwidth and for a given signal to noise ratio. This theory can also be used for the otf-band of a lens, which is $2NA/\lambda$ wide. For white noise added at the detector side, the possible resolvable information content can be written as [36, 30]:

$$I = \left(1 + 2L_x \frac{2NA}{\lambda}\right)\left(1 + 2L_y \frac{2NA}{\lambda}\right) \log_2(1 + \mathrm{SNR})\,, \tag{2.37}$$

with L_x and L_y the dimension of the area on the disc.

This equation shows that the information content that can be transmitted by an optical system depends both on the bandwidth and on the signal to noise ratio. The shape of the optical transfer function is also included by the fact that for optical frequencies that are heavily attenuated the noise will have a relatively higher impact, as white noise is proportional to the square root of the intensity.

2.4.3 Coding techniques

This brings us to the key question in optical data storage:

> *How much information content can be retrieved from a unit surface on the disc by means of an optical detector, for a given numerical aperture, wavelength and noise?*

Apart from using a smaller wavelength and a higher numerical aperture, one of the techniques for increasing the data density on a disc is improving the coding techniques. Till recently signal processing was hardly used in optical recording. This is in contrast to magnetic recording, where it is already extensively used [37]. As explained in section 1.2.2 the CD and the DVD use a run length limited coding of the bits coupled with a cross-interleaved Reed-Solomon (CIRC) detection. There is however much research on using more efficient signal processing. In Blu-ray and the competing HD-DVD format, an extra step is added: The partial response maximum likelihood (PRML) and Viterbi coding [38, 39, 40]. These methods try to guess the current bit, based on the knowledge of the previously detected bits. In chapter 4, we describe a similar but simpler technique to extract the bit patterns with the waveguide detection method.

Recent developments showed also a new type of coding mechanisms, Turbo codes [41, 42], which allow efficient coding very near to the Shannon limit. The disadvantage of these codes is that they work on long blocks, increasing the latency. Secondly they need more calculation power than the traditional coding techniques. They are however already in use for satellite communications and might also be implemented in future optical data storage [43].

2.5 Conclusions

Resolution can be quantified in different ways. The traditional Rayleigh limit regards the distance between two just resolvable point sources. The optical transfer function gives the frequency response of an optical system, and normally has a cut-off frequency. Spatial frequencies higher than this cut-off frequencies cannot be detected. These figures of merit are not always related, as by changing the transmission of the lens pupil function, the resolution in the Rayleigh sense can be decreased to zero without increasing the cut-off frequency. Eventually the resolution is limited only by random and systematic errors. The aim is hence not only to improve the optical bandwidth, but also an optimization of the information transmission within this band. In chapter 3 we will continue with a thorough analysis of resolution in an optical scanning microscope, including the new waveguide detection method. The optimization of the information content with this method will be analysed in chapter 4.

Chapter 3

Analysis of the optical scanning microscope

In this chapter we will give an analysis of the scanning microscope. In the first section the principles of a general scanning microscope will be outlined. In the second section the point spread function and the optical transfer function for the proposed waveguide detection method, the conventional DVD read-out system and of the confocal microscope will be derived. After some analytic approximations, the optical properties of these three systems will analysed in more detail in the next sections. To conclude this chapter a few systems will be compared by their point and step response.

3.1 A general scanning microscope

3.1.1 What is a scanning microscope

In contrast to a conventional optical microscope the scanning microscope does not deliver an image of a fixed sample, but provides an electrical signal by probing a small region on a scanning sample. In general a scanning microscope consists of a light source that illuminates a part of the sample, and a detector that detects only a part of the light reflected from (or transmitted trough) the sample (Fig. 3.1). The illumination and detection can be done by free space propagation, by a lens system or by both. By scanning the sample along one or more axes, a different part of the sample will be illuminated, and likewise at the detection side, light coming from different parts of the sample will be captured by the detector. The signal at the detector will hence change

as a function of the scanning position of the sample. The analysis of this signal can provide us the optical properties of the sample. Existing examples of a scanning microscopes include the DVD detection system, sometimes called extended detector system or type I scanning microscope, and the confocal microscope, sometimes called type II scanning microscope. We will discuss in detail a new type of a scanning microscope, the waveguide detection system.

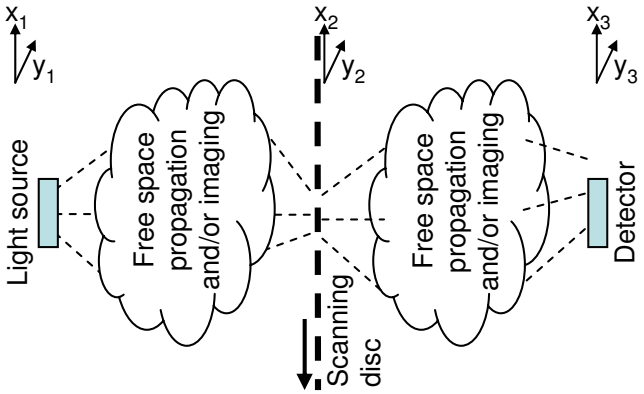


Figure 3.1: The light path in an optical scanning microscope. A light source is guided through free space of an imaging system towards the disc, where it forms small spot. The reflected light is captured back onto a detector (For reasons of clarity the above figure has been drawn as a transmission setup). By scanning the disc, different bit patterns pass into the central light spot, and the signal at the detector changes accordingly.

The resolution in a scanning microscope is defined differently than in a conventional microscope. In a conventional microscope the resolution is defined by the relationship between the optical field in the object plane and the optical field in the image plane. In a scanning microscope, the resolution is the relationship between the features on a sample and the electrical signal at the detector as a function of time. In this section a brief analysis of a general scanning microscope will be given. Later in this chapter, the DVD detection system, the confocal system and the waveguide system will be analysed more thoroughly.

3.1.2 Rigourous calculations

For rigourous simulations taking into account the vectorial character of the electromagnetic field and multiple reflections in between the waveguide and the disc, one can use the eigenmode expansion method. For

a detail description of this method we refer to appendix C, but essentially the method can be summarized as follows: it describes the electromagnetic field in a closed metal box as a linear combinations of the eigenmode of a number of z-invariant sections as shown on Fig. 3.2. If one excites the downwards propagating zeroth order mode, one can calculate the excitation towards the upwards propagating the zeroth and first order mode as a function of different positions of the bit patterns. There are however some problems with these calculation. Firstly the metal box introduced to discretize the eigenmodes influence the electromagnetic field by reflecting the light at the borders. This effect can however be reduced by adding two anti-reflection layers, based on the 'perfectly matched layers'-technique[44]. Secondly the section with the bits has multiple cores with a higher refractive index, this leads often to degeneracy within the set of eigenmodes. One should therefore always carefully check the numerical stability of the solution.

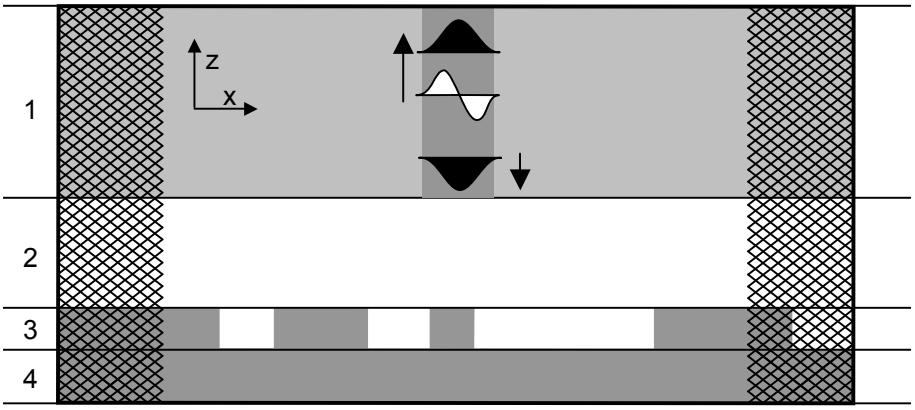


Figure 3.2: Rigourous calculations for the waveguide detection method in the near-field. The waveguide is scans at at small distance over the disc. The waveguide and a part of the disc are placed in a closed metal box, which is subdivided into a number of z-invariant sections. The signal picked up with the waveguide modes reflected from the disc can now rigourously be calculated with the eigenmode expansion method.

Fig. 3.3 shows the results of a numerical calculation of the zeroth and first order mode excitation, when the disc is illuminated with a zeroth order mode. If the minimal bit size is not too small, the zeroth order mode excitation 'follows' the bit pattern itself and the first order mode excitation marks the edges of the bit pattern. For smaller features the link between the bit pattern and the resulting mode excitations is

less obvious. Similarly to the results obtained with a scalar approximation later on in this thesis, discriminating bit patterns with the mode excitations becomes increasingly more difficult. For an air gap larger than one wavelength, spatial periods smaller than $2/\lambda$ are not detected any more.

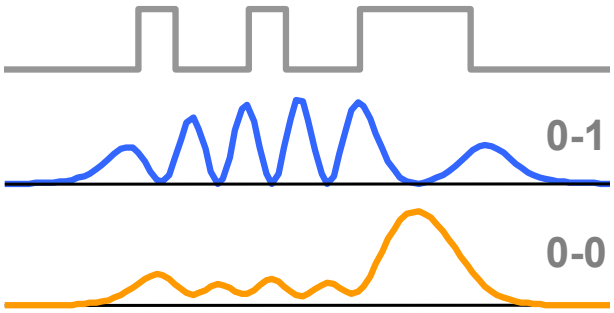


Figure 3.3: Excitation of the zeroth and first order mode by reflection from the disc, which is illuminated with the zeroth order mode of the same waveguide.

3.1.3 Approximations

To simplify the calculations and to enhance the physical insight we made some approximations. In a real system these approximations can not always be taken for granted, and this results in some loss in generality. However the qualitative results remain the same.

No multiple reflections A first approximation is that we assume that there are no multiple reflections in between the illuminating source, the disc, and the detector. With a correct design this should not be the case but in practice parasitical reflections may deteriorate the detector signal of the scanning microscope.

Scalar diffraction approximation The second approximation is more important and says that the optical field can be described by a scalar function. This is correct in free space, but does not describe very accurately the light transmission at media interfaces. In the scanning microscope these are the waveguides, the disc and the lenses. A full vectorial theory can be found in literature [45, 46] but results in more complicated equations, which reduce the physical insight. In the design of a real system these should however be included.

Polarization Very closely linked with the scalar diffraction approximation is the polarization of the light. In free space, light exists in two fully decoupled orthogonal polarizations. Our assumption is that this is also the case in other media. This means among others that there is no coupling in between the two polarizations, nor along the optical light path, nor by reflecting at the disc surface. In reality this is not always true. Some types of magneto-optical recording make even use of a polarization angle shift at the disc reflection to read out the data. Including the polarization dependence of the disc reflection and/or other elements in the optical light path would make all equations more complex.

Disc reflection function In general an optical disc consists of small blocks with a different refractive index closely packed together, forming a 3D structure on a flat surface. More specific the optical disc used in conventional optical recording is a close packing of two types of layered structures forming a bit pattern along a spiral track. In principle one would need vector diffraction to model such a disc structure [47]. We choose however to use a scalar model by replacing the disc by a (complex) amplitude transmission function of the disc, $r(x, y)$, independent of the incident field. For partially coherent light this means a multiplication of the mutual intensity with the reflection function $R(x, y; x', y') = r(x, y)\, r^*(x', y')$.

Especially for features smaller than the wavelength, this scalar approximation deviates severely from the full vectorial description. This should not be a problem as long as one keeps in mind that the final description of the optical resolution of the detection system should be fine-tuned with the reflection properties of the disc.

3.1.4 Resolution of a scanning microscope

The resolution of a scanning microscope can be calculated by following the light path from illumination to detection. For a description of this light path the partially coherent equations from section 2.2.2 can be used. This derivation is a generalization of the derivation of the so called type I and type II scanning microscopes [48, 49]. In a scanning microscope there are elements which operate linearly with the amplitude and other that operate linearly with the intensity. For a full coverage we use therefore the mutual intensity as defined in section 2.2.2.

From illumination to detection we use the following definitions.

- The mutual intensity of the illumination is $J_{illum}(x_1, y_1; x_1', y_1')$. Most scanning microscopes use a coherent light source, but in some cases incoherent light is used.
- At the detector side a similar function, $J_{detect}(x_3, y_3; x_3', y_3')$, describes the detector sensitivity. An ordinary detector gives an electric signal proportional to the intensity:

$$J_{detect}\big(x_3, y_3; x_3', y_3'\big) = p_{detect}(x_3, y_3)\,\delta\big(x_3 - x_3'\big)\,\delta\big(y_3 - y_3'\big)\,,$$

  with $p_{detect}(x_3, y_3)$ equal to the detector aperture. The detector sensitivity is assumed to be 1. As will be described later the waveguide used as a detector, operates linearly with the amplitude of the light, and can hence be called coherent detection.
- As described in section 2.2.2, a general light path from illumination source to the disc can be described by the point spread function $PSF_{12}(x, y; x', y')$.
- The light path from the disc towards the detector can analogously be written as $PSF_{23}(x, y; x', y')$.
- Finally the disc reflection (or transmission) is $R(x_2, y_2; x_2', y_2')$.

Following the light path from illumination source to the disc we find the mutual intensity in the plane just before the disc:

$$J_{2^-}\big(x_2, y_2; x_2', y_2'\big) = \iiiint J_{illum}\big(x_1, y_1; x_1', y_1'\big) \times PSF_{12}\big(x_2 - x_1, y_2 - y_1; x_2' - x_1', y_2' - y_1'\big)\,\mathrm{d}x_1\,\mathrm{d}x_1'\,\mathrm{d}y_1\,\mathrm{d}y_1'. \quad (3.1)$$

The disc is modelled by a linear reflection (or transmission) function. The scanning position of the disc is described by the parameters x_s and y_s. The field reflecting from the disc can then be written as:

$$J_{2^+}\big(x_2, y_2; x_2', y_2'; x_s, y_s\big) = J_{2^-}\big(x_2, y_2; x_2', y_2'\big)\,R\big(x_2 - x_s, y_2 - y_s; x_2' - x_s, y_2' - y_s\big)\,. \quad (3.2)$$

The light path to the detector implies again a convolution:

$$J_{3^-}\big(x_2, y_2; x_2', y_2'; x_s, y_s\big) = \iiiint J_{2^+}\big(x_2, y_2; x_2', y_2'; x_s, y_s\big) \times PSF_{23}\big(x_3 - x_2, y_3 - y_2; x_3' - x_2', y_3' - y_2'\big)\,\mathrm{d}x_2\,\mathrm{d}x_2'\,\mathrm{d}y_2\,\mathrm{d}y_2'. \quad (3.3)$$

Finally the electric signal at the detector is found by integration of the mutual intensity with the detector sensitivity:

$$I(x_s, y_s) = \iiiint J_{3^-}(x_3, y_3; x_3', y_3'; x_s, y_s) \times J_{detect}(x_3, y_3; x_3', y_3')\,\mathrm{d}x_3\,\mathrm{d}x_3'\,\mathrm{d}y_3\,\mathrm{d}y_3' \quad (3.4)$$

Combining Eqs. (3.1–3.4), and after changing the integration order one finds:

$$\begin{aligned} I(x_s) = &\iiiint R(x - x_s, y - y_s; x' - x_s, y' - y_s) \\ &\times PSF(x, y; x', y')\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y', \\ \text{with } &PSF(x_2, y_2; x_2', y_2') = \\ &\Big[\iiiint J_{illum}(x_1, y_1; x_1', y_1') \\ &\quad \times PSF_{12}(x_2 - x_1, y_2 - y_1; x_2' - x_1', y_2' - y_1')\,\mathrm{d}x_1\,\mathrm{d}x_1'\,\mathrm{d}y_1\,\mathrm{d}y_1'\Big] \\ &\times \Big[\iiiint J_{detect}(x_3, y_3; x_3', y_3') \\ &\quad \times PSF_{23}(x_3 - x_2, y_3 - y_2; x_3' - x_2', y_3' - y_2')\,\mathrm{d}x_3\,\mathrm{d}x_3'\,\mathrm{d}y_3\,\mathrm{d}y_3'\Big]. \end{aligned} \quad (3.5)$$

In this equation, the generalized partial coherent point spread function $PSF(x, y; x', y')$ fully describes the properties of the optical system. The disc reflection function $R(x, y; x', y')$ is only dependent on the properties of the disc. In general the intensity I is always real and positive. In section 3.2.3, we will however extend Eq. (3.5) to negative intensity. In any case from $I \in \mathbb{R}$ follows that $PSF(x, y; x', y') = PSF^*(x', y'; x, y)$. Using the Tilde-operator defined by Eq. (A.23) in appendix A, we define the optical transfer function as $OTF(f, g; f', g') = \widetilde{PSF}(f, g; f', g')$. Eq. (3.5) can then be rewritten in the frequency domain:

$$I(x_s) = \iiiint \widetilde{R}(f, g; f', g')\, OTF(-f, -g; -f', -g') \times \exp\{-2\pi j\,[(f - f')\,x_s + (g - g')\,y_s]\}\,\mathrm{d}f\,\mathrm{d}g\,\mathrm{d}f'\,\mathrm{d}g' \quad (3.6)$$

Again one can prove that $OTF(f, g; f', g') = OTF^*(f', g'; f, g)$. Without loss of generality, it is assumed that the disc is only scanned in the x-direction, which means $y_s = 0$. In section 3.2 the PSF and OTF will

be calculated for different optical systems: the waveguide system, the DVD system, and the confocal system.

Examples In section 3.2, we will give a full analysis of the waveguide detection method, extended detector method and the confocal microscope. From Eq. (3.5) one can however already directly derive some results.

Assume at the detection and the illumination side, a circular aberration-free lens. From Eq. (2.27), one finds disregarding a proportionality factor that:

$$PSF_{12}(x,y;x',y') = PSF_{23}(x,y;x',y') = \frac{J_1\left(\frac{2\pi NA}{\lambda}r\right)}{\frac{2\pi NA}{\lambda}r}\frac{J_1\left(\frac{2\pi NA}{\lambda}r'\right)}{\frac{2\pi NA}{\lambda}r'} \quad (3.7)$$
$$\text{with } r = \sqrt{x^2+y^2} \text{ and } r' = \sqrt{x'^2+y'^2}$$

For a very narrow illumination source and a very narrow detection aperture, one has:

$$\begin{aligned} J_{illum}(x_1,y_1,x_1',y_1') &= \delta(x_1)\,\delta(y_1)\,\delta(x_1')\,\delta(y_1') \\ J_{detect}(x_3,y_3,x_3',y_3') &= \delta(x_3)\,\delta(y_3)\,\delta(x_3')\,\delta(y_3')\,. \end{aligned} \quad (3.8)$$

Combining these in Eq. (3.5), one finds:

$$PSF(x,y;x',y') = \left[\frac{J_1\left(\frac{2\pi NA}{\lambda}r\right)}{\frac{2\pi NA}{\lambda}r}\frac{J_1\left(\frac{2\pi NA}{\lambda}r'\right)}{\frac{2\pi NA}{\lambda}r'}\right]^2. \quad (3.9)$$

If all other parameters are the same, but one uses a very large detector this becomes:

$$J_{detect}(x_3,y_3,x_3',y_3') = \delta(x_3-x_3')\,\delta(y_3-y_3')\,, \quad (3.10)$$

and after some calculations:

$$PSF(x,y;x',y') = \frac{J_1\left(\frac{2\pi NA}{\lambda}r\right)}{\frac{2\pi NA}{\lambda}r}\frac{J_1\left(\frac{2\pi NA}{\lambda}r'\right)}{\frac{2\pi NA}{\lambda}r'}\frac{J_1\left(\frac{2\pi NA}{\lambda}(r'-r)\right)}{\frac{2\pi NA}{\lambda}(r'-r)}. \quad (3.11)$$

This shows already the differences between a confocal microscope and an extended detector system. In section 3.2 will we will return on these issues.

3.1.5 Interpretation of the partially coherent OTF

The partially coherent optical transfer function is a general measure of the resolution, it shows to what extent spatial frequencies from the patterns on the disc are retrieved at the detector side. A concrete interpretation of the partially coherent optical transfer function is less trivial. Some insight can be found by investigating the influence of the OTF on a one-dimensional grating. Assume the disc has a constant reflection and a one-dimensionally modulated additional reflection coefficient:

$$r(x,y) = [1 + 2\varepsilon\cos(2\pi f_d x)]\,. \tag{3.12}$$

One finds:

$$\begin{aligned}\widetilde{R}(f,g;f',g') = \delta(g)\delta(g')\,[\delta(f) + \varepsilon\delta\,(f-f_d) + \varepsilon\delta\,(f+f_d)] \\ \times\left[\delta(f') + \varepsilon^*\delta\left(f'-f_d\right) + \varepsilon^*\delta\left(f'+f_d\right)\right]\end{aligned} \tag{3.13}$$

and combining Eq. (3.6) and Eq. (3.13) the intensity can be written as:

$$\begin{aligned}I(x_s) &= |\varepsilon|^2\left[OTF(f_d,0;-f_d,0)\,e^{4\pi j f_d x_s} + OTF(-f_d,0;f_d,0)\,e^{-4\pi j f_d x_s}\right] \\ &+ \varepsilon\left[OTF(f_d,0;0,0)\,e^{2\pi j f_d x_s} + OTF(-f_d,0;0,0)\,e^{-2\pi j f_d x_s}\right] \\ &+ \varepsilon^*\left[OTF(0,0;-f_d,0)\,e^{2\pi j f_d x_s} + OTF(0,0;f_d,0)\,e^{-2\pi j f_d x_s}\right] \\ &+ \left[OTF(0,0;0,0) + |\varepsilon|^2 OTF(f_d,0;f_d,0) + |\varepsilon|^2 OTF(-f_d,0;-f_d,0)\right].\end{aligned} \tag{3.14}$$

By defining some new parameters

$$\begin{aligned}\mathcal{C}_0(f_d) &= OTF(0,0;0,0) \\ \mathcal{C}_{1r}(f_d) &= OTF(f_d,0;0,0) + OTF(0,0;-f_d,0) \\ \mathcal{C}_{1i}(f_d) &= OTF(f_d,0;0,0) - OTF(0,0;-f_d,0) \\ \mathcal{C}_{2a}(f_d) &= OTF(f_d,0;f_d,0) + OTF(-f_d,0;-f_d,0) \\ \mathcal{C}_{2b}(f_d) &= OTF(f_d,0;-f_d,0)\,,\end{aligned} \tag{3.15}$$

Eq. (3.14) can be rewritten in a simpler form. For a given f_d one finds:

$$I(x_s) = \mathcal{C}_0 \tag{3.16a}$$

$$+\,2\,|\mathcal{C}_{1r}\Re(\varepsilon) + j\mathcal{C}_{1i}\Im(\varepsilon)|\cos\{2\pi f_d x_s + \angle[\mathcal{C}_{1r}\Re(\varepsilon) + j\mathcal{C}_{1i}\Im(\varepsilon)]\}\,. \tag{3.16b}$$

$$+\,|\varepsilon|^2\,(\mathcal{C}_{2a} + 2\,|\mathcal{C}_{2b}|\cos[4\pi f_d x_s + \angle(\mathcal{C}_{2b})])\,. \tag{3.16c}$$

Inspecting this equation one finds that there are three main components in the detector signal:

- Component (3.16a) is independent of x_s. This component is a zeroth order signal, and gives a measure of the overall reflectivity but does not help to discriminate the displacement of the grating.
- Component (3.16b) is roughly proportional to $|\varepsilon|$. It is a first order signal and varies along x_s with the same frequency as the grating. The amplitude of the modulation is however decreased and a phase lag is introduced. If ε is real the amplitude decrease and phase lag are defined by C_{1r}. For imaginary ε, C_{1i} has this role. For complex ε both parameters influence the detector signal.
- Component (3.16c), the second order signal, is proportional to $|\varepsilon|^2$. It has two components. A constant component defined by $\mathcal{C}_{2a}$ and a component with the double frequency as the grating. Again the amplitude and phase lag are a function of f_d.

In general the modulation depth of the grating, $|\varepsilon|$, is smaller than one. This means that the third component is relatively weak in comparison to the first and second component. For stronger gratings this component is however not negligible.

The above equations can be generalized for a complex signal modulation $r(x,u) = 1 + \varepsilon\rho(x)$:

$$I(x_s) = OTF(0,0;0,0) \tag{3.17a}$$

$$+ 2\,|\varepsilon|\,\Re\left[\int \widehat{\rho}(f)\, e^{j\angle\varepsilon}\, OTF(f,0;0,0)\, e^{j2\pi f x_s}\,\mathrm{d}f\right] \tag{3.17b}$$

$$+ |\varepsilon|^2 \iint |\widehat{\rho}(f)|^2\, OTF\big(f,0;f',0\big)\, e^{j2\pi(f-f')x_s}\,\mathrm{d}f\,\mathrm{d}f' \tag{3.17c}$$

If the imaging would have no distortion nor a resolution limit, then $OTF(f,g;f',g')$ would be identical to 1 and one would have

$$I(x_s) = 1 \qquad \text{(constant intensity)} \tag{3.18a}$$

$$+ \Re[\varepsilon\rho(x)] \qquad \text{(signal)} \tag{3.18b}$$

$$+ |\rho(x)|^2 \qquad \text{(cross terms)}. \tag{3.18c}$$

Similarly as with the single grating, there are three components: the zeroth order component gives a measure of the overall reflectivity. The first order component shows the original signal on the disc, distorted

by the optical system. The amplitude decrease and phase lag of the frequency components of the signal by the optical system are defined by the cross section $OTF(f, 0; 0, 0)$ of the OTF and the phase of ε. For a shallow modulation, with $|\varepsilon| \ll 1$, the second order signal, which includes the cross terms in between the different spatial frequencies, is negligible. This is however not the case for strong modulation. In this case the entire OTF is important. As a figure of merit for the second order signal, we will use the $\mathcal{C}_{2b}$ cross section without losing sight of the whole OTF.

3.2 Derivation of the PSF and OTF

In this section a general equation for the point spread function, PSF and the optical transfer function, OTF will be derived for the proposed waveguide detection method, the extended detector method and the confocal microscope.

3.2.1 Illumination, detection and optical light path

To derive the properties of the optical scanning microscope we first run through the illumination, the detection and the optical light path. The illumination is done by a coherent laser source, be it in a conventional DVD setup, a confocal microscope or in the new waveguide detection method described in section 1.3. This means that the optical fields can be added up.

The waveguide The field transmitted through the waveguide-air interface by a waveguide mode $m_i^{wg}(x, y)$ will be written as $m_i^{air}(x, y)$. According to the mode expansion method (see appendix C), this field can be expanded as a sum of the eigenmodes of free space, the plane waves $\phi_i^{air}(x, y)$. In contrast to the real eigenmodes of the waveguide, $m_i^{wg}(x, y)$, the hypothetical modes $m_i^{air}(x, y)$ are in general not orthogonal and not normalized.

$$m_i^{wg}(x, y) \rightarrow m_i^{air}(x, y) = \sum_{j=1}^{N} r_{i,j}^{wg,air} \phi_i^{air}(x, y) \tag{3.19}$$

Once the $m_i^{air}(x, y)$ are calculated, it is possible to derive the excitation of the real waveguide modes by an arbitrary field $\psi^{air}(x, y)$ incident on

the waveguide, by calculating the overlap integral of the incident field with the hypothetical waveguide modes $m_i^{air}(x,y)$:

$$\begin{aligned} \psi^{air}(x,y) \rightarrow \psi^{wg}(x,y) &= \sum_{j=1}^{N} \mu_i m_i^{wg}(x,y) \\ \text{with } \mu_i &= \iint \psi^{air}(x,y)\, m_i^{air,*}(x,y)\, \mathrm{d}x\, \mathrm{d}y. \end{aligned} \tag{3.20}$$

This equation can be derived as follows: First we expand $\psi^{air}(x,y)$ into plane waves, the eigenmodes of free space.

$$\begin{aligned} \psi^{air}(x,y) &= \sum_{j=1}^{N} \kappa_j \phi_j^{air}(x,y) \\ \text{with } \kappa_j &= \int \psi(x,y)\, \phi_j^{air}(x,y)\, \mathrm{d}x\, \mathrm{d}y \end{aligned} \tag{3.21}$$

Each of these plane wave ϕ_j^{air} will excite some of the eigenmodes of the waveguide :

$$\phi_j^{air}(x,y) \rightarrow \sum_{i=1}^{N} t_{j,i}^{air,wg} \phi_i^{wg}(x,y). \tag{3.22}$$

Combining Eq. (3.21) and Eq. (3.22) one finds:

$$\begin{aligned} \mu_i &= \sum_{j=1}^{N} \kappa_j\, r_{j,i}^{air,wg} \\ &= \sum_{j=1}^{N} \left(\iint \psi(x,y)\, \phi_j^{air}(x,y)\, \mathrm{d}x\, \mathrm{d}y \right) r_{j,i}^{air,wg} \\ &= \iint \psi(x,y) \left(\sum_{j=1}^{N} \phi_j^{air}(x,y)\, r_{j,i}^{air,wg} \right) \mathrm{d}x\, \mathrm{d}y \\ &= \iint \psi(x,y) \left(\sum_{j=1}^{N} \phi_j^{air}(x,y)\, r_{i,j}^{wg,air} \right) \mathrm{d}x\, \mathrm{d}y \\ &= \iint \psi(x,y)\, m_i^{air}(x,y)\, \mathrm{d}x\, \mathrm{d}y \end{aligned} \tag{3.23}$$

which proves Eq. (3.20). Remind that this equation only tells to what extent a specific waveguide mode is excited, it does not provide information on where the rest of the light is scattered to. In the rest of the text we assume that these parasitical reflections are radiated away.

Detectors A conventional detector results in an electrical signal, proportional to the squared absolute value of the incident optical field, integrated over the detector area. The proportionality factor depends on the detector sensitivity and will be set to 1. This means:

$$I = \iint |\psi(x,y)|^2 \, p_{detect}(x,y) \, \mathrm{d}x \, \mathrm{d}y, \tag{3.24}$$

with $p_{detect}(x,y)$, the detector aperture, and $\psi(x,y)$ the incident field.

Imaging by lens or near-field propagation As described in section 2.3 imaging by a lens or propagation in free space can be written by the same equation:

$$\psi_2(x,y) = \gamma \iint \psi_0(\gamma x_1, \gamma y_1) \, psf_{lens}(x - x_1, y - y_1) \, \mathrm{d}x_1 \, \mathrm{d}y_1. \tag{3.25}$$

With γ, the linear demagnification factor from the waveguide (or detector) to the disc. psf_{lens} is the point spread function of the imaging system psf_{lens}, but can also represent the point spread function related to free space propagation over a small air gap. The point spread function from the waveguide to the disc will be written as $psf_{lens}(x,y)$ and the point spread function of the imaging from the disc to the waveguide will be written as $psf'_{lens}(x,y)$. From Eq. (2.21) one finds that

$$psf'_{lens}(x,y) = \frac{1}{\gamma^2} psf_{lens}\left(\frac{x}{\gamma}, \frac{y}{\gamma}\right). \tag{3.26}$$

To simplify the equations further the following definitions are used:

$$\begin{aligned}
m^{\gamma}_{illum}(x,y) &= \gamma m^{air}_{illum}(\gamma x, \gamma y) & \widehat{m^{\gamma}_{illum}}(f,g) &= \frac{1}{\gamma} \widehat{m^{air}_{illum}}\left(\tfrac{f}{\gamma}, \tfrac{g}{\gamma}\right) \\
m^{\gamma}_{detect}(x,y) &= \gamma m^{air}_{detect}(\gamma x, \gamma y) & \widehat{m^{\gamma}_{detect}}(f,g) &= \frac{1}{\gamma} \widehat{m^{air}_{detect}}\left(\tfrac{f}{\gamma}, \tfrac{g}{\gamma}\right) \\
p^{\gamma}_{conf}(x,y) &= \gamma^2 p_{conf}(\gamma x, \gamma y) & \widehat{p^{\gamma}_{conf}}(f,g) &= \widehat{p_{conf}}\left(\tfrac{f}{\gamma}, \tfrac{g}{\gamma}\right)
\end{aligned} \tag{3.27}$$

3.2.2 Waveguide detection method

The waveguide detection method, as described in section 1.3, uses a multimodal waveguide to detect the patterns on the disc. Fig. 3.4 gives a schematic view: A laser excites one or more waveguide modes. These

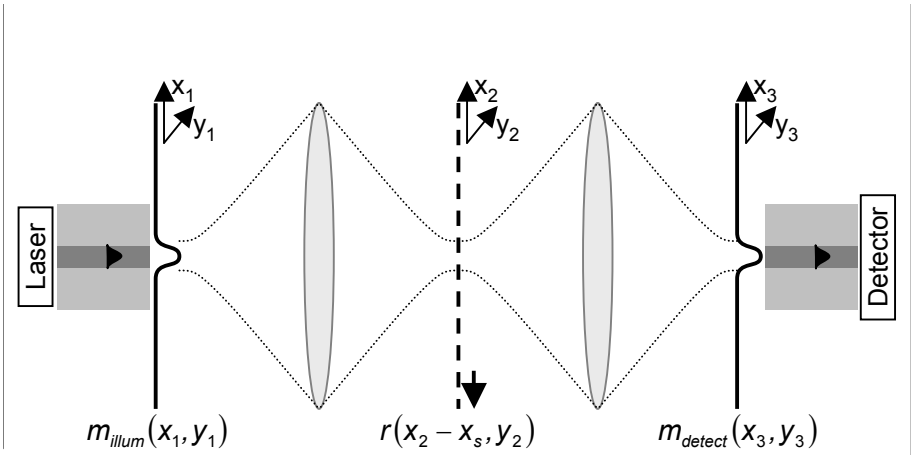


Figure 3.4: A schematic and simplified graph showing the optical light path in the waveguide detection system. For reasons of clarity the graph has been unfolded to a transmission system.

propagate through free space or through a lens system and illuminate a small spot on the disc. The reflection is imaged back onto the waveguide, where it is picked up by the modes of the same waveguide.

Using Eq. (3.19), the field just outside the illumination waveguide can be written as a sum of multiple modes:

$$\psi_{1^+}\left(x_1, y_1, z = 0^+\right) = \sum_{i=1}^{N} \alpha^i m_i^{air}(x_1, y_1) = m_{illum}^{air}(x_1, y_1)\,. \qquad (3.28)$$

This field is imaged onto the disc by a lens or propagates towards the disc in free space. Using the definitions as described in section 2.3.2 the optical field on the disc can be written as a convolution of the optical field at the waveguide with the point spread function psf_{lens}:

$$\psi_{2^-}\left(x_2, y_2, z = z_d^-\right) = \iint \gamma m_{illum}^{air}(\gamma x_1, \gamma y_1)\, psf_{lens}(x_2 - x_1, y_2 - y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1. \qquad (3.29)$$

The disc is considered as an amplitude transmission function $r(x, y)$. As the disc is scanned over a scanning position x_s therefore this reflection function becomes $r(x - x_s, y)$. The field reflected from the disc can hence be written:

$$\psi_{2^+}\left(x_2, x_s, y_2, z = z_d^+\right) = \psi_{2^-}\left(x_2, y_2, z = z_d^-\right) r(x_2 - x_s, y_2)\,. \qquad (3.30)$$

By propagation through the air gap or through a lens system the reflected field is imaged onto the detection waveguide and forms:

$$\begin{aligned}
&\psi_3(x_3, x_s, y_3) \\
&\quad = \iint \frac{1}{\gamma}\psi_{2^+}\left(\frac{x_2}{\gamma}, x_s, \frac{y_2}{\gamma}\right) psf'_{lens}(x_3 - x_2, y_3 - y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2 \qquad (3.31)\\
&\quad = \iint \psi_{2^+}(x_2, x_s, y_2)\,\gamma psf'_{lens}(x_3 - \gamma x_2, y_3 - \gamma y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2 \qquad (3.32)
\end{aligned}$$

At the air-waveguide interface the field $\psi_3(x, y)$ excites the different modes in the multimodal waveguide. The excitation coefficient of a waveguide mode can be found by calculating the overlap with the $m^{air}_{detect}(x, y)$ using Eq. (3.20):

$$\chi(x_s) = \iint \psi_3(x_3, x_s, y_3)\, m^{air}_{detect}(x_3, y_3)\,\mathrm{d}x_3\,\mathrm{d}y_3. \qquad (3.33)$$

Combining Eqs. (3.28–3.33) this results in

$$\begin{aligned}
\chi(x_s) = \iint \Bigg\{ \iint \Bigg[& r(x_2 - x_s, y_2) \\
& \times \iint \gamma m^{air}_{illum}(\gamma x_1, \gamma y_1)\, psf_{lens}(x_2 - x_1, y_2 - y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \Bigg] \\
\times \gamma psf'_{lens}(x_3 - \gamma x_2, y_3 - \gamma y_2)\ \mathrm{d}x_2\,\mathrm{d}y_2 & \Bigg\}\, m^{air}_{detect}(x_3, y_3)\,\mathrm{d}x_3\,\mathrm{d}y_3 \qquad (3.34)
\end{aligned}$$

By changing the integration order, replacing psf'_{lens} using Eq. (3.26), and using the definitions in Eq. (3.27) one finds:

$$\begin{aligned}
\chi(x_s) = \iint \Bigg[& r(x_2 - x_s, y_2) \\
& \times \iint m^{\gamma}_{illum}(x_1, y_1)\, psf_{lens}(x_2 - x_1, y_2 - y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \\
& \times \iint m^{\gamma}_{detect}(x_3, y_3)\, psf_{lens}(x_3 - x_2, y_3 - y_2)\,\mathrm{d}x_3\,\mathrm{d}y_3 \Bigg]\,\mathrm{d}x_2\,\mathrm{d}y_2.
\end{aligned} \qquad (3.35)$$

Using the two-dimensional convolution operator $\otimes$ and assuming $psf_{lens}(x, y) = psf_{lens}(-x, -y)$ this can further be simplified to:

$$\begin{aligned}
psf_{detect}(x, y) &= psf_{lens}(x, y) \otimes m^{\gamma}_{detect}(x, y) \\
psf_{illum}(x, y) &= psf_{lens}(x, y) \otimes m^{\gamma}_{illum}(x, y) \\
psf_{tot}(x, y) &= psf_{illum}(x, y)\, psf_{detect}(x, y)\,.
\end{aligned} \qquad (3.36)$$

Eq. (3.35) can be rewritten as

$$\chi(x_s) = \iint r(x - x_s, y) psf_{tot}(x, y)\ \mathrm{d}x\,\mathrm{d}y \tag{3.37}$$

$$= r(x, y) \otimes_x psf_{tot}(x, y) \tag{3.38}$$

The response measured at the detector is proportional to the power in the excited modes and can be written as in Eq. (3.5)

$$\begin{aligned} I(x_s) &= \left| \int r(x - x_s)\, psf_{tot}(x, y)\, \mathrm{d}x \right|^2 \\ &= \iiiint R\big(x - x_s, y; x' - x_s, y'\big)\, PSF\big(x, y; x', y'\big)\, \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y' \end{aligned} \tag{3.39}$$

In this last equation the partially coherent point spread function is

$$PSF\big(x, y; x', y'\big) = psf_{tot}(x, y)\, psf^*_{tot}\big(x', y'\big) \tag{3.40}$$

$$\begin{aligned} \text{with } psf_{tot}(x, y) = &\left[\iint m^{\gamma}_{illum}(x_1, y_1)\, psf_{lens}(x - x_1, y - y_1)\, \mathrm{d}x_1\,\mathrm{d}y_1 \right] \\ &\times \left[\iint m^{\gamma}_{detect}(x_1, y_1)\, psf_{lens}(x - x_1, y - y_1)\, \mathrm{d}x_1\,\mathrm{d}y_1 \right]. \end{aligned}$$

Similarly one can calculate the partially coherent optical transfer function

$$\begin{aligned} OTF\big(f, g; f', g'\big) &= otf_{tot}(f, g)\, otf^*_{tot}\big(f', g'\big) \\ &= [otf_{illum}(f, g) \otimes otf_{detect}(f, g)] \\ &\quad \times \left[otf_{illum}\big(f', g'\big) \otimes otf_{detect}\big(f', g'\big) \right] \end{aligned} \tag{3.41}$$

$$\begin{aligned} \text{with } otf_{tot}(f, g) = &\iint p_{lens}(-\lambda z_{disc} u, -\lambda z_{disc} v)\, \widehat{m^{\gamma}_{detect}}(f - u, g - v) \\ &\times p_{lens}(\lambda z_{disc}\,(u - f)\,, \lambda z_{disc}\,(v - g))\, \widehat{m^{\gamma}_{illum}}(u, v)\, \mathrm{d}u\,\mathrm{d}v \\ otf_{illum}(f, g) = &\iint p_{lens}(-\lambda z_{disc} u, -\lambda z_{disc} v)\, \widehat{m^{\gamma}_{illum}}(u, v)\, \mathrm{d}u\,\mathrm{d}v \\ otf_{detect}(f, g) = &\iint p_{lens}(-\lambda z_{disc} u, -\lambda z_{disc} v)\, \widehat{m^{\gamma}_{detect}}(u, v)\, \mathrm{d}u\,\mathrm{d}v \end{aligned}$$

To understand how the different factors in this equation have an influence on the eventual optical transfer function, the calculation has been visualized in Fig. 3.5. For reasons of clarity, we assumed the lenses to be aberration-free and circular, and the waveguide modes to be Dirac

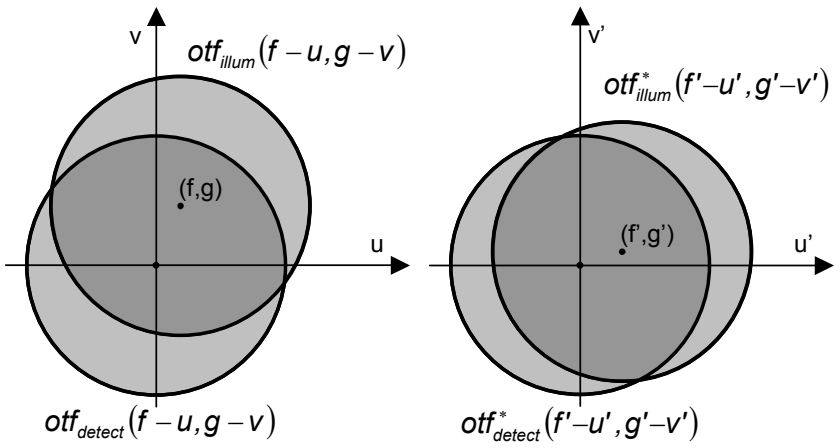


Figure 3.5: Visualization of the calculation of the OTF for the waveguide detection system. The gray circles represent the illumination and detection otf. As a function of f, f', g, g' the product of the area of the two intersection gives the total partially coherent OTF.

functions. Under these assumptions $\widehat{psf_{illum}}(u,v) = \widehat{psf_{detect}}(u,v) = \Pi\left(\frac{\sqrt{u^2+v^2}\lambda}{NA}\right)$, a circular step function with radius $\frac{NA}{\lambda}$. The OTF is proportional to the product of the surface of the areas of the two intersections. It is clear that the overlap area becomes zero as soon as one of the parameters f, g, f' or g' becomes larger than twice the circle radius.

