Een geïntegreerde InP-gebaseerde optische golfgeleiderisolator
op basis van ferromagnetische CoFe-contacten

An Integrated InP-Based Optical Waveguide Isolator
Using Ferromagnetic CoFe Contacts

Mathias Vanwolleghem

Promotor: prof. dr. ir. R. Baets
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Toegepaste Wetenschappen: Natuurkunde

Vakgroep Informatietechnologie
Voorzitter: prof. dr. ir. P. Lagasse
Faculteit Toegepaste Wetenschappen
Academiejaar 2004 - 2005

UNIVERSITEIT
GENT

Examencommissie:

Prof. dr. ir. R. Baets (promotor)	INTEC, Universiteit Gent
Dr. es sciences P. Beauvillain	IEF-CNRS, Université Paris-Sud
Prof. dr. ir. P. Bienstman	INTEC, Universiteit Gent
Prof. dr. ir. M. Burgelman	ELIS, Universiteit Gent
Prof. dr. ir. J. De Boeck	IMEC vzw, Leuven
Prof. dr. ir. F. Olyslager	INTEC, Universiteit Gent
Prof. dr. ir. D. Van Thourhout (secretaris)	INTEC, Universiteit Gent
Prof. dr. ir. R. Verhoeven (voorzitter)	Labo Hydraulica, Universiteit Gent

Universiteit Gent
Faculteit Toegepaste Wetenschappen

Vakgroep Informatietechnologie (INTEC)
Sint-Pietersnieuwstraat 41
B-9000 Gent
België

Tel.: +32-9-264.33.16
Fax: +32-9-264.35.93

Dit werk kwam tot stand in het kader van een specialisatiebeurs toegekend door het IWT (Vlaams Instituut voor de bevordering van het Wetenschappelijk-Technologisch Onderzoek in de Industrie).

ISBN 90-8578-013-6
NUR 959, 926
Wettelijk depot: D/2005/10.500/13

Voor André Vanwolleghem

Dankwoord

In de wetenschap dat deze bladzijden uit het voorliggend werk met een aan zekerheid grenzende waarschijnlijkheid het vaakst gelezen zullen worden, is het niet zonder enige zenuwachtigheid dat ik deze regels tik. Ik heb mij in de voorbije jaren vaak op een bevoorrechte manier omringd en gesteund geweten door een kleurrijke bende van familie, vrienden en collega's die allemaal op hun eigen manier en vaak onbewust een enorme betekenis gehad hebben in de succesvolle afwerking van dit doctoraatsonderzoek. Het zou jammer zijn mocht ik er nu een over het hoofd zien bij de woorden van oprechte dank die ze allemaal verdienen. Nog meer dan de vele stimulerende wetenschappelijke discussies die mij in Brownsiaanse beweging naar het einddoel stuurden, waren het vooral simpelweg hun doodgewone vriendschap van elke dag, hun oprechte interesse, en hun goedbedoelde plagerijen die mij bij dit werk onbewust gesteund en geholpen hebben. Laat hen dus elke letter een beetje als de hunne beschouwen.

Roel, promotor van dit werk, moet ik bedanken voor de wetenschappelijk uitdaging die hij mij toevertrouwde, en zijn blijvend geloof in een succesvolle realizatie van een gek Japans magneto-optisch ideetje. Als enige magneto-optieker in een groep fotonici was het soms moeilijk na een teleurstellend experiment de moed niet op te geven. Roel wist die twijfel altijd in vernieuwde energie om te buigen. Zijn gemotiveerde, en beschermende uitval naar een vreemde Duitse project officer spreekt in dit opzicht boekdelen. Mijn oprechte dank gaat eveneens uit naar Prof. Paul Lagasse die de structuur en het kader waarbinnen dit onderzoek werd gevoerd, heeft mogelijk gemaakt.

Binnen het INTEC team is een magneto-optieker een vreemde vogel. Toch kon ik steeds rekenen op de voortdurende steun en stimulerende discussies van alle collega-onderzoekers. Als partners-in-crime hebben ze allemaal wel bewust of onbewust op een of ander moment een stukje van de puzzel juist gelegd. Ik hoop dat ik bij hen soms hetzelfde gedaan heb. Een heel bijzondere plaats wordt ingenomen door Wouter. Mijn perfecte tegenpool, toonbeeld van rust en orde, wetenschapper pur-sang en begiftigd met een engelengeduld in de meetruimte, altijd bereid aan een meting te twijfelen en voor een zoveelste keer te controleren of we toch maar geen detail over het hoofd gezien hebben. Ik denk te mogen stellen dat we een, zij het bizar, maar zeer complementair team vormden, en dank hem oprecht voor zijn onvervangbare inzet voor het ISOLASER project (en het bijhouden van alle papieren en bestanden die ik telkens opnieuw verloor in mijn nochtans goed geordende kast en harde schijf). Ik ben er van overtuigd dat onder zijn vleugels het project recht op die geïntegreerde laser-isolator module afstevent. Het ga je goed.

Maarten Leys heeft een heel belangrijke rol gespeeld in dit onderzoek. Hij was niet alleen de eerste enthousiaste groeier van een tensile gestrainde well (ver voorbij alle grenzen van mechanische stabiliteit – ik vraag mij nog altijd af hoe hij het voor elkaar gekregen heeft) die ons eveneens in contact gebracht heeft met het team van Jean-Louis Gentner, maar ook een wijze, grijze mentor die mij zowel op wetenschappelijk en levensbeschouwelijk vlak als wat mijn kennis van de rock-and-roll geschiedenis betreft, voortdurend stimuleerde. Mocht hij ooit "*standing on the corner with a suitcase in his hand*", dan weet hij dat er op St-Amandsberg altijd een bord en een bed voor hem klaarstaat. Dries herinnert zich vermoedelijk nog mijn eerste dagen op INTEC toen ik samen met Jan voor hem met een tube colle twee chipkes probeerde aan elkaar te plakken. Niet dat het veel geholpen heeft denk ik, maar in elk geval heeft hij ondertussen ruim meer dan het dubbele teruggedaan voor mij. Zijn bureaudeur stond altijd open en meer dan eens is zijn whiteboard volgeklad door mij met allerlei dingen waar ik zonder hem niet uitgeraakt zou zijn. Bedankt voor je tijd, je geduld, en je vertrouwde rust. Een Rodenbach op jouw gezondheid. Geert, die als bureaugenoot van Dries vaak mijn vervelende vragen mee moest aanhoren, liet zich ook niet onbetuigd in de vele tips. En op de terugweg van bij Dries, was het altijd fijn even het hoofd binnen te steken bij het eiland van Ilse, Gunther, Doos, Wouter, en Hendrik. Ilse had altijd wel net dat ene boek dat je niet vond, en indien niet, dan toch dat luisterend oor bij een van mijn vele klaagzangen. Gunther net weer eens die geniale ingeving, Doos net weer eens die crack-code voor een of ander zeer legaal gedownload voetbalgame. En Hendrik! O Hendrik, stuurman van INTEC. Zonder Hendrik geen meetkamer, zonder Hendrik eeuwig op zoek naar zoek geraakte lensed fibres, GRIN lenzen, Newport manuals, ... Marko has also been around for quite some time in the famous office. He and Nerea have become true friends, who are always ready to help unconditionally (Yes, the wallpaper in our bedroom is still holding!). I wish them all the best, and hope to visit them soon as promised in Bask country. Vreemde wezens werden soms gesignaleerd in de gang van de 41. Ze bleken afkomstig te zijn uit een rijk dat weldra het lot van Atlantis beschoren is. De 39'ers waren een onmisbaar element van de dagelijkse sfeer op INTEC, al was het maar om te horen hoe goed hun kristallen en draden waren. Of om een gezellige discussie over zin en onzin van FoM's. Ik bedank ze allemaal voor hun plagerijen en hun wetenschappelijke weddenschappen. Twee van de opper-39'ers mogen hier niet onvernoemd blijven. C++ goeroe Peter bleef kalm mijn aanhoudende vragen over de interne keuken van de complexe rootsolver van CAMFR beantwoorden en gaf me mijn eerste smaakjes van de wonderen van OO programmering. Misschien komt die magneto-optische uitbreiding van CAMFR er toch nog eens. Lieven, de anders-globalist en python-debugger par excellence, was nooit te beroerd je een stukje van zijn eeuwige enthousiasme af te staan.

Van alle INTEC'ers is er één groep zonder dewelke dit werk nooit hetzelfde zou geweest zijn. De bewoners van de bureau in de gang achteraan rechts, vaak door de inwoners ervan in zwartere dagen beschouwd als het Radeau de la Meduse. Maar vooral het toneel van een gezellig samenhorigheidsgevoel gekruid met de nodige plagende woordenwisselingen. Langzaam druppelt deze lieu de création unique leeg. De Japanner is al een tijdje verdwenen, maar de geest van zijn

wetenschappelijk genie hangt er nog wel rond, af en toe inspirerend tot grote inzichten. De mère poule, die af en toe haar jongens ter orde moest roepen en verzoenend optreden, wil ik graag bedanken voor alles. Haar ruggesteun was vaak onmisbaar. Ik wens haar het allerbeste in haar nieuwe professionele uitdaging, en hoop het beste voor haar, haar Groen!-minnende spin doctor, en haar twee schatten van dochters. De liefhebber van moderne architectuur is ook al een tijdje verdwenen. Maar uiteindelijk was het hij die mij INTEC leerde kennen. Het was leuk met hem op de bureau discussies te kunnen voeren over het belang van het verticalisme in de architectuur van het interbellum. De liefhebber van de betere muziek en de talentvolle handballer met een voorliefde voor Noorse volleybalsters blijven er voorlopig nog een tijdje rondhangen. Ik hoop dat ze het voor de nieuwkomers Jonathan en Frederik, even onvergetelijk maken als het voor mij was. Ik zal de impromptue vrijdagavonduitspattingen met hen missen, en vooral hun nauwgezette controle der aankomst- en vertrekuren van treinen en collega's. Ik wens Franky aka Franky een schitterend doctoraat toe en de "amballer" evenzeer. See you at le Grand Rex.

Laat ik vooral niet vergeten Steven, Liesbet en Steven te bedanken. Zonder hun technologie-expertise was er nooit sprake geweest van een isolator. Bedankt, jullie hebben uiteindelijk die magnetische lasertjes (zoals Steven ze zo graag noemde) gemaakt! Karien, Bernadette, Ilse, Ilse en Isabelle: als het niet voor hulp bij een of andere schandelijk excessieve kostennota was, dan was het altijd leuk met jullie gezellig te roddelen . . . enkel over goede dingen uiteraard. Kristien! Natuurlijk was ik je niet vergeten. Kan je snel even naar de printer kijken straks? Bedankt voor je altijd bereidwillige bijstand. Eddy en Luc zijn net zo onmisbaar als Hendrik in de dagelijkse werking van INTEC. Bedankt voor de zoektocht naar die ene moeilijk te vinden paper. Bedankt voor het geduldig uitgevoerd maatwerk. Vakmanschap waar geen enkel labo zonder mee kan.

ISOLASER deed mij eventjes in IMEC belanden. En alhoewel ik telkens weer opzag tegen "het stapvoets verkeer op de Brusselse binnenring ter hoogte van het verkeersplein in Zaventem", was het aangenaam om aan te komen bij Jo DB, Jo D, Liesbet, Roel, Wouter en Pol. Dank je allemaal voor de hulp en de warme ontvangst en Roel in het bijzonder voor je geduldig sputteren en Liesbet voor je vele tips en tricks bij de MOKE experimenten. En plus ISOLASER m'a donné l'opportunité de rencontrer Bruno, Francois et Béatrice d'Alcatel OPTO+. C'est grâce à eux que le démonstrateur ISOLASER a été réalisé. J'ai été impressionné par leur travail sur la croissance epitaxiale des puits quantiques sous tension. Je les remercie pour leurs efforts technologiques. Et je les souhaite bonne chance. Pierre, Philippe, Bernard et Antoine de l'IEF au centre UPSud à Orsay sont entre temps devenu des collègues sympa soit patron strict mais juste. (Je n'oserais pas rigoler sur ce dernier point-là.) Je leur dois beaucoup de remerciements pour leurs expériences magnéto-optiques. Mais encore plus pour leur acceuil chaleureux et serviable pendant mes premiers jours à Paris, il y a déjà un mois maintenant.

De voorbije jaren heb ik gelukkig niet enkel doorgebracht in de gangen van INTEC! Alhoewel ik hoop vele van de INTEC-ers ook tot mijn vriend(-inn)en te mogen rekenen, toch waren er die bijzondere mensen rondom mij op wie ik altijd kon rekenen en zonder wie het extra-universitaire leven niet hetzelfde was geweest. Ik

weet dat zij de volgende regels aandachtig lezen, dus laat ik beginen met te zeggen dat jullie natuurlijk allemaal op de eerste plaats horen te staan. Vriend sedert 28 jaar, Thomas, tevens broer, bedankt voor je medische zorgen in tijden van lage rugpijn. De jaren in de Martelaarslaan waren onvergetelijk. Aline, je hebt enkel de laatste (lange) rechte lijn van dit werk meegemaakt, maar je oprechte steun samen met Thomas is mij niet ontgaan. Ik hoop dat jij en Thomas een leven vol amour en plezier tegemoet gaan. Job, partner-in-many-crimes, tevens plichtsbewuste getuige, ik vermoed dat je geen reet begrepen hebt van wat ik uitgespookt heb in de voorbije jaren, maar dat is niet belangrijk. Je bereidwillige wekelijkse afslachting op het squashcourt was een fantastische mentale opsteker. Ik wist wel dat je het enkel voor mijn doctoraat deed. Riikka, aangezien je man geen Fins wil leren, zal ik je maar in je moedertaal bedanken. Riikka, kiitokset sinun aina niin hyvien naurujen säestämistä suomalaisista elämänviisauksista, teidän kulinaarisista hemmotteluhetkistänne ja sinun hollannin kielen diminutiivin innovaatisista tulkinnoista. Muuten, haluatko kuulla mikä on mun pointti?
Wollietje, ik zal augustus 2003 met jou en Kathleen nooit vergeten. Hopelijk, komen er nog veel dergelijke avonturen. Je hebt heel veel betekend voor mij, en ook al kan ik niet meer in de Coupure Derby's komen drinken, ik denk dat ze in de Brabantdam nog beter zullen smaken. Leroytje, ons Martelaarsavontuur, eerst met Thomas en daarna met Kathleen erbij, was onvergetelijk. Je eeuwig vrolijke bezoekjes samen met Leuven waren een meer dan welkome afwisseling. Leuven, single malts en een spelletje schaak at any time. Thomas, jonge Corbijn, misschien was op een manier jouw mislukt Ringo avontuur net het butterfly effect dat mij aan dit doctoraat deed beginnen? Woooow, far out. Het was leuk jou en later ook Heidi erbij te hebben. Merci om mij van mijn betere kant te portretteren bij allerhande gelegenheden. Mathieu, jouw zo voorspelbare (eet?)bezoekjes en openingskeuzes waren een leuk steunpunt tijdens de lange schrijfperiode. Bedankt dat jij – toekomstige IM – je af en toe eens van het bord liet spelen tijdens onze blitz-sessies. Olaf, Karine, Klaartje, en Marieke, de zomerse Walprije-avonden waren een ideale ontspanning. Bedankt om mij zo nu en dan te verstrooien met discussies over de grondstoffen van het eiland Catan. Bart en An, creatieve schone broer en ditto zus (?), de zaterdagen en zondagen in Gent, Leuven of Zonhoven waren altijd gezellig. Bedankt voor jullie regietalent.
Een groot woord van dank gaat uit naar mijn lieve familie die mij, zij het soms vanop enige afstand, steeds bemoedigend steunden. Of het nu de vele kaarsen van meter en opa waren, de geïnteresseerde telefoongesprekken met meter, de rust en kalmte en de gezellige monologen van oma, de rijsttaarten van tante Henriette, of de oprechte interesse in de wereld van de glasvezel van tantes en nonkels, neven en nichten, allemaal hebben ze op een manier steun betekend. Miems en pieuws, schoonouders die elk daarover bestaand cliché ontkrachtten, wat was ik geweest zonder de herbronnende rust en kalmte die ik bij jullie telkens weer kan vinden? Kipkap en smurfenvlaai forever! Pieuws, bedankt voor je arendsblik bij het nalezen.
Moeke en papa, het is klaar! Ik wil jullie duizendmaal bedanken voor alle kansen die ik van jullie gekregen heb, voor jullie vele geduld met mijn soms onvoorspelbare wispelturigheid, en voor jullie onvoorwaardeijke steun en liefde.

En Soessie, mijn lieve vrouw met je ogen zo blauw, er zijn geen passende woorden die ook maar van ver kunnen beschrijven wat je de voorbije zeven jaar allemaal betekend hebt voor mij. Dank je voor alles. Koessen in aeternum.

En nu, . . . sampanje !!

Mathias Vanwolleghem, Parijs, 20 februari 2005

Contents

"Au départ, l'art du puzzle semble un art bref, un art mince, tout entier contenu dans un maigre enseignement de la Gestaltttheorie: l'objet visé – qu'il s'agisse d'un acte perceptif, d'un apprentissage, d'un système physiologique ou, dans le cas qui nous occupe, d'un puzzle de bois – n'est pas une somme d'éléments qu'il faudrait d'abord isoler et analyser, mais un ensemble, c'est-à-dire une forme, une structure: l'élément ne préexiste pas à l'ensemble, il n'est ni plus immédiat ni plus ancien, ce ne sont pas les éléments qui déterminent l'ensemble, mais l'ensemble qui détermine les éléments: la connaissance du tout et de ses lois, de l'ensemble et de sa structure, ne saurait être déduite de la connaissance séparée des parties qui le composent: cela veut dire qu'on peut regarder une pièce d'un puzzle pendant trois jours et croire tout savoir de sa configuration et de sa couleur sans avoir le moins du monde avancé: seule compte la possibilité de relier cette pièce à d'autres pièces, et en ce sens il y a quelque chose de commun entre l'art du puzzle et l'art du go; seules les pièces rassembleés prendront un caractère lisible, prendront un sens: considerée isolément une pièce d'un puzzle ne veut rien dire; elle est seulement question impossible, défi opaque; mais à peine a-t-on réussi, au terme de plusieurs minutes d'essais et d'erreurs, ou en une demi-seconde prodigieusement inspirée, à la connecter à l'une de ses voisines, que la pièce disparaît, cesse d'exister en tant que pièce: l'intense difficulté qui a précédé ce rapprochement, et que le mot *puzzle* – énigme – désigne si bien en anglais, non seulement n'a plus de raison d'être, mais semble n'en avoir jamais eu, tant elle est devenue évidence: les deux pièces miraculeusement réunies n'en font plus qu'une, à son tour source d'erreur, d'hésitation, de désarroi et d'attente.
(. . .)"

uit *La vie – Mode d'emploi*
Georges Perec

Nederlandstalige Samenvatting

Nederlandstalige samenvatting

IN de volgende bladzijden wordt een beknopt overzicht gegeven van de belangrijkste verwezenlijkingen en vaststellingen van het voorliggend proefschrift. Verwijzing naar vergelijkingen en figuren gebeurt door middel van de nummering gevolgd in de engelstalige tekst. Voor verdere details en achtergrondinformatie verwijzen we uiteraard naar de engelstalige tekst.

1 Inleiding

1.1 Context

Sedert de beruchte "bubble-burst" van de technologie-aandelen aan het eind van het vorig millennium, is er een aanhoudende druk op de fabrikanten van optische telecom-componenten om de verkoopprijzen te drukken, vaak enkel om te overleven als bedrijf. Die situatie is uiteraard niet houdbaar als de kost van het produktieproces niet volgt. Daarenboven is er uiteraard de simpele economische logica dat wie eenzelfde kwalitatief hoogstaand produkt kan aanbieden aan een lagere prijs een veel groter potentiële clientèle zal bereiken. Het natte droomscenario van de optische telecomindustrie optische vezel naar elke huisdeur te brengen zal zich nooit verwezenlijken als de prijzen van optische metro-componenten niet sterk gereduceerd worden. Bovendien is het extra relevant een dergelijke prijsreductie na te streven specifiek voor metropolitan optische componenten. Niet alleen is dit marktaandeel het grootst, recente marktanalyses tonen aan dat sinds 2000, het vooral dit segment is dat het minst geraakt werd door de technologiecrash. Veruit de belangrijkste metro-component is de transceiver, en meer in het bijzonder de 1300nm laserbron.

De kostprijs van laserdiodes wordt voornamelijk bepaald door het verpakkingsproces omdat de nodige optische alignatie van de subcomponenten een parallelle processing verhindert. Een laserverpakking bevat standaard al snel vier optische "componenten" die gealigneerd moeten worden met micron precisie: de laser zelf, een monitor photodiode, een optische isolator, vaak een externe modulator, en de "deur naar de buitenwereld", het facet van de optische vezel. Elke extra component vergt een hoge apertuurlens en een tijdrovende actieve aligneringsstap. Terwijl de individuele prijs van de componenten vaak niet hoger is dan een tiental €, kan de totale prijs van een laser oplopen tot enkele 1000-en €'s als gevolg van het verpakkingsproces. Prijsreductie moet dus gebaseerd zijn op een vereenvoudiging of een kostenbesparende aanpassing van de verpakking en/of de verpakkingsprocedure.

Dit kan gebeuren door rechttoe rechtaan de laserverpakking en/of de procedure zelf te vereenvoudigen door bijvoorbeeld componenten als de modulator of de isolator weg te laten, of de aligning passief uit te voeren, de vezel ongealigneerd rechtstreeks tegen de laser aan te plaatsen in een daartoe voorzien groef,... Eleganter is echter de functionaliteiten van de verschillende subcomponenten te integreren in een enkele component. Voorbeelden van integratie zijn legio. Men onderscheidt hybride en monolithische integratie. Het onderscheid is soms vaag, maar hier wordt het gedefinieerd als de grens tussen het realiseren van de subcomponenten op een gezamenlijk substraat in een enkele (vaak complexe) groeistap (monolithisch), en het afzonderlijk groeien van de subcomponenten die nadien op een gezamenlijke "drager" permanent met elkaar verbonden worden (hybride).

Integratie en vooral monolithische integratie is om logische redenen het ultieme doel. Zonder functionaliteit op te offeren, bereikt men een component die kleiner is, mechanisch en thermisch stabieler is, nauwelijks gehinderd wordt door aligneringsproblemen, parallelle processing toelaat en vooral goedkoper is.

Tot op heden ontsnapt een enkel type optische functionaliteit aan de mogelijkheid tot integratie, namelijk optische niet-reciproke componenten, zoals optische isolatoren. Een ideale optische isolator is ruwweg gesteld het optische equivalent van een mechanische terugslagklep: terugkerend licht wordt ongeacht zijn polarizatie, zijn intensiteit of zijn golflengte "tegengehouden", terwijl voorwaarts propagarend licht verliesloos doorgelaten wordt. Optische isolatoren zijn onontbeerlijke componenten voor de bescherming van alle type componenten die gevoelig zijn aan toevallige reflecties. Een laser is het voorbeeld bij uitstek. Variaties in de fazeresonantievoorwaarde ten gevolge van reflectie doen de golflengte en het geëmitteerde vermogen fluctueren. Niet enkel lasers maar elke actieve optische component is gevoelig aan reflecties omwille van de licht-drager interactie in dergelijke componenten. Aangezien optische actieve halfgeleiderversterkers verwacht worden een toenemend belangrijke rol te zullen spelen in allerlei volledig optische signaalverwerkende circuits, is het nut van een geïntegreerde optische isolator evident. In het bijzonder voor een 1300nm laserbron zou een dergelijke component een enorme impact kunnen hebben op de prijs, omdat bij die golflengte de optische vezel een minimale dispersie vertoont, en de laser dus geen behoefte heeft aan een extern laag-dispersie-modulatiemechanisme (tenminste niet bij de typische afstanden bij metro-toepassingen). Dit maakt dat voor dergelijke direct-moduleerbare lasers de isolator de enige niet-geïntegreerde component is.

1.2 Magneto-optica

Optische niet-reciprociteit vereist onvermijdelijk de aanwezigheid van een permanent magneetveld. De aanwezigheid van een extern magneetveld breekt de bekende inversiesymmetrie van de Maxwell vergelijkingen (Vgl. (1.1)). Dit brengt ons daarmee in het domein van magneto-optica, of met andere woorden de optische fenomenen die optreden bij de interactie van licht met een gemagnetizeerd medium.

1.2.1 Beschrijving van magneto-optische media

De oorsprong van alle magneto-optische fenomenen kan op een microscopisch niveau verklaard worden als het gevolg van de Zeeman-splitting van de energieniveaus. Kwantummechanische transitieberekeningen tonen dan aan dat de symmetrie ten opzichte van linkshandige en rechtshandige fotonen gebroken is en dat dus de polarizeerbaarheid van het medium gyrale bijdragen krijgt. Dit vertaalt zich rechtstreeks naar de macroscopische constitutieve relaties van een gemagnetiseerd medium door de aanwezigheid van niet-diagonale anti-symetrische elementen in de magnetische permeabiliteitstensor $\underline{\mu}$ en de diëlektrische tensor $\underline{\epsilon}$ (zie (1.2)). De precieze symmetrie-eigenschappen worden volledig bepaald door de rotationele symmetrie rond de aanwezige magnetizatie en door het principe van Onsager (en in kristallijne materialen door de kristalsymmetrie). Bij optische frequenties wordt de $\underline{\mu}$-tensor doorgaans gelijk aan de vacuum permeabiliteit wegens de afwezigheid van Larmorprecessie. Magneto-optische fenomenen zitten dus volledig vervat in de niet-diagonaal elementen van $\underline{\epsilon}$. Deze zogenaamde elektrogyrotropie wordt beschreven door een element van de vorm ig met g de complexe gyrotropie constante. In magneto-optica zijn tekenconventies van nog groter belang dan in isotrope optica, gelet op de twee extra keuzemogelijkheden voor de tekens in enerzijds de complexe g en anderzijds in de antisymmetrische elementen van $\underline{\epsilon}$. We verwijzen hiervoor naar de engelstalige tekst.

Elk magneto-optisch effect is in essentie het gevolg van de gyrotrope anisotropie van het MO medium. Het gevolg is dat de ontaarding van de twee fundamentele polarizaties van het licht opgeheven wordt, en dat bovendien op een manier die afhankelijk is van de propagatierichting van het licht relatief ten opzichte van de magnetizatie. In het algemeen zijn de zogenaamde normale modes of eigenpolarizaties resp. links- en rechtshandig elliptisch gepolariseerd. Deze polarizaties en hun propagatieconstante worden gevonden als oplossing van de Fresnelvergelijking (zie Vgl. (1.5)). Er zijn twee bijzondere configuraties waarbij de polarizatie van de normale modes eenvoudig wordt.
In de Faraday geometrie ($\mathbf{k} \parallel \mathbf{M}$) zijn de normale modi links- en rechtshandig circulair gepolariseerd (en hebben in eerste orde een verschil in effective index van g/n). In de Voigt of Cotton-Mouton geometrie ($\mathbf{k} \perp \mathbf{M}$) behouden de normale modi hun lineaire polarizatie (i.e. een TM polarizatie met $\mathbf{H} \parallel \mathbf{M}$ en een TE polarizatie met $\mathbf{E} \parallel \mathbf{M}$), en krijgt enkel de TM mode een MO correctie: namelijk van tweede orde in g op de effectieve index en een longitudinale veldcomponent van eerste orde in g. De Faradaygeometrie is dus niet-reciprook in transmissie, terwijl de Voigt geometrie enkel niet-reciprook is in reflectie (tengevolge van de eerste orde longitudinale veldcomponent die niet-reciproke randvoorwaarden veroorzaakt).

1.2.2 Fenomenologie van magneto-optica

MO effecten worden over het algemeen ingedeeld in transmissie- en reflectie-effecten. De fenomenologie van elk effect wordt eenvoudig verklaard door het invallend signaal te ontbinden in de eigenpolarizaties van het medium (rekening houdend met de propagatierichting). Elke eigenpolarizatie propageert met zijn typische snelheid, of wordt gereflecteerd met zijn typische reflectiecoëfficiënt. Als

gevolg van de faze- en amplitudeverschillen tussen de eigenpolarizaties bij deze transmissie- en reflectiefenomenen, zal de polarizatietoestand van het invallend signaal wijzigen. Elk magneto-optisch fenomeen verklaart zich dus door het ontstaan van MO correcties op de elementen van de Jones matrix van het transmissiemedium of de reflecterende interface. Deze MO correcties manifesteren zich niet enkel in het optreden van niet-diagonale elementen maar ook in correcties op de diagonale elementen. Zo ontstaan er bijvoorbeeld bij transmissie in de Voigt geometrie geen niet-diagonaal elementen in de Jones matrix (want de eigenpolarizaties in Voigt geometrie zijn nog steeds de lineare TE en TM polarizaties), maar wel een verschil in de propagatiefactor op de diagonale positie (die in isotrope media perfect gelijk zijn). Voor een meer gedetaileerde en kwantitatieve beschrijving van magneto-optische fenomenen verwijzen we naar de engelstalige tekst en naar het excellente standaardwerk van Zvezdin [1]. De behandeling van MO transmissie- en reflectiefenomenen gaande van enkelvoudige interfaces tot complexe multilagenstacks met arbitrair gemagnetiseerde MO materialen kent een uitgebreide literatuur. De lezer vindt uitstekende behandelingen in [2–4].

1.2.3 Magneto-optische materialen

De courant gebruikte materialen in magneto-optische toepassingen zijn de ferromagnetische metalen en hun legeringen, en ferrimagnetische granaten.
Ferromagnetische metalen zijn historisch gezien het eerst aangewend bij de studie van MO data-opslag, en dat omwille van hun uitstekende MO sterkte in reflectie en hun relatief eenvoudige depositie op een variëteit aan substraten met een relatief goede controle over hun magnetische anisotropie eigenschappen. De Kerr polarizatierotatie (en -ellipticiteit) kan significant variëren van legering tot legering, maar is over het algemeen van de orde van 1°. Ferromagnetische metalen geven van alle MO materialen de grootste MO reflectie-effecten, maar worden dermate gehinderd door hun hoge optische absorptie (orde $10^5 \mathrm{cm}^{-1}$) dat ze nog maar moeilijk hun weg gevonden hebben naar geïntegreerde magneto-optische circuits (alhoewel dit werk daar verandering probeert in te brengen).
Ferrimagnetische granaten (molecuulformule:(RE)FeO_3, RE:rare-earth element) daarentegen zijn spinel-achtige materialen met de unieke eigenschap dat ze een relatief hoge Faradayrotatie (tot 1000-3000° cm^{-1}) combineren met een quasi-transparentie ($<50\mathrm{cm}^{-1}$) in de IR band. Dat maakt van hen ideale kandidaten voor de realizatie van geïntegreerde magneto-optische circuits. Er is omwille van die eigenschappen een heel grote technologische expertise ontwikkeld over de jaren op het gebied van epitaxiaalgroei van deze materialen en het engineeren van hun magneto-optische sterkte en magnetische anisotropie door gepaste substitutie van de ijzer-ionen en de vreemde-aard-ionen door andere (niet-)magnetische ionen. De belangrijkste beperking van granaten is dat ze enkel gegroeid kunnen worden op andere granaatsubstraten, dat ze een relatief grote temperatuursgevoeligheid kennen (omwille van hun ferrimagnetische karakter), en dat ze zich niet eenvoudig een planaire magnetische anisotropie laten opdringen (waardoor ze onvermijdelijk een externe magneet nodig hebben om in Faraday of transversale Kerr geometrie gebruikt te kunnen worden). In het algemeen geldt dat ferrimag-

netische granaten op hun eentje verantwoordelijk zijn voor het ontstaan van de onderzoekstak geïntegreerde magneto-optica, maar dat hun MO sterkte op zijn minst een orde lager is dan die van metalen en dat hun integratie met belangrijke halfgeleider platforms voorlopig zorgen baart.

1.3 Optische isolatoren

1.3.1 Vrije ruimte optische isolatoren – specificaties

Het principe van een vrije ruimte optische isolator is genoegzaam gekend (zie Fig.1.2). Een niet-reciproke 45° Faraday rotator geplaatst tussen polarizatoren die exact 45° ten opzichte van elkaar gedraaid zijn is de basis polarizatieahankelijke uitvoering. Een polarizatie-onafhankelijke versie is geschetst in Fig. 1.3. Vrije ruimte isolatoren zijn commercieel ingeburgerde componenten (typische kostprijs basisversie: 100–200€). Ze zijn beschikbaar voor een ruim golflengtegebied en bereiken typische isolatieniveaus van 40 à 50dB voor een typische golflengteband van 20nm rond de centrale golflengte in een verpakking van enkele cm's. Tabel 1.1 somt typische specificaties op van polarizatie-onafhankelijke optische isolatoren.

1.3.2 Geïntegreerde optische golfgeleiderisolatoren

Pogingen tot een planaire golfgeleiderversie van een optische isolator zijn bijna even oud als geïntegreerde optica zelf.

Niet-reciproke TE/TM koppelaars

De eerste theoretisch gerapporteerde voorstellen (1972) waren logischerwijze de golfgeleiderequivalenten van de succesvolle bulk Faraday rotator. De eerste succesvolle experimentele uitvoering van een dergelijke golfgeleider Faraday isolator werd echter pas gerapporteerd in 1988 (met slechts een isolatie van 13dB) [5]! De reden hiervoor is vrij eenvoudig. De niet-reciprook gekoppelde TE en TM modi zijn in planaire golfgeleiders moeilijk in faze-overeenstemming te krijgen. En er bestaan in een golfgeleider geen modi met een polarizatie die 45° t.o.v. elkaar getilt is. Een golfgeleiderversie van een Faradayrotator moet dus "in serie" geplaatst worden met een reciproke rotator (in golfgeleiderversie). Alhoewel later nog verbeteringen werden gerapporteerd (bv. [6]), bleef het fundamentele probleem bestaan van heel lage fabricatietolerantie en van het gebrek aan integreerbaarheid wegens de nood aan een golfgeleiderversie van polarizatoren. Dit deed de onderzoekers naar alternatieven zoeken.

Niet-reciproke faze shifters

Bij het gebruik van het transversale Kerr effect, dat een niet-reciproke correctie op de propagatieconstante van TM modi veroorzaakt (of fazeshift), kan een isolator gerealizeerd worden door een niet-reciproke Mach-Zehnder interferometer (MZI) (zie Fig. 1.7), die leidt tot destructieve interferentie in de achterwaartse richting. Een dergelijke layout is veel toleranter in fabricatie want werkt in een enkele golfgeleidermode (en behoeft dus geen faze-overeenstemming). Bovendien wordt de isolatie bekomen door destructieve interferentie en is er dus geen behoefte aan golfgeleiderpolarizatoren. Een groot voordeel is ook dat deze component geen

YIG[1] geleidende laag behoeft (aangezien het Kerr effect een reflectie-effect is), maar kan uitgevoerd worden met een halfgeleider kernlaag. Bovendien kan de niet-reciproke fazeshift in theorie ook geïntroduceerd worden voor de TE mode. Belangrijkste nadeel is echter dat het Kerr effect over het algemeen zwakker is dan het Faraday effect. De tot nu toe best gerapporteerde TM niet-reciproke MZI bereikt slechts een isolatie van 20dB (in een lengte van meer dan 1cm) [7].

1.3.3 Integratie van ferrimagnetische granaten op standaard halfgeleiderplatforms

Het onderzoek naar granaat-gebaseerde geïntegreerde optische golfgeleiderisolatoren bevond zich bij het begin van het nieuwe millennium ietwat in een crisissituatie. Uiteindelijk is het eindelijke doel integratie met een halfgeleiderplatform, maar de beste layout (Faraday rotator) bleek hoegenaamd niet integreerbaar. Een niet-reciproke MZI kan uitgevoerd worden zonder polarizatoren en is dus wel degelijk integreerbaar, maar leidt tot veel grotere componenten en heeft behoefte aan een redelijk ingewikkelde magnetizatie over die grote afstand. Toch zijn inmiddels pogingen gestart tot integratie. Tot op heden is de meeste vooruitgang geboekt door de groep van Yokoi. Zij maken gebruik van een MZI uitgevoerd in InGaAsP met wafergebonde YIG-mantellagen op de armen van de interferometer, en een actief-passief integratie door oppervlakte-selectieve-epitaxie. Dit laat een semi-monolithische integratie met een laser toe (zie Fig. 1.9). Hun werk evolueert gestaag, en kende reeds verscheidene optimalizaties. Succesvolle bonding werd reeds gerapporteerd [8], net als zeer recent ook de actief-passief integratie [9]. Maar een werkzame geïntegreerde module heeft tot op heden niet het daglicht gezien.

Tabel 1.2 geeft een helder overzicht van de voor-, nadelen en bereikte resultaten in granaat-gebaseerde planaire golfgeleiderisolatoren.

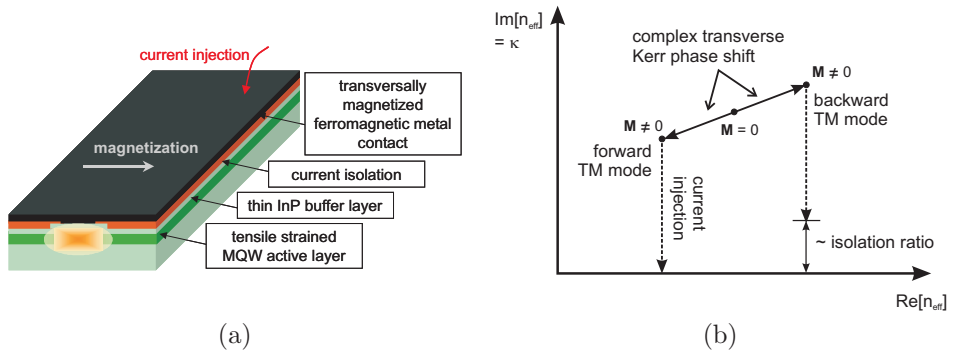
1.4 Innovatief concept voor een geïntegreerde optische isolator: de ferromagnetisch-metaal-gecontacteerde optische halfgeleiderversterker

De moeizame vooruitgang na 25 jaar onderzoek naar een geïntegreerde optische isolator, deden Ando en Nakano aan het einde van de vorige eeuw quasi gelijktijdig op een nieuw idee komen [10, 11]. Een Kerr faze shifter is duidelijk beter qua integreerbaarheid maar te zwak indien uitgevoerd in granaten en heeft bovendien wafer-bonding en passief-actief integratie nodig. Waarom niet het veel sterker Kerr effect benutten van een ferromagnetisch metaal?

1.4.1 Principe en voordelen

Het principe is geschetst in Fig. 1. De kernidee van Nakano's concept is dat de niet-reciproke Kerr shift ten gevolge van een transversaal gemagnetizeerd fer-

[1] YIG staat voor Yttrium Iron Garnet, het populairste ferrimagnetische granaatmateriaal.



Figuur 1: (a): Schematische voorstelling van Nakano's en Ando's innovatief geïntegreerd isolator concept; en (b): zijn werkingsprincipe.

romagnetische metaal door zijn verlieshebbend karakter een niet verwaarloosbaar imaginair deel krijgt, i.e. het modaal verlies van de TM modi wordt niet-reciprook.

Als een standaard III-V halfgeleider voorzien wordt van een dergelijke transversaal gemagnetiseerde metallische mantellaag, ontstaat er een propagatierichting met een lager modaal verlies. Door de golfgeleider actief uit te voeren als een halfgeleiderversterker en tegelijk het metaal als injecterend Ohms contact aan te wenden kan het residuele verlies in die richting gecompenseerd worden, terwijl het modale verlies in de andere richting gelijk blijft aan tweemaal het imaginaire deel van de Kerr shift. Een elegant isolatorconcept dat duidelijk vele van de bestaande problemen van granaat-gebaseerde Kerr fazeshifters oplost.

1. De component heeft geen behoefte aan een extern magneetveld wegens de ferromagnetische eigenschappen van de magneto-optische laag.

2. De imaginaire Kerr faze shift van een ferromagnetische metaal is over het algemeen een grootte-orde sterker dan de zuiver reëele fazeshift van ijzergranaten, wat tot kortere devicelengte leidt.

3. Het hier geschetste concept behoeft geen interferometrische layout. De isolatie-extinctie is een rechtstreekse modale eigenschap. Het gedrag van deze component als functie van golflengte wordt volledig bepaald door de golflengte-afhankelijkheid van de versterker en van het magneto-optisch effect van de metaallaag. Dit in tegenstelling tot het gedrag bij een niet-reciproke YIG-MZI. Bovendien kan, door te werken met een polarizatieselectieve versterkende laag, deze component bidirectioneel niet-transparant uitgevoerd worden voor TE modi. Dit in tegenstelling tot YIG-componenten waar in principe gereflecteerd TE licht "ongehinderd" doorgelaten wordt.

4. Het grootste voordeel van het Nakano concept ligt in de eenvoudige integreerbaarheid. Ferromagnetische metalen kunnen redelijk "eenvoudig" met goede magnetische eigenschappen gedeponeerd worden op standaard III-V

halfgeleiders. Belangrijker nog is dat de isolator zelf een actieve device is, zodat er geen behoefte meer is aan actief-passief integratie. Het device is in essentie een niet-reciproke halfgeleiderversterker die rechtstreeks monolitisch geïntegreerd kan worden met een laser met dezelfde versterkende laag.

In hun oorspronkelijke berekeningen voorspelden Nakano en Ando een isolerend vermogen (i.e. imaginaire Kerr fazeshift) van ca. 150dB/cm bij een materiaalwinst van 1500–2000/cm voor de compensatie van het voorwaarts verlies. De door hun voorgestelde structuren zijn verre van optimaal, maar evenmin onrealistisch, net als hun bekomen waarden voor de materiaalwinst. Beide papers verdienen dus wel degelijk dieper bestudeerd te worden.

1.4.2 Uitdagingen

Toch is het voorstel van Ando en Nakano niet simpelweg een rechttoe rechtaan te volgen recept voor een eenvoudige integreerbare golfgeleiderisolator. Er zijn een aantal veronderstellingen aan verbonden waarvan de evidentie niet op voorhand vaststaat. Deze uitdagingen zijn geïllustreerd in Fig. 1.12.
Aan de ene kant is er behoefte aan een performant ferromagnetisch metaal-halfgeleidercontact, met een goed Ohms gedrag, een voldoende sterke magneto-optische parameter bij de gebruikte golflengte, en een geschikte planaire magnetische anisotropie. Deze eigenschappen zijn slechts in geringe mate eerder experimenteel aangetoond. Er zijn geen bekende voorbeelden van Ohmse contacten met behulp van ferromagnetische metalen. Bovendien zijn de magneto-optische eigenschappen van die metalen nauwelijks gekend in het telecomgolflengtegebied. Enkel de magnetische anisotropie van dunne film ferromagneten is reeds onderzocht. Maar er moet onderlijnd worden dat in de hier beschouwde specifieke vorm (lange smalle contacten) de vormanisotropie niet gering zal zijn.
Daarnaast behoeft de component een actieve laag die zeer hoge TM polarizatieselectieve materiaalwinst kan leveren. Als de spontane TE emissiefactor onderdrukt wordt kan het device absorberend gehouden worden voor TE licht en wordt er bovendien belet dat TE licht ladingsdragers verbruikt, en/of ruis toevoegt aan het signaal. Het is gekend dat dit in theorie kan door de QW's onder roostermisaanpassing te groeien, namelijk met een kleine trekspanning. Het gebruik van trekspanning in een QW versterker is expermenteel minder onderzocht dan het gebruik van drukspanning. De juiste limieten van materiaalwinst zijn dus nog niet gekend.

1.5 Doelstelling en opbouw van dit werk

Het doel van dit werk is op de eerste plaats een eerste wereldwijd bewijs te leveren van de haalbaarheid van het theoretisch concept van Ando en Nakano. Daarbij wordt het ultieme doel van integratie met een DFB laser wel in het achterhoofd gehouden maar niet primordiaal gesteld, gelet op de hierboven vermelde uitdagingen die zich nog stellen. Om dezelfde reden wordt er geen specifiek getal voor de te bereiken isolatie vooropgesteld, alhoewel er moet gezegd zijn dat de component geen zin heeft als die niet op zijn minst 25dB haalt bij een benodigde

stroomdichtheid van niet veel meer dan 5kA/cm^2. We beperken ons hierbij tot de metro telecomgolflengte van 1300nm, om de redenen aangehaald in de inleiding. Er kunnen drie subtaken geformuleerd worden.

- *Het bestuderen en optimalizeren van het ferromagnetische metaal/halfgeleidercontact, zowel wat betreft optische, magneto-optische (bij 1300nm) en magnetische eigenschappen.*
- *Het bestuderen en karakterizeren van actieve TM polarizatieselectieve halfgeleiderlagen (bij 1300nm)*
- *Het ontwerpen, fabriceren en karakterizeren van een ferromagnetisch-metaalgecontacteerde optische halfgeleiderversterker met het doel als geïntegreerde optische golfgeleiderisolator te gebruiken*

Dit werk is zeer sequentieel opgebouwd. In Hoofdstuk 2 wordt een grondige studie gemaakt van de modale modellering van niet-reciproke magneto-optische golfgeleiders. In Hoofdstuk 3 worden de theoretische aspecten van de verschillende bouwstenen van deze component bestudeerd. Aandacht zal hierbij besteed worden aan de verschillende karakterizatiemethodes om de performantie van deze bouwstenen juist in te schatten. Dit hoofdstuk zal afsluiten met een gefundeerd optimalizatie-ontwerp van de lagenstructuur voor deze component. De laatste twee hoofdstukken zullen dan focussen op de experimentele bevindingen van de subcomponenten (Hoofdstuk 4) en de karakterisatie van het niet-reciproke gedrag van de eerste prototypes van de ferromagnetische-metaal-gecontacteerde halfgeleiderversterker/isolator (Hoofdstuk 5). Een aantal conclusies en enige suggesties voor verder werk sluiten deze scriptie af (Hoofdstuk 6).

1.6 Publicaties

De resultaten behaald in dit doctoraatsonderzoek zijn gepubliceerd in verscheidene peer-reviewed internationale tijdschriften:

- M. Vanwolleghem, W. Van Parys, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "Experimental demonstration of nonreciprocal amplified spontaneous emission in a CoFe clad semiconductor optical amplifier for use as an integrated optical isolator." *Appl. Phys. Lett.,* vol. 85(18), pp. 2980–2982, 2004.
- K. Postava, M. Vanwolleghem, D. Van Thourhout, R. Baets, Š. Višňovský, P. Beauvillain, and J. Pištora, "Modeling of a novel InP-based monolithically integrated magneto-optical waveguide isolator." *J. Opt. Soc. Am. B.,* vol. 22(1), pp. 261–273, 2005.
- M. Vanwolleghem, P. Gogol, B. Bartenlian, P. Beauvillain, J. Harmle, L. Lagae, J. Pistora, K. Postava, S. Visnovsky, and R. Wirix-Speetjens, "Magneto-optical parameters of $Co_{90}Fe_{10}$ and $Co_{50}Fe_{50}$ ferromagnetic thin films for 1.3μm integrated isolator." submitted to *J. Appl. Phys.,* (April 2005).

- F. Lelarge, B. Dagens, C. Cuisin, O. Le Gouezigou, G. Patriarche, W. Van Parys, M. Vanwolleghem, R. Baets, and J.-L. Gentner, "GSMBE growth of GaInAsP/InP 1.3μm-TM-lasers for monolithic integration with optical waveguide isolator." *J. Cryst. Growth,* vol. 278(1–4), pp. 709–713, 2005.

- W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "Development of a ferromagnetic transition metal Ohmic p-type contact for a novel integrated optical isolator conept." submitted to *Electron. Lett.,* (April 2005).

Verschillende resultaten werden ook gepresenteerd op internationale en nationale conferenties en workshops:

- M. Vanwolleghem, W.F.H. Van Parys, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens, and J. De Boeck, "Experimental verification of a novel integrated isolator concept." in the Proceedings of *29th European Conference on Optical Communications (ECOC),* post-deadline papers 6, p. 78–79, Rimini, Italy, September 2003.

- M. Vanwolleghem, W.F.H. Van Parys, D. Van Thourhout, R. Baets, O. Gauthier-Lafaye, F. Lelarge, B. Thedrez, R. Wirix-Speetjens, and J. De Boeck, "First Experimental Demonstration of a monolithically integrated InP-based waveguide isolator." in the Proceedings of the *Optical Fibre Communication (OFC '04),* p. TuE6, Los Angeles, USA, February 2004.

- W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, J. Decobert, B. Dagens, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "Demonstration of 81dB/cm isolation on an InP-based optical waveguide isolator." oral presentation at the *12th European Conference on Integrated Optics (ECIO),* Grenoble, France, April 2005.

- W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, J. Decobert, B. Dagens B. Thedrez, R. Wirix-Speetjens and L. Lagae "InP-Based Monolithically Integrated Optical Waveguide Isolator with 32 dB/cm Isolation", in the Proceedings of the *17th Annual Meeting of the Laser and Electro-Optics Society (LEOS),* paper TuY5, pp. 386–387, Puerto Rico, November 2004.

- M. Vanwolleghem, M. Leys, J. Das, L. Lagae, J. De Boeck, P. Van Daele, R. Baets, "Ferromagnetic-metal-based InGaAs(P)/InP optical waveguide isolator : steps towards experimental validation." in the Proceedings of the *14th Annual Meeting of the Laser and Electro-Optics Society (LEOS),* pp. 294–295, San Diego, USA, November 2001.

- M. Vanwolleghem, W.F.H. Van Parys, S. Verstuyft, R. Wirix-Speetjens, L. Lagae, J. De Boeck, and R. Baets, "Ferromagnetic-metal-based InGaAs(P)/InP optical waveguide isolator: electrical and magneto-optical characterisation." in the Proceedings of the *2002 Annual Symposium of the IEEE/LEOS Benelux Chapter,* pp. 282–285, Amsterdam, The Netherlands, 2002.

- W.F.H. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens, and L. Lagae, "Experimental Verification of a Novel Integrated Optical Isolator Concept." in the Proceedings of *Bianisotropics 2004 – 10th Conference on Complex Media and Metamaterials*, pp. 224–227, Ghent, Belgium, September 2004.

- F. Lelarge, B. Dagens, C. Cuisin, O. Le Gouezigou, G. Patriarche, W. Van Parys, M. Vanwolleghem, R. Baets, and J.-L. Gentner, "GSMBE growth of GaInAsP/InP 1.3μm-TM-lasers for monolithic integration with optical waveguide isolator." in the Proceedings of the *13th International Conference on Molecular Beam Epitaxy(IC-MBE),* p. ThA2.2, Edinburgh, Scotland, August 2004.

- M. Vanwolleghem, "A novel monolithically integrated InP-based optical isolator - from theory to practice: major achievements of the ISOLASER project." *IST-OPTIMIST international workshop,* Athens, Greece, June 2004.

- M. Vanwolleghem, "A novel integrated isolator concept." *COST 267 Workshop,* Ghent, Belgium, May 2001.

- M. Vanwolleghem, and R. Baets, "A novel type of integrable non-reciprocal photonic waveguides devices." *1ste doctoraatssymposium Faculteit Toegepaste Wetenschappen,* Ghent, Belgium, December 2000 *(in Dutch).*

Het eerste experimentele succes werd ook opgenomen in de nieuwsbrief van het IMEC onderzoeksinstituut:

- W. Van Parys, and M. Vanwolleghem, "Experimental demonstration of a novel type of optical isolator." *IMEC Newsletter* vol. 39, p. 7, July 2004.

Tenslotte mondde de begeleiding van een studentenproject tijdens de loop van dit onderzoek uit in een internationale publicatie:

- P. Bienstman, E. Six, M. Roelens, M. Vanwolleghem, and R. Baets, "Calculation of bending losses in dielectric waveguides using eigenmode expansion and perfectly matched layers." *IEEE Photon. Technol. Lett.,* vol. 14(2), pp. 164–166, 2002.

2 Modellering van niet-reciproke magneto-optische golfgeleiders

2.1 Reciprociteit

Centraal in de modellering van niet-reciproke golfgeleiders staat het concept van reciprociteit. Vaak wordt dit begrip verward met bidirectionaliteit.

Om het belangrijke verschil te onderlijnen wordt in sectie 2.1.1 het Lorentzreciprociteitstheorema opnieuw mathematisch afgeleid. Vgl. (2.5) en (2.6) geven er de bekende differentiaal- en integraalvorm van, geldend voor velden en bronnen in een medium 'a' en velden en bronnen in het medium 'b', zijnde het toegevoegde (*adjoint*) medium van 'a'. Veel interessanter dan de mathematische afleiding is de fysische interpetatie van het reciprociteitstheorema. Dit is geschetst in Fig. 2.1. Het hypothetisch experiment van Rayleigh-Carlson drukt uit dat tengevolge van het Lorentzreciprociteitstheorema in een reciprook elektromagnetisch probleem (zoals een golfgeleider maar evengoed een verstrooier) *reactie* (i.e. de gegenereerde elektromagnetische velden) ongevoelig is aan een omwisseling van meetpunt en bronpunt. In een niet-reciprook elektromagnetisch probleem geldt hetzelfde maar dan mutueel met het toegevoegde (*adjoint*) probleem. Lorentzreciprociteit zegt dus niets over enig modaal gedrag. Het onderstelt zelfs niet de aanwezigheid van een golfgeleider in het elektromagnetisch probleem. Bidirectionaliteit daarentegen is een eigenschap van het modale spectrum van een golfgeleider. Een golfgeleider is bidirectioneel als voor elke mode met een bepaalde (complexe) propagatieconstante er een modale oplossing bestaat met exact de tegengestelde propagatieconstante (dus propagerend in de tegengestelde richting). Als die mode bovendien een profiel heeft dat exact het spiegelbeeld is van de originele mode dan wordt de golfgeleider bovendien spiegelend genoemd.

Er kan bewezen worden dat een reciproke golfgeleider steeds bidirectioneel is, maar het omgekeerde geldt niet. Algemener geldt dat een bianisotrope golfgeleider (die dus niet-reciprook kan zijn) steeds mutueel bidirectioneel is met zijn toegevoegde. In sectie 2.1.2 wordt verder nog uiteengezet onder welke voorwaarden algemeen bianisotrope golfgeleiders spiegelend worden met zichzelf of met hun toegevoegde. Deze voorwaarden zijn interessant toepasbaar bij het doorrekenen van dergelijke golfgeleiders, aangezien een correcte normalizatie van het modale spectrum van een algemene niet-reciproke bianisotrope golfgeleider de kennis van het spectrum van de toegevoegde golfgeleider vereist. Meer details over allerhande symmetrieën in modale spectra van bianisotrope golfgeleiders vindt de lezer in [12].

2.2 Modale oplossingen van niet-reciproke golfgeleiders

Vanaf hier wordt de behandeling beperkt tot niet-reciproke golfgeleiders waarbij de magnetische permeabiliteit die van het vacuum is, en de dyades $\underline{\xi}$ en $\underline{\zeta}$ nul zijn. De niet-reciprociteit zit dus volledig vervat in $\underline{\epsilon}$. Een grondig overzicht van de modelleringsmethodes voor deze belangrijke klasse golfgeleiders wordt hier gegeven.

2.2.1 Greense functies

Greense functietechnieken zijn zeer populair in elektromagnetische verstrooiingsproblemen. Een Greense functie is een drie-dimensionele dyade (dus 9 componenten!) die de respons van het elektromagnetisch probleem beschrijft bij een Dirac bronterm (zie Vgl. (2.24)). Men kan de berekening van de Greense dyade enigszins verlichten door een scalaire Greense functie in te voeren (zie Vgl. (2.26) en (2.28)). Het Greense functie-concept is een zeer interessante mathematische constructie. Er kan rigoreus bewezen worden dat de berekening van de polen en de vertakkingslijnen van de spectrale Greense functie het volledige modale spectrum van een golfgeleider oplevert [12], waarbij de polen corresponderen met de geleide modi en de vertakkingslijnen met het continuum van stralingsmodi. Bovendien leggen de vertakkingslijnen (wiens vorm en plaats uniek bepaald wordt door de stralingsvoorwaarde) op een heel duidelijke manier het verschil vast tussen zogenaamde onechte "leaky" modi en "echte" geleide modi van de golfgeleider. De toegevoegde waarde van de scalaire spectrale Greense functie van een golfgeleider is dus heel hoog. Echter is het berekenen ervan voor golfgeleiders die complexer zijn dan simpele isotrope golfgeleiders, mathematisch heel moeilijk, voornamelijk omdat het niet altijd duidelijk is welke randvoorwaarden er moeten opgelegd worden aan de verschillende grensvlakken. Meer over Greense functietechnieken vindt de lezer in de engelstalige tekst en uiteraard in [12]. Bij planaire (i.e. 1-dimensionele) niet-reciproke golfgeleiders opgebouwd uit stuksgewijze homogene lagen is het toch mogelijk het modale eigenwaardeprobeem (beschreven door het stelsel sterk gekoppelde partiële differentiaalvergelijkingen van (2.20) met $\mathbf{j}(\mathbf{r}) = 0$) rechtreeks in de velden op te lossen zonder een toevlucht te moeten nemen tot het berekenen van polen en vertakkingslijnen van Greense dyades. De techniek die dit toelaat is een veralgemening van het bekende Transfer Matrix Formalisme.

2.2.2 Transfer Matrix Formalisme (TMF) voor niet-reciproke magneto-optische golfgeleiders

Inleiding Het centrale idee van het TMF is geschetst in Fig. 2.2. Het principe bestaat erin in een zogenaamde transfermatrix op te stellen die beschrijft hoe de lokale voorwaarts en achterwaarts propagerende vlakke golven[2] (i.e. de eigenmodi in de homogene lagen) zich "voortplanten" doorheen de lagenstruktuur. Voorwaarts en achterwaarts slaat in deze context op de propagatierichting loodrecht t.o.v. de grensvlakken tussen de verschillende lagen en heeft niets te maken met de propagatierichting in de golfgeleider (die uiteraard parallel is aan de grensvlakken). Door geschikte randvoorwaarden op te leggen aan de hypothetische wanden waarin de golfgeleider opgesloten zit (of op oneindig bij een open golfgeleiderprobleem), kan uiteindelijk een resonantievoorwaarde afgeleid worden die niets anders is dan de dispersievergelijking van de golfgeleider. Dit principe is genoegzaam gekend in isotrope golfgeleiders, en de afleiding van de transfermatrix is in dit geval redelijk eenvoudig. Dit is voornamelijk het geval omdat de eigenmodi in elke homo-

[2] met een vaste propagatiefactor $\mathrm{e}^{-j\beta z}$ in de voortplantingsrichting van de golfgeleider. De propagatieconstante β in deze factor is uiteraard de nog onbekende eigenwaarde van de gezochte golfgeleidermodi.

gene laag de bekende TE en TM polarizaties zijn en de randvoorwaarden aan de grensvlakken[3] deze polarizaties ongekoppeld laten aan deze grensvlakken tussen de verschillende lagen. Als gevolg hiervan kan een afzonderlijke dispersierelatie voor TE en TM modi afgeleid worden (met behulp van ongekoppelde 2-bij-2 matrices). In een algemene niet-reciproke golfgeleider is dit niet meer het geval, en zijn de voorwaarts en achterwaarts propagerende lokale eigenmodi elliptisch gepolariseerd en koppelen die onderling aan elk grensvlak.

Uitbreiding TMF In sectie 2.2.2.2 wordt de mathematische rigoureuze behandeling uiteengezet die leidt tot de uitbreiding van het TMF voor niet-reciproke slabgolfgeleiders. De lezer wordt verwezen naar de engelstalige tekst voor de details. We zullen hier volstaan met het aanstippen van enkele belangrijke conclusies van deze uitbreiding.

In elke homogene laag zijn er vier elliptisch gepolariseerde vlakke golf eigenpolarizaties (2 voorwaarts en 2 achterwaarts), die ongekoppeld propageren doorheen de homogene laag. Deze eigenpolarizaties worden gevonden als de eigenvectoren van de zogenaamde systeemmatrix van de anisotrope laag (zie Vgl. (2.36) en (2.40)). De amplitudes van deze 4 eigenpolarizaties vormen lokaal in elke laag een viervector. De polarizatie van elk van deze 4 eigenoplossingen wordt beschreven door een Cartesische viervector in de tangentiële veldcomponenten (zie Vgl. (2.38)). Er zijn dus twee coördinatensystemen: enerzijds het Cartesisch systeem van de tangentiële veldcomponenten en anderzijds het lokaal geldende coördinaatsysteem van de amplitudes van de eigenpolarizaties, ook wel het "wavefield" systeem genoemd. De transformatie tussen beide systemen wordt beschreven door de matrix opgebouwd uit de Cartesische eigenvectoren van de systeemmatix van elke laag (zie opnieuw Vgl. (2.38)). Het is duidelijk dat in elke laag het "wavefield" systeem het bevoorrechte systeem is voor de propagatie, terwijl het Cartesische systeem het bevoorrechte systeem is voor het opleggen van de randvoorwaarden aan de grensvlakken tussen de media. Door dus telkens gepast van het ene systeem op het andere over te gaan, kan sequentieel de volledige transfermatrix van de anisotrope multilagenstack opgebouwd worden (zie Vgl. (2.46)).

Dispersievergelijking en modeprofielen Eens de uitgebreide transfermatrix opgesteld, is het een kleine stap om de uiteindelijke dispersievergelijking te bekomen. Voor de afleiding hiervan dient een onderscheid gemaakt te worden tussen open en gesloten golfgeleiders. Gesloten golfgeleiders zijn ingesloten tussen twee perfect geleidende wanden (magnetisch of elektrisch). De dispersievergelijking wordt in dat geval dus gevonden door in de transfermatrixrelatie, die de tangentiële veldcomponenten aan beide uiteinden van de stack met elkaar verbindt, de elektrische resp. magnetische velden gelijk aan nul te stellen, afhankelijk van de aard van de wand (perfect elektrisch geleidend of perfect magnetisch geleidend). De overige veldcomponenten zullen dan enkel verschillend van nul zijn als een determinant van een 2-bij-2 matrix nul wordt.
Bij een open golfgeleider, gaat men analoog te werk, maar dan wel in het "wave-

[3] namelijk de continuïteit van de tangentiële veldcomponenten

field" systeem. Nu stelt men die "wavefield" componenten gelijk aan nul die corresponderen met de onfysische bijdragen "vanuit oneindig". Dit betekent dat er in het superstraat geen "wavefield" component mag aanwezig zijn die vanuit $+\infty$ afkomstig is, en in het substraat geen component die vanuit $-\infty$ afkomstig is. Opnieuw leidt dit tot een determinantvergelijking van een 2-bij-2 submatrix van de transfermatrix.

Het berekenen van de modale profielen gebeurt vrij eenvoudig eens de transfermatrix gekend is. En dat is nu het geval, aangezien de net opgeloste dispersievergelijking toelaat de transfermatrix numeriek door te rekenen. Het enige dat hiervoor vereist is, is een startwaarde voor de velden op een willekeurig punt in de multilagenstack. Een dergelijke startwaarde kan gevonden worden door een gekende randvoorwaarde toe te passen in het beschouwde punt. Deze startwaarde kan dan "gepropageerd" worden doorheen de multilagenstack met behulp van de transfermatrix. In sectie 2.2.2.4 wordt in meer detail uiteengezet hoe een startwaarde afgeleid wordt uit de dispersievergelijking, en hoe deze procedure numeriek stabiel gemaakt kan worden. Tenslotte wordt er ook dieper ingegaan op het normalizatieprobleem van modale profielen in niet-reciproke anisotrope golfgeleiders.

Voorbeeld – een Faraday rotator Deze uitbreiding van het TMF voor niet-reciproke anisotrope slabgolfgeleiders is geïmplementeerd in een Mathematica omgeving. Als testcase wordt in sectie 2.2.2.5 een eenvoudige golfgeleider Faraday rotator doorgerekend.

2.2.3 Storingsrekeningformalisme

Schets van de theoretische achtergrond Vaak zijn de termen in de diëlektrische tensor die de niet-reciprociteit veroorzaken, voldoende klein zodat hun effect perturbatief kan gemodelleerd worden. In sectie 2.2.3 wordt een van de populairste perturbatietechnieken uiteengezet. Ze is een uitbreiding van de techniek van Vassalo [13].

De basisgedachte van deze zogenaamde modale expansiemethode bestaat erin de niet-reciproke storingen in de diëlektrische tensor te behandelen als een "storingsstroombron" voor de originele ongestoorde golfgeleider. Deze storingsstroombron is een functie van de gezochte velden (zie Vgl. (2.63)). Door gebruik te maken van de compleetheid van het modale spectrum van de ongestoorde golfgeleider, kunnen de onbekende velden (en dus ook de storingsbronterm) ontwikkeld worden in de ongestoorde eigenmodi. Toepassing van het Lorentz theorema, dat een verband geeft tussen enerzijds de ongestoorde eigenmodi (bronloos ondersteund in de ongestoorde golfgeleider) en de gezochte velden veroorzaakt door de storingsbron, op deze expansie levert uiteindelijk een oneindige set gekoppelde eerste orde differentiaalvergelijkingen in de expansiecoëfficiënten, de zogenaamde evolutievergelijking (zie Vgl. (2.83) en (2.84)). Deze set differentiaalvergelijkingen is niets anders dan een herformulering van het elektromagnetisch probleem, maar ditmaal in termen van scalaire expansiecoëfficiënten en niet in termen van twee vectoriële velden. Als de storing longitudinaal homogeen is, dan zal diagonalizatie van deze vergelijking uiteindelijk de eigenwaarden en eigenmodi leveren van

de “echte” golfgeleider. In praktijk is dit quasi onmogelijk, omdat dit een diagonalizatie vereist van een oneindig dimensionele matrix. Het is op dit punt dat perturbatietechnieken aangewend worden. We verwijzen de lezer voor alle details hieromtrent naar de engelstalige tekst (meer bepaald naar sectie 2.2.3.4). Ook voor meer details aangaande de afleiding van de evolutievergelijking verwijzen we de lezer naar deze en voorafgaande secties. We volstaan hier enkel met op te merken dat er eenvoudig bewezen wordt dat de diagonaalelementen van de koppelmatrix de eerste orde perturbatieve correcties op de eigenwaarden van de ongestoorde modi voorstellen, terwijl de niet-diagonaalelementen duiden op een perturbatieve koppeling tussen de ongestoorde modi. Dit is een alom bekend resultaat uit storingsrekening.

Klassificatie magneto-optische golfgeleidereffecten Wanneer de evolutievergelijking expliciet uitgeschreven wordt voor de bijzondere klasse van niet-reciproke golfgeleiders bestudeerd in dit werk, namelijk magneto-optische golfgeleiders, dan kan een elegante klassificatie gemaakt worden van de verschillende niet-reciproke magneto-optische golfgeleiderfenomenen als functie van de precieze richting van de magnetizatie in de golfgeleider en van de symmetrie van het magnetizatieprofiel. Een dergelijk klassificatie wordt mogelijk gemaakt door in detail te bestuderen welke elementen verschillend van nul worden in de koppelmatrix van de evolutievergelijking. Van nul verschillende diagonaalelementen duiden op een niet-reciproke magneto-optische correctie op de propagatieconstanten van de ongestoorde modi (m.a.w. zogenaamde niet-reciproke fazeshifts), en van nul verschillende niet-diagonaalelementen duiden op niet-reciproke magneto-optische mode-en/of polarizatieconversie van de ongestoorde modi. Tabellen 2.3 en 2.4 geven een helder overzicht van deze verschillende magneto-optische golfgeleidereffecten, en tabelleren de verschillende formules voor eerste orde perturbatieve modellering ervan. Deze tabel toont duidelijk hoe bijvoorbeeld ook voor een TE mode een niet-reciproke fazeshift kan opgewekt worden, namelijk door een anti-symmetrisch magnetizatieprofiel in polaire configuratie aan te wenden. In sectie 2.2.3.5 wordt een iets diepgaandere discussie gegeven over hoe deze formules effectief leiden tot niet-reciprook gedrag.

2.3 Equatoriale magneto-optische golfgeleiders

2.3.1 Inleiding

De rest van dit hoofstuk spendeert extra aandacht aan de bijzondere klasse magneto-optische golfgeleiders bestudeerd in dit werk, namelijk equatoriaal symmetrisch gemagnetiseerde golfgeleiders. In de sectie over perturbatietheorie is er aangetoond dat in een dergelijke configuratie het modale spectrum van een slabgolfgeleider perfect gescheiden blijft in TE en TM modes, en dat enkel de TM modes een niet-reciproke correctie op hun propagatieconstante ondergaan. Dit kan ook rigoreus bewezen worden vertrekkende vanuit de Hemholtzvergelijking (zie Vgl.(2.106) en (2.111)). Deze vergelijkingen tonen aan dat er geen magneto-optische termen zijn in de TE vergelijking en dat de TM vergelijking een MO term heeft die oneven is

in de propagatieconstante!
Wanneer een realistische 2D golfgeleider beschouwd wordt zoals bijvoorbeeld een ribgolfgeleider, dan is het geweten uit isotrope golfgeleidertheorie dat het ongestoorde spectrum in essentie geen zuiver TE of TM karakter meer heeft, maar dat de modi eerder hybride gepolariseerd zijn. Het is wel zo dat vaak een semivectoriële benadering kan doorgevoerd worden met als ansatz dat deze hybride modi zich weer opsplitsen in een set met een dominant TE karakter en een set met een dominant TM karakter. Tabel 2.3 toont echter aan dat bij een symmetrische equatoriale MO storing er mogelijk een koppeling optreedt tussen de verschillende semivectorieel gepolariseerde modi (derde rij laatste kolom in deze tabel). M.a.w. het is mogelijk dat door deze koppeling de semivectoriële benadering van de ongestoorde isotrope golfgeleider niet meer geldt in de magneto-optische golfgeleider. In sectie 2.3.1 wordt er echter rigoureus bewezen dat dit niet het geval is. M.a.w. als de ongestoorde ribgolfgeleider voldoet aan de semi-vectoriële benadering blijft dit gelden voor de MO golfgeleider: het modale spectrum blijft zijn semivectorieel TE en TM karakter behouden én enkel de TM modi ondergaan een niet-reciproke fazeshift.

2.3.2 Modellering van symmetrische equatoriale magneto-optische golfgeleiders

Storingsrekening De storingsformule voor de niet-reciproke fazeshift van semivectoriële TM modi in equatoriaal gemagnetizeerde MO golfgeleiders (derde rij, tweede kolom in Tabel 2.4) wordt in meer detail bestudeerd. Meer bepaald wordt er onderzocht hoe de verschillende complexe delen van de magneto-optische constante het reëel respectievelijk het imaginair deel van de niet-reciproke fazeshift beïnvloeden. Aangezien in het algemeen zowel de magneto-optische constante g als de ongestoorde elektrische velden complex kunnen zijn, is dit niet zo eenvoudig vast te stellen. De eenvoudigste situatie doet zich voor als de golfgeleider verliesloos is. In dat geval zijn de niet-diagonale elementen van de diëlektrische tensor zuiver imaginair (wegens de hermiticiteit van de diëlektrische tensor) en is de magneto-optische constante dus zuiver reëel. Bovendien zijn de transversale en de longitudinale veldcomponenten in een verliesloze golfgeleider perfect in kwadratuur. Met andere woorden het integrandum van de storingsformule (2.117) is zuiver imaginair, wat in combinatie met de imaginaire voorfactor, een zuiver reële fazeshift oplevert. Uiteraard is deze mathematische argumentatie een redelijke omslachtige bevestiging van wat op puur fysische gronden duidelijk was: een verliesloze golfgeleider onderworpen aan een verliesloze magneto-optische storing zal uiteraard niet leiden tot een absorptie-effect.
Iets ingewikkelder wordt het wanneer de golfgeleider verlieshebbend is, of door een verlieshebbend magneto-optisch effect of doordat de ongestoorde golfgeleider verlieshebbend is, of een combinatie van beide. Er zijn dus verschillende scenario's denkbaar, maar in essentie komt het er steeds op neer dat het reëel deel van het integrandum (dat verantwoordelijk is voor de absorptieshift) gedomineerd wordt door het imaginair deel van g, terwijl het imaginair deel van het integrandum (dat verantwoordelijk is voor de niet-reciproke fazeshift) gedomineerd wordt door

het reëel deel van g. Nog ingewikkelder wordt het als men tracht het teken van het niet-reciprook effect (absorptieshift of fazeshift) te voorspellen in functie van de tekens van de complexe delen van g en van het veldproduct $E_x^{(0),TM} E_z^{(0),TM}$ van de ongestoorde mode. Dit is voornamelijk het gevolg van het feit dat, zoals uitgelegd in Hoofdstuk 1, alle vier de tekencombinaties voor de complexe delen van g fysisch zinvol zijn. Een bescheiden poging tot klassificatie wordt gegeven in Tabel 2.6. Deze tabel moet echter met de nodige voorzichtigheid gebruikt worden. De conclusies gelden enkel als het reëel deel van het veldproduct verwaarloosbaar is t.o.v. het imaginaire deel.

TMF formalisme Zoals hierboven uiteengezet, zal in een equatoriale MO golfgeleider het karakteristieke TE en TM karakter van het golfgeleiderspectrum behouden blijven. Als gevolg hiervan splitst het veralgemeende 4-bij-4 TMF formalisme van sectie 2.2.2.2 opnieuw op in twee sub-TMF's, die elk op zich afgeleid kunnen worden zoals bij een isotrope golfgeleider. De essentie is dat er helemaal geen behoefte meer is aan het berekenen van de lokale eigenpolarizaties. Het is op voorhand geweten dat die TE en TM zullen zijn. Als gevolg hiervan is er geen behoefte meer om te werken met twee verschillende coördinaatsystemen zoals in sectie 2.2.2.2, aangezien de randvoorwaarden aan de grensvlakken impliciet inbegrepen zijn in de scalaire Fresnel reflectie- en transmissiecoëfficiënten geldend aan deze grensvlakken. Merk op dat in het veralgemeende TMF in feite de randvoorwaarden ook via reflectie- en transmissiecoëfficiënten opgelegd worden, maar dat die in dit geval zeer complexe 2-bij-2 matrices worden tengevolge van de eigenpolarizatiemixing.
De basisopdracht is dus het afleiden van de veralgemeende Fresnel coëfficiënten aan het grensvlak tussen (in het algemeen) twee equatoriaal gemagnetiseerde MO media. Er is op gewezen in Hoofdstuk 1 dat voor deze bijzondere MO configuratie de TE Fresnel coëfficiënten ongewijzigd blijven, terwijl de TM coëfficiënten een niet-reciproke MO correctie krijgen. De afleiding van deze niet-reciproke TM Fresnel coëfficiënten is een typische magneto-optica tekstboek oefening. Ze wordt in detail uiteengezet in sectie 2.3.2.2. Merken we tot slot nog op dat de propagatie van de lokale voorwaartse en achterwaartse vlakke golven in elke laag van de equatoriale MO slabgolfgeleider, voor TE golven identiek is aan de isotrope propagatie. Voor TM golven daarentegen komt er een tweede orde MO correctie wegens het Voigt effect (zie sectie 1.2.2.2). Dit alles onderlijnt nogmaals dat de niet-reciprociteit in equatoriale MO slabgolfgeleiders volledig op het conto te schrijven is van de grensvlakken! En dat er voor TE modi hoegenaamd geen effect is. Hiermee zijn alle elementen verzameld om de totale TM transfermatrix te berekenen (zie Vgl. (2.135) en (2.136)).

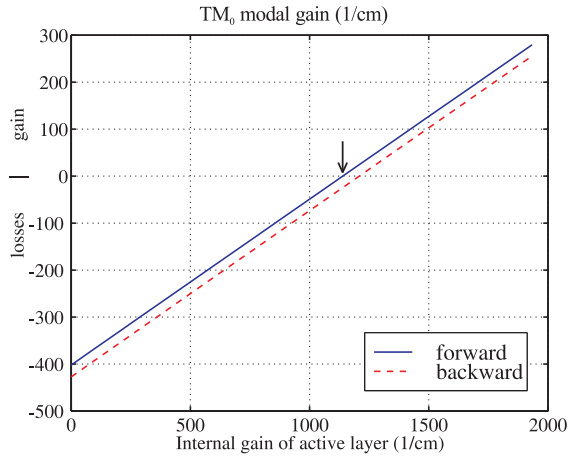
Het is nu een kleine stap om tot de uiteindelijke dispersievergelijking te komen. Afhankelijk van het open of gesloten karakter van de golfgeleider wordt deze gegeven door hetzij een bepaald matrixelement van de 2-bij-2 transfermatrix, hetzij door een lineaire combinatie van alle vier de matrixelementen – namelijk een lineaire combinatie die op een geschikte manier de muurrandvoorwaarden oplegt. Dit is een heel summiere samenvatting van een redelijke subtiele oefening. In werkelijkheid moet er heel voorzichtg rekening gehouden worden met het optreden

van vertakkingslijnen en polen in de dispersievergelijking om op de juiste manier de eigenlijke golfgeleidermodi te vinden. Er kan rigoureus bewezen worden dat in feite de complexe dispersievergelijking in het complexe β-vlak aanleiding geeft tot een viervoudige meerwaardigheid. Bij het zoeken naar nulpunten met behulp van tracking algoritmes of contourintegratiemethodes moet er dus omzichtig te werk gegaan worden, zoniet riskeert men onfysische nulpunten van de dispersievergelijking te detecteren. Dit alles vereist het gebruik van technieken uit de complexe analyse zoals Riemann oppervlakken en ontvouwing door middel van conforme "mapping". De studie hiervan is in detail beschreven in de engelstalige tekst. Eens de correcte nulpunten (of met andere woorden de propagatieconstanten) geïdentificeerd zijn, kan de transfermatrix numeriek uitgerekend worden (uiteraard op de juiste manier rekening houdend met de meerwaardigheid van de complexe functies), en kan op dezelfde manier als in het algemene geval van sectie 2.2.2.2 het modale profiel bekomen worden. Opnieuw moet hier de nodige aandacht geschonken worden aan numerieke instabiliteiten bij het "transfereren" van een startveld doorheen de stack.

Dit gespecialiseerde geval van het TMF werd net als de veralgemeende versie geïmplementeerd in een Mathematica omgeving.

2.3.3 Benchmark

Als afsluiting van dit uitgebreide theoretisch hoofdstuk wordt in sectie 2.3.3 een vergelijkende studie gemaakt van beide hier ontwikkelde modelleringstechnieken, specifiek toegepast op de in dit werk bestudeerde niet-reciproke TM complexe fazeshifter. Met andere woorden, er wordt bestudeerd in hoeverre de perturbatieve aanpak voldoende nauwkeurig is. Er is immers geen eenduidige vuistregel die stelt tot op welk niveau een storing van de isotrope diëlektrische constante als "voldoende klein" beschouwd kan worden om perturbatieve modellering te rechtvaardigen. Dit is niet enkel vanuit theoretisch oogpunt interessant, maar ook om puur praktische reden. Het is immers duidelijk dat het perturbatieve algoritme veel efficiënter is dan het rigoureuze TMF (zowel qua berekeningstijd als qua geheugengebruik). Het storingsalgoritme vereist enkel de berekening van een isotrope golfgeleider gecombineerd met een eenvoudige overlapintegraal, terwijl het rigoureuze TMF de oplossing vereist van een complexe niet-reciproke dispersierelatie in het volledige β-vlak (gecombineerd met de nodige conforme transformatie om onfysische gebieden af te zonderen), en dat laatste moet bij elke variatie van om het even welke golfgeleiderparameter. Het belangrijkste voordeel van het rigoureuze schema is dat het eenvoudig toelaat modale profielen uit te rekenen. Er wordt in de engelstalige tekst uitgelegd waarom de storingstechniek dit niet zo eenvoudig toelaat. De gemodelleerde equatoriale MO slabgolfgeleider is een hypothetische vereenvoudigde versie van de SOA/isolator. De details zijn geschetst in Fig. 2.17. Er wordt een bulk actieve laag verondersteld i.p.v. een MQW laag omringd door geleidingslagen, en er is geen gedopeerde halfgeleidercontactlaag opgenomen onder het ferromagnetische metaal. Dit eenvoudige model volstaat voor deze vergelijkende studie, en voor een eerste numerieke studie van het gedrag

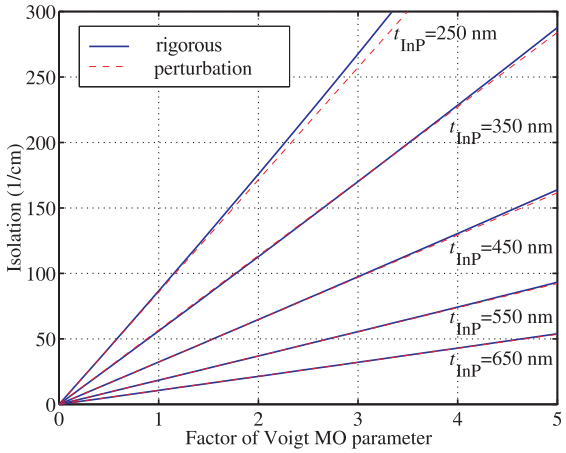


Figuur 2: Rigoureus berekende TM modale winst als functie van de interne materiaalwinst bij een dikte van 500nm voor de InP bufferlaag. Bij een interne winst van ongeveer 1100/cm is de component transparant in voorwaartse richting, terwijl in de achterwaartse richting er een verlies is van ongeveer 25/cm.

van deze component. De winst in de actieve laag wordt in rekening gebracht door een positief imaginair deel van de brekingsindex, en wordt dus in het storingsformalisme mee opgenomen in de ongeperturbeerde isotrope golfgeleider!

Fig. 2 toont hoe, bij een vaste waarde voor de bufferlaag tussen de actieve laag en het ferromagnetisch contact, de modale winst of het verlies van de TM grondmode verschilt in voorwaartse en achterwaartse richting en dat in functie van de winst in de actieve laag. Deze curve is op een rigoureuze wijze berekend. Het valt op dat de niet-reciprociteit quasi-constant is voor een breed bereik van winstniveau's. Deze curve bevestigt dat de eenvoudige visie op de werking van de component geschetst in sectie 1.4.1 niet helemaal onrealistisch is. Het transversaal MO Kerr effect introduceert een bepaalde niet-reciproke absorptieshift en de winst in de actieve laag transfereert deze shift naar een regime van dat van een niet-reciproke versterker ($g_{\mathrm{InGaAsP}} > 1100/cm$), een isolator ($g_{\mathrm{InGaAsP}} \approx 1100/cm$), of een niet-reciproke absorber ($g_{\mathrm{InGaAsP}} < 1100/cm$). Het belangrijkste voordeel van deze vaststelling is dat storingsrekening, als die geldig is, een heel efficiënte manier wordt om de isolator te modelleren. Immers, het volstaat het transparantiepunt te zoeken in de isotrope golfgeleider en dan 1 enkele maal de storingsintegraal uit te rekenen. Het TMF daarentegen moet voor elke nieuwe waarde van de interne materiaalwinst een niet-reciproke dispersierelatie herberekenen. Merk ook op dat de winstniveaus in principe niet haalbaar zijn voor bulk actieve lagen. Zoals gezegd, dienen de simulaties van deze sectie voornamelijk als een test van de ontwikkelde tools en niet als designsimulaties.

Om na te gaan of storingsrekening geldig is wordt de isolatie berekend voor verschillende waarden van de dikte van de bufferlaag en voor verschillende hypo-



Figuur 3: Vergelijking van het storingsalgoritme en TMF. Isolatie wordt geplot als functie van de kunstmatige versterkingsfactor f van de MO Voigt parameter ($Q' = fQ_{\mathrm{CoFe}}$) voor verschillende waarden van de InP buffer.

thetische sterkten van de MO parameter (bekomen door de grootte van de bekende constante met een factor te vermenigvuldigen). Dit is geschetst in Fig. 3. Het valt op dat enkel bij heel dunne bufferlagen, m.a.w. als de geleiding van de TM mode in de actieve laag zwaar beïnvloed wordt door de plasmonische geleiding aan het metaal, er een afwijking begint op te treden tussen TMF en perturbatieberekeningen. Maar zelfs dan enkel bij hypothetisch hoge waarden voor de MO parameter. Bovendien wordt er verderop getoond (zie Fig. 2.21) dat die heel dunne bufferlagen slechts theoretisch interssant zijn, maar niet praktisch toepasbaar aangezien ze uiteraard onrealistisch hoge winstniveaus vereisen voor voorwaartse transparantie. Het is dus frappant dat zelfs voor ferromagnetische metalen, die gyrotropieconstanten vertonen die over het algemeen "slechts" een grootte-orde (of minder) kleiner zijn dan de isotrope constanten, de eerste orde perturbatieve modellering voldoende nauwkeurig is.

In deze sectie wordt nog wat dieper ingegaan op het gedrag van de component bij variatie van de dikte van de bufferlaag en de geleidingslaag, en de dikte van de ferromagnetische filmlaag (zie Figuren 2.20, 2.21, en 2.22). We verwijzen de lezer naar de engelstalige tekst hiervoor. Deze simulaties zijn uiteraard interessant, maar de echte designsimulaties zullen aan bod komen in het volgende hoofdstuk. Hier dienen ze voornamelijk als tests voor de ontwikkelde tools. De waargenomen tendenzen zijn immers redelijk logisch. Wel willen we kort wijzen op het bijzondere gedrag bij heel dunne metaallagen. Als gevolg van een plasmonisch interferentie-effect is er een minimale dikte vereist alvorens het niet-reciprook absorptie-effect merkbaar wordt! Het is ook interessant te letten op het feit dat deze component een sterke asymmetrie vertoont qua brekingsidexprofiel, en dat deze asymmetrie zich sterk laat voelen bij afnemende bufferdikte. Hierdoor moet in principe de dikte van de geleidende laag continu aangepast worden om te vermijden dat de

golfgeleider in cut-off gaat.

2.4 Besluit

Dit uitgebreide theoriehoofdstuk heeft de modellering van niet-reciproke golfgeleiders in detail behandeld. Startend bij een correcte definitie van het begrip, via de gedetailleerde afleiding van een rigoureus en een perturbatief modelleringsalgoritme, tot uiteindelijk de specializatie van deze algoritmes naar het geval van magneto-optische golfgeleiders en meer in het bijzonder equatoriale MO golfgeleiders. Alle ontwikkelde tools zijn geïmplementeerd en getest. Een belangrijke conclusie daarbij was dat voor het bijzondere geval van equatoriale MO golfgeiders eerste orde storingsrekening voldoende nauwkeurig is.

3 Design

Dit hoofdstuk zal één voor één het ontwerp behandelen van de subcomponenten van de bestudeerde component, met de focus op de in sectie 1.4.2 aangehaalde uitdagingen voor deze subcomponenten. Aangezien veel van de vereiste eigenschappen een nauwkeurige karakterizatie vereisen, zal er ook de nodige aandacht besteed worden aan de ontwikkeling en de theoretische achtergrond van de karakterisatiemethodes. Tot slot van dit hoofdstuk zal het ontwerp van de specifieke MQW lagenstruktuur voor de SOA/isolator behandeld worden.

3.1 Ferromagnetisch metaal/halfgeleidercontact

In deze uitgebreide sectie wordt er dieper ingegaan op de theoretische aspecten en karakterizatiemethodologie van de drie belangrijke eigenschappen van het ferromagnetische metaal/halfgeleidercontact: de magneto-optische sterkte, de magnetische eigenschappen en de elektrische eigenschappen.

3.1.1 Veralgemeende magneto-optische ellipsometrie

Er is slechts heel weinig geweten over de magneto-optische sterkte van magneto-optische ferromagnetische metalen in het IR. Buiten het karakterisatiewerk van Krinchik in de jaren '60 op de basistransitiemetalen Co, Fe en Ni [14], is er geen enkele bron die de magneto-optische sterkte van deze metalen, hun legeringen en hun samenstellingen met andere elementen gemeten heeft in het spectrale gebied onder 1.5eV. De basistechniek voor karakterisatie van magneto-optische constanten is, net als voor de karakterisatie van de isotrope brekingsindex, ellipsometrie. Sectie 3.1.1 bestudeert in detail de uitbreiding van klassieke isotrope ellipsometrie naar veralgemeende magneto-optische ellipsometrie.
Het basisprincipe van ellipsometrie bestaat erin de optische eigenschappen van een reflectieve stack van materialen te meten door de verandering van de polarizatietoestand van een invallende lichtbundel (met gekende polarizatie) te detecteren. Op die manier wordt de Jonesmatrix van de stack gemeten, en dus, via combinatie met een theoretische reflectietool, de (magneto-)optische indices van de verschillende lagen in de stack (na een niet-lineaire fittingprocedure). De basiselementen nodig voor een algemene ellipsometrische opstelling zijn geschetst in Fig. 3.1. In essentie komt het erop neer dat een combinatie aan optische elementen nodig is die hetzij in het invallend lichtpad, hetzij in het gereflecteerd lichtpad, een willekeurige polarizatietoestand kan transformeren in elke andere willekeurig polarizatie. Er wordt bewezen dat hiervoor twee elementen nodig zijn: een polarisator en een compensator (i.e. een optisch element dat een instelbare vertraging introduceert tussen zijn snelle en zijn trage as). De volgorde speelt geen rol. Door deze combinatie bijvoorbeeld in het reflectieve pad op te nemen, en in het invallende een bekende lineaire polarizatie te creëren (met behulp van een polarizator) kan in principe de detectie steeds geannuleerd worden door de instellingen van de polarizator en de compensator te variëren . Dit is het basisprincipe van "*null ellipsometry*". Uit de kennis van de invallende polarizatie en de *nulling* instellingen van de compensator/analyzer kan de Jones matrix afgeleid worden. Null ellipsometrie is heel

omslachtig en ruisgevoelig. Een tweede klasse ellipsometrische technieken is de zogenaamde fotometrie. Bij fotometrie wordt de informatie gehaald uit variaties van de lichtintensiteit bij variaties van de instellingen van de optische elementen. Dus in plaats van te trachten te intensiteit te annuleren (en zo dus heel precies de polarizatietoestand te meten), haalt men informatie over de Jonesmatrix van de reflectieve stack uit intensiteitsvariaties als functie van bijvoorbeeld de fazevertraging van de compensator, de invalshoek van het licht, de azimuth setting van de analyzer, ... Deze variaties kunnen dynamisch zijn of statisch. In het eerste geval spreekt men van dynamische fotometrie en bekomt men de informatie na eerst een Fourierontbinding van de variërende intensiteit (met behulp van een lock-in amplifier) toe te passen. In het tweede geval spreekt men van statische fotometrie en meet men dus rechtstreeks een of ander "bekend" analytisch verband tussen intensiteit een experimentele setup parameter.

In de secties 3.1.1.1 en 3.1.1.2 worden een dynamische, resp. statische fotometrische techniek specifiek gericht op het karakteriseren van magneto-optische reflectieve stacks uit de doeken gedaan. Voor de details verwijzen we naar de engelstalige tekst. Hier zullen enkel de grote lijnen geschetst worden. In beide methodes wordt het sample onderworpen aan een longitudinaal magneetveld. Volgens de uiteenzetting van Hoofdstuk 1 zal dus het longitudinaal Kerr effect optreden en wordt het MO sample beschreven door een volledig 2-bij-2 Jones matrix. De doelstelling is dus de genormaliseerde elementen van de Jones matrix te meten, r_{pp}/r_{ss}, r_{sp}/r_{ss} en r_{ps}/r_{ss}. Het volstaat de genormaliseerde elementen te meten omdat de globale fazefactor geen extra informatie bevat over de (magneto-)optische indices. Het is welgekend dat deze complexe ratios niets anders zijn dan de isotrope polarizatierotatie en -ellipticiteit, en de magneto-optische polarizatierotatie en ellipticiteit voor $s-$ resp. $p-$polarizatie.

Dynamische magneto-optische fotometrie In de dynamische opstelling wordt de fazevertraging van de compensator harmonisch gevarieerd. Er wordt bewezen door middel van Jones calculus dat, door de analyzer op 45° te plaatsen t.o.v. de compensator (die zelf in s- of p-configuratie gealigneerd is) de complexe delen van de ellipsometrische ratios gevonden worden bij respectievelijk een cosinus en een sinus term van de gemoduleerde fazevertraging. Als de fazevertraging harmonisch gemoduleerd is, dan kunnen deze cosinus en sinus term ontbonden worden via een Fourier-Bessel expansie. De verschillende delen van de complexe polarizatierotatie worden dan gevonden bij respectievelijk de tweede en de eerste harmonische van de intensiteit (zie Vgl. (3.17)). Meer details over deze heel beknopte behandeling van dynamische ellipsometrie worden gegeven in sectie 3.1.1.1. Daar wordt er ook dieper ingegaan op de methodes die ter beschikking staan om de meetmethode te calibreren. Immers, het is niet evident de amplitude van de fazevertragingsmodulatie exact te bepalen. Als gevolg daarvan is de onzekerheid op de Bessselse functies in Vgl. 3.17 niet onbelangrijk. Er bestaan echter twee elegante manieren om een gecalibreerde meting van de polarizatierotatie uit te voeren zonder de Besselse functies te moeten berekenen.

Statische magneto-optische fotometrie De statische methode beschreven in sectie 3.1.1.2 is een recente magneto-optische uitbreiding van de isotrope *rotating analyser methode*. De compensator wordt weggelaten en de gevarieerde experimentele parameter is de azimuth van de analyser (over π). Er kan gemakkelijk ingezien worden dat dit toelaat de polarizatietoestand van het gereflecteerde licht te bepalen (op het teken van de ellipticiteit na). Er kan bewezen worden dat de bekomen periodieke intensiteitsfunctie slechts twee Fouriercomponenten heeft, die rechtsreeks verbonden zijn met de genormaliseerde elementen van de Jones matrix. De methode werkt in principe voor arbitraire magneto-optische multilagenstacks, maar leidt enkel tot analytische uitdrukkingen in het bijzonder geval van een enkele interface. Het belangrijkste voordeel van deze methode is dat ze heel eenvoudig implementeerbaar is. Daarentegen, is ze quasi-"oncalibreerbaar" in tegenstelling tot de vorige fotometrische techniek. De relevante effecten (rotatie, ellipticiteit) zitten op een ingewikkelde manier verscholen in de periodieke intensiteitsfunctie. Terwijl in de dynamische techniek deze grootheden rechtstreeks en gecalibreerd gemeten worden.

A priori schatting van de MO sterkte – een utopie?! Het is niet echt ideaal als de magneto-optische sterkte van een MO metaal enkel kan verkregen worden door een ellipsometrische karakterisatie. Ideaal zou zijn de magneto-optische sterkte op een of andere manier ab initio theoretisch ruwweg te kunnen schatten op basis van gekende materiaalparameters of door een theoretische berekening. Het laatste is in principe mogelijk via tweede orde tijdsafhankelijke storingsrekening gecombineerd met bandstructuurberekeningen. Echter, dergelijke berekeningen zijn gigantisch moeilijk alleen al voor de zuivere ferromagnetische metalen laat staan voor hun legeringen.
Een meer empirische benadering, waarbij de magneto-optische eigenschappen gerelateerd worden via een semi-klassiek Drude-Lorentz geleidingsmodel aan gekende of eenvoudig meetbare materiaaleigenschappen zoals levensduur, toestandsdichtheden en effectieve massa's van de ladingsdrager, ... is geprobeerd door een aantal auteurs. Echter het was onmogelijk enige consequente trend vast te stellen tussen de gyrotropie parameter en deze geleidingseigenschappen. Dit bracht Busschow ertoe te concluderen dat het "ontwerp" en de keuze van interessante MO materialen onvermijdelijk de "weg van de bedelaar" moet volgen. Met andere woorden: elk nieuw MO materiaal moet onvermijdelijk ellipsometrisch gekarakteriseerd worden. Zelfs binnen een bepaald legeringssysteem is het in principe niet zo dat die legering met de hoogste saturatiemagnetizatie de sterkste MO effecten zal produceren.
De extensieve magneto-optische ellipsometrische studie van Busschow van een 200-tal MO materialen in het zichtbare en het zeer nabije IR ($\lambda < 830$nm) spectrum deed hem concluderen dat vanuit puur MO opzicht het equi-atomische $Co_{50}Fe_{50}$ een van de sterkste kandidaten is. Een vaststelling die later door ons ook bevestigd werd in het telecom-IR (zie Hoofdstuk 4).

3.1.2 Magnetische anisotropie

De studie van de magnetische eigenschappen van het gekozen CoFe systeem komt voor onze toepassing neer op het nagaan van de magnetische anisotropie van dunne CoFe films. Meer bepaald moet de planaire anisotropie bestudeerd worden van lange, smalle contactpatronen. In het ideale geval is voor onze toepassing een stabiele "in-plane" magnetizatie met een hoge remanentie in de richting loodrecht op de lengte-as van het contact gewenst. Om een dergelijke "langwerpige" permanente transversale dunne-film-magneet te kunnen realiseren, is het niet enkel nodig rekening te houden met kristallijne anisotropie eigenschappen van het gebruikte materiaal maar ook en vooral met demagnetiserende oppervlakte-effecten. Bovendien is de magnetische anisotropie van dunne metallische films sterk afhankelijk van pre- en post-depositie "behandelingen". Deze aspecten worden in theorie bestudeerd in sectie 3.1.2.

Fenomenologie van magnetische anisotropie Het bestuderen van magnetische anisotropie is in principe een magnetostatisch energieprobleem, namelijk hoeveel energie is er nodig in elke spatiale richting om het sample van een gedemagnetiseerde toestand in saturatie te brengen. Een dergelijke studie van magnetische anisotropie is echter heel omslachtig, en meestal gebeurt de karakterisatie van de anisotropie op basis van een studie van de richtingsafhankelijkheid van de magnetische hysteresiscurves. Een uitgesproken anisotropie wordt weerspiegeld door sterk vierkante hysteresiscurves in zogenaamde zachte richtingen, terwijl in de harde richtingen de hysteresiscurves eerder S-vormig en gesloten zijn. In sectie 3.1.2.2 wordt daartoe de basisprincipes van ideale magnetische hysteresis herhaald aan de hand van het Stoner-Wohlfarth probleem in een enkel magnetisch domein met uniaxiale kristallijne anisotropie. We verwijzen de lezer naar de engelstalige tekst voor de uiteenzetting over de berekening van de harde en zachte as hysteriscurves in geïdealiseerde monodomein samples. Daar wordt ook uitgelegd hoe in realistische polykristallijne samples (met een random oriëntatie van de zachte as) de hysteresis in principe enkel anisotroop kan worden indien bepaalde maatregelen getroffen worden vóór de depositie (textureren van het hostmateriaal), en/of tijdens de depositie (aanleggen van een magnetisch veld tijdens sputteren) en/of na de depositie van het ferromagnetisch materiaal (legering onder magnetisch veld). Er wordt uitgelegd hoe in principe voor onze toepassing, enkel het aanleggen van een magnetisch veld tijdens depositie zin heeft.

Vormanisotropie Meer nog dan de kristallijne anisotropie zal voor onze toepassing de anisotropie ten gevolge van de demagnetizerende oppervlaktepolen het gedrag bepalen van de hysteresiscurves in de verschillende richtingen. Inderdaad, aan elk grensvlak tussen twee verschillende (magnetische of niet-magnetische) materialen (dus bijvoorbeeld ook domeinwanden), zullen er magnetische polen gecreëerd worden tengevolge van de vereiste divergentieloosheid van de magnetische inductie. Deze magnetische polen wekken een magnetisch veld op dat het

veld dat hun creatie veroorzaakt heeft, tegenwerkt. Dit demagnetiserend veld is dus logischerwijze des te sterker naarmate de magnetostatische energie nodig om in het materiaal de magnetische dipolen in een bepaalde richting te aligneren, groter wordt. Het is intuïtief duidelijk dat dit eerder het geval is voor een situatie waarbij "veel dipolen voornamelijk naast elkaar" in een bepaalde richting moeten gealigneerd worden dan in een situatie waarbij "veel dipolen in elkaars verlengde" in een bepaalde richting moeten gealigneerd worden. Dit alles wordt uitgedrukt door de zogenaamde dimensieloze demagnetizatiefactor N_{dem} ($0 < N_{\text{dem}} < 1$) die de sterkte van het demagnetizerend veld relateert met de sterkte van de magnetizatie in die richting, $H_{\text{dem}} = -N_{\text{dem}}M$. In sectie 3.1.2.3 worden benaderde uitdrukkingen gegeven voor de demagnetizatiefactoren van rechthoekige dunne film (monodomein) samples.
De uiteindelijke invloed van het demagnetizerend veld op de anisotropie vertaalt zich in een schuiftransformatie van de hysteresiscurves, die een rechtsreeks gevolg is van het feit dat het interne veld kleiner is dan het aangelegde veld tengevolge van het demagnetizatieveld.

Karakterisatie van magnetische anisotropie Uiteindelijk wordt er in sectie 3.1.2.4 uiteengezet welke experimentele methodologie er voorhanden is om magnetische anisotropie te karakteriseren. In principe kan dit enkel door rigoureus de anisotropieconstantes te meten door magnetostatische energiemetingen met behulp van een magnetometer of een magnetische koppelmeter op gedemagnetizeerde samples. Maar zoals hierboven uitgelegd neemt men meestal genoegen met het karakteriseren van de magnetische hysteresis in verschillende spatiale richtingen. De "kwaliteit" van de hysteresiscurve wordt uitgedrukt door middel van een dimensieloos getal S tussen 0 en 1 dat de "vierkantheid" van de hysteresiscurve uitdrukt. Er wordt op gewezen hoe de traditionele definitie van S, namelijk de remanentieratio, in principe tekortschiet, en dat het beter is een algemenere uitdrukking voor S te gebruiken gegeven door Vgl. (3.36).

Ontwerpregel Rekening houdend met het feit dat de anisotropie in de door ons bestudeerde toepassing voornamelijk bepaald zal worden door vormeffecten, wordt op basis van de benaderde formules voor de demagnetizatiefactor voor dunne rechthoekige magnetische films, een rudimentaire ontwerpregel afgeleid. Deze is gebaseerd op het minimaliseren van de schuifhoek. Als voorzichtige vuistregel wordt er vooropgesteld dat voor de magnetische contacten voor de SOA/isolator het raadzaam is een w/t verhouding te hanteren die op zijn minst groter is dan 500.

3.1.3 Ohmse ferromagnetische metaal/halfgeleidercontacten

Het laagresistief Ohms karakter van het CoFe/halfgeleidercontact is de derde en laatste karakteristiek van cruciaal belang voor de bestudeerde component. In dit opzicht moet er vermeld worden dat het realizeren van een elektrisch contact met behulp van het magneto-optisch ferromagnetisch metaal geen "conditio sine qua non" is. Er is geen enkel reden waarom het elektrisch contact voor de versterker

ook tegelijk het niet-reciprook effect moet leveren. En er zijn verschillende redenen bedenkbaar waarom dit eigenlijk geen logische keuze is. De behoefte aan een hooggedopeerde lage bandgap halfgeleidercontactlaag is het meest overtuigende contra-argument. Een dergelijke laag zal onvermijdelijk een afschermend effect hebben, en dus in zekere mate de overlap met het niet-reciproke materiaal doen afnemen. Anderzijds is het zo dat als men elektrisch contact en de niet-reciproke functie gescheiden wil houden, het bijna onvermijdelijk is de stroom lateraal te injecteren, en het niet-reciproke materiaal centraal-symmetrisch te behouden om het effect niet te doen afnemen. In elk geval lijkt de suggestie om het magneto-optisch metaal ook meteen als elektrisch contact aan te wenden nog altijd het meest elegant. Op deze manier verschilt de processing van de device nauwelijks van die van een standaard optische halfgeleiderversterker. Het is pas als zou blijken dat het verkrijgen van een Ohms cotact te veel offers aan magneto-optische zijde zou vereisen, dat het noodzakelijk wordt aan alternatieve schema's te denken.

In sectie 3.1.3.2 wordt eerst een korte herhaling gegeven van de theorie van metaal/halfgeleidercontacten. Daar wordt aangetoond dat de belangrijkste manier om een dergelijk contact een goed Ohms karakter te bezorgen gebeurt door de tunnelcomponent van de stroom dominant te maken t.o.v. thermionische diffusiecomponent. Hierdoor neemt de saturatiestroom van de Schottky diode sterk toe, en ontstaat er een ruim stroominterval waar de stroom-spanningsrelatie quasi-lineair is met een lage resistiviteit. De meest voor de hand liggende manier om de tunnelstroomcomponent te doen toenemen is door de dopering van de halfgeleider te verhogen (tot niveaus $10^{+19}\mathrm{cm}^{-3}$).
Het maximaliseren van de saturatiestroom door de Schottky barrière te verlagen is slechts schijnbaar een mogelijke techniek. De werkfuncties van de meeste metalen varieert niet sterk, en bovendien wordt de hoogte van de Schottky barrière voornamelijk bepaald door oppervlakte-vallen. Deze zorgen ervoor (zoals door Bardeen bewezen) dat voor de meeste III-V covalente halfgeleiders het Fermi-niveau gepind is op een derde van de bandgap, onafhankelijk van de werkfunctie van het metaal. Dit onderlijnt bovendien waarom het voor Ohmse contacten op III-V halfgeleiders van uitzonderlijk belang is te werken met een halfgeleidermateriaal met een zo laag mogelijke band gap.

Karakterizatie van contactresistiviteit De meting van de contactresistiviteit wordt vaak stiefmoederlijk behandeld. Vaak wordt daartoe een niet geschikte teststructuur gebruikt die leidt tot een onder- of overschatting van de resistiviteit, omdat belangrijke geometrische effecten niet in rekening gebracht worden bij het experiment. Nochtans kan het verschil tussen een specifieke contactresistiveit van bijvoorbeeld $10^{-7}\Omega\mathrm{cm}^2$ of $10^{-5}\Omega\mathrm{cm}^2$ een wereld van verschil betekenen in het gebruik van de component, ook al zijn beide waarden zeker niet groot te noemen. Het is dus van groot belang een goed model te gebruiken om de elektrostatische potentiaaldistributie te modelleren in de gebruikte teststructuur. Al te vaak wordt hiervoor een te simpele 0D of 1D benadering hiervoor gebruikt. In sectie 3.1.3.3 wordt een elektrostatisch model ontwikkeld dat met slechts beperkte benaderingen de potentiaaldistributie simuleert in 2D contacttheststructuren. Voor de details van

deze afleiding verwijzen we de lezer naar de engelstalige tekst. Essentieel is dat de potentiaaldistributie beschreven wordt door een partiële 2D Helmholtzvergelijking (zie Vgl. (3.48)) waarbij de contactresistiviteit als een parameter verscholen zit in de zogenaamde transfertlengte. Die laatste parameter drukt uit welke afstand een majoritaire ladingsdrager typische aflegt in de halfgeleiderdifussielaag onder het contact alvorens de contactinterface te overbruggen. Het is duidelijk dat deze parameter des te groter is naarmate de specifieke contactresistiviteit groter wordt. Daarnaast wordt de transfertlengte uiteraard ook bepaald door de weerstand van de halfgeleiderdiffusielaag zelf. Naarmate deze toeneemt stijgt de neiging van de ladingsdragers om "snel de contact barrière te overbruggen" en daalt dus de transfertlengte. Met behulp van dit model (en geschikte randvoorwaarden) kan de potentiaaldistributie in allerhande 2D teststructuren gesimuleerd worden, en kan dus de weerstand van deze structuren berekend worden, hetzij numeriek (in 2D contacten) hetzij analytisch (als het 2D contact vereenvoudigd kan gemodeleerd worden door een 1D model). En dit als functie van de transferlengte. Door deze gesimuleerde weerstanden te fitten (via de transferlengte) aan gemeten weerstanden op teststructuren met gekende dimensies, geometrie en halfgeleiderweerstand, kan heel nauwkeurig de contactresistiviteit geëxtraheerd worden.
In sectie 3.1.3.3 wordt uiteengezet hoe men daarbij praktisch te werk gaat, namelijk welke structuren hiervoor typisch in aanmerking komen, hoe deze al dan niet gehinderd worden door parasitaire effecten, en hoe een weerstandsmeting zich vertaalt in een equivalente simulatiewaarde (zie daartoe Fig. 3.11). De belangrijkste conclusie is dat de structuur die het beste compromis levert tussen experimentele eenvoud en robuustheid ten opzichte van parasitaire effecten gegeven wordt door de zogenaamde Cross Kelvin Resistor structuur. Deze vereist onvermijdelijk een volledige 2D simulatie. Een Matlab tool die de 2D partiële differentiaalvergelijking oplost met behulp van eindige differenties werd daartoe ontwikkeld.

Standaard Ohmse p-contacten – gerapporteerde resistiviteiten De karakterisatie van de CoFe contacten volgens de procedure geschetst in voorgaande sectie zal worden gerapporteerd in Hoofdstuk 4. Als referentie wordt in Tabel 3.2 in sectie 3.1.3.4 een overzicht gegeven van de performantie van standaard metaal/halfgeleidercontacten op het InGaAsP/InP materiaalsysteem.
Het meest populaire contactschema maakt gebruik van een p^{++} ($>10^{+19}cm^{-3}$) gedopeerde $In_{0.53}Ga_{0.47}As$ laag (zijnde de InP-roosteraangepaste samenstelling met de laagste bandgap in het InGaAsP systeem, namelijk $\approx$0.75eV) en een metallische dubbellaag bestaande uit een metaal met een zo laag mogelijke werkfunctie (Ti en W zijn populaire keuzes) en een nobelmetaal ter bescherming en als stabilisator. Vaak wordt het contact na processing onderworpen aan een snelle thermische legeringsstap met de bedoeling lage-gap compounds te creëeren aan de interface door een thermisch gestimuleerde metallurgische conversie. Recordwaarden voor de resistiviteit zijn in de orde van 10^{-7}–$10^{-8}\Omega cm^2$. Er wordt niet verwacht dat CoFe contacten deze waarden zullen benaderen, temeer daar onze specifieke toepassing vereist dat de halfgeleidercontactlaag transparant is voor de gebruikte golflengte, met andere woorden een band gap vertoont hoger dan 0.95eV. Maar als er als vage eis gesteld wordt dat het contact in onze devices zeker niet

meer dan 1Ω mag bijdragen tot de totale serieweerstand, dan is een contactresistiviteit voor de CoFe-contacten van de orde van $10^{-5}\Omega\text{cm}^2$ (rekening houdend met een device lengte van verschillende mm's) zeker niet slecht te noemen.

3.1.4 Conclusies

De drie uiteenlopende functies die het ferromagnetisch metaal moet vervullen zijn uitgebreid bestudeerd van fenomenologisch en theoretisch oogpunt. Voor alle drie de aspecten is een gefundeerde methodologie aangereikt om het gedrag te schatten naar waarde. De belangrijkste conclusie echter is dat het een utopie is het ferromagnetisch metaal te ontwerpen of beter te kiezen op basis van eerste principes. De voornaamste reden hiervoor lijkt de puur praktische overweging te zijn dat theoretische modellen hun waarde zo goed als volledig verliezen eens het metaal niet meer perfect kristallijn is maar polykristallijn wordt. Maar zelfs voor perfect kristallijne materialen is het a priori theoretisch voorspellen van de magneto-optische sterkte een utopie, zelfs met semi-klassieke empirische modellen. De keuze voor CoFe moet dus in dit opzicht meer gezien worden als één van gezond verstand (en een heel klein beetje een gok).

Als afsluitende opmerking moet er op gewezen worden dat de behandeling steeds is uitgegaan van het geïdealiseerde geval van een perfect vlak ongestoord grensvlak tussen twee bulk materialen. Met andere woorden er werden geen oppervlakte-effecten in rekening gebracht (met uitzondering natuurlijk van de behandeling van Ohmse contacten waar er op gewezen werd dat precies die oppervlakte-effecten het Fermi niveau pinnen). Voor de magnetische anisotropie is het duidelijk dat oppervlakte effecten enkel een rol zullen spelen als er een belangrijke graad is van epitaxie. Zoniet zal de bijdrage tot de anisotropie van de korrelgrenzen veel belangrijker zijn. Epitaxie van CoFe op InP wordt niet echt verwacht gelet op hun grote verschil in roosterconstante. Voor de magneto-optische eigenschappen kan gesteld worden dat in principe de toestandsdichtheid in het metaal (of m.a.w. de geleidingseigenschappen en dus de optische en magneto-optische constanten) aan het grensvlak sterk gewijzigd zal zijn. Maar een dergelijke perturbatie zal zich niet verder dan enkele monolagen uitstrekken. In het slechtste geval onstaat er een absorberende metallische laag van enkele nm die zijn ferromagnetische en magneto-optische eigenschappen verloren heeft. In volgend Hoofdstuk zal aangetoond worden dat een dergelijke dode magnetische laag niet veel dikker is dan 2nm voor CoFe/InGaAs(P) contacten, m.a.w. voldoende dun om geen grote invloed te hebben op het gedrag van de component.

3.2 Bandgap engineering: roostermisaangepaste quantumputten

Naast een geschikt een ferromagnetisch metaal/halfgeleidercontact steunt de goede werking van het bestudeerde geïntegreerde isolatorconcept op de beschikbaarheid van een sterke polarizatie-selectieve actieve laag. Er is op gewezen in sectie 1.4.2 dat dit de groei vereist van roostermisaangepaste multi-quantumput lagen (MQW),

meer bepaald met een materiaal voor de quantumput (QW) dat een lagere roosterconstante heeft dan die van het omringend materiaal. De theoretische studie van QW's onder trekspanning (of drukspanning) is een volledig onderzoeksgebied op zich. In sectie 3.2.2 wordt dan ook enkel op kwalitatieve en semi-kwantitatieve manier de theoretische principes uiteengezet die bewijzen hoe spanning de winsteigenschappen van QW's "engineert". We verwijzen de lezer naar de samenvattende uiteenzetting van de engelstalige tekst, en volstaan hier met een schets van de belangrijkste principes. Een goede inleidende uiteenzetting van de hier geschtetste principes van bandstructuur engineering vindt de lezer in [15] en in de papers van Yablonovitch en Adams, de bedenkers van strain engineering [16, 17].

3.2.1 Principes van QW's onder trekspanning

Optische winst Het vertrekpunt is de bekende formule voor de optische winst in een halfgeleidermateriaal verkregen door de alombekende "gouden regel van Fermi" in een tweede-orde tijdsafhankelijke quantummechanische storingsbereking van de interactie van een elektromagnetisch veld met de bandenstructuur van een halfgeleidermateriaal onder populatie-inversie.

$$g(\hbar\omega) = \left(\frac{1}{\hbar\omega}\right)\frac{\pi e^2\hbar}{\epsilon_0 c m_0^2}\frac{n_g}{n^2}|M_T|^2\rho_{\mathrm{red}}(\hbar\omega - E_g)(f_c - f_v) \tag{1}$$

Deze formule leert ons dat het maximaliseren van de winst vereist dat de populatie-inversie moet optreden bij energieniveaus waar de gereduceerde toestandsdichtheid[4] zo groot mogelijk wordt. In een bulk materiaal (waar de toestandsdichtheid gradueel toeneemt als een vierkantswortel van de energieseparatie tussen de recombinerende niveau's) is dit enkel mogelijk door eerst "veel lage niveaus op te vullen". In een QW heeft de toestandsdichtheid echter een trapstructuur. De winst is altijd maximaal aan de rand van de banden en krijgt dus een heel steile sprong van zodra het materiaal transparantie bereikt. De quasi-Fermi niveaus moeten niet ver voorbij transparantie "geduwd" worden om de winst snel te doen toenemen. Dit drukt uit dat de zogenaamde differentiële winst van een QW vele malen groter is dan die van een bulk materiaal. Anderzijds is het echter zo dat door de trapstructuur van de toestandsdichtheid het aantal ladingsdragers dat nodig is om transparantie te bereiken ($f_c - f_v = 0$) veel groter is, en vooral dat dit nog wordt verergerd door het verschil in effectieve massa tussen de gaten en de elektronen. Hierdoor ligt het quasi-Fermi niveau van de lichtste ladingsdragers bij transparentie al redelijk diep in de conductieband (om ladingsneutraliteit te garanderen). Door deze asymmetrie in massa zal bovendien de gereduceerde toestandsdichtheid die van de lichtste band benaderen. M.a.w. het winstgedrag wordt bepaald door het bezettingsgedrag van de lichtste band. En aangezien bij transparentie het quasi-Ferminiveau al redelijk diep in deze lichte band ligt, zal de differentiële winst een stuk lager zijn (dan in het hypothetische geval met even zware gaten als elektronen). Dit alles is geschetst in Fig. 3.13.

[4] die voor de eenvoud evenredig kan beschouwd worden met de bandtoestandsdichtheid. Dit is echter enkel rigoureus geldig voor parabolische banden.

Het engineeren van de winst in een QW heeft dus als voornaamste opdracht de bandenstructuur zo symmetrisch mogelijk te maken, of m.a.w. de massa's van de ladingsdragers trachten gelijk te maken. En bovendien de massa's zo klein mogelijk te maken. Een kleinere effectieve massa leidt immers tot een lager toestandsdichtheid en dus tot lagere ladingsdragerdichtheid bij transparantie. In het algemeen is de uiteindelijk doelstelling te komen tot een bandenstructuur die gekarakteriseerd wordt door een ideale trapvormige bandtoestandsdichtheid met gelijke "traphoogtes" in beide banden. Om te begrijpen hoe dit gedrag kan benaderd worden door spanning te introduceren in de QW, is het nodig te herhalen hoe de bandenstructuur van een QW eruitziet. Dit wordt op een semi-kwantitatieve manier uiteengezet in sectie 3.2.2.2. We volstaan hier met het herhalen van de belangrijkste vaststellingen.

QW bandstructuur Het kernpunt in de discussie is de structuur en meer bepaald de effectieve massa van de valentieband. Het is geweten dat de valentieband in een bulk materiaal aan de rand ontaard is in een lichte gatenband en een zware gatenband, en dat weg van de bandrand deze twee banden door de spin-orbit interactie met elkaar koppelen. Deze koppeling is echter niet drastisch in een bulk materiaal en de hoogst gelegen band zal steeds een uitgesproken zwaar karakter behouden. In een QW zal door het quantizatieprobleem de koppeling echter veel sterker worden doordat de vereiste resonantie voor de eigenoplossingen in de QW opsluitingsrichting een subtiele lineaire combinatie van alle bulkoplossingen veroorzaakt. Dus in een roosteraangepaste QW structuur veroorzaakt de quantizatie twee voor de optische winst nadelige effecten. De zware gatenband ligt steeds ver boven de lichte gaten band, en de heel sterke band-mixing zorgt ervoor dat weg van de bandrand de kromming van de banden sterk varieert! Met andere woorden als functie van de recombinatie-energie zal de toestandsdichtheid verre van stapvormig zijn. Een gedrag dat dus helemaal niet het hierboven geschetste ideale gedrag benadert. Merk wel op dat in vergelijking met de bulkstructuur een QW actieve laag wel voordeliger blijft, wegens de steile onset van de winst (door datzelfde gedrag van de toestandsdichtheid). Een voorbeeld van een typische bandenstructuur in een QW is geschetst in Fig. 3.15.

QW bandstructuur bij roostermisaanpassing Om de winstkarakteristieken van QW's te verbeteren moet er dus een manier gevonden waardoor reeds in de bulkstructuur de ontaarding tussen de zware en lichte gatenbanden opgeheven wordt zodat vertrekkende van deze situatie bij de overgang naar een quantumstructuur de mixing tussen de banden sterk gereduceerd wordt. We gaan hier niet in op de details (zie daarvoor de engelstalige tekst), maar er kan bewezen worden dat door een uniaxiale vervorming van het rooster te creëren door het materiaal met een roostermisaanpassing te groeien, precies dit effect verkregen wordt. Bovendien wordt er ook bewezen dat dit hand in hand gaat met een wijziging van de effectieve massa van de banden. Als de vervorming een afname van de roosterconstante veroorzaakt zal de zware gatenband bovenaan te komen liggen en zal de effectieve massa van deze gaten dalen in het vlak van de vervorming. Bij een trekvervorming zal de lichte gatenband bovenop te komen liggen, en zal zijn in-

plane massa weliswaar toenemen maar opmerkelijk lager blijven dan de massa van een zwaar gat in het roosteraangepaste geval. Dit alles is schematisch voorgesteld in Fig. 3.16. Spanning verlaagt dus niet enkel de hoeveelheid van band-mixing maar verlaagt ook globaal gezien de massa's van de hoogst gelegen band (en verhoogt dus zoals gewenst de symmetrie tussen de conductie- en de valentieband)! De positieve invloed van spanning op de bandenstructuur is geschetst in Fig. 3.17.

Polarizatie-selectieve versterking Naast de positieve invloed op de bandenstructuur en dus op de winstkarakteristieken, is veruit het belangrijkste gevolg van het introduceren van spanning in een QW de invloed op de transitiesterkte tussen de conductie- en de valentieband. Dit is een gevolg van het feit dat nu de Bloch functie van de hoogst gelegen valentieband een uitgesproken zwaar of licht karakter heeft (voor een breed bereik aan transitie-energieën omdat band-mixing sterk onderdrukt is), en dat in combinatie met het feit dat de golfvector van de recombinerende conductie- en valentietoestanden door het quantizatie-effect een uitgesproken spatiale gemiddelde richting heeft (namelijk parallel aan de quantizatie-as). In sectie 3.2.2.3 wordt bewezen hoe als gevolg hiervan in een QW onder drukspanning de TE transitiesterkte in de buurt van de bandrand toeneemt met 50% t.o.v. een bulk actieve laag[5] terwijl de TM transitiesterkte quasi volledig onderdrukt is. Als er hoegenaamd geen band-mixing zou optreden zou deze onderdrukking zelfs perfect zijn. In een QW onder trekspanning is de toename van de TM transitiesterkte zelfs 100% (en kan bij toenemende spanning zelfs toenemen tot 200%), terwijl de TE transitiesterkte met 50% onderdrukt wordt. Een meer rigoureus kwatitatief bewijs van deze uitspraken wordt gegeven in de engelstalige tekst.

Kritische dikte en type II gedrag Al het bovenstaande mag het gebruik van spanning in QW's als een wondermiddel doen lijken, toch zijn er belangrijke beperkingen. De meest voor de hand liggende beperking is de mechanische limiet. Een dunne laag kan niet onbeperkt spanning opnemen of kan bij een bepaalde spanning niet onbeperkt dik gegroeid worden en zal vroeg of laat relaxeren door de vorming van dislocatie-defecten. Dergelijke dislocaties gedragen zich als verliesbron zowel optisch als elektrisch en vernietigen uiteraard het goede gedrag van een QW. Matthews en Blakeslee hebben een rigoureuze limiet berekend voor de kritische spanning of de kritische dikte waarboven een QW onder spanning zal relaxeren. De formule van Matthews is niet handig in gebruik omdat ze de kennis vereist van heel wat kristallografische parameters van het materiaal. Een meer bruikbare vuistregel werd afgeleid door Andersson [18] en geldt voor een ruim gamma aan III-V halfgeleiders. Hij stelde vast dat een QW onder spanning defectvrij kan gegroeid worden zolang het dikte-spanning produkt kleiner blijft dan 10–20nm%. Dit geldt voor één enkele laag onder spanning. In een MQW gebied accumuleert de spanning en kan er toch relaxatie optreden ook al blijft elke laag onder de limiet van Andersson. In principe kan de totale spanning van een

[5] waarvoor eveneens in sectie 3.2.2.3 bewezen wordt dat slechts 1 uit de 3 ladingsdragers bijdraagt tot versterking van een willekeurige polarizatie.

roostermisaangepaste MQW laag tot nul herleid worden door de bufferlagen een compenserende spanning te geven. Dit vermijdt natuurlijk niet dat elke individuele laag aan de Andersson regel moet blijven voldoen. Bovendien zal in dit geval ook de valentiebandstructuur van de bufferlagen wijzigen en dus ook de quantummechanische berekening van de valentiebandniveaus in de QW's!

Een tweede beperking op het introduceren van spanning in QW's zit in het mogelijk optreden van zogenaamd type II gedrag. Onder type II gedrag wordt verstaan dat de gaten en de elektronen in spatiaal verschillende gebieden van de actieve laag gebonden zitten doordat de conductieband offset of de valentiebandoffset negatief is geworden. Vaak wordt dit verkeerdelijk gezien als een gevolg van de invloed van de spanning op de bandoffsets. Dit is echter niet juist. Bowers heeft theoretisch en experimenteel aangetoond dat de variatie van de band offset als functie van spanning meer te wijten is aan de invloed van de compositionele variatie op de bandstructuurparameters dan aan de invloed van de spanning zelf. Inderdaad, om spanning en dus roostermisaanpassing te introduceren in een QW moet de samenstelling van het materiaal gewijzigd worden. De veranderingen van de bandalignering en de band gap als gevolg hiervan zijn een grootte-orde groter dan diezelfde variaties puur ten gevolge van de vervorming alleen. In elk geval is het zo dat in het InGaAsP materiaal systeem (met zijn zeer lage CBO:VBO verhouding) de conductieband snel negatief kan worden bij trekspanning, omdat bij dalende In molfractie de band gap snel stijgt en de CBO daalt.

3.2.2 Materiaalsystemen

Voor de realizatie van een QW onder trekspanning met een emissiegolflengte rond 1300nm zijn er een aantal halfgeleidermateriaalsystemen beschikbaar. Fig. 3.18 toont dat op GaAs substraten dit enkel mogelijk is door het gebruik van het nog steeds in volle ontwikkeling zijnde GaInNAs materiaalsysteem. Op het InP substraat zijn er meerdere materiaalsystemen beschikbaar. Op het eerste zicht lijkt het het eenvoudigst gebruik te maken van het ternaire $In_xGa_{1-x}As$ materiaalsysteem. Dit systeem heeft echter slechts een enkele vrijheidsgraad, waardoor het quasi onmogelijk wordt tegelijk een bepaalde emissiegolflengte bereiken en nog een zekere vrijheid te hebben in de dikte van de wells. Thijs heeft bovendien berekend dat door de lage CBO in het InGaAsP/InP materiaalsysteem de vervorming kleiner moet blijven dan 1.6% om type II gedrag te vermijden. Als gevolg daarvan is de bulk band gap steeds kleiner dan 0.95 en kan de emissiegolflengte enkel gehaald worden door de QW's heel dun (orde 2nm) te maken. Wat noch technologisch evident is en noch vanuit optisch standpunt tot goed gedrag zal leiden. Bovendien is het fundamenteel onmogelijk om met dit ternaire systeem het 1.55μm gebied te halen met een QW onder rek.
Het is beter gebruik te maken van quaternaire systemen aangezien die een extra vrijheidsgraad introduceren, en dus toelaten de vervorming en de emissiegolflengte afzonderlijk te variëren (binnen zekere grenzen). Dit is schematisch voorgesteld in Fig. 3.20 voor de twee meest voor de hand liggende quaternaire systemen op InP, nl. $In_{1-x}Ga_xAs_yP_{1-y}$ en $Ga_xAl_yIn_{1-x-y}As$. Deze grafieken plotten de bulk

band gap energy contouren en vervormingscontouren in functie van de compositionele parameters. In combinatie met de kennis van de bandalignering in beide materiaalsystemen kunnen deze grafieken gebruikt worden om de gequantiseerde energieniveaus, bandenstructuren en Bloch functies te bereken volgens de procedure van sectie 3.2.2.2. Dit is geen evidente opdracht. Vooral de kennis van de bandalignering als functie van vervorming en compositie stelt een zware opdracht. Bowers heeft hiertoe een analytisch model ontwikkeld [19]. De belangrijkste conclusie is dat er verwacht mag worden dat het $Ga_xAl_yIn_{1-x-y}As$ systeem zich heel wat beter zal gedragen dan het $In_{1-x}Ga_xAs_yP_{1-y}$ systeem, omwille van zijn veel hogere CBO. Dit zorgt voor een betere elektronenconfinement in de QW's en dus een betere interne quantumefficiëntie.

3.2.3 Ontwerp en karakterizatiemethodes van MQW structuren onder spanning

Ontwerp In principe is het mogelijk met de aangereikte procedures een MQW onder spanning volledig theoretisch te modelleren. Dit is echter geen evidente opdracht, want hiertoe is het achtereenvolgens nodig de bandalignering te kennen, het quantummechanisch probleem op te lossen om de energieniveaus bij de bandrand te kennen, de bandstructuurdispersie te berekenen (en bijhorende Blochfuncties en envelopes), de optische transitiematrixelementen te berekenen, en de quasi-Ferminiveaus doorheen de heterojunctie te kennen. Bovendien zijn er voor een enkel QW reeds niet minder dan 5 parameters die moeten geoptimaliseerd worden (2 compositionele parameters voor de QW en 2 voor de buffer, en de dikte van de QW). En al deze parameters zijn sterk gecorreleerd door de randvoorwaarden opgelegd door de gewenste emissiegolflengte en de vervormings- en type II-limieten. Tenslotte zal een dergelijke berekening enkel de winst opleveren in functie van de Ferminiveauseparatie (of dus de ladingsdragerdichtheid in de QW). De ladingdragerdichtheid in de wells is meestal niet gekend, wel de geïnjecteerde stroomdichtheid. Maar deze laatste is niet eenvoudig te relateren aan de ladingsdragersconcentratie zonder kennis van de niet-radiatieve recombinatiefenomenen, namelijk Auger en Shockley-Read-Hall recombinatie. Het is dus quasi-onmogelijk het winstgedrag van een QW onder spanning theoretisch te optimalizeren. Beperkte pogingen hiertoe leidden enkel tot een theoretische bevestiging van wat door vele auteurs reeds empirisch was vastgesteld, namelijk dat het winst-stroomverband in een QW voor voldoende lage stromen kan benaderd worden door een logaritimisch verband

$$g = G_0 \frac{J_0}{\eta_i} \ln(\frac{\eta_i J}{J_0}), \tag{2}$$

Ondanks al het bovenstaande zijn er toch een beperkt aantal theoretische ab initio berekeningen gebeurd van de optische winst van InGaAsP wells onder trekspanning. Deze dienden louter om een indicatie te hebben van de theoretisch haalbare winstniveaus, omdat op het tijdstip van het eerste generatieontwerp van de SOA/isolatoren er nog geen experimentele verificatie van deze winst voorhanden was. De resultaten van deze 6×6 Luttinger-Kohn $\mathbf{k} \cdot \mathbf{p}$ simulaties gecombineerd met de berekening van de optische transitiematrixelementen

zijn voorgesteld in Fig. 3.21. De gesimuleerde structuur is een 12 nm dikke QW ($In_{0.48}Ga_{0.52}As_{0.78}P_{0.22}$, $\lambda_g = 1.3\mu m$) onder -1.1% rekspanning omringd door 2 barrierlagen onder een compenserende +0.3% drukspanning ($In_{0.93}Ga_{0.07}As_{0.22}P_{0.78}$, $\lambda_g = 1.05\mu m$). De berekende waarden voor de optische winst lijken veelbelovend (5000/cm TM materiaalwinst voor een $N = 4 \times 10^{+18}\text{cm}^{-3}$). Het is echter niet evident in te schatten met welke stroomdichtheid dit overeenkomt, en latere experimentele resultaten zullen bevestigen dat dit winstniveau voor een InGaAsP materiaalsysteem praktisch niet haalbaar is.

Karakterizatiemethodes Aangezien de vorige sectie uitvoerig aantoont dat een ab initio optimalizatie van een MQW onder spanning een quasi onmogelijke opdracht is, is ervoor gekozen de MQW op zeer pragmatische wijze te ontwerpen. Namelijk door een hele reeks actieve lagen te groeien met als gevarieerde parameters de QW vervorming, het aantal QW's, de spanningscompensatie in de bufferlagen, ... en deze verschillende actieve lagen met elkaar te vergelijken op basis van hun winstparameters volgens Vgl. (2) (namelijk de differentiele winst G_0, de transparantiestroomdichtheid J_0, en de interne quantum efficiëntie η_i). Om deze te meten kan men op verschillende manieren te werk gaan. Een vaak gebruikte methode voor winstkarakterizatie is de Hakki-Paoli techniek. Hierbij worden uit de Fabry-Perot resonanties in het spontane emissiespectrum van een actieve caviteit het modale winst/verliesspectrum van de caviteit afgeleid. De Hakki-Paoli methode kent een aantal belangrijke nadelen, waarvan de belangrijkste zijn de onmogelijkheid om de interne quantumefficiëntie te meten, en de moeilijkheid om het intern verlies te meten. Het is enkel mogelijk dit verlies in de lange-golflengtelimiet af te leiden en aan te nemen dat het constant blijft over het volledig opgemeten spectrum.
Een meer geschikte methode om de winst van een actieve laag te karakteriseren is door lasermetingen. Door de differentiële efficiëntie te meten als functie van de lengte van de laser kan het intern verlies bij de lasergolflengte afgeleid worden, alsook de interne quantumefficiëntie. In combinatie met de verschillende gemeten drempelstroomdichtheden leidt dit tot een winst-stroom relatie waaraan het logaritmisch model kan gefit worden (zie Vgl. (3.76) en (3.77)). Deze methode lijdt dus niet onder de nadelen van de HP methode. Het intern verlies wordt rechtstreeks gemeten op de gewenste golflengte en de interne quantum efficiëntie volgt rechtstreeks als de $L = 0$ limiet van de $(1/\eta_d, L)$-metingen. De bovenstaande procedure staat of valt echter met de veronderstelling dat noch het intern verlies, noch de interne quantumefficiëntie afhankelijk zijn van de lengte van de lasercaviteit. En deze veronderstelling wordt steeds twijfelachtiger naarmate de lengte afneemt. Bij korte lasercaviteiten stijgt immers de nodige ladingsdragersconcentratie voor laserwerking heel snel. Dit zorgt ervoor dat het belang van niet-radiatieve recombinatiemechanismen snel toeneemt (Auger recombinatie $\sim N^3$), net als het belang van optische absorptie door vrije ladingsdragers en intervalentiebandabsorptie. In feite moet er dus in bovenstaande extractiemethode eigenlijk gewerkt worden met $\alpha_{\text{int}}(L)$ en $\eta_{\text{int}}(L)$ [20], maar het berekenen van deze functionele afhankelijkheid kan enkel met geavanceerde modellen die zelf-consistent simultaan het stroomtransport door de heterostructuur, de warmteflux, de optische winst en de optische

golfgeleiding oplossen.
In elk geval moet voorzichtig omgesprongen worden met de winstkarakterizatie van een actief materiaal. De vaststellingen van Piprek maken duidelijk dat in feite in de logaritmische winst-stroomrelatie van Vgl. (2) er beter gewerkt wordt met een stroomdichtheid afhankelijke $\eta_{\text{int}}(J)$ en $J_0(J)$, of m.a.w. dat de logaritmische relatie met constante parameters slechts geldt in de buurt van de transparantie-stroomdichtheid.

3.3 Ferromagnetisch-metaal-bedekte-MQW-halfgeleiderversterker: ontwerp van de lagenstructuur

3.3.1 Inleiding

De uiteindelijke taak bestaat erin na te gaan of onder de randvoorwaarden vastgesteld na de theoretische studie van de subcomponenten, er een lagenstructuur voor de SOA/isolator kan ontworpen worden die een voldoende sterke niet-reciproke absorptieshift veroorzaakt zonder de nood aan een te hoge stroominjectie. Een aantal inleidende opmerkingen zijn op hun plaats hierbij. Eerst en vooral zijn alle simulaties uitgevoerd in een 1D model, m.a.w. een slabgolfgeleidermodel, en zijn de 2D simulaties beperkt tot het vinden van de etsdiepte die de golfgeleiders monomodaal houdt tot op praktische breedtes van 2.5 μm. De optimalizatie van het niet-reciprook lagenontwerp wordt voornamelijk bepaald door het ontwerp van de transversale lagenstructuur, met name een optimaal compromis vinden tussen opsluiting in de actieve laag en geleiding aan de metaalinterface. De laterale dimensie, meer bepaald de etsdiepte, zal enkel een algemene reductie veroorzaken van de optische geleiding ter hoogte van het magneto-optisch contact, en geen andere niet-reciproke effecten introduceren. Ten tweede zijn alle simulaties uitgevoerd met behulp van het snelle, efficiënte perturbatie-algoritme, omdat het benchmark voorbeeld van vorig hoofdstuk aangetoond heeft dat in alle praktische situaties het verschil tussen dit algoritme en de rigoureuze aanpak verwaarloosbaar is.
Er zijn twee generaties ontworpen. Op het ogenblik van de eerste generatie waren de “randvoorwaarden” voor de subcomponenten enkel theoretisch geëvalueerd (met uitzondering van het Ohms contact). Er was een theoretische kennis van de geïdealiseerde winstkarakteristieken van de MQW actieve laag onder trekspanning (zie Fig. 3.21), en enkel een vage indicatie van de grootte-orde van de magneto-optische gyrotropie van CoFe legeringen. Het was wel theoretisch vastgesteld dat er een goed Ohms contact kon gevormd worden met een slechts beperkte optische absorptie door een quaternair/ternair hybride halfgeleiderlaag contactlaag te gebruiken (zie verder), en dat het in principe mogelijk was de negatieve effecten van magnetische vormanisotropie tegen te gaan door een voldoende breed planair ferromagnetisch contact te gebruiken. Bij de tweede generatie waren alle “subcomponent-randvoorwaarden” experimenteel vastgelegd.

3.3.2 Generatie 1: InGaAsP/$\mathbf{Co_{90}Fe_{10}}$ isolator

Bij de eerste generatie werd er gekozen voor het InGaAsP materiaalsysteem voor de MQW onder trekspanning. Het is pas later dat er experimenteel werd vast-

gesteld dat het AlGaInAs materiaalsysteem veruit de ideale kandidaat is. Het gebruikte MO materiaal was het op dat moment enige technologisch beschikbare, maar experimenteel niet gekarakteriseerde, $Co_{90}Fe_{10}$.
De details van de te optimaliseren lagenstructuur zijn gegeven in Tabel 3.6. Enkel de diktes van de InP bufferlaag en die van de twee geleidingslagen rondom het MQW gebied kunnen gevarieerd worden. De samenstelling van de MQW lagen en de geleidingslagen zijn "onaanraakbaar" om de theoretische berekende winst-eigenschappen en elektrische injectie in het actieve gebied niet in het gedrang te brengen. Hetzelfde geldt voor de dikte en de samenstelling van de hoog-gedopeerde halfgeleidercontactlaag.
Gelet op de vele onzekerheden betreffende het winstgedrag van de gebruikte MQW (en in feite ook van de magneto-optische en optische eigenschappen van $Co_{90}Fe_{10}$), is een alternatieve ontwerpstrategie gebruikt. In plaats van te zoeken naar het punt in de driedimensionale parameterruimte (gevormd door de drie diktes) met behulp van een goed gedefinieerde "Figure-of-Merit" (FoM) (zie daartoe Vgl. (3.78)), is er eerder voor gekozen om bij een vast gekozen InP buffer dikte een vermoedelijk theoretisch haalbaar winstniveau vast te leggen en onder die vooronderstellingen de dikte van de geleidingslagen te variëren voor maximale niet-reciproke absorptie. Op deze manier zijn vijf lagenstructuren ontworpen die elk als het ware een zogezegde optimaal ontwerp opleveren afhankelijk van de winst die beschikbaar zou kunnen zijn in de MQW. Het zwakke punt in de procedure is natuurlijk de link (of beter het gebrek daaraan) tussen InP bufferdikte en veronderstelde theoretisch haalbare winst. Inderdaad, eenzelfde winst wordt misschien nog optimaler (in de zin van een FoM) gebruikt bij dunnere (of dikkere) InP bufferlagen. Een ietwat rigoureuzer ontwerp op basis van FoM optimalizatie bevestigde echter later dat de gekozen combinaties van InP dikte en haalbare winst verrassend goed overeenkwamen met de voorspellingen van deze iets optimalere ontwerpstrategie. Voor meer details verwijzen we de lezer naar de engelstalige tekst.

De vijf ontworpen lagenstructuren en hun gesimuleerde niet-reciproke gedrag (binnen de grenzen van de onzekerheid op de magneto-optische constanten van CoFe) zijn samengevat in Tabel 1.

3.3.3 Generatie 2: AlGaInAs/$Co_{90}Fe_{10}$ isolator

Het ontwerp van de lagenstruktuur van de tweede generatie gebeurde met een veel nauwkeurigere kennis van de eigenschappen van de subcomponenten. Er was experimenteel een winst-stroom relatie beschikbaar voor AlGaInAs MQW onder trekspanning en de optische en magneto-optische constanten van $Co_{90}Fe_{10}$ waren gemeten met voldoende nauwkeurigheid (zie verder). Dit alles liet toe de lagenstructuur (waarvan de details opgesomd zijn in Tabel 3.8) te ontwerpen volgens een FoM type strategie. Bovendien kan in deze FoM nu ook effectief een realistische gebruiksparameter van het device gebruikt worden, namelijk de nodige stroomdichtheid in plaats van de niet-toegankelijke QW materiaalwinst. Dit laat toe de FoM te definiëren in functie van de belangrijkste deviceparame-

design	t_{InP}	$t_{\mathrm{SCH_l}}$ (nm)	$t_{\mathrm{SCH_u}}$ (nm)	$\Delta\alpha$ (dB/mm)	g_{QW} (1/mm)
safe	450	120	100	1.9–2.8	300
normal	300	150	150	3–5	450
risk	250	155	170	4–6	600
6QW	300	100	50	4–5.5	300
9QW	300	50	50	3.5–5.3	170
CoFe	t_{CoFe} (nm)	Q_{CoFe} worst case		Q_{CoFe} best case	
	100	0.02+i0.02		0.03+i0.03	

Tabel 1: Ontwerp van de 5 eerste-generatie demonstratoren. De benamingen van de 3QW ontwerpen drukken uit in hoeverre een transparante werking van de corresponderende lagenstructuur steeds minder zeker wordt (op basis van de winstberekeningen van Fig. 3.21).

ters, namelijk de lengte van het device om een bepaald isolatie niveau te halen en de totale absolute benodigde stroom om voor datzelfde isolatieniveau een voorwaarts transparante werking te bekomen. Als vooropgesteld wordt dat dit produkt moet geminimaliseerd worden bekomen we volgende FoM definitie.

$$FoM = I_{\mathrm{tot,tr}} \times L \sim J_{\mathrm{tr}} L^2 = \frac{J_{\mathrm{tr}}}{(\Delta\alpha)^2} \tag{3}$$

Gebruik makend van deze FoM definitie is de lagenstructuur geoptimaliseerd over de driedimensionele parameterruimte gevormd door de InP bufferdikte en de diktes van de twee geleidingslagen. Er is geen globale optimalizatie uitgevoerd maar eerder een soort van "sampling" procedure waarbij de evolutie van de FoM bij een vaste combinatie van diktes voor de twee geleidingslagen "getrackt" werd naarmate de dikte van de InP bufferlaag afneemt (en dit dus "gesampled" voor verschillende combinaties van de diktes voor de geleidingslagen). Op die manier wordt meer fysisch inzicht verworven in het gedrag van het device als functie van zijn belangrijkste parameter, zijnde de dikte van de InP bufferlaag. Uiteraard is deze sampling procedure niets minder dan een economischere versie (qua rekentijd) van een globale optimalizatie. Voor meer details verwijzen we de lezer naar de engelstalige tekst.
Een opmerkelijke conclusie van deze simulatiestrategie is de uitgesproken tolerantie van de lagenstruktuur van dit specifieke device t.o.v. de dikte van de geleidingslagen. Belangrijker is echter de vaststelling dat er duidelijk een punt optreedt waar het hoegenaamd geen zin heeft de dikte van de InP bufferlaag verder te doen afnemen[6] en dat dit punt rechtstreeks kan gerelateerd worden aan het saturerende gedrag van de logaritmische winst-stroomdichtheid van de actieve laag. Dit duidt er duidelijk op dat de beperkende factor in de bestudeerde SOA/isolator niet zozeer de sterkte van het magneto-optisch metaal is, maar eerder het winstgedrag van de actieve MQW onder trekspanning, want het is dit gedrag dat bepaalt vanaf welke dikte voor de InP bufferlaag niet-reciprociteit saturatie begint op te treden!

[6] Dit punt komt uiteraard overeen met het minimum van de FoM curve.

Er zijn drie lagenstrukturen ontworpen. Alle drie hebben ze dezelfde combinatie van diktes voor de geleidingslagen. Dit betekent uiteraard dit slechts één ervan optimaal zal zijn in de zin van de FoM. De twee andere zijn toegevoegd om na te gaan in hoeverre het mogelijk is om eventueel toch tot hogere niet-reciproke absorptie te komen ten koste van een veel hogere (niet optimaal benutte) stroomdichtheid, en vooral of het stroom-winst model zwaar overschat wordt bij deze hoge stroomdichtheden (herinner de discussie van de vorige subsectie). De lagenstrukturen zijn gegeven in Tabel 2.

t_{InP} (nm)	$t_{\mathrm{SCH_l}}$ (nm)	$t_{\mathrm{SCH_u}}$ (nm)	$\Delta\alpha$ (dB/mm)	J_{tr} (kA/cm^2)
345	0	15	5.6	6
295	0	15	8.25	20
280	0	15	9.55	30

Tabel 2: Ontwerp van de drie tweede-generatie 9QW InAlGaAs demonstratoren. Enkele de eerste lagenstruktuur is optimaal in de zin van de FoM definitie (3)*. De details van de volledige lagenstruktuur zijn te vinden in Tabel 3.8.*

3.3.4 Conclusie

Het is louter op basis van de ontwerpsimulaties niet vast te stellen dat een AlGaInAs SOA/isolator zich veel beter zal gedragen dan een InGaAsP SOA/isolator, wegens de vele onzekerheden in het ontwerp van dit laatste materiaalsysteem. De experimentele karakterisatie van InGaAsP MQW's die gegeven zal worden in het volgende hoofdstuk, heeft echter wel a posteriori aangetoond dat AlGaInAs materiaal een veel beter differentieel winstgedrag vertoont en lagere QW transparantiestroomdichtheden heeft. Het is dus veilig te stellen dat de isolator zich veruit beter zal gedragen met een AlGaInAs actieve laag. De waarden getabuleerd in Tabel 2 (en de waarden voor de totale lengte en totale absolute stroom voorgesteld in Fig. 3.28) zijn dus een indicatie van wat bij benadering de ultieme performantie is van de bestudeerde component (in het achterhoofd houdend dat het om geïdealiseerde 1D slabsimulaties gaat).
De enthousiaste voorspellingen van Nakano moeten dus blijkbaar ietwat bijgesteld worden. Zonder een verdere verbetering in het winstgedrag van AlGaInAs QW's onder trekspanning en een verbetering van de magneto-optische sterkte van CoFe (of een ander MO materiaal) zal dit isolatorontwerp niet beter doen dan 25dB in een lengte van 5mm en een totale stroom van 500mA. Dit moet wel ietwat bijgesteld worden aangezien later experimenteel is vastgesteld dat de equi-atomische samenstelling $Co_{50}Fe_{50}$ bij benadering 60% sterker is wat gyrotropie betreft, en dat er nog heel wat ruimte is voor optimalizatie van de actieve AlGaInAs MQW laag (niet enkel wat aantal QW's betreft maar ook op gebied van procentuele trekvervorming) (zie verder). Een voorzichtige herschatting op basis hiervan zou dan een 25dB SOA/isolator impliceren met een lengte van 2–3mm en een totale stroom van 200mA !

4 Experimentele karakterisatie van de bouwstenen van de SOA/isolator

In dit hoofdstuk wordt het gedrag van de subcomponenten van de SOA/isolator één voor één experimenteel gekarakteriseerd. Hierbij wordt extensief gebruik gemaakt van de karakterizatiemethodes die gedetailleerd beschreven zijn in het vorige hoofdstuk.

4.1 Karakterizatie van ferromagnetische CoFe metaallegeringen

Zoals vroeger aangeduid is er om praktische redenen gekozen te werken met CoFe legeringen. Evengoed echter zijn er theoretische indicaties dat dit materiaalsysteem een van de ideale kandidaten is wanneer het aankomt op sterke gyrotropie. Twee legeringsamenstellingen zijn onderzocht: $Co_{90}Fe_{10}$ (dat een kristallijne hcp structuur vertoont) en het equi-atomische $Co_{50}Fe_{50}$ (met een bcc kristalstructuur). De metallische films zijn gedeponeerd door sputterdepositie. Om een planaire uniaxiale magnetische anisotropie te induceren werd indien nodig een magneetveld aangelegd tijdens sputtering. In de meeste karakterisatie-experimenten zijn de films gedeponeerd op hooggedopeerde InGaAs(P) zoals uiteindelijk ook het geval zal zijn in de finale devicestructuur, m.a.w. een uitgesproken (1 0 0)-textuur. In geen enkel geval werd gebruikt gemaakt van een "precursor" laag met een "betere" textuur om de magnetische anisotropie te verbeteren. De films zijn polykristallijn met een korrelgrootte van de grootte-orde van 10nm. Er werden geen problemen ondervonden bij het definiëren van lange smalle rechthoekige metallische CoFe contacten door middel van standaard lift-off technieken.

4.1.1 Optische en magneto-optische constanten

De optische en magneto-optische constanten van beide CoFe legeringen werden gemeten bij 1300nm met behulp van beide fotometrische technieken beschreven in sectie 3.1.1.

Dynamische fotometrische karakterizatie De dynamische fotometrische methode die gedetailleerd beschreven is in sectie 3.1.1.1, werd geïmplementeerd in een geavanceerde experimentele opstelling (zie Fig. 4.1) waarbij het magneto-optisch effect in longitudinale configuratie wordt aangewend. De magneto-optische ellipsometrische verhoudingen r_{sp}/r_{pp} en r_{ps}/r_{ss} werden gemeten volgens de gecalibreerde methode beschreven in het vorige hoofdstuk. Voor de isotrope ellipsometrische verhouding r_{pp}/r_{ss} werd echter een andere, meer klassieke techniek aangewend, namelijk de rechtstreekse absolute meting van de s- en p-reflectiviteit als functie van de invalshoek. Dit heeft het voordeel dat alle optische elementen in de opstelling weggelaten kunnen worden op de polarisator na. Het nadeel is natuurlijk dat het een absolute methode is, en er dus een zeer nauwkeurige

alloy	n	κ	$\Re(g)$	$\Im(g)$	FoM = $\frac{\Im(g)}{\kappa}$
$Co_{90}Fe_{10}$	4.4	4.77	-1.75	1.0	0.21
$Co_{50}Fe_{50}$	3.15	4.55	-1.72	-1.65	0.36

Tabel 3: Gemeten optische en magneto-optische CoFe constanten (bij 1300nm). De FoM is gedefinieerd als de verhouding van het imaginair deel van de gyrotropie (∼isolatie) tot de extinctiecoëfficiënt(∼nodige stroom).

vermogencalibratie nodig is[7]. Er wordt in de engelstalige tekst uitgelegd dat de meting van de reflectiviteit slechts tweede orde bijdragen heeft van de gyrotropie. Er is vastgesteld dat indien het metaal niet beschermd wordt door een gecontroleerd gedeponeerde passivatielaag, de natuurlijke oxidatie van het CoFe belet de complexe brekingsindex voldoende nauwkeurig te bepalen. Het belang van een heel nauwkeurige bepaling van de isotrope brekingsindex wordt onderlijnd door de formules die de magneto-optische elipsometrische verhoudingen verbinden met de gyrtropieconstante (zie Vgl. 4.1). Een te grote onzekerheid op de isotrope constanten in combinatie met de normale experimentele onzekerheid op de ellipsometrische MO verhouding zal door foutenpropagatie een nog grotere onzekerheid opleveren op de gyrotropie constante.
Beide legeringen werden voorzien van een Al/Al_2O_3 passivatielaag die afzonderlijk gekarakteriseerd werd. Dit liet een voldoende nauwkeurige extractie van de isotrope brekingsindex van de CoFe legeringen toe. In combinatie met een gecalibreerde meting van de MO Kerr rotatie en ellipticiteit werd, met behulp van een magneto-optisch multilagen-reflectiealgoritme, de gyrotropieconstante van beide legeringen bepaald. Tabel 3 vat de extractieresultaten samen. Een mogelijke FoM vergelijkt beide legeringen.
Deze resultaten lijken de bevindingen van Busschow [21] te bevestigen dat de equiatomische legeringssamenstelling beter doet dan Co-rijke legeringen. Aangezien de storingsformule leert dat niet-reciproke absorptie bepaald wordt door het imaginair deel van de gyrotropie constante (die 70% sterker is voor de equi-atomische samenstelling), en aangezien de optische absorptie quasi-constant blijft, voorspellen de simulaties van vorig hoofdstuk ruwweg[8] een daling van 40% van de lengte van de isolator en van de totale nodige stroom bij het gebruik van $Co_{50}Fe_{50}$.

Statische fotometrische karakterizatie De statische fotometrische methode van Berger is in principe niet bruikbaar omdat ze veronderstelt dat er slechts één enkele reflectend grensvlak aanwezig is. En vorige sectie toonde aan dat de natuurlijke oxidatie van CoFe niet te verwaarlozen is. Pogingen om met een eenvoudige uitbreiding van een bestaande longitudinale MOKE setup in IMEC (normaal gebruikt voor magnetische hysteresismetingen) toch tot een ruwe schatting van de

[7] Meer bepaald moet er aandacht geschonken worden aan een goede alignering van de optische bundel over het hele bereik van invalshoeken.

[8] Dit is niet volledig correct omdat het reëel deel van de isotrope index ook meespeelt in de opsluiting ter hoogte van het metaal.

gyrotropie van CoFe te komen op een snelle en goedkope manier, faalden. Het bleek niet mogelijk het model van Berger te fitten aan de gemeten intensiteitscurves. Nochtans kan volgens Atkinson in theorie een MO multilaag (metaal+oxide) voorgesteld worden door een equivalente enkele MO laag. Om na te gaan of er dus iets fundamentelers fout was met de gebouwde statische meetopstelling, werden gelijkaardige metingen uitgevoerd op dikke Ni lagen. Dit in de wetenschap dat de optische en magneto-optische constanten van Ni bekend zijn in de literatuur (doorheen het volledige spectrum van zichtbaar tot UV), en dat het geweten is dat de natuurlijke oxidatie van Ni zelf-limiterend is tot een dikte van slechts 2nm. Deze experimenten (zowel uitgevoerd met een HeNe-laserbron als met een 1300nm-bron) leverden vreemd genoeg een aanvaardbare extractie op van de isotrope indices van Ni, maar een flagrante overschatting van de MO constante. Er wordt in de engelstalige tekst uitgelegd waarom het onze overtuiging is dat dit een probleem is van een gebrekkige calibratie van de meetopstelling en hoe dit in de toekomst kan verholpen worden.

4.1.2 Magnetische anisotropie van CoFe op InP-gebaseerde halfgeleiderlagen

De karakterisatie van de magnetische anisotropie is gebeurd op basis van het vergelijken van de hysteresiscurves volgens verschillende spatiale richtingen in de film van het ferromagnetisch metaal. Zoals hierboven uitgelegd is getracht planaire uniaxiale anisotropie te induceren in de CoFe films door deze te sputteren in aanwezigheid van een magneetveld. Om na te gaan in hoeverre dit tot het gewenste magnetisch anisotroop gedrag leidt, werden de hysteresismetingen uitgevoerd enerzijds op niet-gestructureerde films van variërende dikte en anderzijds op gestructureerde rechthoekige CoFe contacten van variërende dikte en breedte. Deze laatste metingen laten toe na te gaan of de geïnduceerde magnetische anisotropie de vormanisotropie kan tegenwerken. Er werden enkel metingen uitgevoerd op $Co_{90}Fe_{10}$-films omdat het $Co_{50}Fe_{50}$-target niet beschikbaar was ten tijde van deze experimenten. De magnetische anisotropie van de equi-atomische legering kan a posteriori gecontroleerd worden via TM-ASE hysteresismetingen (zie volgend hoofdstuk). Er kan wel degelijk een verschil zijn tussen beide legeringen wat anisotropie betreft gelet op hun fundamenteel verschil in kristalstructuur.
Fig. 4.9 toont de via longitudinale MOKE gemeten hysteresiscurves in verschillende planaire richtingen en voor verschillende diktes van "blanket" CoFe-films. De resultaten zijn ietwat verrassend. Slechts voor heel dunne CoFe-films (tot max. 10nm) wordt de gewenste planaire uniaxiale magnetische anisotropie waargenomen. Vanaf 20nm verdwijnt blijkbaar het effect van het magnetisch bias veld tijdens sputtering. Er is geen afdoende verklaring voor dit fenomeen. Een mogelijke verklaring zou epitaxie van CoFe kunnen zijn, maar dat is niet onderzocht. In elk geval bewijst de mogelijkheid om een voldoende sterk Kerr signaal te bekomen bij films zo dun als 5nm dat een eventuele "dode magnetische laag" aan het grensvlak tussen CoFe en het ternaire InGaAs, zeker dunner is dan 5nm. Latere absolute magnetizatiemetingen via AGFM hebben aangetoond dat binnen de nauwkeurigheid van de AGFM methode deze dode laag zelfs verwaarloosbaar is.

De toenemende planaire magnetische isotropie als functie van de dikte van de blanket CoFe-films doet vermoeden dat de vormanisotropie in lange rechthoekige CoFe-contacten niet zal kunnen tegengewerkt worden bij diktes boven 20nm. Fig. 4.11 toont via AGFM gemeten planaire hysteresiscurves in de richting parallel aan de lengte van het contact en in de richting loodrecht op het contact voor drie verschillende breedtes van de contacten en voor een dikte van 20nm van de films. Dit is de dikte waarbij voor de blanket films de magnetische planaire anisotropie verdwijnt. De resultaten zijn ietwat vreemd in het licht van de waargenomen anisotropie op de blanket films. Er wordt weliswaar zoals verwacht waargenomen dat met afnemende breedte de schuiftransformatie als gevolg van de vormanisotropie de remanentie in de richting loodrecht op de lengte van het contact doet afnemen, terwijl in de richting parallel aan het contact de remanentie stijgt. Maar vreemd genoeg wordt bij heel brede contacten ($\geq 50\mu$m) de gewenste uniaxiale planaire anisotropie (voor 20nm dikke films) gedeeltelijk herwonnen, in tegenstelling tot wat is waargenomen voor de blanket films. Dit wordt bevestigd door de gemeten waarden voor de "squareness" S^* (zie Tabel 4.4). Er is geen verklaring voor dit gedrag. Bovendien is het ook niet duidelijk waarom de waarden voor het coërcieve veld bij de rechthoekige structuren slechts 25Oe bedraagt in vergelijking met de 50Oe van de blanket films. In elk geval kan geconcludeerd worden dat in het uiteindelijke device er nauwlettend moet toegezien worden op de bescherming tegen magnetische strooivelden en dat er moet gezorgd worden voor een zo breed als technologisch mogelijk contact.

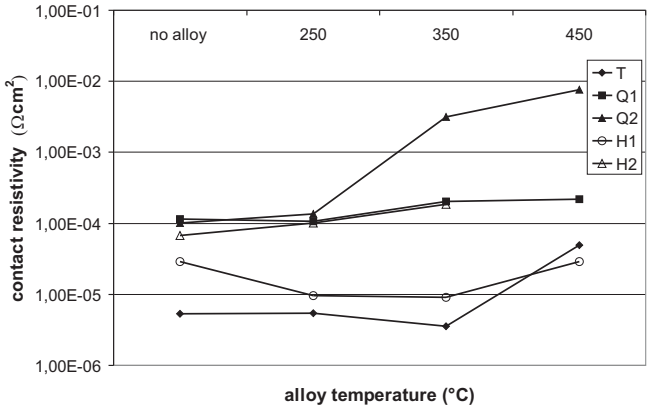
4.1.3 P-type Ohmse CoFe-contacten

Onderzochte contactlagen Er is uitgelegd in sectie 3.1.3.1 dat de realizatie van een goed Ohms contact met behulp van ferromagnetische metalen, meer een kwestie is van te werken met een halfgeleidercontactlaag met een zo laag mogelijke band gap en van zo hoog mogelijke doperingsconcentraties te gebruiken in die laag dan van iets te verwachten van de werkfunctie van het metaal (omwile van Bardeen's 1/3 regel). Bovendien werd er opgemerkt dat voor de specifieke bestudeerde toepassing het gebruik van het standaard ternaire InGaAs in principe niet mogelijk is, wegens de hoge optische absorptie voor 1300nm licht. Daarom is een doorgedreven studie uitgevoerd naar mogelijke alternatieve contactlagen. Enerzijds is er nagegaan of hoge band gap (transparante) quaternaire InGaAsP ($\lambda_g = 1.17\mu$m) lagen kunnen gebruikt worden[9]. En anderzijds is er onderzocht of een hybride combinatie van een dikke transparante quaternaire laag met een zo dun mogelijke aborberende lage band gap ternaire laag kan gebruikt worden. Deze laatste combineert dan het beste van beide werelden. Tabel 4.5 vat de details van de verschillende teststructuren samen. Naast het 100nm InGaAs standaardcontact als referentie, zijn er twee quaternaire structuren gegroeid met verschillende dikte, om na te gaan hoe dun de contactlaag kan gehouden worden, en twee hybride contactstructuren met een verschillende dikte voor de ternaire capping om na te gaan hoe dun deze absorberende lage gap capping kan worden zonder te veel van

[9]Hierbij moet rekening gehouden dat er een belangrijke optimalizatie nodig is van de groei omdat de oplosbaarheid van p-dopanten afneemt met de band gap van de halfgeleider.

zijn goede Ohms karakter op te offeren.

Teststructuren en resultaten Het masker met de teststructuren bevat zowel CKR als TLTR structuren. De laatste werden enkel gebruikt om de vierkantsweerstand van de halfgeleiderdiffusielaag te bepalen, omdat, zoals uitgelegd in de engelstalige tekst in sectie 3.1.3.3, de transmissielijn structuur te veel lijdt onder afrondingsfouten. De CKR structuren (waarvan er vier verschillende geometrieën op het masker staan) werden gebruikt in combinatie met de ontwikkelde Matlab tool om de contactresisitiviteit van de vijf verschillende contactlagen te bepalen volgens de procedure geschetst in 3.1.3.3. Hierbij moet de nodige aandacht geschonken worden aan een visuele inspectie van de geprocessede contactstructuren omdat kleine afwijkingen op de geometrische parameters tot schijnbaar geometrie-afhankelijke en dus foutieve waarden voor ρ_c kunnen leiden. Een voorbeeld daarvan is gegeven in de engelstalige tekst.
Elke lagenstructuur werd bovendien onderworpen aan een snelle thermische behandeling (RTP) op drie verschillende legeringstemperaturen om na te gaan of de resistiviteit kan verlaagd worden door de thermische vorming van lage band gap legeringen aan het metaal/halfgeleidergrensvlak. Fig. 4 vat de resultaten van deze experimenten samen.



Figuur 4: CKR geëxtraheerde specifieke contactresistiviteit van 50nm dikke sputtergedeponeerde $Co_{90}Fe_{10}$ op de vijf verschillende InP-gebaseerde halfgeleiderlagen van Tabel 4.5 in functie van de RTP temperatuur.

Uit deze figuur blijkt duidelijk dat het referentie ternaire contact moeilijk te verslaan is, maar dat vreemd genoeg het volstaat een heel dunne (15nm) absorberende lage band gap ternaire capping toe te voegen aan een hoog-gedopeerde quaternaire laag om een resistiviteit te bekomen die het gedrag van de ternaire structuur benadert. De gemeten contactresistiviteiten zijn niet van topkwaliteit maar een contactresistiviteit van $10^{-5}\Omega\text{cm}^2$ is meer dan voldoende laag voor de hier beschouwde

toepassing (waarbij de contactoppervlakte snel in de orde van $5\times10^{-5}\text{cm}^2$ zal zijn). Dus ook al zal een puur quaternair transparant contact nog steeds "slechts" tot een serieweerstand van de orde 1–2Ω leiden, toch is het verkieslijk een hybride contact te gebruiken. De sterke daling in contactresisitiviteit overcompenseert de geringe toename in optische absorptie.

4.2 Karakterizatie van TM-selectieve versterkende MQW's onder trekspanning

Zoals aangeduid is het quasi-onmogelijk de actieve roostermisaangepaste MQW laag theoretisch te ontwerpen. In plaats van daarvan werd de studie van de invloed van trekvervorming in de QW's op de winstkarakteristieken pragmatisch uitgevoerd door een hele reeks MQW's te groeien met variërende parameters zoals procentuele vervorming, aantal wells, compensatie in de bufferlagen,... – en deze te vergelijken via winstmetingen op geprocessede laserstructuren. Het is eveneens aangeduid dat theoretisch verwacht wordt dat het InAlGaAs systeem zich beter zal gedragen dan het InGaAsP materiaalsysteem, voornamelijk wegens de betere conductiebandoffset in dit systeem. Voor beide systemen werd een doorgedreven optische winststudie doorgevoerd. De epitaxiaalgroei van beide materiaalsystemen is uitgevoerd door collega-onderzoekers in het researchcentrum van Alcatel Opto+. We geven hier een kort overzicht van de voornaamste resultaten in de evolutie van het ontwikkelde materiaal. Voor een gedetailleerdere uiteenzetting verwijzen we uiteraard naar de engelstalige tekst.

4.2.1 1300nm InGaAsP MQW onder trekspaning

TM lasing Eerst en vooral werd nagegaan vanaf welk vervormingsniveau TM-preferentiële versterking optreedt voor 1300nm wells. Hierbij werd nog niet te veel aandacht geschonken aan optimale groeicondities voor de wells en de bufferlagen. Dat er een minimale vervorming nodig is voor TM-versterking is een logisch gevolg van het feit dat het quantizatie-effect de shift van de lichte gaten band tegenwerkt. Voor meer details verwijzen we naar de engelstalige tekst.
Er werd vastgesteld dat de LH-HH crossing optreedt ergens tussen -0.33% en -1.0%, en vermoedelijk in de buurt van -0.7%. Ondanks de niet-geoptimaliseerde groei, werd er voor de -1% structuur reeds een radiatieve transparantiestroomdichtheid per well waargenomen van ongeveer 150A/cm^2 en een differentiële winst van 50cm/kA. Enkel de interne quantumefficiëntie was een eerder ontgoochelende 50%. Er wordt een belangrijke elektronlek vermoed door de intrinsiek lage CBO van InGaAsP QW onder rek (typisch $\frac{\Delta E_c}{\Delta E_v} \approx 0.2 : 0.8$).
De gemeten karakteristieken zijn geen wereldrecordwaarden, maar zijn anderzijds al voldoende in de buurt van de waarden voor de meer gebruikte en beter ontwikkelde MQW's onder druk. Er is op gewezen in het vorig hoofdstuk dat in theorie trekvervorming in alle opzichten nog meer dan drukvervorming de voordelen van roostermisaanpassing uitbuit. Maar om een historische reden is er nooit veel aandacht geschonken aan TM-versterking.

Materiaaloptimalizatie In een tweede run werd enerzijds de groei van de -1% InGaAsP wells verder geoptimaliseerd en anderzijds nagegaan of de winstkarakteristieken intrinsiek konden verbeterd worden door de spanning in de wells (en de spanningscompensatie in de buffers) verder op te drijven.
Met het geoptimaliseerde -1% materiaal (details zie Tabel 4.9) is er nagegaan tot hoeveel wells er gegroeid konden worden zonder aan intrinsieke winstperformantie in te boeten door materiaaldefecten. Structuren met tot 15 wells zijn gegroeid. Enkel voor dit laatste geval zijn er structurele defecten waargenomen. De voornaamste doelstelling hier is uiteraard een zo hoge mogelijke optische opsluiting in de versterkende lagen na te streven. Er werd eerst en vooral vastgesteld dat de optimalizatie van de epitaxiaalgroei van het materiaal[10] ertoe geleid heeft dat de differentiële efficiëntie met een factor 2 toegenomen is, terwijl de radiatieve transparantiestroomdichtheid quasi gehalveerd is! Vreemd genoeg is er weinig verbetering merkbaar qua interne quantumefficiëntie (van iets meer dan 50% naar 60%) en bovendien wordt zelfs waargenomen dat deze interne efficiëntie vanaf 9 wells snel begint te dalen.
De verdere optimalizatiepogingen door de well vervorming en/of de compensatiespanning in de buffers op te drijven (zie Tabel 4.10) hebben niet geleid tot drastische verbeteringen in de winstkarakteristieken. De interne quantum efficiëntie lijkt niet boven de 60% uit te komen en de differentiële winst en radiatieve transparantiestroomdichtheid lijken zelfs optimaal voor -1.1% vervorming (resp. ≈100cm/kA en ≈50A/cm^2 per well). Dit lijkt het theoretisch vermoeden te bevestigen dat het InGaAsP materiaalsysteem te veel gehinderd wordt door de beperkte CBO offset en dat de intrinsieke sterkte van de roostermisaanpassing iets te zwak is (meer bepaald de axiale deformatiepotentiaal).

4.2.2 1300nm InAlGaAs MQW onder trekspaning

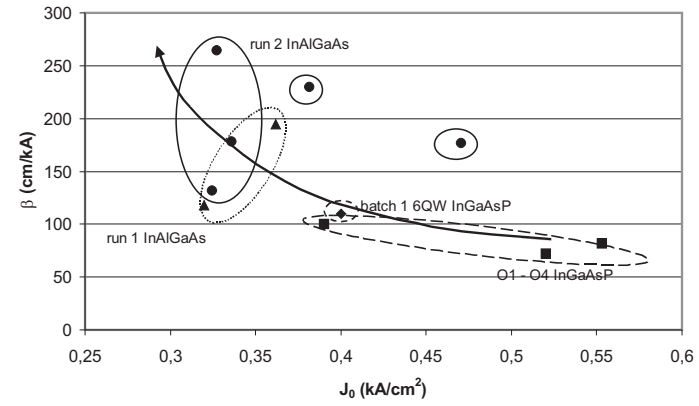
De overstap naar het InAlGaAs systeem gebeurde nadat vastgesteld was dat het winstgedrag van de InGaAsP wells niet verbeterd kon worden. Ook al zijn de waargenomen winstkarakteristieken van het beste InGaAsP materiaal zeker niet teleurstellend te noemen, toch tonen de winstcurves van Fig. 4.20 dat bij de benodigde modale winsten (die in de orde van 150–200cm^{-1} zijn) de QW winst reeds begint te satureren, wat gepaard gaat met allerlei vervelende band-filling effecten. Van het InAlGaAs systeem wordt een veel hogere differentiële winst en een veel lagere QW transparantiestroomdichtheid verwacht, omdat de VBO offset veel lager is en de axiale deformatiepotentiaal sterker is. Het eerste effect laat toe 1300nm QW's onder spanning te ontwerpen die geen HH band geleiden zodat niet alleen er quasi geen band-mixing optreedt maar ook de ladingsdragerdichtheid bij transparantie veel lager is omdat de zware gatenband niet moet opgevuld worden. Het tweede effect versterkt de LH-HH splitting en helpt op die manier niet alleen het type II gedrag voor de zware gaten te induceren maar verhoogt bovendien ook de C-LH TM transitiesterkte (door een sterkere menging met de SO-band [22]).

[10] en van de layout van het actieve materiaal door de band gap van de buffer iets op te drijven (betere electron confinement). De dikte is lichtjes aangepast na vaststelling dat de golflengte iets off-target was en er is vastgesteld dat de werkelijke vervorming eerder -1.1% bedroeg.

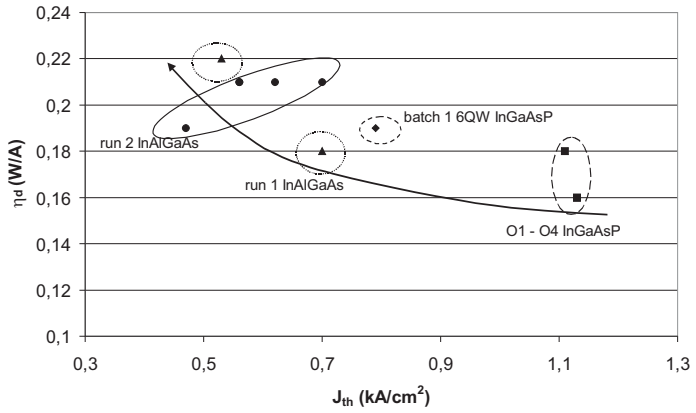
InAlGaAs MQW – type II voor HH1 De eerste gegroeide InAlGaAs MQW's beoogden een verbetering van de winstkarakteristieken t.o.v. InGaAsP door type II gedrag te induceren voor de HH1 band. Hiervoor werden drie types 6QW-structuren gegroeid met -1.25% vervorming in de wells en +0.6% compensatie in de barriers ($E_g = 1.12$eV). Met de typische band offset verhouding $\frac{\Delta E_c}{\Delta E_v} = 0.72 : 0.28$ voor InAlGaAs en de waarde van 2eV voor de axiale deformatiepotentiaal b, is het direkt duidelijk dat de QW steeds type II zal zijn voor de zware gaten. De drie structuren verschilden in geleidingslagen rond het MQW gebied. Gelet op de asymmetrische band offsets is het beter aan de p-zijde een Al-gebaseerde geleidingslaag te gebruiken en aan de n-zijde een P-gebaseerde laag. Er zijn drie combinaties geprobeerd: symmetrische Al/Al en P/P geleidingslagen en het mogelijk elektrisch betere Al/P. De resultaten zijn echter voor alle drie quasi even goed. Vergelijking van de differentiële efficiëntie en de drempelstroomdichtheid van BA FP lasers met dezelfde lengte en aantal QW's als voor de beste InGaAsP spreken boekdelen. De drempelstroomdichtheid is gehalveerd en de externe efficiëntie stijgt met niet minder dan 40%. De winstcurve afgeleid uit deze lasermetingen voorspelt een radiatieve transparantiestroomdichtheid van hooguit 40A/cm^2 per well (bij een interne quantum efficiëntie van 70%) en een differentiële winst van meer dan 220cm/kA. Er moet echter omzichtig omgesprongen worden met deze winstparameters. In de engelstalige tekst wordt duidelijk uitgelegd waar de risico's liggen bij extrapolatie naar hoge stroomdichtheden.