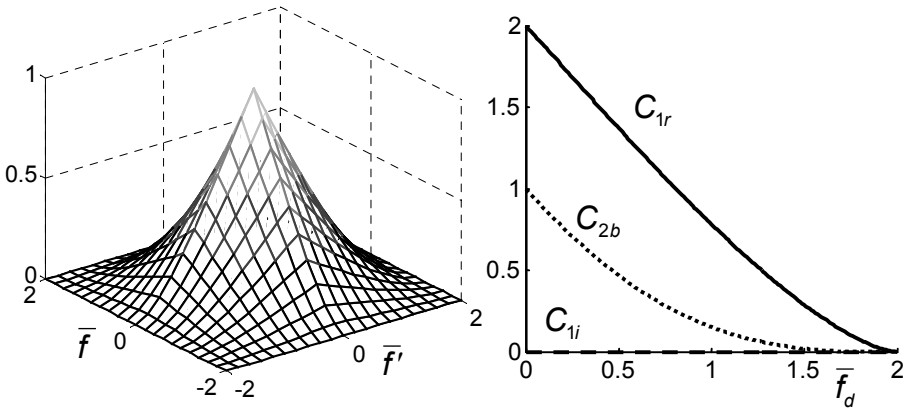


Figure 3.6: On the left the normalized partially coherent OTF along the x-axis of the waveguide detection system for circular aberration-free lenses and point-like waveguide modes. On the right the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15).

In this specific case the overlap areas can easily be calculated and lead to:

$$\begin{aligned} OTF(f,g;f',g') &= OTF(\overline{q};\overline{q'}) \\ &= \left[\frac{2}{\pi}\arccos\left(\frac{\overline{q}}{2}\right) - \frac{\overline{q}}{2}\sqrt{1-\left(\frac{\overline{q}}{2}\right)}\right] \\ &\quad \times \left[\frac{2}{\pi}\arccos\left(\frac{\overline{q'}}{2}\right) - \frac{\overline{q}}{2}\sqrt{1-\left(\frac{\overline{q'}}{2}\right)}\right] \end{aligned} \tag{3.42}$$

$$\begin{aligned} &\text{with } \overline{q} = \sqrt{\overline{f}^2 + \overline{g}^2}, f = \frac{fNA}{\lambda}, \text{ and } g = \frac{gNA}{\lambda} \\ &\text{and } \overline{q'} = \sqrt{\overline{f'}^2 + \overline{g'}^2}, f = \frac{f'NA}{\lambda}, \text{ and } g' = \frac{gNA}{\lambda} \end{aligned} \tag{3.43}$$

$OTF(f,g;f',g')$ has been plotted along the x-axis ($g = g' = 0$) in Fig. 3.6, together with the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15). These principal cross sections give the amplitude of the three components in the detector signal when the disc is a shallow grating. Two components have the same frequency as the grating: $\mathcal{C}_{1r}$ is proportional with the real part of the grating reflectivity and $\mathcal{C}_{1i}$ with the imaginary part. The third component, with amplitude $\mathcal{C}_{2b}$, has a doubled frequency.

3.2.3 Extended detector method

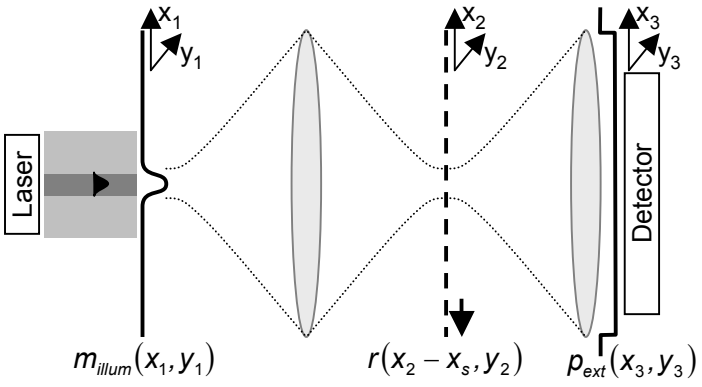


Figure 3.7: A schematic and simplified graph showing the optical light path in a DVD system. A small illumination source is imaged onto the disc, the reflected field is captured with a large detector.

Similarly to the derivation of the waveguide detection system one can derive the optical properties of the DVD system, or extended detector system. Fig. 3.7 shows a schematic view of the optical path in a DVD system. The disc is illuminated by a spot $\psi_{2^-}(x_2, x_s, y_2)$ and the reflected field is captured in the Fourier plane by a detector as large as the aperture of the detection lens. In a real DVD system the configuration for illuminating the disc is a little different from the illumination with a waveguide but the properties of the actual spot are very similar. The reflection by the disc can be written as:

$$\psi_{2^+}(x_2, x_s, y_2) = \psi_{2^-}(x_2, y_2)\, r(x_2 - x_s, y_2)\,. \tag{3.44}$$

The big difference is now that the reflected field is directly picked up by a detector. The exact position of the detector along the Z-axis is not important, as long as all the light that passes the lens is picked up by the detector. For reasons of clarity we placed the detector just behind the lens, and define the detector aperture, p_{ext}, and lens aperture, p_{lens}, to be equal. Eq. (2.36) learns that the field at this position is the Fourier transform of the field reflected from the disc multiplied with the aperture function of the lens, $p_{lens}(x_3, y_3)$.

$$\psi_{3^-}(x_3, x_s, y_3) = \frac{p_{lens}(x_3, y_3)}{\lambda z_{disc}} \iint \psi_{2^+}(x_2, x_s, y_2) \times exp\left[\frac{-2\pi j}{\lambda z_{disc}}\,(x_2 x_3 + y_2 y_3)\right] \mathrm{d}x_2\, \mathrm{d}y_2 \tag{3.45}$$

The intensity at the detector is then

$$I(x_s) = \iint |\psi_{3^-}(x_3, x_s, y_3)|^2 p_{ext}(x_3, y_3)\, \mathrm{d}x_3\, \mathrm{d}y_3, \tag{3.46}$$

or

$$I(x_s) = \iiiint r(x_2 - x_s, y_2)\, r^*\left(x_2' - x_s, y_2'\right) \psi_{2^-}(x_2, y_2)\, \psi_{2^-}^*\left(x_2', y_2'\right) \times \frac{1}{(\lambda z_{disc})^2} \iint p_{ext}(x_3, y_3)\, p_{lens}(x_3, y_3)\, p_{lens}^*(x_3, y_3) \times exp\left\{\frac{-2\pi j}{\lambda z_{disc}}\left[\left(x_2 - x_2'\right) x_3 + \left(y_2 - y_2'\right) y_3\right]\right\} \mathrm{d}x_3\, \mathrm{d}y_3\, \mathrm{d}x_2\, \mathrm{d}y_2\, \mathrm{d}x_2'\, \mathrm{d}y_2'. \tag{3.47}$$

The lens aperture is equal to the detector size, and can be omitted. Eventually we find an equation similar to that of the waveguide system

Eq. (3.5):

$$I(x_s) = \iiiint R\big(x - x_s, y; x' - x_s, y'\big)\, PSF_{tot}(x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y' \tag{3.48}$$

$$\begin{aligned}\text{with } PSF\big(x,y;x',y'\big) = \frac{1}{(\lambda z_{disc})^2}\psi_{2^-}(x,y)\,\psi^*_{2^-}\big(x',y'\big) \\ \times \widehat{p_{ext}}\left(\frac{x-x'}{\lambda z_{disc}}, \frac{y-y'}{\lambda z_{disc}}\right).\end{aligned} \tag{3.49}$$

If we assume that the illumination is done as in the waveguide system we can replace $\psi_{2^-}(x,y)$ by:

$$\psi_{2^-}(x,y) = \iint m^{\gamma}(x_1,y_1)\, psf_{lens}(x-x_1, y-y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1. \tag{3.50}$$

After some calculations (see appendix B.2) we find:

$$\begin{aligned}OTF\big(f,g;f',g'\big) = \iint \psi_{2^-}(f+u,g+v)\,\psi_{2^-}\big(f'+u,g'+v\big) \\ \times p_{ext}(\lambda z_{disc}u, \lambda z_{disc}v)\,\mathrm{d}u\,\mathrm{d}v,\end{aligned} \tag{3.51}$$

or written out

$$\begin{aligned}OTF\big(f,g;f',g'\big) = \iint p_{ext}(\lambda z_{disc}u, \lambda z_{disc}v) \\ \times \widehat{m^{\gamma}_{illum}}[f+u,g+v]\, p_{lens}[\lambda z_{disc}\,(f+u)\,, \lambda z_{disc}\,(g+v)] \\ \times \widehat{m^{\gamma}_{illum}}^{\,*}\big[f'+u,g'+v\big]\, p_{lens}\big[\lambda z_{disc}\,\big(f'+u\big)\,, \lambda z_{disc}\,\big(g'+v\big)\big]\,\mathrm{d}u\,\mathrm{d}v.\end{aligned} \tag{3.52}$$

Similarly as for the waveguide detection system in section 3.2.2 one can understand the influence of the different factors by visualizing the calculation of Eq. (3.51) as an overlap integral of three two dimensional functions. For reasons of clarity we assumed in Fig. 3.8 aberration-free circular lenses and a point-like illumination source. Under these assumptions the three functions are circular step functions with radius $\frac{2NA}{\lambda}$. And the OTF can be visualized by the overlap region of three identical circles within the (u,v)-plane. The calculation is somewhat harder than for the waveguide detection system. The OTF along the x-axis ($g = g' = 0$), shown on Fig. 3.9, can however easily be derived

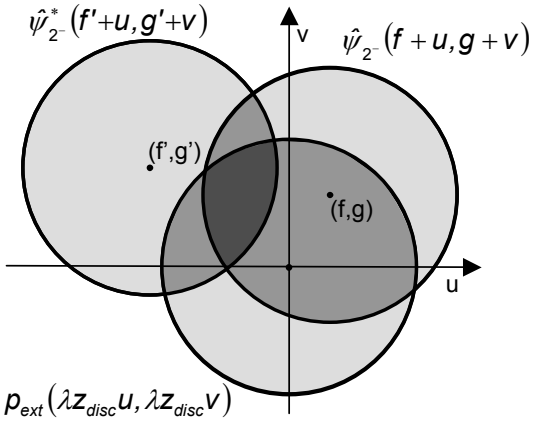


Figure 3.8: The calculation of the OTF for the extended detector system can be visualized by the intersection of three functions in the two-dimensional plane (u, v). For aberration-free circular lenses the OTF is proportional to the area of the intersection of three circles.

as:

$$OTF\left(f, 0; f', 0\right) = \frac{2}{\pi} \arccos\left(\frac{\overline{q}}{2}\right) - \frac{\overline{q}}{2}\sqrt{1 - \left(\frac{\overline{q}}{2}\right)} \qquad (3.53)$$
$$\text{with } \overline{q} = \min\left[\max\left(0, \overline{f}, \overline{f'}\right) - \min\left(0, \overline{f}, \overline{f'}\right), 2\right].$$

3.2.4 Confocal microscope

The working principle of the confocal microscope is very similar to that of the waveguide system. Both in the imaging path and in the detection path a pinhole blocks most of the light [50]. Fig. 3.10 shows a schematic view of the optical path in a confocal microscope. The actual difference between both systems even disappears for very small pinholes.

Similar to the waveguide detection system, the disc is illuminated by a spot $\psi_{2^-}\left(x_2, y_2, z = z_d^-\right)$ and the reflected field is imaged on a pinhole. The actual configuration for illuminating the disc is in reality a little different than the illumination with a waveguide but the eventual properties of the actual spot are very similar and can be described by Eq. (3.29). Using these equations we find the same optical field in plane

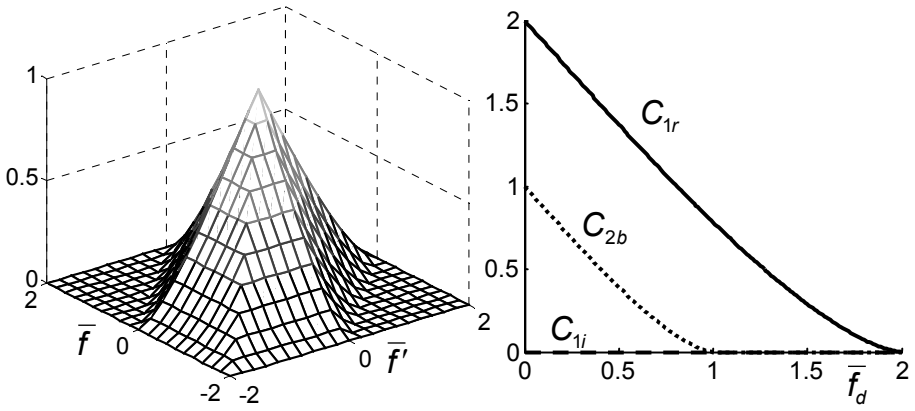


Figure 3.9: On the left the normalized partially coherent OTF along the x-axis of the extended detector system for circular aberration-free lenses and point-like waveguide modes. On the right the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15).

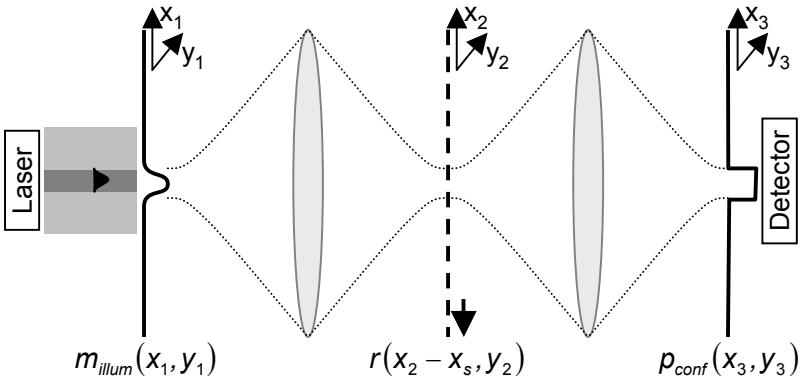


Figure 3.10: A schematic and simplified graph showing the optical light path in a confocal microscope.

3^- as for the waveguide detection method:

$$\psi_3(x_3, x_s, y_3) = \iint \psi_{2^-}(x_2, y_2)\, r(x_2 - x_s, y_2) \times \gamma psf'_{lens}(x_3 - \gamma x_2, y_3 - \gamma y_2)\, \mathrm{d}x_2\, \mathrm{d}y_2. \quad (3.54)$$

The difference with the waveguide system is that the field is not picked up by a waveguide but by a pinhole $p_{conf}(x_3, y_3)$, placed just before the

detector:

$$I(x_s) = \iint |\psi_3(x_3, x_s, y_3)|^2 p_{conf}(x_3, y_3)\,\mathrm{d}x_3\,\mathrm{d}y_3. \tag{3.55}$$

In an actual configuration the detector is placed further away from the pinhole with or without a supplementary lens system in between. As long as all the light form the pinhole is captured this leads however not to a different detector signal [49].

Filling in Eq. (3.54) the intensity becomes

$$\begin{aligned} I(x_s) = &\iint \\ &\times \iint \psi_{2^-}(x_2, y_2)\, r(x_2 - x_s, y_2)\, \gamma psf'_{lens}(x_3 - \gamma x_2, y_3 - \gamma y_2)\,\mathrm{d}x_2\,\mathrm{d}y_2 \\ &\times \iint \psi^*_{2^-}(x'_2, y'_2)\, r^*(x'_2 - x_s, y'_2)\, \gamma psf'^*_{lens}(x_3 - \gamma x'_2, y_3 - \gamma y'_2)\,\mathrm{d}x'_2\,\mathrm{d}y'_2 \\ &\times p_{conf}(x_3, y_3)\,\mathrm{d}x_3\,\mathrm{d}y_3, \end{aligned} \tag{3.56}$$

which by replacing $psf'_{lens}(x, y)$ (Eq. (3.26)) and using the definitions from Eq. (3.27) forms

$$\begin{aligned} I(x_s) = &\iiiint R(x - x_s, y; x' - x_s, y') \\ &\times PSF(x, y; x', y')\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y', \end{aligned} \tag{3.57}$$

$$\begin{aligned} \text{with } PSF(x, y; x', y') = &\psi_{2^-}(x, y)\, \psi^*_{2^-}(x', y') \\ &\times \iint psf_{lens}(x_3 - x, y_3 - y)\, psf^*_{lens}(x_3 - x', y_3 - y') \\ &\times p^{\gamma}_{conf}(x_3, y_3)\,\mathrm{d}x_3\,\mathrm{d}y_3. \end{aligned} \tag{3.58}$$

If one assumes the same illumination proces as for the waveguide system one can fill in the illumination spot ψ_{2^-} by Eq. (3.29).

$$\begin{aligned} PSF(x, y; x', y') = &\iint m^{\gamma}_{illum}(x_1, y_1)\, psf_{lens}(x - x_1, y - y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1 \\ \times &\iint p^{\gamma}_{conf}(x_3, y_3)\, psf_{lens}(x_3 - x, y_3 - y)\, psf^*_{lens}(x_3 - x', y_3 - y')\,\mathrm{d}x_3\,\mathrm{d}y_3 \\ &\times \iint {m^{\gamma}_{illum}}^*(x'_1, y'_1)\, psf^*_{lens}(x' - x'_1, y' - y'_1)\,\mathrm{d}x'_1\,\mathrm{d}y'_1. \end{aligned} \tag{3.59}$$

Some more calculations (see appendix B.3) lead to:

$$\begin{aligned} OTF(f,g;f',g') = \iiiint & \widehat{p^{\gamma}_{conf}}(u'-u,v'-v) \\ & \times p_{lens}(-u\lambda z_{disc},-v\lambda z_{disc})\, p^{*}_{lens}(-u'\lambda z_{disc},-v'\lambda z_{disc}) \\ & \times \widehat{\psi_{2^-}}(f-u,g-v)\, \widehat{\psi_{2^-}}^{*}(f'-u',g'-v')\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}u'\,\mathrm{d}v', \end{aligned} \quad (3.60)$$

or after filling in ψ_{2^-} from Eq. (3.29):

$$\begin{aligned} OTF(f,g;f',g') = \iiiint & \widehat{p^{\gamma}_{conf}}(u'-u,v'-v) \\ & \times p_{lens}(-u\lambda z_{disc},-v\lambda z_{disc})\, p^{*}_{lens}(-u'\lambda z_{disc},-v'\lambda z_{disc}) \\ & \times p_{lens}[(u-f)\,\lambda z_{disc},(v-g)\,\lambda z_{disc}]\, p^{*}_{lens}\left[(u'-f')\,\lambda z_{disc},(v'-g')\,\lambda z_{disc}\right] \\ & \times \widehat{m^{\gamma}_{illum}}(f-u,g-v)\, \widehat{m^{\gamma}_{illum}}^{*}(f'-u',g'-v')\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}u'\,\mathrm{d}v'. \end{aligned} \quad (3.61)$$

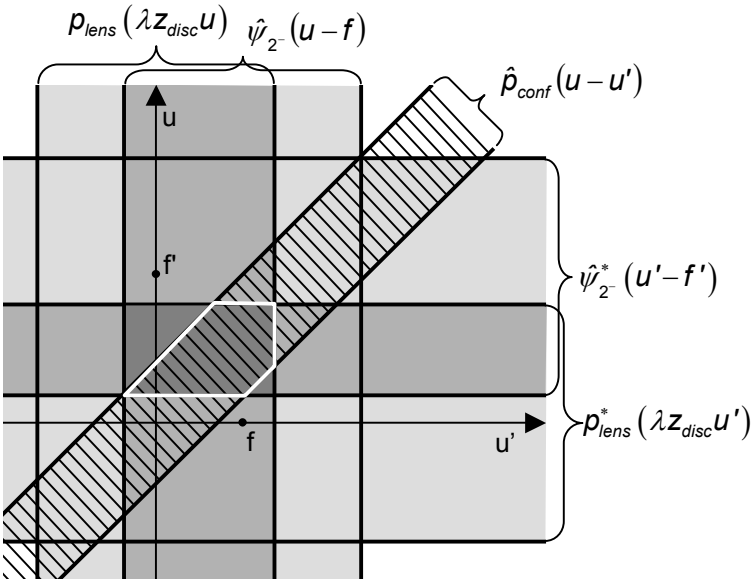


Figure 3.11: The calculation of the OTF for the confocal microscope along the x-axis can be visualized by the intersection of five functions in the two-dimensional plane (u,u'). The OTF is proportional to the area of the intersection, marked with a white border.

The overlap integral is now a combination of five two dimensional functions and four integration variables. For reasons of clarity we reduced the equation to 1 dimension for visualization in Fig. 3.11. In the

graph we assumed aberration-free rectangular lenses and point-like illumination source. Under these assumptions $p_{lens}(u\lambda z_{disc}) = \widehat{\psi_{2^-}}(u) = \Pi\left(\frac{u\lambda}{NA}\right)$. These form gray shaded areas. The overlap is the dark central rectangle. The intensity is the overlap of this rectangle with the hatched region which shows the spatial frequencies picked up by the confocal pinhole.

3.3 Extensions and approximations

3.3.1 Reduction to two dimensions

In an actual system the circular symmetry of the lenses and the cartesian coordinates for the disc structure lead to complex equations. To give a physical insight we therefore will reduce the equations to two dimensions. To make this possible we need to make some additional approximations. We assume that it is possible to split the two dimensional point spread function into two components, $PSF(x, y; x', y') = PSF_x(x, x')\, PSF_y(y, y')$. This is the case if one can write the following functions as a product of a function along the x-axis and one along the y-axis.

$$psf_{lens}(x, y) = psf_{coh,x}(x)\, psf_{coh,y}(y) \tag{3.62}$$

$$m^{air}_{illum}(x, y) = m^{air}_{illum,x}(x)\, m^{air}_{illum,y}(y) \tag{3.63}$$

$$m^{air}_{detect}(x, y) = m^{air}_{detect,x}(x)\, m^{air}_{detect,y}(y)\,, \quad \text{(waveguide detection method)} \tag{3.64}$$

$$p_{conf}(x, y) = p_{conf,x}(x)\, p_{conf,y}(y)\,, \quad \text{(confocal detection method)} \tag{3.65}$$

$$p_{ext}(x, y) = p_{ext,x}(x)\, p_{ext,y}(y)\,, \quad \text{(extended detector method)} \tag{3.66}$$

For the waveguide modes $m^{air}_{illum}(x, y)$ and $m^{air}_{detect}(x, y)$ we used the effective index theory, explained in more detail in section C.3. For $psf_{lens}(x, y)$, $p_{conf}(x, y)$, $p_{ext}(x, y)$ this is possible if we assume rectangular apertures and lenses instead of a circular symmetry.

A comparison between Fig. 3.6 and Fig. 3.18 shows that a rectangular lens results in the same cut-off frequency as a circular lens. The descent of the OTF towards higher spatial frequencies for the circular lens is however steeper than for the rectangular lens. Which results in a narrower point spread function. When comparing two different detection methods one should therefore always use the same rectangular

approximation for the lenses in each method. Although an extrapolation of the results for the rectangular lenses, to circular lenses may quantitatively not exact, the qualitative results are equal.

The intensity at the detector side can then be written as:

$$I(x_s) = \iint PSF_x\big(x;x'\big)\, R_x\big(x-x_s;x'-x_s\big)\,\mathrm{d}x\,\mathrm{d}x' \tag{3.67}$$
$$\text{with } R_x\big(x-x_s,x'-x_s\big) = \iint PSF_y\big(y;y'\big)\, R\big(x-x_s,y;x'-x_s,y'\big)\,\mathrm{d}y\,\mathrm{d}y'.$$

Assume now that the disc consists of tracks with rectangular pits. Under scalar approximation the disc can be modelled as a 2D reflection function. The reflectivity of the land surrounding the pits is A and the reflectivity of the pit-area is B, with an additional phase shift ϕ induced by the difference in optical path length. If the depth of the pit is $\frac{\lambda}{4n}$, $\phi = \pi$. The width of the pit is W_p and $bp(x)$ is the bit sequence. $bp(x) = 1$ for a pit and a $bp(x) = 0$ for no pit (land). The 2D complex reflection function of the disc, $r(x,y)$, can then be written as

$$r(x,y) = A + \Big(Be^{j\phi} - A\Big)\, bp(x)\, \Pi\left(\frac{y}{W_p}\right). \tag{3.68}$$

Before applying Eq. (3.67), we make a few definitions:

$$\iint PSF_y\big(y;y'\big)\,\mathrm{d}y\,\mathrm{d}y' = |Y|^2 \tag{3.69}$$

$$\iint PSF_y\big(y;y'\big)\,\Pi\!\left(\frac{y}{W_p}\right)\mathrm{d}y\,\mathrm{d}y' = \eta\,|Y|^2 \tag{3.70}$$

$$\left[\iint PSF_y\big(y;y'\big)\,\Pi\!\left(\frac{y}{W_p}\right)\Pi\!\left(\frac{y'}{W_p}\right)\mathrm{d}y\,\mathrm{d}y'\right] = (1-\zeta)\,|\eta Y|^2. \tag{3.71}$$

For fully coherent detection systems, such as the waveguide detection system, $PSF_y(y;y')$=$psf_y(y)\,psf_y^*(y')$, and $\zeta = 0$. For other systems this is only true for very large W_p, or very small W_p. A numeric evaluation of ζ for an aberration-free extended detector system with rectangular lenses shows however that even for intermediate values of W_p the deviation is smaller than 6%. (Fig. 3.12) Under the approximation $\zeta = 0$, the disc reflectivity can be written as a function only varying along the x-axis:

$$\begin{aligned} R_x\big(x,x'\big) &= r_x(x)\, r_x^*\big(x'\big) \\ r_x(x) &= YA - \eta Y\left[A - Be^{j\phi}\right] bp(x). \end{aligned} \tag{3.72}$$

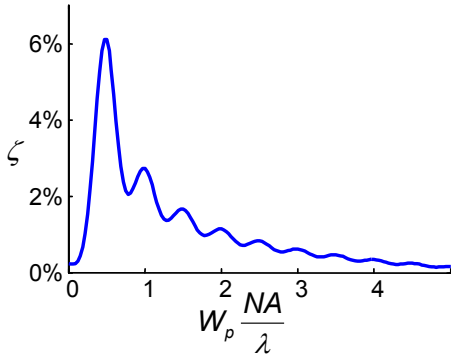


Figure 3.12: The deviation function ζ in Eq. (3.71) as a function of the track width, for the extended detector system. This function gives a figure of merit for the error involved in approximating the 2D-optical transfer function by a 1D-optical transfer function.

If we assume that the spot covers half of the pit, and the reflectivity of land and pit are equal in amplitude ($\eta = \frac{1}{2}$ and $A = B$), then the effective reflectivity for a pit would be zero and the effective reflectivity for a land would be $AY\left(1 + e^{j\phi}\right)/2$.

Rewriting Eq. (3.72) in the form of Eq. (3.12) ($r_x(x) \propto 1 + \varepsilon\rho(x)$) we find:

$$\varepsilon = \frac{A - Be^{j\phi}}{2A/\eta - A + Be^{j\phi}} \qquad \text{and} \qquad \rho(x) = 1 - 2bp(x)\,, \tag{3.73}$$

or with ($\eta = \frac{1}{2}$ and $A = B$)

$$\begin{aligned} \varepsilon &= \frac{1 - e^{j\phi}}{3 + e^{j\phi}} \\ &= \frac{1}{2}\left[1 - e^{j\theta(\phi)}\right] \qquad \theta(\phi) = \arcsin\left(\frac{4\sin\phi}{5 + 3\cos\phi}\right). \end{aligned} \tag{3.74}$$

Fig. 3.13 shows ε as a function of ϕ. For $\phi = \pi$, ε is real. For other values of ϕ, ε has also an imaginary compound. As described in section 3.1.5, the phase of ε is important. For a real ε, the first order detector signal is only proportional to $\mathcal{C}_{1r}$, for complex ε there is also an influence from $\mathcal{C}_{1i}$. For a disc with multiple tracks the reflection of the disc can be written as

$$r(x,y) = A + \sum_{k=-n}^{n}\left(Be^{j\phi} - A\right) bp_k(x)\,\Pi\left(\frac{y + kW_t}{W_p}\right). \tag{3.75}$$

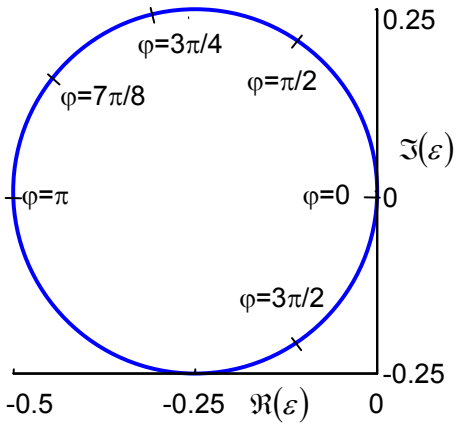


Figure 3.13: Effective modulation depth (ε) as a function of the phase difference between pits and lands(ϕ).

Analogically to a single track the intensity for a coherent optical system can be written by a simple equation using an adapted bit pattern:

$$I(x_s) = \iint r_x(x)\, r_x^*(x')\, PSF_x(x, x')\, \mathrm{d}x\, \mathrm{d}x'$$

$$\text{with} \begin{cases} r_x(x) = AY + Y \sum_{k=-n}^{n} \eta_k bp_k(x) \left(Be^{j\phi} - A \right) \\ \eta_k Y = \iint PSF_y(y; y')\, \Pi\left(\dfrac{y + kW_t}{W_p} \right) \mathrm{d}y\, \mathrm{d}y'. \end{cases} \tag{3.76}$$

Using the equations in this section the y-dimension can be fully omitted. In the remainder of this chapter we will therefore only describe the point spread function and optical transfer function in the x-dimension.

3.3.2 Tilted waveguide

A possible extension to the scanning waveguide detection method, proposed by Pereira [51], is to tilt the waveguide with respect to the disc by an angle θ. One can simulate the tilted waveguide by finding a straight waveguide with the same optical field. One way of doing so is replacing the field of the tilted waveguide by a range of point sources (Fig. 3.14a).

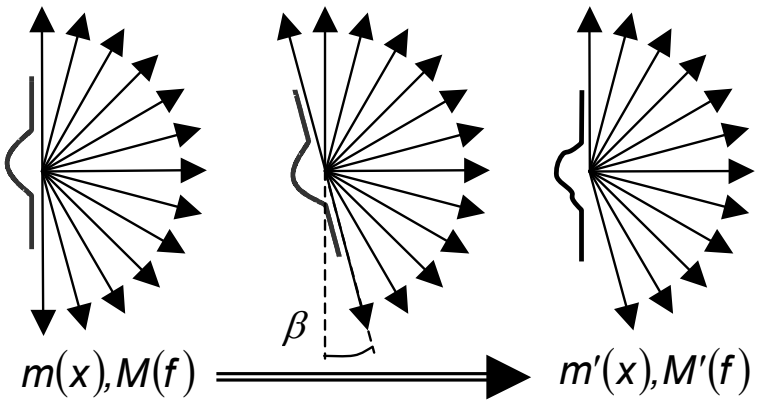


Figure 3.14: Two methods for calculating a straight waveguide equivalent of a tilted waveguide. On the left: point sources, on the right: plane waves.

Similarly as in Eq. (2.33) the optical field at a distance z, in two dimensions can be written as:

$$psf_{lens}(x) = \frac{e^{j\frac{\pi}{4}}}{\sqrt{\lambda r}} \exp\left(-j\frac{2\pi}{\lambda}r\right)\frac{z}{r}. \tag{3.77}$$

Let $m_t(y)$ be the tilted waveguide and $m(x)$ the equivalent straight waveguide mode at $z = z_0$. Then one finds:

$$m(x) = \int m_t(y)\, psf\left(x - y\cos(\theta)\,, z_0 - \sin(\theta)\right) \mathrm{d}y. \tag{3.78}$$

In case there is an imaging system the obvious choice for z_0 is zero. For the near-field there is no unambiguous value possible, but one could make do with $z_0 = a/2$, a being the mode diameter.

When using a lens system and when only propagating waves are transmitted, one can also calculate the equivalent straight waveguide in a different way. The optical field of the tilted waveguide can be expanded into plane waves. By turning the coordinates these can also be interpreted as the plane waves of the straight waveguide. One has then:

$$\widehat{m}(f_x) = \begin{cases} \widehat{m_t}\left(f_0\cos\left[\left(\arccos\frac{f_x}{f_0}\right) + \theta\right]\right) & f_x \in \left[-f_0, f_0\cos\left(\frac{\pi}{2} - \theta\right)\right] \\ 0 & f_x \in \left[f_0\cos\left(\frac{\pi}{2} - \theta\right), f_0\right]. \end{cases} \tag{3.79}$$

With $f_0 = 2\pi/\lambda$. For very narrow waveguides, the waveguide modes can be seen as a single point source and tilting the waveguide has no

effect. For larger waveguides the left and right part of the waveguide mode will be increasingly out of focus. On Fig. 3.15 the effect of tilting the waveguide is shown for a waveguide profile $\exp\left(-\pi\overline{x}^2/4\right)$.

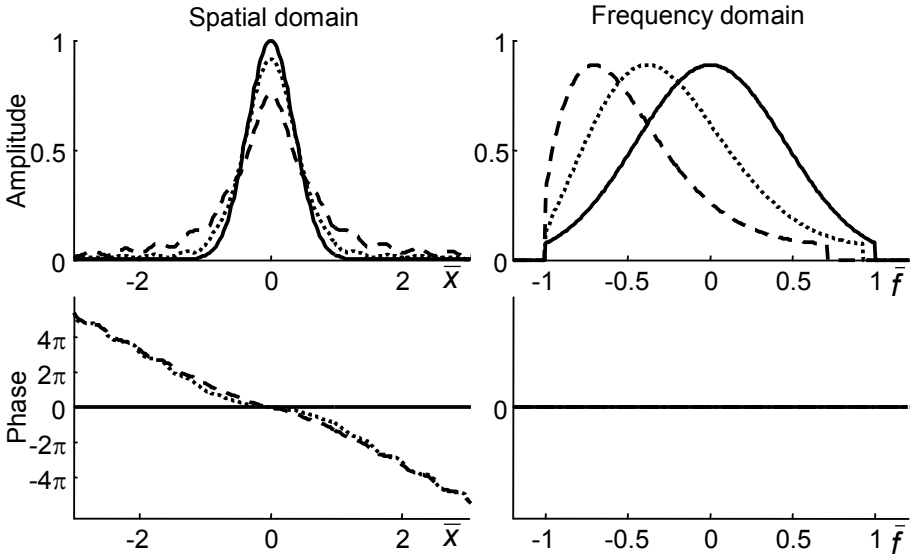


Figure 3.15: Amplitude and phase of the straight waveguide equivalents of a tilted waveguide, for a few tilt angles. On the left in the spatial domain, on the right in the frequency domain. The waveguide mode profile of the untilted waveguide is $\exp\left(-\pi x^2/4\right)$ for different tilt angles. In straight line $\theta = 0$, in dotted line $\theta = \pi/8$, and in dashed line $\theta = \pi/4$.

3.3.3 Double waveguide

As an extension to the multimodal waveguide one could consider a structure consisting of two single mode waveguides placed within a short distance on the chip. The idea is to let these modes interfere within a photonic IC. By adjusting the phase delay θ in between these modes different combinations can be formed. In principle two unequal waveguides could be used. We will keep ourselves however to two identical waveguides spaced a distance d from each other. The combination of these two modes at the detector is seen as a new mode:

$$m_a(x) = m\left(x - \tfrac{d}{2}\right) e^{+j\frac{\theta}{2}} \qquad m_b(x) = m\left(x - \tfrac{d}{2}\right) e^{-j\frac{\theta}{2}} \tag{3.80}$$

$$m_d(x) = m_a(x) + m_b(x) \qquad \widehat{m_d}(f) = 2\widehat{m}(f)\cos\left(\pi df + \tfrac{\theta}{2}\right). \tag{3.81}$$

If one would look to the two separate waveguide as a single waveguide structure, one would find two supermodes. As shown on Fig. 3.16 the zeroth order supermode, $m_0^S \sim m_a + m_b$, can be approximated by the sum of the two modes of the separate waveguides if they are in phase. If they are in anti-phase the sum is approximately equal to the first order supermode, $m_1^S \sim m_a - m_b$.

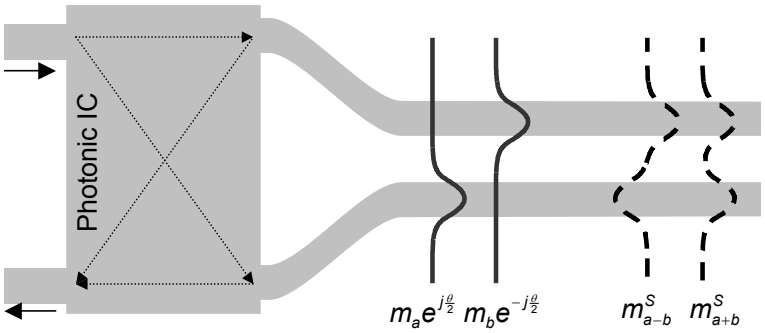


Figure 3.16: Two single mode waveguides in proximity form an equivalent of a single bimodal waveguide. If the modes in both waveguides are in phase they form the zeroth order supermode. If the modes in both waveguides are in anti-phase they form the first order supermode. The phase differences in between the two modes can be controlled with a photonic IC.

3.3.4 Approximations

To give a deeper insight in which parameters influence the optical systems, it is very interesting to find an analytic solution for the point spread function or the optical transfer function. To do so a few approximations have to be made for the lenses, the waveguide modes and/or the apertures.

Lens For the lens we used an aberration-free rectangular lens as in Eq. (2.25), which gives

$$otf_{lens}(f) = \Pi\left(\frac{f\lambda}{2NA}\right) \qquad psf_{lens}(x) = \frac{2NA}{\lambda}\frac{\sin\left(\frac{2\pi NA}{\lambda}x\right)}{\left(\frac{2\pi NA}{\lambda}x\right)}. \tag{3.82}$$

Waveguide modes To a certain extent the waveguide modes can be approximated by Gaussian functions or its derivatives. We will use the Gaussian function definition $\mathcal{G}\left(\frac{x}{a}\right) = \exp\left(-\pi\frac{x^2}{a^2}\right)$ as defined by

Eq. (A.4) in the appendix A . For reasons of clarity we will use the normalization

$$\int \left|m^{air}(x)\right|^2 \mathrm{d}x = 1. \tag{3.83}$$

This is not entirely correct as the shape and amplitude of the modes changes after transmission through the waveguide-air interface. For qualitative results this is however a reasonable approximation.

The zeroth order waveguide mode can be approximated by a single Gaussian function. For later use the Fourier transform is also given:

$$m^{0,air}(x) \approx \sqrt{\frac{\sqrt{2}}{a}}\,\mathcal{G}\Big(\frac{x}{a}\Big) \tag{3.84}$$

$$\widehat{m^{0,air}}(f) \approx \sqrt{a\sqrt{2}}\,\mathcal{G}(af)\,. \tag{3.85}$$

The first order waveguide mode is approximated by the derivative of the zeroth order mode. After normalization using Eq. (3.83) we find

$$m^{1,air}(x) \approx \frac{a}{\sqrt{\pi}}\frac{\mathrm{d}m^{0,air}(x)}{\mathrm{d}x} = -\sqrt{2}\sqrt{\frac{\sqrt{2}}{a}}\left(\frac{\sqrt{2\pi}}{a}x\right)\mathcal{G}\Big(\frac{x}{a}\Big) \tag{3.86}$$

$$\widehat{m^{1,air}}(f) \approx j\sqrt{2}\sqrt{a\sqrt{2}}\left(\sqrt{2\pi}af\right)\mathcal{G}(fa)\,. \tag{3.87}$$

One would be tempted to approximate the second order mode with the second derivative of a Gaussian, $G''\big(\frac{x}{a}\big)$. The second order waveguide mode has however three lobes that are approximately equal in amplitude, which is not the case for $G''\big(\frac{x}{a}\big)$. A better approximation is formed by a weighted sum of $G''\big(\frac{x}{a}\big)$ and $G\big(\frac{x}{a}\big)$:

$$m^{2,air}(x) \approx \frac{1}{\sqrt{2}}\left(m^{0,air}(x) + \frac{a^2}{\pi}\frac{\mathrm{d}^2 m^{0,air}(x)}{\mathrm{d}x^2}\right) \tag{3.88}$$

$$\approx -\frac{1}{\sqrt{2}}\sqrt{\frac{\sqrt{2}}{a}}\left[1 - 2\left(\frac{\sqrt{2\pi}}{a}x\right)^2\right]\mathcal{G}\Big(\frac{x}{a}\Big) \tag{3.89}$$

$$\widehat{m^{2,air}}(f) \approx \frac{1}{\sqrt{2}}\sqrt{a\sqrt{2}}\left[1 - 2\left(\sqrt{2\pi}af\right)^2\right]\mathcal{G}(fa)\,. \tag{3.90}$$

Remark that the approximations for the waveguide modes are orthogonal, as should be:

$$\int m^{i,air}(x)\, m^{j,air,*}(x)\,\mathrm{d}x = \begin{cases} 1, & \text{for } i = j \\ 0, & \text{for } i \neq j. \end{cases} \tag{3.91}$$

As an example a comparison for 2D slab waveguide with parameters $w_{wg} = \lambda$, $n_{core} = 3.25$ and $n_{clad} = 3.1$ with the given approximations with $a = 1$ is given in Fig. 3.17. The fit is not perfect. For other waveguide widths and/or different refractive indices, the fit may even be less favourable, but for qualitative analytic results the approximations are good enough.

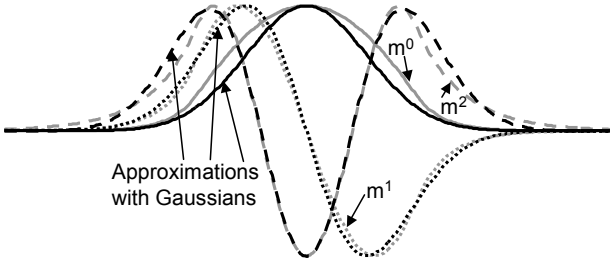


Figure 3.17: An approximation of the waveguide modes by Gaussian functions. Gray solid, dotted and dashed line are respectively the zeroth, first and second order mode. The black solid, dotted and dashed lines are the approximations by Gaussian functions.

For later use we also define a hypothetical mode, $m^{02,air}(x)$, which is a weighted sum of zeroth and second order mode. In principle such a mode could be formed using a photonic IC.

$$m^{02,air}(x) = \xi m^{0,air} - m^{2,air}(x) \tag{3.92}$$

$$= \left(\xi - \frac{1}{\sqrt{2}}\right) m^{0,air}(x) - \frac{a^2}{\sqrt{2}\pi} \frac{\mathrm{d}^2 m^{0,air}(x)}{\mathrm{d}x^2} \tag{3.93}$$

$$\widehat{m^{02,air}}(f) = \left(\xi - \frac{1}{\sqrt{2}}\right) \sqrt{a\sqrt{2}}\, \mathcal{G}(af) + \frac{4a^2\pi}{\sqrt{2}} f^2 \sqrt{a\sqrt{2}}\, \mathcal{G}(af)\,. \tag{3.94}$$

Remark that for $\xi = \frac{1}{\sqrt{2}}$, $m^{02,air}(x)$ is proportional to the second derivative of the zeroth order mode.