Variatie van de trekvervorming Recent werd door de onderzoekers van Alcatel een doorgedreven studie uitgevoerd naar mogelijke verdere verbeteringen van het winstgedrag van het Al-systeem door de spanning in de QW verder op te drijven. De resultaten zijn op zijn zachtst gezegd indrukwekkend te noemen. In Fig. 4.23 worden de winstparameters (β en J_{th} voor 600μm 6QW BA FP lasers) geplot als functie van de procentuele vervorming in de wells. De differentiële winst blijft toenemen met vervorming (tot 250cm/kA !) en materiaalinspectie laat vermoeden dat de kritische vervorming nog steeds niet bereikt is.

Fig. 5 vat de evolutie van het ontwikkelde actieve materiaal samen. De trend van toenemende differentiële winst en afnemende transparantiestroomdichtheid enerzijds, en toenemende externe differentiële effciëntie en afnemende drempelstroomdichtheid voor 600μm lange 6QW BA FP lasers is duidelijk.



(a)



(b)

Figuur 5: Vergelijking van de karakterisieken van alle ontwikkelde 6QW winstmaterialen (voor de labels van de verschillende types verwijzen we naar de engelstalige tekst). (a): Differentiële winst versus transparantiestroomdichtheid; (b): Externe efficiëntie versus drempelstroomdichtheid voor 600µm lange BA FP lasers.

4.3 Conclusies

De experimentele resultaten voor het gedrag van de subcomponenten gepresenteerd in dit hoofdstuk laten vermoeden dat in principe alle ingrediënten aanwezig zijn om het innovatieve isolatorconcept succesvol te realizeren. De geëxtraheerde optische en magneto-optische parameters van de twee CoFe legeringen in combinatie met de ontwerpsimulaties bevestigen de voorspellingen van een relatief compacte 25dB isolator. Er lijkt geen reden te zijn waarom het CoFe niet tegelijkertijd als

elektrisch contact zou gebruikt kunnen worden. De contactresistiviteit is meer dan aanvaardbaar en optische afscherming kan grotendeels vermeden worden door een hybride contactlaag te gebruiken. Enkel de waargenomen anisotropie baart wat zorgen. Er zal verderop blijken dat de nood aan een redelijk brede zo vlak mogelijke CoFe film een niet te onderschatten technologische uitdaging vormt.
Zo mogelijk nog beter nieuws wordt geleverd door de geobserveerde winstkarakteristieken van InAlGaAs MQW lagen onder trekvervorming. De actieve TM-selectieve winstlaag riskeerde immers de grote struikelblok van het device te zijn. De winstmetingen op het Al-gebaseerde systeem nemen deze zorgen grotendeels weg. Heel hoge differentiële winst en heel lage transparantiestroomdichtheid is waargenomen in -1.25% 1300nm InAlGaAs QW's. Modale winsten tot 200cm^{-1} voor stroomdichtheden lager dan 5kA/cm^2 zouden hierbij geen probleem mogen vormen. Bovendien lijkt alles erop te duiden dat de limieten van de winstmogelijkheden van InAlGaAs nog niet bereikt zijn.

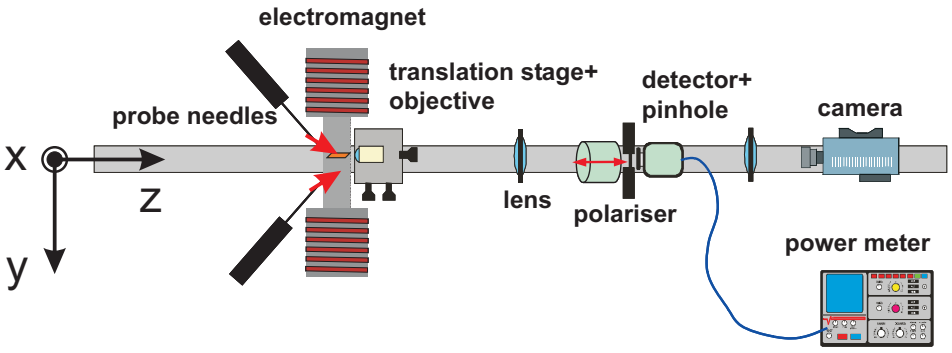
5 Experimentele karakterisatie van de SOA/isolator

Dit laatste hoofdstuk is de culminatie van al het theoretische en experimentele werk beschreven in de vorige hoofdstukken. Het eerste wereldwijde kwalitatieve en kwantitatieve bewijs van het geïntegreerde isolatorconcept van Nakano wordt gegeven.

5.1 Karakterizatiemethodes voor niet-reciproke transversale MO Kerr absorptie in een SOA

5.1.1 Experimentele opstelling – Kwalitatieve karakterisatie

Fig. 6 toont de basisversie van de experimentele opstelling voor de karakterisatie van de niet-reciproke optische absorptie in een actieve structuur. Bij detectie van de versterkte spontane emissie (ASE) aan één enkel "vast" facet zal het gedecteerde licht de eigenschappen hebben van een voorwaartse resp. een achterwaartse propagerende mode afhankelijk van de zin van het aangelegde magneetveld. Dit is intuïtief duidelijk maar kan evengoed rigoureus bewezen worden op basis van het Onsager principe. Voor details van de uitvoering van de opstelling verwijzen we naar de engelstalige tekst.



Figuur 6: Basisversie van de experimentele opstelling voor de karakterizatie van de niet-reciproke optische absorptie in een SOA/isolator. Het laboratorium frame wordt gebruikt om voorwaarts en achterwaarts licht te relateren aan de richting van de magnetizatie. Bij conventie wordt het gededecteerde licht (volgens de positieve $z-as$) "voorwaarts" genoemd als de magnetizatie in de positieve $y-richting$ wijst.

Deze opstelling laat niet enkel een relatieve meting toe van de optische niet-reciprociteit (onafhankelijk van precieze absolute experimentele omstandigheden zoals alignatie) maar ook een direkte kwalitatieve beoordeling van het effect. Immers, door het aangelegde magneetveld te variëren tussen de saturatiewaarden van het CoFe contact zal de magnetizatie de karakteristieke hysteresis van het

ferromagnetisch contact doorlopen. Detectie van de TM-component van de versterkte spontane emissie zal dus leiden tot een emulatie van de hysteresiscurve. Aan de andere kant kan de consistentie van de meting nagegaan worden door onder variërend magneetveld te controleren of de TE-component onbeïnvloed blijft. Deze kwalitatieve metingen waren het eerste succesvolle bewijs van de niet-reciproke werking van de component. Het is bovendien duidelijk dat de "amplitude" van deze TM ASE hysteresis op een of andere manier gelinkt is aan de grootte van het niet-reciprook effect. Om een kwantitatieve meting te kunnen doorvoeren moet dus de verhouding van de saturatieniveaus van de TM ASE hysteresiscurves gerelateerd worden aan een model voor de geëmitteerde TM ASE waarin o.a. de niet-reciproke absorptie als fitting parameter optreedt. De onafhankelijke variabelen zijn uiteraard de geïnjecteerde stroom en de lengte van de SOA.

5.1.2 Methodes gebaseerd op versterkte spontane emissie

Twee dergelijke ASE fitting modellen zijn ontwikkeld afhankelijk van het feit of de SOA caviteit al dan niet resonante reflecties toelaat. In sectie 5.1.2 worden de details van de afleiding van beide modellen gegeven. We volstaan hier met aan te duiden dat de specifieke werking van deze component toelaat gebruik te maken van een heel eenvoudig model om de propagatie van de ASE te beschrijven. Aangezien de SOA/isolator in essentie bedreven wordt in de buurt van transparantie (of zelfs ver eronder als er niet voldoende winst kan geleverd worden) en er zich dus geen grote fotondichtheden opbouwen, mag er aangenomen worden dat er geen longitudinale winstsaturatie optreedt noch enige kruiscorrelatie tussen de spontane emissie geleid in verschillende modi. Het volstaat dus de propagerende ASE te beschrijven door een stel incoherent gesommeerde, exponentieel groeiende/dalende modi. Noteer hierbij dat het in principe niet belangrijk is te weten hoeveel ASE er lokaal in een mode gegenereerd wordt. Deze zogenaamde spontane emissiefactor wordt immers toch weggedeeld bij het opstellen van de uitdrukking voor de verhouding van de ASE saturatieniveaus. Bemerk wel dat dit enkel rigoureus geldt als er geen longitudinale winstsaturatie optreedt.
Afhankelijk van het feit of er al dan niet resonante reflecties optreden in de caviteit wordt de verhouding van de geëmitteerde ASE (bij het schakelen van de contactmagnetizatie tussen zijn twee saturatietoestanden) dan gegeven door

$$\boxed{S(J,L|A,B,\Delta\alpha) = \frac{A\ln(J) - B - \frac{\Delta\alpha}{2}}{A\ln(J) - B + \frac{\Delta\alpha}{2}} \frac{1 - e^{(A\ln(J)-B+\frac{\Delta\alpha}{2})L}}{1 - e^{(A\ln(J)-B-\frac{\Delta\alpha}{2})L}}} \tag{4}$$

of door

$$\boxed{\begin{aligned} S(J,L|A,B,\Delta\alpha) = {} & \frac{e^{(A\ln(J)-B+\Delta\alpha/2)L}}{e^{(A\ln(J)-B-\Delta\alpha/2)L}} \cdots \\ & \left(\frac{\frac{1-e^{-(A\ln(J)-B+\Delta\alpha/2)L}}{(A\ln(J)-B+\Delta\alpha/2)} - |r|^2 \frac{1-e^{+(A\ln(J)-B-\Delta\alpha/2)L}}{(A\ln(J)-B-\Delta\alpha/2)}}{\frac{1-e^{-(A\ln(J)-B-\Delta\alpha/2)L}}{(A\ln(J)-B-\Delta\alpha/2)} - |r|^2 \frac{1-e^{+(A\ln(J)-B+\Delta\alpha/2)L}}{(A\ln(J)-B+\Delta\alpha/2)}} \right) \end{aligned}} \tag{5}$$

Hierin zijn A en B de fitting parameters die het 0^{de} orde modale winst/verliesgedrag van de SOA beschrijven, via een standaard logaritmische afhankelijkheid. B verzamelt hierbij alle "isotrope verliesmechanismes" terwijl $\Delta\alpha$ uiteraard de niet-reciproke absorptieshift is. In de uitdrukking voor een resonante caviteit wordt de facetreflectiviteit behandeld als een gekende parameter, omdat ervaring leert dat het behandelen van deze parameter als een extra fitting parameter tot convergentieproblemen leidt. De meetprocedure bestaat er dus in de verhouding S te meten voor variërende geïnjecteerde stroomdichtheden en eventueel een reeks devices met verschillende lengte[11] en de vergelijking gelijktijdig aan alle meetpunten te fitten. In de engelstalige tekst wordt verder uitgelegd hoe dit monomodaal model, indien foutief toegepast op een multimodale golfgeleider tot een onderschatting zal leiden van het niet-reciprook effect. Tenslotte wordt er ook een alternatieve methode uiteengezet, gebaseerd op Hakki-Paoli metingen van de 0^{de} modale winst/verlies. Op die manier wordt niet alleen het aantal fitting parameters gereduceerd tot enkel de niet-reciproke absorptieshift, maar worden ook eventuele convergentieproblemen ten gevolge van het te eenvoudige winst/verliesmodel ($A\ln(J) - B$) vermeden, aangezien deze laatste in dit geval rechtstreeks gemeten wordt. Het enige probleem met deze HP-gebaseerde meetmethode is dat ze heel tijdrovend is, in het bijzonder als de lichtbundel heel zwak is (wat meestal het geval is, aangezien het merendeel van de meetpunten zich ver onder transparantie bevindt).

5.1.3 Niet-reciproke lasing – karakterisatie van de SOA/isolator performantie in een communicatiesysteem

De karakterisatie wordt heel wat eenvoudiger als de winst in de caviteit voldoende groot is om de roundtrip verliezen te compenseren. In dat geval zal er niet-reciproke lasing optreden, waarbij de externe differentiële efficiëntie afhankelijk wordt van de zin van het magneetveld in het ferromagnetisch magneto-optisch contact. De verhouding van voorwaarts tot achterwaarts lasing vermogen kan heel snel afgeleid worden als

$$\boxed{\Delta\alpha = \frac{2}{L}\ln\left(\frac{P_f}{P_b}\right)} \tag{6}$$

Er dient hier wel bij opgemerkt te worden dat deze uitdrukking natuurlijk enkel "exact" geldig is op de longitudinale mode die lasert omdat enkel voor die mode vergelijking (5.14) geldt. In principe vereist dit dus het gebruik van spectraal geresolveerde detectie, maar meestal volstaat het de voorwaarts-achterwaarts verhouding te evalueren voldoende ver boven drempel.

Bovenstaande methodes zijn voldoende nauwkeurig voor de experimentele karakterisatie van de niet-reciproke absorptieshift en dus het isolatieniveau van het device. Echter, voor dit specifieke geïntegreerde isolatorconcept zijn er een aantal karakteristieken die extra aandacht verdienen. Immers, door de intrinsiek actieve uitvoering van een puur passieve optische functie, bestaat het risico dat

[11] maar dezelfde breedte aangezien de modale optische niet-reciproke absorptie uiteraard varieert met de breedte van de golfgeleider.

in een geïntegreerde laser/isolator module, de isolator in deze versie niet enkel de extinctieverhouding van het lasersignaal sterk degradeert, maar ook door zijn uitvoering als SOA de optische signaal-tot-ruis-verhouding verlaagt. Als gevolg hiervan kan de isolator in deze uitvoering een extra vermogenpenalty toevoegen aan de nodige receivergevoeligheid om een voldoende lage bitfoutrate te behouden bij transmissie over optische vezel. Fig. 5.7 toont schematisch een uitbreiding van de meetopstelling die dergelijke transmissie-experimenten moet toelaten (naast de basiskarakterisatie van de optische niet-reciprociteit). Binnen het tijdsbestek van dit onderzoek is slechts de basisuitbreiding gebouwd, namelijk die die toelaat de niet-reciprociteit van extern geïnjecteerd CW TM licht te detecteren. De meer geavanceerde systeemkarakterisatie zal het onderwerp vormen van een opvolgend doctoraatsonderzoek. Daarnaast moet natuurlijk ook nagegaan worden of de component effectief erin slaagt de lasermodule te beschermen, m.a.w. of de drempel voor coherentie-collapse effectief toegenomen is. Daartoe volstaat het een regelbare spiegel opgenomen in een directionele koppelaar aan het detectiefacet te plaatsen.

5.2 Experimentele resultaten

5.2.1 Processing

Zoals uitvoerig toegelicht in de inleiding van dit werk is veruit het grootste voordeel van dit isolatorconcept zijn “eenvoudige” integreerbaarheid met standaard III-V DFB lasermodules. Dit is voornamelijk een rechtstreeks gevolg van het feit dat deze component in principe niets anders is dan een SOA met een ietwat exotischer metaal als extra contact. Echter, het is precies deze extra metallizatiestap die mogelijk zorgen baart. De reden daarvoor is geschetst in Fig. 5.8. Zoals in standaard laserprocessing, vereist deze metallizatie een kritische alignatie met submicronprecisie van een isolatiemasker met de geëtste ridges. Voor standaard ridge lasers is een kleine alignatiefout doorgaans geen probleem, maar voor de SOA/isolator ridges die gekenmerkt worden door een heel lage aspectratio (als gevolg van het hoge reële deel van de brekingsindex van het metaal in combinatie met de dunne InP claddinglaag) kan een dergelijke alignatiefout drastische gevolgen hebben. In het slechtste geval wordt hierdoor een deel van het metaalcontact gevormd naast de ridge, en wordt er stroom geïnjecteerd in een gebied waar de mode niet geleid is. Het is duidelijk dat heel bijzondere aandacht zal moeten gespendeerd worden aan dit technologisch probleem, aangezien het niet eenvoudig op te lossen valt zolang een ridge layout wordt gebruikt voor de SOA/isolator.
De rest van de processing is redelijk standaard. Er dient enkel nog opgemerkt te worden dat voor alle devices een eenvoudig FP lasermasker is gebruikt zonder daarbij te werken met getilte facetten noch deze laatste te voorzien van een AR coating. Het is duidelijk dat er voldoende karakterisatiemiddelen ter beschikking staan om de niet-reciproke caviteit te karakteriseren zelfs in aanwezigheid van resonanties, en het is uiteraard eenvoudiger een structuur met rechte facetten uit te meten. Enkel voor de meer geavanceerde karakterisatie-experimenten moet er gedacht worden aan het vermijden van deze resonanties.

5.2.2 Generatie 1: CoFe/InGaAsP SOA/isolator

Zoals toegelicht in sectie 3.3.2.1 zijn er vijf eerste-generatie structuren ontworpen. Hun details zijn terug te vinden in de Tabellen 3.6 en 3.7. Er is ook op gewezen dat de gesimuleerde waarden voor de niet-reciproke absorptie en de nodige materiaalwinst met enige voorzorg moeten bekeken worden wegens de vele onzekerheden ten tijde van het ontwerp.

Op alle structuren zijn succesvolle TM ASE hysteresismetingen uitgevoerd (zie Fig. 5.9). Deze kwalitatieve waarnemingen van de geïntegreerde optische niet-reciprociteit golden als het eerste wereldwijde bewijs van de werking van het nieuwe concept. Tegelijk echter brachten ze meteen een aantal problemen aan het licht. De waargenomen remanentie was niet in overeenstemming met de metingen van vorig hoofdstuk ondanks het feit dat de gekarakteriseerde devices een redelijke hoge w/l verhouding hadden (de breedte van de metaal lift-off structuren op het contactmasker is steeds 50μm). Dit was een duidelijke aanduiding dat er iets fout was met het metaalcontact. De lage remanentie doet vermoeden dat de effectieve breedte van het CoFe-contact veel lager is dan 50μm. Bovendien werd er bij de hysteresismetingen vastgesteld dat de geëmitteerde vermogens heel laag waren.
Dit werd nadien bevestigd bij de kwantitatieve ASE-fitting metingen. Vreemd genoeg werd er geen enkele trend vastgesteld tussen de gedecteerde vermogenniveaus en de dikte van de InP bufferlaag. Dit in combinatie met de lage remanentie deed des te meer vermoeden dat er iets grondigs was fout gegaan bij de processing van het CoFe contact. En dat hierdoor op redelijk willekeurige manier een groot verlies aanwezig was in bijna alle geprocessede devices. Dit werd bovendien nog eens extra bevestigd door de gemeten voorwaarts-achterwaarts ASE verhouding S (zie Fig. 5.11). De gemeten verhoudingen bleken slechts heel matig te variëren met lengte en geïnjecteerde stroomdichtheid. Enkel de 3QW "risk" device vormde hierop een uitzondering. Anderzijds is het wel zo dat met beide ASE modellen (4) en (5) een heel goede convergentie werd waargenomen bij de niet-lineaire kleinste kwadraten fit. De resultaten van deze fitting zijn samengevat in Tabel 5.2. De geëxtraheerde waarden voor de fittingparameters zijn enerzijds bevreemdend en anderzijds een bevestiging van eerdere vermoedens. De B-parameter die alle verliesparameters bundelt (op de niet-reciproke absorptieshift na) is voor alle structuren hoog tot heel hoog en volgt geen enkele trend tussen de verschillende lagenstructuren onderling. Het verlies van de "normal" 3QW structuur is bijvoorbeeld dubbel zo hoog als dat van de "risk" 3QW structuur, terwijl de eerste een veel dikkere InP bufferlaag heeft. Hetzelfde beeld wordt waargenomen voor de niet-reciproke absorptieshift $\Delta\alpha$. Ook hier ontbreekt elke logische trend, en heeft opnieuw de lagenstructuur met de dikste InP bufferlaag het grootste niet-reciprook effect!(?) Anderzijds zijn de waarden voor de differentiële winstparameter A wel redelijk goed in overeenstemming met karakterisatieresultaten van het vorige hoofdstuk (zie Tabel 4.10 en Fig. 4.20). Dit doet opnieuw vermoeden dat het probleem zich niet zozeer bij de actieve laag situeert maar eerder te zoeken is bij het metaalcontact.
Om zeker te zijn dat er geen probleem was met het fittingmodel zelf, werden

op een aantal devicestructuren Hakki-Paoli metingen uitgevoerd om rechtstreeks het ongeperturbeerde modale verlies te bepalen. Deze metingen bevestigden de conclusies van de ASE fittingmeting en bewezen de geldigheid van de gebruikte modellen. SEM inspectie van de facetten leverde uiteindelijk de verklaring (zie Fig. 5.14 en 5.15). De hierboven aangehaalde foutieve alignatie van het isolatiemasker heeft ertoe geleid dat het isolatiemateriaal opengeëtst werd naast de ridges. Op die manier bevindt er zich CoFe heel dicht bij de geleidende laag van het device, wat naast de hoge verliezen tegelijk ook de daarmee gepaard gaande hogere waarden voor $\Delta\alpha$ verklaart. Enkel de risk 3QW en de 9QW devices leken quasi vrij van deze processingfout[12]. Naast de alignatiefout onthulden de SEM foto's ook waarom de remanentie in alle devices veel lager was dan verwacht op basis van de vormanisotropiemetingen van het vorig hoofdstuk. De dikte van het gebruikte polyimide isolatiemateriaal is dermate hoog dat het zo goed als zeker is dat het "verval" aan de rand van de contactvensters niet gevolgd kan worden door het slechts 50nm dikke CoFe. Met andere woorden de magnetische flux wordt onderbroken aan de randen van het contactvenster leidend tot een magnetisch contact met slechts de breedte van de ridge i.p.v. het lift-off masker en waardoor de w/l verhouding inderdaad veel kleiner is.

De eerste-generatie demonstratoren waren een partieel succes. De beste geobserveerde waarden (2.0 dB/mm en een modaal voorwaarts verlies van 45/cm bij een stroomdichtheid van 25kA/cm^2)[13] zitten schijnbaar ver van de simulatiewaarden af. Het zijn echter wel de allereerste kwantitatieve bewijzen van het nieuw concept. Daarenboven bestonden er ten tijde van het ontwerp veel onzekerheden. Een herschatting van deze ontwerpen gebruik makend van de achteraf bepaalde magneto-optische eigenschappen van CoFe en van het (beperkte) winstgedrag van InGaAsP QW's onder trek, bewijst dat deze eerste resultaten uiteindelijk toch niet zo ver van de eigenlijke ontwerpwaarden afzitten! Tabel 4 vergelijkt de gemeten waarden met de herschatting van het eerste generatie ontwerp. Merk hier wel bij op dat de ontwerpwaarden gelden voor een geïdealiseerd 1D model en dat de gemeten waarden bekomen zijn op een 2D ridge die bovendien slechts voor 70% bedekt is met CoFe (zoals de SEM foto's duidelijk maken).

5.2.3 Generatie 2: CoFe/InAlGaAs SOA/isolator

De tweede generatie SOA/isolatoren maken gebruik van het veel performantere InAlGaAs materiaalsysteem, en konden ontworpen worden met een veel nauwkeurige kennis van de magneto-optische constante van CoFe en van het stroom-winstgedrag van de actieve laag. Bovendien werd de opgedane ervaring van de experimenten op de eerste generatie nuttig aangewend om het processingschema lichtjes aan te passen. Het gebruikte isolatiemasker werd "verschoven" ten opzichte van het ridge

[12] Het vreemde gedrag van de 9QW devices (verliezen zoals de risk structuur maar toch heel lage $\Delta\alpha$) kan vermoedelijk verklaard worden door een te ondiepe ets waardoor niet alleen de opsluiting ter hoogte van het metaal veel kleiner is dan verwacht maar ook stroomspreiding de modale winst doet afnemen.

[13] de devices met een processingfout worden hierbij buiten beschouwing gelaten, aangezien hun veel hogere $\Delta\alpha$ niet praktisch bruikbaar is wegens hun gigantisch verlies.

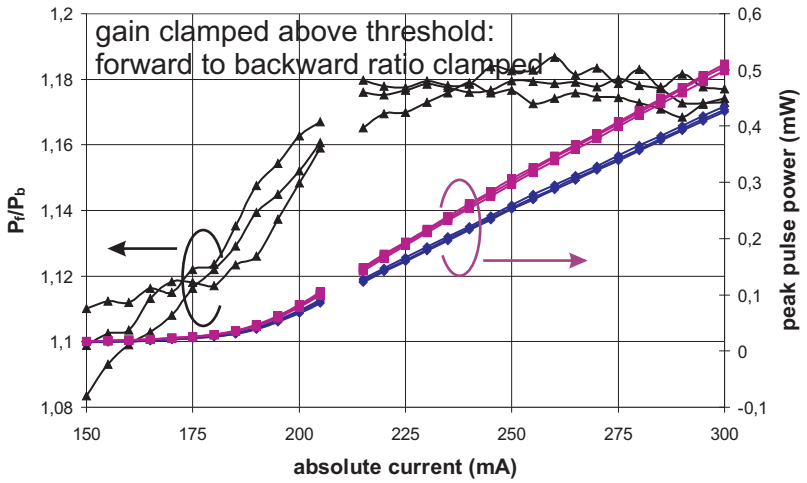
layer	t_{InP}	A	B	$A\ln(J)-B$ (for $J=20\mathrm{kA/cm^2}$)	$\Delta\alpha_{\mathrm{meas}}$	$\Delta\alpha_{\mathrm{sim}}$
	(nm)	(1/mm)	(1/mm)	(1/mm)	(dB/mm)	(dB/mm)
"normal"	300	4.0	42	-30	4.4	7.6
"risk"	250	4.5	21	-7.5	2.0	9.3
6QW	300	8.4	52	-26.8	6.5	8.7

Tabel 4: Herschatting van het eerste generatie ontwerp gebruik makend van de nadien gemeten MO constante van $Co_{90}Fe_{10}$ (zie sectie 4.1.1) en vergelijking met de gemeten waarden.

etsmasker: elke ridge werd voorzien van een contactvenster met een breedte die oorspronkelijk overeenkomt met de "vorige" ridgefamilie met een kleinere breedte (zie Tabel 5.4). Op die manier wordt de alignatie van het isolatiemasker veel toleranter en vermijdt men het risico op depositie van CoFe naast de ridge. Bovendien werd Al_2O_3 gebruikt als isolatiemateriaal. Dit kan gedeponeerd worden met een veel betere controle op de dikte, zodat het risico op onderbreking van de magnetische flux in de gedeponeerde CoFe contacten verkleind wordt. Het nadeel is natuurlijk dat de ridges slechts partieel bedekt zijn met magneto-optisch materiaal. In totaal zijn er 24 families van structuren geprocessed (3 lagenstructuren, 2 CoFe legeringen en 4 ridgebreedtes). De details van de ontworpen lagenstructuren zijn gedetailleerd in de Tabellen 3.8 en 3.9.
Fig. 5.18 toont hoe als gevolg van de aangepaste processing de hysteresiscurve nu wel een vierkant gedrag heeft dat overeenkomt met de waarnemingen van vorig hoofdstuk. De remanentie bedraagt bijna 90%, zodat de isolatie in remanentie slechts weinig zal verschillen van die in saturatie. Er werden veel hogere TM ASE vermogens gedetecteerd. Dit werd verwacht op basis van het veel sterkere winstgedrag van de InAlGaAs wells, maar is natuurlijk ook gedeeltelijk een gevolg van het feit dat de ridges slechts voor typisch 50% bedekt zijn met metaal. De breedste structuren (7 en 9μm ridgebreedte) vertoonden zelfs niet-reciproke laserwerking! Fig. 7 toont hiervan een voorbeeld.

Boven drempel klemt de winst in de niet-reciproke caviteit en klemt dus de verhouding van het voorwaarts tot het achterwaarts vermogen. Zoals hierboven uitgelegd laat dit een heel eenvoudige karakterisatie van de niet-reciproke absorptie toe. De geobserveerde niet-reciproke absorptieshift op alle lasende devices is samengevat in Tabel 5.

De verschillende tendenzen die voorgesteld zijn in deze tabel mogen niet blindelings veralgemeend worden. Het is zo dat het waarnemen van een duidelijke trend binnen een familie devices eerder uitzonderlijk was. Vaak werden belangrijke variaties op de drempelstroomdichtheid en de differentiële efficiënties waargenomen. Zeer vermoedelijk is dit gevolg van een multimodaal effect. Door de partiële bedekking van de ridges is het zeer waarschijnlijk dat de eerste (of nog hogere) orde TM mode het eerst de laserdrempel bereikt, omdat die duidelijk veel minder metaalverlies voelt. Dit werd bevestigd door verre veldmetingen. Aangezien deze hogere orde modi één of meerdere knoopvlakken hebben binnen de ridge, is het



Figuur 7: Voorbeeld van niet-reciproke laserwerking op een $Co_{50}Fe_{50}$ "R2" sample (ridgebreedte = 9μm, lengte = 520μm). Gebruik makend van (6) wordt een niet-reciproke absorptieshift gevonden van 6.45/cm of 2.8dB/mm.

family	$\Delta\alpha$ (dB/mm)		
$Co_{90}Fe_{10}$, 7μm	O	R1	R2
	1.95	1.52	3.26
$Co_{90}Fe_{10}$, R1	5μm	7μm	9μm
	3.47	1.52	1.09
R2, 5μm	$Co_{90}Fe_{10}$	$Co_{50}Fe_{50}$	
	1.95	2.82	

Tabel 5: Waargenomen trends voor $\Delta\alpha$ als functie van de verschillende device parameters. De eerste kolom duidt aan binnen welke familie van devices de trend geobserveerd is.

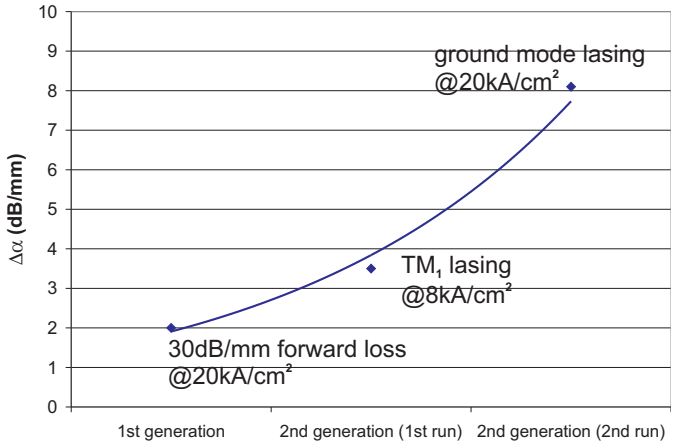
duidelijk dat hun caviteitsverlies (dat moet gecompenseerd worden om laserwerking te krijgen) zeer gevoelig is aan de precieze positie van het metaal. Vorige sectie heeft duidelijk gemaakt dat submicron variaties op deze positie bijna onvermijdelijk zijn. De beperkte trends die dus voorgesteld worden in Tabel 5, zijn eerder kwalitatief dan kwantitatief te interpreteren. Toch zijn ze tot op zekere hoogte te verklaren. Het gebruik van de equi-atomische CoFe legeringssamenstelling levert inderdaad een verbetering van bijna 70%. En de trend als functie van de ridgebreedte is te verklaren door de hogere orde lasing. De centrale metaalbedekking van de ridge blijft quasi 50% maar naarmate de breedte van de ridge toeneemt zal uiteraard de overlap van de eerste orde mode met het metaal afnemen. In elk geval is het beste resultaat (3.47dB/cm) een verbetering van 75% ten opzichte van de

eerste generatie, en is dit resultaat heel waarschijnlijk een ernstige onderschatting van de intrinsiek haalbare $\Delta\alpha$ gelet op de hogere orde lasing.
Op deze lasende devices werd ook voor het eerst niet-reciproke transmissie waargenomen van extern geïnjecteerd licht. De waargenomen niet-reciprociteit lijkt in grote lijnen de resultaten van de laserkarakterisatie te voorspellen. Dit resultaat kwam er echter op het eind van dit doctoraatstonderzoek en de meetprocedure voor deze transmissie-experimenten vereist nog heel wat optimalizatie. Dit zal uiteraard een van de aandachtspunten zijn in het opvolgend onderzoek van Van Parys [23]. In elk geval is de waarneming van niet-reciprociteit op extern geïnjecteerd licht zo mogelijk een nog overtuigender bewijs van de aanwezigheid en de werking van het niet-reciprook effect (voor zover de ASE metingen dat al niet waren).

Bij het afronden van deze scriptie werden de eerste resultaten waargenomen van een tweede processingrun op deze tweede-generatie-lagenstructuren. De geobserveerde niet-reciprociteit is een dermate belangrijke doorbraak dat we niet kunnen nalaten er hier kort melding van te maken. Meer hierover zal uiteraard heel gedetailleerd aan bod komen in het werk van Van Parys.
Zoals uit alle experimenten tot hier toe gebleken is, is de grootste technologische uitdaging bij de fabricatie de SOA/isolatoren het realizeren van een voldoende breed, voldoende planair CoFe contact dat effectief de volledig ridgebreedte bedekt. De oplossing gebruikt voor de tweede generatie (partiële bedekking) pakt het probleem slechts gedeeltelijk aan, en creëert een nieuwe problematiek door hogere orde geleiding toe te laten. Hierdoor wordt bovendien de ware intrinsieke sterkte van de devices gemaskeerd omdat niet nagegaan kan worden wat de niet-reciprociteit is van de grondmode. Om daar een idee over te krijgen werd in deze tweede processingrun het CoFe metaal als etsmasker gebruikt. Dit gaat uiteraard ten koste van de magnetische remanentie (wegens de heel lage w/l aspectratio CoFe contacten; zie Fig. 5.22) maar vermijdt anderzijds wel hogere orde mode lasing, omdat nu het verlies van een eventuele hogere orde mode nauwelijks zal verschillen van dat van de grondmode terwijl de overlap met de QW's van deze hogere orde mode veel lager blijft dan die van de grondmode. De niet-reciproke lasing die opnieuw werd waargenomen leverde in een eerste experiment al meteen een niet-reciproke absorptie op van niet minder dan 8.1dB/mm! Dat is niet enkel een verbetering met een factor vier ten opzichte van de eerste generatie structuren, maar bovendien en vooral bijna on-target met de ontwerpwaarden van Tabel 3.9 (die trouwens enkel voor een slabmode gelden)!

5.3 Conclusies

Dit laatste hoofdstuk is het succesvolle culminatiepunt van het voorbereidende werk beschreven in de vorige hoofdstukken: van de ontwikkeling van het theoretisch framework via het ontwerp van de subcomponenten en de lagenstructuur van de SOA/isolator tot de experimentele karakterizatie van de bouwstenen.
Dit hoofdstuk heeft niet enkel het eerste wereldwijde bewijs geleverd van de haalbaarheid van het innovatieve geïntegreerde isolatorconcept van Nakano, maar heeft vooral aangetoond dat er een goed begrip is van de methodologische karakteriza-



Figuur 8: Evolutie van de niet-reciproke absorptieshift $\Delta\alpha$ voor de verschillende generaties gekaraktiseerde SOA/isolatoren.

tie van deze component en vooral van de verschillende technologische uitdagingen voor de realizatie van de SOA/isolator. Elke nieuwe generatie (of processing) leverde een duidelijke verbetering van de niet-reciproke absorptieshift: van eerste foutief geprocessede devices veraf van transparantie, via verbeterde transparante (maar multimodale) niet-reciprook lasende devices tot monomodale voorwaarts transparante devices met een experimenteel gedrag dat bijna on-target is met de ontwerpwaarde. Fig. 8 schetst deze evolutie.
Ook al rest er nog heel wat optimalizatiewerk alvorens er gesproken kan worden van een component die kan concurreren met bestaande free-space optische isolatoren, toch is het duidelijk dat deze component nog niet het achterste van zijn tong heeft laten zien. Het grootste probleem lijkt de processing van een voldoende breed planair CoFe te zijn. Dit probleem kan elegant vermeden worden door een begraven golfgeleiderstructuur te gebruiken. Dit is trouwens ook de technologie die zich het best zal lenen voor de uiteindelijke integratie met een DFB laser. Bovendien hebben de hier beschreven experimenten duidelijk gemaakt dat de belangrijkste struikelblok niet zozeer de magneto-optische sterkte van het ferromagnetisch metaal is, maar wel het winstgedrag van de actieve MQW laag. De waarden van de haalbare $\Delta\alpha$ voor beide generaties (na herschatting voor de eerste generatie) zijn bijna dezelfde, maar enkel de tweede generatie op basis van InAlGaAs slaagt erin transparantie te bereiken. Voor een 25dB isolator is er echter zelfs voor het InAlGaAs materiaal nog wat verbetering wenselijk. De beste InAlGaAs device haalt immers op dit moment 25dB in een lengte van ongeveer 3mm met een transparantiestroom van om en bij de 1A! De resultaten van vorig hoofdstuk hebben echter aangetoond dat de InAlGaAs lagen nog kunnen verbeterd worden.

6 Conclusies en perspectieven

6.1 Conclusies

Dit doctoraatsonderzoek heeft een innovatief (en misschien zelfs revolutionair) concept voor een eenvoudig integreerbare optische golfgeleiderisolator theoretisch en experimenteel onderzocht. Het concept van Nakano, door hem voorgesteld eind 1999, was op dat moment niet meer dan een interessant theoretisch magneto-optisch ideetje dat (op papier) quasi alle problemen van bestaande ontwerpen voor geïntegreerde golfgeleiderisolatoren kon oplossen, maar dat vooral heel zwaar leunde op een aantal voorheen niet geverifieerde veronderstellingen betreffende actieve lagen en ferromagnetische metalen en hun contact met halfgeleiders. Het device voorgesteld door Nakano is in principe niet gelimiteerd tot de door ons bestudeerde golflengteband rond 1300nm. Toch werd gekozen om enkel te concentreren op een toepassing voor die golflengte omdat enerzijds het vooral voor dergelijke metro-toepassingen is dat de potentiële kostreductie van het idee van Nakano een impact zal hebben, en anderzijds omdat het zeer ambitieus zou zijn om zowel voor 1300nm als voor 1550nm de nodige sterke TM-selectieve actieve MQW lagen te ontwikkelen.

In dit werk hebben we gedetailleerd de route beschreven die we gevolgd hebben van theoretisch concept naar prototype niet-reciproke SOA, de naam geïntegreerde isolator waardig. Een grondig theoretisch framework is ontwikkeld voor de modellering van niet-reciproke optische golfgeleiders in het algemeen en magneto-optische op het transversaal Kerr effect gebaseerde golfgeleiders in het bijzonder. Een eerste modellering van het Nakanoconcept in een vereenvoudigde versie gebruik makend van dit theoretisch kader, heeft duidelijk gemaakt dat alvast vanuit theoretisch oogpunt een praktische realizatie van het idee haalbaar is.

Een doorgedreven theoretische studie van de verschillende subcomponenten heeft geleid tot een beter begrip van hun gedrag, van de manieren om hen te ontwerpen en van de soms subtiele karakterizatiemethodes nodig om dit gedrag te beoordelen. Zo is het bijvoorbeeld verduidelijkt waarom het bijna onmogelijk is een theoretische vuistregel te vinden om een geschikt ferromagnetisch magneto-optisch metaal te kiezen, maar dat eerder elk nieuw ferromagnetisch metaal opnieuw ellipsometrisch moet gekarakteriseerd worden. De veralgemeende ellipsometrische technieken die hiervoor nodig zijn houden een aantal subtiliteiten in vooral omwille van de “kleinheid” van de magneto-optische gyrotrope effecten. De studie van de magnetische anisotropie van het ferromagnetisch contact heeft verduidelijkt hoe het gedrag van dit contact als dunne film magneet zal overheerst worden door de vormanisotropie als gevolg van zijn lange en smalle rechthoekige vorm. Door de typische functie die het ferromagnetisch metaal in dit device moet vervullen (naast die van permanente magneet) kan de vormanisotropie enkel tegengewerkt worden door tijdens de depositie met behulp van een extern magneetveld te trachten een voorkeursinrichting voor de magnetische momenten te induceren. De studie van de Ohmse elektrische contactfunctie van het ferromagnetische metaal heeft vooral

aandacht geschonken aan het ontwikkelen van een nauwkeurige methodologie voor de karakterizatie van de contactresistiviteit. De realizatie van een goed Ohms contact is immers op de eerste plaats "enkel" een kwestie van voldoende hoge doperingsniveaus te halen in de halfgeleiderlagen. De karakterizatie echter wordt te vaak stiefmoederlijk behandeld waarbij de contactresistiviteit met behulp van niet geschikte structuren opgemeten wordt.
De invloed van roostermisaanpassing op het winstgedrag van QW's is grondig bestudeerd. Zonder al te diep in detail te treden kon ingezien worden hoe het gebruik van rekvervroming veel meer voordelen biedt dan enkel een TM-selectieve winst verzekeren. Een Al-gebaseerd (InAlGaAs) bleek hierbij voorlopig (tot aan de komst van nitride-achtige systemen) de kandidaat bij uitstek te zijn voor 1300nm toepassingen.
Tot slot werden op basis van de opgedane kennis twee generaties lagenstructuren voor de SOA/isolator ontworpen. Ook al was ten tijde van het ontwerp niet alle inputdata experimenteel beschikbaar, toch kon als belangrijke conclusie vastgesteld worden dat dit type isolator steeds dicht bij de winstlimiet van de QW's zal opereren. Nakano's oorspronkelijke voorspellingen zijn dan ook lichtjes herschat: een 25dB isolator lijkt theoretisch haalbaar in een lengte van 2–3mm en met een stroomdichtheid voor voorwaartse transparantie van ongeveer 5kA/cm^2.

Het gedrag van de subcomponenten werd één voor één voor experimenteel gekarakteriseerd. Deze experimenten maakten duidelijk dat in principe voor elke subfunctie van de niet-reciproke SOA de nodige performantie aanwezig was. Het gekozen CoFe ferromagnetisch legeringssysteem heeft een gyrotrope constante die voldoende sterk blijkt volgens de ontwerpsimulaties. De magnetische vormanisotropie kan tot op zekere mate gecontroleerd worden en een lage contactresistiviteit kan bekomen worden op een contactlaag die bovendien bijna volledig transparant kan gehouden worden voor 1300nm licht. Heel positieve resultaten werden bekomen op rekvervormde InAlGaAs MQW lagen. Er kan zelfs met enige voorzichtigheid gesteld worden dat het waargenomen winstgedrag van deze lagen de performantie van de tot op heden beste drukvervormde MQW lagen overstijgt.

De laatste stap in dit onderzoek was de succesvolle "samenvoeging" van alle subcomponenten tot een wereldwijd allereerste prototype van de SOA/isolator. Hierbij is niet enkel het eerste bewijs van het concept van Nakano geleverd, maar is eveneens een verregaand inzicht verworven in de methodologie voor de karakterizatie van deze devices en vooral in de technologische uitdagingen om deze componenten te fabriceren. Dit wordt voldoende onderlijnd door de waargenomen evolutie in de performantie van de verschillende generaties van gefabriceerde devices. Een 25dB isolator die werkt bij een lage stroominjectie is er nog niet, maar de evolutie gaande van de eerste kwalitatieve observatie van het effect via TM ASE hysteresis metingen tot een monomodale voorwaarts transparante component met een niet-reciprook absorptieverschil van 8.1dB/mm is opmerkelijk.

6.2 Perspectieven

Dit werk kan beschouwd worden als een eerste stap in de richting van een nieuw revolutionair principe voor een geïntegreerde optische isolator: het aantonen van het principe en van het groeipotentieel. Er moeten nog vele opvolgende stappen gezet worden. Het ontwerp van de component is zeker nog niet optimaal en de effectieve monolithische integratie met een DFB-laser is zelfs voorlopig nog niet verder dan een embryonaal stadium. Bovendien zijn er nog heel belangrijkere deviceparameters die helemaal niet gekarakteriseerd zijn, meer in het bijzonder de reactie van de SOA/isolator op een extern geïnjecteerd gemoduleerd signaal. Dit zijn de drie voornaamste paden voor verder onderzoek.
Het beste InAlGaAs SOA/isolator device haalt op dit moment 25dB in een lengte van ongeveer 3mm met een transparantiestroom van om en bij de 1A. Niet eenvoudig om als idee verkocht te krijgen aan een telecomfabrikant! Toch hebben we redenen om te geloven dat er veel ruimte is voor verbetering van deze karakteristieken.

- De modale winst van de MQW lagen kan nog verbeterd worden, ofwel door het aantal wells te vermeerderen (alhoewel dit vermoedelijk enkel de stroom verder zal opdrijven), maar vooral door verder de trekvervorming van de InAlGaAs wells op te drijven. Een hogere modale winst kan een hogere niet-reciprociteit leveren bij eenzelfde stroom of kan eenzelfde niet-reciprociteit leveren bij een lagere stroom.

- Het gebruik van een begraven structuur zal mogelijk vele bestaande problemen oplossen. Op de eerste plaats denken we hierbij aan de processing van het contact. Maar ook de stroominjectie zal efficiënter zijn en dus ook de interne kwantumefficiëntie.

- Het is niet onmogelijk dat er nog MO materialen (metallisch of niet) gevonden worden met een belangrijk sterkere gyrotropie dan de twee tot nu toe gekarakteriseerde CoFe legeringen. Als het materiaal niet metallisch is kan altijd gedacht worden aan een zijdelingse stroominjectie.

- In dit werk is het niet-reciprook absorptie-effect in zijn meest eenvoudige gedaante gebruikt. Het is echter niet onwaarschijnlijk dat net zoals bij zoveel andere functies in de fotonica de sterkte van dit effect kan opgedreven worden door gebruik te maken van resonantie, bijvoorbeeld door een niet-reciproke absorber op te nemen in een (actieve) ring resonator. Een heel recent nieuw onderzoeksgebied bestaat erin fotonische kristallen te combineren met magneto-optische effecten. Ook hier kunnen vermoedelijk nieuwe isolatorconcepten uit naar voor komen.

Tot slot, zoals hoort bij academisch onderzoek, blijven er uiteraard nog een aantal heel wetenschappelijk theoretische uitdagingen: het uitbreiden van de modelleringstools naar de derde dimensie, een betere modellering van de ladingsdragers-diffusie-effecten en van hogere orde mode effecten,.... En om heel ambitieus te eindigen: het zoeken naar een polarizatie-onafhankelijk versie van deze component die werkt over de hele telecom C-band (knipoog).

Bibliografie

[1] A. K. Zvezdin, and V. A. Kotov, Chapter 2 in *Modern Magnetooptics and Magnetooptical Materials.* (Institute of Physics Publishing, Philadelphia, USA, 1997).

[2] M. Mansuripur, "Analysis of multilayer thin-film structures containing magneto-optic and anisotropic media at oblique incidence using 2×2 matrices." *J. Appl. Phys.,* vol. 67(10), pp. 6466–6475, 1990.

[3] Š. Višňovský, R. Lopuňík, M. Bauer, J. Bok, J. Fassbender, and B. Hillebrands, "Magnetooptic ellipsometry in multilayers at arbitrary magnetization." *Optics Express,* vol. 9(3), pp. 121–135, 2001.

[4] C.-Y. You, and S.-C. Shin, "Generalized analytic formulae for magneto-optical Kerr effects." *J. Appl. Phys.,* vol. 84(1), 541–546, 1998.

[5] K. Ando, T. Okoshi, and N. Koshizuka, "Waveguide magneto-optic isolator fabricated by laser annealing." *Appl. Phys. Lett.,* vol. 53(1), pp. 4–6, 1988.

[6] R. Wolfe, J.F. Dillon,Jr., R.A. Lieberman, and V.J. Fratello, "Broadband magneto-optic waveguide isolator." *Appl. Phys. Lett.,* vol. 57(10), pp. 960–962, 1990.

[7] J. Fujita, M. Levy, R. M. Osgood, Jr., L. Wilkens, and H. Dtsch "Waveguide optical isolator based on MachZehnder interferometer." *Appl. Phys. Lett.,* vol. 76(16), pp. 2158–2160, 2000.

[8] H. Yokoi, T. Mizumoto, M. Shimizu, N. Futakuchi, N. Kaida, and Y. Nakano, "Direct Bonding between Quaternary Compound Semiconductor and Garnet Crystals for Integrated Optical Isolator." *Jpn. J. Appl. Phys. Pt.1,* vol. 38(1A), pp. 195–197, 1999.

[9] K. Sakurai, H. Yokoi, T. Mizumoto, D. Miyashita, and Y. Nakano, "Fabrication of Semiconductor Laser for Integration with Optical Isolator." *Jpn. J. Appl. Phys.,* vol. 43(4A), pp. 1388–1392, 2004.

[10] M. Takenaka, and Y. Nakano, "Proposal of a novel semiconductor optical waveguide isolator." in the Proceedings of 11^{th} *International Conference on Indium Phosphide and Related Materials,* pp. 298–292, Davos, Switzerland, 1999.

[11] W. Zaets, and K. Ando, "Optical waveguide isolator based on nonreciprocal loss/gain of amplifier covered by ferromagnetic layer." *IEEE Photon. Technol. Lett.,* vol. 11(8), pp. 1012–1014, 1999.

[12] F. Olyslager, *Electromagnetic Waveguides and Transmission Lines.* Oxford Engineering Science Series, No. 51 (Clarendon Press, Oxford, 1999).

[13] C. Vassallo, *Optical Waveguide Concepts.* Optical Wave Sciences and Technology Series, No. 1 (Elsevier, Amsterdam, 1992).

[14] G. S. Krinchik, and V. A. Artemjev, "Magneto-optical properties of Ni, Co, and Fe in the ultraviolet visible, and infrared parts of the spectrum." *Sov. Phys.-JETP,* vol. 26(6), pp. 1080–1085, 1968.

[15] S. W. Corzine, R.-H. Yan, and L. A. Coldren, Chapter 2 "Optical Gain in III-V Bulk and Quantum Well Semiconductors." in *Quantum Well Lasers,* (ed. P. S. Zory, Jr.) (Academic Press, San Diego, USA, 1993).

[16] A. R. Adams, "Band structure engineering for low-threshold high effciency semiconductor lasers." *Electron. Lett.,* vol. 22(), pp. 249–250, 1986.

[17] E. Yablonovitch, and E. O. Kane, "Reduction of lasing threshold current density by the lowering of valence band effective mass." *J. Lightwave Technol.,* vol. 4(5), pp. 504–506, 1986.

[18] T. G. Andersson, Z. G. Chen, V. D. Kulakovskii, A. Uddin, and J. T. Vallin, "Variation of the critical layer thickness with In content in strained $In_xGa_{1x}As$-GaAs quantum wells grown by molecular beam epitaxy." *Appl. Phys. Lett.,* vol. 51(10), pp. 752–754, 1987.

[19] T. Ishikawa, and J. E. Bowers, "Band lineup and in-plane effective mass of InGaAsP or InGaAlAs on InP strained-layer quantum well." *IEEE J. Quantum Electron.,* vol. 30(2), pp. 562–570,1994.

[20] J. Piprek, P. Abraham and J. E. Bowers, "Cavity Length Effects on Internal Loss and Quantum Efficiency of Multi-Quantum-Well Lasers." *IEEE J. Sel. Topics Quantum Electron.,* vol. 5(3), pp. 643–647, 1999.

[21] K. H. J. Buschow, P. G. van Engen, and R. Jongebreur, "Magneto-optical properties of metallic ferromagnetic materials." *J. Magn. Magn. Mater.,* vol. 38(1), pp. 1–22, 1983.

[22] G. Jones, and E. P. O'Reilly, "Improved performance of long-wavelength strained bulk-like semiconductor lasers." *IEEE J. Quantum Electron.,* vol. 29(5), pp. 1344–1354, 1993.

[23] W. Van Parys, "Geïntegreerde Optische Isolatoren en Circulatoren Gebaseerd op de Interactie van Licht met Ferromagnetische, Gyrotrope Metalen." Ph. D. thesis (*in preparation*), Ghent University, Belgium, 2007.

English Text

1 Introduction

THIS research focuses on the development of a novel monolithic integration scheme for an optical isolator, or, in a wider perspective, for optical non-reciprocity in general. The central idea behind this novel concept is to exploit the previously well-known strong non-reciprocal properties of ferromagnetic metals, but which have up till now never been considered a serious competitive material in view of their extremely high optical losses.

This introductory chapter presents the specific context explaining the need for an optical isolator and more specifically an integrated version of it. It will be explained why this crucial optical component has evaded any attempt at integration. For this purpose, a short introduction to magneto-optics will be given.

The novel concept will be schematically introduced, indicating its major advantages over existing integration schemes, while at the same time not ignoring its possible pitfalls.

This chapter concludes by an outline of this work and an overview of the achieved scientific output.

1.1 Context

1.1.1 Low-cost metro laser sources

Ever since the infamous year 2000 burst of the technological stock market bubble, the vendors and the manufacturers of optical telecom equipment have been under increasing pressure to drop selling prices in a do-or-die fashion, leading to a (forced) reduction in the average unit price by more than a factor 4 compared to the pre-2000 era [1]. Obviously it is better for the manufacturers to keep on selling albeit with a loss, than not selling at all. This situation can of course not be perpetuated without having the real manufacturing cost following the declined market price. Next to this purely economic survival reality, there is also the simple economic motivation of price reduction in order to reach a larger clientele. After all the often-heard talk of bringing fibre-to-the-home will only materialize if all opto-electronic equipment needed in an optical access network becomes affordable for a customer group that isn't limited to the rich-and-the-few – read, the large business companies. In this way, optical telecom or optoelectronics – with its twenty-something of age – nowadays faces a situation that is very much similar to the one electronics was facing when it had about "the same age". Namely, the situation in the mid-seventies where prices for PC's and related equipment were very high and only a few – primarily businesses – could afford the expense. With the continued improvement of the price/performance ratio of these products, PC's are found today in almost every household in the industrialized world. To achieve this same broad distribution in the telecom industry, the fibre connection to the end-user has to be reduced substantially in price.

Recent market analyses and forecasts show that since the 2000 crash, for the reasons explained above, vendors have to rely increasingly on revenues from optical metropolitan WDM or SDH/SONET equipment. Whereas before 2000 sales of metro WDM equipment only generated a few % of the revenues as compared to 20% for long haul WDM, the situation has completely swung with nowadays (2^{nd} quarter 2003) an almost equal revenue of just over 10% from metro and long haul WDM [2]. In other words income from long haul has halved whereas metro has more than quintupled. The rise of the metro market is logical, as it is here that cost really matters, and as the number of customers, both existing and potential, is much larger. At the same time, the lower cost of metro equipment is accompanied by a lower margin of profit per unit. This analysis thus fits perfectly well with the above argumentation describing the big need for decrease of manufacturing cost of metro equipment. On the one hand, metro is the place where people are still buying. On the other hand it is the place where a large growth potential exists in the shape of the "fibre-to-the-home" dream. The foremost important metro component is the transceiver, more in particular the 1300nm laser source.

1.1.2 Packaging cost reduction through integration

The most important element of the cost structure of laser diodes is the assembly and more specifically the packaging cost. Indeed, while wafer production techniques (for the manufacturing of the laser diode itself) allow mass-production,

packaging remains a cost bottleneck, as it is not supported by parallel processing. The big difference between the much more mature and cheaper microelectronics packaging and photonic component packaging is that the latter faces the extra challenge of the need to reach and maintain optical alignment, concerns with regards to immediate or gradual contamination of the optical path, the handling of often extremely small and fragile parts, and the need to eventually link to an optical fiber. Inside a typical fixed wavelength laser diode package there are at least four optical modules that need to be aligned: the laser diode itself, a monitoring photodiode, an optical isolator and a fibre. If dispersion is an issue and/or wavelength tunability, add to this another low-chirp external modulator and a wavelength locker. Each extra component requires in principle the addition of a high aperture lens (and for the optical isolator two optical polarizers on top) and a time-consuming active optical alignment step. Next to this, the laser package almost always needs to be supplemented with an extra Peltier temperature controller to guarantee wavelength and output power stability in varying temperature conditions. This doesn't really create an extra alignment issue but it increases the total component count of the package and the total power consumption. Assembling all of these individual components into a package doesn't only create an alignment issue, but also poses challenges of mechanical and thermal stability to the manufacturer.

With individual component prices often not higher than several 10's €, all of the above makes clear that the total cost of the laser package (which can amount up to several 1000's €) is for over 90% determined by packaging costs. Cost reduction must therefore be based on simplifying, cost-cutting modifications of either the laser package itself and/or the packaging procedure. There are two ways to achieve this[1].

Packaging simplification By this it is meant

- that any package component that can be sacrificed while maintaining a laser performance still meeting the requirements of the transmission system for which it is used, is left out.
- and/or that one tries to relieve the complexity of the packaging procedure by "clever engineering tricks".

Examples of the former technique are the operation of a laser diode in uncooled conditions [3], and/or isolator-free [4], and/or directly modulated [5], ... Examples of the latter technique are the use of passive alignment techniques by using structurized (i.e. grooved, indented, ...) carrier platforms that quasi self-align the components [6], the use of flip-chip bonding, ... An example that combines both is the direct passive quasi-butt coupling of the laser to the fibre [6], as this is only possible if at least the isolator and the

[1]The references cited in the examples given below are by no means intended to give the most remarkable achievements by serve merely as examples and as starting point to find more similar research work. The work on fabrication simplification (in its broadest sense) of laser diodes is one of the biggest research ares in opto-electronics.

modulator, in the case of external modulation, are sacrificed. Another example is the use of a monolithically integrated spot-size converter in the laser diode in order to narrow the laser's far field and relax the alignment tolerances [3, 7]. Again this makes only sense if no optical elements (isolator, modulator, ...) are needed between the laser output facet and the fibre.

Integrate By this it is obviously meant that one aims at combining two or more of the individual devices needed in the laser package into a single component, by assembling them

- either in a hybrid manner. The terminology "hybridly integrated" has been used in a variety of situations. In this work it is used to describe two dissimilar devices (being dissimilar from a material point of view or from an "operational" point of view – like active-passive integration) that are permanently and rigidly attached to eachother to form a single device with extended functionality.
- or in a monolithic manner. The terminology "monolithically [2] integrated" doesn't need a lot of explanation. It refers to a fabrication scheme where two or more dissimilar devices (mostly active and passive devices) are grown on a same host substrate, using the same epitaxial growth step, or techniques as epitaxial regrowth, selective area growth, quantum well intermixing, ...

Examples of hybrid integration are for instance the permanent butt-coupling integration using epoxies of a wavelength (de-)multiplexer with an amplifier array for a multi-wavelength laser [8]. Other nice examples are the wafer direct bonding of GaAs DBR mirrors onto InP gain materials for the realization of 1.55 μm VCSEL's [9], or the integration of CMOS technology with InP-technology for the realization of a complete Si platform for both electronic and photonic applications [10] , ... Examples of monolithic integration are plenty but one of the best examples is probably the invention and the development of the III-V MQW electro-absorption modulator as an alternative to traditional external $LiNbO_3$ electro-optic modulators [11]. Others include the monolithic active-passive integration of wavelength selective devices such as a multiplexer for the realization of a multiwavelength laser [12]. Monolithic integration pushed to its extremes is the development of an integrated tunable laser-modulator/wavelength converter, enabling the all-optical realization of a switch capable of replacing electronic switches [13].

Both ways to reduce the cost of the laser package have their pro's and con's. It would be to naive to state that integration is the sorcerer's stone that would solve all packaging problems. Often it is simply not possible to integrate dissimilar components onto a single carrier, because it would demand a too high compromise of the individual performances of the components. After all, integration often introduces extra optical and/or electrical boundary conditions that are not present

[2] from the Greek *monos* and *lithos*, meaning *one* and *stone*

when the devices are assembled in a free-space manner. On top of that, it is not always certain whether the fabrication of an integrated component is effectively reducing the total cost of a laser package, or if it allows increased parallel processing. Sometimes the effort needed to create an integrated device can turn out to be much higher both financially and technologically, so that it creates a situation where manufacturing cost is simply shifted. This is more often the case with hybrid integration than with monolithic integration. Still even then the integration cost can after all be justified because it allows to create previously unexisting functionality or because it solves a non-packaging related device problem. A very nice example of this is the above cited realization of a 1.55μm VCSEL through wafer direct bonding[3].
In any case if easy, uncompromising integration is possible, **integration** should always be considered above **simplification through omission of components**. Leaving out subcomponents in a package will always come at a price. However nice it may sound to be able to operate a laser by direct modulation without cooling and without an isolator. It is clear that this ultimately always limits its capabilities. If a similar cost-reduction can be achieved by integrating the modulator-free and the isolator and thus hardly sacrificing the laser's capabilities, the cost saving will have a much greater impact on the market than a modulator and isolator-free device (which might be suspiciously looked upon by the customers). On top of that the main credo of integration is of course footprint reduction. Integration reduces size, quasi eliminates alignment concerns (especially in the case of monolithical integration), enhances the thermal and mechanical stability of the total device, and reduces manufacturing cost by enabling parallel processing of (parts of) the total device, ...

1.1.3 Integrated optical isolators – the last package hurdle?

This drive towards package cost reduction through integration has over the past decades generated a plenitude of devices available in integrated versions. However, there is one class of optical components that is even now still not available in an integrated version, namely non-reciprocal devices. The best known non-reciprocal component is the optical isolator, i.e. the optical equivalent of a mechanical one-way valve. An ideal optical isolator is thus a 2-port unidirectional device that transmits light in one direction without attenuation while perfectly blocking it in the other direction and this in a way *that is independent of the position of the light source with respect to the two ports*. This latter rather artificially formulated statement is needed to underline the non-reciprocity of the component. There are indeed devices thinkable that seemingly act isolating but aren't anymore when

[3]It is well known that VCSELS are very interesting compact low-cost alternatives for traditional edge emitting lasers, but that they are not so easy to realize in the traditional 1550nm region because of the limited refractive index differences in the InP material system. As the evolution and the extension of GaInNAs material towards this wavelength window is slower than expected, wafer bonding of GaAs DBR mirrors is a very nice engineering solution.

source and detector are interchanged[4]. Such devices are simply reciprocal and only block the light under very special return conditions. The non-reciprocity of the optical isolator is a quintessential property of the device. Actual an optical isolator is just a special case of a larger family of non-reciprocal devices, namely ideal N-port circulators. These are devices where every port is intimately connected to another in a circular manner. An isolator is then simply formed by the subset of any two consecutive ports in a N-port circulator (containing N isolators).

The primary application of an optical isolator is protection of optical components whose performance is very "sensitive" to unwanted optical feedback. The laser diode is the most obvious example of a device that benefits greatly from having an isolator at its output. Without any protection against back reflected light, a laser risks of loosing its wavelength and power stability. In the worst case at return levels above the so-called coherence collapse threshold it enters a completely chaotic operation regime with a very broad quasi-incoherent spectrum characterized by a drastic reduction of the coherence length of the laser [14], [15]. It is important to stress that this chaotic behavior also alters the dynamic performances in transmission [16].
Not only a laser diode but any active device can get perturbed by back reflected light. The reason for that is very simple. The presence of gain (or better carriers) introduces an interaction between all present photons inside an active device, as opposed to passive devices where, as long as one stays in the linear regime, there is no interaction between the optical signals. As a result the ideal pass-through behaviour of an active device can get seriously perturbed under the presence of counterpropagating signals. An example is a pass-through SOA-based regenerator or wavelength convertor. Its specific nonlinear power transfer function and its dynamic behaviour can get completely perturbed by reflected light from somewhere further down the optical path. For that reason such a device is most often specifically made unidirectional by adding isolators at its entrance and its output.

In view of the importance of the optical isolator it is clear that an integrated version of it would have a great impact on the cost of a whole range of optoelectronic devices. This applies not only to the traditional transceiver devices, but also to the next-generation all-optical networking components such as crossconnects, regenerators and wavelength convertors. The biggest cost reduction in terms of percentage would of course be for devices in which the optical isolator constitutes the only component that is not integrated. This is especially the case for metro 1.3μm laser diodes. In a metro system (with its transmission distances of typically not more than 100km) price is much more an issue than in long-haul. Therefore, for metro applications it makes much more sense to use the 1300nm minimum fibre dispersion band than the minimum fibre optical absorption C-band[5], because at

[4]This latter remark is rather important as often one can find the nomenclature "isolator" used for devices that are perfectly reciprocal. Examples of these can be found in CO_2 gas lasers where the laser is protected by a polarization selective beam splitter combined with a quarter wave plate. Of course this device is only isolating because of the very specific polarization of the return light, namely a circular polarization with an inverted handedness obtained at reflection.

[5]In optical fibre communication systems the different wavelength windows are denoted by

1300nm the laser can be modulately directly with a considerably lower dispersion penalty than at 1550nm.[6] In order to obtain a comparable dispersion behaviour over the C-band, a 1.55μm laser diode package would have to be equipped with an expensive external low-chirp modulator. On top of this, recent advances in Pr-doped fibre amplifiers allowing amplification at 1300nm, may even kick 1.55 μm out of metro once and for all. In other words, developing an integrated optical isolator for a directly-modulated 1300nm metro laser would really create an important market breakthrough.

From a scientific viewpoint any integration is of course an important development. We don't want to give the impression that an integrated isolator at 1.55μm constitutes less of a challenge, but it is indeed the case that the cost reduction in terms of percentage would be much lower as long as an external modulator is needed for this wavelength. Of course advances in the chirp factor of EA modulators might change this. In any case, today non-reciprocal optical devices constitute the only type of optical components that is not available in integrated form. Their development would in essence mean that finally all possible thinkable optical functionalities could be realized in completely integrated optical subsystems.

1.2 Magneto-optics – some basic concepts

Before giving an overview of the present existing optical isolators and of the state-of-the-art attempts at developing an integrated isolator, some introductory notions of magneto-optics are needed. After all, the introduction of non-reciprocal behaviour in optical devices invariably requires the interaction of light with magnetized matter[7].

1.2.1 Non-reciprocity and origin of magneto-optics

The easiest way to understand that the presence of a permanent magnetic field (or of a material with a net magnetization) causes non-reciprocity is by considering the time inversion symmetry of Maxwell's equations. It is well known that under

letters. S-band: 1485nm → 1520nm; C-band: 1530nm → 1562nm; L-band:1570nm → 1610nm. Recently, the minimum fibre dispersion wavelength window around 1300nm (1280nm → 1320nm) got assigned the letter O.

[6] and in view of the relatively short transmission distances, this cost advantage cannot be offset by working at the absorption minimum.

[7] This is of course an extremely simplified statement. It is not correct to state that optical non-reciprocity can only come about through the use of magneto-optic materials. However, it is correct to state that it is the easiest way to generate non-reciprocity. The other way to generate optical non-reciprocity is by the so-called "magneto-electric" effect theoretically predicted by Landau and Liffschitz. It is a very complex effect that arises through cophasal interaction of magnetic and electric dipoles of certain crystalline media. Landau and Liffschitz have proven that it can typically arise in certain antiferomagnetic materials. These media later became better known as Tellegen media.

the following time reversal transformation

$$\begin{cases} t \rightarrow -t \\ \mathbf{E} \rightarrow \mathbf{E} \\ \mathbf{B} \rightarrow -\mathbf{B} \end{cases} \tag{1.1}$$

the Maxwell's equations are invariant. In other words, out of a forward propagating solution a backward propagating solution with the same properties can be generated by the above transformation. However when a fixed (external) magnetic field is present, this transformation becomes impossible, and the time-reversal invariancy of Maxwell's equations is broken. Such a situation is easily created by incorporating magnetic materials who have a net spontaneous magnetization even in the absence of a magnetic field, i.e. ferro- and ferrimagnetic materials[8]. This way of "generating" optical non-reciprocity in a device thus takes us to the world of magneto-optics. Magneto-optics deals with phenomena arising as a result of interaction between light and matter when the latter is subjected to a magnetic field.
The deeper lying microscopical origin of all magneto-optical effects can always be traced back to the Zeeman splitting of the energy levels of the material in an external magnetic field. As a result the different optical transitions between the energy levels lose their symmetry with respect to the photon spin[9]. This causes the polarizibility of the material to acquire off-diagonal elements, because its response to left-handed ($m = -1$) and ($m = +1$) right-handed photons is different.
This description of the microscopical mechanism of magneto-optics might seem very concise, it is however the most correct description and allows via second-order time-dependent perturbation theory to rigorously calculate the MO spectral properties of any ferromagnetic material (see for example [17]). These calculations are far from straightforward, mainly because they first require the rigorous calculation of the energy bands of the crystalline material, which constitutes especially in ferromagnetic materials with their strong spin-orbit coupling a challenging task. However, the microscopical mechanism responsible for MO effects translates in a very straightforward way to a macroscopical description. It is this description with which most physicians are familiar.

1.2.2 Macroscopical description of MO materials

1.2.2.1 Constitutive relationships

The presence of an external magnetic field in a medium causes on the one hand the electrons to trace helicoidal paths under the influence of an "incoming" electrical field (via the the well-known Lorentz force), and on the other hand a similar

[8]However, it is sufficient that a fixed external magnetic field is present. In principle non-reciprocity can therefore be observed in any medium as soon as it is subjected to a magnetic field. It is intuitively clear that ferro- and ferri-magnets will have much stronger effects, because of the strong magnetic dipole moment of single atoms in these materials.

[9]A photon is a spin-1 boson with three possible spin states, however the longitudinal (m=0) photon state is not physical within a non-relativistic Maxwell theory. It is physical within a relativistic Klein-Gordon quantization of the photon field

helicoidal-like precession of the magnetic dipole moments under the influence of an "incoming" magnetic field (via the well-known Larmor effect). To put it simply, when an electromagnetic wave (with a fixed linear polarization along one of the Cartesian axes) excites a magnetized medium it will "generate" Cartesian components for the electric respectively magnetic polarization that are orthogonal to direction of the electric respectively magnetic field components of the exciting electromagnetic wave. These effects are better known as electrogyration and magnetogyration. The above schematic description makes clear that they manifest themselves as off-diagonal contributions to the dielectric permittivity and the magnetic permeability of the material. As a result the $\underline{\epsilon}$- and $\underline{\mu}$-dyadics get a full tensorial form.

Assume that the fixed magnetization lies along the $z-$axis, and that the MO material in demagnetized state is isotropic[10]. The presence of the magnetization reduces the isotropic spherical symmetry of the material to an uniaxial rotation symmetry around the $z-$axis. Bearing this in mind, it is easy to see, using some very straightforward symmetry considerations, that the permittivity and permeability tensors take on the following form

$$\underline{\epsilon} = \epsilon_0 \begin{pmatrix} \epsilon^1 & +jg & 0 \\ -jg & \epsilon^1 & 0 \\ 0 & 0 & \epsilon^0 \end{pmatrix} \qquad \underline{\mu} = \mu_0 \begin{pmatrix} \mu^1 & +jg_M & 0 \\ -jg_M & \mu^1 & 0 \\ 0 & 0 & \mu^0 \end{pmatrix} \tag{1.2}$$

Here ϵ_0 and μ_0 denote the vacuum permittivity and permeability, while ϵ^0 and μ^0 are the isotropic (i.e. in demagnetized state) values of the relative permittivity and the permeability. The MO contributions to the material dyadics are thus entirely contained in the gyro-electric and the gyromagnetic constants $g(\mathbf{M})$ and $g_M(\mathbf{M})$, and in the change of the isotropic constants to $\epsilon^1 = \epsilon^0 + b(\mathbf{M})$ and $\mu^1 = \mu^0 + b_M(\mathbf{M})$. The MO constants $g(\mathbf{M})$, $g_M(\mathbf{M})$, $b(\mathbf{M})$ and $b_M(\mathbf{M})$ obviously become zero when $\mathbf{M} \to 0$. Using Onsager's principle [18] ($\epsilon_{ij}(\mathbf{M}) = \epsilon_{ji}(-\mathbf{M})$) it is easy to prove that the gyrotropy constants, $g(\mathbf{M})$ and $g_M(\mathbf{M})$, must be odd, first-order functions of $\mathbf{M}$, and that the diagonal MO effects $b(\mathbf{M})$ and $b_M(\mathbf{M})$ must be even second-order functions of $\mathbf{M}$.[11] The first-order off-diagonal elements are typically one or two orders of magnitude lower than the isotropic constants. The second-order MO contributions to the diagonal elements are therefore very often neglected. It is also common practice to work with the so-called Voigt parameter Q (and Q_M) instead of the gyrotropy constant. This parameter is defined as

$$Q = \frac{g}{\epsilon^1} \qquad Q_M = \frac{g_M}{\mu^1} \tag{1.3}$$

The above form for the material dyadics is the only form that is invariant under a rotation around the $z-$axis. Next to that, the fact that the presence of a magnetization along the $z-$axis doesn't affect the $z-$diagonal component of the

[10] This is an assumption that is valid for most practically encountered ferro- and ferrimagnets. For instance because the deposited ferromagnetic metal is polycrystalline.

[11] Third- (or higher) order gyrotropy has never been observed, nor fourth (or higher) order effects in b or b_M.

material dyadics is also logical. An electric field along the z-axis (i.e. parallel to the magnetization) will not cause helicoidal motion of the electrons, as they do not acquire a velocity component perpendicular to the magnetization. A similar argument holds for a magnetic field along the $z-$axis. In other words, if the exciting fields are parallel to the magnetization, the material is sensed as purely isotropic and non-magnetized. For our applications it is not necessary to consider the gyromagnetic contributions to $\underline{\mu}$, for the very simple reason that at optical frequencies the Larmor effect is negligible. As a result, from a magnetic viewpoint the material behaves as if it were a vacuum[12]. Therefore it suffices to consider from here on only the $\underline{\epsilon}$ dyadic. Finally, it is worth pointing out that the inclusion of the imaginary unit on the off-diagonal positions is a choice that is inspired by energy considerations. If one wants to describe a lossless material entirely by real parameters, then the requirement of hermiticity of the material dyadics [19] can only be fulfilled by including the imaginary unit in the off-diagonal elements.

A final remark needs to be made concerning the sign convention in the macroscopical description of MO materials. In principle all material parameters are complex numbers. This introduces a first choice of sign: taking negative or positive imaginary parts as the reference for the material constants. Next to that, the necessary antisymmetry of the off-diagonal MO elements in $\underline{\epsilon}$ obliges us to make a second sign choice: using a $+$ or a $-$ sign for ϵ_{xy}.[13] In principle all four possible sign combinations are viable, at least when correctly interpreted! The problem is that this variety of possible sign schemes has led to a lot of confusion when interpreting experimental data, as it is not always clearly stated by the authors which convention is used.
If it is assumed that in principle there are no gain materials[14] then in pure isotropic optics the situation is simple: in $\epsilon = \epsilon' - j\epsilon''$ a positive value for ϵ'' is effectively a loss parameter when a time dependence $\exp(+j\omega t)$ is used, and for a time-dependence $\exp(-j\omega t)$, loss is described by a positive value for the imaginary part of $\epsilon = \epsilon' + j\epsilon''$. In magneto-optics the situation becomes much more complicated because, as opposed to traditional isotropic optics, the gyrotropy constant of a lossy material can take on a value in any of the four quadrants of the complex plane and still have a physical meaning no matter what sign convention is used for the time-dependence! Therefore, one traditionally uses the same sign convention for $g = g' \pm jg''$ as for $\epsilon = \epsilon' \pm j\epsilon''$. This leaves us with the choice for the $+$ sign or the $-$ sign for ϵ_{xy}. Atkinson [20] has given a clear overview of the different physical implications of the possible sign choices. He concluded purely on the basis of a

[12] That is: from the viewpoint of the incident electromagnetic wave! Of course the value of μ_r is still important for the pure magnetostatic problem that determines the magnetization of the material.

[13] We haven't stressed this above, but obviously once the sign of ϵ_{xy} is chosen those of the other MO off-diagonal elements (ϵ_{xz} and ϵ_{yz}) follow the antisymmetric cyclic permutation of the Levi-Cevitta tensor. This is logical in view of the axial nature of gyrotropy.

[14] which is in fact true as gain can only come about if there are external sources in the electromagnetic problem.

causality principle that the following sign scheme should be the logical one.

$$\begin{aligned} &\boxed{e^{+j\omega t}} \\ &\epsilon = \epsilon' - j\epsilon'' \\ &g = g' - jg'' \\ &\underline{\epsilon} = \epsilon_0 \begin{pmatrix} \epsilon & +jg & 0 \\ -jg & \epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix} \end{aligned} \tag{1.4}$$

Atkinson proved that, with this sign scheme, positive values of phase functions in all kinds of MO effects would indeed correspond to a "lagging behind in time" of the MO response of the material, which is in accordance with the principle of causality[15]. In any case a lossy material will have a $\epsilon'', g'' \neq 0$, but while only a $\epsilon'' > 0$ makes sense for a lossy material, g'' can be both positive or negative.

1.2.2.2 Normal modes

All MO effects arise from the gyrotropic anisotropy of the medium. As with any kind of anisotropy its most important characteristic is that the degeneracy of the two fundamental orthogonal polarization states of light is lifted. In an isotropic material for any propagation direction any set of 2 orthogonal polarizations can be chosen as base states to describe the propagation of an arbitrary polarized wave, ánd both states propagate with the same phase velocity. In an anisotropic material that is no longer the case. Depending on the propagation direction with respect to the crystal axes (or in the here considered case with respect to the magnetization), the base orthogonal polarization states are unique ánd have a different phase velocity. These base states are called the normal modes and can be found as the solution to the so-called Fresnel eigenvalue equation.[16]
This Fresnel eigenvalue equation for a plane wave solution $\mathbf{E}_0 \exp(-jk_0 \mathbf{n}_{\text{eff}} \cdot \mathbf{r})$ in a MO medium, can be found by eliminating the $\mathbf{H}-$field in the Maxwell rotor equations, and substituting the above solution.

$$\left(n_{\text{eff}}^2 \underline{I} - \mathbf{n}_{\text{eff}} \mathbf{n}_{\text{eff}} - \frac{1}{\epsilon_0} \underline{\epsilon} \right) \cdot \mathbf{E}_0 = 0, \tag{1.5}$$

with $\mathbf{n}_{\text{eff}}$ the vector describing the propagation direction having a magnitude equal to the (sought) effective index, and with $\underline{\epsilon}$ the permittivity tensor of (1.2) (with a form corresponding to the known $\mathbf{M}-$direction.) This set of equations will only have a non-trivial solution if its corresponding determinant is zero. This determinantal equation gives the value for the effective index of a plane-wave solution propagating along $\frac{\mathbf{n}_{\text{eff}}}{n_{\text{eff}}}$.

[15] Of course this doesn't imply that a MO effect will always lag behind in time. It only guarantees that obtaining a positive phase factor in any MO calculation will effectively correspond to logical cause-effect relationship.

[16] To be mathematically correct, in an isotropic medium the Fresnel eigenvalue equation leads to a degenerate eigenvalue that has both an algebraic and a geometry multiplicity of 2, i.e. equal to the dimension of the space of transversal electromagnetic waves.

Two very basic examples of the calculation of n_{eff} and the corresponding base polarization states, are those where the propagation direction is either along the magnetization or orthogonal to it. Assume $\mathbf{M}$ to be along the $z-$ direction, so that the form of (1.2) applies.

- $\mathbf{M} \parallel \mathbf{n}_{\text{eff}}$
  This is the well-known *Faraday* geometry. With $\mathbf{n}_{\text{eff}} = (0, 0, n_{\text{eff}})$, i.e. along the $z-$direction, one obtains the following equation.

$$\begin{pmatrix} n_{\text{eff}}^2 - \epsilon & -jg \\ +jg & n_{\text{eff}}^2 - \epsilon \end{pmatrix} \cdot \begin{pmatrix} E_{0,x} \\ E_{0,y} \end{pmatrix} = 0, \qquad E_{0,z} = 0, \tag{1.6}$$

  which leads straightforwardly to the solutions,

$$\text{right circular polarization}: \quad n_{\text{eff}} = \sqrt{\epsilon}\sqrt{1-Q} \quad \mathbf{E}_0 = \begin{pmatrix} 1 \\ +j \\ 0 \end{pmatrix} \tag{1.7}$$

$$\text{left circular polarization}: \quad n_{\text{eff}} = \sqrt{\epsilon}\sqrt{1+Q} \quad \mathbf{E}_0 = \begin{pmatrix} 1 \\ -j \\ 0 \end{pmatrix} \tag{1.8}$$

  Therefore, for plane wave propagation in the $z-$direction the orthogonal base polarization states are the left- and right-circular polarization states. These propagate with different phase velocities, and these velocities are interchanged when the propagation direction is inverted (which is equivalent to $Q \to -Q$)[17], indicating the nonreciprocity of the Faraday effect.

- $\mathbf{M} \perp \mathbf{n}_{\text{eff}}$
  This is the *Voigt* or *Cotton-Mouton* geometry. In this case, with $\mathbf{n}_{\text{eff}} = (n_{\text{eff}}, 0, 0)$, i.e. along the $x-$direction, the eigenvalue equation yields

$$\text{TE polarization}: \quad n_{\text{eff}} = \sqrt{\epsilon} \qquad \mathbf{E}_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{1.9}$$

$$\text{TM polarization}: \quad n_{\text{eff}} = \sqrt{\epsilon}\sqrt{1-Q^2} \quad \mathbf{E}_0 = \begin{pmatrix} -jQ \\ 1 \\ 0 \end{pmatrix} \tag{1.10}$$

  The base polarization states in a Voigt transmission configuration are thus linear, with one state polarized along the magnetization direction and the other perpendicular to it. In MO it is common practice to label the former "TE" and the latter "TM". Also notice that the TM polarization isn't linear in the traditional sense of the word: it has a small first-order Q component

[17] However, remember that when backward propagation direction is considered one has to make a clear distinction between description in the fixed lab frame and the intrinsic frame associated with the wave itself. In the fixed frame a backward propagating $\left(\begin{smallmatrix} 1 \\ +j \end{smallmatrix}\right)$ is actually a left-circularly polarized wave! Therefore when discussing calculations of MO effects it is always wise to clearly state in what reference frame the results are given.

along the propagation direction. This seems to contradict the fact that an EM wave always has to be transversal. However that is of course only correct in isotropic media. Looking at the second Maxwell rotor equation, it is clear that only $\mathbf{D}$ will always be orthogonal to $\mathbf{n}_{\text{eff}}$. It is not difficult to see that with a $\mathbf{M} \perp \mathbf{n}_{\text{eff}}$, $\mathbf{E} = \underline{\epsilon}^{-1} \cdot \mathbf{D}$ can get components along $\mathbf{n}_{\text{eff}}$. In a Voigt transmission experiment the MO effect is of second order in Q, hence it is not only much smaller than the Faraday effect, but also of a purely reciprocal nature! Again, this might seem contra-intuitive, as it was clearly proven at the beginning of this section that the presence of MO media always lead to non-reciprocity. However, there is a first order odd magnetization effect present, namely in the polarization state. This will cause the boundary conditions at the interface between a MO medium (in Voigt geometry) and any other medium (MO or not), to become non-reciprocal. It is well known that it is precisely these boundary conditions that form the basis for the derivation of interface reflection and transmission effects. Therefore, while the Voigt geometry will not give rise to non-reciprocal transmission behaviour, there will be an important non-reciprocal reflection effect. It is also clear that this effect will only be felt by the TM-component of the incident light.

When the propagation direction of the wave has a more general direction with respect to $\mathbf{M}$, the normal modes will be elliptically polarized and the non-reciprocity will be present both in the polarization base states and the phase velocities.

1.2.3 Phenomenology of magneto-optics

MO effects are generally divided into transmission effects and reflection effects. The latter are better known as the magneto-optic Kerr effects.

Transmission effects The best known MO transmission effects are the *Faraday* effect and the *Cotton-Mouton* effect. These have already been touched upon above, when discussing the derivation of the normal modes in a MO medium. It is not our intention to repeat here the derivation of the formulae describing these phenomena. That can be found in any basic textbook on electromagnetism or magneto-optics (see for instance [21]). However, in view of the importance of these effects with respect to a proper understanding of how optical isolators work, we would just like to describe in a qualitative manner how they manifest themselves.
The understanding is always based on the decomposition into the orthognal base states. For example for the Faraday effect, the base states have the Jones vectors $\left(\begin{smallmatrix}1\\-j\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1\\+j\end{smallmatrix}\right)$ (i.e. the left and right circular polarizations). Therefore, a linearly polarized wave (for instance a $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$) will excite both waves to the same amount and in phase. The above treatment has shown that these two states propagate at different phase velocities. Therefore after a distance z, the $x-$polarized Jones vector has evolved as (omitting the normalization constant and the common phase

factor and McLaurin expanding the phase velocities up to first order in Q)

$$\begin{aligned}\begin{pmatrix}E_x\\E_y\end{pmatrix}(z) &\sim \begin{pmatrix}1\\j\end{pmatrix}\exp(jk_0\frac{\sqrt{\epsilon}Q}{2}z)+\begin{pmatrix}1\\-j\end{pmatrix}\exp(-jk_0\frac{\sqrt{\epsilon}Q}{2}z)\\ &= 2\begin{pmatrix}\cos(k_0\frac{\sqrt{\epsilon}Q}{2}z)\\-\sin(k_0\frac{\sqrt{\epsilon}Q}{2}z)\end{pmatrix}\end{aligned} \tag{1.11}$$

Thus the (complex) polarization variable[18] $\chi=\frac{E_y}{E_x}$ is purely real if the MO medium is lossless (i.e. ϵ and Q (or g) real numbers.), and the polarization state thus remains purely linear but is rotated over an angle

$$\theta=-k_0\frac{\sqrt{\epsilon}Q}{2}z \tag{1.12}$$

This is the basic formula for the Faraday effect, in its most simple version. It clearly shows how a linear polarization when propagating along the magnetization of a lossless MO medium will gradually rotate. This effect is also known as magnetic circular birefringence. If the propagation direction is reversed (or equivalently if $\mathbf{M}$ is switched) this rotation will change sign, hence it is non-reciprocal. This is the basic principle of a free space Faraday optical isolator.
When the MO medium is lossy, the left and right circularly polarized base states will not only be phase shifted with respect to eachother, but will also experience different loss, thus their relative strength will change. In other words the medium has magnetic circular dichroism. As a result the incoming linear polarization is not only rotated but also acquires a degree of ellipticity.

The Cotton-Mouton effect can be treated in the same manner. In fact the results will be quasi the same, except for the fact that they are much weaker (because the effect is only of second order), and more importantly, that due to the second-order character, the rotation and ellipticity are reciprocal!

Reflection effects The magneto-optic Kerr reflection effects are classified on the basis of the relative orientation of the magnetization with respect to the plane of incidence. The *transverse* or *equatorial* Kerr effect occurs when the magnetization is orthogonal to this plane. As the incidence plane is by definition orthogonal to the interfacial plane, it follows that in this case the magnetization must lie within the plane of the sample. When the magnetization lies within the plane of incidence, there are two further configurations possible. Either the magnetization is perpendicular to the interfacial plane or it lies within this plane. The former configuration is known as the *polar* Kerr effect, while the latter is labelled as the *longitudinal* or *meridional* Kerr effect. These three types of Kerr configurations are illustrated in Fig. 1.1.

[18] This is the ratio of the components of the Jones vector onto two reference orthogonal base polarization states. Normally one uses the linear Cartesian states x and y of the lab frame, but in reflection one often makes use of the intrinsic s- and p-polarizations. The basics of Jones calculus for the description of effects with polarized light can be found in [22]. We would just like to mention that when χ is known, the angle θ that the polarization makes with the "$x-$"state is found as $\tan(2\theta)=2\Re(\chi)/(1-|\chi|^2)$, while the ellipticity ψ is given by $\sin(2\psi)=2\Im(\chi)/(1+|\chi|^2)$. The proof of this can also be found in the standard work of Azzam [22].

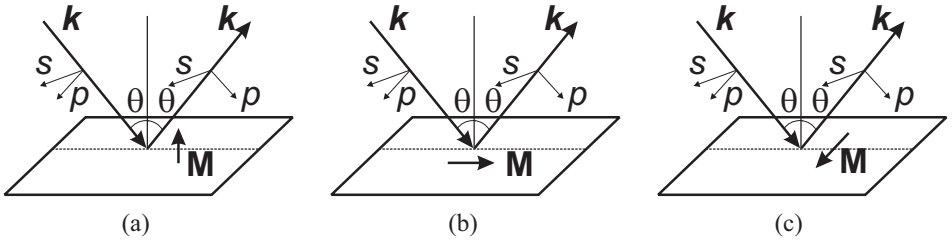


Fig. 1.1: Three types of Kerr effect. (a): polar effect; (b): longitudinal effect; (c): transverse effect. The indicated (s, p) frames correspond to the intrinsic reference frame in which the reflection coefficients are usually defined. It should also be noted that the indicated senses for the magnetization correspond to the references for the different effects. In other words if the magnetization is inverted with respect to what is depicted here, then g (or Q) in the formula has to be replaced by $-g$ (or $-Q$), under the conditions of course that the g's (or Q's) are measured in the same conditions as depicted here ánd that they had their values extracted using the same sign conventions as the ones that one will be using to do MO simulations. All of this serves to illustrate how easily confusion (more in particular sign errors) can arise when one is doing magneto-optics.

The basic principle behind the MO Kerr effects is the fact that in the MO layer the eigenpolarization states have a hybrid character: they are a mix of $s-$ and $p-$waves. Calculations as the ones in the paragraph on "normal modes" in section 1.2.3, have shown that the relative strength of this $s-$ and $p-$character depends on the incidence angle and on the direction of the magnetization (i.e. on Q). It is a well known property of polarimetry [22] that only if the incident polarization state is also an eigenpolarization of the "exit" medium, this polarization will be retained after reflection and transmission. For each of these two eigenpolarizations its reflection (or transmission) can thus be described by its characteristic (complex) Fresnel-like reflection (and transmission) coefficient. The reflection of all other incident polarizations needs to be treated by first decomposing them into the base eigenpolarization states of the system. And because each eigenstate has its unique reflection coefficient, the phase (and amplitude) relationship that existed upon incidence is not retained. As a result the incident polarization changes its state, that is: it acquires additional rotation and/or elipticity. Concretely this means that for instance a $s-$polarized wave will acquire a $p-$component upon reflection at the surface between an isotropic and a MO medium, with in general a phase and amplitude difference towards to the reflected $s-$wave. And the reflected wave acquires both a rotation and an ellipticity due to the MO effect.
The MO Kerr reflection effects thus manifest themselves by the need to describe the reflection (and transmission) phenomena by a full 2×2 Jones matrix in the intrinsic $s - p$ frame. It is intuitively clear that this MOKE rotation and ellipticity is a first order function of Q, because the eigenpolarization states of the MO layer are of this order in Q, and thus changes sign upon magnetization reversal (or propagation inversion). Kerr reflection effects are thus unique mechanisms to induce non-reciprocal coupling phenomena between TE and TM modes in MO

waveguides. All of the above specifically applied to those Kerr configurations where the $s-$ and $p-$polarizations are no longer eigenpolarizations of the MO medium. However, we have seen above that in the special case of Voigt geometry, the eigenpolarizations are still described by $s-$ and $p-$waves! Therefore, there will be no Kerr rotation or ellipticity in the transverse Kerr effect, since the Voigt geometry doesn't mix up the $s-$ and $p-$waves. Actually there is no effect at all for a $s-$wave (except of course a standard isotropic reflection and transmission), as it doesn't even feel the presence of the magnetization[19]. For the p-wave on the other hand we have seen above that the electric field acquires a small non-reciprocal first order component parallel to its wavevector. It is precisely this component that will cause the peculiar $p-$reflection phenomenon in a transverse Kerr configuration. If this component was not present it is easy to see that the reflection coefficient of the $p-waves$ wouldn't differ from the isotropic $p-$Fresnel coefficient. The presence of this component causes a non-reciprocal MO contribution to the boundary conditions used to derive the Fresnel reflection formulae. As a result in a transverse Kerr configuration there is no coupling between the $s-$ and the $p-$waves, but the $p-$waves acquire an extra direction dependent phase shift upon reflection off a lossless transversely magnetized MO material. If on top this the MO material becomes lossy, the loss of the wave upon reflection becomes direction-dependent also. It is clear that this effect will have an important impact on the dispersion equation of a waveguide that contains MO layers in a transverse configuration. It will make this equation and its solutions (i.e. the location of its eigenmodes) direction-dependent for the TM modes. It is precisely this effect that plays a central role in this work.

All of the above qualitative assessments can be quantified in analytical expressions describing the s- and p- MOKE rotations and ellipticities. Again it is not our intention to derive the formulae describing these MO reflection. There is a plentitude of papers describing the theory of magneto-optic Kerr reflection, especially because it plays such a central role in magneto-optical recording. There is a whole arsenal of numerical algorithms, both rigorous and approximative, computing efficiency improving methods, simplifying theories on equivalent media, attempts at exact analytical solutions, ... all developed to model magneto-optic reflection and transmission effects of polarized light incident on multilayer stacks containing a diversity of isotropic, anisotropic and gyrotropic media (with arbitrary magnetization directions). Nice overviews can be found in [23–25]. The main intention of this section was to make the reader familiar with the basic concepts and phenomenology of magneto-optics, and provide sufficient background to understand the principles of existing isolator designs. The primary focus of this work is after all integrated isolator design, and not a primer in magneto-optics. A subsequent chapter will delve more deeply in all the theoretical subtleties and aspects of non-reciprocal waveguides, and specifically those in a transverse Kerr configuration.

[19] Because its electric field is parallel to the magnetization, there can be no gyration.

1.2.4 Magneto-optic materials

Before illustrating all of this it is worthwhile to spend just a few words on what are frequently used MO materials. After all, it is for a large part material problems that explain why an integrated isolator has up till now not seen the break of day.

Basically, as stated at the start of section 1.2.1, any material that can sustain a spontaneous magnetization, will have some magneto-optical response. In fact even a paramagnetic or a diamagnetic sample will give rise to MO effects in the presence of an external magnetic field. Of course there is a relationship between the magnitude of the magneto-optic effect and the magnetization inside the material, but it is not necessarily a linear one. The magneto-optical strength always relates back to the differences in strength of the electrical dipole transitions between the split magnetic levels. This splitting is not necessarily linear in the magnetization, at least not for the ferromagnetic and ferrimagnetic materials, where there are also important contributions from the crystal field and the Heisenberg exchange Hamiltonian term. On top of that in metallic systems there can be large contributions from *intra*-band effects at energies lower than 2eV. These effects depend significantly on the position of the Fermi level and the precise structure of the conduction band near the Fermi level (especially its splitting under the influence of a magnetic field). It is known that a metallic band structure and its calculation is considerably more complex than dielectric crystal spectra or semiconductor band structure calculations, precisely because of the importance and the "multitude" of the electron-electron correlations in metals.

In any case, there are two types of materials that are most often encountered in MO applications: ferromagnetic transition metals (and their alloys) and ferrimagnetic garnets. Recently, a novel class of materials has been added to this list, namely artificially fabricated magnetic composite materials, where magnetic metallic nanoclusters, grains, or microparticles are "created" in a nonmagnetic host material. The interest in these has recently augmented after it was suggested and partly confirmed that the MO effects of these diluted hybrid materials could be remain unchanged (or get even enhanced) with respect to continuous metallic films, while obviously their optical absorption is seriously reduced. We will summarize a few basic properties of these two types of MO materials, keeping an eye on the spectral region of interest for this work (0.75eV-1.1eV).

1.2.4.1 Transition metals and alloys

Ferromagnetic transition metals (especially Co, Fe, Ni, Mn, Cr), their alloys and their compounds with other elements (magnetic or nonmagnetic) have been studied for ages. Mainly because of their relatively easy accessibility and "fabrication", because of their relatively easy to understand and "manipulate" magnetic behaviour, and because the possibility to form alloys and binary or ternary compounds allows for a wide range of tuning parameters. All of this made them historically the first materials to be used in thin film magneto-optical recording. Most of these materials can be rather straightforwardly deposited on any kind of substrate with controlled thicknesses down to the monolayer regime. Their MO properties and optical properties can however vary greatly from metal to metal. Roughly speak-

ing their polar Kerr rotation is of the order of 1 degree (at the blue side of the visible spectrum), which is rather strong. Their optical absorption can easily vary by a factor 2–3 but is in general in the order of magnitude of 10^5cm^{-1}. There hasn't been a lot of characterization work in the infrared part of the spectrum, for the obvious reason that data storage capacity scales inversely with wavelength. In a later chapter we will discuss our efforts in characterizing the optical and magneto-optical properties of a certain class of such a ferromagnetic alloy. When it comes to choosing the best MO metal there is a variety of criteria depending on the precise application (polar Kerr rotation, reflectivity, uniaxial perpendicular or in-plane magnetic anisotropy, Curie temperature, ...). From a pure magneto-optic strength viewpoint Fe, Co and rare-earth/Fe-Co compounds seem to give the best results.

The general observation is that MO ferromagnetic metal alloys are the ideal MO materials for purposes of magneto-optical recording (i.e. needed to operate in a reflection regime), where they can hardly be beaten in MO response by other materials, however for integrated waveguide applications their optical absorption is that high that even their high Kerr effects can't seem to compensate this.

... but the results presented in this work are maybe opening up a new pathway and application for ferromagnetic alloys in integrated devices ...

1.2.4.2 Ferrimagnetic garnets

These are cubic crystals with a very complex spinel-like structure. Each site is occupied by a molecule with formula $\{c^{3+}\}[a^{3+}](d^{3+})\text{O}_{12}$. There are three types of sites for cations. Site c is dodecahedrally surrounded by oxygen atoms and is occupied by a rare-earth or by Y^{3+}, or La^{3+}. The octahedral (a) and tetrahedral (d) sites are occupied by iron. The octahedral and tetrahedral sites couple antiferromagnetically but are not equivalent. As a result the garnet becomes ferrimagnetic. The net magnetization can be controlled by controlled substitution of Fe atoms on the tetrahedral sites or the octohedral sites (depending on the type of substitutive elements used). By this substitution one can tune the net magnetization of the garnet crystal, by causing the magnetization of one or the other sublattice to become dominant. At the crossing point one achieves antiferromagnetical compensation. As a result this also allows to change the sign of the Faraday rotation.

The interest in ferrimagnetic garnets comes from their unique combination of low optical losses and strong Faraday rotation. In fact they are the only type of magneto-optical materials that are quasi-transparent in the spectral region below 1.1eV. Typical optical absorption levels are below 50cm^{-1}, while the observed specific Faraday rotation (see (1.12)) can be anywhere between 50–300 deg/cm (around $h\nu = 1.1\text{eV}$), mostly depending on the type of elements that are used in the dodecahedral sites. Bi and Ce are known to enhance the Faraday rotation dramatically (up to 1000 deg/cm), but also increase the optical absorption. One cannot simply compare values of specific Faraday rotation to polar Kerr rotation angles (which are used to classify the strength of the ferromagnetic metals), because they are entirely different MO effects. However, the gyrotropy constant g

can be calculated using standard formulae for the polar Kerr rotation and specific Faraday rotation, out of these experimental MO quantities. While obviously g will be complex for metals and quasi-real for garnets they can be compared on the basis of their magnitude. In general the absolute value of the gyrotropy constant of garnets will be typically up to two orders of magnitude lower than that of metals.

There has been a long time expertise on all the structural, magnetic and magneto-optic properties of garnets. Their low optical absorption have made them the materials par excellence for bulk optical isolators in optical telecom, but has also spurred further interest in their use as material system for integrated magneto-optical circuits. They have a negligible absorption in the infra-red and the mechanism of element substitution on the different lattice sites allows for a refractive index variation of the same order of magnitude as in InP material system, and also allows for a limited tuning of the MO and magnetic properties. The most important drawback is that thin film garnet devices can only be grown by the non-standard method of Liquid Phase Epitaxy (LPE)[20]. Secondly, the lattice parameters of the spinel unit cell can vary significantly upon element substitution, prohibiting in a way a stress-free growth of thin-film devices. This has serious consequences for the birefringence of the fabricated devices. Thirdly, as opposed to ferromagnetic materials, where the magnetism is typically caused by a very strong direct exchange coupling, ferrimagnetism is mainly caused by the less strong superexchange mechanism and is as a result much more temperature sensitive (sometimes the Néel temperature can be as low as 35°C). This immediately translates to a rather important temperature dependence of the Faraday rotation. Over a range of just 50K around 300K, the specific Faraday rotation of yttrium iron garnet (YIG) has been observed to vary by more than 50%. Finally, the magnetic anisotropy of garnets is typically uniaxial perpendicular instead of in-plane. In plane the magnetic behaviour will be mostly hard-axis-like. And because the most interesting nonreciprocal effects require either longitudinal or transverse magnetization, it is almost inevitable to use an external magnet in garnet based integrated magneto-optical devices to keep the magnetization at the saturation level.

In general, one can state that due to its unique properties the garnet material system is entirely responsible for the "genesis" of the research area of integrated magneto-optics, and ever since entirely monopolizes this research. This has been a self-feeding process. As integrated devices developed, growth techniques improved and engineering of the magnetic, magneto-optic, optical, ... properties through element substitution were invented. At present it is safe to say that ferrimagnetic garnet technology (growth, processing) has reached a maturity comparable to standard semiconductor material platforms for optical applications. However, in the course of this development (started in the mid-70's) engineers and physicians have been pre-occupied with device and material improvement. Only very recently has one come to realize the importance of integration techniques enabling to incorporate the planar garnet waveguide designs with standard optical III-V platforms. This seems to be the next big challenge in integrated garnet magneto-optics. There is obviously no monolithic integration thinkable, and heterogeneous

[20] bulk garnet crystals can be fabricated using for instance a standard Czochralski process.

integration techniques are for the moment progressing really slow. More on this will be presented in section 1.3.2, where the realized integrated devices are reviewed.

1.3 Optical isolators

Now that the basic phenomenology and the most important material systems of magneto-optics have been reviewed, we are in a position to understand how optical isolators operate, how one can devise planar integrated waveguide isolators and what typical challenges they are facing.

1.3.1 Free space isolators – typical specifications

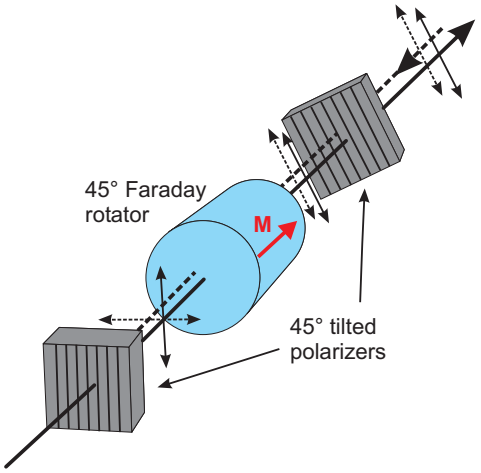


Fig. 1.2: Basic principle of a polarization-dependent free space optical isolator.

The basic principle of a free-space bulk isolator (see Fig. 1.2) is the nonreciprocal Faraday rotation transmission effect discussed in section 1.2.3. In a basic version there are three optical elements: two optical polarizers and a nonreciprocal bulk longitudinally magnetized garnet crystal. By precisely cutting the length of the garnet crystal a rotation of the plane of polarization of 45˚ can be obtained. If the polariser at the output is tilted over 45˚ with respect to the input polarizer, then all light is transmitted. Light that enters the device in the backward direction will due to the non-recprocity of the Faraday rotation again be rotated over +45˚ (in the fixed labframe) or over -45˚ in the intrinsic frame (in which the formula (1.12) was derived), thus arriving at the input polarizer with an orthogonal polarization, and ideally being completely blocked. This optical isolator is obviously polarization-dependent. If the incoming polarization is not known beforehand, an important fraction of light can get lost at the input polarizer. Therefore this simple

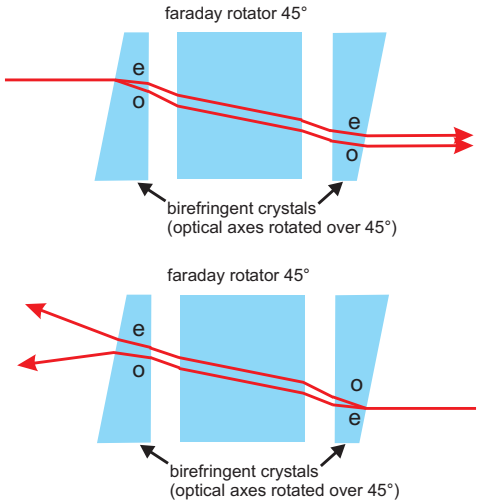


Fig. 1.3: Polarization-independent free space optical isolator using birefringent crystals.

design is most often only used in laser packages. For a polarization-independent protection of active components, one can use different schemes. One can either seperate the TE and TM polarizations by means of a beam splitter and treat them separately. A more elegant way is to replace the polarizers by birefringent crystals. This is illustrated in Fig. 1.3. The optical axes of the birefringent crystals make an angle of 45˚ with eachother, while again the nonreciprocal Faraday rotator induces a 45˚ rotation of the plane of polarization. As a result the ordinary (o) and extra-ordinary (e) components of the incoming light retain their character with respect to the second crystal after passing through the Faraday element. In the backward direction however the Faraday element again rotates the polarization planes over 45˚ (with respect to the fixed labframe), and the o- (resp. an e-) components of the backward travelling light enter the birefringent crystal at the input side as an e- (resp. an o-) wave. In other words in the backward direction a spatial walk-off is created between the two polarization components of the beam.

Free-space bulk optical isolators have been commercially available for over 20 years. Advances in garnet material technology have made it possible to achieve considerable specific Faraday rotation for a very large wavelength range between 250nm and 1700nm. Furthermore, one has nowadays even achieved to control the magnetic anisotropy of garnet bulk crystals by special techniques, so that bias-magnet-free free-space optical isolators are available [26]. An important improvement considering the space that the magnet usually takes up in an isolator package. Table 1.1 summarizes typical specifications of commercially available free-space optical isolators.
These are specifications for a polarization-independent single stage isolator. If specific ultra high protection is needed, narrowband dual-stage isolators are available with isolation ratios up to 60dB. Polarization-dependent isolators also achieve

λ_c	nm	1300, 1500–1600
λ-bandwidth	nm	20
insertion loss	dB	≤ 0.5
peak isolation	dB	42
minimum isolation (at $\pm$20nm)	dB	30
return loss	dB	60
PDL	dB	0.1
PMD	ps	0.2
maximum power density	W/cm^2	1–2
operating temperature	˚C	-10 – +65

TABLE. 1.1: Typical specifications of commercially available free-space optical isolators.

higher isolations (up to 50dB in a single stage). A wavelength bandwidth of around 20nm is typical. Much higher bandwidths are difficult to achieve due to material issues. PDL and PMD stand for respectively polarization dependent loss and polarization mode dispersion. The return loss describes how strong the reflection at the entrance facet (in the forward direction) is suppressed. This is really an issue for an optical isolator, as it can obviously not "isolate" this type of reflections. In a free-space isolator it is not sufficient to AR coat the entrance window (just before the polarizer), as reflections can also occur at the front facet of the Faraday crystal. The package size depends on whether the isolator is equipped with fibre connections, or with high aperture lenses. Typically, a free-space isolator package measures about 2×4cm.

1.3.2 Integrated optical waveguide isolators

Attempts at an integrated optical isolator are almost as old as integrated optics itself. The very first paper on a possible version of an integrated magneto-optical waveguide which could be used as an isolator dates back to 1972 [27]. Since then a multitude of papers have been published on integrated planar MO isolator waveguides. It is not our intention to give a detailed historical overview of this work. This can be found in a series of recently published invited papers [28, 30, 69]. We only want to trace the evolution of the attempts at a planar waveguide version of an optical isolator by highlighting the results obtained and the challenges encountered by the different research groups in this historical quest.

1.3.2.1 Non-reciprocal TE–TM mode convertors

The very first versions of a waveguide optical isolator were in a very logical manner entirely based on the attempt at fabricating a waveguide analogon of a 45˚ non-reciprocal bulk Faraday rotator. It is immediately clear that such a strategy will face some very fundamental theoretical obstacles which are a direct consequence

of the decrease of symmetry when going from a bulk garnet crystal to a layered garnet waveguide.

1. In a waveguide as opposed to a continuous medium not every polarization direction is a possible eigenpolarization. The guided modes come in two possible fixed polarization states, either hybrid or TE or TM. As a result there are no 45 ° separated polarization states, but only orthogonal polarization states! Therefore, the simple bulk isolator described in the previous section will not carry over to a waveguide version, because there the input and output polarizations are at 45° and not at 0° or 90°. In a waveguide version of a Faraday rotator the input and output polarizations *must* be in either TE or TM mode if one aims at a truly integrated component – that is, one that can communicate with other integrated devices (which also operate in either TE or TM).

2. Next to that, the magneto-optically coupled waveguide modes are not phase-matched. They propagate with different phase velocities as opposed to the situation in a bulk medium. This is not only caused by the fundamental waveguide geometric birefringence but also to growth-induced birefringence and stress-birefringence. It is well-known from coupled-mode theory that when two coupled modes are not phase-matched, it is impossible to achieve a total power transfer. On top of that if the phase-mismatched coupling modes have different polarizations, as is the case in a longitudinal MO waveguide, the mode coupling will not only lead to the desired polarization rotation but also to undesired polarization ellipticity. This latter phenomenon is quite annoying, because, while it could be perfectly possible to rotate the (major axis of the) polarization over 45° this might get accompanied by a serious degree of elipticity, which will still leak through the input polarizer!

The first problem can only be solved by adding a second optically active *reciprocal* element "in series" with the *non-reciprocal* optical active element (Faraday rotator). This can be understood by considering the bulk Faraday isolator of Fig 1.2. Adding a -45° reciprocal active rotator will bring the 45° rotated polarization back to 0° thus cancelling out the Faraday rotation, while in the backward direction the reciprocal rotator will cause a rotation that viewed in the fixed labframe will change sign and thus together with the non-reciprocal Faraday rotator (which always causes a +45° in the labframe independent of the propagation direction) adds up to a rotation of 90°. In this way an isolator is created with as input and output polarization a TM-like mode (or a TE-like mode). This is illustrated in Fig. 1.4 for TM mode in- and output. An isolator with TM as the input and TE as the output can be made by using a +45° reciprocal rotator in combination with a +45° Faraday rotator.

One can think of several ways to introduce this necessary reciprocal rotation in a planar waveguide. The easiest way is to use an anisotropic material (e.g. $LiNbO_3$) as a top cladding layer, with an optical axis in the plane of the film and rotated with respect to the TE direction[21]. This is historically the first attempt at

[21]The nice thing about using an anisotropic cladding layer is that not only provides reciprocal mode coupling, but can also be advantageously used to eliminate the shape geometry birefrin-

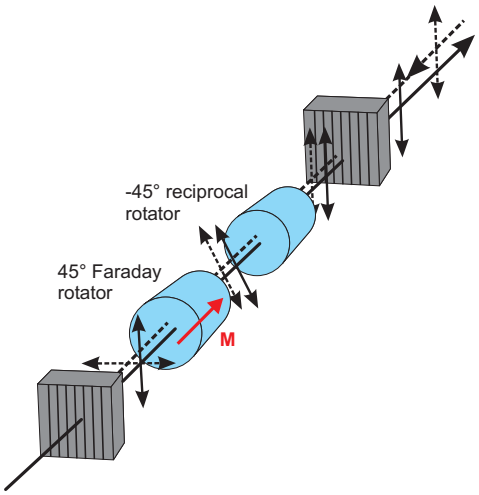


Fig. 1.4: Avoiding the need for 45° separated polarization at the in- and output of an optical isolator.

an integrated Faraday rotator [31, 32] (see Fig. 1.5). These attempts never worked

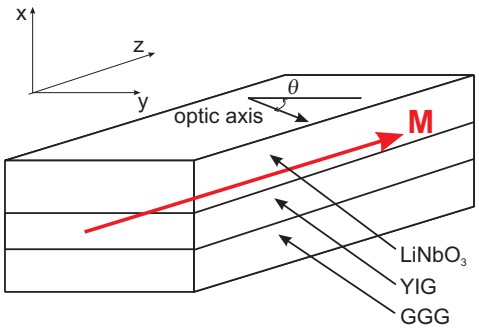


Fig. 1.5: An integrated nonreciprocal TE-TM convertor. By carefully tuning the direction of the optical axis of the anisotropic $LiNbO_3$ cladding layer on top of the longitudinal MO garnet YIG core, the TE-TM coupling can be made to vanish in the forward direction while attaining 100% in the backward direction, or vice versa. If on top of that due to the anisotropy of the cladding layer the TM modes are leaky, one has achieved an isolator that doesn't need a polarizer at the input (after [32]).

for the very simple reasons that 1. the required accuracy on the direction of the optical axis is extremely high, and 2. it is far from straightforward to realize a good optical contact of an anisotropic film on a magneto-optic garnet layer.

---

gence.

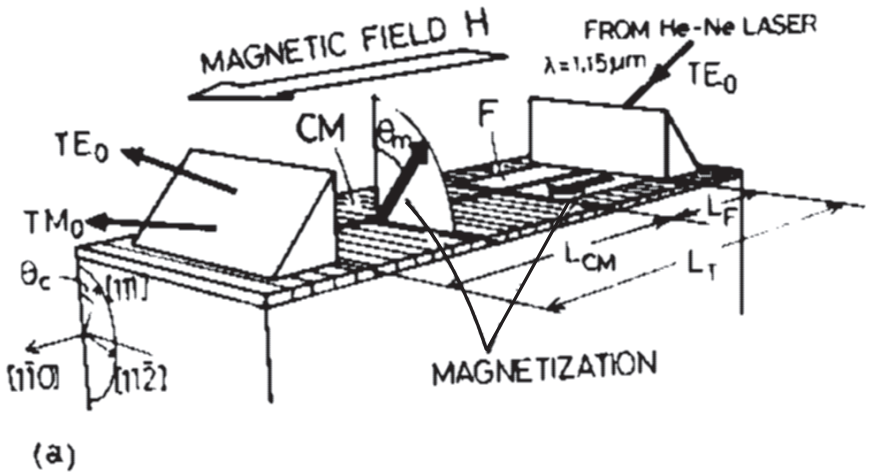


Fig. 1.6: The first successful non-reciprocal TE-TM convertor by Ando. The Voigt (or Cotton-Mouton) region is naturally formed in the Bi:YIG film, as a result of the uniaxial magnetic anisotropy of the bubble domains which form naturally upon growth of YIG. It is tuned by tilting the face normal of the Gallium Gadolinium Garnet (GGG) substrate. Note also that it doesn't matter that Bloch domains are formed, as the Voigt effect is quadratic in the magnetization. By annealing a square region with a pulsed laser source the uniaxial anisotropy could be quasi-eliminated and the magnetization switched to an in-plane direction (i.e. Faraday geometry).

A more elegant way to induce the reciprocal rotation is by using the Voigt geometry physically in series with a Faraday region, by locally changing the direction of the magnetization. It was discussed in section 1.2.2.2 how the Voigt geometry induces a reciprocal MO effect, if the magnetization is tilted in the plane perpendicular to the propagation direction of the light a reciprocal coupling between the TE and TM modes can be achieved (see Fig. 1.6). Ando [33] was the first to achieve this. In fact this was the very first successful fabrication of an integrated optical isolator.
With this design, Ando observed an isolation ratio of up to 13dB with an insertion loss of only 1.8dB. This was an amazing achievement, especially in view of the fact that he did not make any effort to reduce the linear TE/TM birefringence[22]. This also explains why "only" 13 dB was observed. The birefringence limited the TE/TM conversion efficiency. Another slight drawback in Ando's seminal experiment was the need for a rather strong external magnetic field (up to 250Oe) to keep the Faraday region properly magnetized.

[22] Although it has be said that the laser annealing to produce a Faraday region is automatically also a heat treatment which reduces considerably the stress-induced birefringence.

After the success of Ando, most researchers looked into a further optimization of the nonreciprocal TE/TM convertor in different ways. The main concern is the elimination of the linear birefringence, and the development of miniaturized permanent film magnets (as it was quickly clear that the garnet films couldn't do without an external magnetic field). The elimination of the fundamental shape birefringence, which is an unavoidable consequence of the different boundary conditions for the TE and the TM modes at the transversal and lateral guiding interfaces, can in theory be achieved by a meticulous design of the waveguide geometry. It is known for example that in perfectly square symmetric waveguides (for instance a perfect optical fibre without tension or bends), both polarizations are degenerate, as there is obviously no difference between the lateral (TE) and transversal (TM) direction. Such symmetric profiles are however very difficult to fabricate in practice. For raised strip waveguides, in principle the height and the width of the raised profile can be tuned to obtain this same mode degeneracy. However, the tolerances on both the width and the height are quasi-impossible to achieve [34]. Therefore one most often resorted to a post-fabrication tuning technique, that consists of the application of a suitable stress [35], a heat treatment [33], a post-growth etch tuning, or a combination of the latter two [36, 37]. The group of Levy [38, 39] concentrated additional efforts on designing an appropriate thin-film magnet capable of saturating garnet waveguides of several mm's length and depositing it without introducing extra stress in the garnet films. All of these efforts led to integrated Faraday-type optical isolators capable of delivering isolation ratios up to 30dB, with insertion losses not exceeding 2dB.

While all of this work done between the end of the 70's till the early 90's, gave excellent results with regards to isolation ratio, it remained a very daunting task to fabricate a waveguide-type Faraday isolator precisely because of the very stringent phase-matching requirements. On top of that, while one might be able to devise a "perfect" integrated Faraday isolator, it will not work without polarizers at the output and/or the input facets. Therefore, at the beginning of the 90's researchers started focussing their attention on other integrated isolator concepts.

1.3.2.2 Non-reciprocal TM (and TE) phase shifters

One of these other isolator concepts had been suggested right at the start of the quest for an integrated isolator, but was for some reason or another never really considered [40] or sadly enough went unnoticed (because of the unexciting publisher's source ?!). Auracher's idea was very elegant, and yet very simple (which goes alongside with elegance almost always). His idea was to use, instead of the obvious Faraday TE/TM coupling, the nonreciprocal transverse Kerr effect, present when a MO waveguide is magnetized in plane and transversal to the light propagation direction. It was in rough lines explained above how this effect will not cause a coupling of the TE and TM modes, but will on the other hand change the dispersion equation for the TM waveguide modes in a nonreciprocal manner. As a result the TM waveguide modes will have different propagation constants for forward and backward directions. For this reason it is also named the transverse nonreciprocal phase shift. As opposed to the Faraday effect this transverse MO

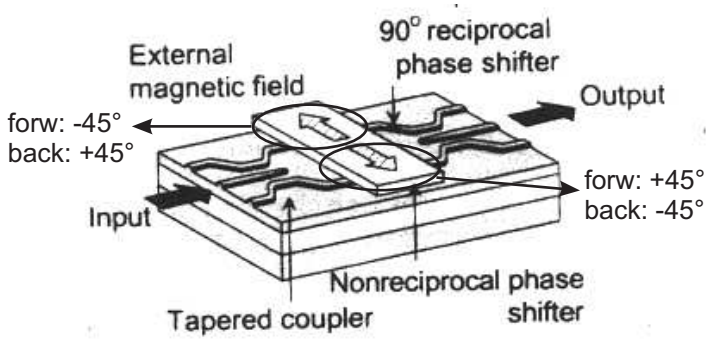


Fig. 1.7: A non-reciprocal TM Mach-Zehnder interferometer using the transverse Kerr phase shift. By placing a well-designed magnetic assembly on top of the two arms of the interferometer as depicted here, a non-reciprocal phase difference of $\pm\pi/2$ between both arms can be obtained (i.e. inducing a non-reciprocal phase shift of $\pm\pi/4$ in each arm.). Inserting an extra reciprocal path length difference corresponding to $\lambda/2 \pm m\lambda$, it is clear that the device will be constructively interfering in the forward direction and destructively in the backward. By a well-designed magnetic assembly it is meant that it induces opposite magnetic transverse fields in both arms and that it magnetizes the garnet ridges over a sufficiently long region so as to obtain a $\pi/4$ phase shift. This design with a tapered coupler for the interfering arms is also more tolerant because it can be optimized to have quasi destructive interference for TE modes (which are not subjected to the transverse Kerr shift), without disrupting the TM interference too much. In this way, it can be avoided that back-reflected TE light enters the device (after [44]).

effect is not a transmission effect[23]. It arises entirely because of the non-reciprocal reflection coupling of the longitudinal and transversal TM modal field components among themselves at the interfaces between regions with different gyrotropy.

Different isolation schemes have been proposed and implemented using this nonreciprocal phase shift. Some rely on the coupling to an evanescent wave in the backward direction [41][24], and others on constructive interference in the forward direction and destructive interference in the backward direction [42–45]. For example, proper arm lengths in a Mach-Zehnder interferometer (MZI), such as in Fig. 1.7, can lead to $\pm 90^\circ$ phase shifts, with a sign depending on the propagation direction. Adding a reciprocal 90° phase shift, achieved by a $\lambda/4$ ($\pm m\lambda$) length difference results in a total phase difference of either 0° or 180° depending on the propagation direction. This integrated isolator design has become over the past

[23]The transverse transmission effect is the Voigt effect, and that is reciprocal.

[24]actually such a layout is a hybrid approach of a transverse Kerr phase shifter and a Faraday waveguide. The transverse Kerr effect shifts the TM mode in the backward direction into phase match with the TE radiation spectrum., and the Faraday effect then effectively couples this backward TM mode with the TE radiation spectrum.

decade the most popular one in the research community, for obvious reasons.

1. This type of device operates within a distinct mode and thus does not require tight control of linear birefringence.

2. Isolation occurs by destructive interference, therefore there is no polarizer needed in the backward direction to extinguish the light. The integration of waveguide-type polarizers is achievable but not so straightforward [46]. Therefore this design is in many ways much easier to integrate.

With these designs typical isolation ratios of up to 20dB have been obtained with insertion losses as low as 2dB. Hertel's group did a lot of theoretical work on the optimization of the layer structure and the gyrotropy profile of transverse Kerr phase shifters, leading to a whole range of alternative isolator designs besides MZI's, namely non-reciprocal directional couplers [47], non-reciprocal MMI's [48], and even a cross-strip isolator [49], which is a vertical version of a non-reciprocal MMI. Most of this work has been reported in the Ph.D. work of Lohmeyer, and his theoretical publications [50–52]. This group also showed theoretically how the TM phase shift can be greatly enhanced by using a MO guiding layer that is a ferrimagnetic double layer [53]. Such a double layer consists of a given garnet composition where the upper (or lower) part has been "overcompensated" through partial Fe substitution by Ga. This leads to a change of the dominance of the tetrahedral magnetic sublattice to the dodecahedral sublattice. This is accompanied by a change of sign of the Faraday rotation. If this gyrotropy discontinuity is suitably introduced at a position corresponding to a change of sign of the longitudinal electrical field component of the TM mode, then a considerable increase of the phase shift can be obtained. This is logical as in fact an extra interface is introduced where the transverse Kerr effect takes place (next to the interfaces at the top and bottom of the MO layer). The exact theoretical explanation of this effect will be given in Chapter 2 where nonreciprocal waveguides and transverse Kerr phase shifters in particular are studied in great detail. Despite this important theoretical work, none of these design efforts has effectively led to an improvement of the 20dB observed by Fujita [43] in a basic non-reciprocal MZI design.

This left the integrated isolator research community at the end of the millennium in a somewhat ambiguous position. On the one hand there was the tried-and-trusted waveguide version of a Faraday rotator isolator giving excellent isolation ratios but being subjected to very stringent fabrication requirements and proving less easy to integrate (due to the need for waveguide polarizers), and on the other hand there was the much more relaxed nonreciprocal MZI based on the nonreciprocal transverse Kerr phase shift, with reported isolation ratios that are on the edge of what is practical but with tolerant design criteria, better perspectives for full integration into planar lightwave circuits, and holding the promise of a possible polarization-independent operation [54, 55][25]. While further improvements of the

[25]We didn't touch upon this in this introductory overview of integrated magneto-optics, but it will be shown in Chapter 2 that there can also occur a nonreciprocal Kerr-like phase shift for TE modes if the gyrotropy is in a polar configuration. In this case the nonzero components of

isolation ratio are continuously being pursued, most research groups shifted their attention at the start of the millennium to the next big challenge for planar MO garnet waveguides, namely the integration on standard photonic and/or electronic platforms. After all it makes little sense to have a planar waveguide isolator, if it can not be integrated on the same substrate as the laser diode. And even then it is preferable that the non-reciprocal devices can be processed in parallel with the III-V semiconductor structure in order to achieve self-alignment.

1.3.3 Iron Garnet integration on standard semiconductor platforms

All of the above reported integrated isolator prototypes make use of ferrimagnetic garnets. And while the growth technology of these materials and the control of their magnetic, optical and magneto-optical properties is very mature nowadays, the problem remains that high-quality garnets can only be grown on other garnet templates, such as gallium gadolinium garnet (GGG) or neodymium gallium garnet (NGG). Therefore, the development of on-chip waveguide isolators faces the challenge of either conceiving an epitaxial-like technology of garnets on standard Si and III-V semiconductor substrates without modification of their magnetic properties or achieving bonding between GGG-grown garnet layer structures on those same substrates.

The former approach is for the moment completely unexplored. There have been some experiments on RF sputtering of garnets on glass and silicon substrates but mostly in connection with magnetooptic recording. On top of that this produced films with large uniaxial perpendicular anisotropy combined with large coercivity. Not the ideal situation if one is aiming at the fabrication of a transverse Kerr phase shifter or a Faraday rotator. The other integration approach has received considerably more attention, mostly due to groundbreaking work of the group of Mizumoto.

Direct wafer bonding of garnets onto InP substrates has been successfuly reported for the first time by Yokoi [59] as early as 1995. He later also achieved successful bonding on quaternary InGaAsP alloys [60]. The bonding technique itself is well-developed since it is almost indispensable for the realization of photonic integrated circuits (PIC). For example, the integration of a $LiNbO_3$ modulator for a standard telecom InGaAsP DFB-laser is unthinkable without wafer bonding techniques. In the case of garnet waveguides the integration faces another challenge. The III-V semiconductors have a refractive index (order 3.2–3.5) which is much higher than that of a garnet (order 2.0–2.3). A direct bonding of a processed garnet structure

---

the dielectric MO tensor $\epsilon_{yz} = -\epsilon_{zy}$ couple the fundamental lateral field component E_y of a TE mode with its longitudinal component E_z (which is, for a TE mode, only different from zero if the guiding profile is twodimensional.). This latter is however in general several times smaller than the longitudinal component of a TM mode. On top of that, E_z is in general antisymmetric with respect to the lateral coordinate (y). Therefore, in order to obtain a nonreciprocal TE phase shift in a raised-strip MO waveguide the gyrotropy profile must be anti-symmetric or at least nonsymmetric with respect to the waveguide axis [56, 57]. A nonreciprocal TE phase shift has been observed and reported but its strength was at least an order of magnitude lower than the TM phase shift [58].

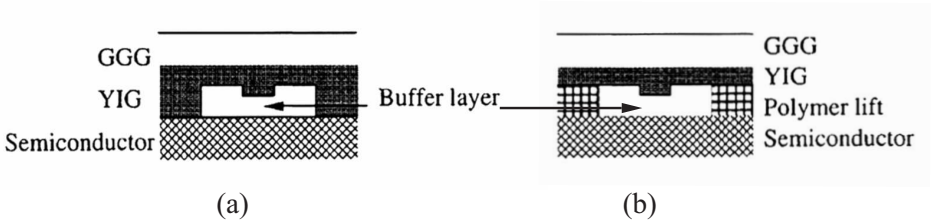


Fig. 1.8: Bonding of garnet structures onto InP substrates with an air buffer layer, using (a) direct wafer bonding, or (b) a "polymer-lift" (after [62]).

onto an InP substrate would therefore lead to unacceptable substrate leakage loss. In order to solve this, two routes have been developed.

The garnet structures are processed first on a GGG host wafer, and then wafer-bonded top-down onto the InP substrate. after which the host substrate is removed by a "suitable technique"[26]. In order to avoid leakage to the semiconductor and also to control the vertical alignment of the bonding, the garnet processing is done within a wide and deeply etched trench of the host substrate. In this way an air buffer is created whose thickness can be controlled by the dimensions of the trench. Another way is to cover first the III-V "recipient" substrate with a polymer layer such as BCB, etch a recess in this polymer layer, and bond the structured garnet onto the "polymer-lift". This method has the advantage that a wider air-gap is implemented more easily than with the direct bonding method, because in the latter method the processing of the patterns gets increasingly difficult as the trench deepens. Its disadvantage is of course that its bonding properties with garnet have not been assessed, and that it is far from certain whether the same thickness (i.e. trench depth) control can be achieved as with the garnet trench. Both methods are illustrated in Fig. 1.8. The main disadvantage of this bottom-up bonding approach is that the garnet structures are processed separately from the III-V structures, so that the alignment both in the vertical and the lateral direction is very critical. While the vertical alignment can be somewhat controlled by the use of etch stop layer, the same doesn't apply to the lateral alignment which has to be done with micron accuracy.

A second more reliable approach is being steadily developed and optimized by the group of Mizumoto. Their concept of solving the problem of optical leakage into the semiconductor layers is original and yet very simple. The functional structure of the integrated isolator, i.e. the MZI, is fabricated in the InGaAsP material system (or any other III-V material system one needs to integrate with) and the nonreciprocal garnet is afterwards bonded on top of this as a cladding layer in which the evanescent tail of the guided TM modes in the InGaAsP structure penetrates. In other words, the role of the MO layer in the guiding structure

[26] A tried and tested method is "thin-film slicing" [61]. This technique relies on the formation of an underlying sacrificial layer in the bulk of the GGG substrate for the release of the crystal film that lies above it. This sacrificial layer is often formed by ion implantation which causes it to have a dramatically higher wet-etch rate than the surrounding material.

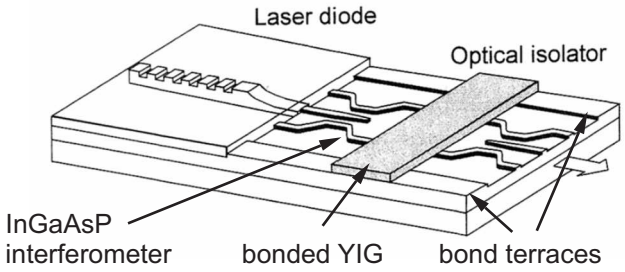


Fig. 1.9: Monolithic laser-isolator integration scheme proposed by Yokoi [45]. The passive interferometer is produced in an InGaAsP material system and can be monolithically integrated with the laser using selective-area growth techniques. The non-reciprocity of the interferometer is obtained by direct wafer bonding of a YIG cladding.

is "transferred" from being a guiding core layer to being a cladding layer. The obvious advantage is that the integration of the isolator with an InP-based active device can be done in a self-aligned and quasi-monolithic manner if techniques such as selective-area epitaxial growth are used to realize the active-passive interface between the MZI and the laser. This is illustrated in Fig. 1.9. This requires of course some preliminary design calculations to optimize this novel layout for the phase shifter. Of particular concern is the optimization of the confinement or overlap of the TM modal profile with the garnet top cladding. In a first design, Yokoi observed only 5.0dB of isolation when simply using Ce:YIG as a cladding for a Q1.3 InGaAsP core layer (0.4μm) [45]. He attributed this limited performance to the asymmetry of the refractive index profile. YIG has a much lower refractive index ($\sim$ 2.2) than the rest of the layers in the InGaAsP-based stack. The modal profile will therefore extend much less into the top cladding but rather in the lower cladding layer and InP substrate. As a result the nonreciprocal phase shift is drastically reduced as compared to a layout with the MO layer as a guiding core. Yokoi elegantly solved this problem by using AlInAs as an undercladding of the InGaAsP core. By subsequently selectively oxidizing this layer, its refractive index is known to be drastically lowered from 3.22 down to a value much closer to that of the garnet index. Yokoi has theoretically assessed the effectiveness of this oxidization process in increasing the non-reciprocal phase shift [63]. Recently, he also successfully fabricated InGaAsP optical waveguides with oxidized AlInAs lower cladding. Using Fabry-Perot resonance measurements he extracted a refractive index for the oxidized Al-layer of 2.45 [64], effectively almost equal to the YIG index. At this moment, this group has successively developed the technology for the final full integration of a laser with a YIG-bonded InGaAsP-based optical isolator. They recently reported successful fabrication and processing of a Fabry-Perot laser monolithically integrated with an optical passive waveguide by selective area epitaxy (SAE) [65]. After the SAE growth, they assessed the feasibility of this integration technique by processing a FP laser with a self-aligned passive outcou-

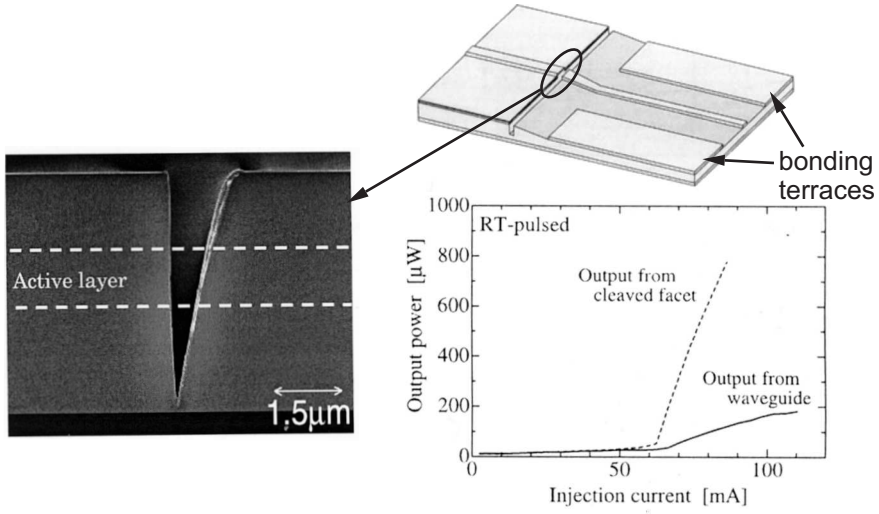


Fig. 1.10: Structure of a laser diode monolithically integrated with a passive waveguide, and the obtained LI characteristics (after [65]). The SEM picture shows the FIB fabricated groove between the active and passive regions of the device with the laser mirror facet on the left. While normally FIB etching leads to two tilted facets, a straight laser facet was obtained by slightly tilting the sample during the etching step. The tilt of the passive facet is about 10° and the width of the groove approximately 300nm.

pling waveguide[27] using standard photolithography and focussed ion beam (FIB) etching to define the laser facet. In parallel, bonding terraces were defined for the later bonding of the garnet cladding. The results of these experiments are shown in Fig. 1.10. Successful laser operation was observed and the excess loss caused by the passive waveguide and the FIB etched groove (diffraction and coupling loss due to the inclined facet) was measured to be 6.5dB. First experiments with a fully integrated laser-isolator device are announced to be underway but are up till now not yet reported.

1.3.4 Conclusion

In this section we have reviewed the evolution of the attempts at a planar optical waveguide isolator and the present state-of-the-art garnet-semiconductor integration techniques. It can be stated that nowadays the know-how on designing and fabricating YIG-based planar optical waveguide isolators is mature. Table 1.2 compares the advantages and disadvantages of the different reported designs of planar YIG-based optical waveguide isolators together with the main achievements.

The same maturity hasn't been reached yet when it comes to the garnet-semiconductor integration technology that is ultimately needed to incorporate the

[27] which would in the final integrated device become the MZI.

	Faraday rotators (FR)	Phase Shifters (PS)
advantages	- strong effect - simple magnetic field configuration - short device lengths - large bandwidth	- no phase matching needed - truly integratable (no polarizers) - possibility to make polarization-independent - reflection effect (III-V core possible) - many degrees of freedom: many routes to optimization, many config's devisable (MZI's, MMI's, dir. couplers)
disadvantages	- need for external magnet - low design tolerance: precise TE/TM phase match needed - extra reciprocal rotator necessary for operation in $\perp$waveguide modes - very difficult to integrate: waveguide polarizers, transmission effect (MO core needed)	- need for external magnet - weak effect: large lengths, difficult to saturate - MZI: $\rightarrow$ **M** in push-pull config $\rightarrow$ limited λ-bandwidth - TM (TE) YIG phase shifter: transparent for TE (TM)
achievements	peak isolation ratio: $>$ 30dB, λ-bandwidth: $>$ 50nm, length: $<$ 2mm, excess loss: $<$2dB [37] BUT: extreme fabrication requirements [36] no integrated device reported	peak isolation ratio: $\approx$ 20dB, λ-bandwidth: $<$ 40nm, length: $>$ 1cm, excess loss: 3–4dB [43] integration on InP [64, 65] and Si [44] platforms underway (first test on InP: only 5dB [45])

TABLE. 1.2: Advantages, disadvantages, and achievements of reported planar YIG-based optical waveguide isolators. The λ-bandwidth is defined is the 3dB isolation ratio bandwidth. The hybrid combination of a Faraday rotator and a phase shifter as reported by Shintaku [41] has not been included in this table. Even though this hybrid FR/PS isolator tries to combine the best of both worlds, the needed magnetization direction (in-plane and slightly tilted with respect to the transverse direction) is not only very difficult to realize but also very intolerant.

waveguide isolators on a standard semiconductor platform.Currently, the technological progress made over the last 10 years by Yokoi and Mizumoto seems very promising, but a fully integrated laser-isolator is still not reported. On top of that it is still not clear whether YIG phase shifters will after all be able to deliver sufficient isolation ratio. Up till now the "disappointing" 19.5dB of Fujita [43] has

not been beaten by other nonreciprocal TM phase shifters. But that design was with a YIG core and can only be integrated using the very unwieldy scheme of Fig. 1.8. To make things worse, the only real isolation experiment on a device that holds the promise of being more or less easy to integrate only yielded an isolation of 5dB [45][28].

It seems that after 25 years of research the quest for an integrated isolator is again at a crossroad. Ten years ago researchers started shifting their attention to the use of non-reciprocal phase shifters instead of waveguide Faraday rotators because of their greater design tolerance and relative larger ease of integration. But despite the efforts and the steady development of integration technologies, the fundamental stumbling block of a limited isolation ratio remained. Results on a nonreciprocal MZI with an oxidized InAlAs undercladding have to be awaited before a "final" verdict can be passed on the performance and feasibility of YIG phase shifters as integrated isolators, but the indications are not rosy.

The ambiguity of the current situation is illustrated by the fact that the group of Levy and Fujita, who hold the isolation record with a non-reciprocal TM phase shifter, abandoned this approach and again picked up the old idea of a waveguide Faraday rotator. However, this time along a completely novel route, namely by trying to greatly enhance the specific Faraday rotation by incorporating the MO materials as defects in a photonic crystal, thereby creating an entirely new domain of research in photonics, namely non-reciprocal magneto-photonic crystal waveguides! This work is still in a theoretical assessment phase [66–69].

1.4 Novel integrated isolator concept– the ferromagnetic-metal-clad semiconductor optical amplifier

The crisis situation in which the isolator research community, in search for an integrated optical waveguide isolator, found itself at the turn of the millennium has spurred several reactions. As explained above, Yokoi's group patiently and steadily continued along their chosen traditional path, while Levy and co picked up an old love (Faraday rotators) and married it to the hot-topic of photonic crystals, thereby giving birth to magneto-photonic crystal research. In between these two extremes, a third way out was suggested independently and almost simultaneously by on the one side Nakano (who also has a big hand in the work of Yokoi) and on the other Ando (who made the first planar waveguide Faraday rotator) [70, 71]. The central idea of their "solution" was very simple: if phase shifters are tolerant in design and allow for integration (because they are based on a MO reflection effect rather than a MO transmission effect so that they can have semiconductor guiding layer), but if at the same time their performance is limited, why not simply stick with a phase shifter but use a stronger MO material?

[28] It has to be admitted though that the design was not optimized due to the lack of a low-index undercladding layer.

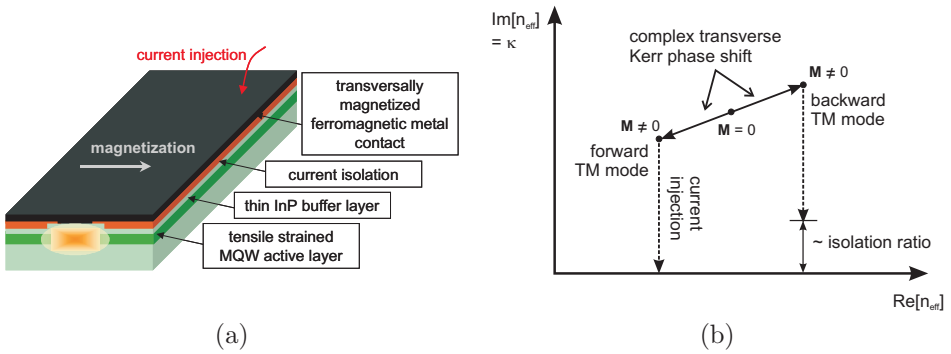


Fig. 1.11: (a): Schematic layout of Nakano's and Ando's novel integrated isolator concept; and (b): its operation principle.

1.4.1 Operation principle, layout and advantages

This sounds very plausible, were it not for the fact that the most obvious candidate for an alternative stronger MO material, namely a ferromagnetic metal, is at the same time strongly absorptive. However, and this is in many ways the stronghold of Nakano's and Ando's suggestion, this lossy character can be exploited in the desired nonreciprocal operation of the device. Indeed, the non-reciprocal Kerr phase shift induced by a **lossy** MO material will now also have an **imaginary** part (on top of the real phase shift)! In other words, besides the "standard" real nonreciprocal phase shift of the guided modes (which forms the basis of a non-reciprocal interferometer), the guided modes will also experience a non-reciprocal modal loss shift. In one direction a guided mode will experience a lower loss than in the opposite direction. If a **reciprocal gain** mechanism is present that precisely compensates the residual loss in the low-loss direction, then the resulting device will operate in a transparent regime in that direction while experiencing a net loss equal to the non-reciprocal absorption shift in the opposite direction. A truly elegant isolator concept!

The most straightforward practical realization of this novel isolator concept is sketched in Fig. 1.11. An amplifying InP-based waveguide structure is provided with a transversely magnetized ferromagnetic film contact, that acts both as magneto-optical gyrotropic cladding layer and as a current injection Ohmic contact. Due its ferromagnetic properties, this metallic film can sustain a permanent magnetization in its plane and transverse to the propagation direction of the guided light, i.e. in a transverse Kerr configuration. As repeatedly explained above this transverse Kerr effect will lift the degeneracy of the forward and backward propagating TM modes of the waveguide. And due to the absorptive nature of the gyrotropic material, the forward and backward propagation TM modes will not only have a different propagation constant ($\Re[n_{\text{eff}}]$), but also a different extinction coefficient ($\Im[n_{\text{eff}}]$). It is intuitively clear that this non-reciprocal absorption shift

will not by itself compensate the optical losses induced by the presence of the metal[29]. By appropriately biasing the underlying semiconductor optical amplifier (SOA) through the ferromagnetic metal contact, the device can be pumped to forward transparency. Furthermore, it is safe to assume that this carrier injection will in a first order approximation leave the modal properties near the metal unchanged, so that the magnitude of the nonreciprocal absorption shift remains constant during the current injection. As a result, when the device is pumped to transparency in the forward direction, it has a loss in the backward direction equal to the non-reciprocal absorption shift.

The design of this SOA/isolator gets down to finding the right compromise between optical TM confinement in the amplifying guiding layer (necessary for forward loss compensation) and confinement near the metal/semicoductor interface (to maximize the non-reciprocity). In a very simplistic view this is merely a question of optimizing the thickness of the spacer layer between the amplifying core and the metal. Roughly stated, decreasing the thickness of this buffer layer will make the necessary material gain as well as the non-reciprocal absorption shift, rise exponentially, not necessarily with the same exponential constant though. If the gain properties of the SOA are known, more in particular the saturation of the material gain of its QW's, then this available gain can be maximally exploited by decreasing the InP spacer layer down to the point corresponding to this gain level. This is of course a very simplified treatment. In reality the interplay between the different waveguide parameters is much more subtle. The design of this device will be treated in considerable detail in Chapters 2 and 3

Ando and Nakano have done some preliminary assessment simulations. They both predicted a non-reciprocal absorption shift in the region of 150dB/cm (or thus a 40dB isolator in a length of just above 2.5mm) at forward transparency gain levels between 1500–2000/cm and a spacer thickness of about 300nm, and showing quasi- flat $\lambda-$band of > 30 nm. These material gain levels necessitate the use of multi-quantum-well (MQW) active layer structures. Their studies served more the purpose of assessing the feasibility of this design. The active MQW layer structure proposed by both of them is far from being optimal (and in the case of Nakano even somewhat hypothetical[30]). However, the layer details are not completely unrealistic, nor are the gain levels. While the reported results of Nakano and Ando are therefore by no means a detailed design example of a novel integrated optical waveguide isolator concept, they can be rightly considered very promising indications that this new concept might constitute an important breakthrough in integrated isolator research. After all, it doesn't require a lot of thought to realize that this novel concept has all the advantages of a traditional phase shifter as summarized in Table 1.2, while at the same time resolving almost all of the mentioned disadvantages.

[29]Roughly stated that would imply that in the dielectric tensor of the metal the off-diagonal elements, i.e. the gyrotropy, would have to be of the same order of magnitude as the diagonal elements, i.e. the isotropic absorption loss. Such a hypothetically strong MO material is up to our knowledge non-existent.

[30]as he considers tensile-strained ternary InGaAs wells emitting at 1550nm. It will be shown in Chapter 3 why it is theoretically impossible to grow such wells.

1. In this design the magnetic layer which provides the magneto-optical effect leading to the isolation, **doesn't need an external magnetic field**. This layer is after all by itself already a ferromagnetic material, and in this way exhibits high remanent magnetisation (along the easy axis) without an external magnetic field. In other words, the "same" permanent film magnet that is normally needed in YIG-based devices solely to saturate the garnet film is now very efficiently used both for its permanent magnetization and its magneto-optic properties.

2. While the transverse Kerr reflection effect still remains weaker than the Faraday transmission effect (which is of course not practical when using metals), it is a well-known fact that the Kerr effects of metals are almost an **order of magnitude stronger than those of garnets**. As a result the specific Kerr phase shift will be also stronger, resulting in a much shorter device length than the one needed in a YIG-based isolator[31].

3. When it was stated that in YIG-based MZI phase shifters one of the main trumps, when compared to the traditional waveguide Faraday rotators, is that such a device operates entirely in a single waveguide mode and that as a consequence no phase-matching between non-reciprocally coupling modes is required, this statement holds even more strongly in this type of device. The SOA/isolator not only operates in a single waveguide mode, but also **doesn't require interfering beams**.
   The isolation behaviour results entirely from the nonreciprocal modal absorption of the guided TM modes. As a result the wavelength transparency of the isolation performance of the device depends on the wavelength dependency of the magneto-optical strength of the metal, the waveguide dispersion of the guided TM modes, the bandwidth of the active layers and the control of the thickness of the buffer layer between the guiding layer and the metal contact. All of these are not critical. Wavelength transparency of at least 30 nm is expected. Since there is no interferometric effect at play, fabrication tolerances are expected to be those of standard SOA devices.
   Secondly, this isolator is not a passive quasi-lossless device. Therefore, it is less important that there is no effect for the TE modes. That is to say, whereas a non-reciprocal YIG-based TM-MZI is essentially bidirectionally transparent for TE light, this device is simply **bidirectionally absorptive for TE light**, if at least the gain properties of the MQW can be engineered to be strongly polarization-dependent. It is known that this can be achieved by growing the QW's under strain. However, this requires careful attention,

[31] Of course this comparison is a bit tricky: in fact it makes no sense to state that the non-reciprocal absorption shift in this SOA/isolator is almost an order of magnitude stronger than the non-reciprocal phase shift in YIG-devices, because then one is comparing an imaginary part of a gyrotropic effect with a real part! Actually, the real part of the complex phase shift of this device is actually of the same order of magnitude as the one in YIG-based phase shifters! But it can in fact be proven that the imaginary part of the complex transverse Kerr effect of a metal is always much stronger than its real part. Therefore it is not completely untrue that a metal will be "stronger" than YIG at least when it comes to the dominant part of the MO effect.

both theoretically and technologically. An important part of Chapters 3 and 4 will be devoted to the subtleties of this.

4. By far the biggest advantage of the SOA/isolator is that it can be **very straightforwardly monolithically integrated**. It was clearly explained that, because the Kerr effect is a reflection effect, a nonreciprocal phase shifter can be designed with a semiconductor guiding layer and a nonreciprocal MO cladding, therefore allowing easier integration. And while the efforts of Yokoi prove indeed this advantage, the design of Nakano and Ando is even still better suited for monolithic integration.
   First of all the MO cladding doesn't need to be wafer-bonded. Ferromagnetic metals can like most metals be sputter-deposited with good properties[32] And secondly, and most importantly, this isolator is an active instead of a passive device! Therefore there is **no question of active-passive integration**. The layer structure is essentially that of an SOA making it particularly suitable for **direct monolithic integration with a DFB laser with exactly the same layer structure** (except for the cladding). Laser and isolator can be grown in the same growth step on the same substrate with just a few standard additional processing steps for the isolator section and without any need for advanced integration techniques (such as selective-area growth). This holds the promise of significant reduction in cost for the production of laser diodes fully integrated with isolators.

1.4.2 Challenges of the novel design

Before overenthusiastically starting to design and fabricate this novel type of optical waveguide isolator, one needs to realize that it successful operation relies entirely on two basic assumptions that were not experimentally assessed at the time of its publication. Experimental success will depend on the combined availability of a "well-behaved" active amplifying region and a "well-behaved" ferromagnetic magneto-optic contact with good interfacial properties towards traditional III-V semiconductors. This is schematically illustrated in Fig. 1.12.

By a "well-behaved" ferromagnetic magneto-optic contact it is meant that the used ferromagnetic metal must have at the same time a sufficiently strong in-plane magnetic anisotropy (to serve as a good permanent magnet), a sufficiently strong gyrotropy at the operation wavelength (to serve as a strongly non-reciprocal absorptive cladding), and a sufficiently low contact resistivity with respect to InP-based semiconductors (to serve as a good Ohmic contact for the underying optical amplifier). Even though it is clear that the latter is purely an interfacial property, the former two are not entirely just material properties. The interface will have a considerable impact on the magnetic anisotropy of the ferromagnetic metal, especially when it is in thin film form. Interfacial effects might also change locally the magneto-optical strength of the metal. These are in this particular case very

[32] after all it is precisely in this way one realizes the thin film magnets that are needed for the YIG-based phase shifters.

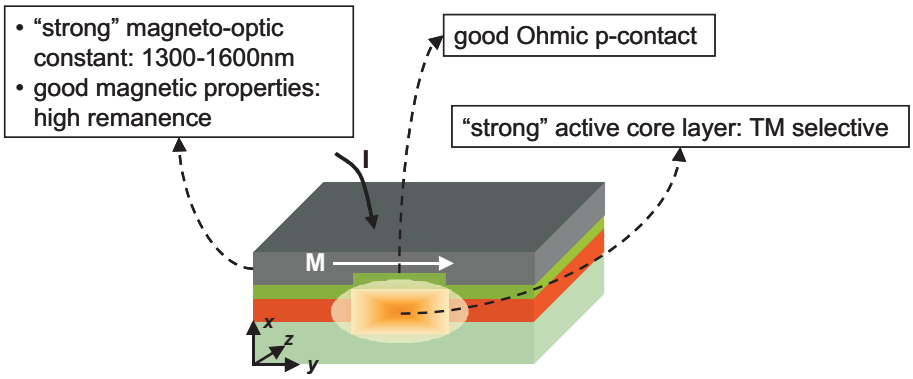


Fig. 1.12: Design and fabrication challenges of the SOA/isolator.

important in view of the fact that the device is based on a reflection-like Kerr effect.

For none of these three effects a proper behaviour is fully guaranteed beforehand. The magnetic anisotropy of ferromagnetic thin films has been of course intensively for many decades now, but it mustn't be forgotten that in the precise geometry considered here (i.e. long stripes with a very high aspect ratio) it is far from clear whether the shape demagnitizing effect can be sufficently suppressed. The magneto-optic strength of ferromagnetic metals at the telecom IR wavelengths are only very limitedly tabulated[33]. And finally, it has never been tried to make an Ohmic contact with ferromagnetic metals.

By a "well-behaved" active amplifying region it is meant that the optical material gain of the QW's must be first of all sufficiently strong but also, and even more importantly, sufficiently polarization discriminating. The latter remark is quite relevant because in the proposed layout the SOA/isolator operates purely in TM polarization. If the gain region is not polarization discriminating, any TE light that accidently enters the device might get amplified and traverse the device "unhindered", creating an undesired situation similar to YIG-based isolators. However, even if the TE light can not reach transparency, its presence (being it from an external source or spontaneously emitted inside the cavity itself) inside the gain medium will still affect the performance of the device, as it will consume carriers and add noise to the signal. Therefore it is primordial to suppress the TE spontaneous emission factor (and thus the TE gain) of the QW's. It will be clearly explained in Chapter 3 how this can be achieved by growing the QW's under tensile strain, i.e. having a lattice constant that is smaller than the host material. The use of strain to engineer the gain properties of SOA's and lasers has been known for many years. However the application of tensile (as a opposed to compressive) strain has been very limited. Even though there is no reason to

[33]For obvious reasons. Up till now the magneto-optical use of metals has been mainly in MO memory applications where it doesn't make a lot of sense to use long-wavelength IR in view of storage capacity.

expect that there are technological issues that aren't present for compressively-strained wells, it is clear that the lack of expertise and general know-how presents a challenge for this device. It is far from experimentally assessed to what gain limits tensile-strained wells can be driven.

Therefore, it is clear that it is not all that straightforward to expect without a doubt an uncomplicated fabrication and performance of the above sketched isolator.

1.5 Goal and outline of this work

It is clear by now that the goal of this doctoral research is to study both theoretically and experimentally this novel isolator concept. In view of all of the above remarks, it is clear that it is risky to set a precise a level of isolation as an ultimate goal. Even though the concept of Nakano sounds very appealing, there is a high degree of uncertainty both on the achievable magneto-optical strengths and the achievable TM material gain. Therefore, even though it should be the ultimate goal to achieve a device that is fully integrated with a DFB laser, that delivers at least >25dB of isolation[34], that can operate uncooled, and thus doesn't need absolute current levels higher than 100mA (or preferrably even 50mA), that will have a coherence collapse threshold of at least >-10dB return loss, and that can operate penalty-free below return levels of -10dB under direct modulation at bitrates[35] of 2.5Gb/s, ..., it would already constitute quite a feat if this novel concept is demonstrated experimentally under forward transparency conditions.

The development of strongly amplifying and polarization discriminating strained QW's already poses a big challenge by itself. For fear of being too ambitious, it is chosen to limit this work to an operation wavelength of 1300nm. It was explained in the introduction to this chapter that metro is not only an increasingly important market (where price really matters), but also that especially for 1300nm DFB lasers an integrated optical isolator could constitute a considerable price reduction, because for this wavelength it makes sense to use direct intensity modulation, leaving the isolator as the only non-integrated component.

The following subtasks can be defined:

To study and optimize the choice of ferromagnetic magneto-optic contact material, and to study, characterize, and optimize the properties of the semiconductor/magneto-optic material interface (at the metro telecom wavelength of 1300nm)
As the desired magneto-optic effects occur mainly at the above mentioned interface, special attention will be paid to the structural, electric, magnetic and magneto-optic quality of the interface.

To study, characterize and optimize active semiconductor layer structures dedicated towards use in ferromagnetic-metal-loaded high gain SOA's operating at 1.3 μm.

[34] which is considered sufficient up to bitrates of STM-16

[35] For the moment the market for 10 Gb/s Black and White transceivers is still relatively small. Reports of 10Gig devices are still in the lab stage. On top of that at 10Gb/s direct modulation an isolation ratio of 25dB will not be sufficient.

Considering the polarization dependence of the operation principle of the proposed design and the needed loss compensation, a considerable amount of effort will be directed towards the realization of high gain, highly polarization discriminating active layers.

To design, fabricate and characterize transversely magnetized ferromagnetic-metal-loaded 1.3 μm SOA's for use as an integrated optical isolating section. Effort will be put into assessment of this key objective of this research towards an optical component meeting the requirements of an advanced optical communication system. Attention must be paid to transparency towards wavelength, input power, temperature and repetition rate.

Finally, on a more global scale and of great importance for a future commercial impact of this type of device, it is of course the ultimate goal to realize a fully monolithically integrated directly modulated DFB laser-isolator that has an increased coherence collapse threshold (for example) and that is operational for any telecom wavelength, for any bit rate, and that has a high output power.

This work is organized as follows. In Chapter 2 an in-depth theoretical treatment will be given of the modelling of non-reciprocal waveguides in general and transversal magneto-optic waveguides in particular. Focus will be on the different techniques available to calculate the guided full modal spectrum (i.e. for forward and backward propagation modes), taking normalization and orthogonality into account. In this chapter, an important part of the work will consist of assessing the validity of perturbation theory modelling for the modelling of the SOA/isolator. The different design subtasks and challenges, indicated in the above section 1.4.2, will be studied one by one on a theoretical basis in Chapter 3. Where necessary, it will also be indicated what the underlying theory is for the characterization methods needed to assess the performance of these subcomponents. For example, the non-trivial task of measuring the small magneto-optic constants by magneto-optic ellipsometry and photometry will be explained. The last two chapters will then entirely focus on the experimental assessment of the building blocks of the SOA/isolator (Chapter 4) and the assessments of the nonreciprocal behaviour of the first prototypes of the SOA/isolator (Chapter 5). Finally, Chapter 6 will offer some conclusions concerning the here achieved results and present perspectives for future work.

1.6 Publications

The work carried out during this thesis has led to a number of publications in international refereed journals:

- M. Vanwolleghem, W. Van Parys, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "Experimental demonstration of nonreciprocal amplified spontaneous emission in a CoFe clad semiconductor optical amplifier for use as an integrated optical isolator." *Appl. Phys. Lett.,* vol. 85(18), pp. 2980–2982, 2004.

- K. Postava, M. Vanwolleghem, D. Van Thourhout, R. Baets, Š. Višňovský,

P. Beauvillain, and J. Pištora, "Modeling of a novel InP-based monolithically integrated magneto-optical waveguide isolator." *J. Opt. Soc. Am. B.,* vol. 22(1), pp. 261–273, 2005.

- M. Vanwolleghem, P. Gogol, B. Bartenlian, P. Beauvillain, J. Harmle, L. Lagae, J. Pistora, K. Postava, S. Visnovsky, and R. Wirix-Speetjens, "Magneto-optical parameters of $Co_{90}Fe_{10}$ and $Co_{50}Fe_{50}$ ferromagnetic thin films for 1.3μm integrated isolator." submitted to *J. Appl. Phys.,* (February 2005).

- F. Lelarge, B. Dagens, C. Cuisin, O. Le Gouezigou, G. Patriarche, W. Van Parys, M. Vanwolleghem, R. Baets, and J.-L. Gentner, "GSMBE growth of GaInAsP/InP 1.3μm-TM-lasers for monolithic integration with optical waveguide isolator." *J. Cryst. Growth,* vol. 278(1–4), pp. 709–713, 2005.

- W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "Development of a ferromagnetic transition metal Ohmic p-type contact for a novel integrated optical isolator conept." submitted to *Electron. Lett.,* (February 2005).

This work was also presented at a number of international and national conferences and workshops:

- M. Vanwolleghem, W.F.H. Van Parys, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens, and J. De Boeck, "Experimental verification of a novel integrated isolator concept." in the Proceedings of *29th European Conference on Optical Communications (ECOC),* post-deadline papers 6, p. 78–79, Rimini, Italy, September 2003.

- M. Vanwolleghem, W.F.H. Van Parys, D. Van Thourhout, R. Baets, O. Gauthier-Lafaye, F. Lelarge, B. Thedrez, R. Wirix-Speetjens, and J. De Boeck, "First Experimental Demonstration of a monolithically integrated InP-based waveguide isolator." in the Proceedings of the *Optical Fibre Communication (OFC '04),* p. TuE6, Los Angeles, USA, February 2004.

- W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, J. Decobert, B. Dagens, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "Demonstration of 81dB/cm isolation on an InP-based optical waveguide isolator." accepted for oral presentation at the *12th European Conference on Integrated Optics (ECIO),* Grenoble, France, April 2005.

- W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, J. Decobert, B. Dagens B. Thedrez, R. Wirix-Speetjens and L. Lagae "InP-Based Monolithically Integrated Optical Waveguide Isolator with 32 dB/cm Isolation", in the Proceedings of the *17th Annual Meeting of the Laser and Electro-Optics Society (LEOS),* paper TuY5, pp. 386–387, Puerto Rico, November 2004.

- M. Vanwolleghem, M. Leys, J. Das, L. Lagae, J. De Boeck, P. Van Daele, R. Baets, "Ferromagnetic-metal-based InGaAs(P)/InP optical waveguide isolator : steps towards experimental validation." in the Proceedings of the

14th Annual Meeting of the Laser and Electro-Optics Society (LEOS), pp. 294–295, San Diego, USA, November 2001.

- M. Vanwolleghem, W.F.H. Van Parys, S. Verstuyft, R. Wirix-Speetjens, L. Lagae, J. De Boeck, and R. Baets, "Ferromagnetic-metal-based InGaAs(P)/InP optical waveguide isolator: electrical and magneto-optical characterisation." in the Proceedings of the *2002 Annual Symposium of the IEEE/LEOS Benelux Chapter,* pp. 282–285, Amsterdam, The Netherlands, 2002.

- W.F.H. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens, and L. Lagae, "Experimental Verification of a Novel Integrated Optical Isolator Concept." in the Proceedings of *Bianisotropics 2004 – 10th Conference on Complex Media and Metamaterials*, pp. 224–227, Ghent, Belgium, September 2004.

- F. Lelarge, B. Dagens, C. Cuisin, O. Le Gouezigou, G. Patriarche, W. Van Parys, M. Vanwolleghem, R. Baets, and J.-L. Gentner, "GSMBE growth of GaInAsP/InP 1.3μm-TM-lasers for monolithic integration with optical waveguide isolator." in the Proceedings of the *13th International Conference on Molecular Beam Epitaxy(IC-MBE),* p. ThA2.2, Edinburgh, Scotland, August 2004.

- M. Vanwolleghem, "A novel monolithically integrated InP-based optical isolator - from theory to practice: major achievements of the ISOLASER project." *IST-OPTIMIST international workshop,* Athens, Greece, June 2004.

- M. Vanwolleghem, "A novel integrated isolator concept." *COST 267 Workshop,* Ghent, Belgium, May 2001.

- M. Vanwolleghem, and R. Baets, "A novel type of integrable non-reciprocal photonic waveguides devices." *1ste doctoraatssymposium Faculteit Toegepaste Wetenschappen,* Ghent, Belgium, December 2000 *(in Dutch).*

The first experimental success was also reported in the periodical newsletter of the IMEC research institute:

- W. Van Parys, and M. Vanwolleghem, "Experimental demonstration of a novel type of optical isolator." *IMEC Newsletter* vol. 39, p. 7, July 2004.

During the timeframe of this research tutoring of a student research task has resulted in the following publication

- P. Bienstman, E. Six, M. Roelens, M. Vanwolleghem, and R. Baets, "Calculation of bending losses in dielectric waveguides using eigenmode expansion and perfectly matched layers." *IEEE Photon. Technol. Lett.,* vol. 14(2), pp. 164–166, 2002.

References

[1] Strategies Unlimited, "Telecom Lasers, Transceivers, and Transponders – Market Review and Forecast 2003," January 2003.

[2] S. Penzo , D. Cooperson , and A. Dwivedi, "RHK Market Update Report – Optical Networks Global," June 2003.

[3] D. Klotzkin, K. Kojima, N. Jordache, N. Chand, P. Kiely, M. Chien, M. Han, E. Michel, S. Ustin, S. Roycroft, D. Melville, R. Kunkel, and L. Ketelsen, "High-Speed Directly Modulated FabryPerot and Distributed-Feedback Spot-Size-Converted Lasers Suitable for Passive Alignment, Unisolated Operation, and Uncooled Environments up to 85 °C." *J. Lightwave Technol.,* vol.21(1), pp. 69–78, 2003.

[4] F. Grillot, B. Thedrez, J. Py, O. Gauthier-Lafaye, V. Voiriot, and J. L. Lafragette, "2.5 Gbit/s transmission characteristics of 1.3μm DFB lasers with external optical feedback." *IEEE Photon. Technol. Lett.,* vol.14(1), pp. 101-103, 2002.

[5] B. Thedrez, J.M. Rainsant, N. Aberkane, V. Voiriot, S. Hubert, J.L. Lafragette, L. Roux, M.F. Martineau, F. Gaborit, B. Fernier, and D. Domingues, "2.5 Gbit/s Error-floor free transmission with high extinction ratio using directly modulated 1.55μm tapered lasers." in the *Proceedings of the Optical Fibre Communication Conference 2000,* vol.1, pp. 184–186, Baltimore, USA, 2000.

[6] B. Fernier, K. Adams, C. Artigue, T. Barrou, A. Göth, E. Grard,W. Jörg, D. Keller, J. L. Lafragette, A. Lestra, P. Pagnod, S. Rabaron, J. M. Rainsant, J. Scherb, D. Toullier, D. Tregoat, and W. Rehm, "1.3 μm low cost plastic module for 622Mbit/s transmission at 85°C." in the *Proceedings of the 24*$^{\text{th}}$ *European Conference on Optical Communication,* vol.1, pp. 445–446, Madrid, Spain, 1998.

[7] M. Galarza, K. De Mesel, S. Verstuyft, D. Fuentes, C. Aramburu, M Lopez-Amo, I. Moerman, P. Van Daele, and R. Baets, "Mode-expanded 1.55μm InP-InGaAsP Fabry-Perot lasers using ARROW waveguides for efficient fiber coupling." *IEEE J. Select. Topics Quantum Electron.,* vol. 8(6), pp.1389–1398, 2002.

[8] D. Van Thourhout, A. Van Hove, T. Van Caenegem, I. Moerman, P. Van Daele, R. Baets, X.J.M. Leijtens, and M.K. Smit, "Packaged hybrid integration phased-array multi-wavelength laser." *Electron. Lett.* vol. 36(5), pp. 434–435, 2000.

[9] A. Black, A.R. Hawkins, N.M. Margalit, D.I. Babic, A.L. Holmes, Jr., and Y.-L. Chang, "Wafer fusion: materials issues and device results." *IEEE J. Select. Topics Quantum Electron.,* vol. 3(3), pp.943–951, 1997.

[10] I. Christiaens, G. Roelkens, K. De Mesel, D. Van Thourhout, and R. Baets, "Thin film devices fabricated with BCB waferbonding." accepted for publication in *J. Lightwave Technol.,* 2005 (to be published).

[11] N. Froberg, G. Raybon, A.M. Johnson, Y.K. Chen, T. Tanbun-Ek, R.A. Logan, A. Tate, A.M. Sargent, K. Wecht, P.F. Sciortino,Jr, U. Koren, B.I. Miller, M.G. Young, and M. Chien, "Multi-gigabit short pulse generation from integrated DBR laser/modulators." in the *Proceedings of the 7*$^{\text{th}}$ *Annual Meeting of Laser and Electro-optics Society,* Boston, USA, 1994.

[12] C.H. Joyner, C.R. Doerr, L.W. Stulz, J.C. Centanni, and M. Zirngibl, "Low-Threshold Nine-Channel Waveguide grating router-based continuous wave transmitter." *J. Lightwave Technol.,* vol. 17(4), pp. 647–651, 1999.

[13] M.L. Mašanović, V. Lal, J.S. Barton, E.J. Skogen, L.A. Coldren, and D.J. Blumenthal, "Monolithically integrated Mach-Zehnder interferometer wavelength converter and widely tunable laser in InP." em IEEE Photon. Technol. Lett., vol.15 (8), pp.117–119, 2003.

[14] F. Favre, "Theoretical analysis of external optical feedback on DFB semiconductor laser." *IEEE J. Quantum Electron.,* vol. QE-23(1), pp. 81-88, 1987.

[15] D. Lenstra, B. H. Verbeek, and A. J. Den Boef, "Coherence collapse in single-mode semiconductor lasers due to optical feedback." *IEEE J. Quantum Electron.,* vol. QE-21(6), pp. 674-679, 1985.

[16] R. B. Clarke, The effect of reflections on the system performances of intensity modulated laser diodes." *J. Lightwave Technol.,* vol. 9(6), pp. 741-749, 1991.

[17] P.M. Oppeneer, T. Maurer, J. Sticht, and J. Kübler, "*Ab initio* calculated magneto-optical Kerr effect of ferromagnetic metals: Fe and Ni." *Phys. Rev. B,* vol. 45(19), pp. 10924–10933, 1974.

[18] L. Onsager, "Reciprocal relations in irreversible processes." *Phys. Rev.,* vol. 37, pp. 405-426, 1931.

[19] F. Olyslager, *Electromagnetic Waveguides and Transmission Lines.* Oxford Engineering Science Series, No. 51 (Clarendon Press, Oxford, 1999).

[20] R. Atkinson, and P. H. Lissberger, "Sign conventions in magneto-optical calculations and measurements." *Appl. Opt.,* vol. 31(28), pp. 6076–6081, 1992.

[21] A. K. Zvezdin, and V. A. Kotov, Chapter 2 in *Modern Magnetooptics and Magnetooptical Materials.* (Institute of Physics Publishing, Philadelphia, USA, 1997).

[22] R. M. A. Azzam, and N.M. Bashara, *Ellipsometry and Polarized Light.* (North-Holland Physics Publishing, Elsevier Science Publishers, Amsterdam, The Netherlands, 1987).

[23] M. Mansuripur, "Analysis of multilayer thin-film structures containing magneto-optic and anisotropic media at oblique incidence using 2×2 matrices." *J. Appl. Phys.,* vol. 67(10), pp. 6466–6475, 1990.

[24] Š. Višňovský, R. Lopuňík, M. Bauer, J. Bok, J. Fassbender, and B. Hillebrands, "Magnetooptic ellipsometry in multilayers at arbitrary magnetization." *Optics Express,* vol. 9(3), pp. 121–135, 2001.

[25] C.-Y. You, and S.-C. Shin, "Generalized analytic formulae for magneto-optical Kerr effects." *J. Appl. Phys.,* vol. 84(1), 541–546, 1998.

[26] V.J. Fratello, S.J. Licht, and C.D. Brandle, "Innovative improvements in bismuth-doped rare-earth iron garnet Faraday rotators." *IEEE Trans. Magn.,* vol. 32(5), pp. 4102–4107, 1996.

[27] P.K. Tien, R.J. Martin, R. Wolfe, R.C. Le Craw, and S.L. Blank, "Switching and modulation of light in magneto-optic waveguides of garnet films." *Appl. Phys. Lett.,* vol. 21(8), pp.394–396, 1972.

[28] M. Levy "The on-chip integration of magnetooptic waveguide isolators." *IEEE J. Sel. Top. Quantum Electron.,* vol. 8(6), pp. 1300–1306, 2002.

[29] M. Levy, "Nanomagnetic route to bias-magnet-free, on-chip Faraday rotators." *J. Opt. Soc. Am. B,* vol. 22(1), pp. 254–260, 2005.

[30] H. Dötsch, N. Bahlmann, O. Zhuromskyy, M. Hammer, L. Wilkens, R. Gerhardt, P. Hertel, and A.F. Popkov "Applications of magneto-optical waveguides in integrated optics: review." *J. Opt. Soc. Am. B,* vol. 22(1), pp. 240–253, 2005.

[31] J. Warner, "Nonreciprocal Magnetooptic Waveguides." *IEEE Trans. Microwave Theory Tech.,* vol. MTT-23(1), pp. 70–78, 1975.

[32] S. Yamamoto, Y. Okamura, and T. Makimoto, "Analysis and design of semileaky-type thin-film optical waveguide isolator." *IEEE J. Quantum Electron.,* vol. QE-12(12), pp. 764–770, 1976.

[33] K. Ando, T. Okoshi, and N. Koshizuka, "Waveguide magneto-optic isolator fabricated by laser annealing." *Appl. Phys. Lett.,* vol. 53(1), pp. 4–6, 1988.