Apertures The aperture in front of the detector in the extended detector system is identical to the aperture of the lens system. Both are approximated by a rectangle function with width D: $p_{ext}(x) = \Pi(\frac{x}{D})$. Within the approximations used we have $\frac{D}{z_{disc}} = NA$. The Fourier transform can easily be calculated: $\widehat{p_{ext}}(f) = a_{ext}\,\mathrm{sinc}(a_{ext} f)$. Similarly the confocal aperture, $p_{conf}(x)$, can be approximated by the rectangle function $\Pi\left(\frac{x}{a_{conf}}\right)$.

3.4 Waveguide detection method

In this section we will use the approximations of previous section to analyse the optical transfer function of the waveguide detection systems for a few configurations. In the first subsection the otf and psf for the zeroth order for the illumination and detection will be derived. In the section subsection we will inspect the difference if the illumination and/or the detection mode is a first order mode. A third section covers the properties for higher order modes. In the last subsections we will describe the optical properties of the near-field configuration, the tilted waveguide and the double waveguide detector.

3.4.1 Zeroth order mode

Using the general equation of the OTF for the waveguide detection method from Eq. (3.41), the far-field OTF for the rectangular lens can be written as:

$$\begin{aligned} OTF(f;f') &= otf_{tot}(f)\, otf^{*}_{tot}(f') \\ &\text{with } otf_{tot}(f) = \int_{-\infty}^{\infty} \Pi\!\left(\frac{u-f}{2NA}\lambda\right) m^{\gamma}_{illum}(u-f) \\ &\qquad \times \Pi\!\left(\frac{u}{2NA}\lambda\right) m^{\gamma}_{detect}(u)\,\mathrm{d}u, \end{aligned} \tag{3.95}$$

or even simpler:

$$otf(f) = \begin{cases} 0, & \text{for } |f| > \frac{2NA}{\lambda} \\ \int_{-\frac{NA}{\lambda}+\max(0,f)}^{\frac{NA}{\lambda}+\min(0,f)} \widehat{m^{\gamma}_{illum}}(u-f)\, \widehat{m^{\gamma}_{detect}}(u)\,\mathrm{d}u, & \text{for } |f| < \frac{2NA}{\lambda}. \end{cases} \tag{3.96}$$

By filling in the expressions for the waveguide modes the optical transfer function can be calculated.

If the zeroth order mode is used for the illumination as well as for the detection one finds relatively simple analytic solution:

$$\begin{aligned} otf^{0,0}(f) &= \frac{a\sqrt{2}}{\gamma} \int_{-\frac{NA}{\lambda}+\max(0,f)}^{\frac{NA}{\lambda}+\min(0,f)} \mathcal{G}\left[\frac{a}{\gamma}(u-f)\right] \mathcal{G}\left[\frac{a}{\gamma}u\right] \mathrm{d}u \\ &= \frac{a\sqrt{2}}{\gamma} \int_{-\frac{NA}{\lambda}+\max(0,f)}^{\frac{NA}{\lambda}+\min(0,f)} \mathcal{G}\left[\frac{a}{\gamma\sqrt{2}}f\right] \mathcal{G}\left[\frac{a}{\gamma\sqrt{2}}(2u-f)\right] \mathrm{d}u \\ &= \frac{1}{2}\,\mathcal{G}\left(\frac{a}{\gamma\sqrt{2}}f\right) \left[\operatorname{erf}\left(\sqrt{\frac{\pi}{2}}\frac{a}{\gamma}\left[2u-f\right]\right)\right]_{u=-\frac{NA}{\lambda}+\max(0,f)}^{u=\frac{NA}{\lambda}+\min(0,f)} \\ &= \mathcal{G}\left(\frac{a}{\gamma\sqrt{2}}f\right) \operatorname{erf}\left[\sqrt{\frac{\pi}{2}}\frac{a}{\gamma}\left(\frac{2NA}{\lambda}-|f|\right)\right], \end{aligned} \tag{3.97}$$

which can further be simplified by normalizing some of the parameters. Using

$$\overline{f} = \frac{f\lambda}{NA}, \qquad \overline{x} = \frac{xNA}{\lambda}, \qquad \text{and } \overline{a} = \frac{aNA}{\gamma\lambda}, \tag{3.98}$$

one finds:

$$otf^{0,0}(\overline{f}) = \mathcal{G}\left(\frac{\overline{a}\overline{f}}{\sqrt{2}}\right) \operatorname{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right] \tag{3.99}$$

$$\begin{aligned} OTF^{0,0}(\overline{f},\overline{f'}) = \mathcal{G}\left(\frac{\overline{a}\overline{f}}{\sqrt{2}}\right) \mathcal{G}\left(\frac{\overline{a}\overline{f'}}{\sqrt{2}}\right) \\ \times \operatorname{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right] \operatorname{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f'}|\right)\right]. \end{aligned} \tag{3.100}$$

For small $\overline{a}$ this is:

$$otf^{0,0}(\overline{f}) = \sqrt{2}\overline{a}\left(2-|\overline{f}|\right) \tag{3.101}$$

$$psf^{0,0}(\overline{x}) = 4\sqrt{2}\overline{a}\,\operatorname{sinc}^2(2\pi\overline{x})\,. \tag{3.102}$$

Fig. 3.18 shows the optical transfer function for $\overline{a} \ll 1$ and normalized to 1. It is maximal for $\overline{f} = \overline{f'} = 0$ and decreases steadily towards $|\overline{f}| = 2$ or $|\overline{f'}| = 0$.

A closer look to $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ learns:

$$\mathcal{C}_{1r}(\overline{f}) = 2\,\mathcal{G}\left(\frac{\overline{a}\overline{f}}{\sqrt{2}}\right) \operatorname{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right] \tag{3.103}$$

$$\mathcal{C}_{1i}(\overline{f}) = 0 \tag{3.104}$$

$$\mathcal{C}_{2b}(\overline{f}) = \mathcal{G}\left(\sqrt{2}\overline{a}\overline{f}\right) \operatorname{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right] \operatorname{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right]. \tag{3.105}$$

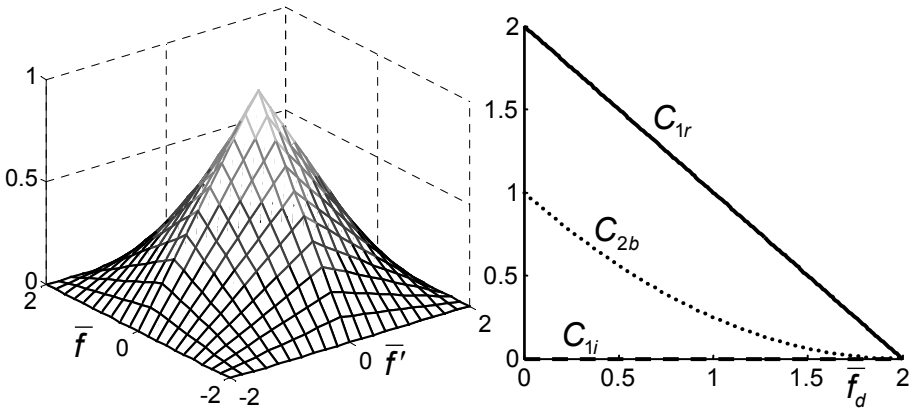


Figure 3.18: On the left the normalized partially coherent $OTF^{0,0}(\overline{f},\overline{f'})$ for small waveguide limit. On the right the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15).

This means that for a simple grating the first order signal and the second order signal both decrease with increasing spatial frequency of the grating. If the modulation of the grating, ε is purely imaginary, there is only a second order signal as $\mathcal{C}_{1i} = 0$

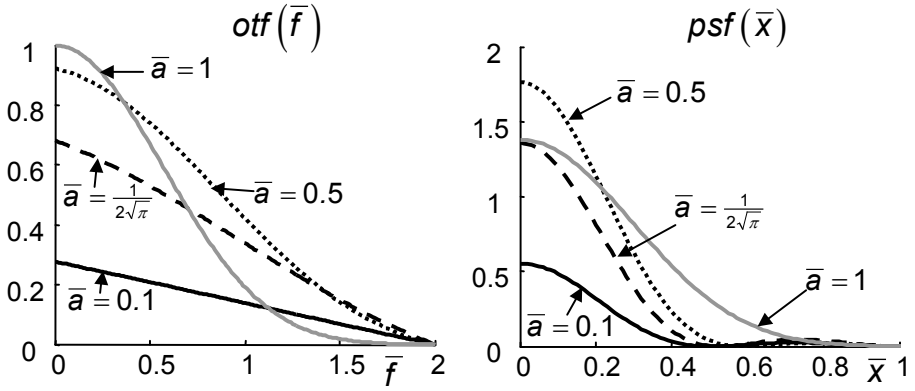


Figure 3.19: The $otf^{0,0}(\overline{f})$ and $psf^{0,0}(\overline{x})$ for a few values of the normalized waveguide width, $\overline{a}$. For reasons of clarity only one half is shown, the otf and the psf are both symmetric.

On Fig. 3.19 the influence is shown of the normalized width of the waveguide, $\overline{a}$, on $otf^{0,0}(\overline{f})$ and $psf^{0,0}(\overline{x})$. For very small values of $\overline{a}$, $otf^{0,0}$ has a triangular shape but the signal is relatively weak. For larger widths the overall otf increases but for higher spatial frequencies the

otf decreases significantly. For very broad waveguides the zeroth order mode can be seen as a plane wave. This means the sample is illuminated by a single plane wave and only the light propagating back in this direction is captured. This means only the zeroth order diffraction is detected and the otf is a Dirac function. The graph suggest that there is an optimum $\overline{a}$ at which the slope of the $otf^{0,0}(\overline{f})$ cross section at $\overline{f} = 2$ is maximal. By taking the derivative one finds that this is the case for $\overline{a} = \frac{1}{2\sqrt{\pi}}$.

Using a beam splitter it would in principle be possible to use a different waveguide for the illumination as for the detection. This would solve the problem of parasitical reflections but one would loose the easy alignment (compared to the confocal microscope). If we define $\overline{a_1} = \frac{a_1 NA}{\gamma\lambda}$ and $\overline{a_2} = \frac{a_2 NA}{\gamma\lambda}$ as respectively the normalized widths of illumination and detection waveguide, we have:

$$\begin{aligned} otf^{0,0}(\overline{f}) = \sqrt{\frac{2\overline{a_1}\overline{a_2}}{a1^2 + a2^2}} \exp\left(-\pi\overline{f}^2 \frac{\overline{a_2}^2\overline{a_1}^2}{\overline{a_1}^2 + \overline{a_2}^2}\right) \\ \times \frac{1}{2}\left\{\operatorname{erf}\left[\sqrt{\pi}\frac{\overline{a_2}^2}{\sqrt{\overline{a_1}^2 + \overline{a_2}^2}}\left(\frac{\overline{a_1}^2 + \overline{a_2}^2}{\overline{a_2}^2} - |\overline{f}|\right)\right]\right. \\ \left. + \operatorname{erf}\left[\sqrt{\pi}\frac{\overline{a_1}^2}{\sqrt{\overline{a_1}^2 + \overline{a_2}^2}}\left(\frac{\overline{a_1}^2 + \overline{a_2}^2}{\overline{a_1}^2} - |\overline{f}|\right)\right]\right\}. \end{aligned} \quad (3.106)$$

For $\overline{a_1} = \overline{a_2}$ one finds Eq. (3.99) back. It is interesting to inspect what happens at extreme values of $\overline{a_1}$ and $\overline{a_2}$. If $\overline{a_2} \ll 1$, one finds:

$$\begin{aligned} otf^{0,0}(\overline{f}) = \sqrt{\frac{2\overline{a_2}}{\overline{a_1}}} \exp\left(-\pi\overline{f}^2\overline{a_2}^2\right) \\ \times \left\{\operatorname{erf}\left[\sqrt{\pi}\overline{a_1}\right] + \operatorname{erf}\left[\sqrt{\pi}\overline{a_1}\left(1 - |\overline{f}|\right)\right]\right\}. \end{aligned} \quad (3.107)$$

And if one also takes the limits for $\overline{a_1}$ one has:

$$otf^{0,0}(\overline{f}) = \begin{cases} \sqrt{2\overline{a_1}\overline{a_2}}\left(2 - |\overline{f}|\right), & \text{for } \overline{a_1} \ll 1 \\ \sqrt{\frac{2\overline{a_2}}{\overline{a_1}}} \exp\left(-\pi\overline{f}^2\overline{a_2}^2\right) \Pi(\overline{f}), & \text{for } \overline{a_1} \gg 1. \end{cases} \quad (3.108)$$

For $\overline{a_1}, \overline{a_2} \ll 1$ one has again the triangularly shaped otf. The proportionality constants show a decreasing overall throughput for smaller widths. For $\overline{a_2} \ll 1 \ll \overline{a_1}$ the illumination is a plane wave and the detection is a point, which explains the rectangular shape of the otf similar to the coherent transfer function of a single lens. Fig. 3.20 shows the otf and psf for a few intermediate values of $\overline{a_1}$ and $\overline{a_2}$.

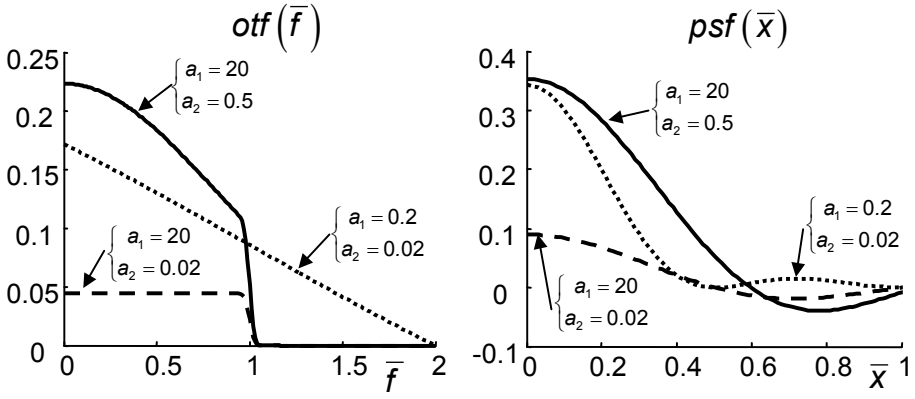


Figure 3.20: The $otf^{0,0}(\overline{f})$ and $psf^{0,0}(\overline{x})$ for the waveguide detection systems for two different waveguide widths, $\overline{a_1}$ and $\overline{a_2}$. For reasons of clarity only one half is shown, the otf and the psf are both symmetric.

3.4.2 First order mode

If one keeps the zeroth order mode for illumination and uses the first order mode for detection, one finds analogously:

$$\begin{aligned} otf^{0,1}(\overline{f}) &= -j2\overline{a}^2\sqrt{2\pi}\int_{-1+\max(0,\overline{f})}^{1+\min(0,\overline{f})} \mathcal{G}[\overline{a}\,(u-f)]\,u\,\mathcal{G}[\overline{a}u]\,\mathrm{d}u \\ &= -j\sqrt{\pi}\overline{a}\overline{f}\exp\left[-\frac{\pi}{2}\left(\overline{a}\overline{f}\right)^2\right]\mathrm{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right] \quad (3.109) \\ OTF^{0,1}(\overline{f,f'}) &= \pi\overline{a}^2\overline{ff'}^*\exp\left[-\frac{\pi}{2}\overline{a}^2\left(\overline{f}+\overline{f'}\right)^2\right] \\ &\quad\times\mathrm{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right]\mathrm{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f'}|\right)\right], \quad (3.110) \end{aligned}$$

or for very small $\overline{a}$:

$$otf^{0,1}(\overline{f}) = -j\sqrt{2\pi}\overline{a}^2\overline{f}\left(2-|\overline{f}|\right) \quad (3.111)$$

$$psf^{0,1}(\overline{x}) = 8\sqrt{2\pi}\overline{a}^2\,\mathrm{sinc}^2(2\pi\overline{x})\left[\frac{1}{2\pi\overline{x}}-\frac{1}{\tan(2\pi\overline{x})}\right]. \quad (3.112)$$

As shown in Fig. 3.21 the partially coherent transfer function is zero for $\overline{f}=0$ or $\overline{f'}=0$. The effect is that C_0, C_{1r}, and C_{1i} are all zero. The only signal retrieved from a grating is the weak second order signal C_0. The effect of the width of the waveguide on the otf is similar to the effect on the otf of the zeroth order mode but stronger. For very broad

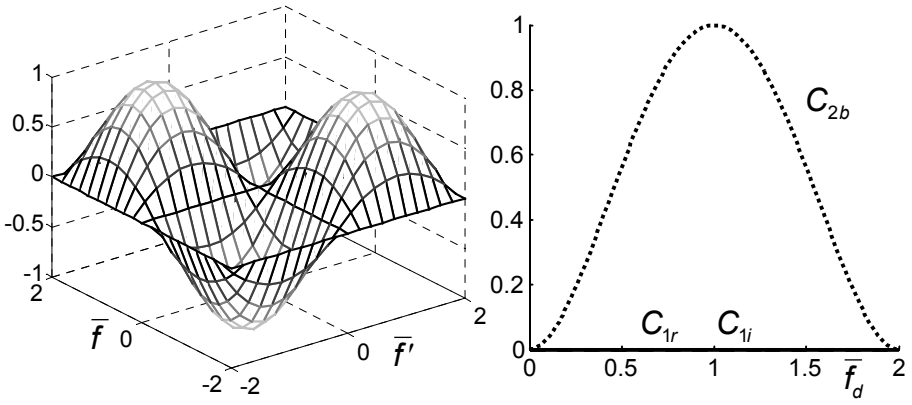


Figure 3.21: On the left the normalized partially coherent OTF for the waveguide detection system (illumination with zeroth order and detection with first order mode), for small waveguide limit: $OTF^{0,1}(f, f')$. On the right the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15).

waveguides the otf is squeezed towards the lower spatial frequencies, for very narrow waveguide modes the overall throughput is reduced. Fig. 3.22 shows the influence of $\overline{a}$ on the otf and psf.

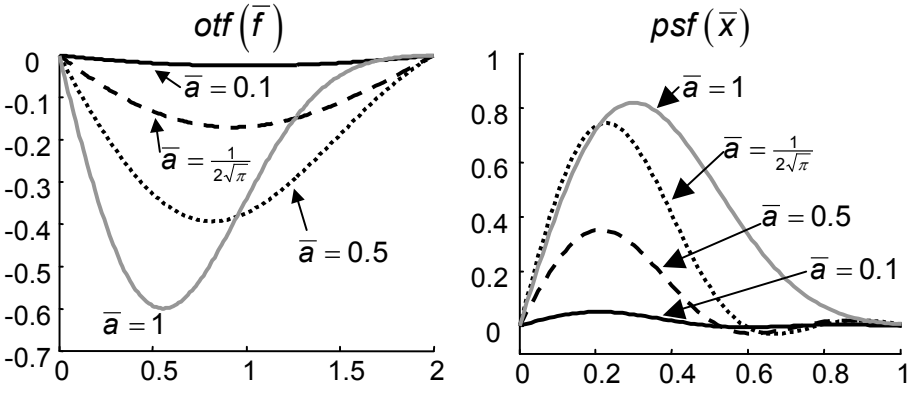


Figure 3.22: The $otf^{0,1}\left(\overline{f}\right)$ and $psf^{0,1}(\overline{x})$ for a few values of the normalized waveguide width, $\overline{a}$. Both function are antisymmetric. The otf is imaginary, and the psf is real.

The detector signal from the first order mode can also be interpreted in a different way. Using Eq. (3.86) and Eq. (3.27), one finds

$$psf_{tot}^{0,1}(x) = psf_{illum}^{0}(x) \left\{ psf_{lens}(x) \otimes \frac{1}{\gamma} m^{1,air}\left(\frac{x}{\gamma}\right) \right\} \tag{3.113}$$

$$= psf_{illum}^{0}(x) \left\{ psf_{lens}(x) \otimes \frac{1}{\gamma}\frac{a}{\sqrt{\pi}}\frac{\mathrm{d}}{\mathrm{d}x}\left[m^{0,air}\left(\frac{x}{\gamma}\right)\right]\right\} \tag{3.114}$$

$$= \frac{1}{\gamma}\frac{a}{\sqrt{\pi}} psf_{illum}^{0}(x)\, \frac{\mathrm{d}}{\mathrm{d}x}\left\{ psf_{lens}(x) \otimes m^{0,air}\left(\frac{x}{\gamma}\right)\right\} \tag{3.115}$$

$$= \frac{1}{\gamma}\frac{a}{\sqrt{\pi}} psf_{illum}^{0}(x)\, \frac{\mathrm{d}}{\mathrm{d}x}\left[psf_{detect}^{0}(x)\right] \tag{3.116}$$

$$= \frac{a}{2\gamma\sqrt{\pi}}\frac{\mathrm{d}}{\mathrm{d}x}\left[psf_{tot}^{0,0}(x)\right], \tag{3.117}$$

or

$$otf_{tot}^{0,1}(f) = -j\sqrt{\pi}\overline{a}\,\overline{f}\, otf_{tot}^{0,0}(x)\,. \tag{3.118}$$

Assume one defines the phase and amplitude of the excitation of the zeroth order mode by $psf^{0,0} \otimes r = A \exp j\phi$. If there would be no diffraction then this would A and ϕ would be the amplitude and phase of the disc reflectivity itself. We will now show that by detecting the zeroth and first order mode it is possible to detect the amplitude and the derivative of the phase. Starting from Eq. (3.5) we have

$$I^{0,0} = \left| psf^{0,0} \otimes r \right|^2 \qquad I^{0,1} = \left| psf^{0,1} \otimes r \right|^2 \tag{3.119}$$

$$\sqrt{I^{0,0}} = A \qquad \sqrt{I^{0,1}} = \frac{a}{2\gamma\sqrt{\pi}}\left| \frac{\mathrm{d}A}{\mathrm{d}x} + A\frac{\mathrm{d}\phi}{\mathrm{d}x}\right| \tag{3.120}$$

From the last equation a solution for the derivative of the phase can be found based on the two output signals $I^{0,0}$ and $I^{0,1}$. There is however ambiguity in the sign of one of the terms, which means we get two solutions form which only one is correct.

$$\frac{\mathrm{d}\phi}{\mathrm{d}x} = \frac{\pm\frac{2\gamma\sqrt{\pi}}{a}\sqrt{I^{0,1}} - \frac{\mathrm{d}\sqrt{I^{0,0}}}{\mathrm{d}x}}{\sqrt{I^{0,0}}} \tag{3.121}$$

One might think that this relationships strongly depends on the specific approximations we made in Eq. (3.86). This is not true as long as the total light throughput is unimportant and the zeroth order and first order mode are respectively symmetric and antisymmetric. The Taylor

expansion of $\widehat{m^{0,\gamma}}(f)$ and $\widehat{m^{1,\gamma}}(f)$ around $f = 0$ can then be written as:

$$\widehat{m^{0,\gamma}}(f) = \frac{1}{\sqrt{\gamma}}\widehat{m^{0,air}}\left(\frac{f}{\gamma}\right) = \frac{1}{\sqrt{\gamma}}\left\{M_0 + \mathcal{O}\left[\left(\frac{f}{\gamma}\right)^2\right]\right\} \tag{3.122}$$

$$\widehat{m^{1,\gamma}}(f) = \frac{1}{\sqrt{\gamma}}\widehat{m^{1,air}}\left(\frac{f}{\gamma}\right) = j\frac{1}{\sqrt{\gamma}}\left\{M_1\frac{f}{\gamma} + \mathcal{O}\left[\left(\frac{f}{\gamma}\right)^3\right]\right\} \tag{3.123}$$

and according to Eq. (3.96) the

$$otf^{0,1}(f) = \int_{-NA/\lambda+\max(0,f)}^{NA/\lambda+\min(0,f)} \widehat{m^{0,\gamma}}(u-f)\,\widehat{m^{1,\gamma}}(u)\,\mathrm{d}u \tag{3.124}$$

$$= j\frac{M_0M_1}{\gamma}\left[\frac{u}{\gamma} + \mathcal{O}\left[\left(\frac{u}{\gamma}\right)^3\right]\right]_{-NA/\lambda+\max(0,f)}^{NA/\lambda+\min(0,f)} \tag{3.125}$$

$$= j\frac{M_0M_1}{2\gamma^2}f\left(2\frac{NA}{\lambda} - |f|\right) + \mathcal{O}\left(\gamma^{-4}\right) \tag{3.126}$$

and similarly

$$otf^{0,0}(f) = \int_{-NA/\lambda+\max(0,f)}^{NA/\lambda+\min(0,f)} \widehat{m^{0,\gamma}}(u-f)\,\widehat{m^{0,\gamma}}(u)\,\mathrm{d}u \tag{3.127}$$

$$= \frac{M_0M_0}{\gamma}\left[1 + \mathcal{O}\left[\left(\frac{u}{\gamma}\right)^2\right]\right]_{-NA/\lambda+\max(0,f)}^{NA/\lambda+\min(0,f)} \tag{3.128}$$

$$= \frac{M_0M_0}{\gamma}\left(2\frac{NA}{\lambda} - |f|\right) + \mathcal{O}\left(\gamma^{-3}\right) \tag{3.129}$$

By choosing the demagnification factor γ sufficiently large one finds that $otf^{0,1}(f) = j2\pi f\frac{M_1}{M_0}\frac{1}{4\pi\gamma}otf^{0,0}(f)$ and subsequently because of the properties of the Fourier transform that $psf^{0,1}(x) \propto \frac{\mathrm{d}}{\mathrm{d}x}\left[psf^{0,0}(x)\right]$. Using a very high γ has however two big disadvantages. Firstly, the detector signals from zeroth and first order mode are severely reduced, which deteriorates the signal to noise ratio at the detector side. Secondly the ratio in between the zeroth order mode signal and the first order mode signal is decreased, which reduces the precision for calculating ϕ in Eq. (3.121).

The optical transfer function for illumination as well as for detection with the first order mode is a little more complex:

$$otf^{1,1}(\overline{f}) = \left\{ -\sqrt{2}\overline{a} \exp\left[-\frac{\pi \overline{a}^2}{2} \left(2 - |\overline{f}|\right)^2 \right] \left(2 - |\overline{f}|\right) + \left(1 - \pi \overline{a}^2 \overline{f}^2\right) \operatorname{erf}\left[\sqrt{\frac{\pi}{2}} \overline{a} \left(2 - |\overline{f}|\right) \right] \right\} \exp\left(-\frac{\pi \overline{a}^2 \overline{f}^2}{2} \right) \quad (3.130)$$

$$= \frac{2}{3} \sqrt{2} \pi \overline{a}^3 \left(\sqrt{3} + 1 + |\overline{f}|\right) \left(\sqrt{3} - 1 - |\overline{f}|\right) \left(2 - |\overline{f}|\right) + \mathcal{O}\left(\overline{a}^5\right) \text{ (for } \overline{a} \ll 1) \quad (3.131)$$

The last expression shows already that the otf will have a zero at $|\overline{f}| = \sqrt{3} - 1$, and for this reason not very suitable for detection.

3.4.3 Higher order modes

The optical transfer function for illumination with the zeroth order and detection with the second order can be written in closed form as:

$$otf^{0,2}(f) = \left\{ \overline{a} \exp\left[-\frac{\pi \overline{a}^2}{2} \left(2 - |f|\right)^2 \right] (2 - |f|) - \frac{\pi \overline{a}^2 \overline{f}^2}{\sqrt{2}} \operatorname{erf}\left[\sqrt{\frac{\pi}{2}} \overline{a} \left(2 - |f|\right) \right] \right\} \exp\left(-\frac{\pi \overline{a}^2}{2} \overline{f}^2 \right) \quad (3.132)$$

The otf and psf, shown on Fig. 3.23, do not look very favourable compared to these of the zeroth order mode detection (Fig. 3.19). For a small waveguide width the shape of both signals is identical but the second order mode has a lower amplitude. For broader waveguides the signal of the second order mode deteriorates much faster than that of the zeroth order mode.

If we would however, instead of the second order mode, use the hypothetical mode $m^{02,air}(x)$ defined in Eq. (3.92) one finds a totally

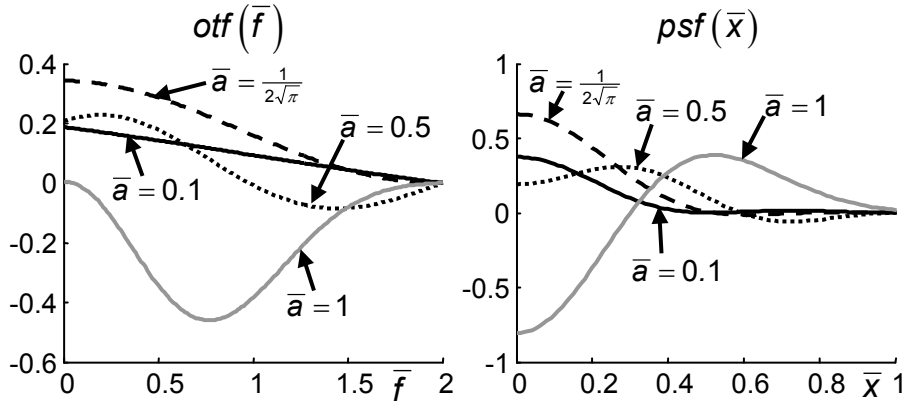


Figure 3.23: The $otf^{0,2}(\overline{f})$ and $psf^{0,2}(\overline{x})$ for a few values of the normalized waveguide width, $\overline{a}$. For reasons of clarity only one half is shown, the otf and the psf are both symmetric.

different otf.

$$otf^{0,02}(f) = \xi otf^{0,0}(f) - otf^{0,2}(f) \tag{3.133}$$

$$= \left(\xi - \frac{1}{\sqrt{2}}\right) otf^{0,0}(f) + 4\pi\overline{a}^3 \int_{-1+\max(0,\overline{f})}^{1+\min(0,\overline{f})} \mathcal{G}[\overline{a}\,(u-f)]\, u^2\, \mathcal{G}[\overline{a}u]\,\mathrm{d}u \tag{3.134}$$

$$= \left(\xi - \frac{1}{\sqrt{2}}\right) \mathcal{G}\left(\frac{\overline{a}\overline{f}}{\sqrt{2}}\right) \mathrm{erf}\left[\frac{\sqrt{\pi}}{2}\overline{a}\,(2-|f|)\right] + \exp\left(-\frac{\pi\overline{a}^2 f^2}{2}\right) \left\{ \frac{\sqrt{2}}{2}\left(1+\pi\overline{a}^2 f^2\right) \mathrm{erf}\left[\frac{\sqrt{\pi}}{2}\overline{a}\,(2-|f|)\right] - \overline{a}\,(2-|\overline{f}|) \exp\left(-\frac{\pi\overline{a}^2}{2}\,(2-|\overline{f}|)^2\right)\right\}$$

$$= \left(\xi - \frac{1}{\sqrt{2}}\right) \left[\sqrt{2}\overline{a}\,(2-|\overline{f}|) + \mathcal{O}(\overline{a}^3)\right] + \frac{4\pi\overline{a}^3}{3}\left[1+(1-|f|)^3\right] + \mathcal{O}(\overline{a}^5)\,. \tag{3.135}$$

If $\left|\xi - \sqrt{2}\right| \ll \overline{a}^2 \ll 1$, the first line in Eq. (3.135) becomes negligible and the OTF attains a shape as in Fig. 3.24. Compared with the OTF with zeroth order detection, the higher spatial frequencies are boosted.

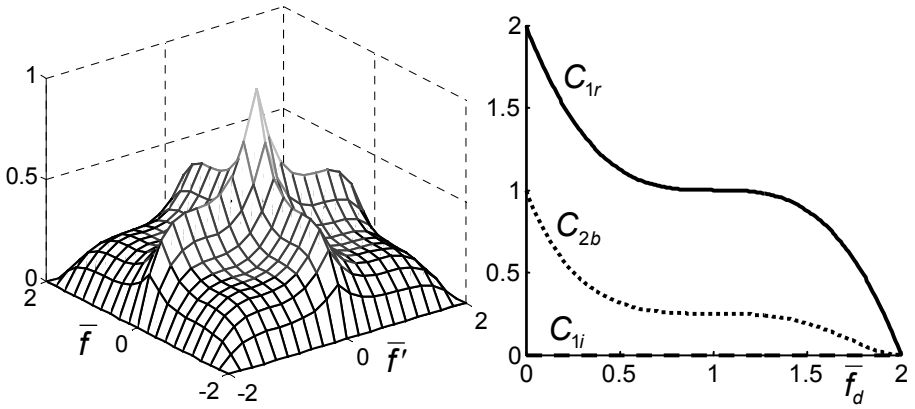


Figure 3.24: On the left the normalized partially coherent $OTF^{0,02}(\overline{f},\overline{f'})$ for $\left|\xi-\sqrt{2}\right| \ll \overline{a}^2 \ll 1$. On the right the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15).

Once again the waveguide width has similar influence on the otf. Broad waveguides give a squeezing towards lower spatial frequencies. An additional effect is that the otf may change sign. For narrow waveguide one finds the ideal shape as in Eq. (3.135), but the overall throughput drops drastically. Fig. 3.25 shows the otf and psf for $\xi=\sqrt{2}$ and a few values of $\overline{a}$.

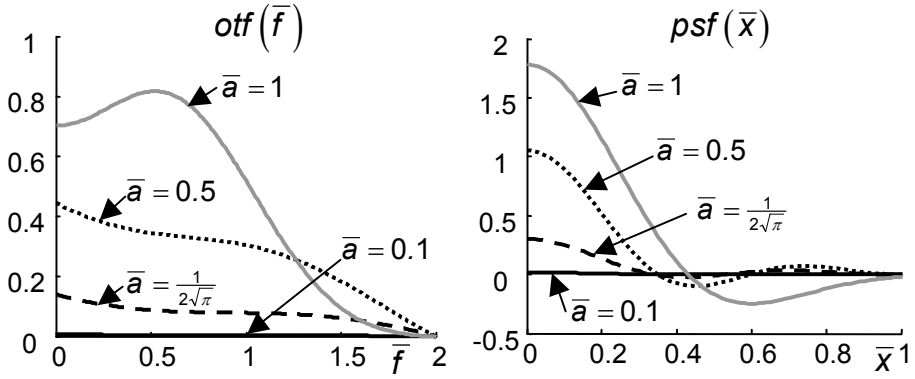


Figure 3.25: The $otf^{0,02}(\overline{f})$ and $psf^{0,02}(\overline{x})$ for a few values of the normalized waveguide width, $\overline{a}$. For reasons of clarity only one half is shown, the otf and the psf are both symmetric.

Similarly as with the first order mode one could think that the specific approximations of the second order modes in Eq. (3.88), are nec-

essary for the improved shape of OTF. Again we can show that in fact any two waveguide modes can be used as long as the total light throughput is not important and a few very general conditions are met. If

- $m^{a,air}(x)$ and $m^{b,air}(x)$ are both symmetric.
- $\int m^{a,air}(x)\,\mathrm{d}x \neq \int m^{b,air}(x)\,\mathrm{d}x$,

it is always possible to find a linear combination of both modes defined as, $m^{c,air}(x) = \xi m^{a,air}(x) - m^{b,air}(x)$, such that $\int m^{c,air}(x)\,\mathrm{d}x = 0$. As a consequence one has $\widehat{m^{c,air}}(0) = 0$. The Taylor expansion of $\widehat{m^{a,\gamma}}(f)$ and $\widehat{m^{c,\gamma}}(f)$ around $f = 0$ can then be written as:

$$\widehat{m^{a,\gamma}}(f) = \frac{1}{\sqrt{\gamma}}\widehat{m^{a,air}}\left(\frac{f}{\gamma}\right) = \frac{1}{\sqrt{\gamma}}\left\{M_a + \mathcal{O}\left[\left(\frac{f}{\gamma}\right)^2\right]\right\} \tag{3.136}$$

$$\widehat{m^{c,\gamma}}(f) = \frac{1}{\sqrt{\gamma}}\widehat{m^{c,air}}\left(\frac{f}{\gamma}\right) = \frac{1}{\sqrt{\gamma}}\left\{M_c\left(\frac{f}{\gamma}\right)^2 + \mathcal{O}\left[\left(\frac{f}{\gamma}\right)^4\right]\right\} \tag{3.137}$$

and according to Eq. (3.96) the

$$otf^{a,c}(f) = \int_{-NA/\lambda+\max(0,f)}^{NA/\lambda+\min(0,f)} \widehat{m^{a,\gamma}}(u-f)\,\widehat{m^{c,\gamma}}(u)\,\mathrm{d}u \tag{3.138}$$

$$= \frac{M_a M_c}{\gamma}\left[\left(\frac{u}{\gamma}\right)^2 + \mathcal{O}\left[\left(\frac{u}{\gamma}\right)^4\right]\right]_{-NA/\lambda+\max(0,f)}^{NA/\lambda+\min(0,f)} \tag{3.139}$$

$$otf^{a,c}(\overline{f}) = \frac{M_a M_c}{3\gamma^3}\left(\frac{NA}{\lambda}\right)^3\left[1 + \left(1 - \left|\overline{f}\right|^3\right)\right] + \mathcal{O}(\gamma^{-5}) \tag{3.140}$$

By setting the demagnification factor γ sufficiently large, one finds the same shape of the otf as in Eq. (3.133).

In principle more symmetric modes could be combined to approximate higher order derivatives. With $n+1$ symmetric modes, one can make an otf of the form

$$otf(f) \propto 1 + \left(1 - \left|\overline{f}\right|\right)^{2n+1} \tag{3.141}$$

For $n = 0$ and $n = \infty$ the psf can be written in closed form as:

$$psf^{n=0}(\overline{x}) \propto \mathrm{sinc}^2(2\pi\overline{x}) \qquad psf^{n=\infty}(\overline{x}) \propto \mathrm{sinc}(4\pi\overline{x}) \tag{3.142}$$

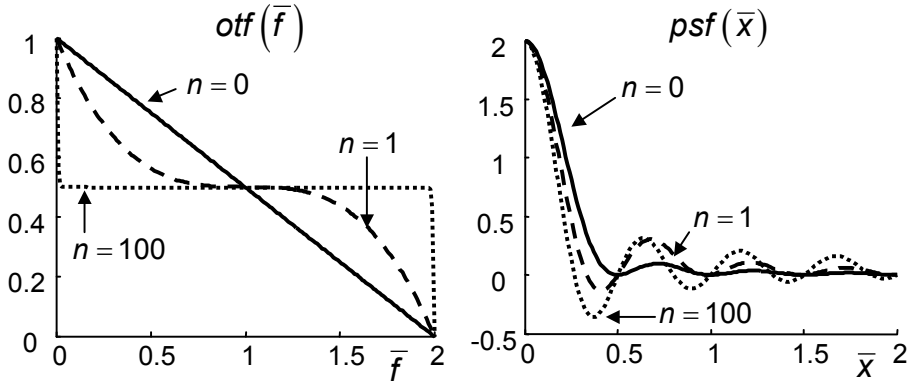


Figure 3.26: The otf and psf for detection with combinations of $n+1$ even higher order modes. For reasons of clarity the otf and psf have been normalized. They are are both symmetric and therefore only shown for positive $\overline{x}$ and $\overline{f}$.

As shown in Fig. 3.26 the more modes are used, the more closely a flat otf is approximated. For $n = 0$ one has only the zeroth order mode and finds the triangularly shaped otf. For $n = 100$ the otf has the form of a rectangle. The graph shows also that the width of the central lobe of the psf decreases with increasing n. This would be a very nice result, but there are some serious problems. The first problem is that one would need a very broad waveguide that guides so many modes and a very accurately designed photonic IC to control all those modes. Secondly the overall throughput decreases with increasing n. The third problem is that while the central lobe of the psf becomes narrower, the side lobes become bigger. Whether such a flat profile is really favourable, is a point of discussion. We will return to this question in section 3.7.

In literature one has already shown that it is possible to obtain such a flat otf by using phase masks in the optical light path [52]. An alternative way to produce a flat otf would be to illuminate the disc with two counter-propagating plane waves, in such a way that a cosine pattern is formed on the disc [53].

$$psf_{illum}(x) = \cos\left(\frac{2\pi NA}{\lambda}x\right) \tag{3.143}$$

$$otf_{illum}(f) = \frac{1}{2}\left[\delta\left(f - \frac{NA}{\lambda}\right) + \delta\left(f + \frac{NA}{\lambda}\right)\right] \tag{3.144}$$

Combined with a small detection aperture this results in:

$$otf_{tot} = otf_{illum}(f) \otimes \Pi\left(\frac{f\lambda}{2NA}\right) = \Pi\left(\frac{f\lambda}{4NA}\right), \tag{3.145}$$

which is the same rectangular otf as for $n = \infty$ with the combination of higher order waveguide modes.

3.4.4 Near-field

Remark that in the 'near-field' approach it is not possible to demagnify the waveguide mode profile as with a lens system. This poses a limit on the effective width of the waveguide mode. In principle one can make a waveguide mode as small as $\lambda/\left(2\sqrt{{n_{core}}^2 - {n_{clad}}^2}\right)$. For a very high contrast waveguide, $n_{core} = 3.5$ and $n_{clad} = 1$, this gives 0.15λ, which is really the smallest possible. More realistic values are at least twice as broad.

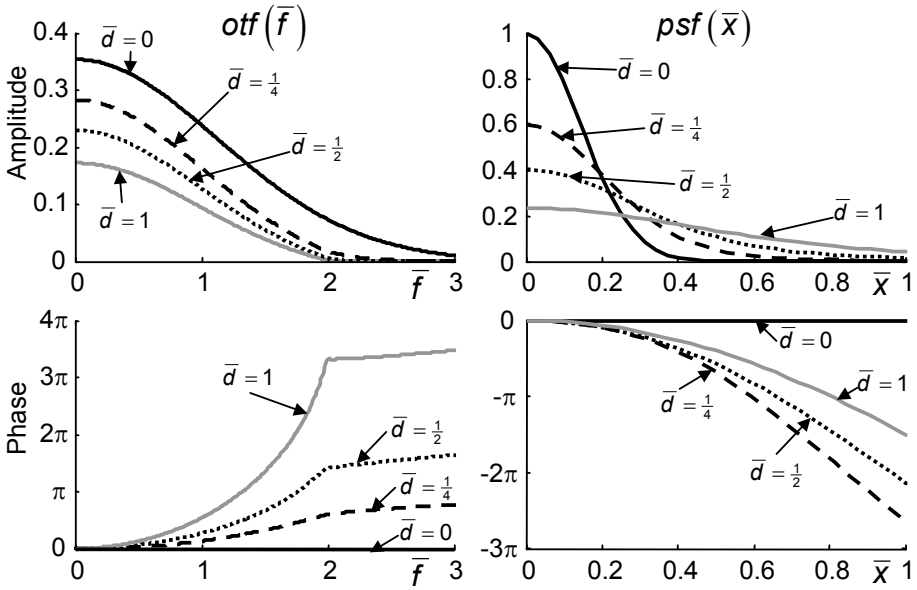


Figure 3.27: Amplitude and phase of the $otf^{0,0}(\overline{f})$ and $psf^{0,0}(\overline{x})$ for near-field detection, as a function of the normalized gap thickness, $\overline{d} = d/\lambda$. The normalized waveguide width is $\overline{a} = \frac{1}{2\sqrt{\pi}}$. For reasons of clarity only one half is shown, the otf and the psf are both symmetric.

Fig. 3.27 shows the otf and psf for a few values of the normalized distance, or gap size, $\overline{d} = d/\lambda$. Remark that because of the finite wave-

guide width even in absence of a gap the otf drops to zero for $\overline{f} = 3$. The cut-off frequency and the amplitude of the otf decrease for increasing gap size. For $\overline{d} = 1$ the advantage of near-field has already vanished. Moreover one sees that in contrast to the far-field otf, also the phase is affected. The larger the gap, the faster the phase of the otf is affected.

We can conclude that sliding the waveguide at a narrow air gap over the disc, can improve the cut-off frequency by approximately 50% compared to a lens with $NA = 1$. The advantage over the far-field is however already lost for gap width larger than one wavelength. This makes it very impractical to implement for removable discs. Moreover the advantage is smaller than with other near-field techniques, because the waveguide mode diameter cannot be made extremely small.

3.4.5 Tilted waveguide

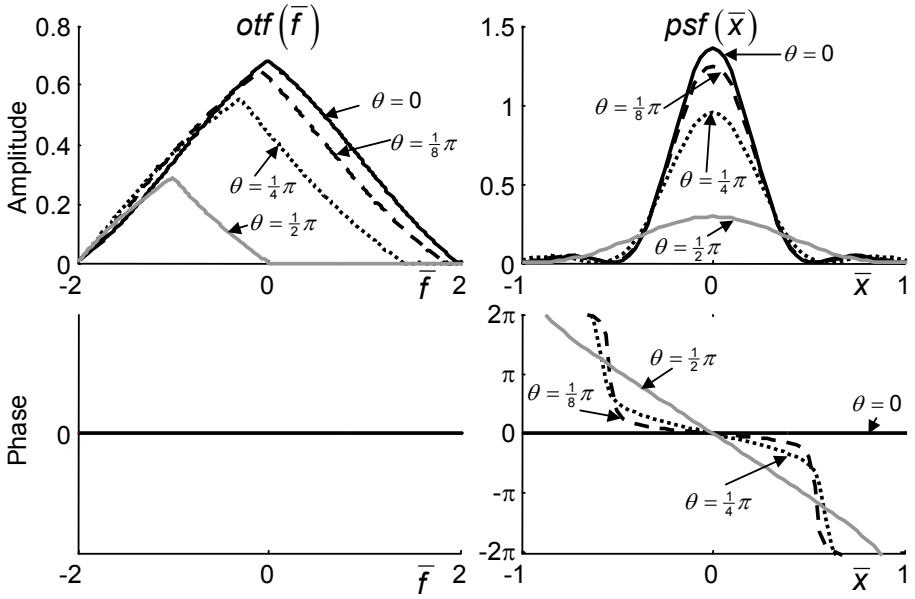


Figure 3.28: Amplitude and phase of the $otf^{0,0}(f)$ and $psf^{0,0}(x)$ for a tilted waveguide, $\overline{a} = \frac{1}{2\sqrt{\pi}}$ and a few tilt angles.

Fig. 3.28 shows the amplitude and phase of the otf and the psf of the tilted waveguide for a few tilt angles. We assumed the same tilt for the illumination and the detection waveguide. Tilting only one of

these gives however similar results. Some of the diffracted orders at the disc surface are not anymore captured at the detector. This results in a broadening of the point spread function.

3.4.6 Double waveguide detector

As described in section 3.3.3, we could use a multiple single mode waveguides closely spaced at a distance d, as an alternative to a single multimodal waveguide. If we use for both waveguides the zeroth order mode, with a mutual phase difference θ_1 for the illumination, and a mutual phase difference θ_2 for the detection, the otf would be:

$$otf^{\theta_1,\theta_2}(f) = 2\exp\left(-\pi\frac{\overline{f}^2\overline{a}^2}{2}\right) \times \left\{\cos\left(\frac{\theta_1-\theta_2}{2}\right)\exp\left(-\pi\frac{\overline{d}^2}{2\overline{a}^2}\right)\Re\left[\operatorname{erf}\left(\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)-j\sqrt{\frac{\pi}{2}}\frac{\overline{d}}{\overline{a}}\right)\right] + \operatorname{erf}\left(\sqrt{\frac{\pi}{2}}\overline{a}\left(2-|\overline{f}|\right)\right)\cos\left(\frac{\theta_1+\theta_2+\overline{df}}{2}\right)\right\} \quad (3.146)$$

For $\overline{a} \ll 1$ this can be simplified to:

$$otf(f) = 4\sqrt{2}\overline{a}\left(2-\left|\overline{f}\right|\right)\cos\left(\frac{\overline{df}}{2}+\theta_1\right)\cos\left(\frac{\overline{df}}{2}-\theta_2\right) \quad (3.147)$$

The combination of separate modes can be seen as a combination of the zeroth and first order mode of the structure comprising both waveguides. It is hence not surprising that for small waveguide separation $\overline{d}$, the otf for $\theta_1 = 0$ and $\theta_2 = 0$ leads to the expression of the zeroth order mode detection found in Eq. (3.99). Similarly for $\theta_1 = 0$ and $\theta_2 = \pi$, one finds the first order mode detection back as in Eq. (3.109). Fig. 3.29 shows how the amplitude and phase of the otf and psf change as a function of the phase difference in between the detection waveguides. The similarity with the tilted waveguide is not incidental, in both cases a phase lag is introduced between the left and right side of the waveguide.

3.5 Extended detector method

For the extended detector system, two operation modes will be described. In the normal operation mode, also call central aperture detection (CAP), the detector captures all the light passing through the

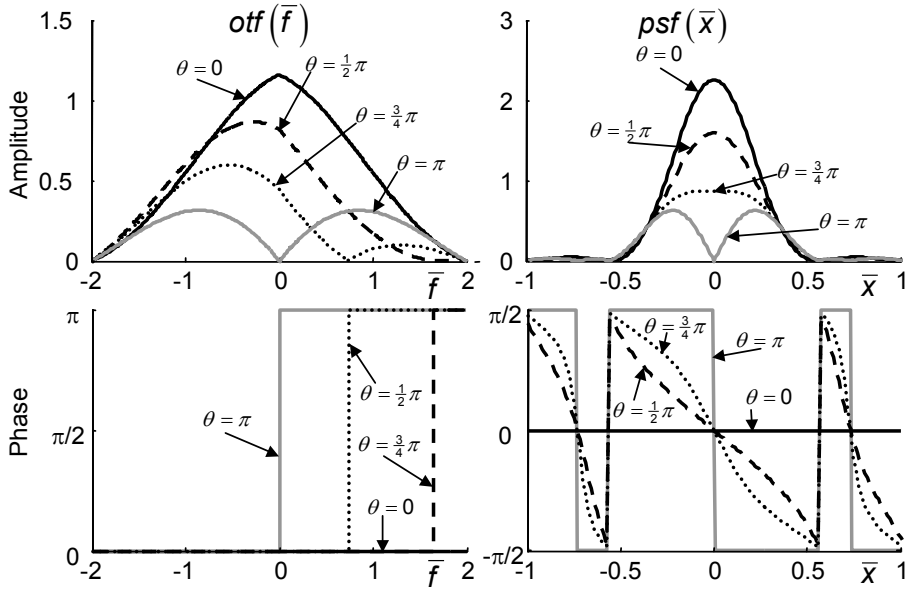


Figure 3.29: Amplitude and phase of the $otf(\overline{f})$ and $psf(\overline{x})$ for double waveguide system, with $\overline{a} = 0.15$, $\overline{d} = 0.3$, $\theta_1 = 0$, $\theta_2 = \theta$.

detection lens. In the push-pull (PP) operation mode, the detector area is split into two halves and the electric signal from these halves is subtracted.

Starting from Eq. (3.52), the OTF for the extended detector method in the far-field is:

$$OTF(f;f') = \int_{-\infty}^{\infty} p_{ext}\left(\frac{u\lambda}{2NA}\right) p_{lens}\left(\frac{u-f}{2NA}\lambda\right) p_{lens}\left(\frac{u-f'}{2NA}\lambda\right) \times \widehat{m^{\gamma}_{illum}}(u-f)\left[\widehat{m^{\gamma}_{illum}}(u-f')\right]^{*} \mathrm{d}u. \quad (3.148)$$

3.5.1 Central aperture detection

In the central aperture detection, the power of all the light reflected from the disc that passes the objective lens is integrated over an extended detector. For a rectangular lens and detector the OTF can then be written as:

$$OTF(f;f') = \int_{-\infty}^{\infty} \Pi\left(\frac{u\lambda}{2NA}\right) \Pi\left(\frac{u-f}{2NA}\lambda\right) \Pi\left(\frac{u-f'}{2NA}\lambda\right) \times \widehat{m_{illum}^{\gamma}}(u-f) \left[\widehat{m_{illum}^{\gamma}}(u-f')\right]^{*} \mathrm{d}u. \quad (3.149)$$

For $|f| > 2NA/\lambda$, $|\overline{f'}| > 2NA/\lambda$ or $|\overline{f} - \overline{f'}| > 2NA/\lambda$ this integral is zero. Otherwise, using the Gaussian approximation of a zeroth order mode for the illumination mode and the normalization as in Eq. (3.98) one finds for the OTF :

$$\begin{aligned} OTF(\overline{f};\overline{f'}) &= \overline{a}\sqrt{2} \int_{-1+\max(0,\overline{f},\overline{f'})}^{1+\min(0,\overline{f},\overline{f'})} \mathcal{G}(\overline{a}(u-\overline{f}))\, \mathcal{G}(\overline{a}(u-\overline{f'}))\, \mathrm{d}u \\ &= \left\{ \mathrm{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2 - |\overline{f}| - |\overline{f'}|\right)\right] + \mathrm{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2 - |\overline{f} - \overline{f'}|\right)\right] \right\} \\ &\quad \times \frac{1}{2}\mathcal{G}\left[\frac{\overline{a}(\overline{f} - \overline{f'})}{\sqrt{2}}\right]. \quad (3.150) \end{aligned}$$

For the main cross sections this results in:

$$\mathcal{C}_{1r} = 2\,\mathrm{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2 - |\overline{f}|\right)\right] \mathcal{G}\left[\frac{\overline{a}\overline{f}}{\sqrt{2}}\right] \quad (3.151)$$

$$\mathcal{C}_{1i} = 0 \quad (3.152)$$

$$\mathcal{C}_{2b} = \mathrm{erf}\left[\sqrt{\frac{\pi}{2}}\overline{a}\left(2 - 2|\overline{f}|\right)\right] \mathcal{G}\left[\sqrt{2}\overline{a}\overline{f}\right] \quad (3.153)$$

For small $\overline{a}$ one finds:

$$OTF(\overline{f};\overline{f'}) = \frac{\overline{a}}{\sqrt{2}}\left(4 - |\overline{f} - \overline{f'}| - |\overline{f}| - |\overline{f'}|\right) \quad (3.154)$$

$$PSF(\overline{x};\overline{x'}) = 8\sqrt{2}\overline{a}\,\mathrm{sinc}(2\pi\overline{x})\,\mathrm{sinc}(2\pi\overline{x'})\,\mathrm{sinc}\left[2\pi\left(\overline{x} - \overline{x'}\right)\right] \quad (3.155)$$

3.5.2 Push-pull detection

In the extended detector system the detector is sometimes cut into two halves as a part of the push-pull detection for the tracking. This is the case for some CD systems and the DVD-RAM system. In these tracking systems the split between the two detector halves lies however parallelly to the tracks, where in our case we will put it orthogonally to

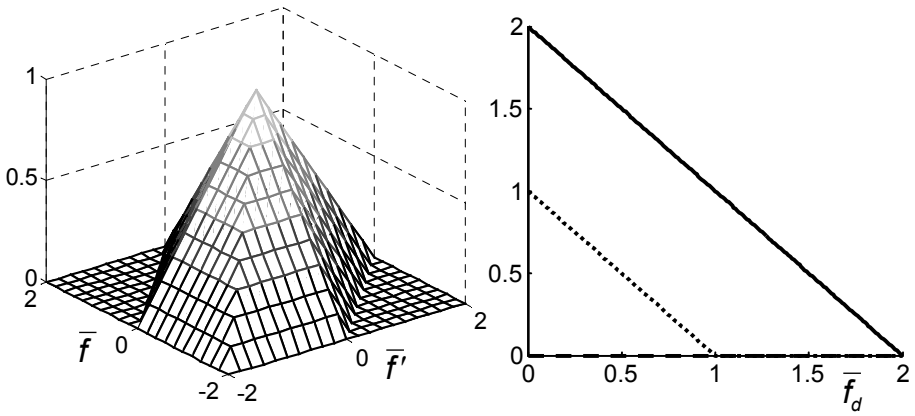


Figure 3.30: On the left the normalized partially coherent OTF for the central aperture signal in an extended detector system for $\overline{a} \ll 1$. On the right the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15).

the track. We start again from Eq. (3.52), and calculate the total OTF by subtracting the OTF for the two detector halves: $OTF^{PP}(f;f') = OTF^{PP+}(f;f') - OTF^{PP-}(f;f')$, with

$$OTF^{PP\pm}(f;f') = \int_{-\infty}^{\infty} \Pi\left(\frac{u\lambda}{NA} \mp \frac{1}{2}\right) \Pi\left(\frac{u-f}{2NA}\lambda\right) \Pi\left(\frac{u-f'}{2NA}\lambda\right) \times \widehat{m^{\gamma}_{illum}}(u-f) \left[\widehat{m^{\gamma}_{illum}}(u-f')\right]^{*} \mathrm{d}u. \quad (3.156)$$

The OTF^{PP+} is non-zero for $\max(0,-\overline{f},-\overline{f'}) + \max(1,\overline{f},\overline{f'}) < 2$ and can within these borders be written in a closed form as:

$$OTF^{PP+}(\overline{f};\overline{f'}) = \frac{1}{2}\exp\left(-\pi\frac{\overline{a}^2\left(\overline{f}-\overline{f'}\right)^2}{2}\right) \left\{\mathrm{erf}\left[\frac{\pi}{2}\overline{a}\left(2-\max\left(\left|\overline{f}-\overline{f'}\right|, 2+\overline{f}+\overline{f'}\right)\right)\right] + \mathrm{erf}\left[\frac{\pi}{2}\overline{a}\left(2-\max\left(\left|\overline{f}-\overline{f'}\right|, -\overline{f}-\overline{f'}\right)\right)\right]\right\} \quad (3.157)$$

The expression for $OTF^{PP-}(\overline{f};\overline{f'})$ is very similar. If one takes the limit for small $\overline{a}$ one easily finds:

$$OTF^{PP}(\overline{f};\overline{f'}) = OTF^{PP+}(\overline{f};\overline{f'}) - OTF^{PP-}(\overline{f};\overline{f'}) \tag{3.158}$$

$$= \sqrt{2}\overline{a}\left\{\max\left[2-\max\left(0,-f,-f'\right)-\max\left(1,f,f'\right),0\right]\right.$$
$$\left.-\max\left[2-\max\left(1,-f,-f'\right)-\max\left(0,f,f'\right),0\right]\right\} \tag{3.159}$$

For the main cross sections one has $\mathcal{C}_{1r} = \mathcal{C}_{2b} = 0$ and

$$\mathcal{C}_{1i} = \begin{cases} 2\sqrt{2}\overline{a}\left|\overline{f}\right|, & \text{for } \overline{f} < 1 \\ 2\sqrt{2}\overline{a}\left(2-\left|\overline{f}\right|\right), & \text{for } 1 < \overline{f} < 2 \end{cases} \tag{3.160}$$

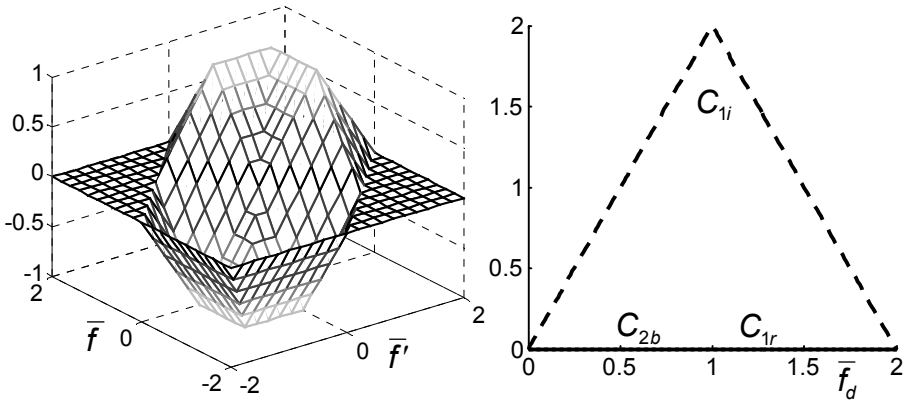


Figure 3.31: On the left the normalized partially coherent OTF for the push-pull signal in an extended detector system for $\overline{a} \ll 1$. On the right the principal cross sections $\mathcal{C}_{1r}$, $\mathcal{C}_{1i}$, and $\mathcal{C}_{2b}$ as defined in Eq. (3.15).

Comparing the OTF of the push-pull signal, shown on Fig. 3.31, with the OTF of the first order mode detection in the waveguide system (Fig. 3.21), there are parallels as well as differences. For both methods the signal is zero for a uniform reflection coefficient and both have their highest sensitivity for gratings with a spatial frequency $\left|\overline{f}\right| = 1$. The big difference is that for the first order mode detection with the waveguide method, the signal detected from a grating is proportional to $\mathcal{C}_{2b}$, which itself is proportional to the squared modulation depth of the grating $|\varepsilon|^2$. The push-pull signal in an extended detector setup on the other hand is proportional to $\mathcal{C}_{1i}$, and hence proportional to $\Im(\varepsilon)$. This means that in case of a purely real grating, there is no push-pull

signal at all. As shown on Fig. 3.13 this is the case for a pit depth of $\lambda/4n$. Therefore the pit depth in a CD system is slightly larger or smaller than $\lambda/4n$ [54]. This is also one of the reasons why for the DVD and BD system differential phase detection is used for the tracking instead of push-pull detection.

3.6 Confocal detection method

For a confocal microscope with rectangular lenses, the OTF in Eq. (3.61), can be rewritten as:

$$OTF(f;f') = \iint \widehat{m^{\gamma}_{illum}}(u-f)\,\widehat{m^{\gamma}_{illum}}^{*}(u'-f')\,\widehat{p^{\gamma}_{conf}}(u'-u) \times \Pi\left(\frac{u\lambda}{2NA}\right)\Pi\left(\frac{u'\lambda}{2NA}\right)\Pi\left(\frac{u-f}{2NA}\lambda\right)\Pi\left(\frac{u'-f'}{2NA}\lambda\right)\mathrm{d}u\,\mathrm{d}u'. \quad (3.161)$$

This means that $OTF(f;f')$ is zero for $|f| > \frac{2NA}{\lambda}$ or $|f'| > \frac{2NA}{\lambda}$. Within the OTF-band it can be written as:

$$OTF(f;f') = \int_{-\frac{NA}{\lambda}+\max(0,f)}^{\frac{NA}{\lambda}+\min(0,f)} \int_{-\frac{NA}{\lambda}+\max(0,f')}^{\frac{NA}{\lambda}+\min(0,f')} \times \widehat{m^{\gamma}_{illum,x}}(u-f)\,\widehat{m^{\gamma}_{illum,x}}^{*}(u'-f')\,\widehat{p^{\gamma}_{conf,x}}(u-u')\,\mathrm{d}u\,\mathrm{d}u' \quad (3.162)$$

$$= \frac{a_{conf}a_1\sqrt{2}}{\gamma}\int_{-\frac{NA}{\lambda}+\max(0,f)}^{\frac{NA}{\lambda}+\min(0,f)} \int_{-\frac{NA}{\lambda}+\max(0,f')}^{\frac{NA}{\lambda}+\min(0,f')} \mathcal{G}\left[\frac{a_1}{\gamma}(u-f)\right] \times \mathcal{G}\left[\frac{a_1}{\gamma}(u'-f')\right]\mathrm{sinc}\left[\frac{a_{conf}}{\gamma}(u-u')\right]\mathrm{d}u\,\mathrm{d}u' \quad (3.163)$$

Despite the analytic approximations Eq. (3.163) cannot written in closed form. Nevertheless one can take the limit for very small or very large values of a_{conf}. For $a_{conf} \ll 1$, $\mathrm{sinc}(a_{conf}f) \simeq 1$ and one finds:

$$OTF(\overline{f};\overline{f'}) = \frac{\overline{a_{conf}}}{2\overline{a_1}}\left\{\mathrm{erf}\left[\sqrt{\pi}\overline{a_1}\right] + \mathrm{erf}\left[\sqrt{\pi}\overline{a_1}\left(1-|\overline{f}|\right)\right]\right\} \times \left\{\mathrm{erf}\left[\sqrt{\pi}\overline{a_1}\right] + \mathrm{erf}\left[\sqrt{\pi}\overline{a_1}\left(1-|\overline{f'}|\right)\right]\right\} \quad (3.164)$$

which is identical to the waveguide detection systems for a small detection aperture (Eq. (3.107)). This is as would be expected, as for a

very small aperture (or waveguide mode), the distinction in between 'coherent' and 'incoherent' detection disappears:

$$\text{'Coherent':}\ \lim_{D\to 0}\left|\int_{-D/2}^{D/2}\phi(x_2)\,\mathrm{d}x_2\right|^2 = |\phi(0)|^2 \tag{3.165}$$

$$\text{'Incoherent':}\ \lim_{D\to 0}\int_{-D/2}^{D/2}|\phi(x_2)|^2\,\mathrm{d}x_2 = |\phi(0)|^2 \tag{3.166}$$

For the other extreme, $a_{conf} \gg 1$, $\mathrm{sinc}(a_{conf} f) \simeq \delta f$ and Eq. (3.163) can be transformed to the equation for the extended detector system (Eq. (3.149)), which is also as expected: The power of the optical field in the entire image plane is identical to the power in the entire Fourier plane. For a large detector, the location along the optical axis has no influence on the detector signal.

3.7 Conclusions

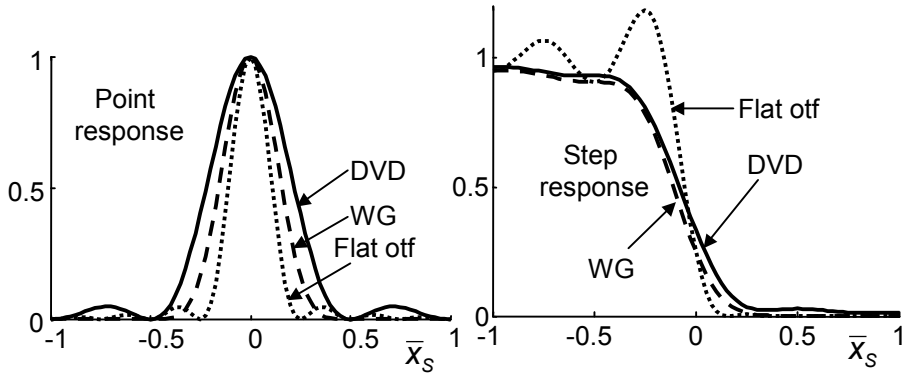


Figure 3.32: Comparison of the signal for a point and edge response. The solid line shows the DVD detection system, the dashed line is the zeroth order detection of the waveguide system, and the dotted line is for a uniform OTF.

In the previous section we have outlined how the optical transfer function for the waveguide detection method, extended detector system and confocal microscope are calculated and how they are influenced by the dimensions of aperture, and by the shape of the waveguide modes. It is harder to define a unambiguous figure of merit to show what system is best. Two approaches are possible. The first

considers how well the electric signal represents the original bit pattern. This measure is however a point of discussion. Consider Fig. 3.32, which shows the point and step response of the waveguide detection with zeroth order mode, the DVD detection system and a system with a flat *otf* as in Eq. 3.141. The point response ($PSF(\overline{x_s};\overline{x_s})$) for DVD detector, waveguide detection system and uniform OTF is respectively: $\mathrm{sinc}^2\, 2\pi\overline{x}$, $\mathrm{sinc}^4\, 2\pi\overline{x}$, and $\mathrm{sinc}^2\, 4\pi\overline{x}$. For the step response,

$$\iint PSF\left(\overline{x}-\overline{x_s};\overline{x'}-\overline{x_s}\right)\mathcal{H}(\overline{x})\,\mathcal{H}\left(\overline{x'}\right)\mathrm{d}\overline{x}\,\mathrm{d}\overline{x'},$$

there are however no analytic expressions. An inspection of the point responses on Fig. 3.32, shows that the flat *otf* clearly provides a much narrower point response, than the waveguide detection with zeroth order mode, which itself is narrower than the that of the DVD system. For the step response function on the other hand, one finds that the signal from the DVD system and from the waveguide detector are very similar. The flat *otf* provides a steeper descent, but at the expense of high side lobes. Moreover there is a large asymmetry between pits and lands. On basis of these arguments Stallinga [55] showed that for a conventional method of extracting the bits from the signal, where a single threshold level marks a 'zero' or a 'one', the *otf* of the DVD system results in the best reliability.

The second approach is using model-based techniques as in section 2.4.2. In principle a channel, being a certain range of spatial frequencies, can hold as much information as defined by the Shannon limit, $\log_2(1+\mathrm{SNR})$. This means that the information capacity depends heavily on the amount and type of noise. In the next chapter we will show how a combined use of the zeroth and first order mode of the waveguide detection system may result in a capacity gain with respect to the DVD and confocal microscope.

Chapter 4

Extracting bit pattern

In this chapter we will show how the signals from different modes can be combined. Simulation results for the zeroth and first order mode show that with a combination of these signals, a lower bit error rate can be achieved than with each of the individual signals. In the first section we will describe how the results of the waveguide detection method are related to those of the conventional disc read-out system. The second section covers the methodology for extracting the bit pattern from the retrieved signals. The third section explains the implementation of the simulations. Finally the results are shown in the fourth section.

4.1 Introduction

4.1.1 Resolution and bit error rate

In a conventional CD or DVD system, the 'zero' and 'one' values from the bit sequence are extracted from the signal at the detector by watching at which points the signal crosses a certain threshold level. These points mark a change from a 'one' to a 'zero' and vice versa [9, 56]. This method is fast and reliable, but a large part of the information contained in the detected signal is dropped. Enhanced methods for reading out bits make it possible to increase the data density as, for instance, in multilevel modulation methods [1, 2]. As described in chapter 2 optical resolution is ultimately not limited by the Raleigh distance but by systematic and random errors. This means that the eventual bit length at which there is still a reliable read-out depends heavily on the noise in the system. Without noise one can resolve features smaller than the cut off frequency of the MTF, $\lambda/(2NA)$. This can be done by methods

based on inverse filtering and analytic continuation of the image spectrum which reconstruct the information outside the MTF band. Unfortunately these methods are very sensitive to small errors in the measured signal, which means it is eventually noise that limits the effective resolution [30]. This means that resolution becomes a somewhat ambiguous measure for an optical storage system. In this section we will therefore use bit error rate (BER) as a figure of merit. The eventual output of a pick-up system is a signal containing the 'zeros' and 'ones'. In an ideal system this series would be identical to the bit pattern on the disc. In a real system some of the detected bits are incorrect. The bit error rate gives the average fraction of these incorrect bits. The bit error rate is not only a function of the optical system, the minimum bit length and the noise but depends also on how the bits are extracted from the detection signals. There exist different types of methods, which are shortly explained in section 2.4.3. We will not dive into the details of these methods but start from a simple parameter fitting algorithm and prove that the waveguide scanning system can read out bits at a lower bit error rate than a conventional system would do in an equivalent situation.

4.1.2 Waveguide detection system

The results in chapter 3 showed that the optical transfer function of the confocal microscope is identical to that of the waveguide detection method for illumination and detection with the zeroth order mode. Moreover measurements have shown that reading out bits from a DVD with a confocal system results in the same jitter values as a conventional DVD system [55].

The idea is now that additionally to the zeroth order mode, more modes can be used to detect the reflected field form the disc. These modes are detected in parallel and guided by the photonic IC towards different detectors, each resulting in an electric signal as a function of time. We will write the signal from the zeroth order mode as $P_0(t)$ and the power in the first order mode as $P_1(t)$. Fig. 4.1 shows these two signals for a sample bit pattern. In general one could use other modes or even combinations of modes as in Eq. 3.92. For reasons of clarity these signals will be written as P_j, with $j = 0 \ldots N-1$.

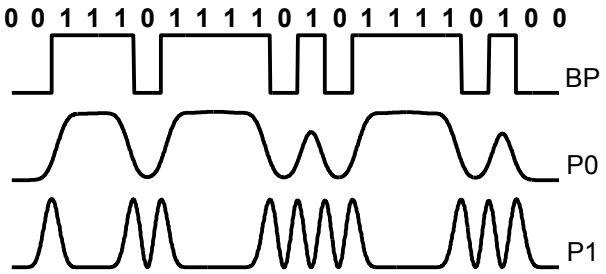


Figure 4.1: The disc is illuminated by the zeroth order mode. The reflected light is picked up by the zeroth and first order mode. The power in those two modes results in two signals dependent on the bit pattern on the disc.

4.1.3 Noise

In a real system the detected signal is distorted by various sources. These can be divided into three categories. The first is often called inter-symbol interference (ISI), and is a measure of the influence on the signal of a given bit from the neighbouring bits on the same track. The second is inter-track crosstalk, which gives the influence of the bits on the neighbouring tracks on the detected signal. These two categories are often not considered as real noise in stricto sensu, as they are inherent to the bit pattern detection itself.

The third category includes the noise from external sources. A non-exhaustive list includes

- Tracking noise: influence of the aberrations in the actuator following the track.
- Focussing errors: influence of the focussing.
- Lenses: influence of the aberrations on the optical properties of the lenses.
- Optical crosstalk: crosstalk in between the zeroth and first order mode.
- Reflections: parasitical reflections inside and at the facets of the photonic IC.
- Detector noise: the noise added by the detector.
- Media noise: irregularities in the shape of the pits on the disc [57].

The inter-symbol interference is automatically included in the simulations when extracting the bit pattern from a long sequence. The signal of a 'zero' in between two 'ones' gives a different P_0 and P_1 than for a 'zero' in between two 'zeros'. The inter-track crosstalk is introduced as in Eq. (3.76). The third category covering the real noise is modelled by three types of distortions on the bit pattern, shown in Fig. 4.2. The additive noise is Gaussian white noise added to the bit pattern, the level noise is noise on the level of reflection at a 'zero' and a 'one', and offset noise is the noise on the place of the 'zero-one' transitions.

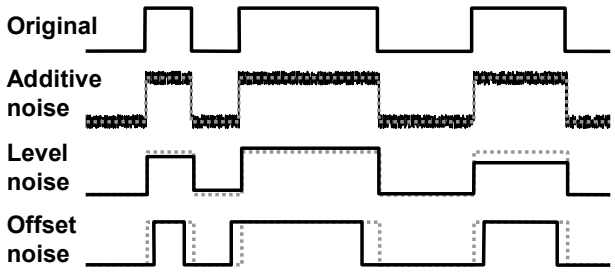


Figure 4.2: Additive noise, level noise and offset noise give model the noise from tracking and focussing errors and media noise.

4.2 Parameter extraction method

4.2.1 Principles

Fig. 4.3 shows the general idea of the parameter extraction method. The signals $P_j(t)$ captured by the detectors are distorted by ISI, crosstalk and noise. To find the original bit pattern that leads to these signals, a range of n candidate bit patterns is proposed: bp_c^k, with $k = 0 \ldots n-1$. For each of these candidates one can calculate the theoretical zeroth and first order signal, which we will call $C_0^k(t)$ and $C_1^k(t)$. In our model we assumed that all possible bit patterns within a given range, can be a possible candidate, which means that for m possible bits there are $n = 2^m$ candidate bit patterns. On Fig. 4.3 this range is marked with a gray dashed rectangle.

To find the correct bit pattern $k_{correct}$ among the n candidate bit patterns, a least squares fitting algorithm is used: for each signal (subindex j), and for each candidate bit pattern (superindex k), the squared differences in between the detected signal $P_j(t)$ and the calculated signals

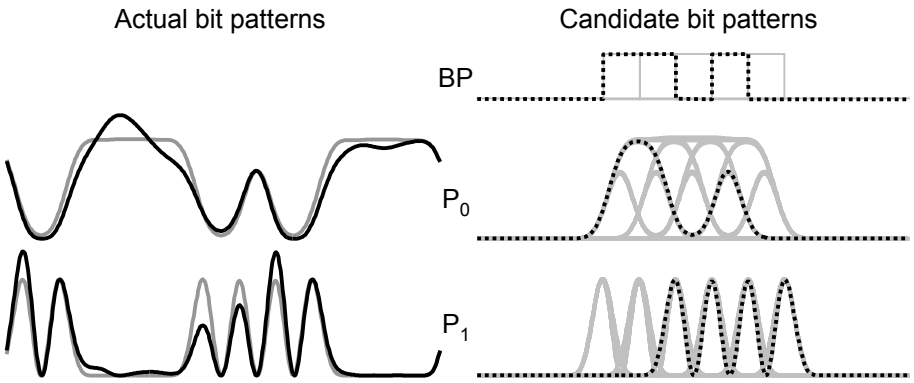


Figure 4.3: Schematic overview of parameter extraction method. In the left column: the detected signals P_j, the theoretical response in gray and the real signals distorted by ISI, crosstalk and noise in black. for reasons of clarity only $P_0(t)$ and $P_1(t)$ have been drawn. In the right column: From k candidate bit patterns the theoretical signals $C_j^k(t)$ are calculated. The idea is to find the bit pattern k, for which the $C_j^k(t)$ resemble best to the measured signals $P_j(t)$.

$C_j^k(t)$ are integrated over a certain range along t. The extent of this range will be discussed further on, and is not necessarily identical to the extent of the candidate bit patterns. These integrals of the square differences will be called overlap values and written S_j^k:

$$S_j^k = \int_{t_l}^{t_u} \left(P_j(t) - C_j^k(t) \right)^2 \mathrm{d}t \tag{4.1}$$

For each signal we define k_j as the superindex k for which the overlap value S_j^k is the lowest. This means

$$S_j^{k_j} = \min_{k=1\ldots n} S_j^k. \tag{4.2}$$

As long as the distortion on the detected signals P_j is low, any of the responses can be used to find a correct bit pattern. This means

$$\forall j : k_j = k_{correct}. \tag{4.3}$$

For larger distortion this may, however, no longer be true. As long as the k_j for the different modes are equal, that bit pattern remains the most logical guess. If however for some i and j, $k_j \neq k_i$, this choice is no longer straightforward. Therefore we have to construct a decision-making algorithm that maximizes the probability of selecting the correct bit pattern. One way of doing so is by constructing a weighted

(linear) combination of the individual overlap values. With β_j as the weighting factor this new overlap value can be written as

$$S_T^k = \sum_{j=0}^{N-1} \beta_j S_j^k. \tag{4.4}$$

The superindex k corresponding with the lowest value of S_T^k is called k_T and is used to select the most probable bit pattern.

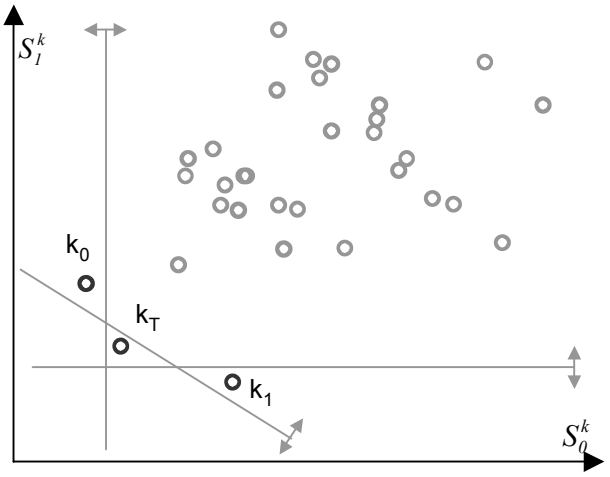


Figure 4.4: For each bit pattern $k = 1 \ldots n$, a dot is plotted with coordinates S_0^k along the horizontal axis and S_1^k along the vertical axis. The black dots represent the k_0, k_1, and k_T. These are the bit patterns for which respectively the S_0^k, S_1^k, and S_T^k are minimal.

The principle is clarified in Fig. 4.4 for the case of two signals: P_0 and P_1. For each bit pattern a dot is plotted with coordinates $\left(S_0^k, S_1^k\right)$. The dot that is closest to the horizontal axis, has the smallest value of S_0^k and determines k_0. The dot closest to the vertical axis, has the smallest value of S_1^k and leads to k_1. If the distortion on the signals P_0 and P_1 from noise and crosstalk is low, these two points will coincide, and give the most probable candidate for $k = k_0 = k_1$. For increasing distortion it is however possible that $k_0 \neq k_1$, and such a selection is not possible anymore. One could select either k_0 or k_1. We choose however for the dot which is closest to the bottom left corner along a diagonal line, defined by the linear combination from Eq. (4.4). In the numerical simulations we will show how the angle of this line, which depends on the relative weight of the β_j, is optimized.

4.2.2 Choice of the candidate bit patterns

Because the point spread function has a certain width, the detected signal of a specific bit is influenced by the neighbouring bits. To reduce this influence the number of candidate bit patterns can be increased. Because of calculation speed it is not possible to have an infinite number of candidate bit patterns. We have set the number of candidate bit patterns to $n = 2^m$. This means if the bit length is d_{bit}, the candidate bit patterns covers only a small section, $m \cdot d_{bit}$ wide. The remainder is filled up with 'zeros'. Fig. 4.3 shows the candidate bit patterns for $m = 5$.

Because it is more likely that the guess on one of the outer bits is incorrect than that such a bit error would occur for the central bits the method has been refined. The candidate bits are divided into two categories: inner candidate bits (m_i) and outer candidate bits (m_o). All candidate bits are used to form the candidate bit patterns bp_c^k, but only the values of the inner bits are stored, the values of the outer bits are thrown away. This means that a candidate bit pattern is correct as long as the inner candidate bits match with the bits on the track regardless of a possible difference in between the outer candidate bits and the bits on the track. In our simulations we made the extent of the integration domain [t_l,t_u] of the overlap values equal to that of the inner bits of the candidate bit patterns. This means $t_u - t_l = m_i \cdot d_{bit}$.

The idea is clarified in Fig. 4.5. The top line shows the original bit pattern [01001101001]. Fig. 4.5b shows the candidate bit patterns centred around the middle bit: 1. For reasons of clarity we used one inner bit and on each side one outer bit, together covering three bits of the original. In the actual simulations we used larger numbers. From the 8 candidate bit patterns, the resulting zeroth order and first order signal are calculated and compared with the measured signals by integrating the squared differences over the section indicated by a gray rectangle.

Fig. 4.5b shows in black solid line the signals corresponding to the correct candidate ([110]). For the zeroth order signal this curve is also the one that fits best with the measured signal. For the first order signal the correct bit pattern reassembles also relatively well with the measured signal, but there is however a better fit, from another candidate ([001]), shown in black dotted line. Combining the fit from zeroth and first order signal, may or may not lead to the correct bit pattern. This is shown in table 4.1: the lowest values selected of S_0^k and S_1^k, point to respectively the bit patterns [110] and [001]. The lowest value of the combination $\beta_0 S_0^k + \beta_1 S_1^k$ depends on the coefficients β_0 and β_1. In

this case using a combination of both modes has no advantage or has even a disadvantage over using only the zeroth order signal. In the next paragraphs a few extensions will be explained that will improve the extraction method. Moreover in section 4.4 will be shown that the advantage of the combination appears only if there are more outer bits m_o, or in case there is crosstalk or noise.

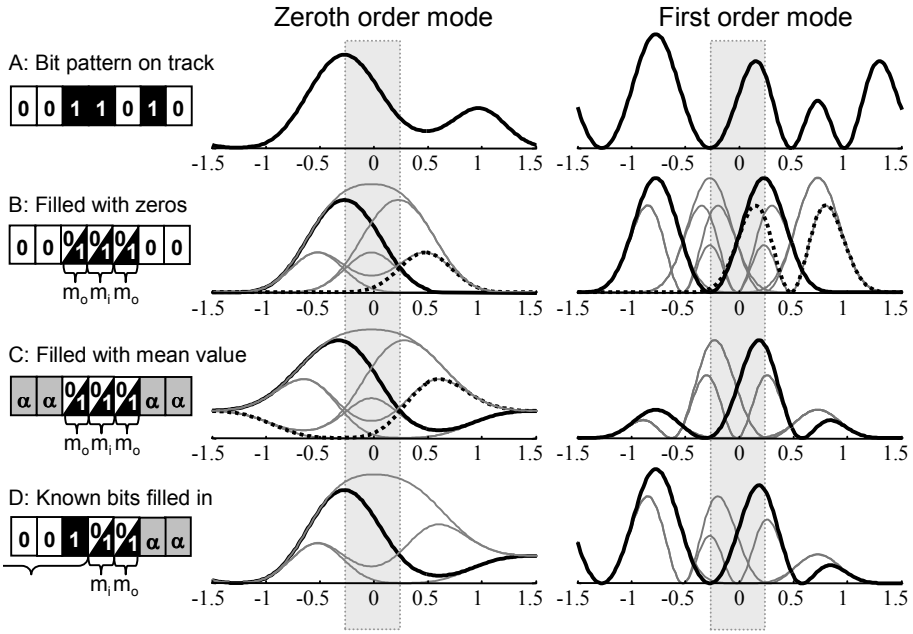


Figure 4.5: Detected signals and calculated signals. Graph A shows the bits on the disc and the measured signal at the detectors. Graph B shows the calculated signals corresponding with a candidate bit patterns covering three bits. The other bits are filled with 'zeros'. Graph C shows signals from candidate bit patterns were the other bits are filled with a mean value of 1/2. Graph D shows the calculated signals where knowledge of previously detected bits has been used to form the candidate bit patterns.

The example of Fig. 4.5 shows a problem with the first order signal: Because the $mtf_{01}(f)$ is zero for $f = 0$, the signal is insensitive to a constant reflection of the disc, it detects only the 'zero-one' transitions. This means that the first order signal cannot discriminate in between [110] and [001]. There is only a slight difference because of the influence of the surrounding bits, which were in both cases filled with 'zeros'.