[34] M. Lohmeyer, N. Bahlmann, O. Zhuromskyy, H. Dötsch, and P. Hertel, "Phase-matched rectangular magnetooptic waveguides for applications in integrated optics isolators: numerical assessment." *Opt. Comm.,* vol. 158, pp. 189-200, 1998.

[35] H. Dammann, E. Pross, and G. Rabe, "Phase matching in symmetrical single-mode magneto-optic waveguides by application of stress." *Appl. Phys. Lett.,* vol. 49, pp. 1755–1757, 1986.

[36] R. Wolfe, V.J. Fratello, and M. McGlashan-Powell, "Elimination of birefringence in garnet films for magneto-optic waveguide devices." *Appl. Phys. Lett.,* vol. 51(16), pp. 1221–1223, 1987.

[37] R. Wolfe, J.F. Dillon,Jr., R.A. Lieberman, and V.J. Fratello, "Broadband magneto-optic waveguide isolator." *Appl. Phys. Lett.,* vol. 57(10), pp. 960–962, 1990.

[38] M. Levy, I. Ilic, R. Scarmozzino, R.M. Osgood,Jr., R. Wolfe, C. J. Gutierrez, and G.A. Prinz, "Thin-Film-Magnet Magnetooptic Waveguide Isolator." *IEEE Photon. Technol. Lett.,* vol. 5(2), pp. 198–200, 1993.

[39] M. Levy, R.M. Osgood,Jr., H. Hegde, F.J. Cadieu, R. Wolfe, and V.J. Fratello, "Integrated Optical Isolators with Sputter-Deposited Thin-Film Magnets." *IEEE Photon. Technol. Lett.,* vol. 8(7), pp. 903–905, 1996.

[40] F. Auracher, and H.H. Witte, "A new design for an integrated optical isolator." *Opt. Comm.,* vol. 13, pp. 435–438, 1975.

[41] T. Shintaku, "Integrated optical isolator based on efficient nonreciprocal radiation mode conversion." *Appl. Phys. Lett.,* vol. 73(14), pp. 1946–1948, 1998.

[42] J. Fujita, M. Levy, R. U. Ahmad, R. M. Osgood, Jr., M. Randles, C. Gutierrez, and R. Villareal, "Observation of optical isolation based on nonreciprocal phase shift in a MachZehnder interferometer." *Appl. Phys. Lett.,* vol. 75(7), pp. 998–1000, 1999.

[43] J. Fujita, M. Levy, R. M. Osgood, Jr., L. Wilkens, and H. Dtsch "Waveguide optical isolator based on MachZehnder interferometer." *Appl. Phys. Lett.,* vol. 76(16), pp. 2158–2160, 2000.

[44] H. Yokoi, T. Mizumoto, and Y. Shoji, "Optical nonreciprocal devices with a silicon guiding layer fabricated by wafer bonding." *Appl. Opt.,* vol. 42(33), pp. 6605–6612, 2003.

[45] H. Yokoi, T. Mizumoto, N. Shinjo, N. Futakuchi N, and Y. Nakano, "Demonstration of an optical isolator with a semiconductor guiding layer that was obtained by use of a nonreciprocal phase shift." *Appl. Opt.,* vol. 39(33), pp. 6158–6164, 2000.

[46] J. Fujita, M. Levy, R. Scarmozzino, and R.M. Osgood,Jr., "Integrated multistack waveguide polarizer." *IEEE Photon. Technol. Lett.,* vol. 10(1), pp. 93–95, 1998.

[47] N. Bahlmann, M. Lohmeyer, O. Zhuromskyy, H. Dötsch, P. Hertel, and A. Popkov, "Nonreciprocal coupled waveguides for integrated optical isolators and circulators for TM modes." *Opt. Comm.,* vol. 161, pp. 330–337, 1999.

[48] O. Zhuromskyy, M. Lohmeyer, N. Bahlmann, P. Hertel, H. Dötsch, and A. Popkov, "Analysis of nonreciprocal light propagation in multimode imaging devices." *Opt. Quantum Electron.,* vol. 32, pp. 885–897, 2000.

[49] M. Lohmeyer, L. Wilkens, O. Zhuromskyy, H. Dötsch, and P. Hertel, "Integrated magneto-optic cross-strip isolator." *Opt. Comm.,* vol. 189, pp. 251–259, 2001.

[50] M. Lohmeyer, "Guided waves in rectangular integrated magnetooptic devices." Ph.D. thesis, (University of Osnabrück, Germany, 1999).

[51] M. Lohmeyer, N. Bahlmann, O. Zhuromskyy, and P. Hertel, "Radiatively coupled waveguide polarization splitter simulated by wave-matching-based coupled mode theory." *Opt Quantum Electron.,* vol. 31, pp. 877–891, 1999.

[52] M. Lohmeyer, N. Bahlmann, O. Zhuromskyy, H. Dötsch, and P. Hertel, "Unidirectional Magnetooptic Polarization Converters." *J. Lightwave Technol.,* vol. 17(12), pp. 2605–2611, 1999.

[53] N. Bahlmann, V. Chandrasekhara, A. Erdmann, R. Gerhardt, P. Hertel, R. Lehmann, D. Salz, F.-J. Schr"oteler, M. Wallenhorst, and H. Dötsch, "Improved Design of Magnetooptic Waveguides for Optical Isolators." *J. Lightwave Technol.,* vol. 16(5), pp. 818–823, 1998.

[54] O. Zhuromskyy, H. Dötsch, M. Lohmeyer, L. Wilkens, and P. Hertel, "Magnetooptical waveguides with polarization-independent nonreciprocal phase shift." *J. Lightwave Technol.,* vol. 19(2), pp. 214–221, 2001.

[55] J. Fujita, M. Levy, R. M. Osgood, Jr., L. Wilkens, and H. Dötsch, "Polarization-independent waveguide optical isolator based on nonreciprocal phase shift." *IEEE Photon. Technol. Lett.,* vol. 12(11), pp. 1510–1512, 2000.

[56] L. Wilkens, D. Trger, H. Dötsch, A. F. Popkov and A. M. Alekseev "Nonreciprocal phase shift of TE modes induced by a compensation wall in a magnetooptic rib waveguide." *Appl. Phys. Lett.,* vol. 79(26), pp. 4292–4294, 2001.

[57] A. F. Popkov, M. Fehndrich, M. Lohmeyer, and H. Dötsch, "Nonreciprocal TE-mode phase shift by domain walls in magnetooptic rib waveguides." *Appl. Phys. Lett.,* vol. 72(20), pp. 2508–2510, 2001.

[58] M. Fehndrich, A. Josef, L. Wilkens, J. KleineBörger, N. Bahlmann, M. Lohmeyer, P. Hertel, and H. Dötsch, "Experimental investigation of the nonreciprocal phase shift of a transverse electric mode in a magneto-optic rib waveguide." *Appl. Phys. Lett.,* vol. 74(20), pp. 2918–2920, 2001.

[59] H. Yokoi, T. Mizumoto, K. Maru, and Y. Naito, "Direct bonding between InP and rare-earth iron-garnet grown on $Gd_3Ga_5O_{12}$ substrate by liquid-phase epitaxy." *Electron. Lett.,* vol. 31(18), pp. 1612–1613, 1995.

[60] H. Yokoi, T. Mizumoto, M. Shimizu, N. Futakuchi, N. Kaida, and Y. Nakano, "Direct Bonding between Quaternary Compound Semiconductor and Garnet Crystals for Integrated Optical Isolator." *Jpn. J. Appl. Phys. Pt.1,* vol. 38(1A), pp. 195–197, 1999.

[61] T. Izuhara, M. Levy, and R.M. Osgood,Jr., "Direct wafer bonding and transfer of 10- mm-thick magnetic garnet films onto semiconductor surfaces." *Appl. Phys. Lett.,* vol. 76(10), pp. 1261–1263, 2000.

[62] T. Izuhara, J. Fujita, M. Levy, and R.M. Osgood,Jr., "Integration of Magnetooptical Waveguides onto a III-V Semiconductor Surface." *IEEE Photon. Technol. Lett.,* vol. 14(2), pp. 167–169, 2002.

[63] H. Yokoi, T. Mizumoto, H. Masaki, N. Futakuchi, T. Ohtsuka and Y. Nakano, "Selective oxidation for enhancement of magneto-optic effect in optical isolator with semiconductor guiding layer." *Electron. Lett.,* vol. 37(4), pp. 240–241, 2001.

[64] Y. Shoji, H. Yokoi, T. Mizumoto, "Enhancement of Magneto-optic Effect in Optical Isolator with GaInAsP guiding layer by Selective Oxidation of AlInAs." *Jpn. J. Appl. Phys.,* vol. 43(2), pp. 590–593, 2004.

[65] K. Sakurai, H. Yokoi, T. Mizumoto, D. Miyashita, and Y. Nakano, "Fabrication of Semiconductor Laser for Integration with Optical Isolator." *Jpn. J. Appl. Phys.,* vol. 43(4A), pp. 1388–1392, 2004.

[66] M.J. Steel, M. Levy, and R.M. Osgood,Jr. "Photonic bandgaps with defects and the enhancement of Faraday rotation." *J. Lightwave Technol.,* vol. 18(9), pp. 1297–1308, 2000.

[67] M.J. Steel, M. Levy, and R.M. Osgood,Jr. "High transmission enhanced Faraday rotation in one-dimensional photonic crystals with defects." *IEEE Photon. Technol. Lett.,* vol. 12(9), pp. 1171–1173, 2000.

[68] M. Levy, H.C. Yang, M.J. Steel, and J. Fujita, " Flat-top response in one-dimensional magnetic photonic bandgap structures with Faraday rotation enhancement." *J. Lightwave Technol.,* vol. 19(12), pp. 1964–1969, 2001.

[69] M. Levy, "Nanomagnetic route to bias-magnet-free, on-chip Faraday rotators." *J. Opt. Soc. Am. B,* vol. 22(1), pp.254–260, 2005.

[70] M. Takenaka, and Y. Nakano, "Proposal of a novel semiconductor optical waveguide isolator." in the Proceedings of 11^{th} *International Conference on Indium Phosphide and Related Materials,* pp. 298–292, Davos, Switzerland, 1999.

[71] W. Zaets, and K. Ando, "Optical waveguide isolator based on nonreciprocal loss/gain of amplifier covered by ferromagnetic layer." *IEEE Photon. Technol. Lett.,* vol. 11(8), pp. 1012–1014, 1999.

2

Modelling of non-reciprocal magneto-optical waveguides

NON-RECIPROCAL electromagnetic waveguides are guides that do not coincide with their adjoint medium [1]. As was discussed in Chapter 1, this automatically implies that the theoretical modelling of such waveguides requires the treatment of Maxwell's curl equations supplemented with full-tensorial constitutive relationships. This obviously severely enhances the difficulty of the modelling task.

In this chapter, we will start by pointing out some important peculiarities of non-reciprocal waveguides and the properties of their eigenmodes. In Section 2.2 an overview will be given of the different modelling approaches for the special case of non-reciprocal, anisotropic waveguides. There, it will be noticed that already for a 1D non-reciprocal waveguide, *ab initio* modelling of the eigenproblem poses several theoretical problems. However, as is often the case in mathematical physics, approximative solution techniques of various order of validity can help in reducing the mathematical difficulties. The best known and most widespread technique is based on the so-called *modal expansion* method. Both ab initio and perturbative modelling algorithms are elaborated, implemented and illustrated by examples.

In the last section 2.3, the developed schemes of section 2.2 will be specialized to the case of transverse magneto-optical waveguides, which occupy a central role in this work. It will be shown how in this case the 1D rigorous modelling tool described in the previous section, can be considerably simplified numerically. Finally, both approaches will be numerically compared by a benchmark simulation of a simplified version of the integrated optical isolator studied in this work. This comparison will reveal the advantages and disadvantages of both approaches.

2.1 Reciprocity

The terms reciprocity, bidirectionality and even mirroring are often mistakenly used to describe the same phenomena. There is however a very clear distinction between the three of them. As an illustration of how easily these labels are mistreated, consider the following correct statements. Every reciprocal waveguide is bidirectional. There exist bidirectional waveguides that are non-reciprocal. All mirroring waveguides are bidirectional. But as a consequence not all reciprocal waveguides are mirroring.

In order to clearly define the framework of this research and because of the very central role that the Lorentz reciprocity theorem fulfills in this work, its derivation is repeated here. The definitions of bidirectionality and mirroring are then compared to that of reciprocity.

2.1.1 Lorentz reciprocity theorem

The derivation of the Lorentz reciprocity theorem that will be given here is not the original one by Lorentz [2]. A more common and modern derivation based on Maxwell's equations in the frequency domain and valid for a general bianisotropic waveguide, is followed [3], [4].

Consider a waveguide 'a', described by the material parameters[1] $\underline{\epsilon}_a(\mathbf{r})$, $\underline{\mu}_a(\mathbf{r})$, $\underline{\zeta}_a(\mathbf{r})$ and $\underline{\xi}_a(\mathbf{r})$, sustaining the electromagnetic fields $\mathbf{e}_a(\mathbf{r})$ and $\mathbf{h}_a(\mathbf{r})$ generated by electric and magnetic sources $\mathbf{j}_a(\mathbf{r})$ and $\mathbf{k}_a(\mathbf{r})$. The Maxwell curl equations describing these fields thus read

$$\boldsymbol{\nabla} \times \mathbf{e}_a(\mathbf{r}) = -j\omega\underline{\mu}_a(\mathbf{r}) \cdot \mathbf{h}_a(\mathbf{r}) - j\omega\underline{\zeta}_a(\mathbf{r}) \cdot \mathbf{e}_a(\mathbf{r}) - \mathbf{k}_a(\mathbf{r}) \tag{2.1a}$$

$$\boldsymbol{\nabla} \times \mathbf{h}_a(\mathbf{r}) = j\omega\underline{\epsilon}_a(\mathbf{r}) \cdot \mathbf{e}_a(\mathbf{r}) + j\omega\underline{\xi}_a(\mathbf{r}) \cdot \mathbf{h}_a(\mathbf{r}) + \mathbf{j}_a(\mathbf{r}) \tag{2.1b}$$

Consider also the 'b' waveguide sustaining the 'b' fields generated by the 'b' sources.

$$\boldsymbol{\nabla} \times \mathbf{e}_b(\mathbf{r}) = -j\omega\underline{\mu}_b(\mathbf{r}) \cdot \mathbf{h}_b(\mathbf{r}) - j\omega\underline{\zeta}_b(\mathbf{r}) \cdot \mathbf{e}_b(\mathbf{r}) - \mathbf{k}_b(\mathbf{r}) \tag{2.2a}$$

$$\boldsymbol{\nabla} \times \mathbf{h}_b(\mathbf{r}) = j\omega\underline{\epsilon}_b(\mathbf{r}) \cdot \mathbf{e}_b(\mathbf{r}) + j\omega\underline{\xi}_b(\mathbf{r}) \cdot \mathbf{h}_b(\mathbf{r}) + \mathbf{j}_b(\mathbf{r}) \tag{2.2b}$$

Using the well-known identity $\boldsymbol{\nabla} \cdot (\mathbf{f}(\mathbf{r}) \times \mathbf{g}(\mathbf{r})) = \mathbf{g}(\mathbf{r}) \cdot (\boldsymbol{\nabla} \times \mathbf{f}(\mathbf{r})) - \mathbf{f}(\mathbf{r}) \cdot (\boldsymbol{\nabla} \times \mathbf{g}(\mathbf{r}))$ one can easily verify using (2.1a) through (2.2b) that[2]

$$\begin{aligned} &\boldsymbol{\nabla} \cdot (\mathbf{e}_a(\mathbf{r}) \times \mathbf{h}_b(\mathbf{r}) - \mathbf{e}_b(\mathbf{r}) \times \mathbf{h}_a(\mathbf{r})) = \\ &-j\omega\mathbf{e}_a(\mathbf{r}) \cdot (\underline{\epsilon}_b(\mathbf{r}) - \underline{\epsilon}_a{}^T(\mathbf{r})) \cdot \mathbf{e}_b(\mathbf{r}) + j\omega\mathbf{h}_a(\mathbf{r}) \cdot (\underline{\mu}_b(\mathbf{r}) - \underline{\mu}_a{}^T(\mathbf{r})) \cdot \mathbf{h}_b(\mathbf{r}) \\ &-j\omega\mathbf{e}_a(\mathbf{r}) \cdot (\underline{\xi}_b(\mathbf{r}) + \underline{\zeta}_a{}^T(\mathbf{r})) \cdot \mathbf{h}_b(\mathbf{r}) + j\omega\mathbf{h}_a(\mathbf{r}) \cdot (\underline{\zeta}_b(\mathbf{r}) + \underline{\xi}_a{}^T(\mathbf{r})) \cdot \mathbf{e}_b(\mathbf{r}) \\ &-\mathbf{e}_a(\mathbf{r}) \cdot \mathbf{j}_b(\mathbf{r}) + \mathbf{e}_b(\mathbf{r}) \cdot \mathbf{j}_a(\mathbf{r}) + \mathbf{h}_a(\mathbf{r}) \cdot \mathbf{k}_b(\mathbf{r}) - \mathbf{h}_b(\mathbf{r}) \cdot \mathbf{k}_a(\mathbf{r}) \end{aligned} \tag{2.3}$$

[1] throughout this work dyadic variables are single underlined

[2] Here one has to pay attention to the dyadic character of the medium parameters. As a result, interchanging the position of the vector fields in the left and right multiplication with a dyad (even though this is equivalent to taking the transpose of a scalar) must be accompanied by taking the transpose of the dyad.

If the materials of waveguide 'b' satisfy

$$\underline{\epsilon}_b(\mathbf{r}) = \underline{\epsilon}_a{}^T(\mathbf{r}) \tag{2.4a}$$
$$\underline{\mu}_b(\mathbf{r}) = \underline{\mu}_a{}^T(\mathbf{r}), \tag{2.4b}$$
$$\underline{\zeta}_b(\mathbf{r}) = -\underline{\xi}_a{}^T(\mathbf{r}), \tag{2.4c}$$
$$\underline{\xi}_b(\mathbf{r}) = -\underline{\zeta}_a{}^T(\mathbf{r}), \tag{2.4d}$$

then (2.3) reduces to

$$\begin{gathered}\boldsymbol{\nabla}\cdot(\mathbf{e}_a(\mathbf{r})\times\mathbf{h}_b(\mathbf{r})-\mathbf{e}_b(\mathbf{r})\times\mathbf{h}_a(\mathbf{r}))= \\ -\mathbf{e}_a(\mathbf{r})\cdot\mathbf{j}_b(\mathbf{r})+\mathbf{e}_b(\mathbf{r})\cdot\mathbf{j}_a(\mathbf{r})+\mathbf{h}_a(\mathbf{r})\cdot\mathbf{k}_b(\mathbf{r})-\mathbf{h}_b(\mathbf{r})\cdot\mathbf{k}_a(\mathbf{r})\end{gathered} \tag{2.5}$$

This very important result is known as the generalized differential form of the Lorentz reciprocity theorem in the frequency domain. It follows directly from Maxwell's equations in their most general form, and is thus the most natural and correct way to introduce orthogonality relationships between waveguide modes, as will be indicated later. The waveguide 'b' satisfying (2.4a) through (2.4d) is called the *adjoint waveguide*. If the adjoint waveguide coincides with the original waveguide, the waveguide is said to be reciprocal.
Normally the Lorentz theorem is seen in its integral form, obtained by integrating (2.5) over an arbitrary volume V with surface S. Using the Gauss divergence integral theorem, one gets

$$\begin{gathered}\iint_S(\mathbf{e}_a(\mathbf{r})\times\mathbf{h}_b(\mathbf{r})-\mathbf{e}_b(\mathbf{r})\times\mathbf{h}_a(\mathbf{r}))\cdot\mathbf{u}_n\mathrm{d}S= \\ \iiint_V[-\mathbf{e}_a(\mathbf{r})\cdot\mathbf{j}_b(\mathbf{r})+\mathbf{e}_b(\mathbf{r})\cdot\mathbf{j}_a(\mathbf{r})+\mathbf{h}_a(\mathbf{r})\cdot\mathbf{k}_b(\mathbf{r})-\mathbf{h}_b(\mathbf{r})\cdot\mathbf{k}_a(\mathbf{r})]\mathrm{d}V\end{gathered} \tag{2.6}$$

Calling a waveguide reciprocal when it is identical to its adjoint waveguide, is the mathematician's interpretation of reciprocity, but does not reveal many physical insights. The most encountered interpretation of (2.6) is based on the so-called Rayleigh-Carlson version of it [5].
Consider a reciprocal waveguide, so that in (2.6) medium 'a' and 'b' coincide, and choose the integration volume V such that the left-hand side of (2.6) vanishes. This can be achieved by considering the entire space as integration volume or by choosing a perfect electric or magnetic conductor as an artificial boundary of the computation window. In the former case all fields will tend to zero due to the radiation condition, in the latter case the boundary conditions on the conductor forces either the tangential[3] electric or the magnetic fields to become zero. The Rayleigh-Carlson version of the Lorentz reciprocity theorem thus becomes

$$\iiint_V[\mathbf{e}_a(\mathbf{r})\cdot\mathbf{j}_b(\mathbf{r})-\mathbf{h}_a(\mathbf{r})\cdot\mathbf{k}_b(\mathbf{r})]\mathrm{d}V=\iiint_V[\mathbf{e}_b(\mathbf{r})\cdot\mathbf{j}_a(\mathbf{r})-\mathbf{h}_b(\mathbf{r})\cdot\mathbf{k}_a(\mathbf{r})]\mathrm{d}V \tag{2.7}$$

[3] as $\mathbf{u}_n$ is the outward pointing normal unit vector to S, the left-hand side of (2.6) will only have tangential field contributions

The integrals in (2.7) can be defined as a sort of inner product of fields and sources. This inner product $\langle a, b \rangle$ is called the *reaction* of the electromagnetic fields 'a', generated by sources 'a', to another set of sources 'b', generating the electromagnetic fields 'b'. The Rayleigh-Carlson theorem thus states that in a reciprocal waveguide the *reaction* is symmetrical to its arguments[4], $\langle a, b \rangle = \langle b, a \rangle$. The experiment depicted in Fig. 2.1 visualizes the meaning of the Rayleigh-Carlson interpretation of reciprocity.

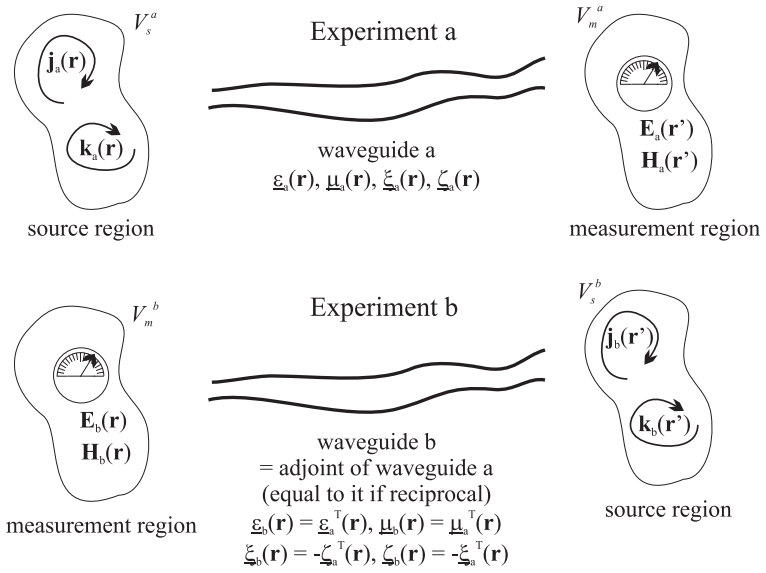


Fig. 2.1: Hypothetical experimental demonstration of the Rayleigh-Carlson reciprocity theorem.

Consider a reciprocal waveguide medium and conduct two experiments. In experiment 'a', a set of sources ($\mathbf{j}_a$, $\mathbf{k}_a$) is placed in a volume ${V_s}^a$ and the generated fields ($\mathbf{e}_a$, $\mathbf{h}_a$) are measured in a volume ${V_m}^a$. In experiment 'b', a set of sources ($\mathbf{j}_b$, $\mathbf{k}_b$) is placed in a volume ${V_s}^b$ of the adjoint medium, coinciding with the measurement volume of experiment 'a', and the generated fields ($\mathbf{e}_b$, $\mathbf{h}_b$) are measured in a volume ${V_m}^b$ of the adjoint medium, coinciding with the generating volume of experiment 'a'. If the medium is reciprocal, both experiments are conducted in the same medium. If not, they are conducted in the same physical space but once with the original media parameters and once with the adjoint media parameters (but with the same spatial dependencies). Assume that the sources for both measurements have identical local amplitude and phase distribution, that is all considered volumes are isomorphic and[5] $\mathbf{j}_a(\mathbf{r}) = \mathbf{j}_b(\mathbf{r}') = \mathbf{j}$, $\mathbf{k}_a(\mathbf{r}) = \mathbf{k}_b(\mathbf{r}') = \mathbf{k}$,

[4] for a non-reciprocal waveguide the symmetry of the reaction also holds, but now the reaction of the fields 'a' is *mutually* symmetric with the sources 'b' in the adjoint waveguide

[5] $\mathbf{r}'$ are the coordinates of the volume ${V_s}^b$

then (2.7) reads

$$\iiint_V [\mathbf{e}_a(\mathbf{r}) \cdot \mathbf{j}(\mathbf{r}) - \mathbf{h}_a(\mathbf{r}) \cdot \mathbf{k}(\mathbf{r})] \mathrm{d}V = \iiint_V [\mathbf{e}_b(\mathbf{r}) \cdot \mathbf{j}(\mathbf{r}) - \mathbf{h}_b(\mathbf{r}) \cdot \mathbf{k}(\mathbf{r})] \mathrm{d}V \quad (2.8)$$

Since the sources and the volumes are chosen completely arbitrary, (2.8) implies that also both sets of fields have the same local amplitude and phase distribution, $\mathbf{e}_a = \mathbf{e}_b$ and $\mathbf{h}_a = \mathbf{h}_b$. This brings us to the most widespread and correct interpretation of the Lorentz reciprocity theorem.

In a reciprocal waveguide reaction (measurement data) is insensitive to the interchange of source and measurement locations. In a non-reciprocal waveguide reaction (measurement data) is mutually insensitive to the interchange of source and measurement locations in the adjoint waveguide.

This is a very important conclusion to this section. Contrary to popular belief and practice, the Lorentz reciprocity theorem doesn't state anything at all about the behavior of modes in waveguides. It is for instance often falsely heard that due to Lorentz reciprocity a reciprocal waveguide that sustains a certain guided mode in one direction of the waveguide will possess an "identical" guided mode in the opposite direction, with the same propagation constant and the same mode profile. The Lorentz reciprocity does not predict this. It is exactly on this point that the term reciprocity is so often confused with the terms bidirectionality and mirroring. These latter two properties do have something to see with eigenmode behavior of waveguides, reciprocity doesn't. In the next section this will be further elucidated.

2.1.2 Reciprocal, bidirectional and mirroring waveguides

A waveguide is called *reciprocal* if all materials are reciprocal. This means that the material dyadics satisfy[6]

$$\underline{\epsilon}(\boldsymbol{\rho}) = \underline{\epsilon}^T(\boldsymbol{\rho}), \quad \underline{\mu}(\boldsymbol{\rho}) = \underline{\mu}^T(\boldsymbol{\rho}), \quad \underline{\zeta}(\boldsymbol{\rho}) = -\underline{\xi}^T(\boldsymbol{\rho}) \quad (2.9)$$

A waveguide is called *non-reciprocal* as soon as one of its materials does not satisfy (2.9).
A waveguide is called *bidirectional*, if for each waveguide mode propagating in one direction there exists another mode propagating in the opposite direction with the same propagation constant.
A waveguide is called *mirroring*, if it is bidirectional and if the modal fields of each set of associated counterpropagating modes are each others mirror images with respect to a perfect conducting mirror.

The latter two properties thus make a statement about the modal behavior of a waveguide, the former doesn't. The difference between bidirectionality and mirroring is important. It is not because a waveguide sustains propagating modes in

[6] $\boldsymbol{\rho}$ is the transverse part of the coordinate $\mathbf{r}$, as it is silently assumed that all considered waveguides are uniform along their axis. For simplicity, this is assumed to be the z-axis.

opposite directions with the same (but with opposite sign) propagation constant, that there exists a relationship between the modal profiles of these eigenmodes. In the following paragraphs we will give a proof [1] that every reciprocal waveguide is bidirectional, but not necessarily mirroring. Then it will be indicated under what conditions a general waveguide can become both bidirectional and mirroring, but thus not necessarily reciprocal.

Consider an eigenmode in a general (bianisotropic) waveguide with complex propagation constant[7] $\gamma_i = j\beta_i$ and described by the electromagnetic fields (in the frequency domain)

$$\mathbf{e}_i(\mathbf{r}) = \exp(-\gamma_i z)(\mathbf{E}_{t,i}(\boldsymbol{\rho}) + E_{z,i}(\boldsymbol{\rho})\mathbf{u}_z) \tag{2.10a}$$
$$\mathbf{h}_i(\mathbf{r}) = \exp(-\gamma_i z)(\mathbf{H}_{t,i}(\boldsymbol{\rho}) + H_{z,i}(\boldsymbol{\rho})\mathbf{u}_z) \tag{2.10b}$$

where the modal profiles are decomposed in their transverse (in the cross-section of the waveguide profile) and longitudinal parts.
Also consider an eigenmode $\tilde{\gamma}_j$, $\tilde{\mathbf{E}}_{t,j}(\boldsymbol{\rho})$, $\tilde{E}_{z,j}(\boldsymbol{\rho})$, $\tilde{\mathbf{H}}_{t,j}(\boldsymbol{\rho})$ and $\tilde{H}_{z,j}(\boldsymbol{\rho})$ of the adjoint waveguide. In an arbitrary piece of length L of this waveguide between the cross-sections at $z = 0$ and $z = L$, the integral form of the generalized Lorentz reciprocity (2.6) reads[8]

$$\big(\exp(-(\gamma_i+\tilde{\gamma}_j)L)-1\big)\iint_S [\mathbf{E}_{t,i}(\boldsymbol{\rho})\times\tilde{\mathbf{H}}_{t,j}(\boldsymbol{\rho})-\tilde{\mathbf{E}}_{t,j}(\boldsymbol{\rho})\times\mathbf{H}_{t,i}(\boldsymbol{\rho})]\cdot\mathbf{u}_z \mathrm{d}S = 0 \tag{2.11}$$

To proof that every reciprocal waveguide is bidirectional, we will start by considering an isotropic waveguide, for which it is common knowledge that this waveguide type is bidirectional and mirroring. I.e. for each eigensolution in an isotropic waveguide $\gamma_{i,0}$, $\mathbf{E}_{t,i,0}(\boldsymbol{\rho})$,$E_{z,i,0}(\boldsymbol{\rho})$, $\mathbf{H}_{t,i,0}(\boldsymbol{\rho})$ and $H_{z,i,0}(\boldsymbol{\rho})$ there exists a mirror solution in the same medium $\tilde{\gamma}_{j,0} = -\gamma_{i,0}$, $\tilde{\mathbf{E}}_{t,j,0}(\boldsymbol{\rho}) = \mathbf{E}_{t,i,0}(\boldsymbol{\rho})$, $\tilde{E}_{z,j,0}(\boldsymbol{\rho}) = -E_{z,i,0}(\boldsymbol{\rho})$, $\tilde{\mathbf{H}}_{t,j,0}(\boldsymbol{\rho}) = -\mathbf{H}_{t,i,0}(\boldsymbol{\rho})$ and $\tilde{H}_{z,j,0}(\boldsymbol{\rho}) = H_{z,i,0}(\boldsymbol{\rho})$. This second solution can also be seen as an eigenmode solution in the adjoint waveguide, since an isotropic waveguide is of course reciprocal and thus its adjoint coincides with itself. Insertion of both solutions in (2.11) yields the well-known normalization condition

$$\frac{1}{2}\iint_S [\mathbf{E}_{t,i}(\boldsymbol{\rho}) \times \mathbf{H}_{t,i}(\boldsymbol{\rho})] \cdot \mathbf{u}_z \mathrm{d}S = 1 \tag{2.12}$$

If the material parameters are changed slightly so that the waveguide becomes (slightly) bianisotropic (and so does the adjoint waveguide), then the propagations constants and modal fields in both original and adjoint waveguide will change by

[7] the usual $\exp(j\omega t)$ time dependence is chosen throughout this work. More fundamentally this is a natural consequence of the time dependence of the Fourier integral transformation needed to construct the frequency domain version of Maxwell's equations.

[8] because we are considering eigenmodes, the sources are of course zero

a small amount

$$\gamma_{i,1} = \gamma_{i,0} + \delta\gamma_{i,1} \tag{2.13a}$$
$$\mathbf{E}_{t,i,1}(\boldsymbol{\rho}) = \mathbf{E}_{t,i,0}(\boldsymbol{\rho}) + \delta\mathbf{E}_{t,i,1}(\boldsymbol{\rho}) \tag{2.13b}$$
$$\mathbf{H}_{t,i,1}(\boldsymbol{\rho}) = \mathbf{H}_{t,i,0}(\boldsymbol{\rho}) + \delta\mathbf{H}_{t,i,1}(\boldsymbol{\rho}) \tag{2.13c}$$
$$\tilde{\gamma}_{j,1} = -\gamma_{i,0} + \delta\tilde{\gamma}_{j,1} \tag{2.13d}$$
$$\tilde{\mathbf{E}}_{t,j,1}(\boldsymbol{\rho}) = \mathbf{E}_{t,i,0}(\boldsymbol{\rho}) + \delta\tilde{\mathbf{E}}_{t,j,1}(\boldsymbol{\rho}) \tag{2.13e}$$
$$\tilde{\mathbf{H}}_{t,j,1}(\boldsymbol{\rho}) = -\mathbf{H}_{t,i,0}(\boldsymbol{\rho}) + \delta\tilde{\mathbf{H}}_{t,j,1}(\boldsymbol{\rho}) \tag{2.13f}$$

Substituting these perturbed solutions in (2.11) yields

$$\begin{aligned}
&\big(\exp[-(\delta\gamma_{i,1} + \delta\tilde{\gamma}_{j,1})L] - 1\big) \\
&\Big(-4 - \iint_S [\delta\mathbf{E}_{t,i,1}(\boldsymbol{\rho}) \times \mathbf{H}_{t,i,0}(\boldsymbol{\rho}) + \mathbf{E}_{t,i,0}(\boldsymbol{\rho}) \times \mathbf{H}_{t,i,1}(\boldsymbol{\rho})] \cdot \mathbf{u}_z \mathrm{d}S \\
&\quad + \iint_S [\mathbf{E}_{t,i,0}(\boldsymbol{\rho}) \times \delta\tilde{\mathbf{H}}_{t,j,1}(\boldsymbol{\rho}) - \delta\tilde{\mathbf{E}}_{t,j,1}(\boldsymbol{\rho}) \times \mathbf{H}_{t,i,0}(\boldsymbol{\rho})] \cdot \mathbf{u}_z \mathrm{d}S \\
&\quad + \iint_S [\delta\mathbf{E}_{t,i,1}(\boldsymbol{\rho}) \times \delta\tilde{\mathbf{H}}_{t,j,1}(\boldsymbol{\rho}) - \delta\tilde{\mathbf{E}}_{t,j,1}(\boldsymbol{\rho}) \times \delta\mathbf{H}_{t,i,1}(\boldsymbol{\rho})] \cdot \mathbf{u}_z \mathrm{d}S\Big) = 0
\end{aligned} \tag{2.14}$$

The second factor, apart from the -4, contains only terms of at least first order in the change in the material parameters. As a result these terms will never be able to cancel out the zero order term -4, and (2.14) will only be fulfilled if the first factor becomes 0, or in other words, if $\delta\gamma_{i,1} = -\delta\tilde{\gamma}_{j,1}$ or if $\gamma_{i,1} = -\tilde{\gamma}_{j,1}$. Thus, we have constructed out of two mirroring eigenmodes in a reciprocal, isotropic waveguide two new eigenmodes in a slightly bianisotropic, not necessarily reciprocal waveguide. One of these two now belongs to the eigenspectrum of the new bianisotropic waveguide, and one belongs to that of its adjoint. And it is proven that both modes have exact opposite propagation constants. Normalizing these new modes along (2.11), the above perturbation procedure can be repeated over and over again, so as to generate any general bianisotropic waveguide. In each step m of this perturbation series $\gamma_{i,m} = -\tilde{\gamma}_{j,m}$.
This proves that every bianisotropic waveguide is *mutually* bidirectional with its adjoint waveguide, or, in particular, that *every reciprocal waveguide is bidirectional.*

The above result is very important for non-reciprocal waveguides, as it shows that once an eigenmode is obtained in such waveguide, automatically one has also found a corresponding mode in the adjoint waveguide, for free. However, the only property that is known for this mode of the adjoint waveguide is that it has the exact opposite complex value for the propagation constant γ. There is in general no special relationship between the profiles of these modes. That is quite annoying, because it is clear from (2.11) that in order to correctly orthonormalize the eigenspectrum of a non-reciprocal waveguide, the exact mode profiles of the eigenmodes of both the original waveguide and the adjoint waveguide need to be known. The computational effort of correctly solving a non-reciprocal waveguide[9]

[9] i.e. building an orthonormalized set of eigenmodes for instance to calculate the forward

not only gets burdened by the "complex" nature of the materials involved but also increases due to the obligatory calculation of two modal spectra.

However, there are certain conditions [1] under which a non-reciprocal waveguide is bidirectional and mirroring with itself. A fortiori, the same will apply for the adjoint waveguide. But even when a (bi-)anisotropic waveguide is bidirectional and mirroring with itself, no prediction can be made about any kind of relationship between the modal profiles of the original waveguide and the adjoint waveguide[10]. The most ideal situation would be that the (bi-)anisotropic waveguide is mutually mirroring with its adjoint. In that case, the eigenmode profiles of the adjoint waveguide do not have to be calculated, since they directly follow from the modal profiles of the original waveguide.
The modal fields in a general bianisotropic waveguide satisfy the following curl equations:

$$\begin{aligned}\boldsymbol{\nabla}_t \times \mathbf{E}_t(\boldsymbol{\rho}) = &- j\omega\mu_{zz}(\boldsymbol{\rho})H_z(\boldsymbol{\rho})\mathbf{u}_z - j\omega\boldsymbol{\mu}_{zt}(\boldsymbol{\rho})\cdot\mathbf{H}_t(\boldsymbol{\rho})\mathbf{u}_z \\ &- j\omega\zeta_{zz}(\boldsymbol{\rho})E_z(\boldsymbol{\rho})\mathbf{u}_z - j\omega\boldsymbol{\zeta}_{zt}(\boldsymbol{\rho})\cdot\mathbf{E}_t(\boldsymbol{\rho})\mathbf{u}_z\end{aligned} \tag{2.15a}$$

$$\begin{aligned}-\gamma\mathbf{u}_z \times \mathbf{E}_t(\boldsymbol{\rho}) + \boldsymbol{\nabla}_t \times E_z(\boldsymbol{\rho})\mathbf{u}_z = &- j\omega\underline{\mu}_{tt}(\boldsymbol{\rho})\cdot\mathbf{H}_t(\boldsymbol{\rho}) - j\omega\boldsymbol{\mu}_{tz}(\boldsymbol{\rho})H_z(\boldsymbol{\rho}) \\ &- j\omega\underline{\zeta}_{tt}(\boldsymbol{\rho})\cdot\mathbf{E}_t(\boldsymbol{\rho}) - j\omega\boldsymbol{\zeta}_{tz}(\boldsymbol{\rho})E_z(\boldsymbol{\rho})\end{aligned} \tag{2.15b}$$

$$\begin{aligned}\boldsymbol{\nabla}_t \times \mathbf{H}_t(\boldsymbol{\rho}) = &\, j\omega\epsilon_{zz}(\boldsymbol{\rho})E_z(\boldsymbol{\rho})\mathbf{u}_z + j\omega\boldsymbol{\epsilon}_{zt}(\boldsymbol{\rho})\cdot\mathbf{E}_t(\boldsymbol{\rho})\mathbf{u}_z \\ &+ j\omega\xi_{zz}(\boldsymbol{\rho})H_z(\boldsymbol{\rho})\mathbf{u}_z + j\omega\boldsymbol{\xi}_{zt}(\boldsymbol{\rho})\cdot\mathbf{H}_t(\boldsymbol{\rho})\mathbf{u}_z\end{aligned} \tag{2.15c}$$

$$\begin{aligned}-\gamma\mathbf{u}_z \times \mathbf{H}_t(\boldsymbol{\rho}) + \boldsymbol{\nabla}_t \times H_z(\boldsymbol{\rho})\mathbf{u}_z = &\, j\omega\underline{\epsilon}_{tt}(\boldsymbol{\rho})\cdot\mathbf{E}_t(\boldsymbol{\rho}) + j\omega\boldsymbol{\epsilon}_{tz}(\boldsymbol{\rho})E_z(\boldsymbol{\rho}) \\ &+ j\omega\underline{\xi}_{tt}(\boldsymbol{\rho})\cdot\mathbf{H}_t(\boldsymbol{\rho}) + j\omega\boldsymbol{\xi}_{tz}(\boldsymbol{\rho})H_z(\boldsymbol{\rho})\end{aligned} \tag{2.15d}$$

Suppose that γ, $\mathbf{E}_t(\boldsymbol{\rho})$, $E_z(\boldsymbol{\rho})$, $\mathbf{H}_t(\boldsymbol{\rho})$ and $H_z(\boldsymbol{\rho})$ is an eigenmode solution of this waveguide, then the equations (2.15) prove that the mirror mode $\hat{\gamma} = -\gamma$, $\hat{\mathbf{E}}_t(\boldsymbol{\rho}) = \mathbf{E}_t(\boldsymbol{\rho})$, $\hat{E}_z(\boldsymbol{\rho}) = -E_z(\boldsymbol{\rho})$, $\hat{\mathbf{H}}_t(\boldsymbol{\rho}) = -\mathbf{H}_t(\boldsymbol{\rho})$ and $\hat{H}_z(\boldsymbol{\rho}) = H_z(\boldsymbol{\rho})$ will be a solution

and backward scattering matrices when considering the propagation of light through a photonic integrated circuit containing pieces of non-reciprocal waveguide and junctions between reciprocal and non-reciprocal waveguide.

[10] this is not completely correct. It is proven in [1] that a linear operator relationship can be found using some very cumbersome dyadic operator algebra. But even then, the solution of the needed dyadic operator equations yielding the linear operators that fix this relationship, will always be approximate since in practice only a finite number of modes can be considered.

of the waveguide with the following material parameters

$$\hat{\epsilon}_{zz}(\boldsymbol{\rho}) = \epsilon_{zz}(\boldsymbol{\rho}) \qquad \hat{\boldsymbol{\epsilon}}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\epsilon}_{zt}(\boldsymbol{\rho}) \tag{2.16a}$$
$$\hat{\boldsymbol{\epsilon}}_{tz}(\boldsymbol{\rho}) = -\boldsymbol{\epsilon}_{tz}(\boldsymbol{\rho}) \qquad \hat{\underline{\epsilon}}_{tt}(\boldsymbol{\rho}) = \underline{\epsilon}_{tt}(\boldsymbol{\rho}) \tag{2.16b}$$
$$\hat{\mu}_{zz}(\boldsymbol{\rho}) = \mu_{zz}(\boldsymbol{\rho}) \qquad \hat{\boldsymbol{\mu}}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\mu}_{zt}(\boldsymbol{\rho}) \tag{2.16c}$$
$$\hat{\boldsymbol{\mu}}_{tz}(\boldsymbol{\rho}) = -\boldsymbol{\mu}_{tz}(\boldsymbol{\rho}) \qquad \hat{\underline{\mu}}_{tt}(\boldsymbol{\rho}) = \underline{\mu}_{tt}(\boldsymbol{\rho}) \tag{2.16d}$$
$$\hat{\xi}_{zz}(\boldsymbol{\rho}) = -\xi_{zz}(\boldsymbol{\rho}) \qquad \hat{\boldsymbol{\xi}}_{zt}(\boldsymbol{\rho}) = \boldsymbol{\xi}_{zt}(\boldsymbol{\rho}) \tag{2.16e}$$
$$\hat{\boldsymbol{\xi}}_{tz}(\boldsymbol{\rho}) = \boldsymbol{\xi}_{tz}(\boldsymbol{\rho}) \qquad \hat{\underline{\xi}}_{tt}(\boldsymbol{\rho}) = -\underline{\xi}_{tt}(\boldsymbol{\rho}) \tag{2.16f}$$
$$\hat{\zeta}_{zz}(\boldsymbol{\rho}) = -\zeta_{zz}(\boldsymbol{\rho}) \qquad \hat{\boldsymbol{\zeta}}_{zt}(\boldsymbol{\rho}) = \boldsymbol{\zeta}_{zt}(\boldsymbol{\rho}) \tag{2.16g}$$
$$\hat{\boldsymbol{\zeta}}_{tz}(\boldsymbol{\rho}) = \boldsymbol{\zeta}_{tz}(\boldsymbol{\rho}) \qquad \hat{\underline{\zeta}}_{tt}(\boldsymbol{\rho}) = -\underline{\zeta}_{tt}(\boldsymbol{\rho}) \tag{2.16h}$$

This waveguide is called the mutually mirroring waveguide.

(2.16) shows that the mutually mirroring waveguide will coincide with the original waveguide if

$$\boldsymbol{\epsilon}_{zt}(\boldsymbol{\rho}) = \boldsymbol{\epsilon}_{tz}(\boldsymbol{\rho}) = 0 \qquad \boldsymbol{\mu}_{zt}(\boldsymbol{\rho}) = \boldsymbol{\mu}_{tz}(\boldsymbol{\rho}) = 0 \tag{2.17a}$$
$$\zeta_{zz}(\boldsymbol{\rho}) = \xi_{zz}(\boldsymbol{\rho}) = 0 \qquad \underline{\zeta}_{tt}(\boldsymbol{\rho}) = \underline{\xi}_{tt}(\boldsymbol{\rho}) = 0 \tag{2.17b}$$

and the other material dyadics are arbitrary.
(2.16) shows that the mutually mirroring waveguide will coincide with the adjoint waveguide if

$$\boldsymbol{\epsilon}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\epsilon}_{tz}(\boldsymbol{\rho}) \qquad \boldsymbol{\mu}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\mu}_{tz}(\boldsymbol{\rho}) \tag{2.18a}$$
$$\underline{\epsilon}_{tt}{}^{T}(\boldsymbol{\rho}) = \underline{\epsilon}_{tt}(\boldsymbol{\rho}) \qquad \underline{\mu}_{tt}{}^{T}(\boldsymbol{\rho}) = \underline{\mu}_{tt}(\boldsymbol{\rho}) \tag{2.18b}$$
$$\zeta_{zz}(\boldsymbol{\rho}) = \xi_{zz}(\boldsymbol{\rho}) \qquad \underline{\zeta}_{tt}(\boldsymbol{\rho}) = \underline{\xi}_{tt}{}^{T}(\boldsymbol{\rho}) \tag{2.18c}$$
$$\boldsymbol{\zeta}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\xi}_{tz}(\boldsymbol{\rho}) \qquad \boldsymbol{\xi}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\zeta}_{tz}(\boldsymbol{\rho}) \tag{2.18d}$$

If one takes the cross-section of the conditions in (2.17) and (2.18), then it is clear that a waveguide can only be mirroring with both itself and its adjoint if

$$\boldsymbol{\epsilon}_{zt}(\boldsymbol{\rho}) = \boldsymbol{\epsilon}_{tz}(\boldsymbol{\rho}) = 0 \qquad \boldsymbol{\mu}_{zt}(\boldsymbol{\rho}) = \boldsymbol{\mu}_{tz}(\boldsymbol{\rho}) = 0 \tag{2.19a}$$
$$\underline{\epsilon}_{tt}{}^{T}(\boldsymbol{\rho}) = \underline{\epsilon}_{tt}(\boldsymbol{\rho}) \qquad \underline{\mu}_{tt}{}^{T}(\boldsymbol{\rho}) = \underline{\mu}_{tt}(\boldsymbol{\rho}) \tag{2.19b}$$
$$\zeta_{zz}(\boldsymbol{\rho}) = \xi_{zz}(\boldsymbol{\rho}) = 0 \qquad \underline{\zeta}_{tt}(\boldsymbol{\rho}) = \underline{\xi}_{tt}(\boldsymbol{\rho}) = 0 \tag{2.19c}$$
$$\boldsymbol{\zeta}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\xi}_{tz}(\boldsymbol{\rho}) \qquad \boldsymbol{\xi}_{zt}(\boldsymbol{\rho}) = -\boldsymbol{\zeta}_{tz}(\boldsymbol{\rho}) \tag{2.19d}$$

This is clearly only possible for a certain class of reciprocal media. Thus, a non-reciprocal waveguide can never be at the same time mirroring with itself and with its adjoint.

All of the above illustrates that the arduous task of modelling non-reciprocal waveguides (in particular the need to solve both the original and the adjoint

waveguide for a proper orthonormalization of the eigenspectrum) can in many practical cases be relieved when some attention is paid to spatial "symmetries"[11] present in the material parameters describing the waveguide.

2.2 Modal solutions of non-reciprocal waveguides

2.2.1 Green's functions

From this point on we will concentrate on non-reciprocal waveguides, operated at optical frequencies and where the non-reciprocity is not caused by the presence of the material dyadics $\underline{\zeta}$ and $\underline{\xi}$. As was indicated in section 1.2.2.1, the former condition results for most practical materials in a scalar magnetic permeability equal to the vacuum magnetic permeability ($\underline{\mu} = \mu_0 \underline{I}$ with $\mu_0 = 4\pi \times 10^{-7}\ \frac{\mathrm{Hy}}{\mathrm{m}}$ in SI units). The latter condition assumes that the waveguide media or the geometry of the waveguide do not exhibit handedness, leading to optical activity[12].

For such waveguides the Maxwell curl equations read

$$\boldsymbol{\nabla} \times \mathbf{e}(\mathbf{r}) = -j\omega\mu_0 \underline{I} \cdot \mathbf{h}(\mathbf{r}) \tag{2.20a}$$

$$\boldsymbol{\nabla} \times \mathbf{h}(\mathbf{r}) = j\omega \underline{\epsilon}(\boldsymbol{\rho}) \cdot \mathbf{e}(\mathbf{r}) + \mathbf{j}(\mathbf{r}) \tag{2.20b}$$

Here $\underline{I}$ is the three-dimensional unit dyadic and for the moment the eigenmode propagator factor $\exp(-\gamma z)$ is included in the electromagnetic fields. Eliminating $\mathbf{h}(\mathbf{r})$ between (2.20a) and (2.20b) leads to

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{e}(\mathbf{r})) = \omega^2 \mu_0 \underline{\epsilon}(\boldsymbol{\rho}) \cdot \mathbf{e}(\mathbf{r}) - j\mu_0\omega \mathbf{j}(\mathbf{r}) \tag{2.21}$$

Using the well-known Levi-Civita representation of the cross product, any cross product of two vectors can be written as

$$(\mathbf{a} \times \mathbf{b})_i = \epsilon_{ijk} a_j b_k = \epsilon_{ijk} a_j \delta_{kl} b_l = ((\mathbf{a} \times \underline{I}) \cdot \mathbf{b})_i \tag{2.22}$$

where the Kronecker delta has been introduced as a representation for the unit dyadic. Thus, the left-hand side of (2.21) can be written as $\frac{1}{\mu_0}\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{e}(\mathbf{r})) = \frac{1}{\mu_0}(\boldsymbol{\nabla} \times \underline{I}) \cdot (\boldsymbol{\nabla} \times \mathbf{e}(\mathbf{r})) = \frac{1}{\mu_0}(\boldsymbol{\nabla} \times \underline{I}) \cdot (\boldsymbol{\nabla} \times \underline{I}) \cdot \mathbf{e}(\mathbf{r})$. Introducing the linear dyadic operator $\underline{L}(\boldsymbol{\nabla}, \boldsymbol{\rho}) = \frac{1}{\mu_0}(\boldsymbol{\nabla} \times \underline{I}) \cdot (\boldsymbol{\nabla} \times \underline{I}) - \omega^2 \underline{\epsilon}(\boldsymbol{\rho})$, (2.21) can be compactly written as

$$\underline{L}(\boldsymbol{\nabla}, \boldsymbol{\rho}) \cdot \mathbf{e}(\mathbf{r}) = -j\omega \mathbf{j}(\mathbf{r}) \tag{2.23}$$

[11] ... and on top of that, we still haven't mentioned all possible bidirectional and/or mirroring behaviors due to geometrical symmetries in the waveguide cross-section. For a detailed account of these the reader is referred to [6]

[12] in order to be completely correct, this latter assumption also excludes the presence of Tellegen media, for which the chirality or the optical activity has nothing to see with a certain handedness, but with a non-electromagnetically caused "bond"between electric and magnetic dipoles. This can for instance be found in certain antiferromagnetic materials [7]. On top of that, we have to be aware that bianisotropic behavior is a natural consequence of the essential Lorentz invariancy of Maxwell's equations [8]. In this way, any moving medium will in its rest-frame exhibit bi-anisotropic consitutive relationships. In other words, the latter condition also assumes the waveguide to be "at the rest", at least, compared to the speed of light.

In order to solve the set of three coupled partial differential equations of (2.23) for arbitrary current sources $\mathbf{j}(\mathbf{r})$, one commonly makes use of the electric-electric Green's dyadic, $\underline{G}_{ee}(\mathbf{r})$,[13] satisfying

$$\underline{L}(\boldsymbol{\nabla}, \boldsymbol{\rho}) \cdot \underline{G}_{ee}(\boldsymbol{\rho}|\boldsymbol{\rho}', z - z') = -\delta(\mathbf{r} - \mathbf{r}')\underline{I} \tag{2.24}$$

The solution of (2.23) can then be written as

$$\mathbf{e}(\mathbf{r}) = j\omega \iiint_V \underline{G}_{ee}(\boldsymbol{\rho}|\boldsymbol{\rho}', z - z') \cdot \mathbf{j}(\mathbf{r}')\mathrm{d}V' \tag{2.25}$$

It is quite annoying that solving (2.24) requires a solution for not less then 9 components! This cannot be avoided. Only when the considered waveguide is homogeneous (or in other words when the waveguide cross-section doesn't depend on the transversal coordinate $\boldsymbol{\rho}$), it is possible to rewrite (2.24) using only a single <u>scalar</u> Green's function (instead of a 9-component dyadic) [9].

$$\det\underline{L}(\boldsymbol{\nabla})G(\mathbf{r} - \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}') \tag{2.26}$$

Because the inverse of a dyadic operator is calculated in the same way as the inverse of a matrix

$$\underline{L}^{-1} = \frac{1}{\det\underline{L}}[\underline{L}^A]^T \qquad \text{(where the adjoint operator is introduced)} \tag{2.27}$$

the dyadic solution of (2.24) can be deduced out of the scalar Green's function satisfying (2.26) by

$$\underline{G}_{ee}(\mathbf{r} - \mathbf{r}') = [\underline{L}^A(\boldsymbol{\nabla})]^T G(\mathbf{r} - \mathbf{r}') \tag{2.28}$$

Solving (2.26) is normally done in the spectral Fourier domain [1]. It is not so difficult to see that the Helmholtz determinant operator in (2.26) is a fourth-order operator in the components of $\boldsymbol{\nabla}$. This implies that even for a homogeneous medium[14], the inverse Fourier transform will run into great mathematical problems. The reason for this is quite simple. The determinant operator $\det\underline{L}(\boldsymbol{\nabla})$ will, due to its fourth-order polynomial character in $\boldsymbol{\beta}$-space, introduce four discrete poles for any one β variable in $G(\beta_x, \beta_y, \beta)$.[15] As a result the inverse transform in β_x (or β_y for that matter) is easy to calculate using Cauchy's residue theorem. However, the second inversion becomes very difficult. This is because after applying the Cauchy theorem in the first inversion, the result contains expressions like $\beta_{x,i}(\beta_y, \beta)$, describing how the β_x poles behave as a function of the other β

[13] In electromagnetics one distinguishes between $\underline{G}_{ee}$, $\underline{G}_{me}$, $\underline{G}_{em}$ and $\underline{G}_{mm}$, depending on whether one solves for the electric fields generated by electric current sources, for the electric fields generated by magnetic current sources, magnetic fields generated by electric current sources, or magnetic fields generated by magnetic current sources. There is obviously a very strong duality connection between them, so one normally solves only for $\underline{G}_{ee}$ or $\underline{G}_{me}$ as the Maxwell rotor equations allow to derive one field out of the other

[14] for which <u>all three</u> components of the nabla operator in the determinant operator can be transformed into complex Fourier β-variables, because the medium has no ρ-dependence

[15] the z component of β is not indexed because it of course represents the standard propagation constant

components. Thus, these expressions are inversions of fourth order polynomials. Even though this is analytically solvable using Cardano's method [10], $\beta_{x,i}$ will be a strong irrational function of β_y and β_z. This means that in the complex β_y-plane for the second inversion these expressions will generate very complex branch-cuts, whose "form" cannot be calculated analytically. This would make in general the second (and a fortiori the third) inversion very hard. Olyslager has proven however that by using spherical coordinates at least the second inversion is performed quasi automatically [11].

For a waveguide (instead of a homogeneous medium) things get even worse, as in this case only a single Fourier transform of (2.26) will be possible, due to the $\boldsymbol{\rho}$-dependence of the material parameters. On top of that, the above reformulation in terms of a scalar instead of a dyadic Green's function, is not possible in waveguide. This is because the determinant operator would contain both coordinates and derivatives with respect to these coordinates. In other words it would be composed out of operators that do not commute, and it can therefore not be defined. As a result, the Fourier transform apparently does not really help in simplifying the formulation of the partial differential equation of (2.26). All that is achieved is a reduction of the original three-dimensional Helmholtz equation to a two-dimensional one.

All of the above might give the wrong impression that a Green's function is really nothing more than a theoretical interesting concept, but has no direct practical value for the modelling of (anisotropic) waveguides. This is only partly correct. Of course it is nigh impossible to calculate a closed form of the Green's function of an arbitrary two- or three-dimensional waveguide. However, for waveguides this is not the main value of the Green's function technique. The real importance of the waveguide Green's function lays in the fact that it allows, due to its definition, to derive *any* electromagnetic field (generated by *any* distribution of sources) that is allowed as a solution of the Maxwell equations for the waveguide under consideration. In other words, it is a kind of "completeness projection operator"! This implies that the poles and the branch-cuts of the Green's function $G(\boldsymbol{\rho}|\boldsymbol{\rho}',\beta)$ will, in the inverse Fourier transform,

$$\begin{aligned} G(\boldsymbol{\rho}|\boldsymbol{\rho}',z-z') &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} G(\boldsymbol{\rho}|\boldsymbol{\rho}',\beta)e^{-j\beta(z-z')}\mathrm{d}\beta \\ &= -j\sum_i \mathrm{Res}_{\beta=\beta_i}[G(\boldsymbol{\rho}|\boldsymbol{\rho}',\beta)]e^{-j\beta_i(z-z')} \\ &\quad -\sum_{c_k}\frac{1}{2\pi}\int_{c_k} G(\boldsymbol{\rho}|\boldsymbol{\rho}',\beta)e^{-j\beta(z-z')}\mathrm{d}\beta \end{aligned} \tag{2.29}$$

generate components that form a *complete* basis of functions in which any electromagnetic field, that can be sustained in the waveguide, can be expanded. Furthermore, it can be proven [1] that the sources needed to generate any one of the components of the Green's function resulting from these poles and infinitesimal parts of the branch-cuts, are zero. In other words, these components are eigenmodes of the waveguide and, because of the completeness property of the Green's function, they form a complete set of functions. Otherwise put, locating all the

poles and the branch-cuts of $G(\boldsymbol{\rho}|\boldsymbol{\rho}', \beta)$ will generate the complete set of eigenvalues of the modal spectrum of the waveguide. The discrete poles correspond of course to the guided modes, while the continuum of modes on the branch-cuts are the radiation modes. It is also interesting to note that this construction immediately solves the often encountered (but unnecessary) controversy about the character of the leaky mode solutions to the Maxwell equations. They are indeed possible solutions of the waveguide problem, but are not a part of the complete set of eigenfunctions (or are not needed to make this set complete). They simply correspond to zeros that are on another sheet of the Riemann-surface, defined by the branch-cuts. This illustrates how the choice of a suitable radiation condition defines the "shape" and location of the branch-cuts, and thus makes the distinction between a normal mode and a leaky mode.

The above clarifies the value of calculating the spectral domain Green's function, $G(\boldsymbol{\rho}|\boldsymbol{\rho}', \beta)$. Even though calculating its inverse is more often than not a tremendously arduous mathematical task, the same doesn't hold for locating its poles and branch-cuts in β-space. And the information provided by this, is nothing less than the complete eigenspectrum of the waveguide. However, even though the added-value of calculating $G(\boldsymbol{\rho}|\boldsymbol{\rho}', \beta)$ is large, there are a few problems involved. Calculating $G(\boldsymbol{\rho}|\boldsymbol{\rho}', \beta)$ for an arbitrary waveguide has its pro's and con's. The main advantage is that (2.26) is a scalar equation and is thus much easier to solve than the set of coupled partial differential equations (2.23). On the other hand, however, it is not so obvious what boundary conditions need to be used for the scalar Green's function at the different medium interfaces in the waveguide. It is indeed not straightforward to establish the connection between the scalar Green's function in (2.26) and the different electromagnetic field components, for which the boundary conditions are well known. This is a rather strong contra argument. And it is indeed a fact that for waveguides a little more complex than the average isotropic waveguide, it is quite hard to calculate the scalar $G(\boldsymbol{\rho}|\boldsymbol{\rho}', \beta)$. In the case of *planar* (bi-)anisotropic waveguides, there are some techniques where it is possible to calculate the eigenspectrum of the waveguide out of the set of coupled differential field equations (2.23) (instead of the scalar equation (2.26)) without too much extra mathematical effort. The most popular of these techniques will be described in the next section, and illustrated by a simple example of a basic waveguide version of a Faraday rotator.

2.2.2 Transfer matrix formalism for planar non-reciprocal magneto-optic waveguides

2.2.2.1 Introduction

The transfer matrix formalism (TMF) finds its origins in the calculation of electromagnetic field scattering at layered structures. The composing elements of these layered structures can be quite general, and do not necessarily need to be stratified, homogeneous materials, but can for instance also be a piece of waveguide, for which the complete eigenspectrum is known. In the latter case, the transfer ma-

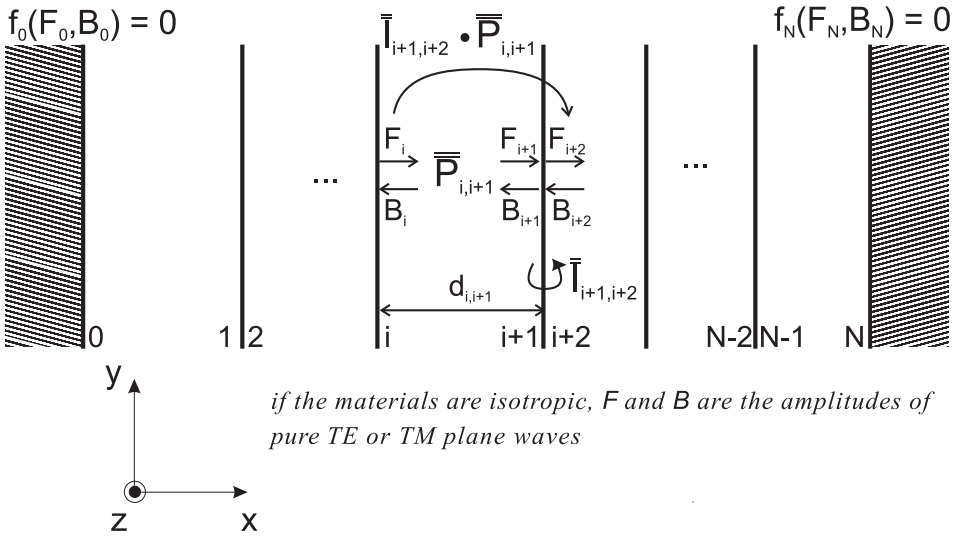


Fig. 2.2: A general layered structure. Nothing is assumed about the nature of the materials nor about the boundary conditions imposed by the walls.

trix formalism is better known under the name of 'wave mode matching' method (WMM) [12]. Here we will consider only the case of layered structures in the strict sense of the word, with the purpose of deriving the dispersion relationship of a planar layered waveguide. Consider the layered structure depicted in Fig. 2.2 for that purpose. Even though most waveguides encountered in opto-electronic IC's, are, strictly speaking, open structures, for obvious numerical purposes (such as the discretisation of the continuum of radiation modes,...) the structure in Fig. 2.2 is enclosed within two boundaries. For the moment their precise nature is not important.

The philosophy behind the TMF is to derive a matrix describing the connection between 'forward' and 'backward' propagating modes at both sides of the layered structure[16]. The term 'modes' has to be read with caution. Here the name refers to the local eigenmodes of each homogeneous layer and not to the eigenmode of the complete layered structure. These local eigenmodes are the solution of the Maxwell equation in a homogeneous medium with an assumed global $e^{-j\beta z}$ behavior in the z-direction, and are plane waves. As already indicated in 2.2.1, for each choice of β the Maxwell equation will have 4 solutions for the x-component of the wavevector, indicating that each of the local forward and backward x-propagating eigenwaves can come in two polarisation states. In the case of isotropic media these are the

[16]The naming 'forward' and 'backward' is in this context quite ambiguous, but is here to be understood as incoming or outgoing waves from an orthogonal viewpoint with respect to the layer interface

well-known TE and TM polarisation.

$$\begin{pmatrix} F_N \\ B_N \end{pmatrix} = \overline{\overline{T}} \cdot \begin{pmatrix} F_0 \\ B_0 \end{pmatrix} \tag{2.30}$$

The main, obvious advantage of the TMF for layered structures is that transfer matrices can easily be concatenated through matrix multiplication[17]. Therefore, it suffices to derive the transfer matrix of an arbitrary 'chunk', in order to derive the framework of the TMF formalism. Two basic transfer matrices have to be deduced for that purpose:

1. the matrix $P_{i,i+1}$, describing the forward and backward x-propagation of the local plane eigenwaves through a stratified layer of thickness $d_{i,i+1}$.

2. the matrix $I_{i+1,i+2}$, describing the connection between 'forward' and 'backward' propagating 'modes' at two sides of an interface between two stratified layers i and j, using the fundamental boundary conditions imposed by Maxwell's equations, i.e. the continuity of the field components tangential to the interface.

Note that for both matrices the only independent variables are the transverse wavevector components $k_{x,i}$, who in their turn are a function of the assumed global β. The total transfer matrix of the layered structure is then simply found as the matrix product of the transfer matrices of all composing chunks (ended by an extra propagation transfer matrix to describe the propagation through the super-(or sub-)strate up to the enclosing right-(or left-)hand boundary).

$$\overline{\overline{T}} = \overline{\overline{P}}_{N-1,N} \cdot \overline{\overline{I}}_{N-2,N-1} \cdot \ldots \overline{\overline{P}}_{2,3} \cdot \overline{\overline{I}}_{1,2} \cdot \overline{\overline{P}}_{0,1} \tag{2.31}$$

The boundary conditions will now impose a certain linear relationship between the forward and backward amplitudes at both ends of the structure, $f_{0|N}(F_{0|N}, B_{0|N}) = 0$. Choosing F and B such that the boundary relationship at one side is satisfied (leaving thus one degree freedom), subsequently calculating the amplitudes at the other boundary using (2.30), and finally substituting these in the boundary relationship valid at this side, will yield the dispersion relationship, whose zeros are the propagation constants of the eigenmodes of the (artificially) closed waveguide of Fig. 2.2.

Up till now, nothing has been assumed about the precise nature of the materials in the slab waveguide of Fig. 2.2. And even though, it is needless to say that this will influence the matrix form of the different transfer matrices, it still seems that the above treatment is quite general. However, that is obviously not correct. Indeed by formally choosing 2-by-2 matrices for the above used transfer matrices, it is in fact silently assumed that the layers of the slab waveguide in Fig. 2.2 leave the two base polarizations for the plane eigenwaves uncoupled. This is always

[17] it is quite interesting to note that viewed in this way the mapping function $T : \mathcal{L} \rightarrow \mathcal{M}$, that maps each possible layered structure onto its characteristic transfer matrix, is group homomorphism between the non-Abelian groups of layered structures, $\mathcal{L}$, and the transfer matrices, $\mathcal{M}$. It has to be noted in this context that a single interface in itself is of course also a group element of $\mathcal{L}$

the case if the materials are isotropic, but for anisotropic materials this is only exceptionally the case[18]. Therefore, even though the two base polarizations will propagate uncoupled through each layer (because they are orthogonal eigenwaves), it is far from clear that

1. the base polarizations will have the same 'character' in each layer of the stack

2. no coupling will occur between the two base polarizations at each interface, even when it concerns an interface between two 'similar' materials.

The latter point is very important, because the derivation of the interface transfer matrix is completely based on the Maxwell boundary conditions expressing the continuity of the tangential field components ($\mathbf{e}_{\|}$, $\mathbf{h}_{\|}$). If the base polarizations do not have a strict TE or TM character anymore, then these boundary conditions will inevitably cause a coupling between the 'forward' and 'backward' amplitudes of both base polarizations. Therefore, as soon as one of the materials in the planar, layered structure becomes just a little more complex than a simple isotropic material, the TMF necessitates the use of 4-by-4 matrices, being able to describe these coupling phenomena.

In the next section we will derive the TMF for arbitrary anisotropic slab waveguides. This derivation is described in extensive detail in [13]. The non-reciprocal magneto-optic waveguides under study in this work are of course just a special case of a general anisotropic waveguide.

2.2.2.2 Transfer Matrix Formalism for planar, anisotropic waveguides

Introduction Before we embark on the extension of the TMF towards anisotropic layered structures, a few introductory remarks have to be made. The considered structure is the one depicted in Fig. 2.3, with the x-coordinate axis chosen transversal to the layer interface. The field components tangential to the interface are denoted by an $\|$-index, and those along the x-axis by an $\perp$-index. Without loss of generality, it is assumed that the eigenmodes propagate in the z-direction, and thus that there is no y-dependence in the modal profiles. Such a y-dependence would only originate from a guided mode whose propagation vector $\boldsymbol{\beta}_{\|}$ has a non-vanishing y-component. This can obviously always be eliminated by an appropriate rotation of the coordinate axis. The equations will however be worked out with an explicit vector form for $\boldsymbol{\beta}_{\|}$. Substituting $\boldsymbol{\beta}_{\|} = \beta \mathbf{u}_z$ at the end will yield the traditional form for the modal equations of an anisotropic slab waveguide.

Reformulation of the Maxwell equations in an homogeneous anisotropic layer The starting point are the Maxwell curl equations of (2.20), explicitly written down in the $\perp$- and $\|$-components, and with $\mathbf{j}(\mathbf{r}) = 0$ and $\boldsymbol{\nabla}_{\|} = -j\boldsymbol{\beta}_{\|}$.

[18] the most obvious examples of these cases are those where the principle axes of the anisotropic material lie along the coordinate axes. Another important example of such a case is of course the transversely magnetized magneto-optic waveguide, so extensively studied in this work! Section 2.3 of this chapter will devote a lot of attention to the traditional TMF applied to this waveguide.

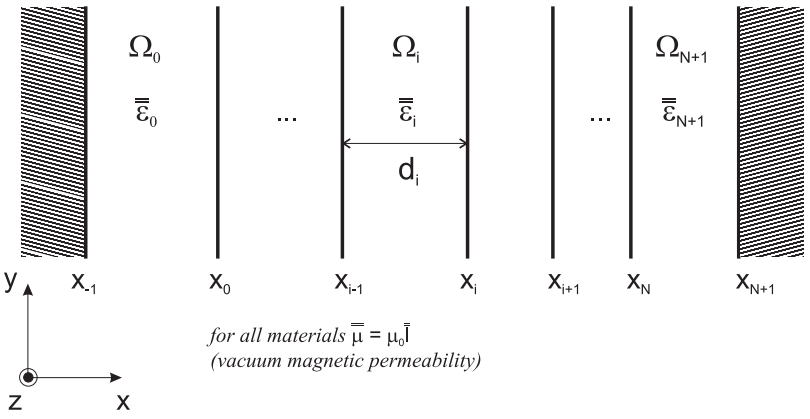


Fig. 2.3: An anisotropic layered slab waveguide. At the considered optical operation frequency, the magnetic permeability tensors will all reduce to the (isotropic) vacuum value.

The explicit $\exp(j\omega t - j\boldsymbol{\beta}_\parallel \cdot \mathbf{r}_\parallel)$ is cancelled out.

$$-j\boldsymbol{\beta}_\parallel \times \mathbf{E}_\parallel(r_\perp) - j\boldsymbol{\beta}_\parallel \times \mathbf{u}_\perp E_\perp(r_\perp) + \mathbf{u}_\perp \times \frac{\mathrm{d}\mathbf{E}_\parallel(r_\perp)}{\mathrm{d}r_\perp} = -j\omega\mu_0(\mathbf{H}_\parallel(r_\perp) + H_\perp(r_\perp)\mathbf{u}_\perp) \tag{2.32a}$$

$$-j\boldsymbol{\beta}_\parallel \times \mathbf{H}_\parallel(r_\perp) - j\boldsymbol{\beta}_\parallel \times \mathbf{u}_\perp H_\perp(r_\perp) + \mathbf{u}_\perp \times \frac{\mathrm{d}\mathbf{H}_\parallel(r_\perp)}{\mathrm{d}r_\perp} = j\omega\underline{\epsilon}(r_\perp) \cdot (\mathbf{E}_\parallel(r_\perp) + E_\perp(r_\perp)\mathbf{u}_\perp) \tag{2.32b}$$

The main idea is to use a well known generally valid lemma from electromagnetics, stating that *once the parallel electromagnetic field components are known along any plane in a homogeneous region, then one can, out of these, derive <u>all</u> field components (thus, both the parallel and perpendicular components) at any point in the homogeneous region.* Therefore, if the plane is chosen along the interfaces of the layered structure, then the Maxwell boundary conditions at each interface are straightforwardly satisfied, and the $\parallel$-components can be propagated throughout the complete layered stack. It will now be shown how this lemma manifests itself in practice. In order not to burden the notation too much, the $\perp$-index will be replaced again by x.
Consider therefore the $\parallel$ and $\perp$-components of the curl equations (2.32) in an arbitrary homogeneous anisotropic layer Ω_i of the stack in Fig. 2.3,

$$\perp \begin{cases} \mu_0 H_x(x) = \big[\mathbf{u}_x \times \dfrac{\boldsymbol{\beta}_\parallel}{\omega}\big] \cdot \mathbf{E}_\parallel(x) \\ \epsilon_{xx} E_x(x) = -\boldsymbol{\epsilon}_{x\parallel} \cdot \mathbf{E}_\parallel(x) + \big[-\mathbf{u}_x \times \dfrac{\boldsymbol{\beta}_\parallel}{\omega}\big] \cdot \mathbf{H}_\parallel(x) \end{cases} \tag{2.33a}$$

and,

$$\parallel \begin{cases} j\boldsymbol{\beta}_\parallel E_x(x) + \dfrac{\mathrm{d}\mathbf{E}_\parallel(x)}{\mathrm{d}x} = j\omega\big[\mathbf{u}_x \times \mu_0 \underline{I}\big] \cdot \mathbf{H}_\parallel(x) \\ j\boldsymbol{\beta}_\parallel H_x(x) + \dfrac{\mathrm{d}\mathbf{H}_\parallel(x)}{\mathrm{d}x} = -j\omega\big[\mathbf{u}_x \times \underline{\epsilon}\big]_{\parallel\parallel} \cdot \mathbf{E}_\parallel(x) - j\omega\big[\mathbf{u}_x \times \boldsymbol{\epsilon}_{\parallel x}\big] E_x(x) \end{cases} \tag{2.33b}$$

This can be written compactly by introducing 4 matrices

$$\overline{\overline{M}}_1 \begin{pmatrix} E_x \\ H_x \end{pmatrix} = \overline{\overline{M}}_2 \begin{pmatrix} \mathbf{E}_\parallel \\ \mathbf{H}_\parallel \end{pmatrix} \tag{2.34a}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} \begin{pmatrix} \mathbf{E}_\parallel \\ \mathbf{H}_\parallel \end{pmatrix} = \overline{\overline{M}}_3 \begin{pmatrix} \mathbf{E}_\parallel \\ \mathbf{H}_\parallel \end{pmatrix} + \overline{\overline{M}}_4 \begin{pmatrix} E_x \\ H_x \end{pmatrix} \tag{2.34b}$$

with

$$\overline{\overline{M}}_1 = \begin{pmatrix} 0 & \mu_0 \\ \epsilon_{xx} & 0 \end{pmatrix} \qquad \text{2-by-2 matrix} \tag{2.35a}$$

$$\overline{\overline{M}}_2 = \begin{pmatrix} \mathbf{u}_x \times \frac{\boldsymbol{\beta}_\parallel}{\omega} & 0 \\ -\boldsymbol{\epsilon}_{x\parallel} & -\mathbf{u}_x \times \frac{\boldsymbol{\beta}_\parallel}{\omega} \end{pmatrix} \qquad \text{2-by-4 matrix} \tag{2.35b}$$

$$\overline{\overline{M}}_3 = j\omega \begin{pmatrix} 0 & \mathbf{u}_x \times \mu_0 \underline{I} \\ -(\mathbf{u}_x \times \underline{\epsilon})_{\parallel\parallel} & 0 \end{pmatrix} \qquad \text{4-by-4 matrix} \tag{2.35c}$$

$$\overline{\overline{M}}_4 = j\omega \begin{pmatrix} -\frac{\boldsymbol{\beta}_\parallel}{\omega} & 0 \\ -\mathbf{u}_x \times \boldsymbol{\epsilon}_{\parallel x} & -\frac{\boldsymbol{\beta}_\parallel}{\omega} \end{pmatrix} \qquad \text{4-by-2 matrix} \tag{2.35d}$$

Eliminating the perpendicular field components in these matrix equations, finally yields the coupled first-order differential equations in the parallel field components,

$$\frac{\mathrm{d}}{\mathrm{d}x} \begin{pmatrix} \mathbf{E}_\parallel \\ \mathbf{H}_\parallel \end{pmatrix} = \Big(\underbrace{\overline{\overline{M}}_3 + \overline{\overline{M}}_4 \overline{\overline{M}}_1^{-1} \overline{\overline{M}}_2}_{\overline{\overline{A}}} \Big) \cdot \begin{pmatrix} \mathbf{E}_\parallel \\ \mathbf{H}_\parallel \end{pmatrix} \tag{2.36}$$

The matrix $\overline{\overline{A}}$ is commonly referred to as the system matrix of the anisotropic medium.

Solution of the system equation In this form the Maxwell equations for a homogeneous, anisotropic layer easily prove the above mentioned lemma. It is well known that such a system of coupled first order linear differential equations is straightforwardly solved by calculating the eigenvectors and eigenvalues of the system matrix $\overline{\overline{A}}$. The similarity transformation that transforms $\overline{\overline{A}}$ into its diagonal form[19], $\overline{\overline{\Gamma}} = \overline{\overline{D}}^{-1} \overline{\overline{A}}\, \overline{\overline{D}}$, is used to rewrite (2.36) as a set of uncoupled first order

[19] It has to be remarked here that there is no straightforward mathematical proof that the system matrix $\overline{\overline{A}}$ of an arbitrary anisotropic medium can be diagonalized. In general, one should therefore rather talk about the Jordan canonical form of the system matrix. However, from a pragmatical point of view it is safe to assume that $\overline{\overline{A}}$ can indeed be diagonalized.

differential equations,

$$\overline{\overline{D}}^{-1}\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix}\mathbf{E}_\parallel\\ \mathbf{H}_\parallel\end{pmatrix}=\overline{\overline{D}}^{-1}\overline{\overline{A}}\,\overline{\overline{D}}\,\overline{\overline{D}}^{-1}\begin{pmatrix}\mathbf{E}_\parallel\\ \mathbf{H}_\parallel\end{pmatrix}=\overline{\overline{\Gamma}}\,\overline{\overline{D}}^{-1}\begin{pmatrix}\mathbf{E}_\parallel\\ \mathbf{H}_\parallel\end{pmatrix}, \tag{2.37}$$

with $\overline{\overline{D}}$ the square 4-by-4 matrix whose columns are the eigenvectors of $\overline{\overline{A}}$, and $\overline{\overline{\Gamma}}$ the 4-by-4 diagonal matrix, whose elements are the corresponding (complex) eigenvalues. Introducing four new scalar unknown functions $w_i(x)$ $(i=1,\ldots,4)$,

$$\begin{pmatrix}\mathbf{E}_\parallel\\ \mathbf{H}_\parallel\end{pmatrix}=\overline{\overline{D}}\cdot\begin{pmatrix}w_1\\ w_2\\ w_3\\ w_4\end{pmatrix}, \tag{2.38}$$

the general solution of (2.36) in the homogeneous layer Ω_i can now be written as[20]

$$\begin{pmatrix}\mathbf{E}_\parallel^{(i)}\\ \mathbf{H}_\parallel^{(i)}\end{pmatrix}(x)=\sum_{j=1}^{4}w_j(x_{i-1^+})\overline{\overline{D}}^{(i)}_{*j}e^{\Gamma_j^{(i)}(x-x_{i-1^+})} \tag{2.39}$$

This shows that the field components parallel to the interfaces, that are solutions of the sourceless Maxwell equations in an arbitrary anisotropic layer of a layered anisotropic stack, are a superposition of four plane waves. These are the eigenwaves of the homogeneous anisotropic space with a wavevector given by $(j\Gamma_k,\boldsymbol{\beta}_\parallel)$ with $k=1,\ldots,4$. Remember that the $\mathbf{r}_\parallel$-dependence of the fields is of course a pure propagator factor, i.e. $\exp(-j\boldsymbol{\beta}_\parallel\cdot\mathbf{r}_\parallel)$. The polarization of each plane eigenwave is described by the eigenvectors of the system matrix $\overline{\overline{A}}$ (the columns of $\overline{\overline{D}}$). The four eigenvalues, $j\Gamma_k$, of the system matrix $\overline{\overline{A}}$ are thus nothing else than the perpendicular component of the propagation constants of the eigenwaves in the homogeneous region. These eigenvalues are the solution of the four-dimensional characteristic equation,

$$\left|\overline{\overline{A}}(\boldsymbol{\beta}_\parallel)-\Gamma\overline{\overline{I}}\right|=0, \tag{2.40}$$

where it has been explicitly indicated that the matrix elements of $\overline{\overline{A}}$ are a function of the chosen value for the propagation constant along the interface plane. Equation (2.40) is a 4-dimensional polynomial in the components of $\boldsymbol{\beta}_\parallel$ and Γ. As mentioned in 2.2.1 this can be analytically solved using Cardano's method. This means that for each $\boldsymbol{\beta}$, there are four values $\Gamma_k(\boldsymbol{\beta})$, $k=1,\ldots,4$, for which the above determinant is zero. These could also be seen as four complex surfaces in the complex $\boldsymbol{\beta}$-vectorspace. In general nothing is known beforehand about the shape or connectivity of these surfaces. However, departing from the isotropic case, for which it is common knowledge that the surfaces are two coinciding, closed complex spheres surrounding the origin of the complex $\boldsymbol{\beta}$-space[21], it is intuitively felt that for most practical anisotropic materials (considered as a small perturbation of an

[20] the superscript (i) refers to the layer number

[21] and thus leading to two groups of degenerate $\Gamma_k(\boldsymbol{\beta})$ with exact opposite values

isotropic material), the zeros $\Gamma_k(\boldsymbol{\beta})$ will come in two groups of values, two with a negative real part, and two with a positive real part. Otherwise put, it is safe to assume that the four surfaces will almost always reduce to two closed ones around the origin. A more rigorous treatment of this assumption (but not a mathematical proof in the strict sense of the word) can be found in [11]. As a consequence the eigenvalues Γ_k (and the corresponding eigenvector columns in $\overline{\overline{D}}$) can always be ordered as $\Re(\Gamma_1) \leq \Re(\Gamma_2) \leq 0 \leq \Re(\Gamma_3) \leq \Re(\Gamma_4)$.
This means that of the four planar eigenwaves in (2.39) those with amplitudes w_1 and w_2 will never grow exponentially in the positive x-direction, while those with amplitudes w_3 and w_4 will never grow exponentially in the negative x-direction. In other words, the first two eigenwaves meet the physical radiation condition for 'upward' propagating waves, while the latter two fulfill that of 'downward' propagating waves. This eigenvalue ordering naturally corresponds to the case of a lossy medium[22]. When the medium becomes lossless, the eigenvalues become strictly imaginary and the ordering is now done in the same manner, but for the imaginary parts. As a result the first two plane eigenwaves will have a negative and purely imaginary Γ, so that the $e^{\Gamma x}$ factor indeed indicates an upward propagating wave. The inverse applies for the last two eigenwaves.
All of the above clarifies the form of the general solution of (2.39) in a way that was of course intuitively to be expected. For any physically possible wavevector $\boldsymbol{\beta}_{\|}$ parallel to the interfaces, the electromagnetic field in any homogeneous anisotropic layer of a layered stack can be decomposed into a linear combination of two 'forward' and two 'backward' propagating plane waves, and for each propagation direction there are two base polarizations[23]. This implies that the matrix $\overline{\overline{D}}$, used in the similarity transformation to diagonalize $\overline{\overline{A}}$, is nothing else but a linear coordinate (or alias) transformation of the 4-dimensional complex vector space of the parallel electromagnetic field components. $\overline{\overline{D}}$ transforms between the coordinate systems described by[24] $\left(E_y, E_z, H_y, H_z\right)$ and $\left(w_1, w_2, w_3, w_4\right)$. The base vectors of the latter coordinate system expressed in the former coordinate system are of course the columns of $\overline{\overline{D}}$, and are nothing else but the $\left(E_y, E_z, H_y, H_z\right)$ coordinate representation of the base polarizations of the anisotropic layer. Also note that the $\left(E_y, E_z, H_y, H_z\right)$ coordinate system is applicable throughout the entire layered stack, whereas the $\left(w_1, w_2, w_3, w_4\right)$ coordinate system only applies locally in its corresponding homogeneous layer. The benefit of working with two different coordinate systems is clear.