A first adaptation of the parameter extraction method is filling up the edges of the candidate bit patterns with a mean value α instead of

Candidate bits	**000**	**100**	**010**	**001**	**110**	**011**	**101**	**111**
Zeroth order: S_0^k	132	105	39	115	<u>0.12</u>	48	0.69	0.60
First order: S_1^k	2.52	2.86	1.30	<u>0.0062</u>	0.18	3.75	1.35	2.44
Comb$_{0.01}$: $\beta_0 S_0^k + \beta_1 S_1^k$	3.82	3.89	1.67	1.15	<u>0.18</u>	4.20	2.03	3.03
Comb$_{0.001}$: $\beta_0 S_0^k + \beta_1 S_1^k$	2.65	2.96	1.34	<u>0.12</u>	.18	3.80	1.42	2.50

Table 4.1: Overlap values in between the measured signals from Fig. 4.5a and the calculated signals corresponding with the candidate bit patterns from Fig. 4.5b. Two combinations are given: Comb$_{.01}$ uses $\beta_0 = 0.01$ and $\beta_1 = .99$, and Comb$_{.001}$ uses $\beta_0 = 0.001$ and $\beta_1 = .999$.

with 'zeros'. For the case $\alpha = 1/2$ (as shown in Fig. 4.5c), the first order signals become degenerated: e.g. the signal from [110] is identical to that of [001]. This may look as a step back, but in fact this prevents that the first order signal may give a higher overlap value for the correct bit pattern than for the inverse bit pattern. By setting $\alpha = 1/2$, these overlap values are equal ($S_1^{[110]} = S_1^{[001]}$ in our example), and it will be the zeroth order signal that will be the deciding factor in $\beta_0 S_0^k + \beta_1 S_1^k$ for any β_0 or β_1. For other differences, e.g. between [100] and [101], the first order signal will have an impact and aid to discriminate in between two different possibilities.

In a second adaptation we use the knowledge of previously detected bits. As the bits are read out in a sequence, the bits on one side have already been detected, and can be glued to the candidate bit patterns instead of filling the area with a mean value α. This method is shown on Fig. 4.5d and has two advantages: As there are only outer bits on one side, the number of candidate bit patterns is decreased form $2^{2m_o+m_i}$ to $2^{m_o+m_i}$, which reduces the calculation time considerably. The second advantage is that the degeneration of the first order signal is solved.

4.3 Implementation

The parameter extraction method has been implemented in five steps in a MATLAB[1] environment.

Step 1 The point spread function $(psf_{tot,x})$ and optical transfer function $(otf_{tot,x})$ of the optical system are calculated by a numerical im-

[1]MATLAB (MATrix LABoratory) is a numerical software package from Mathworks Inc. (http://www.mathworks.com/)

plementation of Eqs. (3.95–3.96). The waveguide modes are calculated with the mode expansion tool CAMFR (Appendix C), and the Fourier transforms are evaluated with the fast Fourier transform (FFT) routines from MATLAB. For a given configuration of waveguides and lenses this calculation has only to be done once. This means the calculation speed is not very important and a fine discretization grid can be used.

Step 2 The signals $C_0^k(t)$ and $C_1^k(t)$ are calculated for all candidate bit patterns bp_c^k ($k = 0 \ldots n-1$). This is done by a convolution of the bit patterns with the point spread functions from step 1, as in Eq. (3.39). For the disc reflectivity Eq. (3.72) is used with $\eta = 1/2$ and $A = B = Y = 1$:

$$r_x^k(x) = 1 - bp_c^k \tag{4.5}$$

$$C_j^k(t) = \left| \int r_x^k(x-t)\, psf_{tot,x}^j(x)\, \mathrm{d}x \right|^2 . \tag{4.6}$$

The scanning speed of the disc has been set to 1. Calculation of Eq. (4.6) can be simplified by using the Fourier domain:

$$C_j^k(t) = \left| \widehat{r_x^k}(t)\, otf_{tot,x}^j(t) \right|^2 . \tag{4.7}$$

Step 3 In a real detection system the detectors capture a signal from the disc. We have simulated this, by generating a random bit pattern, $bp_0(x)$. Additive noise, level noise and offset noise are added, which results in $bp_0'(x)$. Two extra bit patterns, $bp_{-1}(x)$ and $bp_{+1}(x)$, are generated for the side tracks. This results in a simulated disc reflection function

$$r_x^{detect}(x) = bp_0'(x) + \eta_{-1} bp_{-1}(x) + \eta_{+1} bp_{+1}(x) \tag{4.8}$$

η_{-1} and η_{+1} are a measure for the amount of crosstalk. Similarly as in step 2, the resulting detector signals are calculated as:

$$P_j(t) = \left| \widehat{r_x^{detect}}(t)\, otf_{tot,x}^j(t) \right|^2 . \tag{4.9}$$

The advantage of a simulation is that we know in advance which of the candidate bit pattern corresponds to the equivalent section on the track on the disc. Unlike step 1 and step 2, step 3 is executed many times. Therefore calculation speed is also important and a trade-off is made for the discretization grid.

Step 4 The results from step 2 and step 3 are combined by calculating the overlap values S_0^k, S_1^k according to Eq. (4.1):

$$S_j^k = \int_{t_l}^{t_u} \left(P_j(t) - C_j^k(t) \right)^2.$$

The integration domain, [t_l,t_u] is chosen identical to that of the inner bits (m_i) of the candidate bit patterns.

Step 5 The candidate bit pattern corresponding with the minimum overlap value is determined (Eq. 4.10) and it is checked whether the inner bits of that candidate correspond with the bits on the track, for a few different values of the linear combination coefficient $\beta = \beta_0/(\beta_0 + \beta_1)$.

$$S_j^{k_{min}} = \min_{k=1\ldots n} \left[\beta S_0^k + (1-\beta)\, S_1^k \right]. \tag{4.10}$$

$$m_i(k_{min}) \stackrel{?}{=} m_i(k_{correct}) \tag{4.11}$$

We can check whether we found the correct bit pattern because we know which bit pattern $bp_0(x)$ has been used in step 3 to form the 'detected' signals P_0 and P_1. In a real pick-up system, this original bit pattern is however not known and the above algorithm is used to reconstruct it from the detected signals P_0 and P_1.

Steps 3 to 5 are repeated many times and the number of correct and incorrect guesses for a given β are added up. The fraction of the number of incorrect guesses on the total number of guesses is the bit error rate (BER). These simulations were repeated for a series of different bit lengths and types of noise. The bit error rate for the zeroth order signal P_0, is called BER_0. For the first order signal P_1, it is called BER_1.

4.4 Results

Fig. 4.6 shows the bit error rate, as a function of the linear combination coefficient β (As in Eq. (4.10)) The graph shows two different configurations of bit length and type of noise. Fig. 4.6A suggests that the first order signal (P_1) shows a lower bit error rate than the zeroth order signal (P_0) for high crosstalk: $\mathrm{BER}_1 < \mathrm{BER}_0$. For lower crosstalk however, inter-symbol interference becomes dominant factor of distortion, and the zeroth order signal results in a lower bit error rate: $\mathrm{BER}_1 > \mathrm{BER}_0$. This is also the case for additive noise. For most cases, there is however

a minimum bit error rate, $\mathrm{BER}_{\min}$, for $\beta = \beta_{min}$. Unfortunately β_{min} depends on the type and amount of noise. By making an average for different types of noise $\beta_{\mathrm{opt}} = 10^{-2}$ was chosen as optimal value. This choice depends on the expected type and amount of noise but also on the configuration of the waveguide detection method itself.

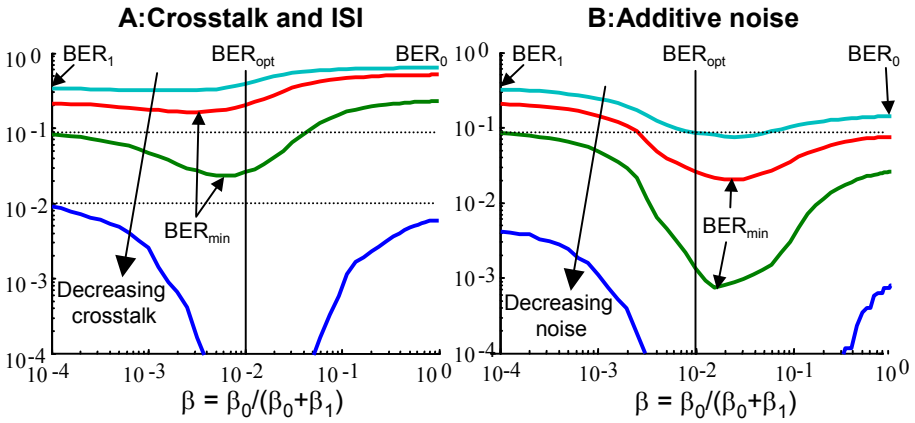


Figure 4.6: Bit error rate as a function of β. The left edge of the curves gives BER_1 and the right edge gives BER_0. On the left graph the bit error rate is shown for different amounts of additive noise and a bit length of $0.15\lambda/NA$, on the right for inter-symbol interference and different amounts of crosstalk and a bit length of $0.25\lambda/NA$.

Ideally the simulations should be done for bit error rates similar to these in a real optical disc system, which are of the order of 10^{-5}. For reasons of calculation speed, we could not perform a very large number of bit patterns. For each noise level and for each bit length we simulated 10^4 bit patterns. This means we can detect bit error rates as low as 10^{-4}. The accuracy of these low bit error rates is however not as good as that of higher bit error rates. Fig. 4.6 shows however a trend that the improvement in the bit error rate is especially big for configurations where the bit error rate is already relatively low, and that the improvement is less considerably for high noise or strong crosstalk. This suggest that the improvement in bit error rate by using a combination of the two modes may even be bigger than in our simulations.

4.4.1 Inter-symbol interference

In absence of noise and side tracks, the detected signals are identical to the theoretical signals and the only relevant influence on the bit er-

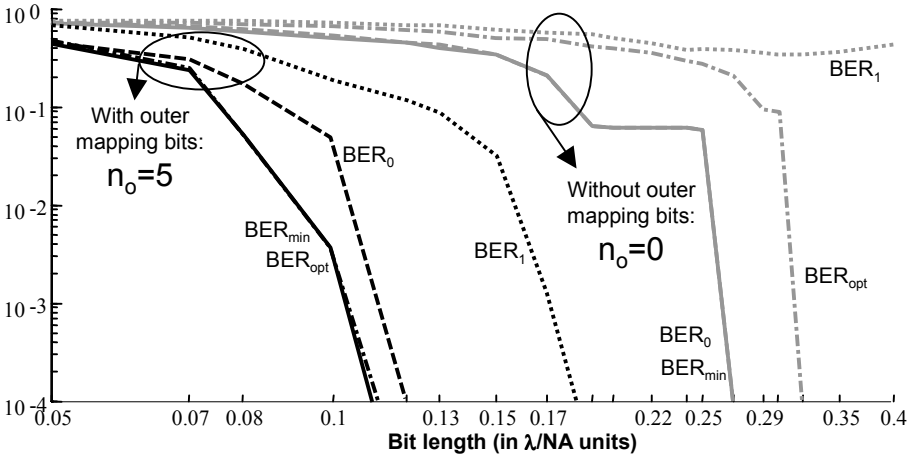


Figure 4.7: In case the detected signal is not distorted, there are still bit errors because of inter-symbol interference (ISI). The bit error rate depends on the bit length, but also on the parameter extraction method. The bit error rate for a bit extraction without outer candidate bits is shown in gray lines. In black lines the method with 5 outer candidate bits is given. In both cases the graph shows BER_0, BER_1, $\mathrm{BER}_{\min}$ and $\mathrm{BER}_{\mathrm{opt}}$.

ror rate is the inter-symbol interference. The bit error rate increases rapidly for decreasing bit lengths, but depends also on the number of candidate bit patterns. As shown on Fig. 4.7 the first order signal is always worse than the zeroth order signal, which is logical because of the broader width of its point spread function. For a high number of outer candidate bits, the bit error rate can however be improved by using a combination of both signals. For a low number of outer candidate bits this is not the case. For these very short bit lengths, the bit error rate derived from the combination of both modes with $\beta = \beta_{\mathrm{opt}}$, $\mathrm{BER}_{\mathrm{opt}}$, is even worse than that of the zeroth order signal alone, BER_0. In the following simulations we used $m_o = 3$ as a trade-off between speed and lower BER.

The graph also shows that even for bit lengths below the diffraction limit ($d_{bit} < \lambda/(4NA)$), the bit error rate can be relatively small. For a given bit length d_{bit}, it is even possible to reduce the bit error rate to any given number ϵ higher than zero, by extending the number of outer candidate bits. It is however only for bit lengths above the diffraction limit, that it is however possible to attain a bit error rate of zero. Below this limit the bit error rate can be made very low but will always be larger than zero, even with a very large number of outer candidate bits.

A single example is sufficient to prove this: assume a track containing an infinite sequences of 'zeros' and 'ones': '101010101010101010'. If the bit length is below $\lambda/(4NA)$, the period (containing a 'one' and a 'zero') is smaller than $\lambda/(2NA)$ and the smallest spatial frequency is then $2NA/\lambda$. This means that the detected signal will be constant as a function of time, regardless of which mode is used for illumination and detection. In that case it will be impossible to say whether the bit at the current position is a 'zero' or a 'one'.

4.4.2 Additive noise

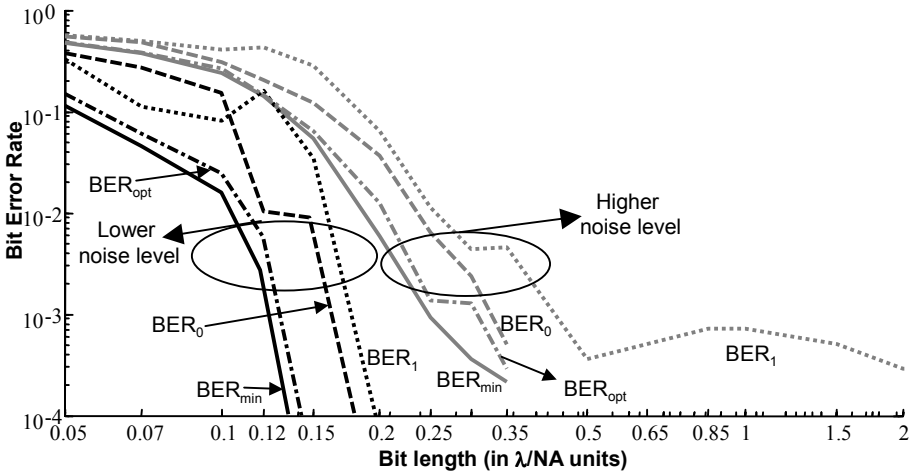


Figure 4.8: Bit error rate for additive white noise. The gray lines show the bit error rate for a disc with a high amount of noise, the black lines show the situation with a lower amount of noise. The solid, dash-dotted, dashed and dotted lines respectively show $\mathrm{BER}_{\mathrm{min}}$, $\mathrm{BER}_{\mathrm{opt}}$, BER_0, and BER_1.

Fig. 4.8 shows the bit error rate as a function of the minimum bit length for two different amounts of additive white noise. In most cases the zeroth order signal leads to the lower bit error rate than the first order signal. Remark that for a high noise level the bit error rate for the first order signal (BER_1) does not drop further for increasing bit lengths. The reason is that the first order signal is only sensitive to changes in between a 'zero' and a 'one'. While only the edges of the bits are detected, the noise is added over the whole length of the bits. The zeroth order mode signal is less sensitive to his kind of noise. But using a combination of both signals reduces the bit error rate even further:

$\mathrm{BER}_{\min}$ and $\mathrm{BER}_{\mathrm{opt}}$ are both smaller than either BER_0 or BER_1. Which means that for a given maximum bit error rate, the bit length can be made smaller.

4.4.3 Crosstalk and inter-symbol interference

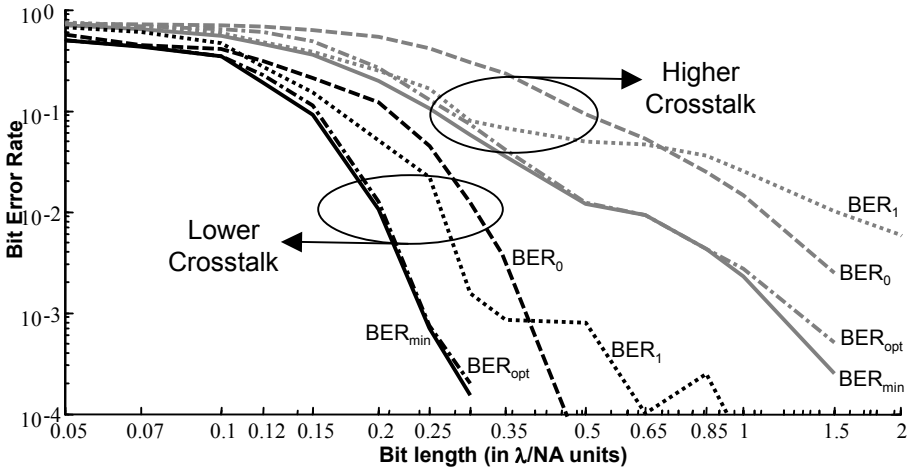


Figure 4.9: Bit error rate as a function of the minimum bit length. The black lines show the bit error rate for a disc with strong crosstalk, and the gray lines for less crosstalk. The inter-symbol interference is equal for both situations. The solid, dash-dotted, dashed and dotted lines respectively show $\mathrm{BER}_{\min}$, $\mathrm{BER}_{\mathrm{opt}}$, BER_0, and BER_1.

Fig. 4.9 shows the bit error rate as a function of the minimum bit length for two different amounts of crosstalk. Whether the zeroth order or the first order signal generates the lowest bit error rate depends on the bit length and the amount of noise. This could already be seen on Fig. 4.6b, which shows the bit error rate for a fixed bit length ($0.25\lambda/NA$), for different values of the combination coefficient β. The first order mode signal tends to be more sensitive to inter-symbol interference and less sensitive to crosstalk. Moreover the first order mode signal has difficulties with very large bit lengths. This explains why BER_1 becomes larger than the BER_0 for large bit lengths. It also explains why for heavy crosstalk this crossing occurs at a larger bit length. Except for very small bit lengths, making use of a combination of both signals reduces the bit error rate even further: $\mathrm{BER}_{\min}$ and $\mathrm{BER}_{\mathrm{opt}}$ are both smaller than either BER_0 or BER_1.

4.5 Conclusions

In this chapter we showed that even with a simple parameter extraction method it is possible to extract bits on the disc, spaced narrower than the diffraction limit, $\frac{\lambda}{4NA}$. Depending on the amount of noise, crosstalk and inter-symbol interference, some of the detected bits will however be incorrect. The ratio of incorrect bits, the bit error rate, gives a measure how well the bits can still be discriminated. In most cases the zeroth order signal, which is identical to that of a confocal microscope, gives a lower bit error rate than the first order signal. Using a combination of both signals however can reduce the bit error rate even further.

This nice result has however still some problems. The most important is power. The power in the first order mode signal (P_1) is much lower than in the zeroth order mode signal (P_0). This explains why the optimal coefficient in the linear combination of Eq. (4.10) is so low: $\beta_{\text{opt}} = 10^{-2}$. This means that the first order mode signal has to be amplified very strongly. One type of noise which was not included in this simulation, namely detector noise, will hence have a much stronger effect on this first order mode signal.

In this chapter we showed the results for illumination with the zeroth order mode and detection with the zeroth and the first order mode. In principle one could use the second or the third order mode to capture a detection signal. It is however questionable whether this will lead to another big improvement in bit error rate. By using the first order mode additionally to the zeroth order mode, we obtain information from the phase as well as the amplitude of the reflected field. Adding also the second order mode signal will not get supplementary information from such an extra parallel channel. Moreover the problem that the first order signal is relatively weak, holds even much more for the second order mode.

Another possible improvement would be not to use the zeroth order mode to illuminate the disc but a combination of more modes to form a flat otf as in Eq. (3.141). This might be an improvement but as described in section 3.7 there are two big problems: intolerance to small variations in the ratio of the illumination modes and decrease in overall power throughput. The third disadvantage, the higher side lobes, may be much less a problem using a parameter extraction method instead of the conventional bit detection method.

For a reliable detection the raw bit error rate should be below a certain value. The raw bit error rate in a conventional compact disc

should be below 10^{-5}, and is further reduced afterwards by error correcting codes. The results in section 4.4 show that by combing zeroth and first order signal, this threshold value can be achieved for smaller bit lengths, or higher noise than when using only one of the signals, which also proves that the waveguide detection method has a potential better resolution than the conventional DVD detection method.

Chapter 5

Photonic integrated circuit

A photonic integrated circuit, often abbreviated as PIC or photonic IC, is a group of components etched or imprinted on a single substrate, offering light generation, manipulation and/or detection. This is in contrast to free space optics, where all components are fabricated and aligned separately. Real photonic ICs combining all three: generation, manipulation and detection, are relatively rare. Mostly one uses a hybrid approach, combining an integrated chip with free space components. In this thesis a photonic IC has been designed and fabricated to manipulate the different modes in the detection and illumination waveguides. An overview is given of the theory, design, and measurements of two components that make it possible to split off the different modes of a multimodal waveguide. In section 5.1, a multimode interference coupler splitting off the odd and the even modes, is outlined. The principles of a mode splitter based on a directional coupler are described in section 5.2. Section 5.3 briefly explains an integrated optical interferometer that can detect accurately the phase difference between the modes of two separate waveguides. The design of the components and the fibre interface are described in section 5.4 and section 5.5.

5.1 Mode splitter based on multimode interference couplers

One way of controlling the modes in a multimodal waveguide has been proposed by Leuthold [3]. The basic principle relies on the different mapping characteristics of symmetric and antisymmetric modes in multimode interference couplers (MMIs). With this method it is pos-

sible to separate the odd and even modes of a multimodal waveguide to different output waveguides.

5.1.1 Multimode interference couplers

The multimode interference coupler [58, 59] is a component that uses mode interference within a multimodal waveguide in order to image the field at the input waveguide(s) onto the output waveguide(s). In general an $N \times M$ MMI has N input and M output waveguides. On Fig. 5.1 a 1×2 MMI is shown. An interesting tool to analyse an MMI is the eigenmode expansion method, described in more detail in appendix C. It is not only a strong tool for design and numerical modelling, but also provides a good physical insight into the mechanism of multimode interference.

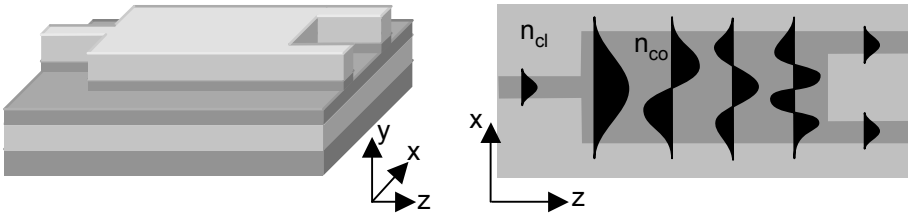


Figure 5.1: A 1×2 MMI and the two-dimensional representation.

For exact numerical modelling one should do a full 3D analysis but the principle can be very well explained in two dimensions, without loss of generality. The reduction from three to two dimensions can be done by the effective index method (appendix C.3). Fig. 5.1 shows a 1×2 MMI. In the two-dimensional representation the layered structure is replaced with two effective refractive indices: n_{co} for the waveguide and the MMI and n_{cl} for the surrounding area. If the refractive index along the x-axis has a step profile, $n(x) = \Pi\left[\frac{x-W/2}{W}\right]$, the eigenmodes may be approximated by

$$\phi(x) = \sin\left[\pi\left(\nu+1\right)\frac{x}{W_{\mathit{eff},\nu}}\right] \text{ with } \nu = 0, 1, 2, \ldots \tag{5.1}$$

The width of the eigenmodes, $W_{\mathit{eff},\nu}$, is larger for the higher order modes because of a lateral penetration depth, associated with the Goos-Hähnchen shift. For high contrast waveguides, all mode widths may however be approximated by the effective width of the fundamental

mode $W_{eff,0}$. This effective width which will be written as W_e and is roughly equal to the width of the MMI. The lateral wave number $k_{x,\nu}$ and the propagation constant β_ν of mode ν are connected by the dispersion equation:

$$k_{x,\nu}^2 + \beta_\nu^2 = k_0^2 n_{co}^2 \qquad \text{with } k_0 = \frac{2\pi}{\lambda} \text{ and } k_{x,\nu} = \frac{(\nu+1)\,\pi}{W_e} \tag{5.2}$$

Using a binomial expansion with $k_{x,\nu} \ll k_0^2 n_{co}^2$, the propagation constant β_ν can be deduced as

$$\beta_\nu \simeq k_0 n_{co} - \frac{(\nu+1)^2\,\pi\lambda}{4 n_{co} W_e^2}. \tag{5.3}$$

This equation gives an approximative relationship between the propagation constants of the different modes:

$$\beta_0 - \beta_\nu = \frac{\nu\,(\nu+2)\,\pi}{3L_\pi} \qquad \text{with } L_\pi = \frac{4 n_{co} W_e^2}{3\lambda} \tag{5.4}$$

The length L_π is the beat length between the two lowest order modes. By this equation one can find already many self-imaging properties of the MMI.

To a certain extent one can make the assumption that the radiative modes of the MMI are not excited. This assumption is not entirely correct but as long as the input waveguide mode is not too small and the MMI has at least a few guiding modes, it is a good approximation. This means the input field $\psi(x, z=0)$ can be written as a composition of only the m guided eigenmodes of the MMI.

$$\psi(x, z=0) = \sum_{\nu=0}^{m-1} c_\nu \phi_\nu(x) \tag{5.5}$$

At a distance L the field profile is a superposition of the individual field profiles of the guided modes. If we drop out a common phase factor, such that the phase of the fundamental mode is zero we find:

$$\psi(x, z=L) = \sum_{\nu=0}^{m-1} c_\nu \phi_\nu(x) \exp[j\,(\beta_0 - \beta_\nu)\,L]. \tag{5.6}$$

After substituting $\beta_0 - \beta_\nu$ by Eq. (5.4) this becomes:

$$\psi(x, z=L) = \sum_{\nu=0}^{m-1} c_\nu \phi_\nu(x) \exp\left[j \frac{\nu\,(\nu+2)\,\pi}{3L_\pi} L\right]. \tag{5.7}$$

Under certain circumstances the field $\psi(x, z = L)$ will be a reproduction of the input field $\psi(x, z = 0)$. Using the symmetry and antisymmetry of even and odd modes with respect to the plane $x = W/2$, Eq. (5.7) shows that $\psi(x, z = L)$ will be an exact image of $\psi(x, z = 0)$ if

$$\exp\left[j\frac{\nu\,(\nu+2)\,\pi}{3L_\pi}L\right] = 1 \text{ or } (-1)^\nu\,. \tag{5.8}$$

The first condition says that the phase changes of the eigenmodes of the MMI over a length L should be a multiple of 2π. Which means the field at $z = L$ is an exact replica of the field at $z = 0$. The second condition says that the even modes are in phase and the odd modes in antiphase, which means the image is mirrored with respect to the plane $x = 0$. The expression $\nu\,(\nu + 2)$ is odd for ν odd, and even for ν even. This means that the first and second condition of Eq. (5.8) will be fulfilled at

$$L = p\,(3L_\pi) \text{ with } p = 0, 1, 2, \ldots \tag{5.9}$$

For a fractional value of p ($p = M/N$, with M and N having no common divisors) one can observe multiple images at the output plane. This is well explained by a paper of Bachmann [60]. This theory will be briefly outlined here as it will be used in next section to explain the operating principles of the mode splitting MMI.

The incident light distribution $\psi(x)$, defined in the real MMI section $0 < x < W$, is extended over the entire x-axis and named $\psi_{in,ext}(x)$, by

$$\begin{aligned} g(x) &= \psi(x) - \psi(-x) \\ \psi_{in,ext}(x) &= \sum_{n=-\infty}^{\infty} g(x - 2nW). \end{aligned} \tag{5.10}$$

or simplified:

$$\psi_{in,ext}(x) = \sum_{n=-\infty}^{\infty} \psi(x - 2nW) - \psi(-x - 2nW). \tag{5.11}$$

Now Bachmann [60] states that if the MMI has a length $L = \frac{M}{N}3L_\pi$, then the light distribution at the end of an MMI, $\psi_{out}(x)$, will be a summation of N images with equal amplitudes $1/\sqrt{N}$:

$$\begin{aligned} \psi_{out,ext}(x) &= \frac{1}{\sqrt{N}}\sum_{t=0}^{N-1}\psi_{in,ext}(x - x_t)\exp(j\varphi_t) \\ \text{with } x_t &= (2t - N)\,\frac{M}{N}W \text{ and } \varphi_t = t\,(N - t)\,\frac{M}{N}\pi \end{aligned} \tag{5.12}$$

An example is given by Fig. 5.2. On the first line the input field $\psi(x)$ is expanded to $\psi_{in,ext}(x)$. The second line gives the output field for an MMI with length $L = 3L_\pi$, which is the single image case, mirrored over the $x = W/2$ axis. The third line gives the output field for an MMI with length $L = 3L_\pi/2$, two images of the input field. The phase of these images is not shown on Fig. 5.2, but as $\varphi_0 = 0$ and $\varphi_0 = \pi/2$, the two images are in quadrature.

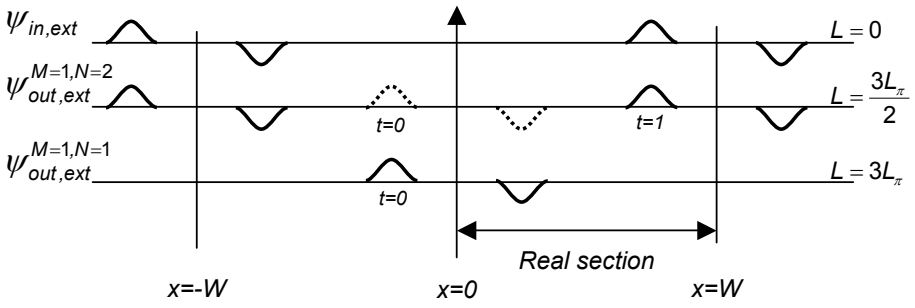


Figure 5.2: Bachmann method: The original input field (the section in between x=0 and x=W of the top line), is extended periodically. According to Eq. (5.12), this extended field is reproduced a N times at different positions and with different phase lag, for an MMI with length $L = 3L_\pi/N$. The interference in between these images gives the final output field. For M=1, there is only one image, $t = 0$, marked in solid line. For M=2, there are two images in quadrature. One in dotted line for $t = 0$, and one in solid line for $t = 1$. The phase of these images is not shown not the graph.

5.1.2 Splitting up odd and even modes

The even and the odd modes in a central input waveguide, will excite respectively only the even and the odd modes of the MMI. In the case of even mode excitation the effective length at which a single image is formed is divided by a factor 4. This can easily be derived by inspecting Eq. (5.8). The stated condition needs only to be fulfilled for even ν. Moreover for even ν: $\mod_4 [\nu(\nu+2)] = 0$. The length periodicity will hence be reduced four times and an exact image will be formed at $L = 3L_\pi/4$. If the excitation of the MMI has however an odd symmetry this principle does not hold any more and the input field will not be imaged at $L = 3L_\pi/4$.

To give a deeper understanding into the different mapping properties of the odd and even modes we will use the above described

method of Bachmann. Following Eq. (5.12) the extended input function $\psi_{in,ext}(x)$ generates a four-fold image at $L = 3L_\pi/4$ and an eight-fold image at $L = 3L_\pi/8$. Because of the central position of the excitation some of these images coincide. A closer inspection of the location x and the phase ϕ of the different images, will show that on some positions there is constructive interference and on others there is destructive interference.

For a symmetric and antisymmetric input field centred around $x = W/2$ one has respectively

$$\psi\left(\frac{W}{2}+x\right) = \psi\left(\frac{W}{2}-x\right), \qquad \text{for symmetric inputs} \tag{5.13}$$

$$\psi\left(\frac{W}{2}+x\right) = -\psi\left(\frac{W}{2}-x\right), \qquad \text{for antisymmetric inputs} \tag{5.14}$$

Using this relations, Eq. (5.10) can be rewritten as

$$\psi_{ext,in}(x) = \sum_{n=-\infty}^{\infty} \psi[x-(2n)\,W] - \psi[x-(2n+1)\,W] \qquad \text{(sym. input)} \tag{5.15}$$

$$\psi_{ext,in}(x) = \sum_{n=-\infty}^{\infty} \psi[x-(2n)\,W] + \psi[x-(2n+1)\,W] \quad \text{(anti-sym. input)}. \tag{5.16}$$

This extended input function is hence composed of two functions with period $2W$:

$$\begin{aligned} f_{real} &= \sum_{n=-\infty}^{\infty} \psi(x-(2n)\,W) \\ \text{and } f_{virtual} &= \sum_{n=-\infty}^{\infty} \psi(x-(2n+1)\,W). \end{aligned} \tag{5.17}$$

These are identical disregarding a mutual shift by half the period, W. For a symmetric input field f_{real} and $f_{virtual}$ are subtracted, for an antisymmetric input field they are added up. The top lines of Fig. 5.3 and Fig. 5.4 show a graph of the extended input function for a symmetric and an antisymmetric input field respectively. Using Eq. (5.12) one can calculate table 5.1, which gives an overview of the position and the phase of the images of this extended input field at a length

	$L = 3L_\pi/4$			
t	0	1	2	3
x	$-W$	$-W/2$	0	$W/2$
φ	0	$3\pi/4$	π	$3\pi/4$

Table 5.1: An MMI with length $L = 3L_\pi/4$ results in four identical images of the input field. This table gives the position and the phase of these images.

$L = 3L_\pi/4$. The second line of Fig. 5.3 and Fig. 5.4 show these images at length $L = 3L_\pi/4$. Because of the lateral shift W, in between f_{real} and $f_{virtual}$, the image of f_{real} for $t = 0$ overlaps with the image of f_{real} for $t = 2$. The same applies for the couples $(t = 1, t = 3)$, $(t = 2, t = 0)$, and $(t = 3, t = 1)$. For symmetric input fields f_{real} and $f_{virtual}$ are subtracted, which means constructive interference is found for positions with opposite phase: $t = 0$ and $t = 2$. For antisymmetric input fields f_{real} and $f_{virtual}$ are added, and constructive interference is found for positions with equal phase: $t = 1$ and $t = 3$.

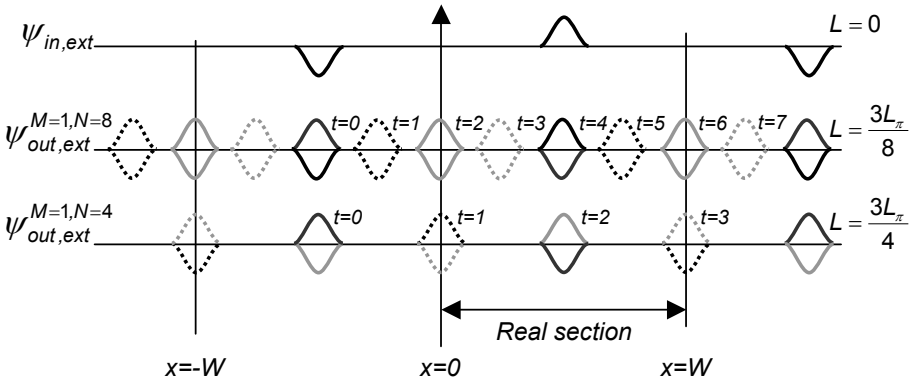


Figure 5.3: Output field for a zeroth order mode excitation of the central input waveguide, according to Bachmann's method. The input field is periodically extended (top line). At a length $3L_\pi/N$ this extended field is replicated N times, each with a different phase (not shown on graph). Some parts of these images overlap and interfere. For $N = 4$, and taking into account the sign of the mode in the different sections, there is constructive interference for $t = 0$ and $t = 2$, which means only one effective image of the input mode (mode stopper and mode splitter). For $N = 8$, there is constructive interference for $t = 1$, $t = 3$, $t = 5$, and $t = 7$, which means two equal images of the input mode (3dB splitter).

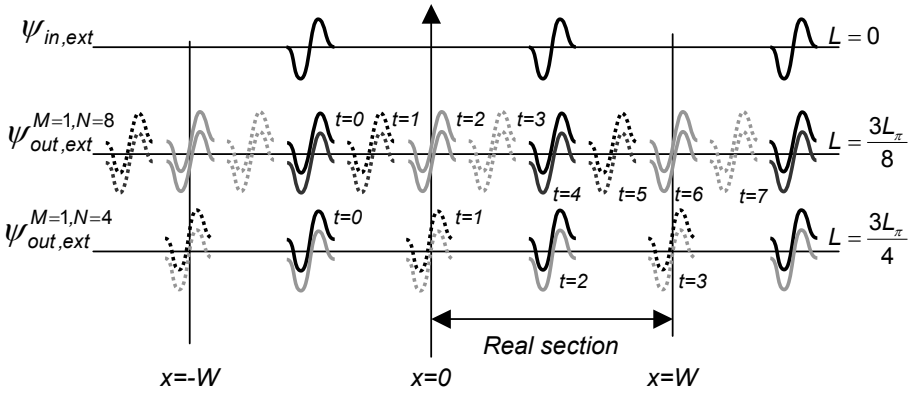


Figure 5.4: Output field for a first order mode excitation of the central input waveguide, according to Bachmann's method. The input field is periodically extended (top line). At a length $3L_\pi/N$ this extended field is replicated N times, each with a different phase (not shown on graph). Some parts of these images overlap and interfere. For $N = 4$, there is constructive interference for $t = 1$ and $t = 3$, which means the first order mode is split into two parts and imaged at the edge of the MMI (mode stopper and mode splitter). For $N = 8$, there is constructive interference for $t = 0$, $t = 2$, $t = 4$, and $t = 6$, and there are three images in the real section: two halves at the edges of the MMI and a complete image in the centre (3dB splitter).

	$L = 3L_\pi/8$							
t	0	1	2	3	4	5	6	7
x	$-W$	$-3W/4$	$-W/2$	$-W/4$	0	$W/4$	$W/2$	$3W/4$
φ	0	$7\pi/8$	$3\pi/2$	$-\pi/8$	0	$-\pi/8$	$3\pi/2$	$7\pi/8$

Table 5.2: An MMI with length $L = 3L_\pi/8$ results in eight identical images of the input field. This table gives the position and the phase of these images.

Using Table 5.2 for the position and the phase of the images at $L = 3L_\pi/8$, one finds analogously that a central input of a symmetric mode gives two images: at $x = W/4$ and at $x = 3W/4$. An antisymmetric mode results in destructive interference at these positions.

5.1.3 Components

In this section we will describe the three components that have been used in the actual design. They will be called the 'mode stripper', the '3dB splitter', and the 'mode splitter'. Each of these components is

based on the imaging properties described in section 5.1.2. The choice of the dimensions depends on the object function one wants to optimize, which on its turn depends on the wanted functionality of the final component. We will explain the design rules along with the components.

5.1.3.1 The mode stripper

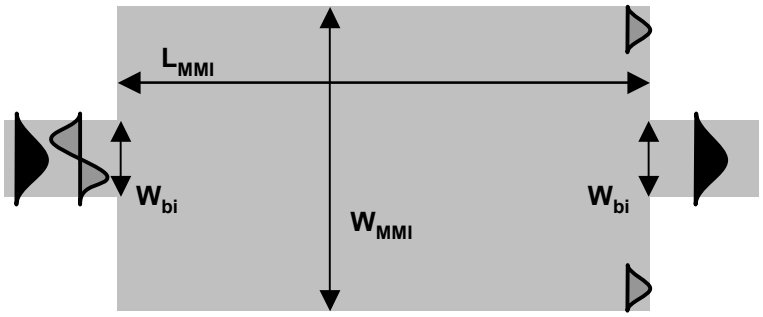


Figure 5.5: The mode stripper is a 1×1 MMI that images the zeroth order input mode to the zeroth order output mode and removes the first order input mode.

The mode stripper is the simplest component. The main objective is to image the zeroth order with no loss and to remove the first order mode completely. As described in section 5.1.2, we can use an MMI with a central in- and output waveguide and a length $L = 3L_\pi/4$ (Fig. 5.5). The zeroth order mode will be imaged at the central output and the first order mode will result in destructive interference at the centre and constructive interference at the edges of the MMI. For the design of the MMI we want to optimize the remaining free parameter, the width (W_{MMI}), in order to minimize the loss in the zeroth order mode and the crosstalk from the first order mode. Moreover we want the component to be as robust as possible to small deviations from the optimal dimensions.

There are two reasons why the width of the MMI should not be too small. For an interference pattern ,there need to be at least two even modes and two odd modes in the MMI, otherwise there is no imaging, only a loss because of coupling to radiative modes. The second reason is that if the width is too small the imaged spots of the first order mode at the edges of the MMI do overlap with the output waveguide, which results in crosstalk.

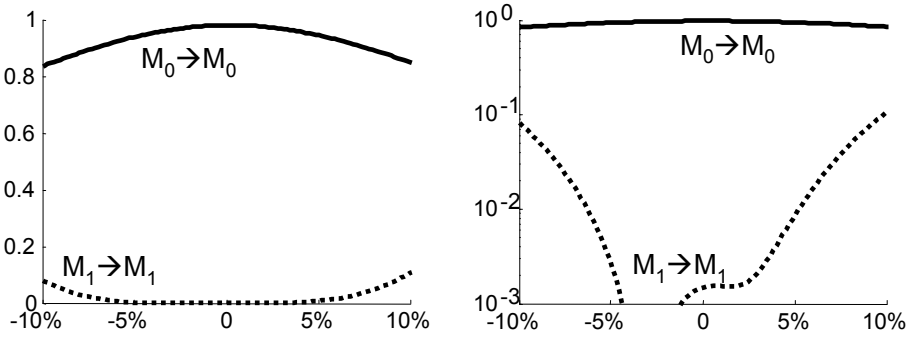


Figure 5.6: Mode transmission tolerances in the mode stripper designed in mask 2 (28.5 µm long and 2.75 µm broad) on a linear and a logarithmic scale. The zeroth order transmission is depicted in solid line and the first order in dotted line. The x-axis shows the relative deviation of the actual length from the designed value.

The reason why we should neither make the MMI too broad is to reduce the sensitivity to small deviations from the designed dimensions. Like with Gaussian beams in free space, the initial slope at which the input mode spreads out inside the MMI, is inversely proportional to the width of the input mode. Because of the symmetry in the component, the slope at which the images spread out at the start, is equal to the slope at which the image is formed at the other side of the MMI. This means that the power loss for a given deviation from the optimal length is roughly independent of the width of the MMI. If we differentiate the optimal length using Eq. (5.4) we find:

$$L = \frac{3L_\pi}{4} = 3\frac{n_{eff}\, W_{MMI}^2}{\lambda}$$

$$\mathrm{d}L = 6\frac{n_{eff}}{\lambda} W_{MMI}\mathrm{d}\, W_{MMI} \qquad \text{and} \qquad \frac{\mathrm{d}L}{L} = 2\frac{\mathrm{d}\, W_{MMI}}{W_{MMI}} \tag{5.18}$$

This means that a variation on the width results in a variation of the optimal length, proportional to the total width of the MMI. In order to minimize the difference between the optimal length and the real length of the MMI, the MMI should be kept small. This short tolerance analysis complies with the full analysis of the fabrication tolerances by Besse [61]. A second advantage is that the component becomes more compact. Fig. 5.6 shows the effect of a 10% deviation of the length of the MMI, which is equivalent to a 5% deviation in the width, for the mode stripper on mask 2.