$\left(E_y, E_z, H_y, H_z\right)$ allows the fields to be expressed in their natural form for application of the Maxwell boundary conditions. This will be the preferred coordinate system to 'transfer' the fields over the different interfaces of the layered stack. This the standard Cartesian coordinate system.

[22] which in reality any medium is. In electromagnetics it is not so evident how a theoretical correct treatment for media having gain should be constructed. More on this can be found in [14].

[23] This is a fundamental physical consequence of the fact that, even though photons are bosons with s = 1, their rest mass is zero and thus the projection m = 0 can never be realised [15], due to the gauge invariance of Maxwell's equations. Therefore photons are particles that can only exist in two states.

[24] for clarity the parallel field components are expressed in their Cartesian components.

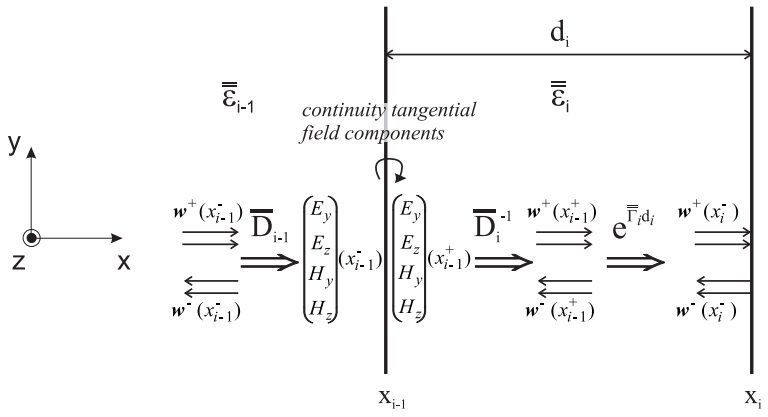


Fig. 2.4: Derivation of the 4×4 *transfer matrix for a general anisotropic chunk.*

(w_1, w_2, w_3, w_4) allows the fields to be expressed in their independently propagating components. This will be the preferred coordinate system to 'transfer' the fields through each layer of the layered stack. This system is also called the wavefield coordinate system. It is important to note that this is a local coordinate system in contrast to the standard Cartesian system which is applicable throughout the layered stack.

The extended transfer matrix formalism All elements have been collected to correctly formulate the framework of the extended transfer matrix formalism. Consider therefore again Fig. 2.3. The aim is to derive the transfer matrix connecting the parallel field components at the 'left' enclosing computational boundary with the parallel field components at the 'right' enclosing computational boundary. For that purpose it suffices to derive the basic matrix correctly describing the transfer through an arbitrary interface and the subsequent homogeneous layer, and then concatenate these 'chunk' matrices. Consider therefore in Fig. 2.4, the layer Ω_i surrounded by interfaces at x_{i-1} and x_i.

The idea is to derive the matrix that correctly describes the transfer of the 2 'forward' and 2 'backward' independent plane wave eigenpolarizations through an anisotropic chunk. In this way we proceed as in the 2×2 TMF formalism for isotropic structures (indicated in 2.2.2.1). The difference is of course that a 4×4-formalism allows for interface coupling between the independent eigenpolarizations.

Just below the interface at x_{i-1} the electromagnetic fields are expressed in the wavefield coordinate system, being nothing else than the fourvector of the amplitudes of the different eigenpolarizations composing the total field. In order to formulate Maxwell boundary conditions the wavefield fourvector must be projected onto the (E_y, E_z, H_y, H_z) coordination system. In this way the Cartesian field fourvector at the other side of the interface is obtained (due to the continuity

of the tangential components).

$$\begin{pmatrix} E_y \\ E_z \\ H_y \\ H_z \end{pmatrix}(x_{i-1^+}) = \begin{pmatrix} E_y \\ E_z \\ H_y \\ H_z \end{pmatrix}(x_{i-1^-}) = \overline{\overline{D}}_{i-1} \begin{pmatrix} w_1 \uparrow \\ w_2 \uparrow \\ w_3 \downarrow \\ w_4 \downarrow \end{pmatrix}(x_{i-1^-}) \tag{2.41}$$

At x_{i-1^+} the Cartesian electromagnetic field components are transformed back into the wavefield coordinate system, so that the plane eigenpolarizations can be independently propagated over a thickness d_i up to x_{i^-}

$$\begin{pmatrix} w_1 \uparrow \\ w_2 \uparrow \\ w_3 \downarrow \\ w_4 \downarrow \end{pmatrix}(x_{i-1^+}) = \overline{\overline{D}}_i^{-1} \begin{pmatrix} E_y \\ E_z \\ H_y \\ H_z \end{pmatrix}(x_{i-1^+}), \tag{2.42}$$

$$\begin{pmatrix} w_1 \uparrow \\ w_2 \uparrow \\ w_3 \downarrow \\ w_4 \downarrow \end{pmatrix}(x_{i^-}) = e^{\overline{\overline{\Gamma}}_i d_i} \begin{pmatrix} w_1 \uparrow \\ w_2 \uparrow \\ w_3 \downarrow \\ w_4 \downarrow \end{pmatrix}(x_{i-1^+}) \tag{2.43}$$

Combining (2.41),(2.43) and (2.42) results in the extended 4-by-4 transfer matrix of an anisotropic chunk,

$$\begin{pmatrix} \mathbf{w} \uparrow \\ \mathbf{w} \downarrow \end{pmatrix}(x_{i^-}) = e^{\overline{\overline{\Gamma}}_i d_i} \overline{\overline{D}}_i^{-1} \overline{\overline{D}}_{i-1} \begin{pmatrix} \mathbf{w} \uparrow \\ \mathbf{w} \downarrow \end{pmatrix}(x_{i-1^-}) \tag{2.44}$$

Concatenating these transfer matrices allows the 4-by-4 transfer matrix of the N-layered anisotropic stack in Fig. 2.3 to be written as

$$\overline{\overline{D}}_{N+1}^{-1} \Big(\prod_{i=1}^{N} \overline{\overline{D}}_i e^{\overline{\overline{\Gamma}}_i d_i} \overline{\overline{D}}_i^{-1} \Big) \overline{\overline{D}}_0, \tag{2.45}$$

relating the 'forward' and 'backward' propagating waves at x_0^- to those at x_N^+. This form of the generalized transfer matrix makes clear where the coupling comes from. At every interface there occurs a matrix product of the form $\overline{\overline{D}}_{i+1}^{-1} \overline{\overline{D}}_i$. This product describes how the local wavefield coordinate system in layer i is transformed into the one valid in layer $i+1$. Obviously, this transformation will cause coupling between the two sets of eigenpolarizations at both sides of the interface, unless for instance both sets of local eigenpolarizations have a pure TE and TM character[25].

[25] In order to see this, it suffices to realize that for pure TE and TM eigenpolarizations, the $\overline{\overline{D}}$-matrix can be brought in a block diagonal form through a similar column and row permutation, $\begin{pmatrix} \star & \star & 0 & 0 \\ \star & \star & 0 & 0 \\ 0 & 0 & \star & \star \\ 0 & 0 & \star & \star \end{pmatrix} = P_{23} P_{34} \begin{pmatrix} \star & 0 & 0 & \star \\ 0 & \star & \star & 0 \\ 0 & \star & \star & 0 \\ \star & 0 & 0 & \star \end{pmatrix} P_{34} P_{23}$. As a consequence its inverse can also be 'block-diagonalized' through the same permutations, because the inverse of a block-diagonal matrix is again block-diagonal. So that, finally, the product of a pure TE/TM $\overline{\overline{D}}$-matrix with the inverse of a pure TE/TM $\overline{\overline{D}}$-matrix, will again be a pure TE/TM $\overline{\overline{D}}$-matrix. This proves that an interface between layers having pure TE and TM eigenpolarizations, will leave these polarizations uncoupled. Of course, this is only a special case, and it is intuitively clear that also no coupling will occur if the eigenpolarizations in both layers are mutually orthogonal with eachother.

By introducing the matrix $\overline{\overline{S}}_i = \overline{\overline{D}}_i e^{\overline{\overline{\Gamma}}_i d_i} \overline{\overline{D}}_i^{-1}$, (2.45) can be rewritten compactly as

$$\overline{\overline{D}}_{N+1}^{-1} \Big(\prod_{i=1}^{N} \overline{\overline{S}}_i \Big) \overline{\overline{D}}_0 \tag{2.46}$$

As a last step, before deriving the waveguide dispersion equation out of this generalized 4×4 transfer matrix, it is very important to pay some attention to the numerical stability of this procedure.
As discussed above, the eigenvalues of the system matrices $\overline{\overline{A}}_i$, will come in pairs of positive real-valued and negative real-valued Γ. In other words the diagonal propagator matrices $e^{\overline{\overline{\Gamma}}_i d_i}$, will contain both exponential increasing and exponential decreasing terms. If no precautions are taken the elements of the stack matrix of (2.46) will consist of sums of (very) small and (very) large numbers. Therefore the stack matrix of (2.46) needs to be numerically stabilized. The ordering of the eigenvalues, as discussed in the previous paragraph, thus forces us to factor out the exponential scalar with the largest positive exponent $e^{\overline{\overline{\Gamma}}_{4,i} d_i}$ in every $\overline{\overline{S}}_i$-matrix, leaving us with a numerically stabilized matrix-scheme[26]

$$\overline{\overline{Q}}_i = e^{-\overline{\overline{\Gamma}}_{4,i} d_i} \overline{\overline{S}}_i \tag{2.47}$$

2.2.2.3 The dispersion equation

After having derived a numerically stable version of the stack transfer matrix of a general anisotropic layered stack, we are in a position to correctly determine the waveguide dispersion equation whose zeros will be the waveguide modes of the layered anisotropic 1-dimensional stack. The independent variable of this equation is the assumed $\boldsymbol{\beta}_{\|}$-wavevector in the invariant direction of the slab waveguide. As was pointed out at the beginning of this section, it will be assumed that $\boldsymbol{\beta}_{\|} = \beta \mathbf{u}_z$. This avoids having another continuum in the eigenspectrum of the slab waveguide, next to the continuum of radiative eigenmodes.
In order to derive the waveguide dispersion equation out of the generalized 4×4 stack transfer matrix, appropriate boundary conditions have to be defined. The precise form of the boundary conditions depends on whether the waveguide is defined in an *open* or a *closed* configuration.

- An **open** waveguide has no computational boundaries. Both sub- and superstrate layers are considered to be semi-infinite. The boundary conditions are defined on a fundamental physical basis, i.e. the waveguide modes, no matter whether they are radiation modes or guided modes, are sourceless solutions of Maxwell's equations. Therefore, there can be no 'incoming' wavefield contributions in the substrate or the superstrate layers.
  The transfer matrix relation between the wavefield vectors at the substrate

[26] if a layer has purely imaginary Γ eigenvalues, then there is no apparent need for this stabilization. However, as in this case the above stabilization procedure gets down to adding a global phase factor to the stack matrix, there is no 'harmful' influence on the derivation of the waveguide dispersion equation.

interface and those at the superstrate interface, then reads, using (2.46) and (2.47),

$$\begin{pmatrix} \mathbf{w}\uparrow \\ \mathbf{0} \end{pmatrix}(x_{N^+}) = e^{\sum_{i=1}^{N}\Gamma_{4,i}d_i}\underbrace{\overline{\overline{D}}_{N+1}^{-1}\left(\prod_{i=1}^{N}\overline{\overline{Q}}_i\right)\overline{\overline{D}}_0}_{\overline{\overline{T}}}\begin{pmatrix} \mathbf{0} \\ \mathbf{w}\downarrow \end{pmatrix}(x_{0^-}) \tag{2.48}$$

If the here defined T-matrix is written as 2×2-block matrix, then the above equation will only have non-trivial solutions, if the determinant of the (2,2)-block is zero,$\left|T_{22}\right| = 0$. The zeros of this equation will in theory come in both a discrete set (the guided modes) and a continuum (the radiation modes). This latter is obviously the result of the presence of branch-cut(s) in (2.47). In a general anisotropic waveguide the modes will have hybrid polarization, with all 6 electromagnetic field components different from zero.

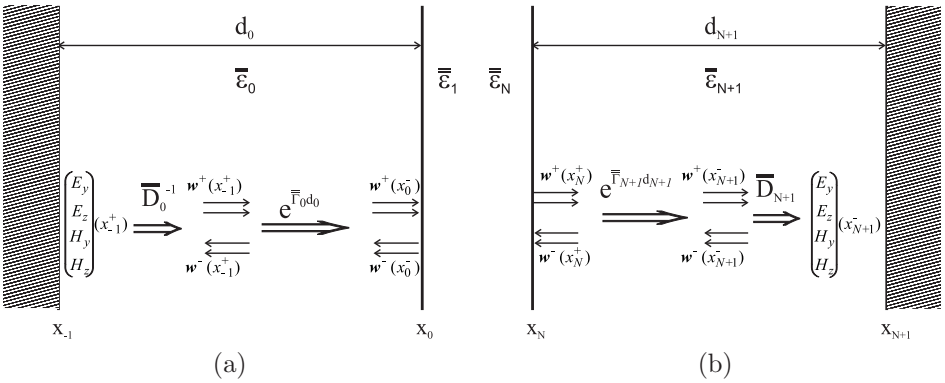


Fig. 2.5: (a) calculation of the substrate stack matrix; (b) calculation of the superstrate stack matrix

- A **closed** waveguide is enclosed within two (hypothetical) walls, which are impenetrably for electromagnetic waves and thus completely shield the waveguide from the external world. Typically, these walls are perfect electric or magnetic conductors. Such walls are also called 'hard' walls, because they reflect all incident power. The eigenspectrum of a closed waveguide will be purely discrete, but infinite. The discretisation of the continuum of radiation modes is the most important motivation to introduce these hard walls. In this way one eliminates the numerical difficulties associated with branch-cuts. The main disadvantage of using a hard wall is the occurrence of parasitic reflections of the radiation modes in eigenmode expansion propagation calculations, which can completely obliterate the meaningfulness of the simulation results (see e.g. [12] for convincing examples of this).
  If the walls are placed far enough from the guiding layers, their influence on the modal properties of the guided modes will be negligible. In order to derive the dispersion equation, one first needs to extend the stack transfer matrix of

(2.47) to incorporate the local eigenwave propagation through the now finite sub- and superstrate layers. This extended stack matrix is then expressed in the Cartesian coordinate system, in order to relate the electromagnetic field components at both walls. This is illustrated in Fig. 2.5 The stabilized transfer matrix relation between the parallel Cartesian electromagnetic field components at the substrate wall and those at the superstrate wall, then reads, using (2.46) and (2.47) and the matrices derived in Fig. 2.5,

$$\begin{pmatrix} E_y \\ E_z \\ H_y \\ H_z \end{pmatrix}(x_{N+1}) = e^{\sum_{i=0}^{N+1}\Gamma_{4,i}d_i}\underbrace{\overline{\overline{D}}_{N+1}e^{\overline{\overline{\Gamma}}_{N+1}d_{N+1}}}_{\text{superstratematrix}}\overline{\overline{D}}_{N+1}^{-1}$$
$$\Big(\prod_{i=1}^{N}\overline{\overline{Q}}_i\Big)\overline{\overline{D}}_0\ \underbrace{e^{\overline{\overline{\Gamma}}_0 d_0}\overline{\overline{D}}_0^{-1}}_{\text{substratematrix}}\begin{pmatrix} E_y \\ E_z \\ H_y \\ H_z \end{pmatrix}(x_{-1}), \quad (2.49)$$

Or, more compactly

$$\begin{pmatrix} E_y \\ E_z \\ H_y \\ H_z \end{pmatrix}(x_{N+1}) = e^{\sum_{i=0}^{N+1}\Gamma_{4,i}d_i}\underbrace{\Big(\prod_{i=0}^{N+1}\overline{\overline{Q}}_i\Big)}_{\overline{\overline{T}}}\begin{pmatrix} E_y \\ E_z \\ H_y \\ H_z \end{pmatrix}(x_{-1}) \quad (2.50)$$

The dispersion equation is now found by applying the boundary conditions on the field vectors at both walls, imposed by the character of these walls. An electric wall requires vanishing parallel electric field components. A magnetic wall requires impose vanishing parallel magnetic field components. This means four different wall combinations. Considering the above defined T-matrix again as a 2×2-block matrix, non-trivial solutions for the four mentioned wall combinations will be the zeros of the determinants of the four different 2×2 blocks of T. For instance, the dispersion equation of a waveguide enclosed within two electric walls is found as $\det(T_{12}) = 0$.

2.2.2.4 Modal profiles

To calculate the modal profile, when the propagation constant of a mode of the non-reciprocal anisotropic waveguide is found, the most common approach is to use the so-called T-scheme [12]. The "T" stands for "transfer", hinting at the basic idea behind the T-scheme. It is in essence based on the "transfer" of a known field distribution at a certain position in the slab waveguide (most often at the walls) throughout the entire waveguide cross-section using the transfer matrices (2.44). This is numerically the most economical approach. However, it is not so difficult to see that it involves some risks of numerical instability, as one is basically transferring the field throughout the entire stack by means of "building" up the entire stack matrix, for which it was pointed out above that it inevitably leads to

the addition of exponentially increasing and exponentially decreasing factors. The risk of rounding errors due to limited machine precision obviously increases with the "transfer distance" into the stack. Remark also that of course there is no way of numerically stabilizing the transfer matrices when calculating field profiles in a way as was done for the dispersion equation. As opposed to the dispersion equation, where the exponential increasing propagators could be factored out without influencing the zeroes of the equation, for the calculation of mode profiles this is not possible, since it is essentially the subtle balance of the addition of the propagator factors that determines the mode profile. A possible alternative to calculate the field profiles that is numerically more stable will be indicated. However it will be seen that this comes at a heavy price of considerable increased numerical effort. The starting point in the T-scheme is the determination of the wavefield amplitudes $\mathbf{w}$ at a position in the waveguide where a boundary condition is defined. For an open waveguide this is either at the left-hand side of the substrate interface, x_{0^-}, or at the right-hand side of the superstrate interface, x_{N^+}. For a closed waveguide this is at one of the perfect conducting walls.

Field profiles in open waveguides In an open waveguide, the radiation boundary condition states that at x_{0^-} there are no 'upward' wavefield components $\mathbf{w}\uparrow$, no 'downward' wavefield components $\mathbf{w}\downarrow$ at x_{N^+}. Using $\overline{\overline{T}}$ defined in (2.48) has led to the dispersion equation $\overline{\overline{T}}_{22}\begin{pmatrix} w_1\downarrow \\ w_2\downarrow \end{pmatrix}(x_{0^-}) = 0$ or $\det(\overline{\overline{T}}_{22}) = 0$. For a zero of this determinant dispersion equation (i.e. a waveguide mode), one of both equations in the former expression can be used to derive a relationship between the wavefield amplitudes $w_1\downarrow$ and $w_2\downarrow$ at x_{0^-} of the found mode.

$$w_2\downarrow(x_{0^-}) = -\frac{(\overline{\overline{T}}_{22})_{11}}{(\overline{\overline{T}}_{22})_{12}} w_1\downarrow(x_{0^-}) \tag{2.51}$$

With this expression and the chunk matrices (2.44) (with $d_i = x_i - x_{i-1}$), the complete mode profile can be calculated throughout the entire stack in terms of the modal amplitude, $A = w_1\downarrow(x_{0^-})$. At any point the transferred wavefield components can be transformed into the traditional Cartesian field components with the local transformation matrix $\overline{\overline{D}}$.

Field profiles in closed waveguides In a closed waveguide, for example – without loss of generality – a waveguide enclosed within two electric walls, the boundary conditions state that at x_{-1^+} and at x_{N+1^-}, the Cartesian tangential electric field components must be zero. This has led to the dispersion equation $\det(\overline{\overline{T}}_{12}) = 0$ (with $\overline{\overline{T}}$ as defined in (2.50)). This allows to determine a relationship between the tangential magnetic field components at the substrate wall.

$$\begin{aligned} &\overline{\overline{T}}_{12}\begin{pmatrix} H_y \\ H_z \end{pmatrix}(x_{-1^+}) = 0 \\ &H_z(x_{-1^+}) = -\frac{(\overline{\overline{T}}_{12})_{11}}{(\overline{\overline{T}}_{12})_{12}} H_y(x_{-1^+}) \end{aligned} \tag{2.52}$$

Using the substrate matrix (defined in (2.50)), the field components at x_{-1^+} can be propagated up to x_{0^-} and any other position in the substrate. From there, the chunk matrices (2.44) (with $d_i = x - x_{i-1}$) can be used. At the superstrate interface, x_{N^-}, the wavefield components can be further propagated into the superstrate using $e^{\overline{\overline{\Gamma}}_{N+1}(x-x_N)}\overline{\overline{D}}_{N+1}^{-1}\overline{\overline{D}}_N$. For this particular case (two electric walls), the field amplitude is obviously chosen to be $A = H_y(x_{-1^+})$. Again, at any point the profile can be expressed in terms of the Cartesian components using the appropriate transformation matrix, $\overline{\overline{D}}$.

If the waveguide stack comprises a lot of layers and/or the layers are appreciably thick with respect to the wavelength, the propagation of the wavefield components starting at the substrate can easily loose precision as explained above. The deeper one propagates into the stack, the less accurate the results become. There is no simple way to avoid this. The only solution one has, is to avoid propagation over large distances. This can be done when the stack has a symmetry plane. Then the modes might also have a certain symmetry. The idea is then to formulate a boundary condition in the symmetry plane, and derive new dispersion relationships, one for each symmetry eigenvalue of the modes. This would allow to perform the above described transfer procedure over only half the stack thickness. However in anisotropic waveguides, it is not easy to predict what the parities of the different Cartesian field components will be, because of the tensorial character of the dielectric permittivity. If there will be any at all! Indeed, a full tensorial form for $\underline{\epsilon}$ indicates that there are certain "preferred" directions in the anisotropic layers. And these can have in general an arbitrary direction. It is therefore to be expected that even in a geometrical mirror symmetric slab waveguide, only under certain conditions the modes will exhibit an even and odd character. The reader is again referred to [1] and [6] for a more detailed review of mode symmetries in (bi-)anisotropic waveguides[27].

Another approach is to restart the derivation of the dispersion equation, but now not in terms of the wavefield behaviour at the boundaries of the waveguide (which is an external viewpoint of guiding) but in terms of a "resonance" condition that the wavefields must fulfill inside the waveguide (i.e. an internal viewpoint of waveguiding). The principle is illustrated in Fig. 2.6. The outgoing wavefields at the right-hand side resp. the left-hand side of the interface considered in Fig. 2.6 can be considered to be generated by 2×2 reflection matrices, $\overline{\overline{R}}_{\text{left}}$ resp. $\overline{\overline{R}}_{\text{right}}$. These matrices not only take all multiple reflection and transmission effects within the substacks to the right and the left into account, but also the initial conditions at the far sides of the stack (these being conducting walls or semi-infinite sub- and/or superstrate layers). For a derivation of these matrices the reader is referred to [13]. There it is also proven that these reflection matrices are always numerically stable.

$$\mathbf{w}(x_{i^+}) \uparrow = \overline{\overline{R}}_{\text{left}}\mathbf{w}(x_{i^+}) \downarrow \qquad \mathbf{w}(x_{i^-}) \downarrow = \overline{\overline{R}}_{\text{right}}\mathbf{w}(x_{i^-}) \uparrow \tag{2.53}$$

On top of these relationships, the wavefields also are subjected to the Maxwell

[27] When the system matrix is completely written out for the three magneto-optic base configurations (see section 2.2.3.5), it can be proven that the modal profiles of such magneto-optic waveguides will exhibit even and odd parity.

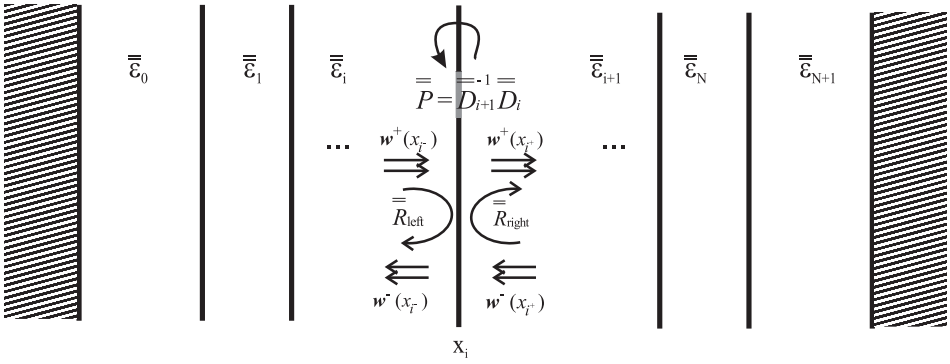


Fig. 2.6: Derivation of the dispersion equation from an internal viewpoint. Using the three pictured matrix relationships between the four unknown sets of wavefields at the interface x_i, a homogeneous resonance-type matrix equation can be obtained for a single set of wavefields. The determinant of this matrix is then the internally derived dispersion equation of the whole stack. The boundary conditions are "hidden" inside the reflection matrices.

boundary conditions at x_i.

$$\begin{pmatrix}\mathbf{w}\uparrow\\ \mathbf{w}\downarrow\end{pmatrix}(x_{i^+}) = \underbrace{\overline{\overline{D}}_{i+1}^{-1}\overline{\overline{D}}_i}_{\overline{\overline{P}}}\begin{pmatrix}\mathbf{w}\uparrow\\ \mathbf{w}\downarrow\end{pmatrix}(x_{i^-}) \tag{2.54}$$

Combining (2.53) and (2.54), allows to eliminate all three of the four 2-vectors, yielding the following homogeneous 2×2 set of linear equations (with the block matrices $\overline{\overline{P}}_{ij}$ defined in (2.54))

$$\underbrace{\left(\overline{\overline{P}}_{11} - \overline{\overline{R}}_{\text{left}}\overline{\overline{P}}_{21} + \overline{\overline{P}}_{12}\overline{\overline{R}}_{\text{right}} - \overline{\overline{R}}_{\text{left}}\overline{\overline{P}}_{22}\overline{\overline{R}}_{\text{right}}\right)}_{\overline{\overline{A}}}\mathbf{w}\uparrow(x_{i^-}) = 0 \tag{2.55}$$

This can be considered as a dispersion equation derived from an internal viewpoint. Note again that all boundary conditions at the far ends of the stack are implicitly included via the reflection matrices. The modes are as usual found as the zeroes of $\det(\overline{\overline{A}})$ (as defined in (2.55)). And one of the two linear equations can be used to express one of the wavefield components as a function of the other. The entire wavefield 4-vector at x_{i^-} is then found as

$$\begin{pmatrix}\mathbf{w}\uparrow\\ \mathbf{w}\downarrow\end{pmatrix}(x_{i^-}) = \begin{pmatrix} 1 \\ -\frac{\overline{\overline{A}}_{11}}{\overline{\overline{A}}_{12}} \\ (\overline{\overline{R}}_{\text{right}})_{11} - (\overline{\overline{R}}_{\text{right}})_{12}\frac{\overline{\overline{A}}_{11}}{\overline{\overline{A}}_{12}} \\ (\overline{\overline{R}}_{\text{right}})_{21} - (\overline{\overline{R}}_{\text{right}})_{22}\frac{\overline{\overline{A}}_{11}}{\overline{\overline{A}}_{12}} \end{pmatrix} w_1\uparrow(x_{i^-}) \tag{2.56}$$

This wavefield can now be propagated in the same manner as described above (using the chunk matrices). The obvious advantage is that this dispersion equation can be derived anywhere in the stack. This position can be chosen appropriately in order to avoid propagation of the "starting" field over large distances when calculating the modal profiles. But this increase in numerical stability comes obviously at a large price, in view of the need of the extra calculation of the reflection matrices.

Up till now nothing has been said about a proper normalization of the modal profiles. In section 2.1.2 it was indicated that a proper normalization requires the modal profiles of the adjoint waveguide.

$$-\frac{1}{4N_i}\iint_S [\mathbf{E}_{t,i}(\boldsymbol{\rho}) \times \tilde{\mathbf{H}}_{t,j}(\boldsymbol{\rho}) - \tilde{\mathbf{E}}_{t,j}(\boldsymbol{\rho}) \times \mathbf{H}_{t,i}(\boldsymbol{\rho})] \cdot \mathbf{u}_z \mathrm{d}S = 1 \tag{2.57}$$

with mode j the mode in the adjoint waveguide that is mutually bidirectional with the mode i in the original waveguide (i.e. $\tilde{\beta}_j = -\beta_i$). Note that in general this normalization has to be performed both for forward and backward propagating modes i, since in general the modal spectrum of the original waveguide is not bidirectional, and even if it was, still the modal profiles of the opposite travelling mode can look entirely different. Therefore, a single-mode waveguide will require two normalization integrals and four modal profiles[28]. The normalization integral for the backward travelling modes i must have the opposite sign in order to be consistent with the isotropic case, where backward travelling modal profiles are the mirror images of forward travelling modal profiles. This is the reason why a so-called normalization constant N_i has been introduced in (2.57), with $N_i = \begin{cases} +1 & \text{for forward modes} \\ -1 & \text{for backward modes} \end{cases}$.
On top of that, it is clear, that (2.57) on its own can not entirely fix the amplitude of a mode in a non-reciprocal anisotropic waveguide, since the amplitudes of mode i and mode j can be chosen independently. A complete normalization requires an extra condition, fixing the relationship between a mode in the original waveguide and the corresponding opposite-travelling adjoint mode in the adjoint waveguide[29]. In [1] a possible procedure to derive this relationship is presented, but even that one is approximative (requiring a truncation of the infinite modal spectrum). There seems no easy way out of this theoretical problem. This is not so bad, as the orthogonality of the modes is of course automatically satisfied, and even if the modes are not properly normalized, any mathematical expression that requires normalization can always be properly divided by the Lorentz normalization integral (2.57) (with arbitrary, but fixed chosen amplitudes for every mode i and the corresponding adjoint mode j). An example of this is the application

[28]...and even that is not entirely guaranteed as it is by no means certain that a monomode non-reciprocal waveguide with a bound mode in the positive z-direction will also have a bound mode in the backward z-direction!

[29]Remark that all these normalization subtleties never get any attention in isotropic waveguides, since there the relationship between a mode and its opposite travelling adjoint mode are well-known. They are eachother's mirror image.

of perturbation theory to non-reciprocal, anisotropic waveguides as explained in section 2.2.3.

2.2.2.5 Example - a waveguide Faraday rotator

Here we will give a very basic example to illustrate the above developed formalism. The bound modal spectrum of a lossless three-layered symmetric open, non-reciprocal magneto-optic waveguide in a Faraday configuration will be calculated. Such a waveguide will cause a non-reciprocal rotation of the plane of polarization, and is the waveguide analogy of the Faraday rotator discussed in Chapter 1. This example will illustrate the most important peculiarities of the previous sections: the symmetry of the modal profiles, their mirroring properties (according to (2.17)), the possible normalization, and the non-reciprocity of the induced polarization rotation.

Table 2.1 gives the details of the here used hypothetical simulation example. The wavelength is set to 1300 nm. The cladding layers are assumed to be isotropic, and symmetric, though this is merely to illustrate the symmetry properties of modes in anisotropic waveguides, and not to circumvent the calculation of the full transfer matrix. The core is a hypothetical magneto-optic material subjected to a magnetization along the propagation direction. In Chapter 1 (and further on in section 2.2.3.5 on the perturbation modelling of MO waveguides) it is explained that a MO material in Faraday (or equivalently *longitudinal*) configuration is characterized by a dielectric permittivity tensor of the form $\underline{\epsilon}_{\text{Faraday}} = \begin{pmatrix} \epsilon & +j\epsilon Q & 0 \\ -j\epsilon Q & \epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix}$, with ϵ the scalar dielectric permittivity in the absence of a magnetization, and Q the magneto-optical Voigt parameter. If the material is lossless both parameters are real (as required by the hermiticity of the tensor).

layer	ϵ	Q	d
O (substrate)	9.0	0	semi-infinite
1 (Faraday MO core)	12.25	0.01	510 nm
2 (superstrate)	9.0	0	semi-infinite

TABLE. 2.1: Parameters for the (hypothetical) planar waveguide Faraday rotator.

The formalism described in section 2.2.2.2 is applied by explicitly writing out the system matrices $\overline{\overline{A}}$ along (2.36) for the isotropic layer and the Faraday MO layer, and solving for their respective eigenvectors and eigenvalues. For the here considered case the system matrix reads

$$\overline{\overline{A}} = \begin{pmatrix} 0 & 0 & 0 & -jZ_{\text{vac}}k_0 \\ -k_0 n_{\text{eff}} Q & 0 & jk_0 Z_{\text{vac}}(1 - \frac{n_{\text{eff}}^2}{\epsilon}) & 0 \\ 0 & j\frac{k_0\epsilon}{Z_{\text{vac}}} & 0 & 0 \\ -j\frac{k_0}{Z_{\text{vac}}}\epsilon(1 - \frac{n_{\text{eff}}^2}{\epsilon}) & 0 & -k_0 n_{\text{eff}} Q & 0 \\ -j\frac{k_0}{Z_{\text{vac}}}\epsilon Q^2 & & & \end{pmatrix}, \qquad (2.58)$$

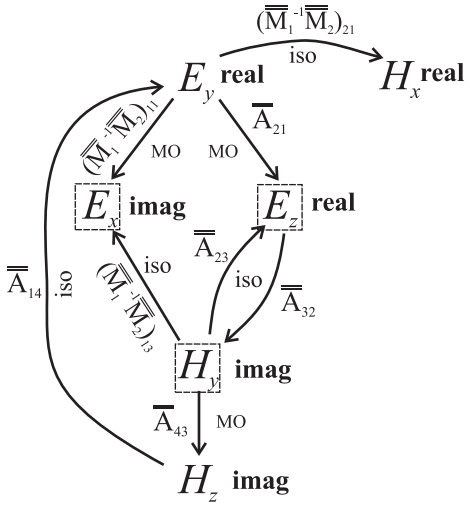


Fig. 2.7: Schematical representation of the origin of the (complex) hybrid nature of the modal profiles in a planar waveguide Faraday rotator as predicted by the system matrices (2.58) *and* (2.59).

while the matrix relating the field components tangential to the interface (y and z) and those perpendicular to it (x) becomes

$$\overline{\overline{M}}_1^{-1}\overline{\overline{M}}_2 = \begin{pmatrix} -jQ & 0 & \frac{Z_{\mathrm{vac}} n_{\mathrm{eff}}}{\epsilon} & 0 \\ -\frac{n_{\mathrm{eff}}}{k_0} & 0 & 0 & 0 \end{pmatrix} \tag{2.59}$$

These matrices convincingly explain the hybrid nature of the modal profiles due to the occurrence of Q. Intuitively, this can be understood as follows. Consider a TE eigenmode (only E_y, H_x and H_z different from zero) in a purely isotropic waveguide and "adiabatically" turn on Q. This causes first of all the creation of a small E_x component (via the MO matrix element $\left(\overline{\overline{M}}_1^{-1}\overline{\overline{M}}_2\right)_{11}$). Next to that, there appears a small MO interaction between E_y and E_z via $\frac{\mathrm{d}}{\mathrm{d}x}E_z \sim \overline{\overline{A}}_{21}E_y$, which in turn causes H_y to increase (via the isotropic matrix element $\overline{\overline{A}}_{32}$). This H_y component will now create additional standard isotropic contributions to E_x (via $\left(\overline{\overline{M}}_1^{-1}\overline{\overline{M}}_2\right)_{13}$) and E_z (via $\overline{\overline{A}}_{23}$), and also a MO contribution to H_z through the MO interaction $\overline{\overline{A}}_{43}$. In a last step this interaction with H_z eventually traces back to where we started, namely at E_y through $\overline{\overline{A}}_{14}$. This reasoning is thus circular, and will eventually reach resonance when the correct modal balance is obtained. This circular argument is schematically depicted in Fig. 2.7. Thus out of the E_y component of a TE mode all other (normally zero) field components are being generated. On top of that, (2.58) and (2.59) also allow to predict the complex nature of the different field components if the waveguide is lossless, because then all material parameters are real. Starting off with a real E_y field and

tracking the matrix interactions described above, allows to predict the imaginary / real character of all field components. This is also illustrated in Fig. 2.7. Note that exactly the same circular argument could be obtained by starting off from the H_y component of a TM mode.

The dispersion equation is obtained by calculating $\left|\det(\overline{\overline{T}})\right|$ with $\overline{\overline{T}}$ as defined in (2.48). Fig. 2.8 plots this dispersion relationship for several values of the thickness of the core as a function of the effective mode index, $n_{\text{eff}} = \beta/k_0$. Due to the lossless character of the waveguide the effective index is purely real.

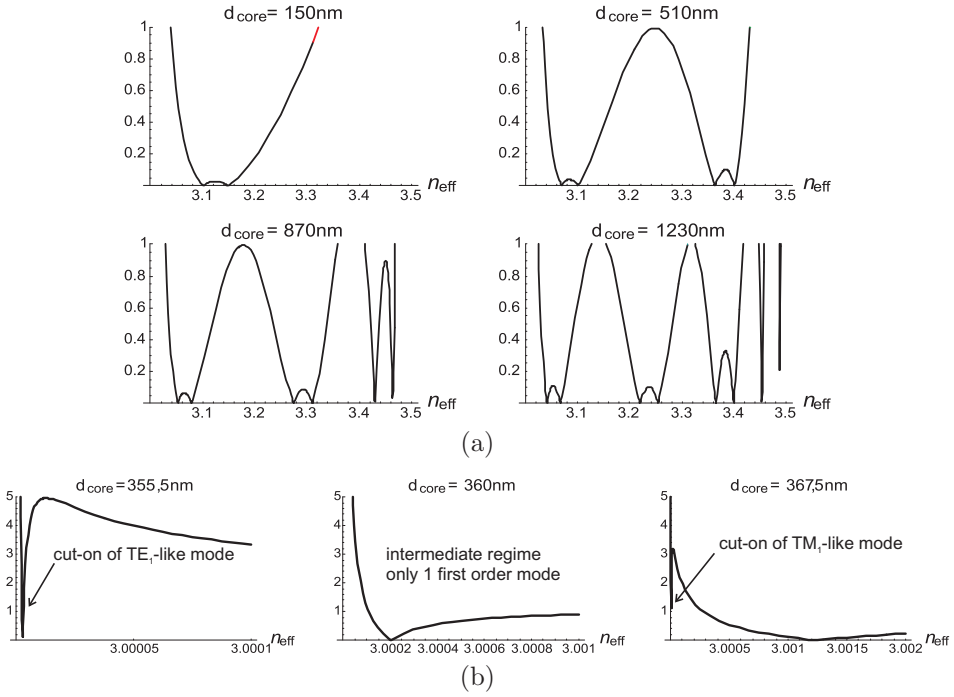


Fig. 2.8: (a) Dispersion relationship for several core thicknesses, evolving from monomodal to quadrimodal regime in steps corresponding to the cut-on thickness $d_{\text{c.o.}}$; (b) Behaviour near cut-on for the first-order modes. At the normal isotropic cut-on thickness (360 nm) only the first TE-like mode has crossed cut-on conditions. Remark that the naming "TE-like" and "TM-like" only indicates that these solutions will converge towards these isotropic modal limits when $Q \to 0$, and has nothing to see with their polarization nature (which is hybrid as will be shown in Fig. 2.10).

First of all it can be remarked that the dispersion relationship correctly predicts the existence of two (hybrid) polarizations for every order of the bound modes. On top of that the multimodality correctly follows the rule for symmetric isotropic waveguides, yielding cut-on of a higher order of the bound modes every time the core thickness is increased by $\Delta d_{\text{c.o.}} = \lambda/2\sqrt{n_{\text{core}}^2 - n_{\text{clad}}^2}$ (in this case this amounts

to $\approx$ 360nm). However, closer inspection near the cut-off thicknesses reveals that there is small correction to the cut-off thicknesses of both polarizations. In a pure symmetric isotropic waveguide, both polarizations (TE and TM in that case) will cut-on simultaneously. Plot (b) in Fig. 2.8 reveals that there is a small difference in the cut-on thicknesses of both polarizations. This is as expected, because, even though the main effect of a Faraday MO layer is a non-reciprocal polarization rotation, there is also a higher order MO change of the effective index of the eigenwaves in the MO layer, with an opposite sign for both polarizations. As a result, the cut-on thickness for both polarizations slightly increases/decreases with respect to the isotropic case. This is illustrated by the occurrence of a small range of thicknesses where only one of the two polarizations has already reached cut-on.

Secondly, investigation of the dispersion relationship at negative values for the effective index proves that this MO waveguide is indeed bidirectional, as predicted by (2.17) ($\boldsymbol{\epsilon}_{tz}(\boldsymbol{\rho}) = \boldsymbol{\epsilon}_{zt}(\boldsymbol{\rho}) = 0$). This is illustrated in Fig. 2.9 for a thickness of 510 nm. It is clearly seen that the dispersion relationship has a perfect even behaviour as a function of n_{eff}. The conditions (2.17) also predict this waveguide to be mirroring. To prove that this is indeed the case, the modal profiles are calculated at a thickness of $d_{\text{core}} = 510$nm.

Fig. 2.10 shows the calculated modal profiles for the ground mode. These modes are calculated using the formalism for open waveguides described in 2.2.2.4. The imaginary field components are indicated by dashed lines, while a full line is used for the real components. Calculating the mode with a backward solution indeed leads to a "mirrored image" of the forward calculated mode[30]. It can be proven that in a symmetrical MO waveguide in Faraday configuration, the y field components must have mutually the same parity, just as the z field components, and both groups must have opposite parity. And (2.59) proves that the x-fields only depend on the y-fields, and thus share their parity. All of this is successfully observed in Fig. 2.10. Furthermore, the observed behaviour of the different field components is in accordance with the Maxwell boundary equations. The discontinuity of the E_x component at the core boundaries is increased with respect to the isotropic case due to the extra MO contribution $\left(\overline{\overline{M}}_1^{-1}\overline{\overline{M}}_2\right)_{11}$ of the core, $\Delta E_x = -jQE_y(\pm d/2) - Z_{\text{vac}} n_{\text{eff}} \Delta\epsilon/(\epsilon_{\text{core}}\epsilon_{\text{clad}}) H_y(\pm d/2)$ (note that both terms are negative imaginary because H_y is also imaginary). The discontinuity of the derivative of E_z at the core boundaries is in agreement with the isotropic case (where $\overline{\overline{A}}_{23}$ changes sign at this boundary). Again the MO contribution ($\overline{\overline{A}}_{21}$) to this discontinuity provides a further increase.
As pointed out above, the modes have a hybrid polarization. None of the six field components is negligible. The reasoning has indicated that the main effect of the MO layer in this lossless case has been to create a small[31] perpendicular field com-

[30] Note that this could also be directly predicted by inspecting the matrices (2.58) and (2.59) with $n_{\text{eff}} \rightarrow n_{\text{eff}}$ and checking how the different field components should behave in order for this substitution to be a symmetry operation.

[31] Looking at the relative values in Fig. 2.10 of E_x and E_y, this "smallness" could be questioned. However, the hypothetical MO material considered here ($Q = 0.01$) is already a very strong one. Comparing this to the values given in Chapter 1, it is observed that only ferromagnetic metals behave this strongly magneto-optically spoken, but then again these also have a very strong

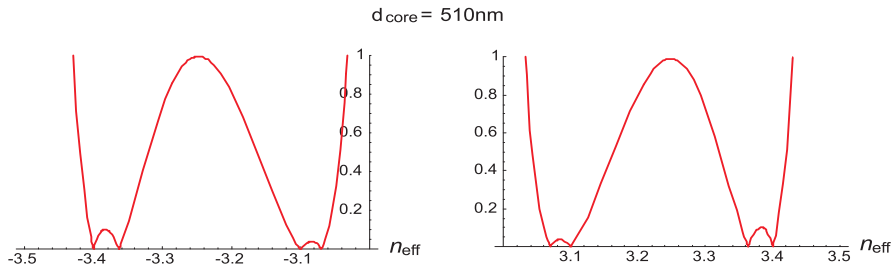
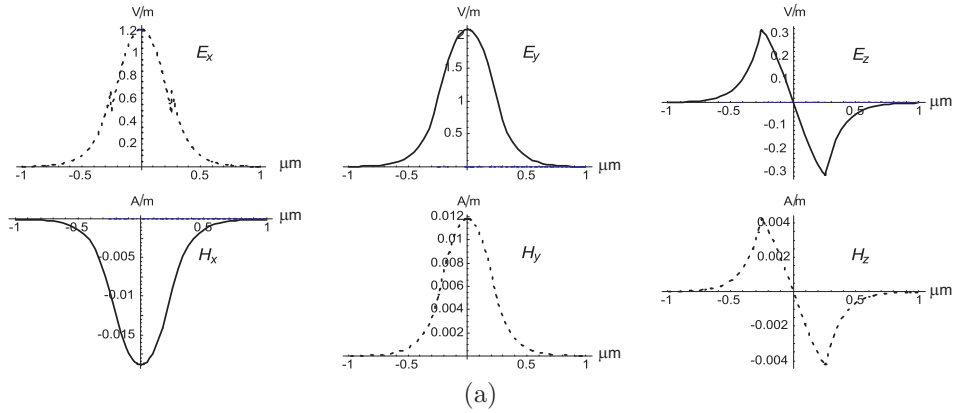
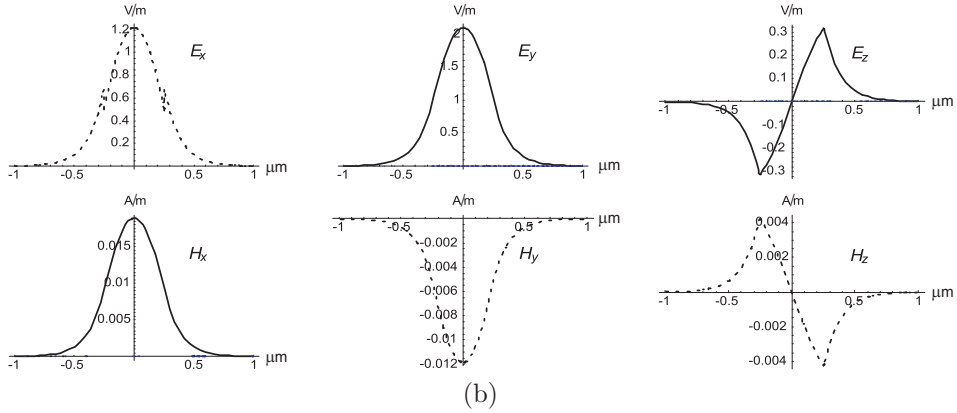


Fig. 2.9: Bidirectionality of the planar waveguide Faraday rotator. For every modal solution $\beta_i = n_{i,\mathrm{eff}} k_0$ there exists an opposite travelling solution with $\beta_j = -\beta_i$.



(a)



(b)

Fig. 2.10: (a) Modal profiles of the zero-order mode ($n_{\mathrm{eff}} = 3.40024$) in the forward direction ($d_{\mathrm{core}} = 510\mathrm{nm}$)(b) and of the corresponding backward travelling bidirectional mode (in the same coordinate system!). As can be seen, the mirroring properties are indeed fulfilled, $\boldsymbol{E}_{t,\mathrm{bw}}(\boldsymbol{\rho}) = \boldsymbol{E}_{t,\mathrm{fw}}(\boldsymbol{\rho})$, $E_{z,\mathrm{bw}}(\boldsymbol{\rho}) = -E_{z,\mathrm{fw}}(\boldsymbol{\rho})$, $\boldsymbol{H}_{t,\mathrm{fw}}(\boldsymbol{\rho}) = -\boldsymbol{H}_{t,\mathrm{fw}}(\boldsymbol{\rho})$ and $H_{z,\mathrm{bw}}(\boldsymbol{\rho}) = H_{z,\mathrm{fw}}(\boldsymbol{\rho})$.

ponent in quadrature. In other words, to transform the original linear polarization to an elliptical one. The hybrid ground-mode depicted in Fig. 2.10 is described by a Jones matrix $\binom{E_x}{E_y} = \binom{+jb}{a}$ with both a and b positive. With the typical Jones calculus conventions, this mode thus has a left-elliptical polarization. Intuitively, it is expected that the other mode with the same zero-order symmetry will have a right-elliptical polarization. This is successfully checked in Fig.2.11 and suggests to label the modes as $\circlearrowleft_i$ (and $\circlearrowright_i$) with $i = 0, 1, 2, \ldots$ indicating the number of nodal points (or thus the symmetry).
The modal profiles given in Fig.2.11 are normalized. This normalization has been performed by calculating the opposite-travelling mode in the adjoint waveguide (obtained by $Q \rightarrow -Q$). It was remarked in the previous section, that normally this normalization is only possible if an extra relationship between the modes in the original waveguide and the adjoint waveguide is available (otherwise the single normalization integral (2.57) does not suffice to fix two amplitudes), and that this more often than not poses a gigantesque theoretical challenge. In this case however, such a relationship is straightforwardly obtained. Indeed, due to Onsager's principle [16] the adjoint permittivity profile of a magneto-optic waveguide is obtained by reversing the magnetization direction. And because the waveguide is also symmetric with respect to the y-axis, $Q \rightarrow -Q$ is equivalent to a physical rotation (with respect to the fixed coordinate system) of the waveguide over π radians with y as a the rotation axis. It is known from physics that an active transformation of any physical problem will transform the solutions of the original problem accordingly into a solution of the transformed problem [17]. The transformed problem is the adjoint waveguide, and as a result the transformed solution is none else but the mutual backward propagating adjoint mode of the original waveguide mode. Therefore the amplitude of the adjoint mode is no longer a degree of freedom in the normalization integral but is equal to the amplitude of the original mode. The transversal components of the transformed mode (needed for (2.57)) are of course obtained by a rotation of the original mode, i.e. $\tilde{E}_x(x) = -E_x(-x)$, $\tilde{E}_y(x) = E_y(-x)$, $\tilde{H}_x(x) = -H_x(-x)$ and $\tilde{H}_y(x) = H_y(-x)$. The normalized modal profiles are then found by dividing them by the square root of the normalization integral.

The final step is to check whether the polarization rotation of this waveguide is indeed non-reciprocal. In order to visualize this properly one ideally would like to calculate the overlap of a linearly polarized mode (say for instance an isotropic TE mode) with the modes of the Faraday rotator and track how the total E_x and E_y components evolve in the positive/negative direction. A rigorously correct derivation of the overlap integrals at a junction between (in general) two anisotropic waveguides is not straightforwardly generalized out of the isotropic case. This is of course because this junction itself has an intrinsic anisotropic character. A correct treatment therefore requires not only the knowledge of the adjoint modes in both directions at both sides of the junction, but also the adjoint junction. As a result every junction is characterized by no less than 8 (!) scattering matrices.

imaginary component for Q due to their losses. A purely real $Q = 0.01$ is thus very hypothetical and is only used here to provide a convincing example of non-reciprocal Faraday rotation, and to justify the use of rigorous modelling rather than perturbational modelling.

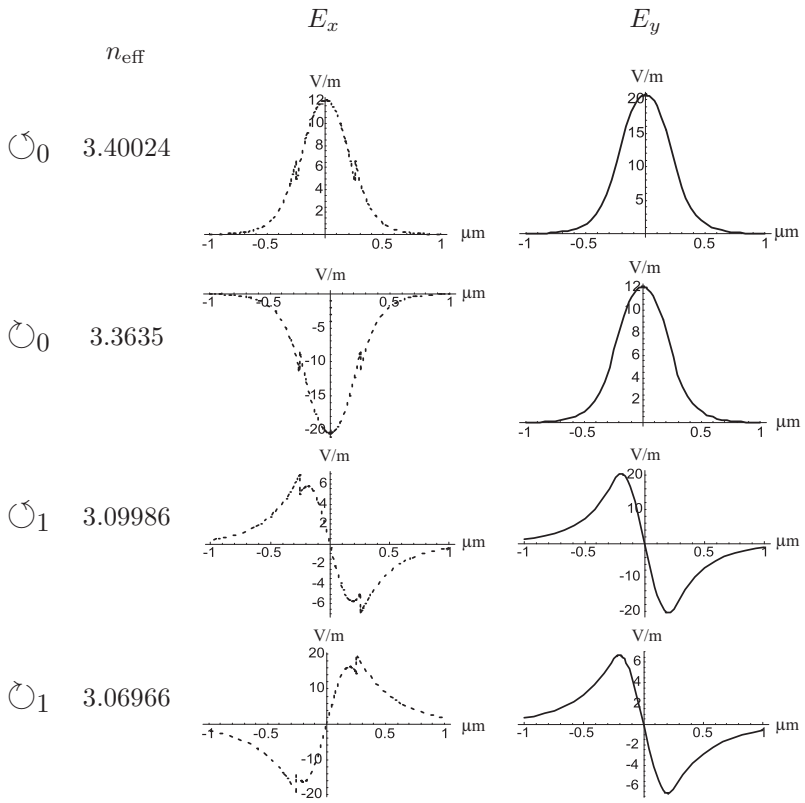


Fig. 2.11: Normalized modal profiles of the transverse electric field components for the symmetric waveguide Faraday rotator of Table 2.1. Note that the right elliptic polarized modes originate from isotropic TM modes with the same order, and the left elliptic polarized modes from the corresponding isotropic TE modes. This is illustrated by the fact that in these normalized profiles, the $\circlearrowleft$ modes have a dominant E_y component, while for the $\circlearrowright$ modes E_x is dominant field component. On top of that, these profiles prove that the normalization is correct, since $E_{y,\circlearrowright_i}$ has exactly the same strength as $E_{x,\circlearrowright_i}$.

If the junction is abrupt (as is most often the case) the adjoint scattering matrices are no longer needed, but still both the original and the adjoint normalized modal spectrum in both forward and backward direction of the two joined waveguides is necessary. The derivation of the formulae for these anisotropic, non-reciprocal scattering matrices is beyond the scope of this text. The interested reader is referred to [1]. We have implemented Olyslager's formulae for the simple case of an abrupt junction between an isotropic slab waveguide and a non-reciprocal magneto-optic slab waveguide.

The forward and backward scattering matrices for the abrupt junction between the magneto-optic waveguide of Table 2.1 and its isotropic equivalent (obtained

by putting $Q = 0$) are calculated for the four bound modes at this thickness. Note that in fact these two matrices actually correspond to two different junctions; the first being the isotropic waveguide at the "left" and the Faraday rotator at the "right", and the second the Faraday rotator at the "left" and the isotropic waveguide at the "right". In general, these two junctions can be entirely different because the junctions are non-reciprocal. However, in this simple case, both waveguides happen to be mirroring so that both matrices will be exactly the same. Note also that obviously both matrices are calculated within a fixed coordinate frame.

$$S_{\text{iso}\to\text{MO}} = S_{\text{MO}\leftarrow\text{iso}} = \begin{pmatrix} 0.862348 & -0.506347j & -2.1307910^{-6} & -3.903910^{-7}j \\ 0.506344 & 0.862358j & -1.3435810^{-6} & 2.0374110^{-7}j \\ 1.8994910^{-6} & 3.2404810^{-6}j & -0.94688 & 0.32249j \\ -5.5489910^{-6} & -9.4236710^{-6}j & -0.32247 & -0.94693j \end{pmatrix} \tag{2.60}$$

There is no coupling between even and odd modes, and the original TE and TM character of $\circlearrowleft$ and $\circlearrowright$ predominantly causes isotropic TE modes to couple with left elliptic modes, and vice versa for TM modes. Finally, within the accuracy of the calculation, there is negligible reflection at this interface. The squares of the elements in each column add up to ≈ 1. To investigate the polarization rotation, we consider an incident isotropic TE mode, and calculate the evolution of the Jones vector in the forward and the backward direction.

$$\begin{aligned} \begin{pmatrix} E_x(x) \\ E_y(x) \end{pmatrix}(z) = e^{-j\frac{\Delta n^{\pm}_{\text{eff},0}}{2}k_0 z} S_{\text{iso}\to\text{MO},11} \begin{pmatrix} E_{\circlearrowleft_0,x}(x) \\ E_{\circlearrowleft_0,y}(x) \end{pmatrix}(0) \\ + e^{+j\frac{\Delta n^{\pm}_{\text{eff},0}}{2}k_0 z} S_{\text{iso}\to\text{MO},21} \begin{pmatrix} E_{\circlearrowleft_0,x}(x) \\ E_{\circlearrowleft_0,y}(x) \end{pmatrix}(0) \end{aligned} \tag{2.61}$$

with $\frac{\Delta n^{\pm}_{\text{eff},0}}{2} = \pm(\frac{n_{\text{eff},\circlearrowleft_0} - n_{\text{eff},\circlearrowright_0}}{2})$. For logical reasons the odd modes are neglected in this expansion. The common phase factor $e^{-jk_0\frac{n_{\text{eff},\circlearrowleft_0}+n_{\text{eff},\circlearrowright_0}}{2}z}$ will not affect the polarization rotation and is left out. The superscripts indicate in which propagation direction the evolution is considered.

In the backward direction the modal profiles of the E_x and E_y components are the same due to the mirroring properties of the Faraday rotator (see for instance Fig. 2.10). Therefore there is no need for a $\pm$ superscript for the Jones vector profiles at $z = 0$. These same mirroring properties cause $\Delta n_{\text{eff},0}$ to change sign upon a change of a propagation direction. These remarks immediately make clear that the evolution of the Jones vector $\left(\begin{smallmatrix} E_x(x) \\ E_y(x) \end{smallmatrix}\right)(z)$ will be independent of the sign of z. In other words, from the viewpoint of the fixed coordinate system, the rotation of the plane of polarization and the ellipticity angle is equal in the forward and the backward direction. From the viewpoint of the propagating wave itself (which is defined by the orientation of a clock when looking into the beam), the ellipticity and the rotation are opposite, hence non-reciprocal. The evolution of the Jones vector components is illustrated in Fig. 2.12. As can be seen from this figure,

one cannot obtain complete TE/TM polarization conversion. At a 90° rotation of the major axis of the polarization ellipse (which is of course a linear degenerate at $z = 0$) the polarization simultaneously obtains its maximum ellipticity. This is the complete opposite of what happens in a bulk magneto-optic Faraday rotator. The obvious reason behind this behaviour is of course the isotropic phase mismatch (birefringence) between the simultaneously propagating left- and right-polarized eigenmodes. In other words, the effective index difference is not an odd function of the Voigt parameter, but also has an isotropic contribution which induces a second polarization conversion, which in a way counteracts the MO polarization conversion. In a bulk MO medium there is no "preferred direction" (as opposed to a layered medium). As a result both natural polarization states in which light can propagate, will have the same isotropic contribution to their effective index, and the polarization conversion will entirely be triggered by the MO effect alone. And it was explained in Chapter 1 that the pure MO (lossless) Faraday rotation (unhindered by birefringence effects) will always leave the original state of polarization untouched (i.e. add no ellipticity to a linear polarization), hence will allow perfect TE/TM conversion. In a layered waveguide the isotropic birefringence is reduced by increasing the width of the core. In that case the results of bulk magneto-optics will be recovered, not only because $n_{\text{eff},\circlearrowleft,0} - n_{\text{eff},\circlearrowright,0} \sim Q$, but also, and mainly because both $\circlearrowleft_0$ and $\circlearrowright_0$ will be excited to the same amount. Finally, also note that the periodicity of the polarization conversion is found by remarking that in (2.61) the same polarization state will be recovered when both components of the Jones matrix changes their sign, i.e. after each propagation over $2\pi/(k_0 \Delta n_{\text{eff},0})$. This is in perfect agreement with the results of bulk magneto-optics.

This simple example has illustrated most of the basic concepts and subtleties of modelling anisotropic waveguides: the solution of their eigenspectrum, the calculation of the modal profiles, their symmetries, both spatial and spectral (bidirectional behavior), their subtle normalization, the possible non-reciprocical behaviour both of the waveguides themselves and the junction between them, ... The here described algorithm has been implemented for arbitrary anisotropic slab waveguides in a MATHEMATICA environment.

As a final remark we would like to draw the attention of the reader to the problem of modelling anisotropic waveguides with a higher dimensionality, such as ridge waveguides or anisotropic photonic crystal waveguides. Section 2.2.2.2 has made clear that the main difficulty in modelling anisotropic waveguides is the complicated (often non-reciprocal) intermixing of the local eigenpolarizations at each interface. This necessitated the use of 4-dimensional transfer matrices. When an extra dimension is added to the waveguide problem, the resulting waveguide is most often made up of piecewise concatenation of slabs. A very popular method (known as the Transverse Resonance Method or Transverse Mode Matching Method [18]) is then to calculate in the direction perpendicular to the concatenated slabs a "resonance matrix" using all the modal spectra of the concatenated slabs. This is a generalization of the resonance calculation for 1D-slabs, with the important difference that now at each slab interface, matrices of a dimensionality of an order equal to the number of considered slab modes must be considered. Already for isotropic waveguides, this is a considerable increase in numerical effort. However

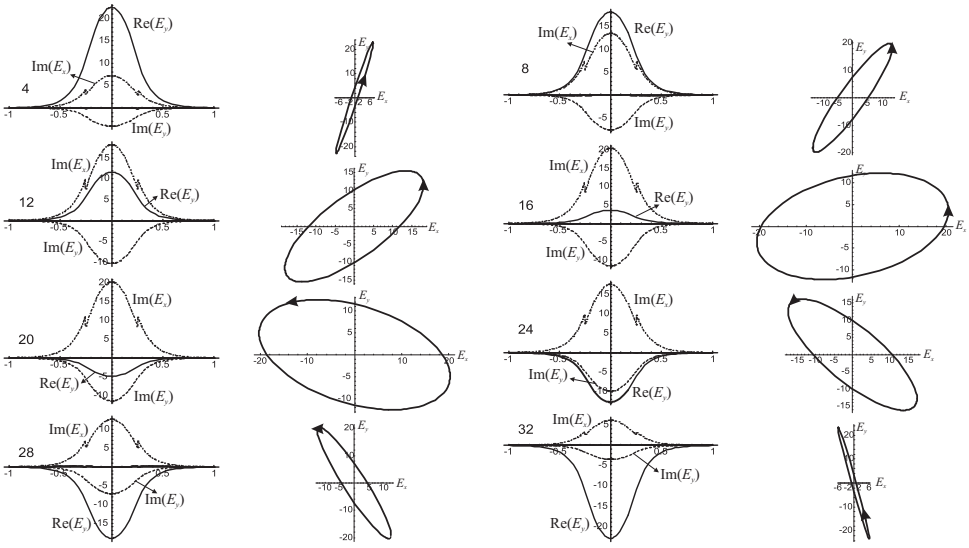


Fig. 2.12: Evolution of the field components (note that $\Re(Ex)$ is everywhere 0) as a function of propagation distance (in μm) when a TE eigenmode is launched at $z = 0$, plotted in a fixed coordinate system, and corresponding polarization ellipse at each point (traced by the field in the center of the waveguide, i.e.$x = 0$). Note that, as explained in the text, a TE mode launched in the backward direction will after a distance $z = -L$ have exactly the same distribution of field components as when launched in the forward direction and propagated over the same distance L. However, when facing the beam this will amount to opposite ellipticity and opposite rotation.

when the concatenated slabs become anisotropic, things get really nasty, as now not only in each slab the adjoint modal spectrum must be solved, but also, and mainly, because at each slab interface one gets an anisotropic waveguide junction. And as was pointed out above, these anisotropic junctions will not only induce important polarization mixing but can on top of that behave non-reciprocally. In any case, it is clear that a generalization of Transverse Mode Matching methods towards anisotropic waveguides is not a task one is easily tempted into.

2.2.3 Perturbation theory modelling

2.2.3.1 Introduction

When the dielectric permittivity profile of the non-reciprocal anisotropic waveguide can be considered to be a 'small' deviation of an isotropic profile, then perturbation techniques enter the picture, when solving for the modal spectrum of the non-reciprocal waveguide. Actually, all one needs to apply any perturbation tech-

nique is an 'unperturbed' starting point for which the modal spectrum is known[32]. Therefore, the theory that will be developed here is more general, and allows the unperturbed waveguide itself to be a general anisotropic waveguide.
The technique that is presented here is called the *modal expansion method* and is, like most perturbation techniques, based on the completeness of the modal spectrum of the unperturbed waveguide. Completeness allows to express the unknown fields in the 'real' waveguide as a (possibly z-dependent) linear combination of the unperturbed modes. The goal is then to derive a (infinite) coupled set of first-order evolution equations for these expansion coefficients, by "also properly expanding the perturbation" in the unperturbed modes. The derivation of these evolution equations is rigorous, and is just a reformulation of the Maxwell equations. Approximative perturbation techniques will only come into play once attempts are made at integrating the set of coupled first-order differential equations.
In section 2.2.3.2, the modal expansion for the unknown fields will be detailed. The evolution equations will be derived in section 2.2.3.3, and finally in section 2.2.3.4 it will be detailed how these can be integrated to various orders in the perturbation.

2.2.3.2 Modal expansion for the perturbed fields

If we use the notation $\underline{\epsilon}_r^{(0)}(\boldsymbol{\rho})$ for the permittivity profile of the unperturbed waveguide and $\underline{\epsilon}_r(\boldsymbol{\rho})$ for the actual permittivity profile, the Maxwell equations to be solved can be written as

$$\begin{aligned} \boldsymbol{\nabla} \times \mathbf{e}(\mathbf{r}) &= -j\omega\mu_0 \mathbf{h}(\mathbf{r}) \\ \boldsymbol{\nabla} \times \mathbf{h}(\mathbf{r}) &= j\omega\epsilon_0 \underline{\epsilon}_r^{(0)}(\boldsymbol{\rho}) \cdot \mathbf{e}(\mathbf{r}) + \mathbf{j}(\mathbf{r}) \end{aligned} \tag{2.62}$$

where

$$\mathbf{j}(\mathbf{r}) = j\omega\epsilon_0 (\underline{\epsilon}_r(\boldsymbol{\rho}) - \underline{\epsilon}_r^{(0)}(\boldsymbol{\rho})) \cdot \mathbf{e}(\mathbf{r}) \tag{2.63}$$

In other words, the unknown fields of the 'real' waveguide are those fields that are sustained by the perturbation current source $\mathbf{j}(\boldsymbol{\rho})$ in the unperturbed waveguide[33]. Formulating the problem in this way, makes it obvious to try and construct a solution by means of a series expansion in the unperturbed eigenmodes. The eigenmodes of the unperturbed (anisotropic) waveguide are written as

$$\mathbf{e}_i^{(0)}(\mathbf{r}) = \mathbf{E}_i^{(0)}(\boldsymbol{\rho}) \exp(-j\beta_i z) \tag{2.64a}$$

$$\mathbf{h}_i^{(0)}(\mathbf{r}) = \mathbf{H}_i^{(0)}(\boldsymbol{\rho}) \exp(-j\beta_i z) \tag{2.64b}$$

[32] This statement is only partially correct. There are several perturbation techniques that do not rely on any particular expansion basis for the unknown fields. These methods are known as global perturbation methods. Their major advantage is that no assumption have to be made about the convergence of the modal expansion series in the here presented perturbation technique [14]. This will be pointed out later in this section

[33] The perturbation current source is of course dependent on the unknown fields themselves. Otherwise, the solution would just be a superposition of the known eigenmodes of the 'unperturbed' waveguide. Therefore the perturbation current source is obviously also an unknown.

As proven in section 2.1.2 the eigenmodes of an anisotropic waveguide are normalized along (2.11). Normalization thus requires the corresponding (contra-directional) adjoint mode ($\tilde{\beta}_i = -\beta_i$) of the adjoint unperturbed waveguide.

$$\tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}) = \tilde{\mathbf{E}}_i^{(0)}(\boldsymbol{\rho}) \exp(j\beta_i z), \tag{2.65a}$$

$$\tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}) = \tilde{\mathbf{H}}_i^{(0)}(\boldsymbol{\rho}) \exp(j\beta_i z), \tag{2.65b}$$

so that the orthonormalization equation reads[34]

$$-\frac{1}{4}\iint (\mathbf{e}_i^{(0)}(\mathbf{r}) \times \tilde{\mathbf{h}}_j^{(0)}(\mathbf{r}) - \tilde{\mathbf{e}}_j^{(0)}(\mathbf{r}) \times \mathbf{h}_i^{(0)}(\mathbf{r})) \cdot \mathbf{u}_z \mathrm{d}S = N_i \delta_{ij}, \tag{2.66}$$

with $N_i = \pm 1$ the normalization constant of the i$^{\text{th}}$ mode.

Due to the completeness of the modes of the unperturbed waveguide, and the fact that the unknown fields are sustained by the 'unperturbed' waveguide according to (2.62), the transversal components of the unknown fields can be written as[35]

$$\begin{pmatrix} \mathbf{e}_t(\mathbf{r}) \\ \mathbf{h}_t(\mathbf{r}) \end{pmatrix} = \sum_{i \in \mathcal{N}} C_i(z) \begin{pmatrix} \mathbf{e}_{t,i}^{(0)}(\mathbf{r}) \\ \mathbf{h}_{t,i}^{(0)}(\mathbf{r}) \end{pmatrix} \tag{2.67}$$

It is fundamental that the expansion coefficients $C_i(z)$ are z-dependent. After all, if that wasn't the case, the sought solution would just be a superposition of eigenmodes of the unperturbed waveguide, and hence an eigenmode itself of this waveguide. Using (2.66), the expansion coefficients are found through[36],[37]

$$C_i(z) = -\frac{1}{4N_i}\iint (\mathbf{e}(\mathbf{r}) \times \tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}) - \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}) \times \mathbf{h}(\mathbf{r})) \cdot \mathbf{u}_z \mathrm{d}S \tag{2.68}$$

The question is if these coefficients can also be used for a series expansion for the longitudinal unknown field components, $e_z(\mathbf{r})$ and $h_z(\mathbf{r})$. It is known from optical waveguide theory that this is the case in every region of the waveguide outside the source regions [19]. For the transversal components the expansion (2.67) even applies in the interior of the sources. However, for the longitudinal field components inside the source region, one needs to take the longitudinal source component $j_z(\mathbf{r})$ into account. And as (2.63) indicates that the source itself

[34] The factor $-\frac{1}{4}$ is needed to achieve consistency with the normalization equation of an isotropic waveguide.

[35] $\mathcal{N}$ is a notation for the complete set of eigenmodes of the unperturbed waveguide. Note that this means both forward and backward propagating eigenmodes! Even if these are just each others mirror image. This is because the unknown fields are excited by a current source, which will in general excite both forward and backward propagating eigenmodes. (let's for the moment disregard the fact that this current source is an unknown itself.)

[36] Remark that only transversal field components contribute to this integral

[37] It is also important to note that this integral and the expansion (2.67) are only defined in the cross-section of the unperturbed waveguide. This is not a trivial remark, since the perturbations do not necessarily have to be material parameter perturbations, but can also take the form of waveguide boundary variations (like for instance, a vibrating fibre or a deformation of a closed metallic waveguide). From now on we will focus on open waveguides with pure material parameter perturbations.

is dependent on the unknown fields, and thus "fills" the entire space, it is clear that it is not so straightforward to see how the expansion coefficients $C_i(z)$ can also be used for a series expansion for $(e_z(\mathbf{r}), h_z(\mathbf{r}))$. If such an expansion can be derived, then the Maxwell equations (2.62) describing the actual waveguide can be reformulated in terms of unknown scalar coefficients $C_i(z)$ depending only on one (longitudinal) coordinate, instead of two three-dimensional vector fields depending on a three-dimensional vector coordinate! A considerable simplification, that comes of course at the price of prerequisite knowledge of the unperturbed eigenmodes. In order to derive such an expansion for $(e_z(\mathbf{r}), h_z(\mathbf{r}))$ in terms of $C_i(z)$ and the unperturbed eigenmodes, consider the z-component of the Maxwell equations (2.62)

$$\boldsymbol{\nabla}_t \cdot (\mathbf{u}_z \times \mathbf{e}_t(\mathbf{r})) = j\omega\mu_0 h_z(\mathbf{r}) \tag{2.69a}$$

$$\boldsymbol{\nabla}_t \cdot (\mathbf{u}_z \times \mathbf{h}_t(\mathbf{r})) = -j\omega\epsilon_0 \boldsymbol{\epsilon}_{3*}^{(0)}(\boldsymbol{\rho}) \cdot \mathbf{e}(\mathbf{r}) - j_z(\mathbf{r}) \tag{2.69b}$$

The notation $\boldsymbol{\epsilon}_{3*}^{(0)}$ indicates the row vector formed by the third row in the permittivity tensor. The same equations hold for the unperturbed eigenmodes, but obviously without the current source component. Thus,

$$\boldsymbol{\nabla}_t \cdot (\mathbf{u}_z \times \mathbf{e}_{t,i}^{(0)}(\mathbf{r})) = j\omega\mu_0 h_{z,i}^{(0)}(\mathbf{r}) \tag{2.70a}$$

$$\boldsymbol{\nabla}_t \cdot (\mathbf{u}_z \times \mathbf{h}_{t,i}^{(0)}(\mathbf{r})) = -j\omega\epsilon_0 \boldsymbol{\epsilon}_{3*}^{(0)}(\boldsymbol{\rho}) \cdot \mathbf{e}_i^{(0)}(\mathbf{r}) \tag{2.70b}$$

Substituting the expansion (2.67) in the equations (2.69a) and (2.69b),

$$\boldsymbol{\nabla} \cdot \Big(\sum_{i\in\mathcal{N}} C_i(z) \mathbf{u}_z \times \mathbf{e}_{t,i}^{(0)}(\mathbf{r}) \Big) = j\omega\mu_0 h_z(\mathbf{r}) \tag{2.71a}$$

$$\begin{aligned} \boldsymbol{\nabla} \cdot \Big(\sum_{i\in\mathcal{N}} C_i(z) \mathbf{u}_z \times \mathbf{h}_{t,i}^{(0)}(\mathbf{r}) \Big) = &- j\omega\epsilon_0 \boldsymbol{\epsilon}_{3*}^{(0)}(\boldsymbol{\rho}) \cdot \sum_{i\in\mathcal{N}} C_i(z) \mathbf{e}_{t,i}^{(0)}(\mathbf{r}) \\ &- j\omega\epsilon_0 \epsilon_{33}^{(0)}(\boldsymbol{\rho}) e_z(\mathbf{r}) - j_z(\mathbf{r}), \end{aligned} \tag{2.71b}$$

and assuming that these series can be derived termwise, the left- hand side of these equations can be written as (using (2.70))

$$\sum_{i\in\mathcal{N}} C_i(z) \boldsymbol{\nabla} \cdot (\mathbf{u}_z \times \mathbf{e}_{t,i}^{(0)}(\mathbf{r})) = j\omega\mu_0 \sum_{i\in\mathcal{N}} C_i(z) h_{z,i}^{(0)}(\mathbf{r}) \tag{2.72a}$$

$$\sum_{i\in\mathcal{N}} C_i(z) \boldsymbol{\nabla} \cdot (\mathbf{u}_z \times \mathbf{h}_{t,i}^{(0)}(\mathbf{r})) = -j\omega\epsilon_0 \boldsymbol{\epsilon}_{3*}^{(0)}(\boldsymbol{\rho}) \cdot \sum_{i\in\mathcal{N}} C_i(z) (\mathbf{e}_{t,i}^{(0)}(\mathbf{r}) + \mathbf{u}_z e_{z,i}^{(0)}(\mathbf{r})) \tag{2.72b}$$

Combining (2.71) and (2.72) finally yields

$$h_z(\mathbf{r}) = \sum_{i\in\mathcal{N}} C_i(z) h_{z,i}^{(0)}(\mathbf{r}) \tag{2.73a}$$

$$j\omega\epsilon_0 \epsilon_{33}^{(0)}(\boldsymbol{\rho}) e_z(\mathbf{r}) = j\omega\epsilon_0 \epsilon_{33}^{(0)}(\boldsymbol{\rho}) \sum_{i\in\mathcal{N}} C_i(z) e_{z,i}^{(0)}(\mathbf{r}) - j_z(\mathbf{r}) \tag{2.73b}$$

If the longitudinal component of the perturbation current source of (2.63) is now written out in its composing field components making use of the known transversal eigenmode expansion (2.67),

$$j_z(\mathbf{r}) = j\omega\epsilon_0(\boldsymbol{\epsilon}_{3*}(\boldsymbol{\rho})-\boldsymbol{\epsilon}_{3*}^{(0)}(\boldsymbol{\rho}))\cdot\sum_{i\in\mathcal{N}} C_i(z)\mathbf{e}_{t,i}^{(0)}(\mathbf{r})+j\omega\epsilon_0(\epsilon_{33}(\boldsymbol{\rho})-\epsilon_{33}^{(0)}(\boldsymbol{\rho}))e_z(\mathbf{r}), \quad (2.74)$$

one finally gets, after substitution of the above equation in (2.73), the expansion series for the longitudinal unknown field components in terms of $C_i(z)$ and the unperturbed eigenmodes. These are summarized together with the transversal expansion series in the following equations and form the basis of the here established modal expansion perturbation method. The expansion for the perturbation current source, resulting from these series, is also included.

$$\mathbf{e}_t(\mathbf{r}) = \sum_{i\in\mathcal{N}} C_i(z)\mathbf{e}_{t,i}^{(0)}(\mathbf{r}) \quad (2.75)$$

$$e_z(\mathbf{r}) = \sum_{i\in\mathcal{N}} C_i(z)\left[\frac{\epsilon_{33}^{(0)}(\boldsymbol{\rho})}{\epsilon_{33}(\boldsymbol{\rho})}e_{z,i}^{(0)}(\mathbf{r}) - \frac{(\boldsymbol{\epsilon}_{3*}(\boldsymbol{\rho})-\boldsymbol{\epsilon}_{3*}^{(0)}(\boldsymbol{\rho}))}{\epsilon_{33}(\boldsymbol{\rho})}\cdot\mathbf{e}_{t,i}^{(0)}(\mathbf{r})\right] \quad (2.76)$$

$$\mathbf{h}_t(\mathbf{r}) = \sum_{i\in\mathcal{N}} C_i(z)\mathbf{h}_{t,i}^{(0)}(\mathbf{r}) \quad (2.77)$$

$$h_z(\mathbf{r}) = \sum_{i\in\mathcal{N}} C_i(z)h_{z,i}^{(0)}(\mathbf{r}) \quad (2.78)$$

$$\mathbf{j}(\mathbf{r}) = j\omega\epsilon_0\delta\underline{\epsilon}(\boldsymbol{\rho})\cdot\sum_{i\in\mathcal{N}} C_i(z)\left[\mathbf{e}_{t,i}^{(0)}(\mathbf{r})+\mathbf{u}_z\Big(\frac{\epsilon_{33}^{(0)}(\boldsymbol{\rho})}{\epsilon_{33}(\boldsymbol{\rho})}e_{z,i}^{(0)}(\mathbf{r})-\frac{\delta\boldsymbol{\epsilon}_{3*}(\boldsymbol{\rho})}{\epsilon_{33}(\boldsymbol{\rho})}\cdot\mathbf{e}_{t,i}^{(0)}(\mathbf{r})\Big)\right] \quad (2.79)$$

2.2.3.3 The modal evolution equations

Now that the modal expansion series have been introduced for the unknown fields, the unknowns of the waveguide problem have been converted from two vector fields into an (infinite) set of scalar expansion coefficients $C_i(z)$. The next step is thus to reformulate the Maxwell equations (2.62) in terms of these coefficients. This can be done by using the Lorentz reciprocity equation (2.5) applied on the unknown fields $(\mathbf{e}(\mathbf{r}), \mathbf{h}(\mathbf{r}))$ and an arbitrary unperturbed adjoint eigenmode i of the adjoint unperturbed waveguide $(\tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}), \tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}))$.

$$\boldsymbol{\nabla}\cdot(\mathbf{e}(\mathbf{r})\times\tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}) - \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r})\times\mathbf{h}(\mathbf{r})) = \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r})\cdot\mathbf{j}(\mathbf{r}) \quad (2.80)$$

Recall that the Lorentz reciprocity equation is derived directly out of the Maxwell rotor equations. Therefore, the above equation is a rigorous equivalent of the original waveguide problem. If both terms of this differential identity are integrated

over an arbitrary cross-section of the actual waveguide[38],

$$\begin{aligned}&\iint \boldsymbol{\nabla}_t \cdot \Big(\mathbf{e}(\mathbf{r}) \times \tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}) - \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}) \times \mathbf{h}(\mathbf{r})\Big)_t \mathrm{d}S + \ldots \\ &\iint \frac{\mathrm{d}}{\mathrm{d}z}\Big(\mathbf{e}(\mathbf{r}) \times \tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}) - \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}) \times \mathbf{h}(\mathbf{r})\Big) \cdot \mathbf{u}_z \mathrm{d}S = \iint \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}) \cdot \mathbf{j}(\mathbf{r}) \mathrm{d}S,\end{aligned} \tag{2.81}$$

and a Green integral theorem is used[39], this Lorentz reciprocity identity reduces to

$$-4N_i \frac{\mathrm{d}}{\mathrm{d}z} C_i(z) = -\oint_c \Big(\mathbf{e}(\mathbf{r}) \times \tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}) - \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}) \times \mathbf{h}(\mathbf{r})\Big)_t \cdot \mathbf{u}_n \mathrm{d}c + \iint \tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}) \cdot \mathbf{j}(\mathbf{r}) \mathrm{d}S \tag{2.82}$$

Here it has been assumed that in the second term of the left-hand side of (2.81) integration and differentiation can be interchanged, so that this term is nothing else but $-4N_i$ times the z-derivative of the expansion coefficient formula (2.68). The contour integral (originating from the first term of (2.81)) will always vanish for the types of perturbation that are considered here. For an open waveguide, for which the contour is to be evaluated at infinity, this is direct consequence of the physical radiation boundary condition at infinity, causing all acceptable fields to decay towards zero at infinity[40]. For a closed waveguide, the contour-integral will disappear if it is assumed that the surface impedance at the enclosing boundary is not changed[41].

By finally inserting the modal expansion for the perturbation current source (see (2.79)) into this integrated lorentz identity, the following (infinite) set of coupled first-order differential equations is obtained

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}z} C_i(z) = -j \sum_{k \in \mathcal{N}} A_{ik} \exp[j(\beta_i - \beta_k)z] C_k(z) \quad (i \in \mathcal{N})} \tag{2.83}$$

with

$$A_{ik} = \frac{\omega \epsilon_0}{4N_i} \iint \Big[\tilde{\mathbf{E}}_i^{(0)}(\boldsymbol{\rho}) \cdot \delta\underline{\epsilon}(\boldsymbol{\rho}) \cdot (\mathbf{E}_{t,k}^{(0)}(\boldsymbol{\rho}) + \mathbf{u}_z \hat{E}_{z,k}^{(0)}(\boldsymbol{\rho}))\Big] \mathrm{d}S \tag{2.84}$$

[38] Remember that we are only considering open waveguides, or closed waveguides without a deformation of the waveguide boundary and without a change of the surface impedance at this boundary. As a consequence both the unperturbed adjoint mode i and the unknown fields exist in the same region and satisfy the same boundary conditions, either at infinity or at the metallic boundary.

[39] $\iint_S \boldsymbol{\nabla}_t \cdot \mathbf{f}_t \mathrm{d}S = \oint_c \mathbf{f}_t \cdot \mathbf{u}_n \mathrm{d}c$ [4] with $\mathbf{u}_n$ the outward pointing normal unit vector to the contour c enclosing the waveguide cross-section S.

[40] For radiating fields $(\mathbf{e}(\mathbf{r}), \mathbf{h}(\mathbf{r}))$ and $(\tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}), \tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}))$ which are only finite at infinity, the occurrence of the contour-integral can be avoided by considering (2.80) in the sense of distributions operating on the set of functions $(\tilde{\mathbf{e}}_i^{(0)}(\mathbf{r}), \tilde{\mathbf{h}}_i^{(0)}(\mathbf{r}))$, now considered as test functions. This implies that these are zero at infinity.

[41] This can be seen by expanding the integrand and writing the electric fields on the boundary contour with the use of the surface impedance of this boundary. The two terms will then cancel because the surface impedance of the adjoint medium is the matrix adjoint of the surface impedance of the original waveguide.

Here the fields of the unperturbed eigenmodes have been explicitly written out as the product of their modal profiles and propagator factor (along (2.64) and (2.65)). $\hat{E}^{(0)}_{z,k}(\boldsymbol{\rho})$ is short-hand notation for the z-component of the modal expansion (2.79) for the perturbation current source, $\frac{\epsilon^{(0)}_{33}(\boldsymbol{\rho})}{\epsilon_{33}(\boldsymbol{\rho})} E^{(0)}_{z,i}(\boldsymbol{\rho}) - \frac{\delta\boldsymbol{\epsilon}_{3*}(\boldsymbol{\rho})}{\epsilon_{33}(\boldsymbol{\rho})} \cdot \mathbf{E}^{(0)}_{t,i}(\boldsymbol{\rho})$.

The evolution equations of (2.83) are a completely rigorous equivalent formulation of the Maxwell equations (2.62). In this form they bare a close resemblance to the set of coupled mode equations that are normally encountered in coupled mode theory. Actually, (2.83) is just a specialized version of the set of first order differential equations derived in coupled-mode theory [20]. The phenomenon of power transfer that is commonly addressed by coupled-mode theory is in this case the exchange of power between the (unperturbed) modes of a single waveguide subjected to a longitudinally homogeneous perturbation. The constant coefficients A_{ik} in this equation can correctly be considered the coupling coefficients induced by the tensorial permittivity perturbation $\delta\underline{\epsilon}$.[42] This becomes even clearer when the evolution equations are further elaborated. If the rapid variation of the propagator factors in (2.83) are included in the expansion coefficients $X_i(z) = C_i(z)\exp(-j\beta_i z)$, one obtains

$$\frac{\mathrm{d}}{\mathrm{d}z} X_i(z) = -j \sum_{k\in\mathcal{N}} \Big(\beta_i \delta_{ik} + A_{ik}\Big) X_k(z) \quad (i \in \mathcal{N}), \tag{2.85}$$

or, in matrix form,

$$\frac{\mathrm{d}}{\mathrm{d}z} \overline{X}(z) = -j(\overline{\overline{G}} + \overline{\overline{A}})\overline{X}(z) \text{ with, } \quad \overline{\overline{G}} = \beta_i \delta_{ik} \tag{2.86}$$

This form is completely equivalent with the equations encountered in coupled-mode theory and confirms that the matrix $\overline{\overline{A}}$ is indeed the coupling matrix. Also remark that this is a constant matrix due to the longitudinal homogeneity of the perturbation.

The next step in this treatment of anisotropic waveguides, is the actual perturbational resolution of the above set of evolution equations. This will be elaborated in the next section.