5.1.3.2 The 3dB splitter

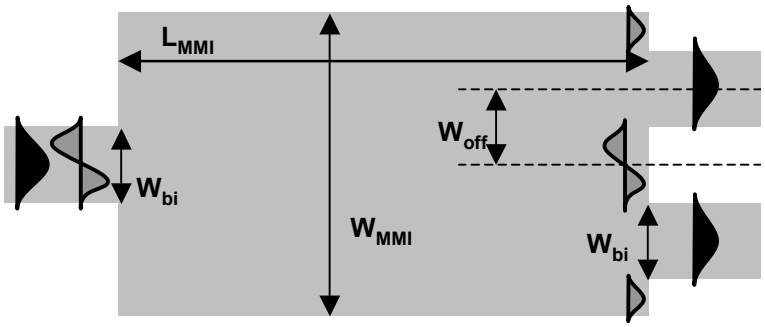


Figure 5.7: The 3dB splitter is a 1×2 MMI that splits the zeroth order mode equally towards two output waveguides and filters the first order mode out.

The 3dB splitter should make two equal images of the zeroth order input mode and remove the first order mode. As described in section 5.1.1 we can use an MMI with a central in- and output waveguide and a length $L = 3L_\pi/8$ (Fig. 5.7). The zeroth order mode will generate two images, one at an offset $-W_{MMI}/4$ and one at $+W_{MMI}/4$. The first order mode will result in constructive interference in the very centre and at the very edges of the MMI. The same considerations as for choosing the dimensions of the mode stripper hold for the 3dB splitter. There is however the additional constraint on the distance between the output waveguides. It should be large enough to prevent the images of the zeroth order mode to overlap. This condition also prevents parasitical coupling in between the waveguides. A stronger constraint is however that the images of the first order mode do not overlap with the output waveguide. These constraints will result in a broader MMI than for the mode stripper. Fig. 5.8 shows the intensity of the optical field within a 1×2 MMI, for a zeroth and for a first order mode excitation. Fig. 5.9 shows the effect of a 10% deviation of the length of the MMI, which is equivalent to a 5% deviation in the width, for a designed mode stripper.

5.1.3.3 The mode splitter

The mode splitter is a component very similar to the mode stripper, with the difference that the image of the first order mode at the edges of the MMI is collected by additional output waveguides. The most crucial aspect of the design of the mode splitter is crosstalk from the zeroth

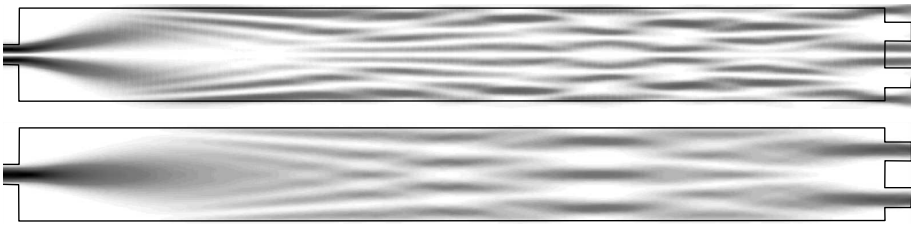


Figure 5.8: Intensity of the optical field in an 1×2 multimode interference coupler. On top for a zeroth order mode excitation, at the bottom for a first order mode excitation.

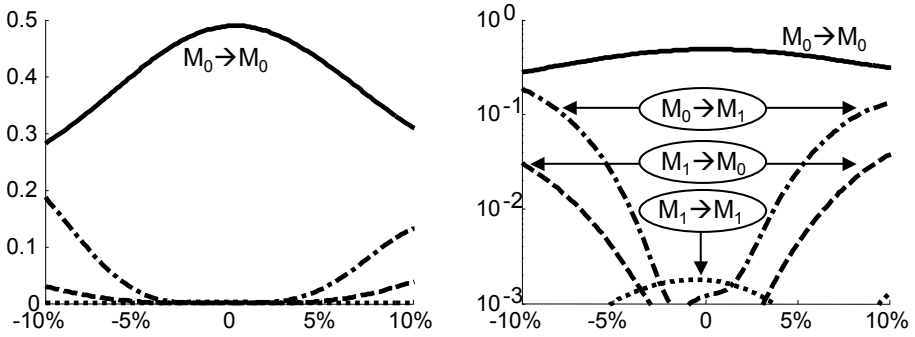


Figure 5.9: Mode transmission tolerances in the 3dB splitter designed for mask 2 (65 µm long and 6 µm broad) on a linear and a logarithmic scale. The transmission of the zeroth order input to the zeroth order mode of the output is in solid line, and to the first order output is in dotted line. The transmission of the first order input to the zeroth order mode of the output is in dashed line, and to the first order output is in dash-dotted line. The x-axis shows the relative deviation of the actual length from the designed value.

order mode input waveguide mode towards the outer output waveguides. Fig. 5.11 shows the effect of a 10% deviation of the length of the MMI, which is equivalent to a 5% deviation in the width, for a designed mode splitter. Remark that the power of the first order mode is split over two output waveguides. In principle one could avoid this power loss by combining these outputs using a supplementary MMI [62].

A comparison of Fig. 5.6 and Fig. 5.11 shows that light throughput of the zeroth order mode in the mode stripper drops to 82% for 10% deviation on the MMI length. For the mode splitter this zeroth order mode light throughput drops to 50%. This is as expected from the theory, which states that broader MMIs are less tolerant to small variations in the MMI dimensions.

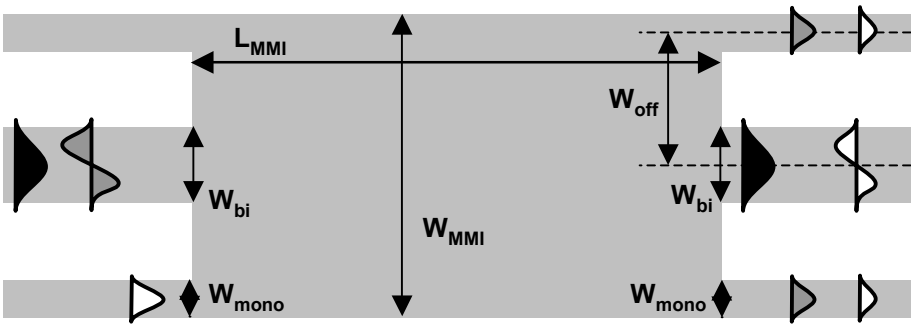


Figure 5.10: The mode splitter is a 3×3 MMI that images the zeroth order mode of the central input waveguide to the zeroth order mode of the central output waveguide. A first order mode in the central input waveguide is imaged onto the zeroth order mode of the outer output waveguides, each with half of the power of the input mode. A zeroth order mode at an outer input waveguide is split into three parts: half of the power is imaged onto the first order mode of the central output waveguide and the remaining half of the power is equally divided over the zeroth order modes of the outer output waveguides.

5.2 Mode splitter based on directional couplers

With the method described in the previous section one can split off the symmetric and the antisymmetric modes. As the method in previous section can only separate the even from the odd modes, we investigated ways to split two symmetric modes to different output channels. A method for doing so is a directional coupler that make uses of the difference in propagation constants between the modes.

A directional coupler is a component in which two waveguides are parallel next to each other. As shown on Fig. 5.12 there is a lead-in section, a coupling section and a lead-out section. As a function of the dimensions and the refractive indices of the two waveguides, light coupled into one waveguide will couple partially or entirely to the other section. In general the directional coupler is less tolerant to variation in the wavelength, polarization, and dimensions of the component [63]. A recent paper showed that the directional coupler and the MMI are two components of a same family [64].

The behaviour of the directional coupler is well explained by two theories. The *coupled mode theory* [65, 66] starts from the individual waveguides and calculates the light transfer between both. The *supermode analysis* [67, 68] starts from the structure as a whole and analy-

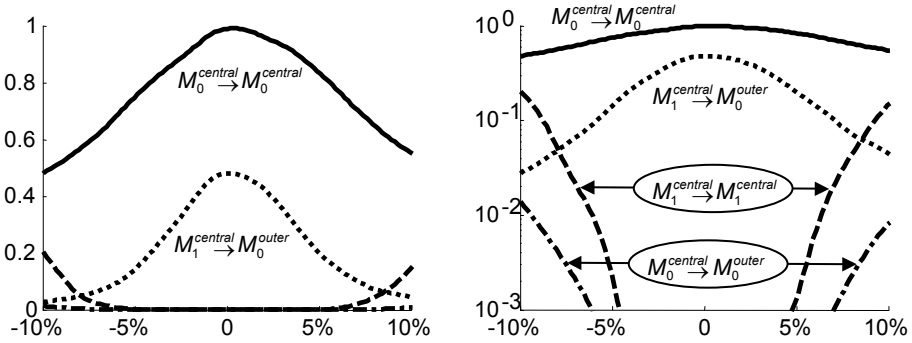


Figure 5.11: Mode transmission tolerances in the mode splitter designed for mask 2 (290 µm long and 9 µm broad) on a linear and a logarithmic scale). The transmission of the zeroth order input to the zeroth order mode of the central output is in solid line, and to the outer output is in dashed line. The transmission of the first order input to the outer output is in dotted line, and to the zeroth order of the central output is in dash-dotted line. The x-axis shows the relative deviation of the actual length from the designed value.

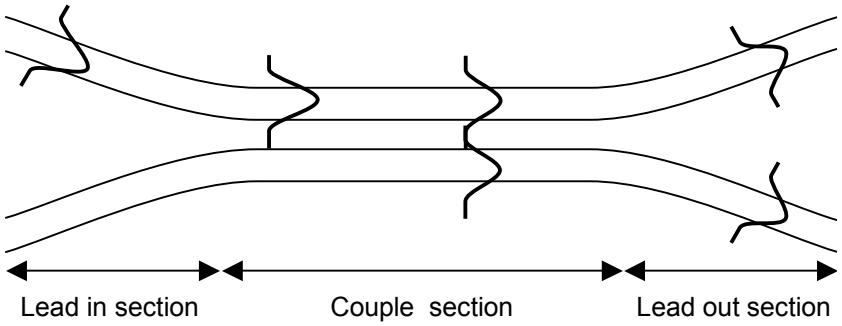


Figure 5.12: A directional coupler consist of two waveguides that join each other (lead-in section), run parallelly along each other in close proximity (couple section) and separate again (lead-out section). Within the couple section light couples from one waveguide to the other. The final coupling depends mainly on the mutual overlap of the individual waveguide modes, their propagation constants and the length of the coupling section.

ses the interference between so called supermodes. These two theories each give a different insight the dynamics of the directional coupler and will therefore briefly be outlined. Without loosing generality, all equations have been reduced to two dimensions for reasons of clarity.

5.2.1 Coupled mode theory

Assume a directional coupler consisting of two parallel monomodal waveguides A and B. The couple mode theory says that the optical field in the waveguide can still be written as a function of the eigenmodes of the individual waveguides $\phi_A(x)$ and $\phi_B(x)$. The mutual influence of the waveguides on each other is included in this theory by adding them as a perturbation on the excitation of the waveguide modes. In principle this theory is therefore only valid for waveguide which are weakly coupled.

The coupling parameter κ gives a measure of the mutual influence and depends on the overlap in between the two modes. If we define $\Delta = \beta_A - \beta_B$ as the difference in the propagation constant of the two modes. The length needed for maximal coupling and the maximal coupled power respectively are:

$$z = \frac{\pi}{\sqrt{4\kappa^2 + \Delta^2}} \tag{5.19}$$

$$P = \frac{4\kappa^2}{4\kappa^2 + \Delta^2}. \tag{5.20}$$

The coupling between the waveguide is complete if $\Delta = 0$, which means the propagation constants of the eigenmodes in the two waveguides should be equal. $\beta_A = \beta_B$. For weakly coupled waveguides κ is small, this means that from Eq. (5.20) follows that there is only coupling in between the waveguides if β_A and β_B are precisely matched. A closer look to Eq. (5.19) also reveals a second problem: If the propagation constants of the two modes are precisely matched, the length of the coupling section is inversely proportional to κ. As κ decreases exponentially with the distance between the waveguides, the coupling length increases exponentially. This means that the directional coupler is very intolerant to variations in the distance between the waveguides.

5.2.2 Supermode analysis

Although the coupled mode theory gives a good insight in the dynamics of the directional coupler, it is only an approximation and is not valid for strongly coupled waveguides. An exact theory of the directional coupler can be based on the eigenmodes of the total structure comprising both waveguides. For two monomodal waveguides the combined structure will contain two guided eigenmodes, which are

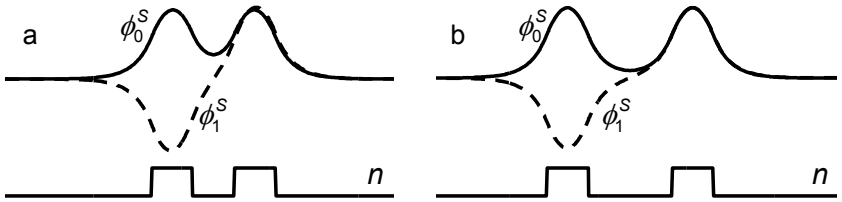


Figure 5.13: Supermodes in two identical monomodal waveguides. a. Strong coupling (small distance between the waveguides). b. Weak coupling (waveguides further apart).

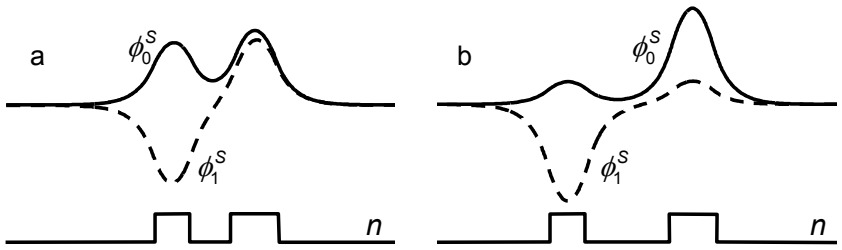


Figure 5.14: Supermodes in two non-identical monomodal waveguides. a. Strong coupling (small distance between the waveguides). b. Weak coupling (waveguides further apart).

called supermodes: $\phi_0^S(x)$ and $\phi_1^S(x)$. In absence of higher order modes the total field can be written as:

$$\psi(x,z) = U_0^S \phi_0^S(x) \exp\left(-j\beta_0^S z\right) + U_1^S \phi_1^S(x) \exp\left(-j\beta_1^S z\right). \qquad (5.21)$$

U_0^S and U_1^S are the excitation coefficients and β_0^S and β_1^S are the propagation constants of the respective supermodes. In general the two propagation constants are different and along the propagation axis there will be alternated destructive and constructive interference in the two waveguides. This interference pattern can be interpreted as a coupling from one waveguide to the other, as in the coupled mode theory. For weakly coupled identical waveguides the supermodes can be seen as the sum and the difference of the individual waveguide modes. When in phase or in anti-phase they form the respective eigenmodes of the single waveguides (Fig. 5.13b). The interference pattern can hence be interpreted as a 100% coupling from one waveguide to the other, as predicted by the coupled mode theory. For strongly coupled waveguide modes this is not any more entirely true (Fig. 5.13b) With an adiabatic taperin in the lead-in and lead-out section it is however still possible

to obtain a high coupling efficiency. In general the minimum crosstalk decreases with a larger component length [63].

Fig. 5.14 shows the supermodes for two different waveguides. The absolute value of the two supermodes are slightly different. This means only part of the light can be coupled from one waveguide to the other. For weakly coupled waveguides (Fig. 5.14b) this effect is much stronger than for strongly coupled waveguides (Fig. 5.14a). This was also predicted by Eq. (5.20).

5.2.3 Splitting off the second order mode

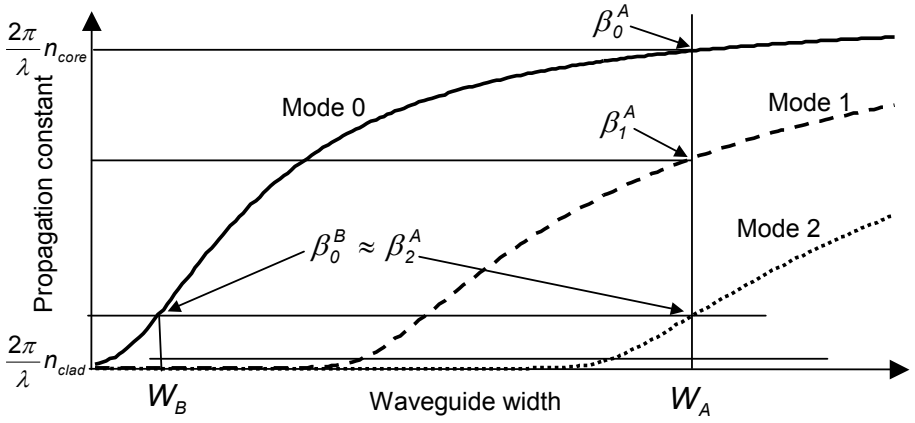


Figure 5.15: The propagation constant as a function of the waveguide width for different modes. The zeroth order mode is drawn with a solid line, the first order mode with a dashed line, and the second order mode with a dotted line.For a good coupling the widths of the waveguides are chosen such that the propagation constant of the zeroth order mode in waveguide B, β_0^B, equals the propagation constant of the second order mode in waveguide A, β_2^A.

Assume one chooses two waveguides, A and B, so that the propagation constant of the second order mode of waveguide A equals the propagation constant of the zeroth order mode of waveguide B. Then a directional coupler composed of those two waveguides will split of the second order mode of waveguide A. To align the two propagation constants one could for instance change the width of one of the waveguides. For an increasing width the waveguide will guide more modes, while the propagation constants of these waveguide modes will steadily rise from $\frac{2\pi}{\lambda} n_{clad}$ to $\frac{2\pi}{\lambda} n_{core}$ (Fig. 5.15). This means that for a width W_A of a trimodal waveguide A, one can find a monomodal

waveguide with width W_B so that $\beta_A^2 = \beta_B^0$, but $\beta_A^1 \neq \beta_B^0$ and $\beta_A^0 \neq \beta_B^0$. Such a structure will have four supermodes (Fig. 5.16). Two of them are the zeroth and first order mode of waveguide A. The other two supermodes are a combination of M_A^2 and M_B^0. As shown in the previous

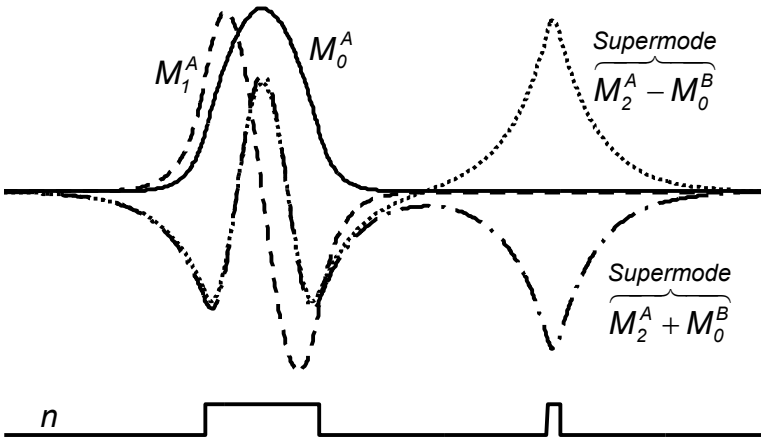


Figure 5.16: Supermodes for a trimodal and a monomodal waveguide. The second order mode of the broad waveguide and the zeroth order mode of the small waveguide have the same propagation constant and form two super modes. The other supermodes are identical to the zeroth and first order mode of the broad waveguide.

section, a directional coupler is very sensitive to small differences between the propagation constants of the two waveguide modes. Fig. 5.17 shows the supermodes for a directional coupler where the width of the small waveguide has increased 1% compared to Fig. 5.16. As the absolute values of the supermodes do now overlap much less, the coupling from M_A^2 to M_B^0 will drop heavily. This problem together with the fact that the coupling length heavily depends on the distance in between the waveguides makes it very difficult to fabricate a directional coupler that fully filters out the second order mode from a trimodal waveguide. As explained in chapter 3 splitting off additionally the second order mode does not add much extra in terms of resolution or noise sensitivity reduction. Therefore the directional coupler has been omitted from the design.

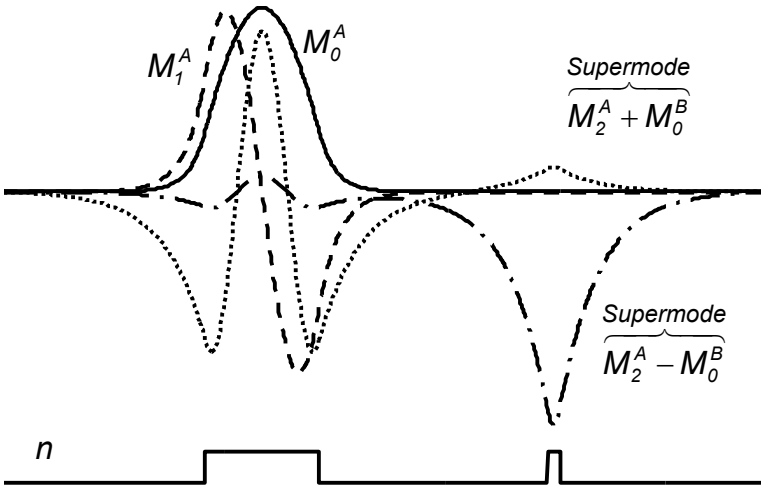


Figure 5.17: Supermodes for a trimodal and a monomodal waveguide. The monomodal waveguide is made 1% wider than would be needed to make equal the propagation constants of the highest modes in both waveguides.

5.3 Optical interferometer

As described in section 3.4.6 one could use two single mode waveguides, instead of a single bimodal waveguide, to pick up the light from the disc. In collaboration with the Optics Department at Delft University of Technology, an optical interferometer was designed that can measure the phase between the two waveguides. The design is based on the phase diversity network developed by Soldano [58].

If we omit the common phase factor, the excitation in both waveguides can be written as A and $B\exp(j\varphi)$. Each of the signals is led to a different 2×2 MMI. Each of these MMIs result in two signals, which have an equal amplitude and a mutual phase difference $\pi/2$. As shown on Fig. 5.18, the four resulting signals are cross-connected to two new 2×2 MMIs. The optical path length of the three of the four signals is equal, the fourth has a phase shift of $\pi/2$. The resulting amplitude and power at the outputs are then:

$$o_1 = -A + jB\exp(j\varphi) \quad O_1 = |o_1|^2 = A^2 + B^2 + 2AB\sin(\varphi) \quad (5.22)$$

$$o_2 = -jA + B\exp(j\varphi) \quad O_2 = |o_2|^2 = A^2 + B^2 - 2AB\sin(\varphi) \quad (5.23)$$

$$o_3 = A - B\exp(j\varphi) \quad O_3 = |o_3|^2 = A^2 + B^2 - 2AB\cos(\varphi) \quad (5.24)$$

$$o_4 = jA + jB\exp(j\varphi) \quad O_4 = |o_4|^2 = A^2 + B^2 + 2AB\cos(\varphi) \quad (5.25)$$

By some basic algebra one can solve these four equations and find

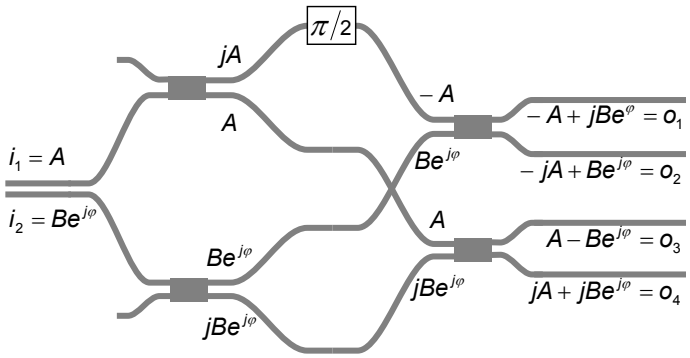


Figure 5.18: The optical interferometer has two signals as input, A and $B\exp(j\varphi)$, and as output four signals, from which the phase difference in between the input signals can be calculated.

closed forms for A, B, and φ:

$$\varphi = \arctan\left(\frac{O_1 - O_2}{O_4 - O_3}\right) \tag{5.26}$$

$$A \text{ and } B = \sqrt{\frac{p \pm \sqrt{p^2 - 4q}}{2}} \tag{5.27}$$

$$\text{with } \begin{cases} p = A^2 + B^2 & = \frac{O_1+O_2+O_3+O_4}{4} \\ q = A^2B^2 & = \frac{(O_1-O_2)^2+(O_4-O_3)^2}{16}. \end{cases} \tag{5.28}$$

5.4 Design

We designed two types of components to check how well a multimodal waveguide pick-up head can resolve small features on a disc. A first mask was made primarily to check experimentally the performance of the mode splitters. With this first mask we aimed at components picking up the reflection from an externally illuminated pattern with a bimodal waveguide. With the second mask we aimed at components that illuminate a pattern and pick up the reflected field with the same multimodal waveguide. This prototype should hence be able to show the full potential of the waveguide detection system. The operation wavelength in a CD, DVD, or blu-ray system is respectively 780 nm, 650 nm, and 405 nm. For reasons of technological feasibility within the time frame of this research we choose however to aim at a proof-of-principle

with components designed for λ=980 nm. For an extension to 450 nm, one should use a material system which is transparent for these smaller wavelengths, such as gallium nitride (GaN) or silica-on-silicon.

5.4.1 Layer structure

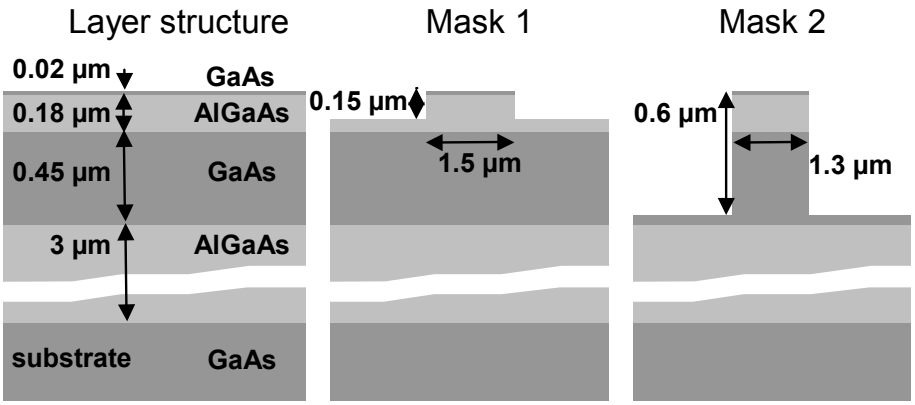


Figure 5.19: The waveguide chip consists of a layer structure of gallium arsenide (GaAs, dark gray layers) and aluminium gallium arsenide (AlGaAs, light gray layers). The vertical and horizontal scale are not identical. The etch depth of mask 1 and mask 2 is respectively 200 nm and 600 nm. The picture shows the smallest waveguide on each mask.

The waveguide chip consists of a layered structure on top of a gallium arsenide (GaAs) substrate. The core layer is made of GaAs, which has a refractive index of n=3.52 for λ=980 nm. The top and bottom cladding are aluminium gallium arsenide ($Al_{0.4}Ga_{0.6}As$) with a refractive index of n=3.31 for λ=980 nm. The waveguide structures are formed by etching away the material around the ridges. For the first mask we etched 200 nm into the core layer. For the second mask we etched 600 nm deep, nearly completely through the core layer. The reason for this deeper etch is to eliminate the stray light that is guided by the slab modes. A graph of the layer structure is shown on Fig. 5.19. The layer structure is an adaptation of a standard layer structure from photonic ICs for telecom applications. Because of the difference in wavelength and refractive indices, the thickness of the layers has been scaled down.

The etch depth of the waveguides, depends on how much modes a waveguide with a given width can guide. The lithographic fabrication process used for the components does not allow waveguides smaller

than $1.3 - 1.5$ µm. For mask 1 the wafer was etched such that the first order mode of a waveguide of 1.5 µm wide is in cut-off, and hence monomodal. For mask 2, the wafer is etched such that a waveguide of 1.3 µm only guides two modes.

5.4.2 Mask 1

The mask layout consists of several basic cells. A schematic view of such a cell is shown on Fig. 5.20. To test the performance of the mode splitter, two MMIs have been placed on the mask. The first MMI is used as a mode combiner, the second as a mode splitter. Both are connected with a central bimodal waveguide. The in- and output waveguides are monomodal. To accommodate for the mode mismatch between the waveguide mode and the fibre mode, a horizontal taper was included at the end of the in- and output waveguide.

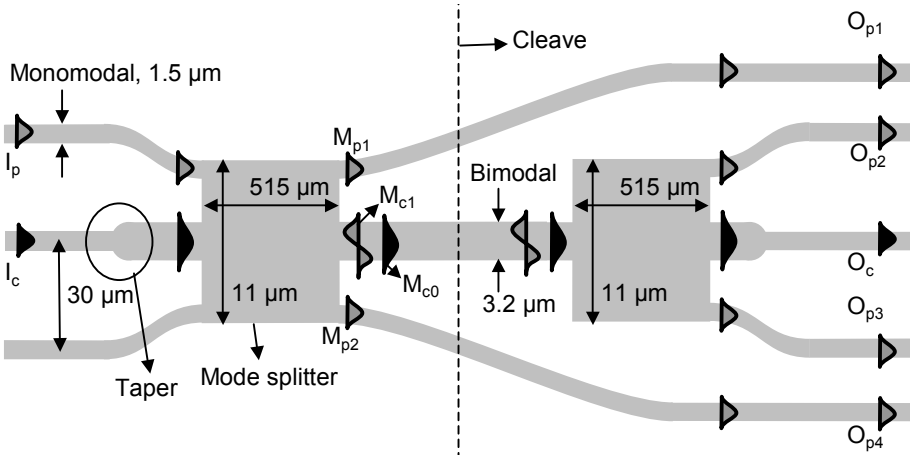


Figure 5.20: Schematic view of a basic cell on mask 1 (not to scale). Two MMIs are connected with a bimodal waveguide, three input waveguides and five output waveguides.

Assume light is introduced in one of the outer input waveguides, I_p, with an amplitude A. As described in section 5.1.2 the mode splitter will then form a first order mode in the central waveguide with amplitude $A/\sqrt{2}$ and twice a zeroth order mode in waveguides M_{p1} and M_{p2}, with amplitude $A/2$. The first order mode is guided towards the second MMI, where it forms two zeroth order modes of amplitude $A/2$. In brief, an excitation of one of the outer input waveguide leads to an

	$I_p = A, I_c = 0$	$I_p = 0, I_c = A$
$O_{p1} = O_{p4}$	$A/2$	β
$O_{p2} = O_{p3}$	$A/2$	β
O_c	β	A

Table 5.3: Excitations of the output waveguides as a function of the excitation of the input waveguide. The crosstalk has been included in a first order approximation.

equal excitation of the zeroth order mode in each of the four outer output waveguides with half the amplitude of the input mode. Similarly an excitation of the central input waveguide results in a zeroth order mode in the central output mode, equally in amplitude to that of the input mode.

Because of small errors in refractive index, etch depth, or dimensions, the mode splitters may not function optimally. If we assume the structure remains fully symmetric two types of crosstalk may show up. The first type will be called α and is the part of the first order mode in the central waveguide that excites a first order mode in the central waveguide at the other side. The second type of crosstalk will be called β and is the part of the zeroth order mode in the central waveguide that excites the outer waveguides. Because of reciprocity this is equal to the crosstalk from an outer input to the central zeroth order mode. The first type of crosstalk is not very important as it can easily be filtered out by tapering the bimodal waveguide down to a monomodal waveguide. The second type can however not be filtered out. Table 5.3 shows the excitations of the output waveguides as a function of the excitation of the input waveguides. By comparing the outputs of the different waveguides one can measure directly the crosstalk within the mode splitter.

In a next stage the chip can be cleaved into two parts along the dashed line in Fig. 5.20. Each of these two parts consists now of a bimodal waveguide with a waveguide-air facet on one side, and a connection through a mode splitter to output waveguides on the other side of the chip.

To accommodate for small fabrication errors the same basic cell has been reproduced several times, with a small variation in one of the parameters. The offset on the MMI width was varied from -0.3 µm to 0.3 µm in steps of 0.1 µm. For each of these widths two basic cells are placed on mask 1. Resulting in 28 cells. These 28 cells were reproduced five times varying the width offset of all structures, i.e. the width of

both the waveguides and the MMIs (from -0.2 µm to 0.2 µm in steps of 0.1 µm). This means that the width of the MMIs is varied twice.

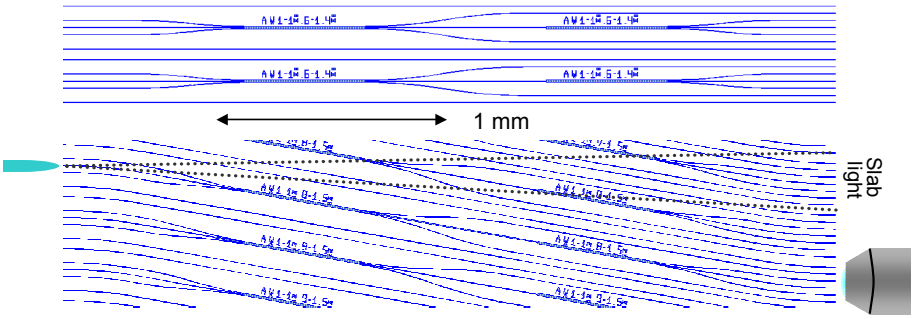


Figure 5.21: Layout of mask 1: on top a detail of two components laying straight on the bottom a detail of a component placed tilted on the mask.

Because of the low etch depth it is very difficult not to excite some of the slab modes, when introducing light in one of the input waveguides. This slab light might interfere with the light in the output waveguides. Therefore one range of components was placed tilted with respect to the input and output waveguides, as shown on Fig. 5.21. The idea is that the parasitical slab light goes straight on while the light inserted in the waveguides is curved to the right.

5.4.3 Mask 2

The mask layout of mask 2 also consists of a number of basic cells. This basic cell is shown on Fig. 5.22. The main difference is the supplementary 3dB splitter, which is needed to enable a fibre interface to both the illumination and for the detection of the zeroth order mode. Because of the deeper etch depth, it is not possible to make the waveguides small enough to cut off the first order mode. Therefore a mode stripper was added at the input and output waveguides. This mode stripper is not necessary at the waveguides connected to the 3dB splitter as this last component strips already the first order mode by itself. The central waveguide connected to the mode splitter should be twice as broad as the outer waveguides, and hence guides four modes. An adiabatic taper was used to reduce the width to that of a bimodal waveguide.

Similarly to the components on mask 1, the crosstalk in the mode splitters can be tested by focusing light into the central or the outer input waveguides and by checking the excitations of the output waveguides. Additionally there might however be some side effects due to

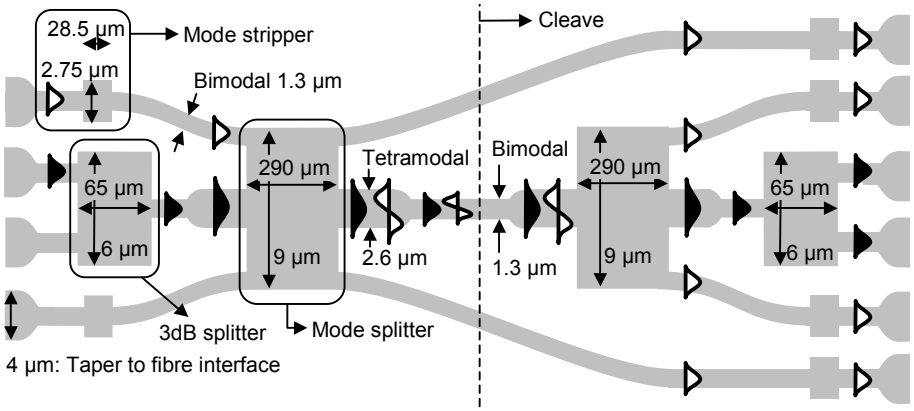


Figure 5.22: Schematic view of a basic cell on mask 2 (not to scale). Two MMIs are connected with a central bimodal waveguide, four input waveguides and six output waveguides.

malfunctioning of the mode stripper and 3dB splitter. These effects are however not very important. This can be shown as follows. Assume that due to a small misalignment the first order mode is excited in one or more of the input waveguides. Most of the first order mode that would thus be introduced in one of the outer input waveguides will be blocked by the mode stripper, the remainder will be imaged onto a third order mode in the tetramodal waveguide, and onto a first order mode in the outer output waveguides. The third order mode will however be radiated away because of the tapering down to a bimodal waveguide, and the first order mode in the outer waveguide will pass a second mode stripper. If a first order mode would however be excited in the central input waveguide, most of it will be blocked by the 3dB splitter, the remainder will be image onto the first order mode. This first order mode is however blocked by the mode splitter. We can conclude that crosstalk in de mode stripper or 3dB splitter are second order phenomena, and that the bottleneck remains the crosstalk of the mode splitter.

Other features To test the power loss of the zeroth order mode in the mode stripper and the 3dB splitter, a few calibration features were added to the mask: broad straight waveguides, narrow straight waveguides, narrow straight waveguides along with 8 consecutive mode strippers, and narrow straight waveguides along with 6 consecutive mode splitters. On the same mask we also included a few optical inter-

ferometers, designed by Sami Musa (Delft University of Technology). These optical interferometers were designed for a shallower etch depth, but could nevertheless be placed on the same mask.

5.4.4 Anti-reflective coating

To reduce the reflections and losses at the waveguide-air and at the waveguide-fibre interfaces we added an anti-reflection (AR) coating at the facets of the components. With an AR-coating the reflection of one mode can be minimized to a small fraction [69]. Unfortunately it is not possible to eliminate the loss to radiative modes. Moreover the optimal thickness and refractive index is different for each mode. This can be understood as follows: the ideal coating for a plane wave with a propagation perpendicular to the interface, has a refractive index n_c, which is the square root of the refractive indices in the two media: $n_c = \sqrt{n_1 n_2}$. The thickness of the coating is $\lambda/(4n_c)$. For an oblique plane wave the optimal coating refractive index and thickness are however different [69].

A waveguide mode is a small spot and can be seen as a composition of different plane waves. This is composition is different for the zeroth and the first order mode, and hence also the optimal coating. Fortunately the eventual shape of the mode outside the waveguide is not very important as has been shown in section 3.4.2. For our samples we used a simple one-layer coating. For the waveguide-fibre interface we used TiO_2 ($n = 2.62$) and for the waveguide-air interface we used HfO_2 ($n = 1.88$).

5.5 Fibre interface

The components on mask 2 were designed to include a fibre interface for inserting and extracting the light from the input and output waveguides (Fig. 5.23).

To connect the fibres with the input and output waveguides on the chip, two methods have been investigated. One method is aligning and gluing each fibre separately to the chip. The advantage of this method is that each fibre coupling can be individually optimized. An important drawback on the other hand is that for individual alignment, the distance between the fibre connections needs to be relatively large. A second method is first lining the fibres up as a fibre array into V-grooves. This fibre array is then attached to the waveguide chip. This way the

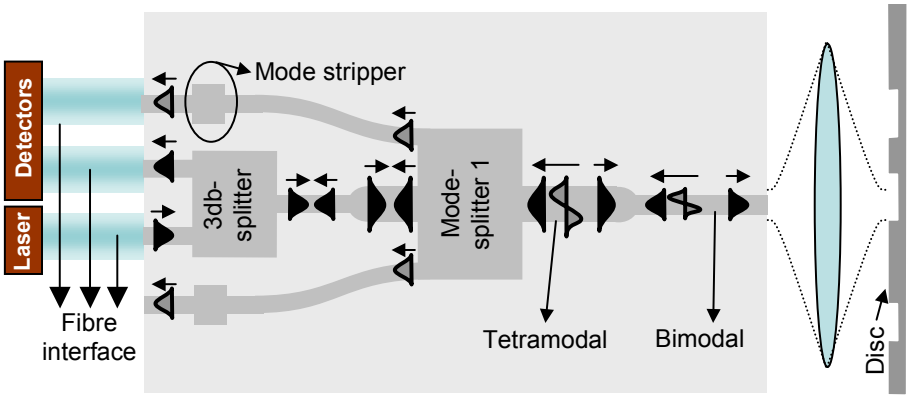


Figure 5.23: A schematic view of the optical light path in the components of mask 2 including the fibre interface to the laser and detector.

fibres can be packed much closer to each other. The alignment is however more difficult.

5.5.1 High NA fibre

A standard optical fibre has a mode diameter of 9 µm. This is much larger than that of the waveguide modes, which have a diameter of approximately 1 µm. To reduce the losses at the fibre-chip interface because of difference in modal spot size, a high NA fibre has been used. This fibre, made by Thorlabs, has a mode diameter of 4 µm, which is more than twice as small as that of a standard optical fibre.

As the lasers and detectors are connected by standard telecom fibre, an additional splice is needed to join the high NA fibres with the standard fibre. Because the modes of the two fibres are very different there will inevitably be some losses at the interface. One can however reduce these losses by forming a tapered transition at the interface. This can be done by extending the duration of the electric arc used for the splice. During this electric arc the core of the High NA fibre diffuses into the cladding. This diffusion results in an adiabatic tapered coupling between the telecom fibre and the High NA fibre [70].

In our setup 980 nm light has been used. Both the standard telecom fibre and the high NA fibre are however optimized for 1550 nm. This misfit in wavelength has two major consequences: a higher radiation loss, and guiding of higher order modes. As the fibre lengths in our setup are however relatively short the losses are negligible and mode

coupling from the ground mode to the higher order modes will be limited. Moreover we prevented coupling from the higher order modes of the fibre to and from the higher order modes of the input and output waveguides, by stripping these waveguide modes.

We found experimentally a loss of 10% loss per splice for 1550 nm light and 20% loss per splice for 980 nm light.

5.5.2 Individual fibre coupling

As explained above, one way of connecting the chip with three fibres is to align each fibre individually to the waveguide chip. Using a piezoelectric controlled translation stage, a fibre is aligned with an input waveguide on the chip. The alignment is done by maximizing the light throughput at the other end of the chip. By gluing the fibre to the chip, the alignment is fixated. This last step poses however some problems. Fig. 5.24 shows a fibre glued to a sample chip with parallel waveguides. The connection between the fibre and the chip proves to be solid enough, but because the connection is so rigid, the fibre itself tends to break very rapidly. This is especially so, because the plastic outer cladding is removed. Therefore the fibre ends need to be reinforced in addition. A simple solution would be to add some supplementary drops of glue. But, because of the cohesion and adhesion forces a drop of glue tends to spread out all over the waveguide chip leaving nothing at the fibre(Fig. 5.25a).

A solution to this problem is first placing and hardening a drop on the fibre a few millimetres away from the fibre facet (Fig. 5.25b). Finally one can connect this drop with the waveguide chip by dropping some glue in between these hardened drops. The the fluid glue is now pulled from two sides: the already hardened drop at the fibre, and the waveguide chip. This ensures that the glue is relatively well spread over the fibre (Fig. 5.25c). This way the fibre is more solidly fixated to the chip. The operation can be repeated to attach more fibres. Because of the diameter of the glue drops the minimal spacing between the fibres is however of the order of 1 mm.