[42] There is actually a small difference between the exact form of the mode coupling coefficients A_{ik}, when calculated with the formula of (2.84), and the form obtained by applying the formulae for the coupling coefficients encountered in standard texts on coupled-mode theory [21]. But this is mainly the result of the choice of trial fields and stationary functionals used for the derivation of the coupled-mode equations. As is pointed by Vassallo [22], it is always possible to reproduce the formulae of (2.84) for the coupling coefficients by changing the stationary functional (that is equivalent with the Maxwell equations for the waveguide problem under study) or by changing the trial field. But this is not a question of which is the "best" formulae. It all depends on the investigated problem. The main difference between the here developed version for the coupling matrix A_{ik} and the one normally encountered in coupled-mode theory, is that the expansion used for the unknown fields ((2.75) to (2.78)) are rigorous, while for the trial fields in the variational formulation of the coupled-mode theory, one normally uses an expansion that is as simple as possible. Most of the time this is just $\sum C_i \mathbf{e}^{(0)}_i(\mathbf{r})$. This is obviously not rigorous, since the closure operator for any waveguide is only defined for 4-component vectors and not for 6-component vectors.

2.2.3.4 Solving the evolution equations

The perturbations that are considered in this section 2.2.3 are longitudinally homogeneous, because the electromagnetic problem to be solved is basically a waveguide problem. Consequently the permittivity profile, including the perturbation, must remain invariant along the propagation direction. As a result the perturbed system is again a waveguide whose eigenmodes have a $\exp(-j\Gamma z)$ z-dependence. This implies that the expansion coefficients, $X_i(z)$, for each one of these "new" eigenmodes, have a known z-dependence, and the evolution equation reduces to a pure matrix eigenproblem[43]. The commonexp$(-j\Gamma z)$ is divided away in the next equation.

$$\Gamma\overline{X} = \left(\overline{\overline{G}} + \overline{\overline{A}}\right)\overline{X} \tag{2.87}$$

The propagation constants and the modal profiles of the eigenmodes of the perturbed waveguide are thus found by calculating the eigenvectors and eigenvalues of the matrix $\overline{\overline{G}}+\overline{\overline{A}}$. Even though this is a theoretically important conclusion, it is not very practical to use for the calculation of the perturbed propagation constants as it requires the prior calculation of a matrix of infinite dimensionality. Therefore, one usually resorts to the use of perturbational schemes as developed in quantum mechanics for the calculation of the eigenvalues. The method that is mostly used is based on the expansion of Γ and $\overline{X}$ in terms of powers of the perturbation, the so-called Rayleigh-Schrödinger technique. However, that technique requires some introductory conventions about how to define a scalar product in the X-vector space, and about how to cope with orthogonality in the different orders of the perturbation . This is too technical for the scope of this text. A correct treatment of this can be found in [23] and [24]. Here, a simplified, pragmatical derivation of this technique will be presented[44].

Suppose a mode i of the unperturbed waveguide is considered. When the perturbation $\delta\underline{\underline{\epsilon}}$ is turned off, all expansion coefficients X will be zero except $X_i = 1$, and the propagation constant Γ in (2.87) will be equal to the unperturbed propagation constant β_i of the unperturbed mode i. If the perturbation is turned on, it assumed that $X_i = 1$ and that $\Gamma - \beta_i$ and $X_k, (k \neq i)$ are small numbers of the order of the perturbation (the matrix elements A_{ij} are of course also of the order of the perturbation). Inserting the assumption $X_i = 1$ in the matrix equation (2.87) gives

$$(\Gamma - \beta_k)X_k = A_{ki} + \sum_{l\neq i} A_{kl}X_l \qquad (\forall k, l) \tag{2.88}$$

The summation is of second order in the perturbation, which can be neglected if

[43] This is of course a very elaborate explanation for the pure mathematical fact that any first order differential matrix equation with constant coefficients, as the one in (2.86), is solved by diagonalizing the coefficient matrix

[44] the technique presented here is only valid for guided modes. The perturbation calculation of the radiation spectrum of a waveguide is not a simple task and requires advanced quantum-mechanical perturbation techniques, such as the Lippman-Schwinger equation [24].

it converges fast enough, so that up to first order in the perturbation one finds

$$\begin{aligned} \Gamma &= \beta_i + A_{ii} + \mathcal{O}(2) \\ X_i &= 1 \\ X_k &= \frac{A_{ki}}{\Gamma - \beta_k} + \mathcal{O}(2) = \frac{A_{ki}}{\beta_i - \beta_k} + \mathcal{O}(2) \qquad (\forall k \neq i) \end{aligned} \tag{2.89}$$

Iterating this procedure allows to construct formulae for Γ and $X_k (k \neq i)$ up to any order in the perturbation. For simplicity it was assumed that the unperturbed guided mode is non-degenerate. Details on discrete, degenerate perturbation theory are again to be found in [24].

Before illustrating these formulae with some practical examples, it is important to make some general remarks.

- No criterion has been provided to check when the perturbation $\delta\underline{\epsilon}$ can be considered weak enough so that the summation in (2.88) can rightly be considered of second order. In other words, to check whether this series indeed converges rapidly. For, if it wouldn't, there is a possibility that the sum behaves globally as a first order term. And then it would have been unjustly been neglected. The perturbation can be considered weak

  - either because the material parameters are only very slightly changed in an otherwise arbitrary volume of the waveguide. This is the most commonly encountered situation. In this case it is realistic to assume that the perturbation formulae of (2.89) do hold;

  - or because the perturbation volume is a very small fraction of the waveguide cross-section, while the perturbation of the material parameters in that region can be arbitrarily large. In this case, the formulae of (2.89) usually run into problems. The reason for this is very simple. If the perturbation of the material parameters is locally strong, the true field components normal to the interface of the perturbation region will show a large extra discontinuity (due to the standard required continuity of $(\underline{\epsilon} \cdot \mathbf{e})_n$). This implies that in the expansion (2.67) many terms will contribute. As a result neglecting the series in (2.88) leads to the unjustified cancellation of a first-order contribution. Using the formulae $\Gamma = \beta_i + A_{ii}$ in this case is as if one makes the assumption that the normal real field components are continuous over the interface! Of course in the previous case, the same "error" is made, but there the extra discontinuous jump of the real field is much smaller and the main perturbational correction comes from the integration over the large volume. In this case (i.e. a thin film but strong material parameter perturbation) the perturbational correction A_{ii} is mainly caused by the interfaces and there a big error occurs. For these cases the formulae can only be corrected by an *ab initio* change in the expansion formulae [19].

- The first-order correction to the modal profile of the guided modes can only be approximatively calculated via (2.89), as it not only requires a discretisation of the continuum of radiation modes, but also a truncation of that discretized part of the spectrum. In that way the practical value of (2.89) is very limited. The main feature of perturbation theory modelling along the lines of this section lies in its straightforward calculation of the shift of the propagation constant. However, the modal expansion method set out in sections 2.2.3.2 and 2.2.3.3 is rigorous. Therefore, rather than calculating the new (often hybrid) eigenmodes (through perturbation theory or diagonalization methods), the derived evolution equations are very often used for the modeling of anisotropic waveguide devices in terms of mode coupling between the modes of the unperturbed waveguide. This will be illustrated in the next section.

2.2.3.5 Non-reciprocal evolution equations — a classification of magneto-optical waveguide configurations and effects

In order to illustrate the results obtained here, the above developed theory will be specialized to the class of anisotropic waveguides under study in this work, namely non-reciprocal magneto-optic waveguides.

It was pointed out in Chapter 1 that magneto-optical materials are characterized by a dielectric permittivity tensor whose form depends on the direction of the static (or quasi-static[45]) magnetization present in this material. For most practical applications it suffices to consider cubic materials with magneto-optic effects of first order in the magnetization, so that according to the Onsager principle [16] ($\epsilon_{ij}(\mathbf{M}) = \epsilon_{ji}(-\mathbf{M})$), the effect of the magnetization on the tensor elements must then be the occurrence of anti-symmetric off-diagonal contributions. Furthermore, because of the considered cubic crystal symmetry, the generating matrices describing this symmetry allow the magneto-optic (MO) effect to be described by a *single* (complex) material parameter. This is the complex magneto-optic Voigt parameter Q, introduced in Chapter 1. Therefore, "turning-on" the magnetization in the MO material leads the following permittivity perturbation

$$\delta\underline{\epsilon}_{MO} = \begin{pmatrix} 0 & jn^2Q(\mathbf{1_M})_z & -jn^2Q(\mathbf{1_M})_y \\ -jn^2Q(\mathbf{1_M})_z & 0 & jn^2Q(\mathbf{1_M})_x \\ jn^2Q(\mathbf{1_M})_y & -jn^2Q(\mathbf{1_M})_x & 0 \end{pmatrix} \tag{2.90}$$

Here n is the refractive index of the MO material, and $\mathbf{1_M}$ is the unit vector along the direction of the magnetization. The dependence of the Voigt parameter on the absolute value of $\mathbf{M}$ has been omitted, because we usually assume the magnetic layer to be saturated. Also, remark that if the magneto-optic material is lossless, then Q is purely real as proven in Chapter 1. That is why the imaginary unit j appears in (2.90). It is needed to guarantee the hermiticity of the perturbation in case of lossless magneto-otic materials. If the Euler angles are introduced, then

[45] Quasi-static in the sense of varying "slowly" enough with respect to the frequency of the electromagnetic phenomena that are considered (in order to avoid non-linear effects). In the case of integrated optics, this means slow on a THz scale!

the permittivity perturbation can be decomposed into its three basic Cartesian components,

$$\delta\underline{\epsilon}_{MO} = \sin\vartheta\cos\varphi\delta\underline{\epsilon}^{x}_{MO} + \sin\vartheta\sin\varphi\delta\underline{\epsilon}^{y}_{MO} + \cos\vartheta\delta\underline{\epsilon}^{z}_{MO} \tag{2.91}$$

The tensor forms for $\delta\underline{\epsilon}^{x}_{MO}, \delta\underline{\epsilon}^{y}_{MO}$and $\delta\underline{\epsilon}^{z}_{MO}$, are obviously found by putting the $\mathbf{1_M}$ unit vector along the respective coordinate axes. In standard magneto-optics terminology the waveguide configurations corresponding to these three base configurations are labelled: *polar*, *transversal* (also *equatorial*) and *longitudinal*[46]. While most of the integrated magneto-optical waveguide devices rely on one of these configurations, magnetizations in one of the coordinate planes or even arbitrary magnetization directions can be interesting as well.

For a classification of the magneto-optic waveguide effects for each of these three base configurations, it is assumed that the waveguide has a standard epitaxial layered structure along the x-direction and a symmetrical lateral guiding profile, such as an etched ridge or buried structure. In other words, a mirror symmetry with respect to x-z plane. Furthermore, to simplify things further it is assumed that the unperturbed waveguide has an isotropic permittivity profile. Naturally, the magneto-optical permittivity perturbation is assumed to be longitudinally homogeneous. Finally, it is also assumed that the magneto-optic perturbation region is symmetrical with respect to the x-z plane. In other words that the magnetization of the waveguide is an even or odd function of y. This may seem like a loss of generality. Even though that may be so, this assumption helps in classifying the different magneto-optical waveguide effects at play. These are completely described by the coupling matrix $\overline{\overline{A}}$ of (2.84). According to the treatise of section 2.2.3.3 the diagonal elements of this matrix denote the perturbational phase shifts (in other words the perturbational change in propagation constants) of the various guided unperturbed waveguide modes, while the off-diagonal elements, according to (2.89) represent the perturbational mode profile changes, which can possibly be also polarization changes, if modes of different polarization couple, through non-zero off-diagonal $\overline{\overline{A}}$ elements. The relative strength of these two phenomena (phase shifts and mode conversion) with respect to each other, depend on the precise form of $\overline{\overline{A}}$. For the waveguide under consideration, (2.84) can be worked out in more detail. Because the unperturbed waveguide is assumed to be isotropic, it is self-adjoint, meaning it is bidirectional, and moreover, as was proven in section 2.1.2, it is mirroring. This implies that the adjoint modes (needed in (2.84)) are the mirror images of their corresponding normal modes, thus in (2.84)

$$\tilde{\mathbf{E}}^{(0)}_{t,i}(\boldsymbol{\rho}) = \mathbf{E}^{(0)}_{t,i}(\boldsymbol{\rho}) \tag{2.92}$$

$$\tilde{E}^{(0)}_{z,i}(\boldsymbol{\rho}) = -E^{(0)}_{z,i}(\boldsymbol{\rho}) \tag{2.93}$$

If the second term between brackets in the expansion of E_z (see (2.76)) is neglected (because it will only lead to second order terms in the elements of $\overline{\overline{A}}$), and taking

[46] Recall that the z-coordinate axis is along the propagation direction of the waveguide chosen and the x-coordinate axis is perpendicular to the epitaxial layers.

into account that the MO permittivity perturbation has no diagonal elements ($\epsilon_{33}(\boldsymbol{\rho}) = \epsilon_{33}^{(0)}(\boldsymbol{\rho})$), then the magneto-optic coupling coefficients read

$$A_{ik}^{MO} = \frac{\omega\epsilon_0}{4N_i} \iint \Big[(\mathbf{E}_{t,i}^{(0)}(\boldsymbol{\rho}) - \mathbf{u}_z E_{z,i}^{(0)}(\boldsymbol{\rho})) \cdot \delta\underline{\epsilon}_{MO}(\boldsymbol{\rho}) \cdot (\mathbf{E}_{t,k}^{(0)}(\boldsymbol{\rho}) + \mathbf{u}_z E_{z,k}^{(0)}(\boldsymbol{\rho})) \Big] \mathrm{d}S \tag{2.94}$$

or, specialized for each of the three configurations,

$$\boxed{\begin{aligned} A_{ik}^{MO,(x)} &= \frac{j\omega\epsilon_0}{4N_i} \iint \Big[g(\boldsymbol{\rho})(E_{y,i}^{(0)}(\boldsymbol{\rho})E_{z,k}^{(0)}(\boldsymbol{\rho}) + E_{y,k}^{(0)}(\boldsymbol{\rho})E_{z,i}^{(0)}(\boldsymbol{\rho})) \Big] \mathrm{d}S \\ A_{ik}^{MO,(y)} &= -\frac{j\omega\epsilon_0}{4N_i} \iint \Big[g(\boldsymbol{\rho})(E_{x,i}^{(0)}(\boldsymbol{\rho})E_{z,k}^{(0)}(\boldsymbol{\rho}) + E_{x,k}^{(0)}(\boldsymbol{\rho})E_{z,i}^{(0)}(\boldsymbol{\rho})) \Big] \mathrm{d}S \\ A_{ik}^{MO,(z)} &= \frac{j\omega\epsilon_0}{4N_i} \iint \Big[g(\boldsymbol{\rho})(E_{x,i}^{(0)}(\boldsymbol{\rho})E_{y,k}^{(0)}(\boldsymbol{\rho}) - E_{x,k}^{(0)}(\boldsymbol{\rho})E_{y,i}^{(0)}(\boldsymbol{\rho})) \Big] \mathrm{d}S \end{aligned}} \tag{2.95}$$

Here jg is introduced as short hand notation for jn^2Q. It is also assumed that the unperturbed modes are properly normalized, along (2.12). If that is not the case, then all of the above integrals need to be divided by $\frac{1}{2N_i} \iint \left[\mathbf{E}_{t,i}(\boldsymbol{\rho}) \times \mathbf{H}_{t,i}(\boldsymbol{\rho}) \right] \cdot \mathbf{u}_z \mathrm{d}S$.

These equations show that in general the unperturbed modes will be subjected both to phase shifts and mode conversion. Only in the longitudinal configuration, no phase shift is possible. If the functions in the integrand (the magneto-optic profile and the unperturbed field profiles) and the integration domain itself show spatial symmetries, then due to symmetrical invariance of the integral, a further classification of the mode coupling coefficients becomes possible. Most frequently, integrated waveguide devices are mirror-symmetrical with respect to the x-z plane, so that the unperturbed waveguide modes also exhibit a definite parity under y-coordinate inversion. The unperturbed waveguide modes will then as well have a definite even or odd parity.

In general, it must be assumed that the electric and magnetic field profiles of the unperturbed waveguide modes are hybrid. I.e. all six field components are different from zero. The modes do not show a pronounced transverse electric or transverse magnetic polarization. This behaviour is all the more pronounced, as the lateral refractive index difference becomes larger[47], or as the waveguide permittivity profile gets more square. In these cases, the waveguide modes of the mirror symmetric waveguide, can only be classified on the basis of their y-symmetry. It is easy to show that the Maxwell equations for symmetric, isotropic waveguides demand equal parity for E_y, H_x and H_z, and reversed parity for E_x, E_z and H_y. The common convention [25] is to name a hybrid mode with even E_y, H_x and H_z *symmetric* (s), and a hybrid mode with these components odd *antisymmetric* (a). If the hybrid character of the modes decreases (for instance because the lateral index contrast decreases), then the modes will have an increasingly well-defined polarization character, and a particular transverse field component will become distinctly dominant. On top of their symmetric label, the waveguide modes will

[47] At very high lateral index differences, such as in deeply etched waveguide, the hybrid character is again reduced (unless of course the ridge width is comparable to the guide layer thickness), because then the field strengths are negligible near the etched interface.

then be named according to their dominant transverse field component. This is summarized in Table 2.2. The numbering of the well-polarized modes is based on

	E_x	E_y	E_z	H_x	H_y	H_z	
(s)	−	+	−	+	−	+	TE_0,TM_1,TE_2,TM_3,...
(a)	+	−	+	−	+	−	TM_0,TE_1,TM_2,TE_3,...

TABLE. 2.2: Naming conventions (s) versus (a) (and extra label TE versus TM for non-hybrid modes) based on the parity (+/-) of the different field components under the reflection operation $y \rightarrow -y$ for waveguides with mirror symmetry with respect to the x-z plane.

the symmetry of their dominant transverse component, E_y for TE modes and H_y for TM modes. Therefore, because of the convention for the label "symmetric" and "anti-symmetric" used here, only TE modes with an even index will indeed be symmetric. TM modes with an odd index are symmetric. For example, the first order TM mode is labelled symmetric, even though its H_y profile is odd, because its field components belonging to E_y, H_x and H_z have positive parity!

Using these mode symmetry considerations, together with the assumed symmetry of the magneto-optic profile, it is not so difficult to see that for each of the three basic configurations of the magneto-optic profile ($\delta\underline{\epsilon}^x_{MO}, \delta\underline{\epsilon}^y_{MO}$and $\delta\underline{\epsilon}^z_{MO}$) half of the mode conversion coefficients disappear, and, depending on the symmetry of the magneto-optic profile $g(x,-y) = s_{MO}g(x,y)$ the phase shifts can vanish. All of this is summarized in the Table 2.3.

Table 2.3 can be specialized for the most common case of monomodal symmetric waveguides. Because the waveguide is laterally symmetric, there will of course always be two guided modes! There will, therefore, be a fundamental symmetric hybrid mode and a fundamental antisymmetric hybrid mode. Or, if the lateral index contrasts are not too large, a zero order TE mode and a zero order TM mode. According to Table 2.2 the symmetric mode will have a dominant TE character, while the antisymmetric mode will be TM dominant. In other words, the first two columns of Table 2.3 translate to the Table 2.4.

For well polarized modes, or semivectorial basic fields, the terms in smallprint are negligible. Merely one entry remains in each row. Thus each orientation of the magnetization has a single dominant effect. A symmetric MO profile in transversal configuration causes TM mode phase shifts. The polar configuration yields phase shifts for TE modes, provided that the MO profile is antisymmetric. And a magnetization adjusted along the longitudinal direction drives the TE/TM polarization conversion. Note however, that with a proper symmetry for the MO profile, TE/TM conversion can also occur in the equatorial and the polar configuration. However, these latter two configurations involve z-components of the fields, which are known to be an order of magnitude lower. Therefore the most pronounced magneto-optic effect will be the longitudinal TE/TM conversion, as both the TE and TM phase shift formulae also include z-components. Before illustrating these applications of perturbation theory modelling with a few examples, it is worthwhile to investigate whether the here obtained coupling coefficients

	s_{MO}	PS (s), PS (a)	(s)/(a)–CV
$\delta\underline{\epsilon}^{x}_{MO}$	+	0	$p\iint g(E^{(0)}_{y,s}E^{(0)}_{z,a}+E^{(0)}_{y,a}E^{(0)}_{z,s})\mathrm{d}S$
	–	$2p\iint gE^{(0)}_{y}E^{(0)}_{z}\mathrm{d}S$	0
$\delta\underline{\epsilon}^{y}_{MO}$	+	$-2p\iint gE^{(0)}_{x}E^{(0)}_{z}\mathrm{d}S$	0
	–	0	$-p\iint g(E^{(0)}_{x,s}E^{(0)}_{z,a}+E^{(0)}_{x,a}E^{(0)}_{z,s})\mathrm{d}S$
$\delta\underline{\epsilon}^{z}_{MO}$	+	0	$p\iint g(E^{(0)}_{x,s}E^{(0)}_{y,a}-E^{(0)}_{x,a}E^{(0)}_{y,s})\mathrm{d}S$
	–	0	0
	s_{MO}	$(\mathrm{s})_i/(\mathrm{s})_k$–CV, $(\mathrm{a})_i/(\mathrm{a})_k$–CV	
$\delta\underline{\epsilon}^{x}_{MO}$	+	0	
	–	$p\iint g(E^{(0)}_{y,i}E^{(0)}_{z,k}+E^{(0)}_{y,k}E^{(0)}_{z,i})\mathrm{d}S$	
$\delta\underline{\epsilon}^{y}_{MO}$	+	$-p\iint g(E^{(0)}_{x,i}E^{(0)}_{z,k}+E^{(0)}_{x,k}E^{(0)}_{z,i})\mathrm{d}S$	
	–	0	
$\delta\underline{\epsilon}^{z}_{MO}$	+	0	
	–	$p\iint g(E^{(0)}_{x,i}E^{(0)}_{y,k}-E^{(0)}_{x,k}E^{(0)}_{y,i})\mathrm{d}S$	

TABLE. 2.3: Coupling coefficients (2.95) *exploiting the y-mirror symmetry of the waveguide. PS: phase shift for a hybrid symmetric or antisymmetric mode; (s)/(a)–CV: mode conversion between a symmetric and an antisymmetric hybrid mode;* $(s)_i/(s)_k$*–CV,* $(a)_i/(a)_k$*–CV: mode conversion between two symmetric modes (labelled i and k) or two antisymmetric modes (labelled i and k);* s_{MO}*: parity of the magneto-optic profile; p: short-hand notation for* $\frac{j\omega\epsilon_0}{4N_i}$*.*

predict nonreciprocal behavior. Both integrands in the formulae for the TE and TM phase shift[48] change their sign, when the field profiles for backward modes are substituted. Remember that for an isotropic waveguide, the backward electric field mode profile remains unchanged for the transversal components, and changes sign for the longitudinal component. However, the normalization constant (which is hidden in the factor p) also changes sign when switching between forward and backward modes. As a result the phase shift for forward and backward modes has the same sign! According to (2.89), this means that the perturbation correction on the backward propagation constant is the same as for the forward propagation constant. In other words, the absolute values of both perturbationally corrected propagation constants differ, and the behaviour is indeed non-reciprocal.

$$\begin{aligned}\Gamma^{\mathrm{fw}}_{\mathrm{TE|TM}} &= \beta^{\mathrm{fw}}_{\mathrm{TE|TM}} + A^{\mathrm{fw}}_{\mathrm{TE,TE|TM,TM}}\\ \Gamma^{\mathrm{bw}}_{\mathrm{TE|TM}} &= \beta^{\mathrm{bw}}_{\mathrm{TE|TM}} + A^{\mathrm{bw}}_{\mathrm{TE,TE|TM,TM}}\\ &= -(\beta^{\mathrm{fw}}_{\mathrm{TE|TM}} - A^{\mathrm{fw}}_{\mathrm{TE,TE|TM,TM}})\end{aligned} \tag{2.96}$$

[48] Well-polarised modes are assumed, though the conclusions remain valid if also the column entries in smallprint contribute.

	s_{MO}	TE–PS	TM–PS
$\delta\underline{\epsilon}^x_{MO}$	+	0	0
	–	$2p\iint g E_y^{(0),TE} E_z^{(0),TE}\mathrm{d}S$	${}_{2p\iint g E_y^{(0),TM}}{}_{E_z^{(0),TM}\mathrm{d}S}$
$\delta\underline{\epsilon}^y_{MO}$	+	${}_{-2p\iint g E_x^{(0),TE}}{}_{E_z^{(0),TE}\mathrm{d}S}$	$-2p\iint g E_x^{(0),TM} E_z^{(0),TM}\mathrm{d}S$
	–	0	0
$\delta\underline{\epsilon}^z_{MO}$	+	0	0
	–	0	0
	s_{MO}	TE/TM–CV	
$\delta\underline{\epsilon}^x_{MO}$	+	$p\iint g(E_y^{(0),TE} E_z^{(0),TM} + {}_{E_y^{(0),TM}}{}_{E_z^{(0),TE}})\mathrm{d}S$	
	–	0	
$\delta\underline{\epsilon}^y_{MO}$	+	0	
	–	$-p\iint g({}_{E_x^{(0),TE}}{}_{E_z^{(0),TM}} + E_x^{(0),TM} E_z^{(0),TE})\mathrm{d}S$	
$\delta\underline{\epsilon}^z_{MO}$	+	$p\iint g({}_{E_x^{(0),TE}}{}_{E_y^{(0),TM}} - E_x^{(0),TM} E_y^{(0),TE})\mathrm{d}S$	
	–	0	

TABLE. 2.4: Coupling coefficients for a monomodal waveguide (analogon to Table 2.3). If the modes are well-polarized (or in a semivectorial approach) the terms in smallprint become negligible. s_{MO}: parity of the magneto-optic profile; p: short-hand notation for $\frac{j\omega\epsilon_0}{4N_{(TE|TM)}}$.

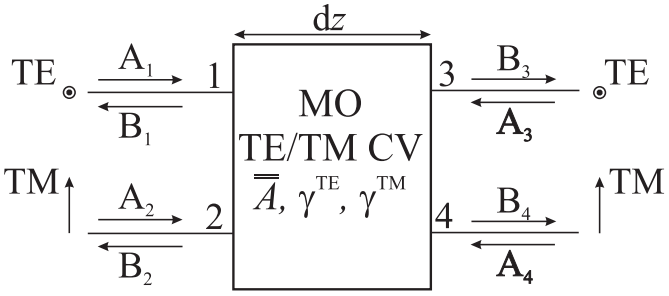


Fig. 2.13: Magneto-optic polarization convertor as a 4-port

For the TE/TM mode conversion, non-reciprocity should be reflected by the asymmetry of the scattering matrix of the 4-port schematically depicted in Fig. 2.13.

To construct the scattering matrix of this 4-port, consider the coupled mode equations (2.86) applied to an infinitesimal part dz of a magneto-optic waveguide with a polarization converting configuration (symmetric polar, antisymmetric equatorial, and symmetric longitudinal magneto-optic profile). It is assumed that only

the fundamental TE and TM mode are interacting.

$$\mathrm{d}X^{\mathrm{fw}}_{\mathrm{TE}} = -j\beta_{\mathrm{TE}}X^{\mathrm{fw}}_{\mathrm{TE}}\mathrm{d}z - j\underbrace{A^{\mathrm{fw}}_{\mathrm{TE,TE}}}_{=0\,\text{table 2.4}}X^{\mathrm{fw}}_{\mathrm{TE}}\mathrm{d}z - jA^{\mathrm{fw}}_{\mathrm{TE,TM}}X^{\mathrm{fw}}_{\mathrm{TM}}\mathrm{d}z \tag{2.97}$$

$$\mathrm{d}X^{\mathrm{fw}}_{\mathrm{TM}} = -j\beta_{\mathrm{TM}}X^{\mathrm{fw}}_{\mathrm{TM}}\mathrm{d}z - j\underbrace{A^{\mathrm{fw}}_{\mathrm{TM,TM}}}_{=0\,\text{table 2.4}}X^{\mathrm{fw}}_{\mathrm{TM}}\mathrm{d}z - jA^{\mathrm{fw}}_{\mathrm{TM,TE}}X^{\mathrm{fw}}_{\mathrm{TE}}\mathrm{d}z \tag{2.98}$$

$$\mathrm{d}X^{\mathrm{bw}}_{\mathrm{TE}} = +j\beta_{\mathrm{TE}}X^{\mathrm{bw}}_{\mathrm{TE}}\mathrm{d}z - j\underbrace{A^{\mathrm{bw}}_{\mathrm{TE,TE}}}_{=0\,\text{table 2.4}}X^{\mathrm{bw}}_{\mathrm{TE}}\mathrm{d}z - jA^{\mathrm{bw}}_{\mathrm{TE,TM}}X^{\mathrm{bw}}_{\mathrm{TM}}\mathrm{d}z \tag{2.99}$$

$$\mathrm{d}X^{\mathrm{bw}}_{\mathrm{TM}} = +j\beta_{\mathrm{TM}}X^{\mathrm{bw}}_{\mathrm{TM}}\mathrm{d}z - j\underbrace{A^{\mathrm{bw}}_{\mathrm{TM,TM}}}_{=0\,\text{table 2.4}}X^{\mathrm{bw}}_{\mathrm{TM}}\mathrm{d}z - jA^{\mathrm{bw}}_{\mathrm{TM,TE}}X^{\mathrm{bw}}_{\mathrm{TE}}\mathrm{d}z \tag{2.100}$$

With the conventional definition of the scattering matrix, as the matrix translating the incoming field amplitudes into the 4-port with the outgoing amplitudes out of the 4-port, one gets the following matrix identities (with the notation of Fig. 2.13) for the infinitesimal magneto-optic waveguide of Fig. 2.13

$$\begin{pmatrix} B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 1-j\beta_{\mathrm{TE}}\mathrm{d}z & -jA^{\mathrm{fw}}_{\mathrm{TE,TM}}\mathrm{d}z \\ -jA^{\mathrm{fw}}_{\mathrm{TM,TE}}\mathrm{d}z & 1-j\beta_{\mathrm{TM}}\mathrm{d}z \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \tag{2.101a}$$

$$\begin{pmatrix} A_3 \\ A_4 \end{pmatrix} = \begin{pmatrix} 1+j\beta_{\mathrm{TE}}\mathrm{d}z & -jA^{\mathrm{bw}}_{\mathrm{TE,TM}}\mathrm{d}z \\ -jA^{\mathrm{bw}}_{\mathrm{TE,TM}}\mathrm{d}z & 1+j\beta_{\mathrm{TM}}\mathrm{d}z \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \tag{2.101b}$$

or (after inversion of the (sub-) scattering matrix in the second equation up to first order in $\mathrm{d}z$)

$$\begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1-j\beta_{\mathrm{TE}}\mathrm{d}z & +jA^{\mathrm{bw}}_{\mathrm{TE,TM}}\mathrm{d}z \\ 0 & 0 & +jA^{\mathrm{bw}}_{\mathrm{TM,TE}}\mathrm{d}z & 1-j\beta_{\mathrm{TM}}\mathrm{d}z \\ 1-j\beta_{\mathrm{TE}}\mathrm{d}z & -jA^{\mathrm{fw}}_{\mathrm{TE,TM}}\mathrm{d}z & 0 & 0 \\ -jA^{\mathrm{fw}}_{\mathrm{TM,TE}}\mathrm{d}z & 1-j\beta_{\mathrm{TM}}\mathrm{d}z & 0 & 0 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix} \tag{2.102}$$

Because it is known from the theory of scattering matrices [14] that the reciprocity of a passive N-port requires the symmetry of the S-matrix, (2.102) shows that the above polarization conversion will only be non-reciprocal if one of the following conditions is fulfilled

$$A^{\mathrm{fw}}_{\mathrm{TE,TM}} \neq -A^{\mathrm{bw}}_{\mathrm{TM,TE}} \tag{2.103a}$$

$$A^{\mathrm{fw}}_{\mathrm{TM,TE}} \neq -A^{\mathrm{bw}}_{\mathrm{TE,TM}} \tag{2.103b}$$

With the formulae of Table 2.4, it is found that these conditions are always fulfilled. For the longitudinal MO configuration the coupling coefficient is antisymmetric with respect to an interchange of the polarized modes, but also antisymmetric with respect to an interchange of the propagation direction (due to a change of sign of the normalization constant N hidden in the factor p). Therefore, one finds that $A^{\mathrm{fw},(z)}_{\mathrm{TM,TE}} = A^{\mathrm{bw},(z)}_{\mathrm{TE,TM}}$, which fulfills (2.103). For the even polar MO configuration and the odd equatorial MO configuration, the coupling coefficient is even for an

interchange of the coupling polarizations, but also even for a change of propagation direction! The latter property is a result of the simultaneous change in sign of the normalization constant and the sign of the z-component of the field profile of a backward propagating mode. A a result, one finds again $A_{\mathrm{TM,TE}}^{\mathrm{fw},(x_+|y_-)} = A_{\mathrm{TE,TM}}^{\mathrm{bw},(x_+|y_-)}$, in agreement with (2.103).

The latter result may strike as a little surprising, as it was pointed out in Chapter 1 that in a transversal configuration, i.e. with the magnetization pointing perpendicular to the propagation direction, there is no non-reciprocal polarization rotation, only an even effect in the magnetization. Moreover, this so-called Cotton-Mouton effect does not even induce a polarization conversion in first-order! Only when second order contributions to the $\underline{\epsilon}$ dyadic are taken into account, there can be a polarization conversion. But this conversion is by definition reciprocal. This does not imply that the above results should be suspiciously looked upon. There remains a difference between the phenomenology of bulk magneto-optics, as described in Chapter 1 and that of integrated magneto-optics. For one thing, in integrated magneto-optics there is the simultaneous occurrence of MO transmission effects (Faraday and Cotton-Mouton) during propagation through the MO epitaxial layer(s), ánd MO Kerr reflection effects at the interface of the MO layers in the epitaxial layer structure. As a result, reciprocal transmission effects might be accompanied by non-reciprocal reflection effects. This is precisely what leads to the above conclusions: a non-reciprocal polarization conversion which cannot be explained by considering the transmission effects alone. Another example of this behavior is the nonreciprocal phase shift undergone by TM polarised eigenmodes in an equatorial MO configuration. Once again, this cannot be explained by a transmission effect, but is solely the result of the MO Kerr reflection at the interfaces of the MO layer(s). This interpretation will become strikingly clear in the next section, where focus will be on the modelling of these type of devices.
Finally, it must be indicated that the best consistency check of the formulae of Table 2.4 is to examine whether these formulae reduce to their bulk analogues when the limit is considered of very large MO (guiding) layers and waveguide structures with very low index contrasts. In that case the unperturbed eigenmodes tend to plane wave solutions and the phenomena of bulk magneto-optics should be recovered. For the formulae of the polar and the equatorial configurations, this is immediately clear. All field z-components reduce to zero, and in agreement with bulk magneto-optics, no first order effects are observed. For the longitudinal configuration, the fields in the integrals can with good approximation be replaced by plane waves (in the considered limit of very large MO guiding regions and very weak guidance). Then there is no dependence on the transversal coordinate $\boldsymbol{\rho}$, and the fields in the integral can be considered as constant amplitudes. Taking into account that the modes are normalised according to $\frac{1}{2}\iint(\mathbf{E}\times\mathbf{H})\cdot\mathbf{u}_z\mathrm{d}S$, one finds, using the Maxwell rotor equations that the TE E_y amplitude and the TM H_y amplitude are resp. $\sqrt{\frac{2}{\epsilon_0 ncS_{MO}}}$ and $\sqrt{\frac{2n}{c\mu_0 S_{MO}}}$. Thus the longitudinal coupling coefficient becomes in this case (using again Maxwell's rotor equation to write

E_x^{TM} as a function in terms of H_y^{TM}),

$$\begin{aligned}\kappa = A_{\mathrm{TE,TM}}^{\mathrm{MO},(z)} &= -\frac{j\omega\epsilon_0 g}{4}\iint E_x^{(0),\mathrm{TM}} E_y^{(0),\mathrm{TE}}\mathrm{d}S \\ &= -\frac{j\omega\epsilon_0 g}{4}\frac{c\mu_0}{n}\sqrt{\frac{2n}{c\mu_0 S_{MO}}}\sqrt{\frac{2}{\epsilon_0 ncS_{MO}}}S_{MO} = -\frac{j2\pi c\epsilon_0 g}{2\lambda}\frac{c\mu_0}{n} \\ &= -j\frac{\pi g}{n\lambda}\end{aligned} \tag{2.104}$$

The resulting length for TE/TM conversion is $L_c = \frac{\pi}{2|\kappa|}$ or a specific Faraday rotation of $\Theta_F = \frac{\pi}{2L_c} = |\kappa| = \frac{\pi nQ}{\lambda}$ in perfect agreement with the results of Chapter 1 (see (1.12)).

As a final remark it should be pointed out that intuitive reasoning about non-reciprocal effects should be conducted with care. Intuitively, one would expect a device to behave non-reciprocally if the magnetization of the MO layer breaks the z-inversion invariance of the waveguide. However that is not the case. The above described magneto-optic non-reciprocal effects for a symmetric MO polar configuration or an anti-symmetric MO equatorial configuration are convincing counterexamples. This counterintuitive result might surprise the reader, but it is nothing more than another manifestation of the confusion that often arises when discussing non-reciprocity on the same level as bidirectionality. Even though a magneto-optic waveguide with a symmetrical polar MO configuration has a permittivity profile that is invariant under a 180˚rotation about the x-axis, that only guarantees its bidirectionality. But as was pointed out in great detail in section 2.1.2, bidirectionality does not imply reciprocity.

2.2.4 Conclusion

This section has given an extensive overview of all the subtleties involved when modelling anisotropic waveguides. In section 2.2.1 it has been shown how the complete modal spectrum of such a waveguide still follows from the poles and branch-cuts of a scalar Green's function, even though the Maxwell equations are strongly coupled, and one would thus expect difficult tensorial forms for the Green's functions. The construction of the scalar Green's function is however not so straightforward, mainly because the tensorial interface boundary conditions are not easily brought in connection with the single scalar Green's function. Because for slab isotropic waveguides the Green's function can be calculated by using the Transfer Matrix Formalism, the logical next step was the extension of this formalism for slab (bi-)anositropic waveguides. This is described in section 2.2.2. The formalism developed there is a logical extension of the calculation of the reflection of plane waves at layered (bi)-anisotropic stacks. It has been shown how the dispersion equation follows from a 4-dimensional determinantal equation, and how normalized modal profiles can be calculated. Both for the construction of the dispersion equation and the modal profiles attention has been paid to numerical stability. The successful implementation of this algorithm has been illustrated by a modelling example.

The other main modelling route is based on considering the anisotropic waveguide as a small perturbation departing from an unperturbed waveguide problem whose entire modal spectrum is known. The unperturbed waveguide can itself be anisotropic. The basic idea is then to try and find a solution in terms of a series expansion of the unperturbed modal spectrum. This perturbation formalism has been elaborated in section 2.2.3. It has been indicated how this approach is equivalent to constructing a set of coupled-mode equations. It was then proven that the perturbed eigenspectrum follows from the diagonalization of the coupling matrix. This formalism has been specialized to the case of symmetrical magneto-optic waveguides, leading to a schematic overview of the non-reciprocal waveguide effects at play for the three basic configurations. This formalism has been implemented. For full 2D perturbation calculations it has been implemented in a MATLAB environment and combined with a commercial waveguide solver [26] for the unperturbed waveguide calculations. For slab perturbation calculations the routines were written in the scripting language PYTHON in order to allow for a seamless transition between the in-house developed waveguide modelling tool CAMFR [27] and the perturbation calculations. An example will be given in section 2.3.3, where both rigorous and perturbation modelling will be compared on a benchmark simulation of an integrated isolator.

It is clear that the choice between the perturbation formalism and the rigorous TMF mainly depends on the degree of precision that is needed, at least when slab waveguides are considered. When the waveguide cross-section is two-dimensional or when the waveguide is not longitudinally invariant, rigorous modelling becomes a tough challenge. The here described perturbation scheme is not useful for the calculation of modal profiles. But then again, if perturbation theory is valid, it can be expected that the corrections to the profiles will be irrelevant for applications like overlap integrals, far-field calculations, ... The main advantage of perturbation theory seems to be its small numerical effort (only a few integrals need to be calculated) and its applicability to waveguides with cross-sections with any form and dimensionality. In any case, as a last remark there is no direct a priori criterion to check whether perturbation theory will be valid. In theory, results always need to be double-checked with a rigorous model if available.

2.3 Transverse (or equatorial) magneto-optic non-reciprocal waveguides

2.3.1 Introduction

In Chapter 1 it has been sketched how and why the design of the most important class of integrated non-reciprocal optical devices, namely optical isolators, has evolved from integrated non-reciprocal polarization convertors towards integrated non-reciprocal phase shifters. The device under study in this work also belongs to this class.

As explained in great detail in section 2.2.3.5, both TE and TM waveguide modes can undergo a non-reciprocal phase shift. A planar waveguide is automat-

ically mirror symmetric. With the results of Table 2.4 it is then clear that it is far from obvious to design a magneto-optic profile that will induce simultaneously an identical non-reciprocal phase shift for both polarizations. There have been a few theoretical proposals [28]. But the experimental feasibility of these designs is far from being guaranteed, as they require magnetic domain structures with 90°Néel-like domain walls, or with 180°domain walls between magnetic domains with magnetizations tilted in the x-y plane. In Chapter 1 another approach to achieve polarization-independent operation for a non-reciprocal phase shifter has been mentioned. The main idea there was to separate the TE and the TM non-reciprocal phase shifts in the commonly used Mach-Zehnder layout, and put the shifts in different arms of the interferometers. In that way, the non-reciprocal phase shift in each arm of the interferometer can be separately optimized for its respective polarization. This idea has the best prospects for experimental success, provided that the TE phase shift can be experimentally realized. As Table 2.4 shows, the strongest effect for TE polarizations can be expected for perfectly antisymmetric magneto-optic profiles. As the degree of non-symmetry decreases, the TE phase shift tends to zero. Therefore, in the optimum case this requires a control of the magnetic Bloch domain wall position with an accuracy better than 1μm. This is not a trivial task. And even though present day expertise allows a good control of the magnetic domain structure of epitaxially grown garnet layers, it is far from a trivial task to achieve this control uniformly over lengths of mm's (see for instance [29] for an illustration of observed possible domain configuration perturbations). These observations prove that it is not all that evident to realise an integrated magneto-optic waveguide with a non-reciprocal TE phase shift. Let alone, one with a polarization independent phase shift. For a TM phase shift, the situation is much easier, because it suffices to have a symmetrical equatorial MO configuration. One is not confronted with the technological difficulties of having to control domain patterns with respect to micronsize ridges. This is the main reason why most experimental reports on non-reciprocal phase shifters have been of the TM-type.

The operation principle of the component under study in this work is entirely based on the TM non-reciprocal phase shift occurring in equatorially magnetized magneto-optic waveguides. That alone provides a strong enough argument to study the modelling of this class of magneto-optical waveguides in more detail. It may seem superfluous to spend extra attention to a specialized but admittedly simple version of the general class of magneto-optic waveguides, discussed in the previous section. However, in equatorial configuration, magneto-optic slab waveguides have the interesting peculiarity that the Maxwell waveguide equations again break up into two independent differential equations, one for TE modes and for TM modes, just as in the case of isotropic slab waveguides. This allows the TMF formalism of section 2.2.2 to be considerably simplified, because now the eigenwaves in the homogeneous layers will retain their TE and TM polarization character and no polarization mixing will occur at the interfaces. As a result, the 4×4 matrix formalism reduces to a traditional 2×2 formalism as known from isotropic waveguides. On top of this numerical simplification of the rigorous modelling technique, the study of equatorial MO waveguides also allows a

straightforward comparison of perturbation and rigorous modelling techniques, because the eigenspectrum of the 'actual' waveguide and the 'unperturbed' isotropic waveguide have the same polarization character. For other configurations of MO slab waveguides, the eigenmodes will have a hybrid polarization. In that case, for a comparison of the guided mode eigenspectrum, rigourously obtained through the extended TMF formalism of section 2.2.2, with the results obtained through perturbation theory it is not sufficient to merely calculate the first-order eigenvalues of the coupling matrix $\overline{\overline{A}}$, since the most pronounced first-order effect will be a polarization coupling. A sound comparison in these cases requires a first-order calculation of the eigenvectors of the matrix $\overline{\overline{G}}+\overline{\overline{A}}$ of (2.86). It has been indicated there that the perturbational formulae for that purpose are not very practical.

In the next sections the simplified version of the TMF for equatorial MO slab waveguides will be elaborated, and some details of the perturbational formula (i.e. the first entry on the third row of Table 2.4) for these nonreciprocal TM phase shifters will be indicated. A simplified layout of the integrated optical isolator, studied in this work, will then serve as a comparison benchmark test to check the validity of the perturbation theory approach [30].

Before proceeding, it is worthwhile to investigate whether the pure TM and TE character of the eigenmodes in an equatorial MO slab waveguide is also retained if realistic waveguides with a full 2-dimensional permittivity profile are considered. First of all, it is very straightforward to prove this for an equatorial MO slab waveguide. In a slab waveguide, the modal profiles have no y-dependence and thus $\partial_y = 0$. As a result, the Maxwell waveguide equation (2.21)

$$-\boldsymbol{\nabla}^2\mathbf{e}(\mathbf{r}) + \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{e}(\mathbf{r})) = k_0^2\underline{\epsilon}(x)\cdot\mathbf{e}(\mathbf{r}), \tag{2.105}$$

splits up into two independent sets of partial differential equations

$$\text{TE}: \begin{cases} \left(\dfrac{\mathrm{d}^2}{\mathrm{d}x^2} + k_0^2 n^2(x) - \beta^2\right) E_y(x) = 0 \\ \beta E_y(x) = \omega\mu_0 H_x(x) \\ \dfrac{\mathrm{d}}{\mathrm{d}x}E_y(x) = -j\omega\mu_0 H_z(x) \end{cases} \tag{2.106}$$

$$\text{TM}: \begin{cases} \big(k_0^2 n^2(x) - \beta^2\big)E_x(x) - j\big(k_0^2 g(x) - \beta\dfrac{\mathrm{d}}{\mathrm{d}x}\big)E_z(x) = 0 \\ \dfrac{\mathrm{d}}{\mathrm{d}x}\big(n^2(x)E_x - jg(x)E_z(x)\big) = j\beta\big(jg(x)E_x(x) + n^2(x)E_z(x)\big) \\ j\beta E_x(x) + \dfrac{\mathrm{d}}{\mathrm{d}x}E_z(x) = j\omega\mu_0 H_y(x) \end{cases} \tag{2.107}$$

This proves the independence of TE and TM solutions in equatorial MO slab waveguides. Furthermore, the differential equation for the dominant field component E_y of TE modes, is the traditional Helmholtz equation, proving that in the equatorial slab case there is no non-reciprocal effect for TE modes. The non-reciprocal behaviour of TM modes does not show up in an obvious manner in (2.107). This is easier to detect when the Maxwell slab equations for TM

modes are derived out of the second rotor equation in terms of the dominant TM field component, H_y.

$$\boldsymbol{\nabla} \times \underline{\epsilon}^{-1}(x) \cdot (\boldsymbol{\nabla} \times \mathbf{H}(x)) = k_0^2 \mathbf{H}(x) \tag{2.108}$$

For an equatorial MO profile, the inverse of the permittivity tensor is easily obtained as[49]

$$\underline{\epsilon}^{-1} = \underline{\eta} \begin{pmatrix} \epsilon & 0 & -jg \\ 0 & \epsilon & 0 \\ +jg & 0 & \epsilon \end{pmatrix}^{-1} = \begin{pmatrix} \eta & 0 & j\zeta \\ 0 & \eta & 0 \\ -j\zeta & 0 & \eta \end{pmatrix}, \tag{2.109}$$

with, up to first order in g, $\eta = 1/\epsilon$ and $\zeta = g/\epsilon^2$. The y-component of (2.108) becomes

$$\begin{aligned} k_0^2 H_y &= -j\beta\Big(\eta(\boldsymbol{\nabla} \times \mathbf{H})_x + j\zeta(\boldsymbol{\nabla} \times \mathbf{H})_z\Big) \\ &\quad - \frac{\mathrm{d}}{\mathrm{d}x}\Big(-j\eta(\boldsymbol{\nabla} \times \mathbf{H})_x + \eta(\boldsymbol{\nabla} \times \mathbf{H})_z\Big) \\ &= \beta^2 \eta H_y + \beta\zeta \frac{\mathrm{d}}{\mathrm{d}x} H_y - \beta \frac{\mathrm{d}}{\mathrm{d}x}\big(\zeta H_y\big) - \frac{\mathrm{d}}{\mathrm{d}x}\big(\eta(\frac{\mathrm{d}}{\mathrm{d}x} H_y)\big) \\ &= \Big(-\eta \frac{\mathrm{d}^2}{\mathrm{d}x^2} + \beta^2 \eta - (\frac{\mathrm{d}}{\mathrm{d}x}\eta)\frac{\mathrm{d}}{\mathrm{d}x} - \beta\big(\frac{\mathrm{d}}{\mathrm{d}x}\zeta\big)\Big) H_y \end{aligned} \tag{2.110}$$

$$\Big(\frac{\mathrm{d}^2}{\mathrm{d}x^2} + k_0^2\epsilon - \beta^2 + \epsilon(\frac{\mathrm{d}}{\mathrm{d}x}\eta)\frac{\mathrm{d}}{\mathrm{d}x} + \beta\epsilon(\frac{\mathrm{d}}{\mathrm{d}x}\zeta)\Big) H_y = 0 \tag{2.111}$$

Again the Helmholtz operator is recovered, and now also, the non-reciprocity of the TM waveguide equation is apparent. The last term, which contains the magneto-optic contribution, has an odd behavior with respect to the propagation constant β. In other words, the z-inversion symmetry of the waveguide equation is broken by the MO contribution. A modal TM solution with eigenvalue β_i will not have an associated bidirectional solution with eigenvalue $-\beta_i$.

If the waveguide permittivity (and magneto-optic) profile becomes 2-dimensional, it is known from optical waveguide theory that in general there will be no two independent sets of eigenmodes with each three field components. The two polarization types will transform into two hybrid elliptically polarized eigenmodes, with all 6 field components different from zero. However, for most commonly encountered integrated optical devices, it turns out that usually each hybrid polarization exhibits one set of large and perpendicular electromagnetic transverse field components and one set of small and perpendicular electromagnetic field components. The labels TE and TM are then transposed to these hybrid modes, depending on whether E_x and H_y, or E_y and H_x are the small field components. Based on these observations it has become common practice to use a semivectorial approximation in waveguide modelling. In this approximation, (quasi-)TE modes are found by imposing the ansatz $E_x = 0$, while for (quasi-)TM modes the ansatz $H_x = 0$ is used. All of this is well known for isotropic waveguides, and has been extensively tested on standard benchmark problems, proving that semivectorial mode solvers

[49] As usual optical magnetic anisotropy is neglected. Hence all diagonal elements in $\underline{\epsilon}$ (and, a fortiori, in $\underline{\eta}$) are equal.

achieve an equal level of accuracy as fully vectorial mode solvers [31]. However, when the isotropic waveguide (with a traditional assumed y-mirror symmetry) also obtains a symmetric equatorial magneto-optic profile, the formulae in the third row of Table 2.3, indicate that the magneto-optic mode coupling phenomena may well undermine the semi-polarized state of the eigenmodes. The question is thus whether the semivectorial approximation still holds for symmetric equatorial MO waveguides when these have a realistic 2-dimensional geometric profile. In order to check this, quasi-TM solutions will be enforced and it will be checked *a posteriori* whether $H_x = 0$ was justified.
The Maxwell equation to be solved is

$$\boldsymbol{\nabla} \times \underline{\eta}(\boldsymbol{\rho}) \cdot (\boldsymbol{\nabla} \times \mathbf{H}(\boldsymbol{\rho})) = k_0^2 \mathbf{H}(\boldsymbol{\rho}) \tag{2.112}$$

A quasi-TM solution is enforced by demanding $H_x = 0$ and $H_z = -j\partial_y H_y/\beta$. The latter equation is needed to guarantee a divergence-free magnetic induction. Inserting this ansatz in (2.112) results in a vector equation, for which the z-component is already taken care of by having imposed the divergence-free condition in the previous line. The y-component becomes

$$\begin{aligned} k_0^2 H_y &= \partial_z \left(\eta(\boldsymbol{\nabla} \times \mathbf{H})_x + j\zeta(\boldsymbol{\nabla} \times \mathbf{H})_z\right) \\ &\quad - \partial_x \left(-j\zeta(\boldsymbol{\nabla} \times \mathbf{H})_x + \eta(\boldsymbol{\nabla} \times \mathbf{H})_z\right) \\ &= -j\beta\eta\partial_y H_z + \eta\beta^2 H_y + \cancel{\zeta\beta\partial_x H_y} \\ &\quad + j\partial_x(\zeta\partial_y H_z) - \cancel{\zeta\beta\partial_x H_y} - \beta\partial_x(\zeta)H_y - \partial_x(\eta\partial_x H_y) \\ &= \left(-\partial_x\eta\partial_x - \eta\partial_y^2 + \eta\beta^2 - \beta\partial_x(\zeta) + \partial_x\zeta\partial_y^2/\beta\right) H_y \end{aligned} \tag{2.113}$$

All partial derivatives need to be applied to all terms to their right, except when explicit brackets limit their scope. The fourth term between brackets is the magneto-optic perturbation of this quasi-TM equation. The last term is an additional higher order MO contribution and can be neglected with regard to the previous MO contribution as it is two times suppressed, firstly by a factor $1/\beta^2$ and secondly by the second order partial y-derivative (which amounts to a factor $1/W^2$, where W is a typical lateral extension of the waveguide). Thus, for typical waveguides and typical material systems, this latter term is less than 1% of the already small previous magneto-optic contribution. Rewriting the last line, one obtains the following quasi-TM waveguide equation

$$\Big(\partial_x^2 + \partial_y^2 + k_0^2\epsilon(\boldsymbol{\rho}) - \beta^2 + \epsilon(\boldsymbol{\rho})\partial_x\big(\eta(\boldsymbol{\rho})\big)\partial_x + \beta\epsilon(\boldsymbol{\rho})\partial_x\big(\zeta(\boldsymbol{\rho})\big)\Big) H_y(\boldsymbol{\rho}) = 0 \tag{2.114}$$

In the first four (five) terms the traditional TM Helmholtz operator is again obtained. And the magneto-optic contribution in the last term is again accompanied by an odd factor in the propagation constant β, proving the non-bidirectionality of the quasi-TM modes. In order to check whether the semivectorial ansatz is justified, the x-component of (2.112) is written out under these assumptions ($H_x = 0$

and $H_z = -j\partial_y H_y/\beta$)

$$\begin{aligned} k_0^2 H_x &= \partial_y \left(\eta(\boldsymbol{\nabla}\times\mathbf{H})_z - j\zeta(\boldsymbol{\nabla}\times\mathbf{H})_x\right) \\ &\quad - \partial_z \left(\eta(\boldsymbol{\nabla}\times\mathbf{H})_y\right) \\ &= \partial_y(\eta\partial_x H_y) - j\partial_y(\zeta\partial_y H_z) + \beta\partial_y(\zeta H_y) - j\beta\eta\partial_x H_z \\ &= \cancel{\eta\partial_y\partial_x H_y} + (\partial_y\eta)\partial_x H_y - \cancel{1/\beta\partial_y(\zeta\partial_y^2 H_y)}^{\approx 0} + \beta\partial_y(\zeta H_y) - \cancel{\eta\partial_x\partial_y H_y} \end{aligned} \tag{2.115}$$

The third term in the last equation can be neglected with respect to the fourth for reasons described above. So what is left in the x-component of (2.112), is the small magneto-optic contribution $\beta\partial_y(\zeta H_y)$ and $(\partial_y\eta)\partial_x H_y$. For standard rib waveguides only the sidewalls of the rib contribute to $(\partial_y\eta)\partial_x H_y$. The large index jumps at the etched sidewalls cause the field to be small there, and hence to exhibit no large transversal variations (i.e. in the x-direction). It can thus be concluded that the semi-vectorial approximation is in most cases self-consistent.
Starting from $-\underline{\eta}\cdot\boldsymbol{\nabla}^2\mathbf{E} + \underline{\eta}\cdot\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{E}) = k_0^2\mathbf{E}$, a similar reasoning can be applied to prove that the semivectorial approach is also self-consistent for TE modes with the ansatz ($E_x = 0$). The derivation will not be given, only the resulting equation for the dominant component E_y.

$$\partial_x^2 E_y + \partial_y\big(\eta\partial_y(\epsilon E_y)\big) + k_0^2\epsilon E_y - \beta^2 E_y - \cancel{j\partial_y\big(\eta\partial_x(gE_z)\big)}^{\approx 0} = 0 \tag{2.116}$$

There is a very small magneto-optic contribution in the quasi-TE mode equation. But, what is most important, is that within the validity of the semivectorial approximation, (2.116) is even in β! The waveguide is in other words bidirectional for TE modes, but not mirroring (as $E_z \to -E_z$, $E_y \to E_y$ is obviously not a symmetry operation of (2.116)). Again it can be shown that the validity range of the semivectorial approximation is determined by the smallness of the fields components at the sidewalls of the waveguide.

In conclusion, it is proven that for most commonly encountered integrated optical device layouts, the eigenspectrum of symmetric equatorial MO waveguides retains its (quasi-)TE and (quasi-)TM character, and only the (quasi-)TM modes are subjected to non-reciprocal behaviour. The (quasi-)TE eigenspectrum remains bidirectional.

2.3.2 Modelling symmetric equatorial magneto-optic waveguides

2.3.2.1 Non-reciprocal TM phase shift – first-order perturbation formulae

The first order perturbation non-reciprocal phase shift in symmetric equatorial waveguides, is given by the first formula in the third row of Table 2.4. If the modes are well-polarized or if the semivectorial approximation holds, then Table 2.4 and the discussion of the previous section shows that only the TM modes are subjected

sgn	$\Re[\Delta\beta]$	$\Im[\Delta\beta]$
$+$	$\varphi \nearrow$	$\alpha \searrow$
$-$	$\varphi \searrow$	$\alpha \nearrow$

TABLE. 2.5: Influence of the MO phase shift on the modal behaviour a forward TM mode as a function of the signs of the complex parts of $\Delta\beta$. The modal behaviour in the backward direction is of course found by reversing the tabulated trends.

to a phase shift, and no other magneto-optic effects will take place. For a guided TM mode the first order perturbation shift of the propagation constant becomes,

$$\Delta\beta = -2p \iint g E_x^{(0),TM} E_z^{(0),TM} \mathrm{d}S \tag{2.117}$$

As explained in section 2.2.3.5 the shift is the same for the forward and the backward modes. So that according to (2.96), the total nonreciprocal difference[50] in propagation constant becomes $2\Delta\beta$.
In the most general case $\Delta\beta$ will be complex. So it is worthwhile to pay some attention to the meaning of the signs and the complex parts of $\Delta\beta$. Therefore we need to investigate $\Gamma_{\mathrm{fw}} = \Re[\beta^0] + \Re[\Delta\beta] - j(\kappa^0 - \Im[\Delta\beta])$ and $\Gamma_{\mathrm{bw}} = -\Re[\beta^0] + j\kappa^0 + \Re[\Delta\beta] + j\Im[\Delta\beta] = -\big(\Re[\beta^0] - \Re[\Delta\beta] - j(\kappa^0 + \Im[\Delta\beta])\big)$. Due to the natural sign convention $\exp(j\omega t)$, chosen in this work, a positive value for $\Im[\Delta\beta]$ indicates a decrease of the optical TM modal loss of the forward propagating mode, and consequently an increase of the TM modal loss of the backward propagating mode by the same amount. A positive value for $\Re[\Delta\beta]$ indicates an increase of the phase shift undergone by a forward propagating TM mode when propagating over a certain length L, as compared to the phase shift of this same mode when propagating over the same length in the same waveguide with the magneto-optic perturbation "switched off". In other words, when one arm of a Mach-Zehnder is subjected to an equatorial MO effect, and the other (identical) arm isn't, the second arm has to be made physically longer by a factor $1 + \frac{\Re[\Delta\beta]}{\beta^0}$ in order to have both arms in balance. In the backward direction, the magneto-optic arm will then be lagging behind in phase with respect to the non-MO arm by an amount $2\Re[\Delta\beta]L_{\mathrm{MO}}$. So if in the backward direction one wants the arms to be in 100% destructive imbalance, the length of the MO section should be $L_{\mathrm{MO}} = \frac{\pi}{2\Re[\Delta\beta]}$. All of this is summarized in Table 2.5.

A second aspect to be investigated about the perturbation formula of the nonreciprocal TM phase shift (2.117), is the influence of the different complex parts of the magneto-optical parameter g on $\Delta\beta$. This is not so easy as in the integrand both g and the electric fields can in general be complex. Therefore, to make things

[50] This is commonly called the 'non-reciprocal TM phase shift', but obviously this is a somewhat unlucky name, as it only becomes a phase shift when it is multiplied by a length. However throughout this text we are going to stick with the traditional name, as the context will make clear whether a real phase shift or a propagation constant shift is meant. Note that the latter could be called a specific phase shift.

clear, it is first assumed that the magneto-optical waveguide is lossless. Then it will be investigated how the behaviour changes when small losses are introduced.

When the MO waveguide is lossless, the permittivity tensor must be hermitic throughout the waveguide. This implies that its diagonal elements (the refractive index) must be purely real everywhere, and its off-diagonal elements must be purely imaginary and antisymmetric. In other words, the magneto-optic parameter g must be purely real. The unperturbed waveguide (obtained by putting $g = 0$ everywhere) will then obviously also be lossless. As a result, it is a well known fact from optical waveguide theory, that the longitudinal electric field will be in perfect quadrature with the transverse (i.e. in the cross-sectional plane) electric field components. In general, the amplitude of the electric fields is free up to a complex phase factor, but the choices are limited by the normalization condition, $\frac{1}{2N}\iint \left[\mathbf{E}_t(\boldsymbol{\rho}) \times \mathbf{H}_t(\boldsymbol{\rho})\right] \cdot \mathbf{u}_z \mathrm{d}S = 1$. Therefore, one is obliged to choose purely real transverse electric field components and a purely imaginary longitudinal electric field component[51]. All of this leads to the conclusion that the integrand is purely imaginary. Because E_x and g are purely real and E_z is purely imaginary. But the factor $p = \frac{j\omega\epsilon_0}{4N}$ in (2.117) is also purely imaginary. Thus, in conclusion, for a lossless MO waveguide $\Delta\beta$ is purely real. A TM mode in lossless symmetric equatorial MO waveguide will only undergo a real non-reciprocal phase shift.

When small losses are introduced in the waveguide, the conclusions will change. Now, the unperturbed waveguide will be lossy, and as a result the transverse and longitudinal fields will not be in perfect quadrature anymore. If again the transversal field components are chosen to be real, then this means that E_z will also acquire a small real part. As a result the product E_xE_z in the integrand of (2.117) will have a small real part, and the total formula (taking the imaginary unit in p into account) will have a small imaginary part. In other words, the non-reciprocal phase shift will in this case also establish itself through a non-reciprocal modal absorption change, as opposed to the lossless case where there will only be a non-reciprocal phase change. The interesting question here is which part of the magneto-optical parameter g is mainly responsible for this non-reciprocal absorption shift. This is answered by looking at the complex behaviour of the integrand gE_xE_z, more in particular at the real part. The field product E_xE_z will have a small real part, as discussed above. Therefore, there are two possible scenarios.

MO layer is lossless When the magneto-optic itself is lossless (within a lossy waveguide), g will be purely real, and will therefore have the same influence both on the real and the imaginary part of the integrand. In this case, the relative strength of $\Delta\alpha$(the non-reciprocal absorption shift) versus $\Delta\varphi$ (the non-reciprocal phase change), is only caused by the complex behaviour of the field product E_xE_z in the MO region. As explained, due to the losses in the waveguide, this field product will have a small real part. Or, in other

[51] Actually, this is not entirely true, because when an other phase factor is chosen for the transverse field components, the normalization integral will differ from 1 by exactly the same phase factor. But because the normalization condition is not exactly fulfilled, the perturbation formula needs to be divided by the normalization integral, and as a result the extra phase factor will cancel.

words, there the phase difference between E_z and E_x will slightly deviate from $\pi/2$. However, practical simulations prove that this deviation does not extend far from the waveguide region where the loss is concentrated, and that on top of that, this deviation is never "drastic", even for very high material losses. Thus, even though it is the presence of the (admittedly purely real) MO parameter g in a region of the waveguide that causes the non-reciprocal effect in general, in this scenario it could be stated that the non-reciprocal *absorption shift* is caused by the "closeness" of the *lossless* MO layer to the absorbing isotropic layer. The question of what is lying at the origin of the non-reciprocal absorption shift, the (*real*) MO parameter or the lossy behaviour of the unperturbed eigenmodes, is thus actually more a semantic discussion. For the non-reciprocal phase change ($\Delta\varphi$), the situation is clearer. Because $\Im[E_x E_z]$ is always at least an order of magnitude larger than its real counterpart, and extends throughout the waveguide cross-section, it is undeniably the (real) MO parameter that causes the non-reciprocal phase change $\Delta\varphi$.

MO layer is lossy When the magneto-optic layer itself is lossy, one could argue that there are again two subscenarios: one where the MO layer itself is also the only source of losses in the waveguide, and a second scenario, where other parts of the waveguide contribute to the lossy behaviour. However, this would only have an impact on this discussion, if "turning off" the magneto-optic parameter would also turn "off" its loss mechanism. This is very unlikely, if one thinks of the very complex material model this would require, because in any case the material always has to fulfill the fundamental Kramers-Krönig relationship [32]. Therefore it is safe to assume that a lossy MO material will always have both a complex refractive index and a complex g, ánd that the refractive index will remain complex even without a magnetization present [52]. Therefore, we will only consider this as one scenario. Because the unperturbed waveguide is lossy and the MO layer is lossy, both g and the field product $E_x E_z$ will be complex numbers. The real part of the integral (causing the non-reciprocal absorption shift) will thus be given by $\Re[g]\Re[E_x E_z] - \Im[g]\Im[E_x E_z]$. As stated above, the real part of the field integral will for commonly encountered losses always be at least an order of magnitude lower than its imaginary part. For most commonly encountered MO materials the real and imaginary parts of g have the same order of magnitude. Thus, one can conclude that in this scenario, the non-reciprocal absorption shift is mainly caused by the imaginary part of g. The non-reciprocal phase change ($\Delta\varphi$) is caused by the imaginary part of the integral, and thus mainly by the real part of g, as can be concluded by looking at $\Re[g]\Im[E_x E_z] + \Im[g]\Re[E_x E_z]$.

It is hard to predict how the signs of the complex parts of the MO parameter will influence the non-reciprocal phase shift, because this not only depends on the signs of the complex parts of g but also on the signs of the in general complex field

[52] Of course on a microscopic scale there will always be a random magnetization present unless the material is diamagnetic.

product. As explained in Chapter 1 the situation becomes even more complex when considering the fact that both complex parts of g can exhibit all 4 possible sign combinations (all having realistic physical meanings), as opposed to the refractive index where for instance a negative real part and a positive imaginary part is a spurious mathematical construction.
However, we can predict some influences under certain assumptions. These are summarized in Table 2.6, and are based on the previous conclusions given in Table 2.5, and the perturbation formula (2.117). For simplicity, the real part of the field product is neglected with respect to its imaginary part. This is for most common cases justified. Finally, it is also assumed that the MO perturbation is constant throughout the MO layer, so as to allow g to be brought outside the integral in (2.117). Table 2.6 shows that it is not possible to achieve complete

$\Delta\varphi$ $(\operatorname{sgn}[\Delta\varphi] = \operatorname{sgn}[\Re[(2.117)]])$		
	$\Im[\iint E_x E_z \mathrm{d}S] > 0$	$\Im[\iint E_x E_z \mathrm{d}S] < 0$
$\Re[g] > 0$	+	−
$\Re[g] < 0$	−	+
$\Delta\alpha$ $(\operatorname{sgn}[\Delta\alpha] = -\operatorname{sgn}[\Im[(2.117)]])$		
$\Im[g] > 0$	−	+
$\Im[g] < 0$	+	−

TABLE. 2.6: Influence of the sign of the complex parts of the composing parts of (2.117) *on the non-reciprocal phase shift (from the viewpoint of the forward TM mode).*

control over the occurring non-reciprocal shifts when the material parameter g is completely known. All one can do if the sign of the real/imaginary effect is the opposite of what is wanted, is switch the magnetization (taking however into account that this will also change the sign of the accompanying real/imaginary effect!)[53]. Another possibility is to change the position of the MO perturbation layer in the stack of materials, so as to bring it into a region where the field integral has to correct sign. This latter "trick" is of course under the assumption that the refractive index of the MO layer itself does not drastically influences the unperturbed fields.

As a conclusion, it is worthwhile to summarize the important formulae of this section. The implementation aspects will be discussed in section 2.3.3.

[53]This "trick" will of course only work if the MO material under consideration has no large second-order effects in the magnetization.

The first-order perturbation non-reciprocal TM phase shift is given by[54]

$$\Delta\beta = -j\omega\epsilon_0 \frac{\iint g E_x^{(0)} E_z^{(0)} \mathrm{d}S}{\iint \left[E_x^{(0)} H_y^{(0)} - E_y^{(0)} H_x^{(0)}\right] \mathrm{d}S} \tag{2.118}$$

The important formulae for isolator design that can be derived out of this non-reciprocal phase shift, are the necessary length L_π for a (real) non-reciprocal π phase shift (when designing a non-reciprocal Mach-Zehnder interferometer) and the non-reciprocal absorption difference $\Delta\alpha$ (when designing a non-reciprocal optical amplifier/absorber, such as the component under study in this work).

$$L_\pi = \frac{\pi}{2\left|\Re[\Delta\beta]\right|} = \frac{\lambda Z_{\mathrm{vac}}}{4} \left| 1/\Im \left[\frac{\iint g E_x^{(0)} E_z^{(0)} \mathrm{d}S}{\iint \left(E_x^{(0)} H_y^{(0)} - E_y^{(0)} H_x^{(0)}\right) \mathrm{d}S} \right] \right|, \tag{2.119}$$

and

$$\Delta\alpha = 4\Im[\Delta\beta] = -\frac{8\pi}{Z_{\mathrm{vac}}\lambda} \Re \left[\frac{\iint g E_x^{(0)} E_z^{(0)} \mathrm{d}S}{\iint \left(E_x^{(0)} H_y^{(0)} - E_y^{(0)} H_x^{(0)}\right) \mathrm{d}S} \right] \tag{2.120}$$

This last formula assumes a positive value for the imaginary part of $\Delta\beta$, so that the backward mode suffers higher absorption. If that is not the case, then, as discussed in the previous paragraph, the magnetization needs to be reversed. The isolation ratio (per unit of length) obtained through this non-reciprocal optical absorption is found through the standard formula $\mathrm{IS} = -10\log(P_{\mathrm{out}}^{\mathrm{bw}}/P_{\mathrm{out}}^{\mathrm{fw}})/L$, assuming that unit power is launched at both facets. This gives

$$\begin{aligned} \mathrm{IS} &= -10\log(\exp(-\Delta\alpha L)) = 10/\ln(10)\Delta\alpha \\ &= -\frac{80\pi}{\ln 10 Z_{\mathrm{vac}}\lambda} \Re \left[\frac{\iint g E_x^{(0)} E_z^{(0)} \mathrm{d}S}{\iint \left(E_x^{(0)} H_y^{(0)} - E_y^{(0)} H_x^{(0)}\right) \mathrm{d}S} \right] \\ &\approx -\frac{0.3}{\lambda} \Re \left[\frac{\iint g E_x^{(0)} E_z^{(0)} \mathrm{d}S}{\iint \left(E_x^{(0)} H_y^{(0)} - E_y^{(0)} H_x^{(0)}\right) \mathrm{d}S} \right] \end{aligned} \tag{2.121}$$

2.3.2.2 Non-reciprocal TM 2×2 Transfer Matrix Formalism

As indicated in the previous section, in equatorial magneto-optic waveguides the modal spectrum retains its TE and TM character, and the slab Transfer Matrix Formalism for such waveguides again assumes its standard 2×2 form. In this section, this simplified MO TMF will be derived.

Because there is no mixing of the base eigenpolarizations — TE and TM in this case — at the interfaces in equatorial MO slab waveguides, there is no need for a coordinate transformation of the electromagnetic fields between the Cartesian system and the wavefield coordinate system at the interfaces (as opposed to

[54] For practical reasons the formula is given for non-normalized fields, as it is not always clear whether the used mode solver properly normalizes the unperturbed fields in the proper way. It has been mentioned that this requires a division of the (2.117) by $-\frac{1}{2N}\iint \left[\mathbf{E}^{(0)} \times \mathbf{H}^{(0)}\right] \cdot \mathbf{u}_z \mathrm{d}S$

the situation for completely arbitrary anisotropic slab waveguides and described in 2.2.2.2). The essential characteristic of such a situation is that one can use directly the amplitudes of the plane wave eigenpolarizations to express the boundary conditions at the interfaces. In essence, in such a 2×2 TMF, one can work directly with the traditional Fresnel interface reflection coefficients (see for instance [12]). The waveguide problem to be solved is again that of Fig. 2.3, with all $\underline{\epsilon}_i$ either equatorial or isotropic. As explained in 2.2.2.1 there are two tasks in deriving a TMF framework:

1. Deriving the effective indices of the plane eigenwave solutions in the homogeneous layers. These are needed for the diagonal propagation matrix $\overline{\overline{P}}$, describing the forward and backward propagation of the local plane eigenwaves through a stratified layer of a certain thickness d.

2. Deriving the interface matrix $\overline{\overline{I}}$, describing the connection between 'forward' and 'backward' propagating 'modes' at two sides of an interface between two stratified layers.

Propagation matrix When the propagating plane waves have no wavevector component along the magnetization direction, the base eigenpolarizations are linear and have either a TE or a TM character. The TE modes have only an electric field component parallel with the magnetization (and a fortiori a magnetic field vector perpendicular to it), while for TM modes the magnetic field is aligned with the magnetization and the electric field is perpendicular to it[55]. It was calculated that in these equatorial layers the TE plane waves have an effective index equal to the layer index, while the TM plane waves show a small even magneto-optic correction $n_{\mathrm{TM}} = n\sqrt{1-Q^2} = \sqrt{n^2 - g^2/n^2}$. Thus, the diagonal entries for the propagation matrices in the TM equatorial MO TMF are known. With an assumed propagation constant β, the entries for the diagonal propagation matrix, describing the transversal propagation in the layers of the slab waveguide, are $\exp(-j\sqrt{k_0^2(n^2 - g^2/n^2) - \beta^2}d)$ for the forward (i.e. in the positive x-direction in Fig. 2.3) propagating local eigenwave, and $\exp(+j\sqrt{k_0^2(n^2 - g^2/n^2) - \beta^2}d)$ for the backward one. Writing this compactly as $\exp(\pm jk_{x,i}d_i)$ for the i^{th} layer, the propagation matrix becomes

$$\overline{\overline{P}}_i = \begin{pmatrix} \exp(-jk_{x,i}d_i) & 0 \\ 0 & \exp(+jk_{x,i}d_i) \end{pmatrix} \tag{2.122}$$

In general these square roots are complex, since both the propagation constant and the material parameters can be complex. For the moment, nothing will be assumed about the characteristics of this complex square root. In a later paragraph, it will be explained how the associated twofold Riemann surface of this square root must be investigated.

Interface matrix For the derivation of the TM interface matrix, consider Fig. 2.14.

[55] albeit with some ellipticity in the x-z plane itself! But that is not really relevant, as for TM modes everything will be expressed in terms of the H_y amplitude.

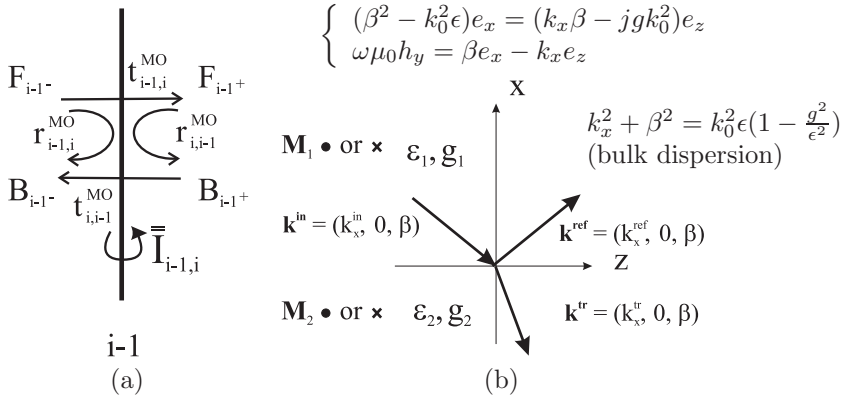


Fig. 2.14: (a) Derivation of the interface matrix $\overline{\overline{I}}_{i-1,i}$; (b) Reflection of a TM plane wave off an interface between two equatorially magnetized MO layers. The Fresnel equations and bulk dispersion valid in each medium are also indicated.

Introducing the MO Fresnel reflection and transmission coefficients $r^{\mathrm{MO}}_{i-1,i}, r^{\mathrm{MO}}_{i,i-1}, t^{\mathrm{MO}}_{i-1,i}$ and $t^{\mathrm{MO}}_{i,i-1}$, with indices referring to Figure 2.3, the following matrix identity can be derived,

$$\begin{cases} F_{i-1^+} = t^{\mathrm{MO}}_{i-1,i} F_{i-1^-} + r^{\mathrm{MO}}_{i,i-1} B_{i-1^+} \\ B_{i-1^-} = r^{\mathrm{MO}}_{i-1,i} F_{i-1^-} + t^{\mathrm{MO}}_{i,i-1} B_{i-1^+} \end{cases} \tag{2.123}$$

$$\begin{pmatrix} F_{i-1^+} \\ B_{i-1^+} \end{pmatrix} = \begin{pmatrix} t^{\mathrm{MO}}_{i-1,i} - r^{\mathrm{MO}}_{i,i-1}(t^{\mathrm{MO}}_{i,i-1})^{-1} r^{\mathrm{MO}}_{i-1,i} & r^{\mathrm{MO}}_{i,i-1}(t^{\mathrm{MO}}_{i,i-1})^{-1} \\ -(t^{\mathrm{MO}}_{i,i-1})^{-1} r^{\mathrm{MO}}_{i-1,i} & (t^{\mathrm{MO}}_{i,i-1})^{-1} \end{pmatrix} \cdot \begin{pmatrix} F_{i-1^-} \\ B_{i-1^-} \end{pmatrix}, \tag{2.124}$$

with the amplitude indices referring to the values taken at either the "lower" or "upper" limits of the interface at x_i. It has to be remarked again that this matrix identity automatically takes the fundamental Maxwell boundary conditions into account, since these are used to derive the reflection and transmission coefficients. The main task therefore is to derive these generalized Fresnel coefficients. Consider for that purpose the right-hand figure in Fig. 2.14. A plane TM wave is incident on an interface between two equatorial MO media 1 and 2 with material parameters ϵ_1 and ϵ_2, and g_1 and g_2. The incident plane wave has a wavevector in the x-z plane $\mathbf{k}^{\mathrm{in}} = (k_x^{\mathrm{in}}, 0, \beta)$. Snell's law demands the z-component of the wavevector of the transmitted and the reflected plane wave to be continuous. Thus $\mathbf{k}^{\mathrm{ref}} = (k_x^{\mathrm{ref}}, 0, \beta)$ and $\mathbf{k}^{\mathrm{tr}} = (k_x^{\mathrm{tr}}, 0, \beta)$. Remark that all wave vectors fulfill the bulk dispersion equality[56] $\mathbf{k}^2 = k_x^2 + \beta^2 = k_0^2 n_{\mathrm{TM}}^2 = k_0^2\epsilon(1 - Q^2) = k_0^2\epsilon(1 - \frac{g^2}{\epsilon^2})$. As a result, obviously one has that $k_x^{\mathrm{ref}} = -k_x^{\mathrm{in}}$.

[56] Remark that the square of a vector is here not in the sense of a norm! One has to be very careful not to oversee the complex character of the vector components.

Using the Fresnel equation for normal plane wave solutions in a bulk medium[57],

$$\left(-\mathbf{kk} + \mathbf{k}\cdot\mathbf{k}\underline{I} - k_0^2\underline{\epsilon}\right)\cdot\mathbf{e} = 0, \tag{2.125}$$

and the first Maxwell rotor equation, a relationship between the e_x, e_z and h_y components can be derived.

$$\begin{cases} e_x = e_z\dfrac{k_x\beta - jgk_0^2}{\beta^2 - k_0^2\epsilon} \\ \qquad = -e_z\dfrac{k_x\beta - jgk_0^2}{k_x^2 + k_0^2(g^2/\epsilon)} \\ \omega\mu_0 h_y = \beta e_x - k_x e_z \end{cases} \tag{2.126}$$

Using (2.126), e_z can be written as a function of h_y, the dominant TM field component. In this way the two field components tangential to the interface, e_z and h_y, can both be expressed as a function of the plane wave amplitude.

$$e_z = \omega\mu_0 h_y / \left[\beta\frac{\beta k_x - jk_0^2 g}{\beta^2 - k_0^2\epsilon} - k_x\right] \tag{2.127}$$

If it is assumed that the incident plane wave has unit amplitude, then by definition of the reflection and transmission coefficients[58] the amplitudes of the reflected and transmitted plane waves will be, $r_{1,2}^{\mathrm{MO}}$ resp. $t_{1,2}^{\mathrm{MO}}$. These generalized Fresnel coefficients follow as a solution of the linear system of equations obtained by applying the Maxwell continuity boundary conditions on h_y and e_z, $h_y^{\mathrm{tr}} = h_y^{\mathrm{ref}} + h_y^{\mathrm{in}}$ and $e_z^{\mathrm{tr}} = e_z^{\mathrm{ref}} + e_z^{\mathrm{in}}$,

$$\begin{cases} t_{1,2}^{\mathrm{MO}} = r_{1,2}^{\mathrm{MO}} + 1 \\ t_{1,2}^{\mathrm{MO}} = r_{1,2}^{\mathrm{MO}}\dfrac{\left[\beta\frac{\beta k_x^{\mathrm{tr}} - jk_0^2 g_2}{\beta^2 - k_0^2\epsilon_2} - k_x^{\mathrm{tr}}\right]}{\left[\beta\frac{-\beta k_x^{\mathrm{in}} - jk_0^2 g_1}{\beta^2 - k_0^2\epsilon_1} + k_x^{\mathrm{in}}\right]} + \dfrac{\left[\beta\frac{\beta k_x^{\mathrm{tr}} - jk_0^2 g_2}{\beta^2 - k_0^2\epsilon_2} - k_x^{\mathrm{tr}}\right]}{\left[\beta\frac{\beta k_x^{\mathrm{in}} - jk_0^2 g_1}{\beta^2 - k_0^2\epsilon_1} - k_x^{\mathrm{in}}\right]} \end{cases} \tag{2.128}$$

[57] dyadic notation **fg** is used for the rank 2 tensors

[58] It is important to remark that there are several ways to define these coefficients. In most texts on electromagnetics (see for instance [33]) the reflection and transmission coefficients are defined as the ratios of the electric field amplitudes. For TE waves no confusion is possible, since there is only one electric field component. TM waves however, have two electric field components and for plane waves neither one of them is dominant. When the media are isotropic both electric field components are in phase, and it is not so difficult to calculate everything in terms of the global electric field amplitude. However, for more complex media it is more logical to define the TM coefficients in terms of the single magnetic field component.
Remark that for isotropic media this will not influence the final formulae for the reflection coefficient, but it will influence the transmission coefficient, as the electric field amplitude and the magnetic field amplitude have a medium dependent proportionality factor: $e \sim \frac{h_y}{n}$, or in other words, $t_{\mathrm{h_y}} = \frac{n_2}{n_1} t_{\mathrm{e}}$.

After tedious algebraic manipulations, one finds that the most general solution to this linear system is,

$$
\begin{aligned}
r_{1,2}^{\mathrm{MO}} = -\frac{(g_1\beta - jk_x^{\mathrm{in}}\epsilon_1)}{(g_1\beta + jk_x^{\mathrm{in}}\epsilon_1)} \times \dots \\
\Big(- g_2^2k_0^2\epsilon_1\big(g_1\beta + jk_x^{\mathrm{in}}\epsilon_1\big) + g_2\beta\epsilon_2\big(g_1^2k_0^2 + (k_x^{\mathrm{in}})^2\epsilon_1\big) + \dots \\
jk_x^{\mathrm{tr}}\epsilon_2\big(g_1^2k_0^2\epsilon_2 + jg_1k_x^{\mathrm{tr}}\beta\epsilon_1 + k_x^{\mathrm{in}}\epsilon_1(k_x^{\mathrm{in}}\epsilon_2 - k_x^{\mathrm{tr}}\epsilon_1)\big)\Big) \Big/ \\
\Big(g_2^2k_0^2\epsilon_1\big(- g_1\beta + jk_x^{\mathrm{in}}\epsilon_1\big) + g_2\beta\epsilon_2\big(g_1^2k_0^2 + (k_x^{\mathrm{in}})^2\epsilon_1\big) + \dots \\
jk_x^{\mathrm{tr}}\epsilon_2\big(g_1^2k_0^2\epsilon_2 + jg_1k_x^{\mathrm{tr}}\beta\epsilon_1 + k_x^{\mathrm{in}}\epsilon_1(k_x^{\mathrm{in}}\epsilon_2 + k_x^{\mathrm{tr}}\epsilon_1)\big)\Big), \quad (2.129)
\end{aligned}
$$

and

$$
\begin{aligned}
t_{1,2}^{\mathrm{MO}} = 2k_x^{\mathrm{in}}\epsilon_1\epsilon_2\frac{(-jg_2\beta + k_x^{\mathrm{tr}}\epsilon_2)}{(-jg_1\beta + k_x^{\mathrm{in}}\epsilon_1)} \times \dots \\
\Big(g_1^2k_0^2 + (k_x^{\mathrm{in}})^2\epsilon_1\Big) \Big/ \Big(g_2^2k_0^2\epsilon_1\big(jg_1\beta + k_x^{\mathrm{in}}\epsilon_1\big) - jg_2\beta\epsilon_2\big(g_1^2k_0^2 + (k_x^{\mathrm{in}})^2\epsilon_1\big) + \dots \\
k_x^{\mathrm{tr}}\epsilon_2\big(g_1^2k_0^2\epsilon_2 + jg_1k_x^{\mathrm{tr}}\beta\epsilon_1 + k_x^{\mathrm{in}}\epsilon_1(k_x^{\mathrm{in}}\epsilon_2 + k_x^{\mathrm{tr}}\epsilon_1)\big)\Big) \quad (2.130)
\end{aligned}
$$

This is the most general form for the Fresnel magneto-optic reflection and transmission coefficient between two equatorial MO media. No approximations have been made, except for the assumption that the magneto-optic tensor has only magnetic linear contributions. Neither the non-reciprocity of these generalized Fresnel coefficients nor their correct limiting behaviour for vanishing $g_1 = g_2 = 0$, are easy to spot in (2.129) and (2.130). However, when neglecting all second and higher order terms in g_1 and g_2 in the nominator and denominator of these formulae, the following first order MO Fresnel coefficients are obtained (after some further simplifications)

$$
r_{1,2}^{\mathrm{MO},\mathcal{O}(1)} = -\frac{(g_1\beta - jk_x^{\mathrm{in}}\epsilon_1)}{(g_1\beta + jk_x^{\mathrm{in}}\epsilon_1)}\frac{(\beta(g_2(k_x^{\mathrm{in}})^2 - g_1(k_x^{\mathrm{tr}})^2) + jk_x^{\mathrm{in}}k_x^{\mathrm{tr}}(k_x^{\mathrm{in}}\epsilon_2 - k_x^{\mathrm{tr}}\epsilon_1))}{(\beta(g_2(k_x^{\mathrm{in}})^2 - g_1(k_x^{\mathrm{tr}})^2) + jk_x^{\mathrm{in}}k_x^{\mathrm{tr}}(k_x^{\mathrm{in}}\epsilon_2 + k_x^{\mathrm{tr}}\epsilon_1))} \quad (2.131)
$$

$$
t_{1,2}^{\mathrm{MO},\mathcal{O}(1)} = j\frac{(g_2\beta + jk_x^{\mathrm{tr}}\epsilon_2)}{(g_1\beta + jk_x^{\mathrm{in}}\epsilon_1)}\frac{2(k_x^{\mathrm{in}})^3\epsilon_1}{(\beta(g_2(k_x^{\mathrm{in}})^2 - g_1(k_x^{\mathrm{tr}})^2) + jk_x^{\mathrm{in}}k_x^{\mathrm{tr}}(k_x^{\mathrm{in}}\epsilon_2 + k_x^{\mathrm{tr}}\epsilon_1))} \quad (2.132)
$$

Two important conclusions can be drawn here.