If one would want to glue three fibres individually to the chip, the width of one component on the chip would need to be at least 3-4 mm. This means that only two or three components can be placed on a single chip, which is $\approx 1\mathrm{cm}^2$. To be able to place the components closer together, we used a trick: curving the output waveguides in such a way that the individual fibres can be attached to three opposite side of the

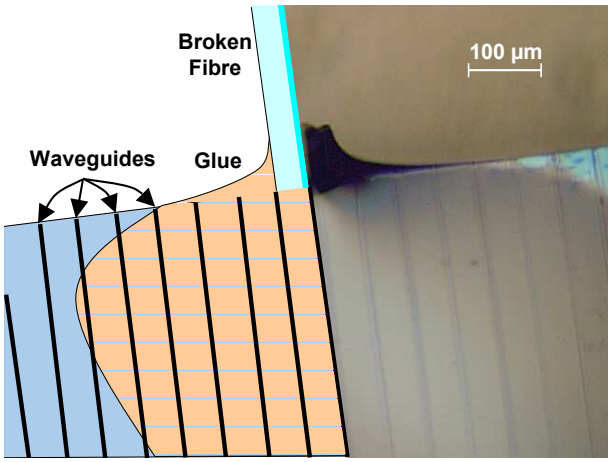


Figure 5.24: A photograph of an individual butt coupling of a fibre to the chip. Without reinforcement the fibres tends to break easily. The left part of shows a schematic drawing to show more clearly the waveguides, glue and fibre.

chip. The remaining fourth side contains the facet of the bimodal waveguide picking up the light from the disc. The new bottleneck is now the minimum distance between two cleaves, which is approximately 2 mm. To put the components close together, we placed two components in between two cleaves (Fig. 5.26). The idea is that one first measures the crosstalk of both components before gluing, and then glues the fibres to the in- and output waveguide of the component with the lowest crosstalk.

5.5.3 Lining up the fibres into V-grooves

A second technique to connect the fibres with the in- and output waveguides is first aligning the fibres into an array of V-grooves and then aligning this array as a whole with the chip.

V-grooves The V-grooves are triangularly shaped trenches etched in a silicon sample. Selectively wet etching a rectangular opening in a mask will result in a triangular etch at an angle of $55°$. The spacing between the V-grooves is 250 µm.

Alignment of fibres into V-grooves The fibres are manually aligned into the V-grooves using tweezers and a microscope. Each time a fibre

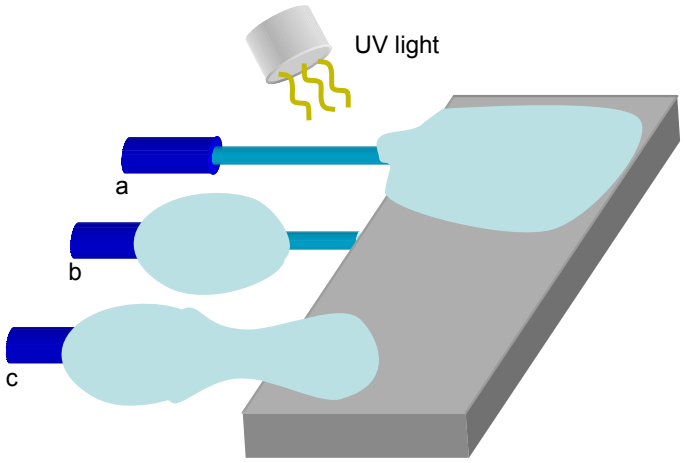


Figure 5.25: Because of the cohesive forces, a drop of glue tends to flow all over the chip, leaving little glue at the fibre-chip interface (a). By first uv-curing a drop a millimetre away from the chip (b), a bridge is created between which a new glue drop sticks (c). This combination is much less fragile then without the reinforcements.

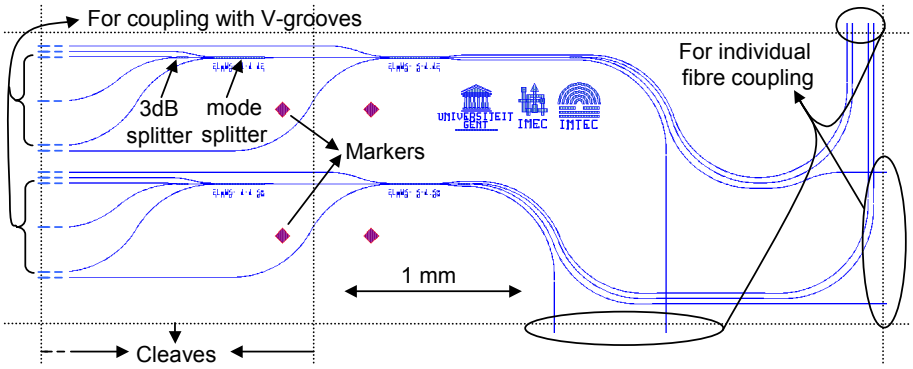


Figure 5.26: A schematic view of the layout of mask 2. The components on the left are coupled with a fibre array in V-grooves. The components on the right are designed to be individually coupled with three fibres.

is placed into the grooves a weight is placed on top of it, to hold the fibre in place. This way the three fibres can be placed into the grooves one by one. To make the alignment permanently the fibres are glued between the V-grooves and a small glass plate, which is meanwhile pressed onto the fibres. The glue is hardened with a UV-source illuminating through

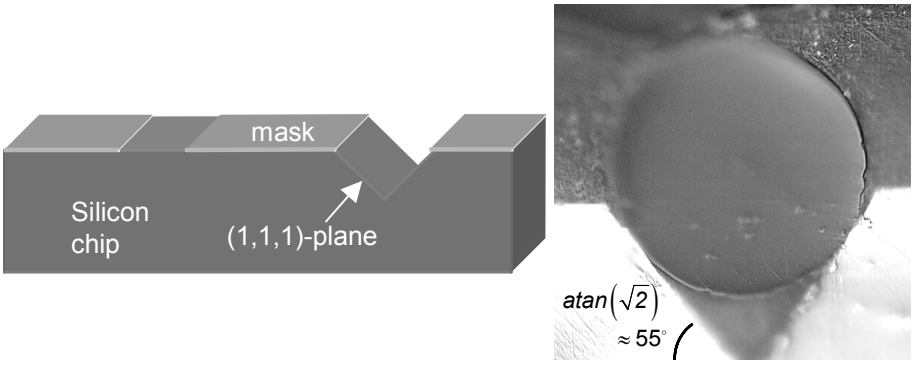


Figure 5.27: Selective etch of a (1,0,0) silicon chip results in V-grooves along the (1,1,1) plane. On the right a photograph of a fibre aligned in a V-groove. On top of the left half a schematic drawing has been added to make the size of the fibre core visible.

the small glass plate, as shown on Fig. 5.28. To reinforce the fragile fibres the whole sample is glued on top of a larger glass plate.

Polishing The lateral alignment of the fibres is very accurate, because of the V-grooves. The longitudinal alignment of the fibres along these grooves is however much less accurate. Therefore the facet with the fibre ends is polished, until the surface is clean and flat. On the right side of Fig. 5.27 the polished facet with one fibre is shown.

Aligning the fibre array with the chip To align the whole component with V-grooves and fibres with the waveguide chip one needs to control the three translational movements and the three rotational movements by micro screws, while maximizing the throughput through each waveguide. By a UV-curable glue the alignment can be made permanently.

5.6 Future extensions

A first extension would be to include other optical functions on the waveguide chip. In principle one could further integrate a laser, detectors and tracking probes onto a single chip, thereby reducing the size of the pick-up head by an order of magnitude an eventually leading to a much more compact optical disc drive.

A second extension would be to use a waveguide which is multimodal horizontally as well as vertically. This would lead to four differ-

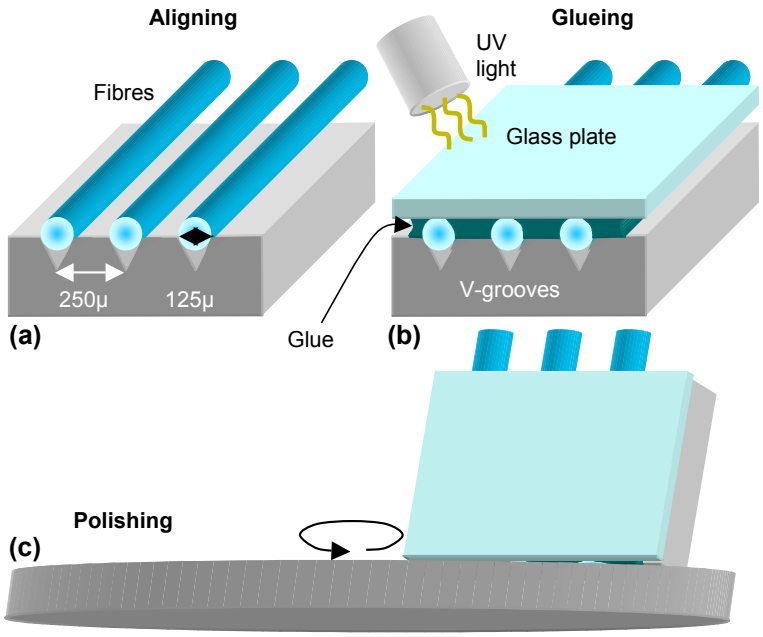


Figure 5.28: The fibres are aligned into V-grooves. Through a glass plate, which presses the fibres into the grooves, glue is hardened to make the alignment permanent. Finally the facets are polished to obtain a clean and flat surface.

ent signals. The low frequency content could be used for tracking and the high frequency content to improve the bit error rate of the zeroth order mode signal.

With the fabrication technology used for the current samples it is rather difficult to make complicated structures in the dimension perpendicular to the waveguide chip. Possible techniques to do so would be wafer-bonding [71, 72, 73], Selective area growth (SAG), polymer technology [74], implantation of oxygen in Silicon (SIMOX[1]) [75], or regrowth. From a theoretical point of view some problems arise as the polarization into the x- and y-dimension will be different. This means that a perfect constructive and destructive interference pattern will occur at a different length in the vertical and horizontal direction [76].

[1]Separation by IMplantation of OXygen

Chapter 6

Measurements

In this chapter we will outline how the components made from mask 1 and mask 2 were tested. In the first section will be described how the isolated chips were measured. From these measurements the mode profiles and crosstalk in the components can be derived. In the second section is outlined how well the multimodal waveguide can discriminate small features on an externally illuminated mask. Finally the measurements are shown, for the multimodal waveguide used both as the illumination source and as the detector.

6.1 Measurements on isolated chips

6.1.1 Setup

Fig. 6.1 shows a schematic view of the measurement setup at the INTEC laboratory at the Ghent University. Basically the setup contains three stages. Stage 1 controls the device that couples the light into the chip, this can either be an objective lens, a lensed fibre, or a butt coupled cleaved fibre. This stage is controlled along the three translational and the three rotational axes. The waveguide chip is placed on stage 2. On stage 3 an objective lens is mounted to couple the light out of the waveguide chip. The lenses on the right, image the light spot onto a camera. The pinhole can be aligned such that only the light of one output waveguide can be isolated. A detector (not shown on the figure) placed just behind the pinhole, can then detect the waveguide mode without interference from other light sources.

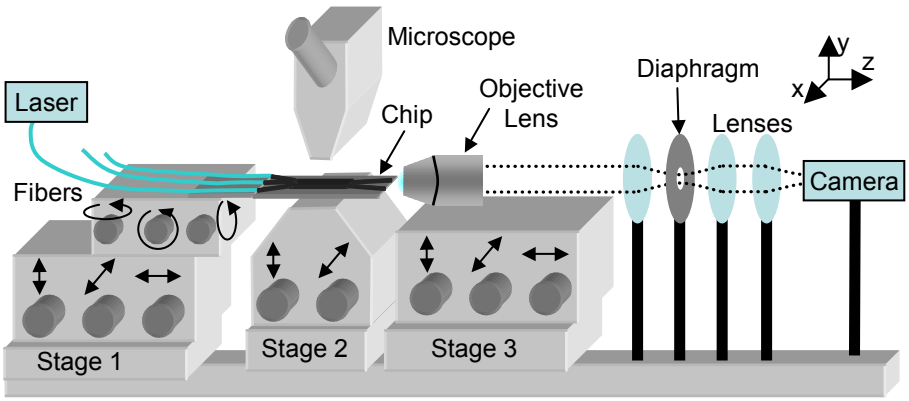


Figure 6.1: Measurement setup for testing the components, as used for aligning the fibre array to the chip. A fibre array (on stage 1) is aligned with a sample (on stage 2), the light is coupled out by an objective lens (stage 3). The light is further guided through a pinhole to the detector.

The laser source for the experiments was a LC91A-20 Fabry-Perot laser from Nortel, which was designed as a pump laser for EDFAs[1], but could be used for this application. It has however the disadvantage of a relatively broad spectral width: 975 nm-980 nm.

6.1.2 Components mask 1

The components on mask 1 are etched 0.15 µm as shown on Fig. 5.19. The MMI is 515 µm long by 11 µm broad, as shown on Fig. 5.20. This component is placed a few times on the mask with a variation on the width of the MMIs and the waveguides.

Crosstalk As described in section 5.4.2 the crosstalk within the mode splitter can be measured by combining two identical mode splitters into one component. By inserting the light in one of the input waveguides, the crosstalk is measured by the ratio in between the power in the output waveguides. This measurement can be repeated for each of the input waveguides. Fig. 6.2 shows the crosstalk for each measurement and the average of the measured crosstalk for each component. Because some of the components were scratched and could not be measured, some values are missing. Both the graph for the straight components as for the tilted components show an increasing crosstalk for broader

[1] erbium doped fibre amplifiers

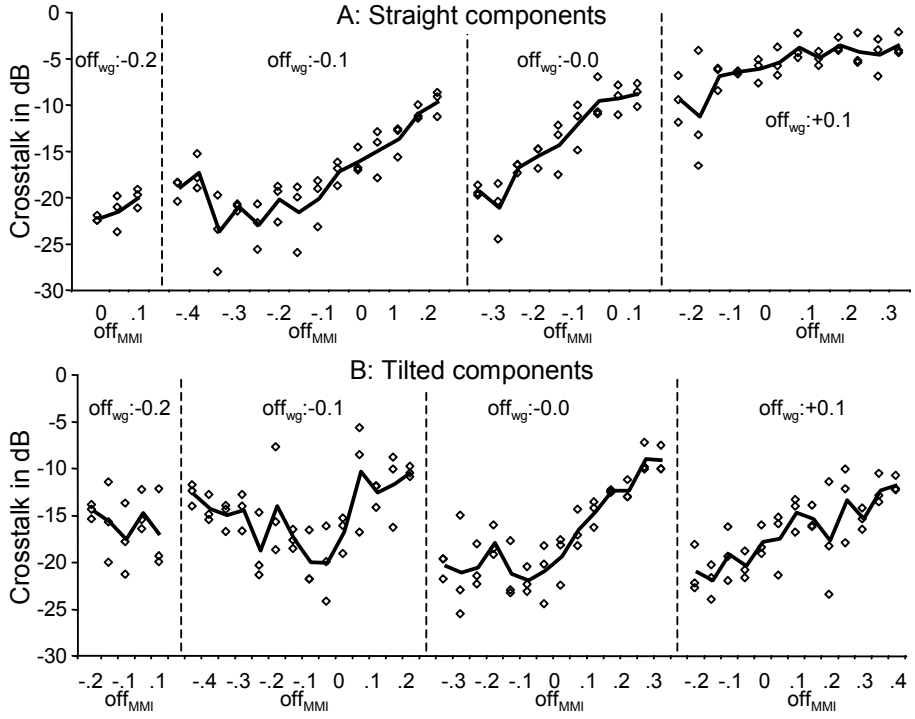


Figure 6.2: Crosstalk in the mode splitters on mask 1: ratio of excitation of the outer output waveguides by the zeroth order mode in the central input waveguide over that of the same output by the first order mode. On top the straight components, at the bottom the tilted components. The diamonds indicate the crosstalk, calculated from the ratio of the output waveguide excitations for the illumination of one of the input waveguides. The solid line gives an average of these three measurements. off_{WG} is the offset on the width of the waveguides with respect to the designed width. off_{MMI} is the offset on the MMI width.

MMIs (larger off_{MMI}). This suggests that the fabricated components were broader than expected from the design. Or another possibility: the waveguides are etched slightly to deep. Nevertheless the best components showed a crosstalk level down to less than -22 dB.

Lensed fibre scanning An alternative method to check how well the mode splitters work is by scanning a lensed fibre in front of the bimodal waveguide. When the lensed fibre is moved laterally in front of the multimodal waveguide, the excitation of the different modes changes proportionally with the overlap integral in between the wave-

guide mode profile and the spot of the lensed fibre. If the lensed fibre is situated exactly in the middle of the input waveguide, only the zeroth order mode will be excited, otherwise both modes will be excited. Using the mode splitter the excitations of these two modes can be measured at the output waveguides. We used the component with the lowest crosstalk (off_{MMI}=-0.3 µm, and off_{wg}=-0.1 µm). The resulting mode profiles on Fig. 6.3a_2 are in relatively good agreement with expected mode profiles. Output O_{p4} shows however some artifacts on the left, which may be due to some coupling of parasitical slab light into that output waveguide.

We performed the same experiment on the total component, which is shown on Fig. 6.3b_1. The output connected to the zeroth order mode shows again a nice waveguide mode profile. We see however that at the outer output waveguides a first order mode profile is visible. This is unexpected, as we designed the input and output waveguide in such a way that the first order mode is in cut-off. One possibility for the output at the outer output waveguides is that the input waveguides have become bimodal because they are broader than designed. Another cause might however be that the slab light excited by the lensed fibre, couples into the taper in between the facet and the MMI.

6.1.3 Components mask 2

The components on mask 2 are etched 0.6 µm as shown on Fig. 5.19. The dimensions of the MMIs are shown on Fig. 5.22. This component is placed a few times on the mask with a variation on the width of the MMIs.

Calibration structures As described in section 5.4.3, we included a few calibration structures on mask 2, to give an additional check on the performance of the fabricated components. The light throughput of a series of 8 concatenated mode strippers and of a series of 6 concatenated mode splitters is compared with the light throughput of a simple waveguide. Fig. 6.5 shows a comparison of the simulation results and the measurements. The measurements can be brought into agreement with the simulations, if one assumes that the real structures are 0.3 µm broader than designed. Fig. 6.5 shows a photograph taken by an scanning electron microscope (SEM) that confirms this assumption. It shows that the waveguide facet, which was designed to be 4 µm wide, is actually 4.4 µm wide.

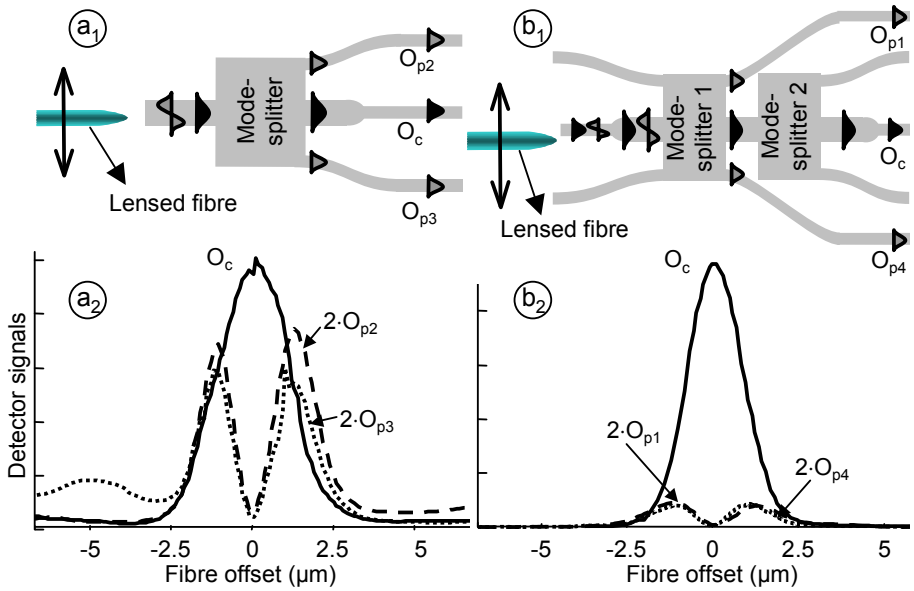


Figure 6.3: When a lensed fibre is moved laterally in front of the input waveguide, the modes of that waveguide are excited according to the mode profile. With the mode splitter these excitation can be measured. On the left the results for scanning in front of the central bimodal waveguide are shown. On the right the same experiment is done on the whole component. This input waveguide is not designed to be bimodal, but this experiment suggests that it is nevertheless the case.

Crosstalk As described in section 5.4.3 the crosstalk in the mode splitters of mask 2 can be measured by comparing the ratio in the output waveguides excitations, for a given input waveguide excitation. Each components has four input waveguides and leads thus to four measurements. Fig. 6.6 shows the four individual crosstalk measurements with diamonds and an average for each component by a black line. The crosstalk is lowest for the smallest components, which could be expected as the measurements of the calibration structures already showed that the fabricated structures were broader than designed. But even for components with the smallest width, the crosstalk is considerably higher than in the components of the first mask and goes not below -17 dB.

Lensed fibre scanning Fig. 6.7 shows the output of the outer output waveguide when a lensed fibre is scanned in front of the bimodal wave-

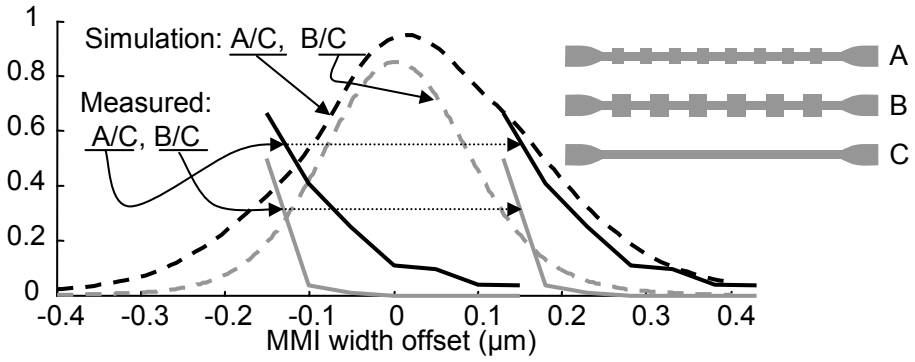


Figure 6.4: Measurements and simulations of the calibration structures on mask 2. The mismatch between simulation and experiment can be solved by assuming that the fabricated structures were 0.3 µm broader than designed.

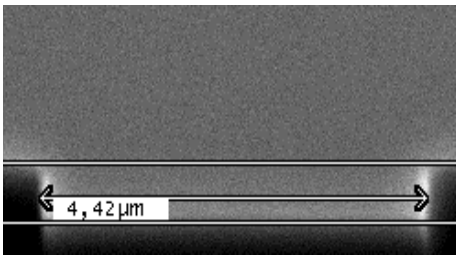


Figure 6.5: Photograph of the waveguide facet by a scanning electron microscope (SEM). The designed width is 4 µm. The photo shows that the fabricated with is 0.4 µm larger.

guide. A first order mode profile is visible, but compared to the result of mask 1 in Fig. 6.3, the signal is much more distorted. This may be partially due to the higher crosstalk, but probably some light from the lensed fibre is also coupled to the tetramodal section of the waveguide further on, or directly into the MMI, explaining the large tail on the right of Fig. 6.7.

6.2 Resolving small features

6.2.1 Setup

To check how well the waveguide can discriminate small features, a mask with small features has been designed (Fig. 6.8). These features include:

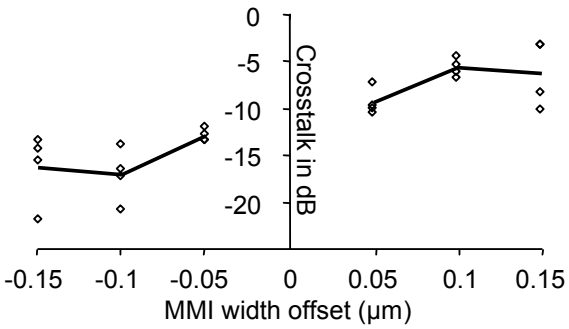


Figure 6.6: Crosstalk in the components on mask 2. The diamonds give the crosstalk, measured by illuminating one input waveguide. The solid line gives an average of the four different measurements. The component with offset zero was broken and could not be measured.

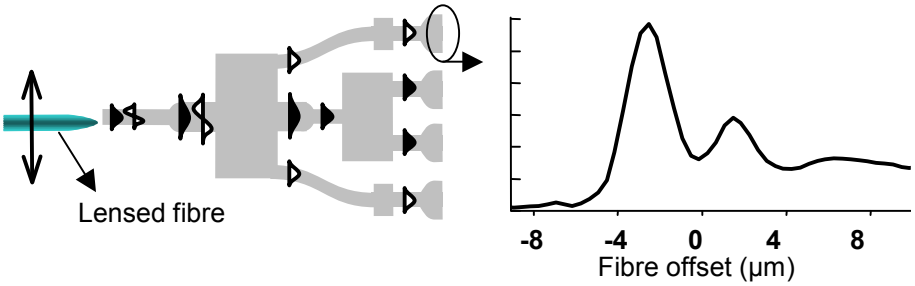


Figure 6.7: The light picked up by the first order mode from a moving lensed fibre. The graph suggests that there is crosstalk from the zeroth order mode, which interferes with the first order mode.

- The big square on the top and the large rectangles are alignment markers to position the mask with respect to the laser spot and waveguide facet.
- On the first line below the square there are 20 different gratings. The duty cycle is $1/2$ and the grating period varies from 10λ to $\lambda/2$.
- The second line contains different line pairs. The width of the lines varies from $\lambda/2$ to $\lambda/4$ and the line distance varies from 4λ to $\lambda/4$.
- The bottom line contains all possible variations of five bits. To save space, bit patterns which can be made identical by translation or mirroring are omitted. This means there are 13 combina-

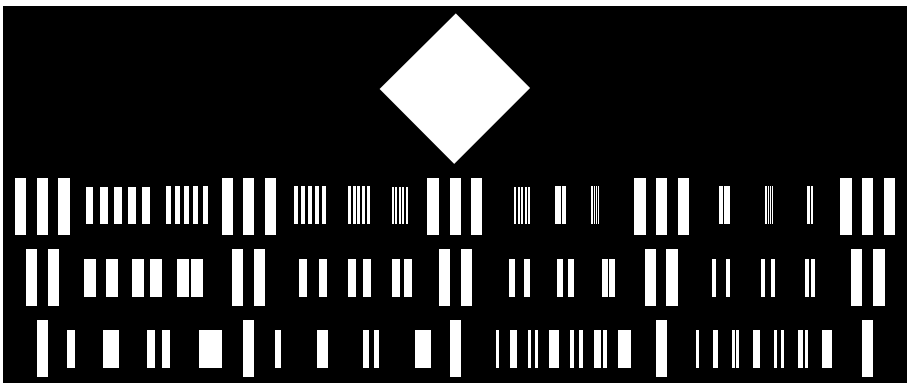


Figure 6.8: A schematic view of the mask with small features. For reasons of clarity, the graph is not drawn to scale and not all features are included in the graph. The big square on the top and the large rectangles are alignment markers. The top line contains gratings, the second line contains line pairs, and the bottom line contains bit patterns.

tions left. (1, 11, 101, 111, 1001, 1011, 1111, 10001, 10011, 10101, 10111, 11011, and 11111)

There exist four possible setups, to use a multimodal waveguide to detect the features on the mask, as shown on Fig. 6.9. In the top two graphs, the multimodal waveguide picks up light transmitted by an externally illuminated mask. Fig. 6.9A shows the near-field setup, and Fig. 6.9B is the far-field setup. The other graphs, Fig. 6.9C and D, show the eventual waveguide detection method: the mask is illuminated by the zeroth order mode of a waveguide and the reflected light from the mask is picked up by the same waveguide. In the far-field setup (Fig. 6.9D) it is possible to put a chopper in the light path, to filter out the parasitical reflections inside the photonic IC.

With the components of mask 1 it is not possible to simultaneously illuminate the mask and detect the light picked up with the same waveguide. The main reason is that the zeroth order mode is only connected to an output waveguide. But even if we used an external splitter and used that output waveguide also as input waveguide, the output waveguides of the first order mode are to closely spaced to that of the zeroth order mode. Therefore we first choose for a measurement setup of type A, shown on Fig. 6.10[2]. The mask is illuminated with a laser, the trans-

[2]This measurement setup was built at the OPTICA department at the Delft University of Technology in collaboration with Silvania Pereira and Sami Musa.

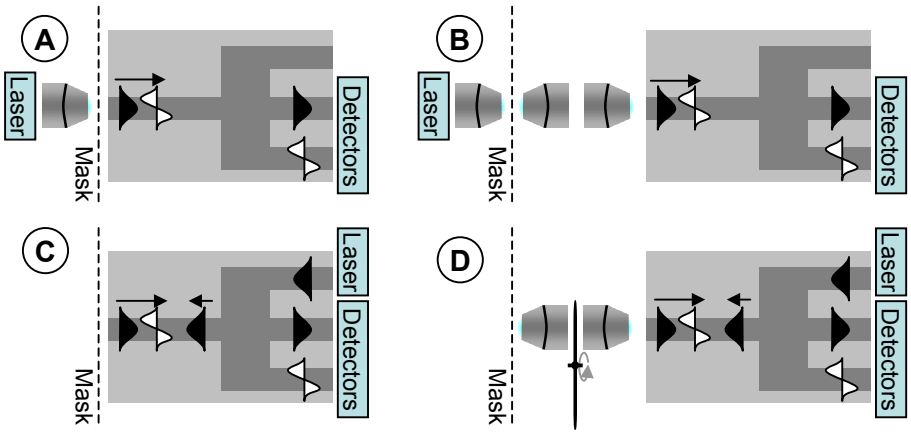


Figure 6.9: Four measurement setups for resolving small features. One can work in the near-field (A and C) or in the far-field (B and D). Another division is whether the mask is illuminated externally (A and B) or by the same waveguide as the detection (C and D).

mitted light is picked up by a multimodal waveguide in front of the mask, and the two modes of the multimodal waveguide are detected by a camera or detector. To align the setup, we went through three steps. Firstly, stage 2 is removed from the setup and the objective lens and mask are aligned in such a way that a relatively large spot of a few 10 µm wide illuminates the mask. This can be checked by imaging the mask on the camera using the optics on stage 3. The spot is centred in the large square on the mask. Secondly, the waveguide is placed back on stage 2 and aligned such that the excitation of the zeroth order mode is maximized. Finally the mask is moved such that the different features on the mask pass in front of the multimodal waveguide. On the camera or detector the signal from the zeroth and first order mode are captured. The gap in between the mask and the multimodal waveguide can be controlled by a capacitive sensor. To calibrate the zero-level of this sensor, the mask can be moved in such a way that there is contact with the waveguide.

6.2.2 Measurements

With the setup shown on Fig. 6.10, we measured to what extent the light transmitted by an externally illuminated grating, is picked up by the zeroth and the first order mode of a multimodal waveguide.

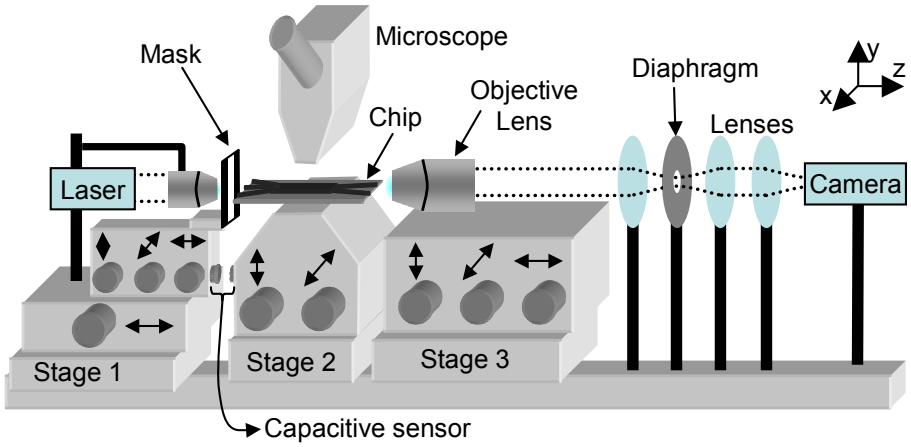


Figure 6.10: Measurement setup for resolving small features (Fig. 6.9A). A laser illuminates a mask through an objective lens. This light is picked up by a waveguide. The output is captured by an objective lens and guided to the detector.

Zeroth order mode Fig. 6.11 shows the excitation of the zeroth order mode detected when moving the mask in front of the waveguides. The graphs shows the result for different gratings. Within the measurement, we set the gap to 1 µm. A fit of the measurement results to the simulations shows however that the real gap is nearly 20 µm wide. This difference is caused by the fact that it is very hard to ensure that the mask and the waveguide run 100% parallel with the translation of the mask along the x- and y-axis. A small deviation in angle of the 2 cm wide mask will enlarge the gap in between the waveguide and the mask.

First order mode Fig. 6.12 shows the excitation of the first order mode detected by moving the grating with spatial period 10λ in front of the multimodal waveguide. In the right column a simulation is shown of the zeroth and first order mode excitation using the gap size 20 µm, found in the measurements of the zeroth order mode. A comparison between the measurement and the simulations suggests that a part of the zeroth order mode signal has been added to the first order mode signal. As the amplitude of the zeroth order mode signal is much bigger, a relatively small crosstalk may already distort the first order mode signal significantly. The interference of the first order mode and crosstalk

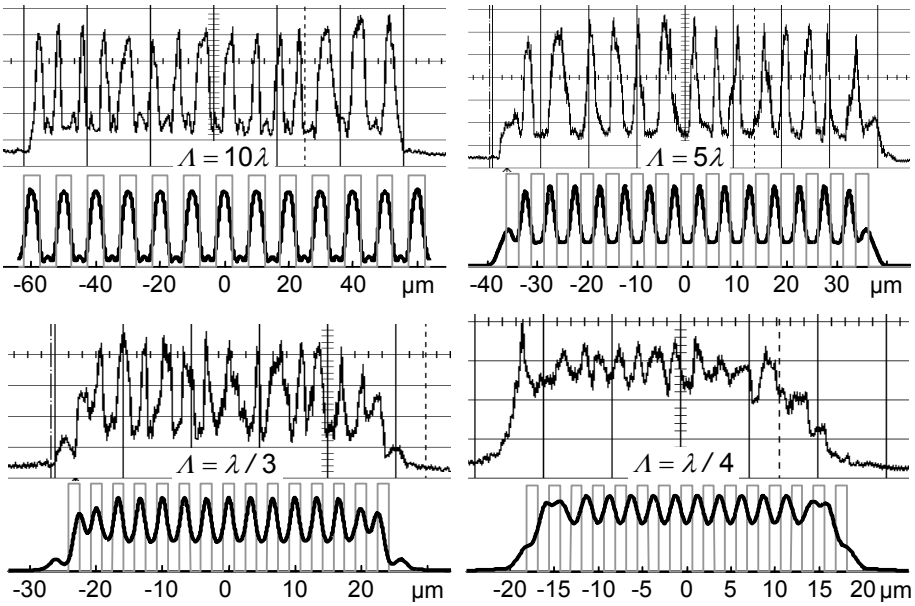


Figure 6.11: Detection with the zeroth order mode of externally illuminated gratings. The graph shows the measurements on top and the simulations at the bottom. The grating period in the four subfigures is: 10λ, 5λ, $\lambda/3$, $\lambda/4$. As the grating were moved manually the x-coordinates are only qualitatively correct. The simulations were fitted to the measurements and showed a gap of 20λ between grating and waveguide chip.

from the zeroth order mode can lead to the asymmetric pattern of the measured first order mode signal.

6.3 Reflection measurements

The setup of Fig. 6.1 was rebuilt, to measure how the components of mask 2 the mask can be used both to illuminate a disc and pick up the reflected light. Instead of a mask we used a gold coated standard fibre tip as reflector (125 µm diameter). This has two reasons. Firstly, the surface of the mask has some roughness and would scatter the light rather than reflecting it. Secondly, it is much easier to manipulate a small fibre tip than a larger mask. A detailed photograph of the setup on Fig. 6.13 shows the fibre array, the waveguide chip and the gold coated fibre tip.

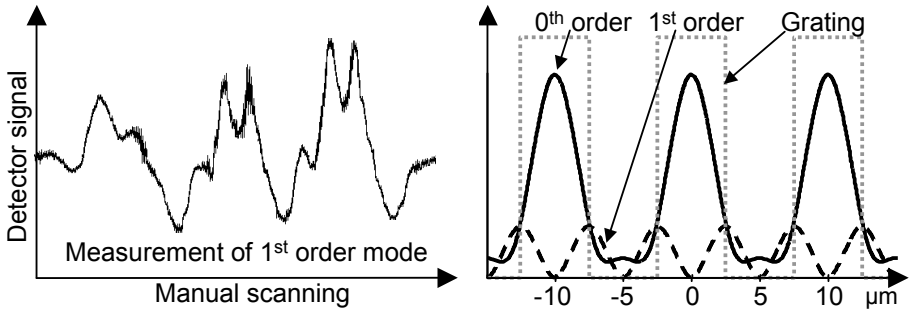


Figure 6.12: Detection with the first order mode of an externally illuminated grating with period 10λ. The gratings were moved manually, which means the x-coordinates of the measurement curve are not very precise. On the right a simulation for a gap of 20λ between grating and waveguide chip is shown for zeroth and first order mode.

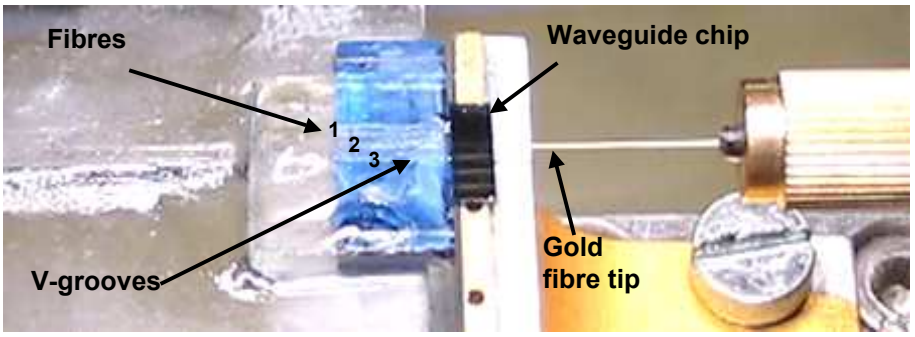


Figure 6.13: Detail of setup for a reflection measurement with the components of mask 2. Fibres in V-grooves are aligned with the input and output waveguides of the component on the sample. A gold coated fibre is aligned to the multimodal waveguide facet at the other side of the component. Fibre 1 is used to illuminate with the zeroth order, fibre 2 and fibre 3 detect respectively the zeroth order and the first order mode pick up in reflection.

A check of the alignment of the fibre array with the waveguide showed that 3% of the light inserted into the fibre 1 and 2 could be detected by an objective lens and detector at the other side of the sample. Fibre 3 showed a much lower throughput. The parasitical reflection inside the photonic IC or at the facet of the multimodal waveguide resulted in a reflection of 0.03%. If we put the gold coated fibre tip in contact with the facet of the multimodal waveguide, this reflection raised to 0.1%. This measured signal is however to weak to give a clue on how well small features can be resolved in reflection.

6.4 Conclusions

The measurements on the isolated chips showed that it is possible to make mode splitters with a crosstalk level between zeroth and first order mode as low as -22 dB. The crosstalk of the components of mask 2 was however higher. With the best components of mask 1 we tried to quantify the resolution limits of the waveguide detection method at a different set-up. These measurements did however not deliver very nice results. Because of the difficult alignment it was not possible to maintain a very small air gap in between the waveguide and the mask. Moreover the results for the first order mode showed a lot of crosstalk, probably from the slab light. The reflection measurements were especially hampered by the very low overall light throughput.

Chapter 7

Conclusions

In this thesis we showed the potential of a multimodal waveguide as a detection means in general and as pick-up head for optical discs specifically. With the so called waveguide detection system, a multimodal waveguide illuminates a small spot on the disc. The reflected light is captured by each of the modes of the same waveguide. The different modes are guided to separate detectors through a photonic integrated circuit, generating separate electrical signals. The optical properties of the waveguide detection system have been analysed extensively, and compared with those of the conventional DVD read-out system and the confocal microscope. As a test-case we also implemented a model for extracting the bits on a disc in presence of noise, showing that combining the signals from two different modes can decrease the bit error rate. We showed the analysis and design of a photonic IC splitting up the zeroth and first order mode. The design has been fabricated, tested, and measured.

As a summary we conclude with some drawbacks and potentials of the waveguide detection system, together with future perspectives. Whether the system might be included in a future generation of optical drives remains however a question and depends as much on the optical data storage market and support from major companies as on the optical properties of the disc read-out system itself.

7.1 Drawbacks

7.1.1 Light throughput

The major drawback of the waveguide detection system is the low light throughput compared to the extended detector system. This is primarily caused by three factors: The first reason is the intrinsic loss, due to the pin-hole like nature of the detector. A relatively large part of the light is scattered away at the interface. This loss is very severe for the first order mode detection. A second reason is the loss in the mode splitters. However theoretically it should be possible to make a mode splitter that splits off the two modes without loss, the practical demonstrations showed that it is hard to do so in reality. The third reason is the waveguide air-interface. As a mode propagates from the waveguide into the air or the other way around. Part of the light is reflected back. This loss can be heavily reduced by an anti-reflective coating but can never be zero for all modes at the same time.

7.1.2 Crosstalk

A second important drawback is the fact that there are a few sources of crosstalk. A first source of crosstalk is the mode splitter itself. We experimentally showed that the crosstalk can be as low as -22 dB. The question remains however whether this would be sufficient. A second source of crosstalk comes from the illumination mode. This crosstalk is low, but as the illumination signal is much bigger than the detection signal a small crosstalk can already be devastating for the bit error rate.

7.1.3 Compatibility and speed

Finally there are two issues that are unrelated to the potential discriminable minimal feature size on the disc, but are also important factors for a potential use in the fourth generation optical disc system.

The first issue is the compatibility. Contrary to hard disks, where every manufacturer can use a different technology, optical disc technology should be standardized world-wide, and should ideally be backwards compatible with earlier generations of optical discs. The waveguide detection method is heavily different from the conventional extended detector method, which makes it harder to gain acceptance compared to its direct competitors like phase masks.

The second issue is speed. Compared to the conventional method of processing the signals from the detectors, the waveguide detection method makes extensive use of heavy overlap calculations. Whether this may become a severe problem depends on the processing speed of future capacity of the computer hardware.

7.2 Potential and future perspectives

7.2.1 Enhanced resolution

With the waveguide detection method it is possible to simultaneously extract information from both the amplitude and the phase of the optical field reflected from the disc. In chapter 4 we showed for the far-field case, that combining the zeroth order and the first order mode signals can be combined using a simple parameter extraction algorithm. The results show a considerable decrease in bit error rate, which opens the way for a decrease in the feature size and an increase in data density. Depending on the noise the gain in lateral density could be as large as 25%.

In principle the waveguide detection method could also be used as pick-up head sliding at a narrow air gap over the disc. For very narrow waveguides this could in principle approximately give a 50% increase in cut-off frequency for very small air gaps. Because the waveguide mode diameter cannot be made very small, the gain in cut-off frequency is smaller than with other near-field techniques. Moreover the advantage of the near-field approach over a lens system is lost for gap widths larger than one wavelength. This is already many times smaller than the current flying heights.

7.2.2 Compatibility with multi-layer techniques

Due to its confocal nature, the waveguide detection method has a very good sectioning capacity. This means that only that layer of the disc that is in focus, will be captured by the waveguide detector. This sectioning capacity will also reduce the effect of scratches or dust on the disc surface.

7.2.3 Integration and miniaturization

In principle it should be possible to fit all optical functions including the light source and detectors on a single chip. This would not only in-

crease the robustness and decrease the cost in case of mass fabrication, but opens also the way to smaller pick-up heads. This might be very important for optical drives in mobile devices. Such an optical drive could easily fit in a PC-card [77].

7.2.4 Parallelization

Integrating the optical pick-up head in a single chip opens also the way to parallel recording and read-out. The cost of placing several laser, detectors, multimodal waveguides, and mode splitters in a row compared to a single combination is relatively low compared to the cost of the packaging and the optical disc drive as a whole. This parallelization can provide a big increase in read-out speed.

7.2.5 Extension to three dimensions

Within this thesis the focus was on increasing the data density on a track of bits considering the distance between the tracks as a constant. As the bit error rate has improved it is possible to decrease the inter-track distance instead of decreasing the bit lengths within a track. A next possible step would be to use a waveguide that is not only bimodal along the track but is also bimodal in the radial direction of the disc. Such a waveguide would hence guide four different modes. If one would be able to make a photonic IC that splits up all those modes, the bit error rate could be further reduced. Moreover one could use the low frequency components of the modes that are odd in the radial direction of the disc as push-pull signal for the tracking.

Appendix A

Notations and definitions

In this appendix we describe a few functions, and conventions. Some basic functions are sometimes defined in different ways. To solve this ambiguity, we give here an overview of the definitions used in this thesis. At the end of this appendix a list is added with the most important symbols used in this thesis.

A.1 Functions

In the text we make use of the following functions

Step function
$$\mathcal{H}(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2} & \text{for } x = 0 \\ 1 & \text{for } x > 0 \end{cases} \tag{A.1}$$

Ramp function
$$\mathcal{R}(x) = x\,\mathcal{H}(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases} \tag{A.2}$$

Rectangle function:
$$\Pi(x) = \mathcal{H}\big(\tfrac{1}{2} - x\big) = \begin{cases} 1 & \text{for } |x| < \frac{1}{2} \\ \frac{1}{2} & \text{for } |x| = \frac{1}{2} \\ 0 & \text{for } |x| > \frac{1}{2} \end{cases} \tag{A.3}$$

Gaussian function:
$$\mathcal{G}(x) = \exp\big(-\pi x^2\big) \tag{A.4}$$

Error function:
$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\big(-t^2\big)\,\mathrm{d}t \tag{A.5}$$

Sinc function:
$$\operatorname{sinc}(x) = \frac{\sin(x)}{x} \tag{A.6}$$

By extension we define

$$\mathcal{H}(x,y) = \mathcal{H}(x)\,\mathcal{H}(y) \qquad \mathcal{R}(x,y) = \mathcal{R}(x)\,\mathcal{R}(y) \tag{A.7}$$

$$\Pi(x,y) = \Pi(x)\,\Pi(y) \qquad \mathcal{G}(x,y) = \mathcal{G}(x)\,\mathcal{G}(y) \tag{A.8}$$

Some properties

$$\int_c^d \mathcal{G}(ax+b)\,\mathrm{d}x = \frac{\mathrm{erf}[\sqrt{\pi}\,(ad+b)] - \mathrm{erf}[\sqrt{\pi}\,(ac+b)]}{2a} \tag{A.9}$$

$$\int_{-\infty}^{\infty} \mathcal{G}(ax+b)\,\mathrm{d}x = \frac{1}{a} \tag{A.10}$$

$$\widehat{\mathcal{G}}(ax+b) = \frac{\exp(2\pi b x)}{a}\,\mathcal{G}\Big(\frac{x}{a}\Big) \tag{A.11}$$

$$\mathcal{G}(ax+b)\,\mathcal{G}(cx+d) = \mathcal{G}\left[\frac{|ad-bc|}{\sqrt{a^2+c^2}}\right]\mathcal{G}\left(\sqrt{a^2+c^2}\,x + \frac{ab+cd}{\sqrt{a^2+c^2}}\right) \tag{A.12}$$

$$\mathcal{G}(ax+b) \otimes \mathcal{G}(cx+d) = \frac{1}{\sqrt{a^2+c^2}}\,\mathcal{G}\left(\frac{ac}{\sqrt{a^2+c^2}}x + \frac{ad+bc}{\sqrt{a^2+c^2}}\right) \tag{A.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\,\mathcal{G}(ax+b) = -2\pi\,(ax+b)\,G(ax+b) \tag{A.14}$$

A.2 Complex numbers

Assume a complex number $c = a + jb$ with $a, b \in \mathbb{R}$. Then the real part is defined $\Re(c) = a$, the imaginary part is defined as $\Im(c) = b$, the complex adjunct is defined as $c^* = a - jb$, the amplitude is defined as $|c| = \sqrt{a^2+b^2}$ and the phase is defined as $\angle(c) = \arctan(b/a)$.

A.3 Convolution

$$h(x) = f(x) \otimes g(x) = \int f(x-u)\,g(u)\,\mathrm{d}u = \int f(u)\,g(x-u)\,\mathrm{d}u \tag{A.15}$$

By extension we define the two dimensional convolution operator as:

$$\begin{aligned} h(x,y) = f(x,y) \otimes g(x,y) &= \iint f(x-u, y-v)\,g(u,v)\,\mathrm{d}u\,\mathrm{d}v \\ &= \iint f(u,v)\,g(x-u, y-v)\,\mathrm{d}u\,\mathrm{d}v \end{aligned} \tag{A.16}$$

A.4 Fourier transform

In the scientific community different definitions are used for a Fourier transform. The Fourier transform and the inverse Fourier transform used in this thesis are defined by:

$$\widehat{\psi}(f) = \mathcal{F}[\psi(x)](f) = \int \psi(x) \exp(-2\pi j f x)\, \mathrm{d}x \tag{A.17}$$

$$\psi(x) = \mathcal{F}^{-1}\Big[\widehat{\psi}(f)\Big](x) = \int \widehat{\psi}(f) \exp(2\pi j f x)\, \mathrm{d}f \tag{A.18}$$

By extension the two dimensional Fourier transform is defined as:

$$\widehat{\psi}(f,g) = \iint \psi(x,y) \exp[-2\pi j\,(fx+gy)]\, \mathrm{d}x\, \mathrm{d}y \tag{A.19}$$

$$\psi(x,y) = \iint \widehat{\psi}(f,g) \exp[2\pi j\,(fx+gy)]\, \mathrm{d}f\, \mathrm{d}g \tag{A.20}$$

Properties:

$$\widehat{\widehat{\psi}}(x,y) = \psi(-x,-y) \tag{A.21}$$

$$\widehat{A \otimes B} = \widehat{A}\widehat{B} \tag{A.22}$$

A.5 Tilde operator

Within this thesis we make use of another similar operator, which we will call the 'Tilde' operator and is defined by:

$$\widetilde{\Psi}\big(f;f'\big) = \int \Psi\big(x;x'\big) \exp\big[-2\pi j\,\big(fx - f'x'\big)\big]\, \mathrm{d}x\, \mathrm{d}x' \tag{A.23}$$

$$\Psi\big(x;x'\big) = \int \widetilde{\Psi}\big(f;f'\big) \exp\big[2\pi j\,\big(fx - f'x'\big)\big]\, \mathrm{d}f\, \mathrm{d}f' \tag{A.24}$$

Or in two dimensions:

$$\widetilde{\Psi}\big(f,g;f',g'\big) = \iint \Psi\big(x,y;x',y'\big)\, \mathrm{e}^{-2\pi j(fx+gy-f'x'-g'y')}\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}x'\, \mathrm{d}y' \tag{A.25}$$

$$\Psi\big(x,y;x',y'\big) = \iint \widetilde{\Psi}\big(f,g;f',g'\big)\, \mathrm{e}^{2\pi j(fx+gy-f'x'-g'y')}\, \mathrm{d}f\, \mathrm{d}g\, \mathrm{d}f'\, \mathrm{d}g' \tag{A.26}$$

A basic property is :

$$\text{If } \Psi\big(x;x'\big) = \psi(x)\,\psi^*\big(x'\big) \text{ then } \widetilde{\Psi}\big(f;f'\big) = \widehat{\psi}(f)\,\widehat{\psi}^*\big(f'\big)\,. \tag{A.27}$$

A.6 Notations

In this section an overview is given of some symbols and notations used in this thesis.

λ	Wavelength.
NA	Numerical aperture.
$\psi(x,y)$	The complex amplitude of the optical field.
$J\big(x,y;x',y'\big) =$	
$\langle\psi(x,y)\,\psi\big(x',y'\big)\rangle$	The mutual intensity.
$I(x_s)$	Detector signal.
$psf_{incoh}(x)$	Incoherent point spread function.
$psf(x) = psf_{coh}(x)$	Coherent point spread function.
$PSF(x)$	Partial coherent point spread function.
γ	Linear demagnification factor.
$r(x,y)$	Complex reflection coefficient of the disc.
$R\big(x,y;x',y'\big) =$	
$r(x,y)\,r^*\big(x',y'\big)$	Generalized reflection function.
d_{bit}	Minimum bit size.
$P_j(t)$	Detected power, along mode j.
$C_j^k(t)$	Calculated power of pattern k, along mode j.
$S_j^k =$	
$\int_{t_l}^{t_u}\Big[P_j(t) - C_j^k(t)\Big]^2\mathrm{d}t$	Overlap integral.
k_j	The bit pattern which minimizes S_j^k.
$k_{correct}$	The bit pattern on the disc.
$S_T^k = \sum_{j=0}^{N-1}\beta_j S_j^k$	Linear combination of the overlap integrals.
$\beta = \beta_0/\left(\beta_0+\beta_1\right)$	Fraction of zeroth order in total.
BER_j	Bit error rate, obtained using mode j.
BER_{min}	Minimum BER using all detection modes.
BER_{opt}	BER for a fixed $\beta = \beta_{opt}$.

Appendix B

Mathematical derivations

In this appendix we include some intermediate calculations from chapter 3 that are not needed for insight in the final equations, and would clutter the text unnecessary.

B.1 Waveguide detection systems calculations

$$\widehat{PSF}(f,g;f',g') = \widehat{psf_{tot}}(f,g)\,\widehat{psf^{*}_{tot}}(f',g') \tag{B.1}$$

$$\begin{aligned}
\widehat{psf_{tot}}(f,g) &= \mathcal{F}[psf_{tot}(x,y)] && \text{(B.2)}\\
&= \mathcal{F}\big\{\big[psf_{lens}(x,y)\otimes m^{\gamma}_{illum}(x,y)\big] \\
&\qquad \times \big[psf_{lens}(x,y)\otimes m^{\gamma}_{detect}(x,y)\big]\big\} && \text{(B.3)}\\
&= \big\{\mathcal{F}[psf_{lens}(x,y)]\,\mathcal{F}\big[m^{\gamma}_{illum}(x,y)\big]\big\} \\
&\qquad \otimes \big\{\mathcal{F}[psf_{lens}(x,y)]\,\mathcal{F}\big[m^{\gamma}_{detect}(x,y)\big]\big\} && \text{(B.4)}\\
&= \Big\{\widehat{psf_{lens}}(f,g)\,\widehat{m^{\gamma}_{illum}}(f,g)\Big\} \\
&\qquad \otimes \Big\{\widehat{psf_{lens}}(f,g)\,\widehat{m^{\gamma}_{detect}}(f,g)\Big\} && \text{(B.5)}
\end{aligned}$$

B.2 Extended detector calculations

If we start with Eq. (3.49),

$$PSF(x,y;x',y') = \frac{1}{(\lambda z_{disc})^2}\psi_{2^-}(x,y)\,\psi^{*}_{2^-}(x',y')\,\widehat{p_{ext}}\left(\frac{x-x'}{\lambda z_{disc}},\frac{y-y'}{\lambda z_{disc}}\right), \tag{B.6}$$

and write the Fourier transform $\widehat{p_{ext}}$ explicitly we find:

$$\begin{aligned} OTF\left(f,g;f',g'\right) &= \frac{1}{\left(\lambda z_{disc}\right)^2}\iiiint \psi_{2^-}(x,y)\,\psi_{2^-}^*\left(x',y'\right) \\ &\times \iint p_{ext}(u_3,v_3)\exp\left\{-2\pi j\left[u_3\frac{x-x'}{\lambda z_{disc}}+v_3\frac{y-y'}{\lambda z_{disc}}\right]\right\}\mathrm{d}u_3\,\mathrm{d}v_3 \\ &\times \exp\left[-2\pi j\left(fx+gy-f'x'-g'y'\right)\right]\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}x'\,\mathrm{d}y' \end{aligned} \tag{B.7}$$

Changing the integration order and using the shift theorema we find:

$$\begin{aligned} OTF\left(f,g;f',g'\right) &= \iint \widehat{\psi_{2^-}}\left(f+\frac{u_3}{\lambda z_{disc}},g+\frac{v_3}{\lambda z_{disc}}\right) \\ &\times\widehat{\psi_{2^-}}^*\left(f'+\frac{u_3}{\lambda z_{disc}},g'+\frac{v_3}{\lambda z_{disc}}\right)p_{ext}(u_3,v_3)\,\mathrm{d}u_3\,\mathrm{d}v_3 \end{aligned} \tag{B.8}$$

After filling in

$$\psi_{2^-}(x,y) = \iint m^{\gamma}(x_1,y_1)\,psf_{lens}(x-x_1,y-y_1)\,\mathrm{d}x_1\,\mathrm{d}y_1, \tag{B.9}$$

$$\text{and } \widehat{\psi_{2^-}}(f,g) = \widehat{m^{\gamma}_{illum}}[f,g]\;p_{lens}(-f\lambda z_{disc},-g\lambda z_{disc}) \tag{B.10}$$

and substituting u_3 and v_3 by $u\lambda z_{disc}$ and $v\lambda z_{disc}$ Eq. (B.8) can be rewritten as

$$\begin{aligned} OTF\left(f,g;f',g'\right) &= \iint p_{ext}(\lambda z_{disc}u,\lambda z_{disc}v) \\ \times p_{lens}[\lambda z_{disc}\left(f+u\right),\lambda z_{disc}\left(g+v\right)]\,&p_{lens}\left[\lambda z_{disc}\left(f'+u\right),\lambda z_{disc}\left(g'+v\right)\right] \\ &\times\widehat{m^{\gamma}_{illum}}[f+u,g+v]\,\widehat{m^{\gamma}_{illum}}^*\left[f'+u,g'+v\right]\mathrm{d}u\,\mathrm{d}v, \end{aligned} \tag{B.11}$$

which is identical to Eq. (3.52).

B.3 Confocal calculations

From Eq. (3.58), one finds by using some Fourier transforms that

$$\begin{aligned} PSF\big(x,y;x',y'\big) &= \psi_{2^-}(x,y)\,\psi_{2^-}^{*}\big(x',y'\big) \\ &\times \iint\!\!\iint \widehat{psf_{lens}}(u,v)\exp\{-j2\pi\left[(x-x_3)\,u,(y-y_3)\,v\right]\}\,\mathrm{d}u\,\mathrm{d}v \\ &\quad\times \iint \widehat{psf_{lens}^{*}}\big(u',v'\big)\exp\big\{j2\pi\left[\big(x'-x_3\big)\,u',\big(y'-y_3\big)\,v'\right]\big\}\,\mathrm{d}u'\,\mathrm{d}v' \\ &\quad\times p_{conf}^{\gamma}(x_3,y_3)\,\mathrm{d}x_3\,\mathrm{d}y_3 \end{aligned} \tag{B.12}$$

$$\begin{aligned} PSF\big(x,y;x',y'\big) &= \psi_{2^-}(x,y)\,\psi_{2^-}^{*}\big(x',y'\big) \\ &\times \iint\!\!\iint \widehat{psf_{lens}}(u,v)\exp\{-j2\pi\,(xu,yv)\}\,\mathrm{d}u\,\mathrm{d}v \\ &\quad\times \iint \widehat{psf_{lens}^{*}}\big(u',v'\big)\exp\big\{j2\pi\,\big(x'u',y'v'\big)\big\}\,\mathrm{d}u'\,\mathrm{d}v' \\ &\quad\times \exp\big\{j2\pi\left[\big(u-u'\big)\,x_3,\big(v-v'\big)\,y_3\right]\big\}\,p_{conf}^{\gamma}(x_3,y_3)\,\mathrm{d}x_3\,\mathrm{d}y_3 \end{aligned} \tag{B.13}$$

$$\begin{aligned} PSF\big(x,y;x',y'\big) &= \psi_{2^-}(x,y)\,\psi_{2^-}^{*}\big(x',y'\big) \\ &\quad\times \iiiint \widehat{psf_{lens}}(u,v)\,\widehat{psf_{lens}^{*}}\big(u',v'\big)\,\widehat{p_{conf}^{\gamma}}\big(u'-u,v'-v\big) \\ &\quad\times \exp\{-j2\pi\,(xu,yv)\}\exp\big\{j2\pi\,\big(x'u',y'v'\big)\big\}\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}u'\,\mathrm{d}v' \end{aligned} \tag{B.14}$$

and one finally finds:

$$\begin{aligned} OTF\big(f,g;f',g'\big) &= \iiiint \widehat{\psi_{2^-}}(f-u,g-v)\,\widehat{\psi_{2^-}}^{*}\big(f'-u',g'-v'\big) \\ &\times \widehat{psf_{lens}}(u,v)\,\widehat{psf_{lens}^{*}}\big(u',v'\big)\,\widehat{p_{conf}^{\gamma}}\big(u'-u,v'-v\big)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}u'\,\mathrm{d}v' \end{aligned} \tag{B.15}$$

which is identical to Eq. (3.61)

Appendix C

Eigenmode expansion method

Expansion into eigenmode is a standard technique, in the field of integrated optics. This appendix gives a brief outline of this technique.

C.1 Introduction

The eigenmode expansion method is a technique to calculate the propagation of a monochromatic optical field through a set of sections with z-invariant refractive index. For a wavelength λ, the time dependency is $\exp\left(j\frac{2\pi c}{\lambda}t\right)$. The eigenmode expansion method is based on a vectorial optical **E**- and **H**-field. Within this thesis we assumed however a scalar approximation, which says that the optical field can be expanded into two independent polarizations. We define the scalar optical field as the electric field component for TE-polarization (electric field component lays in the plane of the waveguide chip) and as the magnetic field component for TM-polarization (magnetic field component lays in the plane of the waveguide chip).

For a z-invariant section it is possible to find field profiles that have a z-dependency as follows:

$$\Phi_\nu(x,y,z) = \phi_\nu(x,y)\exp(-j\beta_\nu z)\,. \tag{C.1}$$

The field profiles, $\phi_\nu(x,y)$, are called eigenmodes and maintain their general shape during propagation. β_ν is called the propagation constant. In general the eigenmodes form a finite discrete set of guided modes and an infinite continuous set of radiative modes. For practical

applications one considers however only eigenmodes within a hypothetical metal box placed around the simulated structure. In this case a infinite, but discrete, set of modes is sufficient to form a complete orthogonal set. The eigenmodes are defined disregarding an arbitrary multiplication factor. We choose to normalize them in such a way that the power flow of a propagating mode is normalized to 1. This means:

$$\iint \phi_\nu(x,y)\,\phi^*_{\nu'}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \begin{cases} 0, & \text{for } \nu \neq \nu' \\ \beta_\nu, & \text{for } \nu = \nu' \text{ and TE} \\ \frac{1}{\beta_\nu}, & \text{for } \nu = \nu' \text{ and TM} \end{cases} \tag{C.2}$$

As the eigenmodes form a complete set, which means a general optical field $\psi(x,y,z=0)$ can be expanded in terms of these eigenmodes. In principle one needs all the modes, but for practical purposes the first N modes are sufficient:

$$\psi(x,y,z=0) = \sum_{\nu=0}^{N} a_\nu \phi_\nu(x,y) \tag{C.3}$$

The coefficients a_ν can be calculated as using Eq. (C.2):

$$a_\nu = \iint \psi(x,y,z=0)\,\phi_\nu(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{C.4}$$

By multiplying each eigenmode with its propagation factor, one can

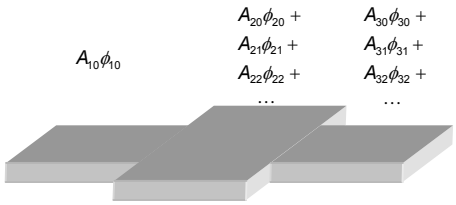


Figure C.1: Expansion into eigenmodes. A three-dimensional structure is subdivided in longitudinal invariant sections. In each of these sections one can construct eigenmodes. These are optical field profiles that retain their shape while propagating in this section. At each interface the total optical field needs to be expanded in the eigenmodes of the next section.

calculate easily the propagation of the total field ψ in the forward direction:

$$\psi(x,y,z=z_1) = \sum_{\nu=0}^{N} a_\nu \phi_\nu(x,y)\exp(-j\beta_\nu z). \tag{C.5}$$

All equations above are valid within one z-invariant section. At the interface between two sections an optical field is partially reflected and partially transmitted. An eigenmode in section A, forward propagating towards section B, will result in a reflected field that can be expanded in backward propagating eigenmodes of section A, and in a transmitted field that can be expanded in forward propagating eigenmodes of section B. The excitation coefficients depends on the mutual overlap integral in between the eigenmodes of both sections and on the refractive index profile. The same can be done for an eigenmodes from section B propagating towards the interface:

$$\phi_i^{A^+}(x,y) \rightarrow \sum_{j=1}^{N} t_{i,j}^{A,B}\phi_j^{B^+}(x,y) + r_{i,j}^{A,B}\phi_j^{A^-}(x,y) \tag{C.6}$$

$$\phi_j^{B^-}(x,y) \rightarrow \sum_{i=1}^{N} t_{j,i}^{B,A}\phi_i^{A^-}(x,y) + r_{j,i}^{B,A}\phi_i^{B^+}(x,y) \tag{C.7}$$

Using the reciprocity theorem and the power conservation law one can find different relationship between these coefficients:

$$\text{Reciprocity: } \begin{cases} \forall A,B,i,j : t_{i,j}^{A,B} = t_{j,i}^{B,A} \\ \forall A,B,i,j : r_{i,j}^{A,B} = r_{j,i}^{A,B} \end{cases} \tag{C.8}$$

$$\text{Power conservation: } \begin{cases} \forall A,B,i : \sum_{j=1}^{N} \left(\left| r_{i,j}^{A,B} \right|^2 + \left| t_{i,j}^{A,B} \right|^2 \right) = 1 \end{cases} \tag{C.9}$$

These coefficients are sufficient to calculate the reflection and transmission for any arbitrary field, by expanding it into the eigenmodes.

C.2 Eigenmode calculation

In general In a non-uniform space it is a non-trivial task to calculate the eigenmodes. There exist however software packages, both commercial and free open source software. The main packages we have used are Fimmwave-Fimmprop3d from Photon Design[1] and the open source package CAMFR[2] [44], developed by the photonics research group at the Ghent University. Fimmwave solves the eigenmode problem in a

[1] Photon design, http://www.photond.com/

[2] CAMFR, CAvity Modelling FRamework, http://camfr.sourceforge.net/

three-dimensional closed box. CAMFR was used to solve the electromagnetic problems in two dimensions.

The accurateness of the eigenmode expansion technique depends heavily on the box size, the boundary conditions and the number of modes that is taken into account in the calculations. Optimizing these parameters is an expertise on itself.

In free space A special case is propagation in free space. In this case the eigenmodes are plane waves represented by $\Phi(\mathbf{r}) = \exp(-j\mathbf{k}\cdot\mathbf{r})$, where $\mathbf{k}$ is the wave vector. Using a double index f and g ($f = \frac{k_x}{2\pi}$ and $g = \frac{k_y}{2\pi}$) one finds:

$$\Phi_{f,g}(x,y,z) = \exp[-j2\pi\,(fx+gy)]\exp[-j\beta_{f,g}z] \tag{C.10}$$

$$\text{with } \beta_{f,g} = \begin{cases} 2\pi\sqrt{\frac{1}{\lambda^2} - f^2 - g^2}, & \text{if } f^2+g^2 \leq \frac{1}{\lambda^2} \\ -j2\pi\sqrt{f^2+g^2-\frac{1}{\lambda^2}}, & \text{if } f^2+g^2 \geq \frac{1}{\lambda^2} \end{cases} \tag{C.11}$$

The coefficients for decomposition of a general field $\psi(x,y)$ into plane wave eigenmodes can be found by taking the Fourier transform of the original field.

$$a_{f,g} = \widehat{\psi}(f,g) = \iint \psi(x,y)\exp[-j2\pi\,(fx+gy)]\,\mathrm{d}y\,\mathrm{d}x \tag{C.12}$$

C.3 Effective index method

Very often it is possible to reduce the calculations heavily by only taking into account two dimensions: the dimension along the mode propagation and one dimensions perpendicular to the mode propagation and within the plane of the waveguide chip. This can be done by replacing the refractive index profile in the dimension orthogonal to the waveguide chip by a uniform effective index. There is no single method for calculating an effective index that gives the best approximation, but one of them is to replace it with the effective index of the fundamental mode of the slab waveguide. In any case, one should be cautious with the polarization. As written in section C.1 the TE-polarization for the 3-dimensional waveguide is defined as the polarization with the electric field component in the plane of the waveguide chip. After simplifying

the structure to two dimension the **E**-field is orthogonal to the slab-layers, as shown on figure C.2. This means the 2-dimensional equations should be solved for TM-polarization in order to approximate the TE-polarization in the 3-dimensional waveguide.

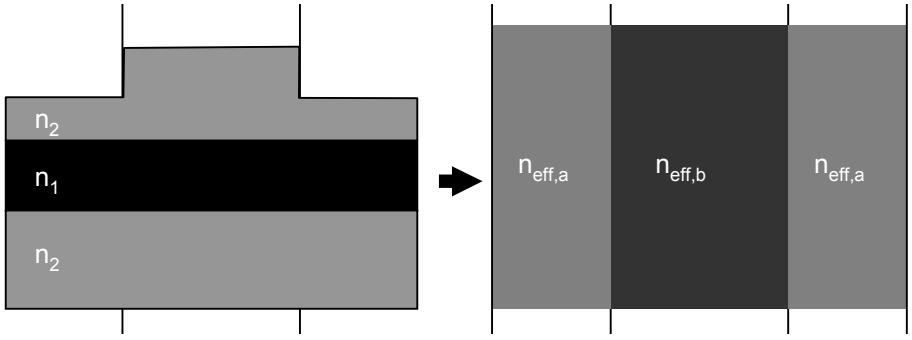


Figure C.2: The effective index method: A two-dimensional waveguide cross-section is subdivided in section that have a planar structure. The propagation constant of the modes in these planar structures are used to define a so called effective refractive index, n_eff. In a second step the lateral confinement and propagation constant of the real mode is approximated by a mode in between layers with these effective refractive index.

Figures

Tables

Bibliography

[1] S. Spielman, B.V. Johnson, G.A. McDermott, M.P. O'Neill, C. Pietrzyk, T. Shafaat, D.K. Warland, and T.L. Wong. Using pit-depth modulation to increase capacity and data transfer rate. *SPIE Proceedings*, 3109:98–104, 1997.

[2] S.W. McLaughlin, Y.C. Lu, C. Pepin, and D. Warland. Multilevel dvd: coding beyond 3 bits/data-cell. In *International Symposium on Optical Memory and Optical Data Storage Topical Meeting*, pages 380–382, 2002.

[3] J. Leuthold, R. Hess, J. Eckner, P. A. Besse, and H. Melchior. Spatial mode filters realized with multimode interference couplers. *Opt. Lett.*, 21(11):836–838, June 1996.

[4] R. Bradshaw and C. Schroeder. Fifty years of ibm innovation with information storage on magnetic tape. *IBM journal of Research and Development*, 47(4):373–383, July 2003.

[5] H.J. Mamin, R.P. Ried, B.D. Terris, and Rugar D. High-density data storage based on the atomic force microscope. *Proceedings of the IEEE*, 87(6):1014–1027, June 1999.

[6] P. Vettiger, G. Cross, M. Despont, U. Drechsler, U. Durig, B. Gotsmann, W. Haberle, M.A. Lantz, H.E. Rothuizen, R. Stutz, and G.K. Binnig. The "millipede" - nanotechnology entering data storage. *IEEE Transactions on Nanotechnology*, 1(1):39–55, March 2000.

[7] E. Spitz, J.P. Huignard, and C.P. Puech. Early experiments on optical disc storage. *IEEE Journal of Selected Topics in Quantum Electronics*, 6(6):1413–1418, November/December 2000.

[8] A.E. Bell. The dynamic digital disk. *IEEE Spectrum*, 36(10):28–35, October 1999.

[9] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K. Schouhamer Immink. *Principles of Optical Disc Systems*. Adam Hilger Ltd, Bristol, United Kingdom, 1985.

[10] J. Braat, P. Dirksen, and Janssen A.J.E.M. *Optical imaging and microscopy: techniques and advanced systems*, chapter Diffractive Read-Out of Optical Discs. 87. Springer Series in Optical Sciences, 2002.

[11] H.A. Jones-Bey. High-density storage options compete. *Laser Focus World*, 40(6):31, June 2004.

[12] Immink K. Efm coding: Squeezing the last bits. *IEEE Transactions on Consumer Electronics*, 43(3):491–495, August 1997.

[13] W. Coene, H. Pozidis, M. Van Dijk, J. Kahlman, R. Van Woudenberg, and B. Stek. Channel coding and signal processing for optical recording systems beyond dvd. *IEEE Transactions on Consumer Electronics*, 37(2):682–688, March 2001.

[14] Milster T.D. Near-field optics: A new tool for data storage. *Proceedings of the IEEE*, 88(9):1480–1490, September 2000.

[15] Q. Wu, L.P. Ghislain, and V.B. Elings. Imaging with solid immersion lenses, spatial resolution, and applications. *Proceedings of the IEEE*, 88(9):1491–1498, September 2000.

[16] P. Herget, T.E. Schlesinger, and D.D. Stancil. Domain position detection mammos. In *Proceedings of the International Symposium on Optical Memory and Optical Data Storage Topical Meeting, 2002.*, pages 22–24, July 2002.

[17] W.C. Hsu, M.R. Tseng, and P.C. Tsai, S.Y. Kuo. Blue-laser readout properties of super-rens disc with inorganic-write-once recording layer. In *International Symposium on Optical Memory and Optical Data Storage Topical Meeting*, pages 192 – 194, July 2002.

[18] J. Kim, I. Hwang, D. Yoon, I. Park, D. Shin, T. Kikukawa, T. Shima, and J. Tominaga. Super-resolution by elliptical bubble formation with ptox and aginsbte layers. *Applied Physics Letters*, 83(9):1701–1703, September 2003.

[19] Y. Kawata, R. Juskaitis, T. Tanaka, T. Wilson, and S. Kawata. Differential phase-contrast microscope with a split detector for the readout system of a multilayered optical memory. *Applied Optics*, 35(14):2466–2470, May 1996.

[20] N. Shida, K. Suga, T. Higuchi, and T. Iida. Super high density optical disc by using multi-layer structure. In *Conference Digest of Optical Data Storage, 2000.*, pages 27 – 29, May 2000.

[21] J. Ashley, M.P. Bernal, G.W. Burr, H. Coufal, H. Guenther, J.A. Hoffnagle, C.M. Jefferson, B. Marcus, R.M. Macfarlane, R.M. Shelby, and G.T. Sincerbox. Holographic data storage. *IBM Journal of Research and Development*, 44(3):341–368, May 2000.

[22] L. Hesselink, S.S. Orlov, and Bashaw M.C. Holographic data storage systems. *Proceedings of the IEEE*, 92(8):1231–1280, August 2004.

[23] T. Zhou, C. Tan, C. Leis, I. Harvey, G. Lewis, T. Wong, and M. ONeill. Multilevel amplitude-modulation system for optical data storage. In *Proc. SPIE Advanced Optical Storage Technology*, volume 4930, pages 7–20, 2002.

[24] W.M.J. Coene. Nonlinear signal-processing model for scalar diffraction in optical recording. *Applied Optics*, 42(32):6525–6535, November 2003.

[25] L. Fagoonee, A. Moinian, B. Honary, A. van der Lee, and W.M.J. Coene. Experimental characterization for the binary two-dimensional optical storage channel. In *IEEE International Conference on Communications*, volume 5, pages 2772–2776, June 2004.

[26] A.H.J. Immink, J. Riani, S.J.L. v. Beneden, J.W.M. Bergmans, M. Ciacci, A. Nowbakht Irani, W.M.J. Coene, A.M. van der Lee, and D. Bruls. Adaptation and timing recovery for two-dimensional optical storage. In *Proceedings of the Optical Data Storage conference*, 2004.

[27] D. Kraemer, B.J. Siwick, and Miller R.J.D. Ultra high-density optical data storage information. *Chemical Physics*, 281(1):73–83, December 2001.

[28] T. Wilson. Image-formation in 2-mode fiber-based confocal microscopes. *Journal Of The Optical Society Of America A-Optics Image Science And Vision*, 10(7):1535–1543, July 1993.

[29] T. Wilson. Coherent methods in confocal microscopy. *IEEE Engineering in Medicine and Biology Magazine*, 15(1):84–91, January-February 1996.

[30] A.J. den Dekker and A. van den Bos. Resolution: a survey. *J. Opt. Soc. Am. A*, 14(3):547–557, March 1997.

[31] M. Born and E. Wolf. *Principles of Optics*. Pergamon Press, 6th edition, 1980.

[32] I. Leizerson, S.G. Lipson, and V. Sarafis. Superresolution in far-field imaging. *Journal Of The Optical Society Of America A-Optics Image Science And Vision*, 19(3):436–443, March 2002.

[33] A.J. den Dekker. Model-based optical resolution. *IEEE Transactions on Instrumentation and Measurement*, 46(4):798–802, August 1997.

[34] M. Shahram and P. Milanfar. Imaging below the diffraction limit: A statistical analysis. *IEEE Transactions on Image Processing*, 13(5):677–689, May 2004.

[35] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October 1948.

[36] I.J. Cox and C.J.R. Sheppard. Information capacity and resolution in an optical-system. *Journal Of The Optical Society Of America A-Optics Image Science And Vision*, 3(8):1152–1158, August 1986.

[37] S.W. McLaughlin. Shedding light on the future of sp for optical recording. *IEEE Signal Processing Magazine*, 15(4):83–94, July 1998.

[38] L.L. McPheters and S.W. McLaughlin. Turbo-coded optical recording channels with dvd minimum mark size. *IEEE Transactions on Magnetics*, 38(1):298–302, January 2002.

[39] H. Hayashi, H. Kobayashi, M. Umezawa, and S. Hosaka. Dvd players using a viterbi decoding circuit. *IEEE Transactions on Consumer Electronics*, 44(2):268–272, May 1998.

[40] White paper blu-ray disc format 1.a physical format specifications for bd-re. Technical report, Blu-ray Disc Founders, August 2004.

[41] C. Berrou, A. Glavieux, and P. Thitimajashima. Near shannon limit error-correcting coding and decoding: Turbo-codes. 1. In *Communications, 1993. ICC 93*, volume 2, pages 1064–1070, May 1993.

[42] C. Berrou. The ten-year-old turbo codes are entering into service. *IEEE Communications Magazine*, 41(8):110–116, August 2003.

[43] C.M. Riggle and S.G. McCarthy. Design of error correction systems for disk drives. *IEEE Transactions on Magnetics*, 34(4):2362–2371, July 1998.

[44] P. Bienstman and R. Baets. Advanced boundary conditions for eigenmode expansion models. *Optical and Quantum Electronics*, 34(5/6):523–540, June 2002.

[45] P. Török and T. Wilson. Rigorous theory for axial resolution in confocal microscopes. *Optics Communications*, 137(1–3):127–135, April 1997.

[46] M.R. Arnison and C.J.R. Sheppard. A 3d vectorial optical transfer function suitable for arbitrary pupil functions. *Optics Communications*, 211(1-6):53–63, October 2002.

[47] W. Liu and Kowarz M.W. Vector diffraction from subwavelength optical disk structures: Two-dimensional near-field profiles. *Optics Express*, 2(5):191–197, March 1998.

[48] C.D. Wright, P.W. Nutter, M.K. Loze, and S.D. Jepson. Computer simulation tools for the design and optimization of optical disk systems. *IEEE Transactions on Magnetics*, 32(4):3154–3164, July 1996.

[49] P.W. Nutter. *Image formation in the scanning optical microscope*. PhD thesis, The Manchester School of Engineering, 1997.

[50] R.H. Webb. Confocal optical microscopy. *Reports On Progress In Physics*, 59(3):427–471, March 1996.

[51] S.F. Pereira, J.A. de Pooter, M. de Haan, and J.J.M. Braat. Towards subwavelength optical detection using waveguides. In H.F. Schouten T.D. Visser, Daan Lenstra, editor, *Proceedings of the IEEE/LEOS Benelux Annual Symposium*, pages 28–31. IEEE/LEOS Benelux Chapter, December 2002.

[52] U. Brand, G. Hester, J. Grochmalicki, and R. Pike. Super-resolution in optical data storage. *Journal of optics a-pure and Applied Optics*, 1(Suppl. S):794–800, December 1999.

[53] J.T. Frohn, Knapp H.F., and A. Stemmer. True optical resolution beyond the rayleigh limit achieved by standing wave illumination. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 97(13):7232–7236, June 2000.

[54] T.D. Milster. New way to describe diffraction from optical disks. *Applied Optics*, 37(29):6878–6883, October 1998.

[55] S. Stallinga. Confocal detection and pupil masks for data storage on optical discs. In *Proceedings of EOS Topical Meeting on Advanced Imaging Techniques, Delft, 2003*, pages 44–51. European Optical Society, October 2003.

[56] M. Mansuripur and G. Sincerobox. Principles and techniques of optical data storage. *Proceedings of the IEEE*, 85(11):1780–1796, 1997.

[57] Mansuripur M. Dependence of capacity on media noise in data storage systems. *Japanese Journal Of Applied Physics Part 1-Regular Papers Short Notes & Review Papers*, 41(3B):1638–1642, March 2002.

[58] L. Soldano. *Multimode interference couplers: design and applications*. PhD thesis, Delft University of Technology, October 1994.

[59] L. Soldano and E. Pennings. Optical multi-mode interference devices base on self-imaging: Principles and applications. *Journal of Lightwave Technology*, 13(4):615–627, April 1995.

[60] M. Bachmann, P.A. Besse, and H. Melchior. General self-imaging properties in N X N multimode interference couples including phase-relations. *Applied Optics*, 33(18):3905–3911, June 1994.

[61] P.A. Besse, M. Bachmann, H. Melchior, L.B. Soldano, and M.K. Smit. Optical bandwidth and fabrication tolerances of multimode interference couplers. *Journal of Lightwave Technology*, 12(6):1004–1009, June 1994.

[62] J. Leuthold, J. Eckner, E. Gamper, P.A. Besse, and Melchior H. Multimode interference couplers for the conversion and combining of zero- and first-order modes. *Journal Of Lightwave Technology*, 16(7):1228–1239, July 1998.

[63] M. Rajarajan, B.M.A. Rahman, and K.T.V. Grattan. A rigorous comparison of the performance of directional couplers with multimode interference devices. *Journal of Lightwave Technology*, 17(2):243–248, February 1999.

[64] S.Y. Lee, S. Darmawan, C.W. Lee, and M.K. Chin. Transformation between directional couplers and multi-mode interferometers

based on ridge waveguides. *Optics Express*, 12(14):3079–3085, July 2004.

[65] A. Yariv. Coupled-mode theory for guided-wave optics. *IEEE Journal of Quantum Electronics*, 9(9):919–933, september 1973.

[66] A. Hardy and W. Streifer. Coupled mode theory of parallel waveguides. *Journal of Lightwave Technology*, 3(5):1135–1146, October 1985.

[67] E. Marom, O. Ramer, and S. Ruschin. Relation between normal-mode and coupled-mode analyses of parallel waveguides. *IEEE Journal of Quantum Electronics*, 20(12):1311–1319, December 1984.

[68] Q. Wang and S. He. A simple, fast and accurate method of designing directional couplers by evaluating the phase difference of local supermodes. *Journal of Optics A: Pure and Applied Optics*, 5(5):449–452, June 2003.

[69] W. Hellmich and P.P. Deimel. Optimal ar-coating for optical waveguide devices. *Journal of Lightwave Technology*, 10(4):469–476, April 1992.

[70] A. Leinse, C.G.H. Roeloffzen, K. Wrhoff, G. Sengo, R.M. de Ridder, and A. Driessen. Low loss fiber to chip connection system for telecommunication devices. In *Proceedings Symposium IEEE/Leos Benelux Chapter*, pages 185–187, December 2001.

[71] B. Liu, A. Shakouri, P. Abraham, and J.E. Bowers. A wavelength multiplexer using cascaded three-dimensional vertical couplers. *Applied Physics Letters*, 76(3):282–284, January 2000.

[72] D. Pasquariello and K. Hjort. Plasma-assisted inp-to-si low temperature wafer bonding. *IEEE Journal of Selected Topics in Quantum Electronics*, 8(1):118–131, January/February 2002.

[73] I. Christiaens, G. Roelkens, K. De Mesel, D. Van Thourhout, and R. Bockstaele. Thin film devices fabricated with bcb waferbonding. To be published in Journal of Lightwave Technology.

[74] S.M. Garner, S.S. Lee, V. Chuyanov, A.T. Chen, A. Yacoubian, W.H. Steier, and Dalton L.R. Three-dimensional integrated optics using polymers. *IEEE Journal of Quantum Electronics*, 35(8):1146–1155, August 1999.

[75] P. Koonath, K. Kishima, T. Indukuri, and B. Jalali. Sculpting of three-dimensional nano-optical structures in silicon. *Applied Physics Letters*, 83(24):4909–4911, December 2003.

[76] D. Khalil and A. Yehia. Two-dimensional multimode interference in integrated optical structures. *Journal of Optics A: Pure and Applied Optics*, 6(1):137–145, January 2004.

[77] S.H. Kim, Y. Yee, J. Choi, H. Kwon, M.H. Ha, C. Oh, and J.U. Bu. A micro optical flying head for a pcmcia-sized optical data storage. In *17th IEEE International Conference on Micro Electro Mechanical Systems*, pages 85–88, January 2004.