1. The first order formulae (2.131) and (2.132) correctly reduce to the isotropic

Fresnel coefficients. When putting $g_1 = g_2 = 0$, one finds[59]

$$r_{1,2}^{\text{iso}} = \frac{(k_x^{\text{in}}\epsilon_2 - k_x^{\text{tr}}\epsilon_1)}{(k_x^{\text{in}}\epsilon_2 + k_x^{\text{tr}}\epsilon_1)} = \frac{(n_1\cos\theta_2 - n_2\cos\theta_1)}{(n_1\cos\theta_2 + n_2\cos\theta_1)} \tag{2.133}$$

$$t_{1,2}^{\text{iso}} = \frac{2k_x^{\text{in}}\epsilon_2}{(k_x^{\text{in}}\epsilon_2 + k_x^{\text{tr}}\epsilon_1)} = \frac{2n_2\cos\theta_1}{(n_1\cos\theta_2 + n_2\cos\theta_1)} \tag{2.134}$$

2. The magneto-optic contributions to the reflection and transmission coefficients underline again the non-reciprocal character of this type of interface. Indeed, these terms are odd with respect to the propagation constant. In a "visual" interpretation, this means that one cannot just reverse the sense of the reflecting and refracting "rays". Such a sense reversal gets down to keeping the k_x's equal and changing the sign of β. (2.131) and (2.132) prove that this operation is not a symmetry operation for the generalized Fresnel coefficients. It is not difficult to see that this will lead to a dispersion equation that is odd in β. Imagine for example a three-layer slab waveguide, consisting of e.g. an isotropic substrate and core (of thickness d_{core}), and cladded by an equatorial MO layer. The resonance condition for guided TM modes requires the sum of the reflection phase shifts at the interfaces between the (isotropic) core and MO cladding, φ_1, and core and (isotropic) substrate, φ_2, and the propagation phase shift, $k_{x,\text{core}} = k_0\sqrt{n_{\text{core}}^2 - \beta^2}d_{\text{core}}$, to add up to a multiple of π. (2.131) clearly shows that φ_1 is not an even function of β, while $k_0\sqrt{n_{\text{core}}^2 - \beta^2}d_{\text{core}}$ is even in β and (2.133) shows that φ_2 is also a quadratic function of β (via k_x). As a result the resonance condition $\varphi_1 + \varphi_2 + k_{x,\text{core}}d_{\text{core}}$, will behave as an odd function of the propagation constant, and therefore be direction-dependent.

All information has been collected to calculate the complete scattering matrix of an equatorial MO slab waveguide. Returning to Fig. 2.3, the scattering matrix of the MO slab is constructed as a concatenated product of "chunk" matrices. These are obtained by the matrix product of a propagation matrix and an interface matrix, the first "chunk" is formed by the substrate interface at x_0 and the first homogeneous layer between x_0 and x_1. This gives rise to the chunk matrix $\overline{\overline{m}}_1$, relating the forward and backward fields at x_{0-} and x_{1-}. The general form of the i^{th} chunk matrix is thus found by combining (2.122) and (2.124)

$$\overline{\overline{m}}_i = \begin{pmatrix} \exp(-jk_{x,i}d_i)(t_{i-1,i}^{\text{MO}} - r_{i,i-1}^{\text{MO}}(t_{i,i-1}^{\text{MO}})^{-1}r_{i-1,i}^{\text{MO}}) & \exp(-jk_{x,i}d_i)r_{i,i-1}^{\text{MO}}(t_{i,i-1}^{\text{MO}})^{-1} \\ -\exp(jk_{x,i}d_i)(t_{i,i-1}^{\text{MO}})^{-1}r_{i-1,i}^{\text{MO}} & \exp(jk_{x,i}d_i)(t_{i,i-1}^{\text{MO}})^{-1} \end{pmatrix} \tag{2.135}$$

The "backward" Fresnel coefficients ($r_{i,i-1}$ and $t_{i,i-1}$) are found by interchanging the material parameters, and by making the interchanges $k_x^{\text{tr}} \to -k_x^{\text{in}}$ and $k_x^{\text{in}} \to -k_x^{\text{tr}}$, in (2.129) and (2.130) (with obviously $k_x^{\text{in}} = k_{x,i-1}$ and $k_x^{\text{tr}} = k_{x,i}$).

[59] Take into account that the transmission coefficient is defined for the h_y amplitude. To find the traditional Fresnel coefficient, one needs to multiply by n_1/n_2 as pointed out above.

The dispersion equation To construct the dispersion equation of the planar equatorial MO waveguide, there are basically two routes. As described in 2.2.2.2, one can opt for a either *closed* or an *open* formulation of the dispersion equation. The main difference is the boundary conditions imposed on the modal solutions in the superstrate and the substrate of the waveguide. In an open waveguide, the proper modal solutions are found by imposing physically acceptable radiation conditions towards infinity, while in a closed waveguide the modal fields or their derivatives are forced to zero at a certain distance inside the sub- and superstrate. As explained in 2.2.2.2, it is common practice to artificially enclose the waveguide within two perfectly conducting walls, and hence to use a closed dispersion equation. However, there are several reasons as to why using an open formulation of the dispersion equation is more interesting [34]. The most convincing of these is the fact that an open dispersion equation will treat proper and improper solutions of the waveguide on the same level. Under proper solutions, one understands those waveguide modes with exponential decay in the sub- and the superstrate (or at least with a finite value at infinity). Improper solutions will have an exponential increasing mode profile in at least one of the two cladding layers. They are sometimes also called leaky modes, and do not belong the complete eigenspectrum of the waveguide. When using an *open* dispersion equation the proper modes and improper modes are separately found in well-defined regions of a Riemann surface [35]. When the waveguide parameters are varied, the open dispersion equation allows to monitor the migration of both the proper and improper modes on the Riemann surface. Every time an improper mode migrates into the region of the Riemann surface that belongs the proper modes, mode "cut-on" is said to occur. Therefore, leaky modes are generally also labelled as proper modes in cut-off. Because an open dispersion equation allows a treatment of proper and improper modes on the same footing, cut-on phenomena can be easily predicted. It can be proven that a closed dispersion equation will never have improper solutions. This might sound as a blessing, but at the same time it is annoying because now cut-on will appear without any warning! Not the most ideal situation for a root-solver when for instance waveguide parameters are scanned. Therefore the subtleties of open and closed dispersion equations will be discussed here.

The open dispersion equation is found by evaluating the elements of the scattering matrix that relates the "forward" and "backward" field amplitudes at the lower side of the substrate interface to those amplitudes at the upper side of the superstrate interface. As pointed out in the previous paragraph, the scattering matrix can easily be constructed by concatenating chunk matrices. By labelling the chunk matrices as in (2.135), and by assuming N layers with the substrate and superstrate labelled as 0 and N+1 respectively, the concatenated product runs from chunk matrix $\overline{\overline{m}}_1$ through $\overline{\overline{m}}_{N+1}$. In order to avoid that this last chunk matrix will also include a propagation in the superstrate, the thickness d_{N+1} can be artificially set to 0. The resulting matrix is thus

$$\begin{pmatrix} F_{N+} \\ B_{N+} \end{pmatrix} = \underbrace{\prod_{i=1}^{N+1} \overline{\overline{m}}_i}_{\overline{\overline{T}}} \begin{pmatrix} F_{0-} \\ B_{0-} \end{pmatrix} \tag{2.136}$$

It should be very strongly emphasized here, that the labels "forward" and "backward" only get a physical meaning when they are related to a certain complex exponential wave behaviour. In the most general case, being layered stacks containing non-transparent (i.e. absorptive or amplifying) materials, a $\exp(-jk_x d)$ behaviour can only be physically considered as a forward plane eigenwave in the transverse direction of the layered structure, if it exhibits an exponential amplitude decay. Or, in other words, if the square root $k_x = \sqrt{k_0^2 n^2 (1-Q^2) - \beta^2}$ has a negative imaginary part, independent of the sign of its real part! However, this does not imply that solutions for the modal propagation constant β leading to k_x with a positive imaginary part should be rejected as such. When constructing the dispersion equation, the terms "forward" and "backward" have no physical meaning, but are merely mathematical solutions to the transverse Helmholtz equation in each homogeneous layer. A "forward" mode being one with $\exp(-jk_x d)$ behaviour, and a "backward" mode, being one with $\exp(+jk_x d)$ behaviour, no matter what the sign of the imaginary part of k_x is. The physical meaning only enters the picture if physically feasible boundary conditions are introduced. Of course, this whole discussion relates to the multi-valuedness of the complex square root k_x, and the fact that each complex square root introduces a double-sheeted Riemann surface in the complex β plane. Therefore, it is important to study the branch-points and branch-cuts of the open dispersion equation, derived out of the elements of (2.136).

Any of the elements of $\overline{\overline{T}}$, can be interpreted as an implicit complex dispersion equation for the modes of an open multilayered structure. Thus in general, there are four possible dispersion relations, depending on which of the four possible combinations of the radiation boundary conditions are used. This is summarized in Table 2.7.

$T_{i,j}$	BC substrate	BC superstrate
(1,1)	$B_0 = 0; \sim e^{-jk_{x,0}(x-x_0)}$	$F_{N+1} = 0; \sim e^{+jk_{x,N+1}(x-x_N)}$
(1,2)	$F_0 = 0; \sim e^{+jk_{x,0}(x-x_0)}$	$F_{N+1} = 0; \sim e^{+jk_{x,N+1}(x-x_N)}$
(2,1)	$B_0 = 0; \sim e^{-jk_{x,0}(x-x_0)}$	$B_{N+1} = 0; \sim e^{-jk_{x,N+1}(x-x_N)}$
(2,2)	$F_0 = 0; \sim e^{+jk_{x,0}(x-x_0)}$	$B_{N+1} = 0; \sim e^{-jk_{x,N+1}(x-x_N)}$

TABLE. 2.7: Four possible dispersion equations $T_{i,j}$ and corresponding radiation boundary conditions in the sub- and the superstrate. $T_{2,2}$, i.e. no "incoming" fields from infinity in the sub- and superstrate, represents the commonly encountered dispersion equation in integrated optics.

It is wrong to conclude that only $T_{2,2}$ is the physically feasible choice for a dispersion relationship allowing the determination of the proper (i.e. bound) waveguide modes. This would only be the case if it is a priori established that the square roots $k_{x,0}$ and $k_{x,N+1}$ have negative imaginary parts. And that is only possibly if the Riemann surface associated with $T_{2,2}$ is completely known, so that the correct sheet can be chosen. Therefore, it is interesting to study the Riemann surface

associated with the elements of $\overline{\overline{T}}$ in little more detail. This will help to prove that the proper waveguide modes can be found by using *any* of the four possible dispersion equations, by investigating the appropriate associated Riemann sheet. A Riemann surface [35] is a multi-sheeted surface in the complex plane on which a multi-valued complex function – here, $T_{i,j}(\beta)$ with $i,j = 1,2$ – can be regarded as single-valued. For every point on the Riemann surface, the function takes on an unique value. It is, in other words, a topological multiple folding of the complex plane. The importance of the Riemann surface is that it allows to describe complex functions, that would otherwise require branch-cuts and branch-points in the "standard" complex plane, by a continuation that is analytical everywhere on this surface. As such it allows to use powerful numerical techniques that are only applicable on analytical functions[60]. Smith and Houde-Walter [34], [38], [39], have studied the Riemann surface of multilayered complex *isotropic* slab waveguides in great detail. Their conclusions will be summarized here, after which it will be indicated how these can be generalized to the case of general (non-transparent), equatorial, magneto-optic slab waveguides.

Riemann surface of isotropic multi-layered slab waveguides If the multi-layered slab waveguide is purely isotropic, then it is easily calculated, using (2.135) and (2.133), that the (TM) chunk matrices take the following form

$$\overline{\overline{m}}_i = \frac{1}{2}\begin{pmatrix} (1+\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}})\exp(-jk_{x,i}d_i) & (1-\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}})\exp(-jk_{x,i}d_i) \\ (1-\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}})\exp(+jk_{x,i}d_i) & (1+\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}})\exp(+jk_{x,i}d_i) \end{pmatrix}, \quad (2.137)$$

with of course $k_{x,i} = \sqrt{k_0^2\epsilon_i - \beta^2}$. Looking at this chunkmatrix it is clear that for all four elements of $\overline{\overline{T}}$ in (2.136), the complex points $k_0^2\epsilon_i$ $(i = 0,\ldots,N+1)$ might introduce both branch-points (through the square roots) and singularities (via the denominator of $\overline{\overline{m}}_i$) in the complex β-plane. Through (2.137) it is also clear that all $k_{x,i}$ dependence of the elements of $\overline{\overline{T}}$ is enclosed in only two chunk matrices, namely, $\overline{\overline{m}}_i$ and $\overline{\overline{m}}_{i+1}$. This means that before it can be concluded that all $k_{x,i}$ are singularities of $\overline{\overline{T}}$, the product of these two matrices must be evaluated. The following result is obtained[61]

$$\overline{\overline{m}}_{i+1}\cdot\overline{\overline{m}}_i = \frac{1}{2}\begin{pmatrix} e^{-j\varphi_{i+1}} & 0 \\ 0 & e^{+j\varphi_{i+1}} \end{pmatrix}\cdot \quad (2.138)$$

$$\begin{pmatrix} (1+\frac{k_{x,i-1}\epsilon_{i+1}}{k_{x,i+1}\epsilon_{i-1}})\cos\varphi_i - j(\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}} + \frac{k_{x,i}\epsilon_{i+1}}{k_{x,i+1}\epsilon_i})\sin\varphi_i & \cdots \\ (1-\frac{k_{x,i-1}\epsilon_{i+1}}{k_{x,i+1}\epsilon_{i-1}})\cos\varphi_i - j(\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}} - \frac{k_{x,i}\epsilon_{i+1}}{k_{x,i+1}\epsilon_i})\sin\varphi_i & \\ (1-\frac{k_{x,i-1}\epsilon_{i+1}}{k_{x,i+1}\epsilon_{i-1}})\cos\varphi_i + j(\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}} - \frac{k_{x,i}\epsilon_{i+1}}{k_{x,i+1}\epsilon_i})\sin\varphi_i & \\ (1+\frac{k_{x,i-1}\epsilon_{i+1}}{k_{x,i+1}\epsilon_{i-1}})\cos\varphi_i + j(\frac{k_{x,i-1}\epsilon_i}{k_{x,i}\epsilon_{i-1}} + \frac{k_{x,i}\epsilon_{i+1}}{k_{x,i+1}\epsilon_i})\sin\varphi_i & \end{pmatrix},$$

[60] One of those techniques is the complex root-finding algorithm of Delves and Lyness [36] (see also [12]) using the Cauchy contour integral theorem [37], which is of obvious interest here for locating the complex zeros of the complex dispersion dispersion equation.

[61] For clarity and compactness the propagator factors of the $i+1^{\text{th}}$ layer are taken out of the matrix and put into a diagonal matrix.

with $\varphi_i = k_{x,i} d_i$.
A single revolution of 2π radians around the assumed branchpoint $k_0^2\epsilon_i$ results in a change of sign of the square root $k_{x,i}$ (and thus also of φ_i). However, (2.138) proves that the elements of $\overline{\overline{m}}_{i+1} \cdot \overline{\overline{m}}_i$ remain unchanged under this sign reversal. Thus it can be concluded that $k_0^2\epsilon_i$ is <u>not</u> a branchpoint of the elements of $\overline{\overline{T}}$ in the complex β plane. Furthermore, looking at the elements of $\overline{\overline{m}}_{i+1} \cdot \overline{\overline{m}}_i$, it is seen that with β in the neighborhood of the assumed singularity $k_0^2\epsilon_i$ (i.e. $k_{x,i} \to 0$), the possible divergent terms in the elements of the matrix product (2.138) behave proportionally to $\frac{\sin 0}{0}$, so that even in an arbitrarily close region of $k_0^2\epsilon_i$ the elements of $\overline{\overline{T}}$ are always finite.
This reasoning can be applied for all $k_0^2\epsilon_i$, except for the two outer ones, $k_0^2\epsilon_{N+1}$ and $k_0^2\epsilon_0$! These latter appear only in one chunk matrix in $\overline{\overline{T}}$ and can thus not be removed in similar fashion. However, it can easily be proven that only $k_0^2\epsilon_{N+1}$ is actually a singular branch-point. $k_0^2\epsilon_0$ is a branchpoint, but the elements of $\overline{\overline{T}}$ remain finite when $k_0^2\epsilon_0$ is approached arbitrarily close. It is well-known (see for example [40]) that a complex square-root is double-valued, and hence results in a two-sheeted Riemann surface. Because the elements of $\overline{\overline{T}}$ have two square root branch points, $k_0^2\epsilon_{N+1}$ and $k_0^2\epsilon_0$, the combination of these two branch points causes the Riemann surface to become four-sheeted. Fig. 2.15 shows a schematical representation of a possible topology for this four-sheeted surface for an arbitrary multilayered isotropic, non-transparent waveguide. In this figure it is also schematically indicated what field types are to be found on each sheet of the Riemann surface for a particular choice of dispersion equation. A "non-traditional" disperion relationship $T_{2,1}$ has been chosen in order to show that all four possible relationships are equivalent. Fig. 2.15 makes it clear that all four elements of $\overline{\overline{T}}$ have the same four values for any given value β^2 – the only difference being the sheet on which these values are located! For example, if a zero for $T_{2,2}$ is found for a certain β located on the upper-right sheet, for the same β a zero will be found on the lower-right sheet for $T_{1,1}$. This is straightforward to understand by considering the change of sign of the square roots occurring upon a change of Riemann sheets. This again underlines that all four possible dispersion relationships are equivalent. If a certain dispersion relationship is chosen, it is straightforward to identify the region (or the sheet) of the Riemann surface where the physically feasible waveguide modes are to be found. This is summarized in Table 2.8, where for every subsheet of the Riemann surface in Fig. 2.15 it is indicated which of the four dispersion equations will have proper mode zeroes within that subsheet. The results in this table are easily understood when combining the tabulated signs of the complex of the square roots in Fig. 2.15 with the exponential boundary behaviour in sub- and superstrate tabulated in Table 2.7. Thus, for example, for the "natural" dispersion equation $T_{2,2}$, the proper modes must be sought in subsheets 2 and 3. However, even though with Table 2.8 it is possible to identify the region of interest on the Riemann surface, it is not possible to simply "separate" this sheet from the rest of the surface. That would imply that one would have to "tear off" a certain part of the surface, which is of course mathematically incorrect. A correct mathematical treatment of any complex function requires by definition the entire

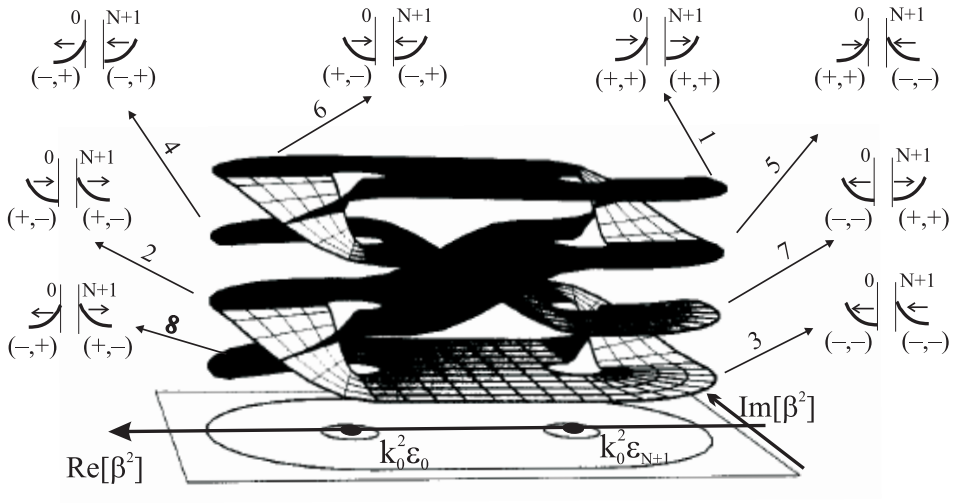


Fig. 2.15: Four-sheeted Riemann surface of an arbitrary multi-layered complex slab waveguide (for simplicity it is assumed that $\Im[k_0^2\epsilon_0] = \Im[k_0^2\epsilon_{N+1}] = 0$) and corresponding subdivision in 8 subsheets, indicated by the signs of the real and imaginary parts of the two square roots, $k_{x,0}$ and $k_{x,N+1}$, ($sgn(\Re(k_x, 0|N+1)) = \pm, sgn(\Im(k_x, 0|N+1)) = \pm$). Also indicated is the field types to be found within the 8 subsheets if a certain combination of radiation boundary conditions is used. In this case the combination of boundary conditions associated with the dispersion relationship $T_{2,1}$, i.e. $\sim \exp(-jk_{x,0}x)$ and $\sim \exp(-jk_{x,N+1}x)$ in substrate and superstrate respectively. Remark that off course the apparent intersections of the surface are artificial. It is obviously impossible to draw a three-dimensional picture of a surface on which any closed contour in the complex plane undergoes a 4π revolution.

associated Riemann surface to be considered! That implies a.o. that contour integrations in the complex plane (for example needed in the root-finding Argument Principal Method of Delves and Lyness [36]), must be performed along the (often complicated) projection of these contours on the entire four-sheeted surface. It is clear that this is not the easiest way to proceed[62]. If one wants to investigate separately the physically feasible regions of the Riemann surface, then this four-sheeted folding of the complex plane must be unfolded onto another complex plane of a new independent complex variable. It has been shown by Smith [39] how this unfolding can be achieved through a conformal mapping of the β-plane onto the u-plane, with $u = \sqrt{k_0^2\epsilon_0 - \beta^2} + \sqrt{k_0^2\epsilon_{N+1} - \beta^2} = k_{x,0} + k_{x,N+1}$. Any of the four dispersion equations can be expressed in terms of this new independent variable u, $T^{\star}_{i,j}(u) = T_{i,j}(\beta^2(u))$, and all 8 subsheets of the 4-sheeted Riemann surface over

[62] On the other hand this mathematical obligation to consider the entire Riemann surface has as an added plus that also the improper (or so-called below cut-off) modes are obtained. And their "cut-on" migration as a function of changing material parameters provide a lot of informative insight in the behaviour of the waveguide [34].

subsheet	$T_{i,j}$	gain/loss
1	(1,1)	loss
2	(2,2)	gain
3	(2,2)	loss
4	(1,1)	gain
5	(2,1)	loss
6	(1,2)	gain
7	(1,2)	loss
8	(2,1)	gain

TABLE. 2.8: Sheet location of the proper waveguide modes for the four different dispersion equations.

the β-plane get unfolded into 8 neighbouring regions of the complex u-plane. This is illustrated in Fig. 2.16 where the Riemann surface and the associated $T_{1,2}$ field types of Fig. 2.15 are projected using the above mapping function. Using the conformal mapping, the complicated double loops traced on the 4-sheeted Riemann surface by projections of contours in the β-plane, now become plane curves in the u-plane, which are much easier to evaluate numerically. As a result, it now becomes possible to use root-finding contour integration methods, such as APM, directly on the region of interest in the u-plane. In other words, the conformal mapping has allowed to "tear off" the Riemann subsheets containing the proper modes.

Riemann surface of MO equatorial multi-layered slab waveguides After this overview of the subtleties of open dispersion equations for isotropic multi-layered slab waveguides, we turn our attention to how the situation changes when the slab waveguide becomes equatorial magneto-optic.

Again the four equivalent dispersion equations are given by the elements of $\overline{\overline{T}}$. But this time the chunk matrices, are considerably more complicated. The difference with the isotropic chunk matrix is the occurrence of the MO Fresnel coefficients (2.129) and (2.130) (or their first-order versions (2.131) and (2.132)). However, the two main features of the chunk matrices, will remain the same – firstly, the possible singularities and branch-points at the square roots $k_{x,i}$, and secondly the fact that all dependence of the $T_{i,j}$ on $k_{x,i}$ is enclosed in just two matrices, $\overline{\overline{m}}_i$ and $\overline{\overline{m}}_{i+1}$. In the previous paragraph it was proven that in a purely isotropic slab, the elements of $\overline{\overline{T}}$ will only have branchpoints at the wavenumbers of the sub- and the superstrate, $k_0^2\epsilon_0$ and $k_0^2\epsilon_{N+1}$. It is intuitively clear that this will not change for MO equatorial slabs (with now branchpoints at $k_0^2\epsilon_{N+1}(1-Q_{N+1}^2)$ and $k_0^2\epsilon_0(1-Q_0^2)$). However, what is not clear is whether again in the product $\overline{\overline{m}}_{i+1}\cdot\overline{\overline{m}}_i$ the possible "interior" branchpoint $k_0^2\epsilon_i(1-Q_i^2)(i=1,\ldots,N)$ is removable when MO equatorial chunk matrices are used. Lengthy calculations prove that this is indeed the case. This implies that the Riemann surface associated with an equatorial MO slab, has (nearly) all the same characteristics as the one of

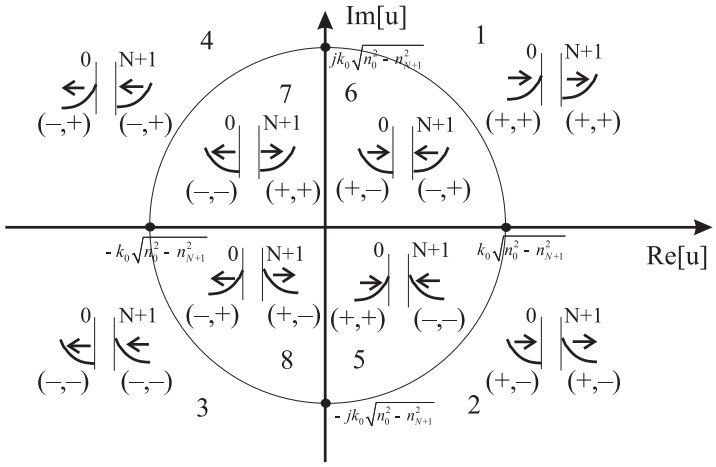


Fig. 2.16: Conformal mapping of Fig.2.15 onto the u-plane, with $u = k_{x,0} + k_{x,N+1}$. *Note how the boundaries between the different Riemann subsheets transform to the u-plane, by investigating the corresponding signs of the comlex parts of* $k_{x,0}$ *and* $k_{x,N+1}$. *For instance on the boundary between sheet 5 and 8, starting at* $k_0^2\epsilon_0$ *and extending towards* $+\infty$ *in Fig.2.15, the terms in the square roots* $k_0^2\epsilon_0 - \beta^2$ *and* $k_0^2\epsilon_{N+1} - \beta^2$ *are purely negative real. The square roots will thus be purely imaginary, and the signs for these can be looked up in Fig. 2.15. There it seen that across this boundary the imaginary sign of both square roots is continuous, and* $\Im(k_{x,0}) > 0$ *and* $\Im(k_{x,N+1}) < 0$. *Thus* $+\infty$ *gets projected onto the origin of the u-plane and* $k_0^2\epsilon_0$ *onto the point* $-j\sqrt{k_0^2\epsilon_0 - k_0^2\epsilon_{N+1}}$. *Similar reasoning can be applied to all other boundaries. With the choice of* $T_{2,1}$ *all proper bound modes are found in a semicircle in the u-plane. This considerably simplifies the root-searching algorithms.*

a purely isotropic slab waveguide. The shape and the physically feasible sheets are the same, and even the same conformal mapping function can be applied. There are however a few subtleties involved. It can also be proven that, as opposed to the isotropic case, the interior wavenumbers introduce poles of order 2. Therefore they are not entirely removable isolated singularities. They are not branchpoints, but still behave as poles. This is however not so dramatic, since infinitesimal small neighbourhoods of $k_0^2\epsilon_i(1 - Q_i^2)$ $(i = 1, \ldots, N)$ can be removed out of the Riemann surface without altering its topology. On top of that the elements of $\overline{\overline{T}}$ contain terms of odd order in β, as a result of the non-reciprocal MO contributions in the Fresnel coefficients. This implies that now the dependence of the dispersion equations $T_{i,j}$ on the modal propagation constant is not solely via the square roots $k_{x,i} = \sqrt{k_0^2\epsilon_i(1 - Q_i^2) - \beta^2}$. As a result in the dispersion equation the independent variable is not β^2, but β. This would imply that the Riemann surface has to be considered on the β-plane instead of the β^2-plane. This is rather annoying since that would influence the topology of this surface, because on the β-plane the branchpoints then split up, $(\pm k_0\sqrt{\epsilon}\sqrt{1 - Q^2})$, resulting in a total of

4 branchpoints. However, this behaviour can be subtly avoided by writing β as $\sqrt{\beta^2}$, and using a sign convention for this square root before any calculation is performed. If one is solving for forward modes, the square root has to be evaluated with positive real parts, and vice versa. In this way the possible ambiguities of the conformal mapping are also avoided, as the independent variable is again β^2. All other conclusions of the previous paragraph remain valid, in particular Table 2.8 summarizing the physically feasible sheets of the Riemann surface and their associated dispersion equation. The non-reciprocity of the dispersion equation is completely included in the above explained "$\beta = \sqrt{\beta^2}$" trick. The dispersion equation and its zeroes will change depending on the sign convention used for the evaluation of this square root.

Closed dispersion equation – wall boundary conditions As already mentioned, most mode solvers make use of a *closed* waveguide dispersion equation. The motivation is of course a discretisation of the radiation continuum, needed for example to model signal propagation through an optical integrated device by means of an eigenmode expansion technique (see for instance [12], [41]).
A *closed* dispersion equation is constructed by imposing artificial boundaries. A boundary condition can be very exotic. It is in any case only a mathematical construct needed to lift the degrees of freedom in the most general solution of the Maxwell equations describing the electromagnetic problem, and expressed by the arbitrariness of the amplitudes of the forward and backward waves in either the sub- or the superstrate of the waveguide. In a strict mathematical sense though, every correct boundary condition to a second order partial differential equation should be reducible to a von Neumann, a Dirichlet or a Cauchy condition[63]. In [12] it is shown that this implies that the most general canonical form for a boundary condition should be a linear relationship between the F and B amplitudes, $F - rB = 0$, where r can be interpreted as a complex amplitude reflection coefficient (or $rF - B = 0$ for a wall in the superstrate layer). Any choice for r can in principal be connected with a von Neumann, a Dirichlet, or a Cauchy boundary condition. Of course, in practice only certain values for r will have a physical meaning, and in common practice one mostly uses a boundary condition with a constant value for r. However, there are perfectly physically feasible boundary conditions where the ad hoc reflection coefficient r is not constant but depends on the incidence angle of the waves. These so-called Transparent Boundary Conditions (TBC) are extremely popular in Beam Propagation algorithms (BPM) in order to provide a kind of "reflectionless" behaviour of the propagating radiation modes in the modelling of integrated devices. TBC however need to be "tailored" in order to meet the modelling needs specific to the device under study[64]. In this context it should be noted also that the nowadays extremely fashionable Perfectly Matched Layer (PML) techniques (see a.o. [43], [44]) do not constitute a

[63] The latter is a weighted average of a the former two [42]

[64] One can easily prove that a TBC will give rise to a Cauchy condition. One very well known example of a Cauchy condition is the boundary condition at infinity in an open waveguide, being the complete absence of reflections. This is nothing else than the radiation condition, which for TE modes is written as $-jk_xE_y + E_y^{'} = 0$

real boundary condition in the strict sense of the word, since they do not impose a mathematical condition at one position in the waveguide.
For pure mode solving purposes, there is no need for these more advanced types of boundary conditions. All that is needed is a "simple" (but physically feasible) boundary condition that will discretisize the radiation continuum. For this purpose one most often makes use of perfect conducting walls. These force the magnetic or the electric tangential field component to zero at the wall. If the wall is positioned far enough of the guiding core of the waveguide, its influence on the modal properties of the proper bound modes will be negligible, and its only effect will be a sampling of the radiation continuum. There are two types of perfect conductors.

Perfect Magnetic Conductor will force the tangential magnetic field component to zero. If the dispersion equation is derived for TM modes (thus with modal amplitudes proportional to H_y as is here the case), this implies $F + B = 0$ or a reflection coefficient $r = -1$. This corresponds to a Dirichlet boundary condition. If the dispersion equation is derived for TE modes (thus with modal amplitudes proportional to E_y), this implies $H_z = 0$ or $F - B = 0$ (and hence $r = 1$). This can be easily derived out of the first Maxwell curl equation applied on TE plane waves, which predicts proportionality between H_z and $\frac{\mathrm{d}E_y}{\mathrm{d}x}$. So, a PMC wall for TE modes is equivalent to a von Neumann condition.

Perfect Electric Conductor will force the tangential electric field component to zero. For TE modes this is a Dirichlet condition leading straightforwardly to $F + B = 0$ or $r = -1$. If a PEC wall is used for TM modes then of course we find the analogon of a PMC wall used for TE modes, namely a von Neumann condition $\frac{\mathrm{d}H_y}{\mathrm{d}x} = 0$ or $F - B = 0$ (and $r = 1$). For the last case it is also interesting to note that this is only correct if the wall is placed in an isotropic material, because only then the second Maxwell curl equation will allow a direct relationship between the derivative of H_y and E_z. For example, 2.127 illustrates the difficulty if the PEC was placed inside an equatorial MO layer. In that the case the boundary condition itself would be non-reciprocal!

In all four cases the reflection coefficient has unit amplitude, underlining once again the hardness of these boundary conditions.

Having derived the wall boundary conditions, the closed dispersion equation now follows directly out of (2.136), by multiplying $\overline{\overline{T}}$ on the left and the right with propagation matrices into the truncated sub- and superstrate, subsequently on the left with a row vector expressing the wall condition at the superstrate and finally choosing amplitudes at the substrate wall, fulfilling the wall boundary condition,

$$\left(r_{\mathrm{super}} e^{-jk_{x,N+1}d_{N+1}} \quad -e^{jk_{x,N+1}d_{N+1})}\right) \cdot \overline{\overline{T}} \cdot \begin{pmatrix} r_{\mathrm{sub}} e^{-jk_{x,0}d_0} \\ e^{+jk_{x,0}d_0} \end{pmatrix} = 0, \qquad (2.139)$$

with $r_{\text{sub|super}} = \pm 1$ depending on whether a PEC or PMC wall is considered. Written out this gives

$$r_{\text{super}} r_{\text{sub}} e^{-jk_{x,N+1}d_{N+1} - jk_{x,0}d_0} T_{11} + r_{\text{super}} e^{-jk_{x,N+1}d_{N+1} + jk_{x,0}d_0} T_{12} - r_{\text{sub}} e^{+jk_{x,N+1}d_{N+1} - jk_{x,0}d_0} T_{21} - e^{+jk_{x,N+1}d_{N+1} + jk_{x,0}d_0} T_{22} = 0 \quad (2.140)$$

In the previous paragraph, the symmetry properties of the Riemann surface were pointed out, and more importantly the fact that for every choice of subsheet only one of the four T elements will have zeroes corresponding to proper modes (see Table 2.8). The zeroes of the other elements in the chosen sheet will correspond to improper or cut-off proper modes of the waveguide. In (2.140) all four T elements are simultaneously present. Therefore, one could argue that for a closed dispersion equation it is possible that when a choice of a subsheet (or a region in the complex u-plane after conformal mapping) has been made the complex root-finding algorithm might converge to improper waveguide modes. However, this is a false argument. In (2.140) the elements of T are weighted by exponential factors. A closer investigation of the "damping" (or "amplifying") behaviour of these factors makes clear that no matter what sheet is chosen, the zeroes of the natural (proper) T element will always be favoured in (2.140). This behaviour gets more pronounced when the thickness of the truncated sub- and superstrate are increased, which is logical as the closed equation should converge to an open dispersion equation. One can actually prove that it is fundamentally impossible to retrieve improper (below cut-off) modes with a closed dispersion equation.

A final remark concerns the numerical stability of the dispersion equation. Just as in section 2.2.2.2, all chunk matrices in a 2×2 TMF can be numerically stabilized by dividing them by the exponential propagator factor with a positive real exponent. This depends on the sign of the imaginary part of the square roots, $k_{x,i} = \sqrt{k_0^2 \epsilon_i (1 - \frac{g_i^2}{\epsilon_i}) - \beta^2}$ $(i = 0, \ldots, N+1)$. For open dispersion equations there is no propagation into the sub- and superstrates. On top of that it was seen that the inner layers introduce no branch-points. Therefore, the sign of the imaginary part of the square root is completely determined by the imaginary sign of $k_0^2 \epsilon_i (1 - \frac{g_i^2}{\epsilon_i}) - \beta^2$, and not by some choice of a Riemannsheet, since there aren't any for the inner layers. If this expression has a positive imaginary part, then $\exp(-jk_{x,i}d_i)$ is the increasing exponential, and vice versa in the opposite case. For root-searching algorithms it is thus important to continuously monitor the imaginary signs of square roots as this obviously changes as a function of the successive root guesses for β.

For closed dispersion equations, the propagator factors in the truncated sub- and superstrate also enter the picture. For these it was seen that the square roots do introduce branch-points, so that the imaginary sign of the corresponding square roots is completely determined by the choice of subsheets of the Riemann surface. When a certain combination of subsheets is chosen, it is immediately clear which propagator will introduce exponential increase. During the root search the successive guesses can again migrate between two subsheets (for example from sheet 1 to 4, or from 2 to 3,... as explained in Table 2.8). But the subsheets are always combined so that the imaginary sign of the positive square root will be the same

(see Fig. 2.15) no matter on which of the two subsheets the root is found. Therefore, for the sub- and superstrate it is always the same propagator factor that induces exponential increase, as opposed to the propagator factors of the inner layers. Hence with closed dispersion equations, a root-searching algorithm only needs to monitor the signs of the square roots of the inner layers for numerical stability.

Modal profiles As was explained in section 2.2.2.4, once the dispersion equation (either open or closed) is solved, the modal profiles can be calculated by "propagating" a correct starting field throughout the layered stack using the chunk matrices. We will not repeat this procedure here as it is obviously just a specialized 2×2 version of the general 4×4 algorithm. We will only summarize the different starting fields for the different types of dispersion equation.

Closed waveguides take $F_{-1^+} = \pm 1$ and $B_{-1^+} = 1$ as their starting field at the substrate wall, the plus sign being applicable to PMC walls and the minus sign to PEC walls, and assuming that we are solving for TM modes. For TE modes, the signs reverse.

Open waveguides take as their starting field $F_{0^-} = 0$ and $B_{0^-} = 1$, or $F_{0^-} = 1$ and $B_{0^-} = 0$ depending which Riemann sheet the proper solutions are sought.

Again, calculation of the modal profiles in this manner will suffer from numerical instabilities if the stack gets thick. But also here it is possible to start the transfer of a starting field somewhere in the middle of the stack, in order to avoid propagation of large distances. As opposed to the situation for the full 4×4 anisotropic TMF, described in 2.2.2.4, it is not necessary to rederive the dispersion relationship from an internal viewpoint. It simply suffices to calculate the left or right substack reflection coefficient at the starting position, say x_i. The starting field at that position then becomes, $F_{i^-} = 1$ and $B_{i^-} = r_{i,\text{right}} F_{i^-}$ (or $F_{i^+} = r_{i,\text{left}} B_{i^+}$ and $B_{i^+} = 1$). These scalar reflection coefficients at an equatorial MO stack are obviously nothing else than the TM diagonal element of the general 2×2 reflection matrices of anisotropic stacks. For their calculation the general approach outlined in [13] can again be used. Needless to say that these coefficients are always numerically stable.
The reason why in the full 4×4 algorithm one first had to rederive the disperion relationship is because the wavefields $\mathbf{w} \uparrow$ and $\mathbf{w} \downarrow$ are two dimensional and not scalar. As a result an extra degree of freedom needs to be eliminated.

Finally, just as in section 2.2.2.4, the normalization requires the adjoint mode to be known as well. Also for this aspect things simplify considerably in the case of equatorial MO waveguides. As explained, this normalization is only possible if an extra relationship is known between the mode and its adjoint, which is in general not straightforward. However, due to Onsager's principle (see section 2.2.3.5 and [16]), the adjoint waveguide is nothing else but the original waveguide physically rotated over 180˚along the x axis. The transversal adjoint slab TM mode components (needed for the normalization integral (2.57)) are thus given

by $\tilde{H}_y(x) = -H_y(x)$ and $\tilde{E}_x(x) = E_x(x)$. Substituted in (2.57) this gives for a slab TM mode $\frac{1}{2}\int E_x H_y \mathrm{d}x$. In other words, TM modes in equatorial slab MO waveguides normalize in the same way as in the isotropic case!

2.3.3 Benchmark

Both modelling techniques have a fundamentally different approach. In order to compare their performance in a practical situation, hypothetical simplified layer structure for the integrated optical isolator under study in this work is modelled using both the rigorous 2×2 MO TMF and the first-order perturbation formulae of section 2.3.2.1.

In chapter 1, the basic concept of the novel isolator design has been schematically explained. There it was indicated that experimental success for this device relies heavily on the availibility of sufficient TM material gain in the active region of the layer structure, and that this is only feasible if a tensile-strained MQW active gain region is used. However, for modelling purposes it is first of all important to determine whether the strength of the MO Voigt parameter of the potential candidates for the ferromagnetic metal contact allows the use of first-order perturbation calculations, regardless of the precise layout of the active region. In other words, it can be expected that, if a simplified layout for the isolator with a hypothetical bulk active region allows the use of perturbation modelling, this would also be the case when the active region is replaced by a realistic MQW region with 2 SCH guiding layers, because roughly speaking the same amount of optical confinement would be realized in the combination SCH+MQW as in the bulk layer. Therefore the simplified 4-layer design for the integrated optical isolator depicted in Fig. 2.17, has been studied. The choice for $Co_{90}Fe_{10}$ as a representative ferromagnetic

air	$n^{(0)} = 1$	
$Co_{90}Fe_{10}$ MO layer	$n^{(1)} = 4.35 - i\,4.75$ $Q = 0.0214 + i\,0.0442$ or thus, $g = 1.749 - i\,1.045$	$d^{(1)} = 100$ nm
InP spacer layer	$n^{(2)} = 3.2019$	$d^{(2)} = 250$ nm
InGaAsP active layer	$n^{(3)} = 3.3971 + ik$	$d^{(3)} = 200$ nm
InP substrate	$n^{(4)} = 3.2019$	

Fig. 2.17: Multilayer structure of the benchmark waveguide isolator. Optical and MO constants for the wavelength of 1300 nm are obtained from [45], [46], [47], [48].

metal in this benchmark example is justified, as it is one the strongest MO metals.

The techniques of sections 2.3.2.2 and 2.3.2.1 have been implemented and applied to the simplified structure of Fig. 2.17. Obviously a slab structure has been chosen to serve as benchmark, because the equatorial TMF of section 2.3.2.2 is

only applicable to slab waveguides. This 2×2 matrix formalism has been implemented in a MATHEMATICA environment. The four-sheeted dispersion equation has been conformally mapped onto the variable u as described above. With a choice of either T_{12} or T_{21} as open dispersion equation the bound TM modes are all found within a semi-circle in the u-plane allowing efficient determination of the eigenvalues.
The integrals in (2.118) needed for the first-order perturbation calculation of the non-reciprocal TM phase shift are numerically evaluated using an extended Simpson quadrature formula [49]. As a rule of thumb, it was found that with a nodal point spacing in the MO metal of at least 5 nm the numerical evaluation of the integral converges. This has been implemented in PYTHON. This is precisely the same scripting language that acts as an interface for CAMFR, the in-house slab waveguide modelling tool that is used for the calculation of the unperturbed eigenmode. As a result, flexible data exchange between the unperturbed mode calculations and the integral calculations is obtained. The modal calculations in CAMFR are typically done with a closed dispersion equation. To avoid any influence of the PMC walls the InP substrate was chosen to be 10 μm and the air superstrate to be 1μm.
It is important to underline again that material gain in the waveguide is in both approaches rigorously included by using a positive imaginary part for the core refractive index.

First of all, it needs to be remarked that this slab waveguide is inherently asymmetric. Therefore, the zero-order TM mode is not always cut-on, as is the case for symmetric slab waveguides. For a 4-layer waveguide there is no analytical expression for the cut-off frequency (or cut-off core thickness). But it is obvious that it will increase with decreasing InP spacer thickness (in other words with an increasing degree of asymmetry)[65]. The same will apply for every higher order TM mode. Their cut-on thickness, will get an extra contribution due to the asymmetry. This means that if the waveguide core thickness is limited to the symmetrical cut-on thickness of TM_1, then the waveguide under study here will always be monomodal, regardless of the thickness of the spacer layer. In this case that amounts to a maximal thickness of $d^{(3)} = \lambda/(2\sqrt{3.3971^2 - 3.2019^2} \approx 600$nm. All simulations are limited to this monomodal case.

Fig. 2.18 shows the rigorously calculated non-reciprocal TM modal gain (i.e. $4\pi\Im(n_{\text{eff}}/\lambda[\text{in cm}])$) of the simplified integrated slab waveguide SOA/isolator of Fig. 2.17, with the thickness of the InP spacer $t^{(2)}$ reset to 500 nm, as a function of the internal gain of the bulk active layer. It is striking to see how for a wide range of gain levels the non-reciprocity remains quasi-constant, suggesting that the guiding effect due to the gain itself is negligible with respect to the index-guiding mechanism. Of course this is also a direct consequence of the plasmonic property of TM polarized light at the InP/CoFe interface [51]. This effect strongly enhances the TM field strength at the InP/CoFe interface, apparently partially

[65] This assessment immediately leads to an alternative design for an integrated optical isolator. As the equatorial MO waveguide effect is to induce a non-reciprocal change of the propagation constant, there will be a certain range of core thicknesses for which an asymmmetric equatorial MO waveguide will be in complete cut-off regime in one direction! This has actually been proposed theoretically, but a numerical calculation has proven this approach to be unfeasible [50].

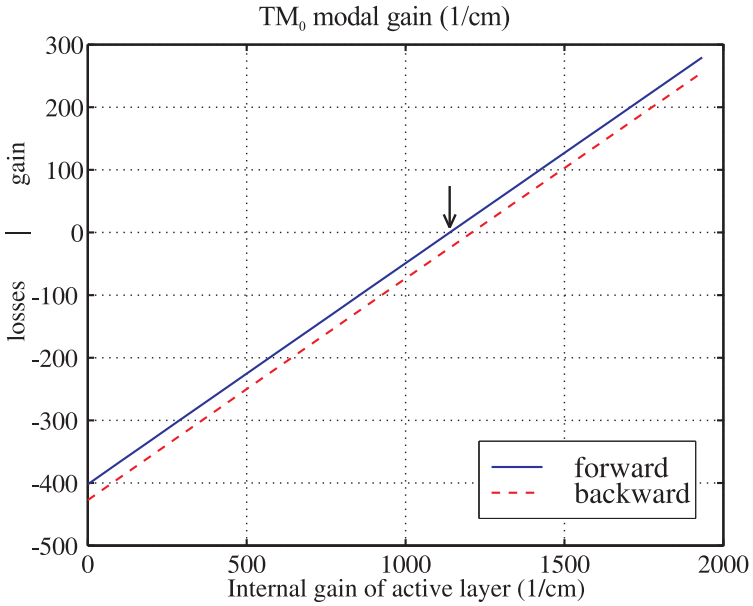


Fig. 2.18: Rigorously calculated TM modal gain as a function of the internal gain of the active layer for the layer structure of Figure 2.17 with the InP spacer thickness set to 500 nm. At an internal gain of about 1100/cm the device behaves transparent in the forward direction while in the backward direction its loss (or in this case, its isolation extinction ratio) amounts to $\approx 25/cm$.

binding the TM mode to the metal, regardless of the gain property of the active layer. Fig. 2.18 also underlines that the rather intuitive and simplistic explanation of the operation principle of this integrated optical isolator, as given in chapter 1, is scientifically sounder than would have been expected at first. Depositing a transversely magnetized ferromagnetic metal contact close to the gain region of a "moderately" pumped active device, induces a non-reciprocal loss difference for the strongly absorbed guided TM mode. In this regime the active device operates as a non-reciprocal absorber. An appropriate forward bias current of the active device "transfers" the loss difference to an operation regime where only the backward propagating TM mode will suffer absorption. The device then truly operates as an isolator. A further increase of the forward bias current finally brings the device into its operation regime of a non-reciprocal SOA.
Another important consequence of this quasi-independence of the MO non-reciprocal effect on the internal gain of the active layer, is that it effectively allows to separate the search for the transparency material gain from the non-reciprocity calculation. This is a very pleasant observation, as the perturbation formalism greatly benefits from this. Indeed, since the equatorial MO TMF treats the waveguide problem *ab initio*, every single simulation point in the curves of Fig. 2.18 requires a full[66]

[66]This is of course slightly exaggerated as it is only the chunk matrices containing the active

recalculation of the MO dispersion equation and mode search. The perturbation formalism could make use of the fact that apparently the non-reciprocal phase shift is independent of the material gain, by first calculating this shift at zero material gain, and subsequently scanning the material gain of the unperturbed waveguide up to the point where the unperturbed loss would be exactly compensated by the non-reciprocal phase shif in the forward direction. The obvious advantage is that this approach only requires one calculation of the perturbation formulae (2.118), and one scan of the propagation constant of an isotropic waveguide mode. This is of course considerably faster than calculating and solving a MO dispersion equation up till transparency. Actually, the rigorous simulations of Fig. 2.18 take about 2 sec [67] for every simulation point (thus about 4 in total, once in the forward direction once in the backward direction). This means that in practice scanning for the transparency point with the rigorous approach easily takes up to a minute. The perturbation calculations on the same machine took less than a second to calculate the overlap integral[68] and only a few seconds to find the transparency point.

These remarks, favouring the perturbation formalism, are naturally only significant if indeed the perturbation calculation predicts the same non-reciprocal index change as the rigorous approach. On top of that it is far from certain that the non-reciprocal effect will always be independent of the internal gain. Actually, a very straightforward example for which this is not the case is when the mode is close to cut-off. In that case gain-guiding rather than index guiding gets the upper hand, and the mode structure is strongly influenced by the amount of gain in the active layer. As a result also the non-reciprocity severely changes as the gain is varied. However, operating the waveguide close to cut-off can hardly be considered an interesting operation regime, as the amount of gain that would be required for transparency would always be unrealistically high. Thus, the main point to be studied is the possible difference between the rigorous calculations and the perturbation calculations.

This is investigated in Fig. 2.19. Here, the absolute strength of the MO Voigt parameter of the CoFe metal film is artificially varied, $Q' = f\, Q_{\mathrm{CoFe}}$ ($f = 1$ corresponding to the actual, experimentally determined Voigt parameter of $Co_{90}Fe_{10}$ [48]). The isolation extinction ratio is calculated both in a rigorous way (as $4\pi n_{\mathrm{eff,b}}/\lambda$) and via the perturbation formula (2.120). This has been done for several values of the InP spacer thickness. The core thickness is kept fixed at 200 nm. Naturally the first-order perturbation calculation has a perfect linear dependence on the MO strength. This first-order behaviour is however also observed in the rigorous TMF calculations. Of course in the origin for $Q \to 0$ both curves are perfectly tangential, but what is really remarkable is that the higher-order effects in the magneto-optic parameter, fundamentally included in the rigorous approach, only start to appear at strong values for Q'. On top of that this divergence is only

layer that are changed. the other chunk matrices of the stack could as well have been cached. However this is of course a software optimization issue. The fact remains that every new simulation is indeed a "new" MO waveguide to be solved in the TMF approach.

[67] on a Pentium III 500Mhz processor

[68] for up to nodal point densities of 1/nm in the quadrature formula.

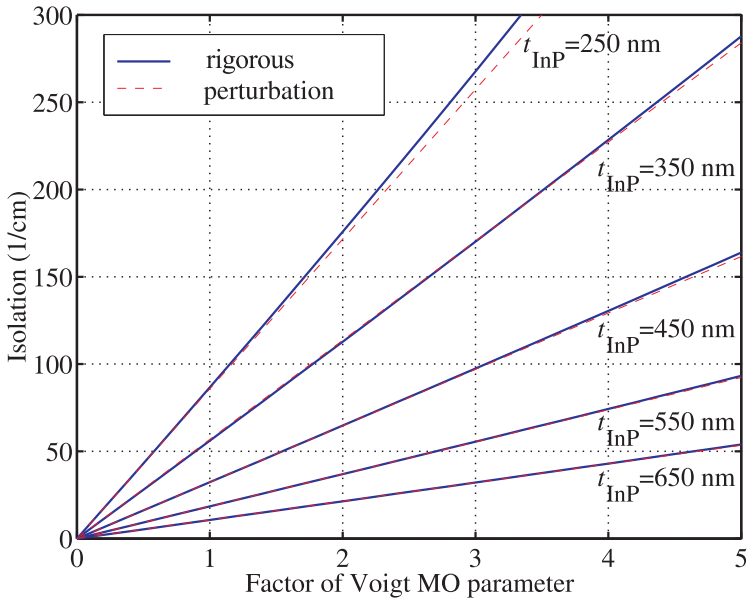


Fig. 2.19: Comparison of perturbation theory and rigorous approach. Waveguide isolation is shown as a function of the artificial enhancement factor of the Voigt magneto-optical parameter $Q' = f\ Q_{\mathrm{CoFe}}$.

notable for small spacer thicknesses (and even then still at hypothetical values that are twice as strong as the actual Q strength). In other words, since the core thickness is kept fixed in this simulations, higher order MO effects will only start playing a role as the mode approaches cut-off. This is not surprising, as this is accompanied by a strong increase of the modal field in the cladding layers. As a result the off-diagonal elements in the coupled-mode matrix, i.e. $\overline{\overline{A}}_{0k}(k \neq 0)$ in (2.84), being the overlap of the mode just above cut-off with the radiation modes, can not be considered small in the perturbation series expansion (2.89). Either the convergence of this series sum will be of first order or they will contribute as a non-negligible second-order contribution.
In any case, Fig. 2.19 makes clear that, except for the above described limit situation, the agreement between rigorous MO TMF modelling and perturbation theory modelling is quasi-perfect. For commonly encountered Voigt parameter strengths and practical device applications, higher order MO effects are negligible, and the perturbation theory formalism is a fast and feasible alternative modelling tool.

Returning now to the simulation of the simplified layout for the integrated optical isolator, consider Figures 2.20 and 2.21. These show the dependency of the waveguide isolation and the necessary internal gain of the active layer on the thickness of the active layer for several values of the InP spacer thickness. Actually, the curves shown here are calculated with both modelling approaches, but their difference is not distinguishable.

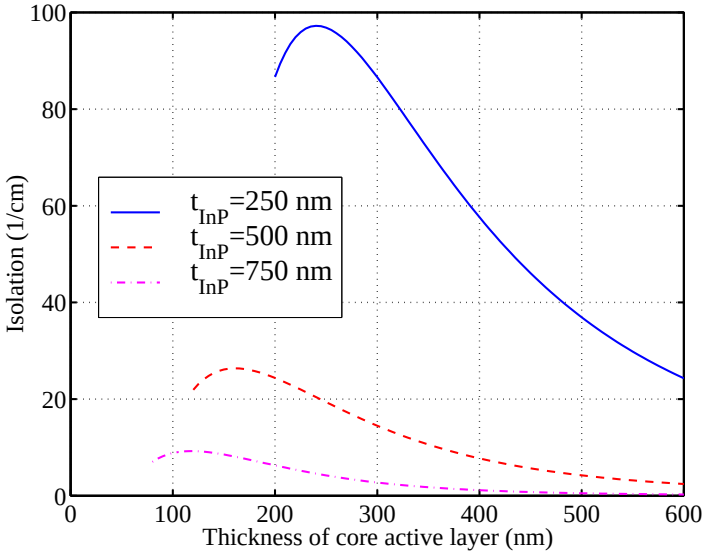


Fig. 2.20: Isolation extinction ratio as a function of core thickness for three InP spacer thicknesses. The metal thickness is as in Fig. 2.17.

Isolation increases for decreasing active layer thickness and reaches a maximum just above cutoff thickness. This cut-off thickness is clearly observed to increase when the InP spacer thickness decreases, or, in other words, when the asymmetry of the waveguide increases. The steep decrease of isolation ratio when the cut-off core thickness is approached, is a result of a strong increase of the modal tail into the substrate when the mode approaches cut-off. As a consequence the relative amount of guided TM light near the MO metal steeply drops. The increase of necessary internal gain of the active layer with decreasing active layer thicknesses does not show this maximum. This is because decreasing the active layer thickness will only result in a smaller overlap with the gain region and hence will always lead to an increasing necessary internal gain. The dependencies for different thicknesses of the InP spacer layer show that both the isolation and the necessary internal gain of the active layer increase for decreasing InP spacer thickness. This is not a surprising result, since a thinner InP spacer thickness allows a stronger TM light confinement near the MO metal. However, due to the asymmetry of the layer structure there will also be a minimum thickness for this InP spacer layer below which the guided mode is cut-off. As a consequence the isolation as a function of InP spacer thickness waveguide will also show a maximum. Consequently, if the maximum achievable internal gain for a fixed active layer thickness is known, the isolation ratio of the device can be optimized by optimizing the spacer layer thickness within the limits of the available gain.

Figure 2.22 shows the effect of the magneto-optic layer thickness. Both the iso-

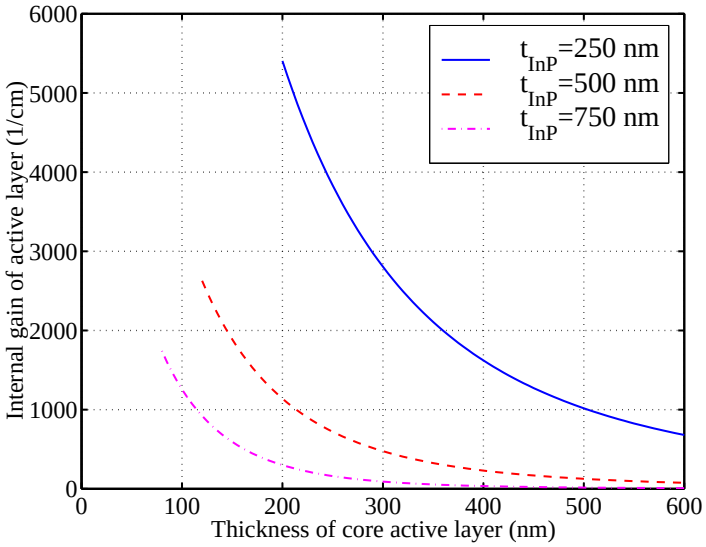


Fig. 2.21: Necessary material gain for forward loss compensation. Same simulation parameters as in Fig. 2.20

lation and necessary gain of active layer increase with increasing $Co_{90}Fe_{10}$ thickness. For high thicknesses the effect is saturated, indicating that the upper interface of the MO metal has become irrelevant. The existence of a maximum and the behavior for very thin MO films originate from effects of phase changes of reflected waves at the interfaces and interference phenomena.

Finally, we would also like to point out, that for this particular device one has to be careful when applying the results of Table 2.6, where the influence of the complex signs of the MO parameter g on the non-reciprocal phase shift are tabulated. The results in this table assume a field integral that is dominantly imaginary. In other words, that the deviation of perfect quadrature between E_x and E_z caused by the presence of losses does not extend considerably into the MO region. For this particular case this is obviously not fulfilled, as the MO region itself is strongly absorbing. As a result the signs of the non-reciprocal MO (real) phase shift and the non-reciprocal absorption shift will be influenced by both the real and imaginary parts of g and of the field integral. For this case, it suffices to mention that the perturbation calculations predict a decrease in optical loss for the forward directionar and are thus in full accordance with the rigorous calculations (see Fig. 2.18), not only as far as magnitude of the effect is concerned but also as far as sign is concerned.

The values for the necessary internal gain, obtained for this simplified isolator layer structure, are for most cases above what is theoretically achievable in bulk active layers. Moreover, in a bulk active layer, there will be no gain discrimination

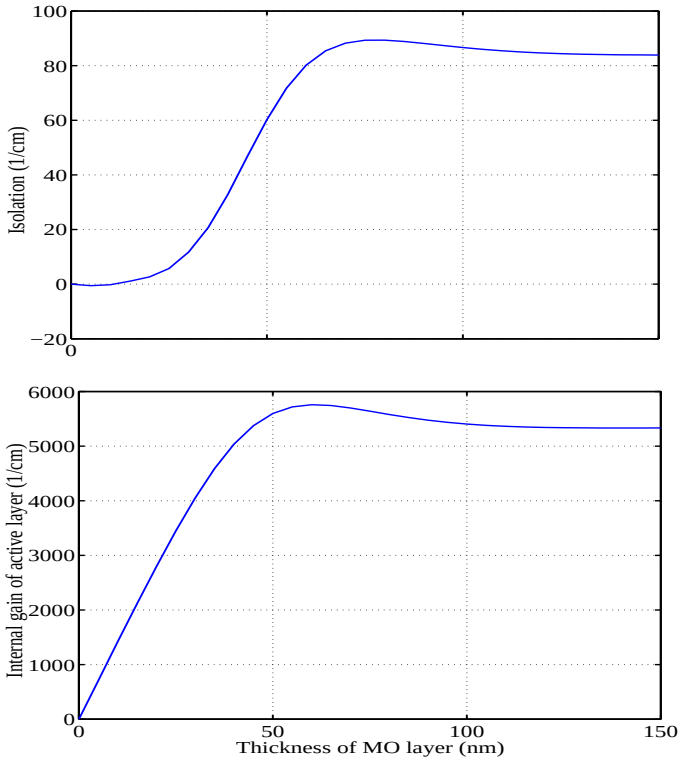


Fig. 2.22: Waveguide isolation and necessary internal gains of the active layer as a function of the MO metal film thickness. Structure of the waveguide is the same as shown in Fig. 2.17.

between TE and TM light, which is needed as pointed out in chapter 1. Therefore, the main goal of this example was to compare the two different modelling approaches and to qualitatively illustrate the influence of the different structural parameters on the behavior of the device. In the chapter 3, it will be numerically shown that this device is indeed feasible if a tensile strained MQW layer structure is used. Even though several other refinements can be included in the isolator model (next to the use of a SCH+MQW active region), the main conclusions will not change. Isolation will not indefinitely increase with decreasing spacer thickness, due to the asymmetry of the waveguide. Possible refinements are the inclusion of optical magnetic anisotropy in the ferromagnetic metal contact, the influence of misalignment of the magnetization, inclusion of an absorbing contact layer and/or magnetic dead layer just below the MO metal, strain-induced optical anisotropy in the gain region, and several carrier-dependent effects on the optical constants of the layers such as anti-guiding, free-carrier absorption,.... Finally, two-dimensional modelling of the device, including the precise layout of the guiding profile in the lateral direction cannot be considered a refinement of

the model. It is indeed obvious that for this particular device the design of the lateral dimension will not add specific non-reciprocal effects that are not already included in the pure transversal modelling. The inclusion of the lateral dimension will only lead to a general decrease of the effect due to a decreased confinement near the MO metal. The lateral design is thus purely a question of finding the right compromise between achieving maximal non-reciprocity (slab waveguide) and obtaining appropriate other 2D waveguide characteristics such as monomodality, surface-recombination (when designing etched rib waveguides), current spreading, ... Note that for other MO non-reciprocal waveguide effects, the lateral design does have an important influence on the modelled non-reciprocity. A Faraday rotator for instance depends heavily on the modal profiles of the interacting modes, and it is clear that the second dimension will play an important role in this.

The benchmark example has made clear that even for the strongest MO materials, the perturbation formalism achieves the same accuracy as the rigorous MO TMF, except in a few limiting cases that are more of an academic than a practical interest. At first sight one would be tempted to give preference to the perturbation formalism. Indeed, it is fast, accurate and not limited to 1D calculations. However, it is not practical for the calculation of modal profiles. Also, for non-reciprocal MO polarization converting waveguides, it is only useful for the derivation of the coupled-mode equations, but not to solve for the non-reciprocal hybrid supermodes. The TMF on the other hand (either the one of section 2.3.2.2 or the general one of section 2.2.2.2) automatically includes all nonlinear higher order effects in the material parameters and always allows the full vectorial calculation of the (super-)modal profiles. As a result it allows to calculate non-reciprocal scattering matrices and other field profile related characteristics. But it is not straightforwardly extended to 2D waveguide cross-sections. When and why to use one or the other modelling formalism is thus by all means a question whose answer depends on the specific needs of the modelling task. Is calculation time an issue? Does one need precise knowledge of the modal profiles? Does one have to calculate scattering matrices with other waveguides sections? Is the precise 2D profile important? ...

2.4 Conclusion

This concludes the extensive chapter on the theoretical modelling of non-reciprocal waveguides. We have given a clear and unambiguous definition of the often confused concepts of reciprocity, bidirectionality and mirroring. An overview was given of the conditions under which a non-reciprocal bianisotropic waveguide can still be bidirectional and/or mirroring. Examples of these have been given throughout this chapter.
The modelling of non-reciprocal waveguides was then restricted to anisotropic waveguides in the optical frequency regime (whose magnetic permeability is that of the vacuum). An in-depth overview has been given of all the subtleties involved when theoretically and numerically solving the eigenproblem of such waveguides.

Theoretical aspects such as completeness of the eigenspectrum, mode orthogonality in non-reciprocal waveguides, and a proper normalization using the adjoint waveguide have been discussed. Two fundamentally different numerical modelling approaches have been elaborated. A 1D rigorous Extended Transfer Matrix Formalism allows the vectorial calculation of the full hybrid modal eigenspectrum of a non-reciprocal waveguide and its adjoint. Sensible use of the bidirectionality and mirroring conditions given in section 2.1 and of any spatial symmetry present in the waveguide problem, often allows a direct derivation of the adjoint and/or backward modal spectrum out of the original forward modal spectrum, herewith seriously relieving the normalization problem and the total computational effort. Illustrations of this are given in the numerical simulation example of a slab waveguide Faraday rotator.
A second modelling approach is the so-called modal expansion technique. It allows a rigorously equivalent reformulation of the Maxwell equations for a non-reciprocal, anisotropic waveguide by expanding the unknown fields in the complete basis formed by the modes a known waveguide. The validity of this expansion is always assured due to the completeness theorem, however the convergence rate of the modal series expansion quickly drops as the difference (the perturbation) between the actual waveguide and the known waveguide increases. There is no a priori criterion available to test for the rate of convergence. The result of the modal expansion technique is a (infinite) set of coupled first-order differential equations for the modal expansion coefficients. Perturbation techniques up to several orders can then be used to calculate the corrections to the initial, unperturbed eigenspectrum. Or the derived set of coupled differential equations can be used for the treatment of the non-reciprocal waveguide in terms of coupled-mode phenomena in the unperturbed waveguide. For the particular case of symmetry-exhibiting magneto-optical waveguides, the non-reciprocal waveguide effects can be elegantly classified by investigation of the elements of the coupling matrix.
Finally, both approaches are specialized to the important case of equatorial magneto-optical waveguides. It was seen that for this class of non-reciprocal waveguides there is no hybridization of the modes. The Maxwell equations still split up into a TE and a TM set. For the slab case this is rigorously correct. For general 2D cross-sections this holds as long as the semivectorial approach holds, just as is the case in isotropic waveguides. This absence of polarization mixing at the layer interfaces seriously simplifies the extended TMF. It was described how a 2×2 isotropic-like TMF is recovered, by deriving the non-reciprocal TM Fresnel reflection and transmission coefficients. The perturbation formulae for this case have been studied in a little more detail, classifying among other things the signs of the expected phase shifts in terms of the MO parameter. An extensive study of a simplified layout for the integrated isolator under study in this work has served as a benchmark for comparison between the two modelling approaches. Next to interesting qualitative insights in the operation principle of this device, the major conclusion of these simulations was of course the apparent perfect equivalence between first-order perturbation theory and rigorous TMF modelling.

References

[1] F. Olyslager, *Electromagnetic Waveguides and Transmission Lines.* Oxford Engineering Science Series, No. 51 (Clarendon Press, Oxford, 1999).

[2] H. A. Lorentz, "The theorem of Poynting concerning the energy in the electromagnetic field and two general propositions concerning the propagation of light." *Verslagen van de Koninklijke Akademie voor Wetenschappen, Amsterdam,* vol. 4, pp. 176-187, 1896.

[3] F. Olyslager, "Properties of and generalized full-wave transmission line models for hybrid (bi)(an)isotropic waveguides." *IEEE Trans. Microwave Theory Tech.,* vol. 44, pp.2064-2075, 1996.

[4] J. Van Bladel, *Electromagnetic Fields.* (Hemisphere, Washington, DC, 1984).

[5] R. W. P. King, "Quasi-stationary and nonstationary currents in electric circuits." In *Encyclopedia of physics,* vol. XVI: *Electric fields and waves* (ed. S. Flügge). (Springer-Verlag, Berlin, 1958).

[6] P. R. McIsaac, "Mode orthogonality in reciprocal and nonreciprocal waveguides." *IEEE Trans. Microwave Theory Tech.,* vol. 39, pp. 1808-1816, 1991.

[7] A. H. Sihvola and S. A. Tretyakov, "Early Studies on Chiral and Bianisotropic Media." Chapter 1 in *Electromagnetic Waves in Chiral and Bi-isotropic Media.* (Artech House, Boston & London, 1994).

[8] H. Verschelde, "Elektromagnetisme en relativiteitstheorie." Chapter 5 in *Relativiteitstheorie en Klassieke Velden.* (in Dutch) Master Course, Ghent University, Ghent, Belgium, 1999.

[9] F. Olyslager and I.V. Lindell, "A Pedigree of Bianisotropic media." *Proc. of 8^{th} International Conference on Electromagnetics of Complex Media,* Lisbon, Portugal, pp. 153-158, 2000.

[10] E. W. Weisstein, "Quartic Equation." From *Mathworld* – A Wolfram Web Resource. http://mathworld.wolfram.com/QuarticEquation.html

[11] B. Jakoby and F. Olyslager, "Asymptotic expansions for Green's dyadics in bianisotropic media." In *Progress in electromagnetic research 12* (ed. J.A. Kong), pp. 277-302, (EMW Publishing, Cambridge, 1996).

[12] P. Bienstman, "Rigorous and efficient modelling of wavelengh scale photonic components." Ph. D. Thesis, Ghent University, Ghent, Belgium, 2001.

[13] B. Baekelandt, "Study of advanced layered and periodic electromagnetic absorbers." Ph. D. Thesis, Ghent University, Ghent, Belgium, 1998.

[14] C. Vassallo, *Optical Waveguide Concepts.* Optical Wave Sciences and Technology Series, No. 1 (Elsevier, Amsterdam, 1992).

[15] H. Verschelde, "Klassieke Velden." Chapter 6 in *Relativiteitstheorie en Klassieke Velden.* (in Dutch) Master Course, Ghent University, Ghent, Belgium, 1999.

[16] L. Onsager, "Reciprocal relations in irreversible processes." *Phys. Rev.,* vol. 37, pp. 405-426, 1931.

[17] K. Heyde, "Rotation in Quantum Mechanics." Chapter 2 in *The Nuclear Shell Model.* (Springer-Verlag, Berlin, 1994).

[18] A. S. Sudbø, "Numerically Stable Formulation of the Transverse Resonance Method for Vector Mode-Field Calculations in Dielectric Waveguides." *IEEE Phot. Tech. Lett.,* vol. 5(3), pp. 342-344, 1993.

[19] C. Vassallo, *Théorie des Guides d'ondes Electromagnetiques (Tome 2).* Collection Technique et Scientifique des Télécommunications, Editions Eyrolles, Paris, 1985.

[20] H. A. Haus, W. P. Huang, S. Kawakami and N. A. Whitaker, "Coupled-Mode Theory of Optical Waveguides." *J. Lightwave Technol.,* vol. LT-5(1), pp. 16-23, 1987.

[21] S. Chuang, "A Coupled Mode Formulation by Reciprocity and a Variational Principle." *J. Lightwave Technol.,* vol. LT-5(1), pp. 5-15, 1987.

[22] C. Vassallo, "About Coupled-Mode Theories for Dielectric Waveguides." *J. Lightwave Technol.,* vol. 6(2), pp. 294-303, 1988.

[23] M. Waroquier, "Storingsrekening." (in Dutch) Part D in *Quantummechanica,* Master Course, Ghent University, Ghent, Belgium, 1997.

[24] J. J. Sakurai, *Modern Quantum Mechanics.* (Addison Wesley, 1993).

[25] M. Lohmeyer, "Guided waves in integrated rectangular magnetooptic devices." Ph. D. Thesis, University of Osnabrück, 1999.

[26] `Fimmwave`, A fully vectorial 2D Mode Solver. http://www.photond.com

[27] P. Bienstman, "CAvity Modelling FRamework." Software code freely available at htttp://camfr.sourceforge.net

[28] P. Hertel, "Theoretische Untersuchungen zum integrierten magnetooptischen Isolator." (in German) Arbeitsbericht des Teilprojekt D10 des Sonderforschungsbereiches 225 der Deutsche Forschungsgemeinschaft, http://www.sfb225.uni-osnabrueck.de

[29] M. Fehndrich, A. Josef, L. Wilkens, J. Kleine-Börger, N. Bahlmann, M. Lohmeyer, P. Hertel and H. Dötsch, "Experimental investigation of the nonreciprocal phase shift of a TE-mode in a magneto-optic rib waveguide." *Appl. Phys. Lett.,* vol.74, pp. 2918-2920, 1999.

[30] K. Postava, M. Vanwolleghem, D. Van Thourhout, R. Baets, Š. Višňovský, P. Beauvillain and J. Pištora, "Modelling of a novel InP-based monolithically integrated magneto-optical waveguide isolator," *J. Opt. Soc. Am. B,* vol. 22(1), pp. 261-273, 2005.

[31] C. Vassallo, "1993–1995 Optical mode solvers." *Opt. Quantum Electron.,* vol. 29(2), pp. 95-114, 1997.

[32] L. D. Landau and E. M. Lifschitz, in *Electrodynamics of Continuous Media (2^{nd} ed.),* pp. 280-283 (Pergamon Press, Oxford, England, 1984).

[33] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light.* (Cambridge University Press, Cambridge, England, 1999).

[34] R. E. Smith and S. N. Houde-Walter, "The Migration of Bound and Leaky Solutions to the Waveguide Dispersion Relation." *J. Lightwave Technol.,* vol. 11(11), pp. 1760-1768, 1993.

[35] E. W. Weisstein, "Riemann Surface." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/RiemannSurface.html

[36] L. Delves and J. Lynnes, "A numerical method for locating the zeros of an analytic function." *Math. Comp.,* vol. 21, pp. 543-560, 1967.

[37] E. W. Weisstein, "Cauchy Integral Formula." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/CauchyIntegralFormula.html

[38] R. E. Smith, S. N. Houde-Walter and G. W. Forbes, "Mode Determination for Planar Waveguides Using the Four-Sheeted Dispersion Relation." *IEEE J. Quantum Electron.,* vol. 28(6), pp. 1520-1526, 1992.

[39] R. E. Smith, G. W. Forbes and S. N. Houde-Walter, "Unfolding the Multivalued Planar Waveguide Dispersion Relation." *IEEE J. Quantum Electron.,* vol. 29(4), pp. 1031-1034, 1993.

[40] H. A. Priestley, *Introduction to Complex Analysis.* (Oxford University Press, Oxford, England, 1985).

[41] J. Haes, "Studie van bundeleigenschappen van laserdiodes en ontwerp van geintegreerde bundelexpansiestrukturen." (in Dutch) Ph. D. Thesis, Ghent University, Ghent, Belgium, 1996.

[42] E. W. Weisstein, "Boundary Conditions." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/BoundaryConditions.html

[43] P. Bienstman and R. Baets "Optical modelling of photonic crystals and VCSELs using eigenmode expansion and perfectly matched layers." *Opt. Quantum Electron.,* vol. 33(4), pp. 327-341, 2001.

[44] H. Derudder, F. Olyslager, D. De Zutter, S. Van den Berghe, "Analysis of waveguide discontinuities using perfectly matched layers." *Electron. Lett.,* vol. 34(22), pp. 2138-2140, 1998.

[45] O. J. Glembocki and H. Piller, in *Handbook of Optical Constants of Solids*, E. D. Palik, ed., Chap. Indium Phospide (InP), pp. 503–516. (Academic Press, 1985).

[46] B. Jensen, in *Handbook of Optical Constants of Solids II*, E. D. Palik, ed., Chap. 6. Calculation of the refractive index of compound semiconductors below the band gap, pp. 125–149. (Academic Press, 1991).

[47] M. Muñoz, T. M. Holden, F. H. Pollak, M. Kahn, D. Ritter, L. Kronik, and G. M. Cohen, "Optical constants of $In_{0.53}Ga_{0.47}As$/InP: Experiment and modelling," *J. Appl. Phys.,* vol. 92, pp. 5878–5885, 2002.

[48] M. Compin, B. Bartenlian, P. Beauvillain, P. Gogol, J. Hamrle, L. Lagae, J. Pistora, K. Postava, S. Visnovsky, and R. Wirix-Speetjens, "Détermination des indices optiques et magnéto-optiques de films minces de CoFe constituant un isolateur intégré à 1.3 μm," In *Colloque Luis Neél CMNM 2004*, (France, 2004).

[49] E. Isaacson, and H. B. Keller, Chapter 7 "Numerical Integration" in *Analysis of Numerical Methods,* (Dover Publications, New York, USA, 1994).

[50] H. Hemme, H. Dötsch and P. Hertel, "Integrated optical isolator based on nonreciprocal-mode cut-off." *Appl. Opt.,* vol. 29(18), pp. 2741-2744, 1990.

[51] H. Raether, *Surface Plasmons – on smooth and rough surfaces and on gratings.* Springer Tracts in Modern Physics, vol. 111 (Springer-Verlag, Berlin, Germany, 1988).

3

Design of the monolithically integrated non-reciprocal Semiconductor Optical Amplifier/isolator

IN this chapter the two key aspects determining the experimental success of the novel type of integrated optical isolator, will be discussed. In Chapter 1 it was explained how the operation principle of the novel integrated isolator design relies heavily on the combined availability of a "well-behaved" active amplifying region and a "well-behaved" ferromagnetic magneto-optic contact with good interfacial properties towards traditional III-V semiconductors. The subtleties of these "well-behaved" characteristics belong to previously little explored, electric, opto-electronic and magneto-optic research areas. These will be studied in this chapter, covering topics ranging from the magnetic anisotropy of magnetic metallic stripe contacts to the optical gain engineering through strain incorporation in quantum wells. Here, we will mainly focus on the theoretical side of these topics. Experimental characterization results for each of these key aspects will be discussed in the next chapter.

Next to the optimization of the building blocks of the device, its performance obviously depends on a proper design of the layer structure combining the gain region and the magneto-optic contact into one transversal guiding structure. This will also be studied in this chapter.

3.1 Ferromagnetic metal/semiconductor interface

It is obvious that a non-reciprocal device first of all needs a sufficiently strong magnetooptic layer. In the here studied concept there are however additional requirements. As the magneto-optic layer must at the same time act as an electrical contact for the underlying SOA, it must also be a good Ohmic contact for the underlying SOA. Finally, as an integrated device, this novel type of isolator must be able to function properly without the need for an external permanent magnet. The metal magnetic contact must in other words behave as a good permanent magnet itself. These electric, magnetic and magneto-optic properties will be studied in this section.

3.1.1 Generalized magneto-optic ellipsometry

Introduction In principle a magneto-optic material is phenomenologically characterized by three polar tensors defined by the crystallographic class of the material, ϵ^0_{ik} (the zero-order permittivity tensor including optical anisotropy), a_{lq} (the gyrotropic tensor defining the gyration vector g_l) and δ_{iklm} (the fourth-order tensor defining the quadratic magnetic optical anisotropy) [1]. For our purposes, it is sufficient to consider the MO material as an optical isotropic material with negligible magnetic optical anisotropy and gyrotropic anisotropy (see section 1.2.2.1). This simplest case can be described by the traditional complex isotropic dielectric constant ϵ and a single complex magneto-optic constant Q, also known as the Voigt parameter. Using the sign convention of Atkinson [2], Q is related to the magnitude of the gyration vector $|\mathbf{g}|$ [1] by $Q = g/\epsilon$. The (magneto-)optical characterization of such a MO material gets down to the measurement of two complex material parameters.

Applied magneto-optics has been around for a long time. Historically speaking its main activity originally has been magneto-optical recording. The first experimental successes date from the late 60's through the early 70's [3], [4]. The obvious spectral region of interest for magneto-optic memory is the blue-side of the visible spectrum. The advent of low-loss optical fibres and the birth of optical telecom in about the same era (mid-70's) has caused research in integrated magneto-optics to be not much younger than its big brother, magneto-optic memory. However, not only has integrated magneto-optics up till now remained mainly in academic milieus, it has also been focussed on the use of MO materials which are optically quasi-transparent in the telecom wavelength region. As a result, while most magneto-optic materials have been extensively characterized in the visible spectrum (for recording purposes), the same effort has been done for only a very few materials in the near infrared. Ferrimagnetic garnets are the main example. Little is known about the magneto-optical constant of, for instance, ferromagnetic metals in the near-infrared, except for the excellent work of a single, old Russian

[1] recall that for a MO material without gyrotropic anisotropy a gyration vector along the z-axis means a magnetization along the z-axis and $|\mathbf{g}| = |\epsilon_{xy}|$.

academic research initiative [5], [6]. For our application, it is clear that a better characterization of ferromagnetic magneto-optic metals and their alloys in the wavelength range of 1300-1600 nm is needed. Just as for the optical characterization of ordinary isotropic materials, the basic technique to achieve this is through ellipsometry.

The idea behind ellipsometry is to measure the optical properties of a reflective stack of materials through the measurement of the change of the polarization state undergone at reflection off this stack by an incident light beam of a known specific polarization, and with a known wavelength and incidence angle. It is one of the most sensitive and reliable techniques for this purpose. As it basically achieves an accurate determination of the normalized Jones reflection matrix of the optical system under study, ellipsometry allows, when combined with an optical model (and sufficient data points[2]), to extract through numerical fitting the optical constants (and thickness) of any layer in a stack of materials. Ellipsometry is a relative measurement technique. As such the components of the Jones matrix of a reflecting surface can only be determined up to a global complex factor. Viewed in this way, ellipsometry only allows the determination of three complex ratios (r_{pp}/r_{ss}, r_{sp}/r_{ss} and r_{ps}/r_{ss}), instead of four absolute complex coefficients. However, for optical parameter extraction this is sufficient. This is because the optical model of an ideal interface (i.e. non-depolarizing and perfectly plane) used for fitting, can also be expressed in a normalized way[3]. Two generalized magneto-optic ellipsometric techniques will be presented here. Combined with a numerical magneto-optic reflection tool as the one developed by Višňovský or Mansuripur [7, 8], these techniques allow precise determination of the optical and magneto-optic constants. Experimental results obtained with these techniques will be presented in Chapter 4.

Ellipsometry – the basic set-up Fig. 3.1 sketches the basic setup of ellipsometry. As with any polarization analysing experiment three optical polarization manipulating components are needed. The role of the polariser and analyser is obvious. The compensator is an optical retarder. It is needed to achieve full control over the incident polarization. Indeed it is not so hard to prove, using Jones calculus, that the polarization output after the polarizer and the compensator is given by (see Fig. 3.1 for the definitions of the different azimuth angles)

$$\begin{aligned}\begin{pmatrix} E_s \\ E_p \end{pmatrix} &= \begin{pmatrix} \cos\theta_c & -e^{-j\delta}\sin\theta_c \\ \sin\theta_c & e^{-j\delta}\cos\theta_c \end{pmatrix} \begin{pmatrix} \cos(\theta_c-\theta_p) \\ -\sin(\theta_c-\theta_p) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta_c-\theta_p)\cos\theta_c + e^{-j\delta}\sin(\theta_c-\theta_p)\sin\theta_c \\ \cos(\theta_c-\theta_p)\sin\theta_c - e^{-j\delta}\sin(\theta_c-\theta_p)\cos\theta_c \end{pmatrix}\end{aligned} \tag{3.1}$$

[2] for instance variable angles of incidence.

[3] This is because in an idealized optical model it is assumed that there are no other polarization dependent reflection effects than those contained within the Fresnel coefficients.

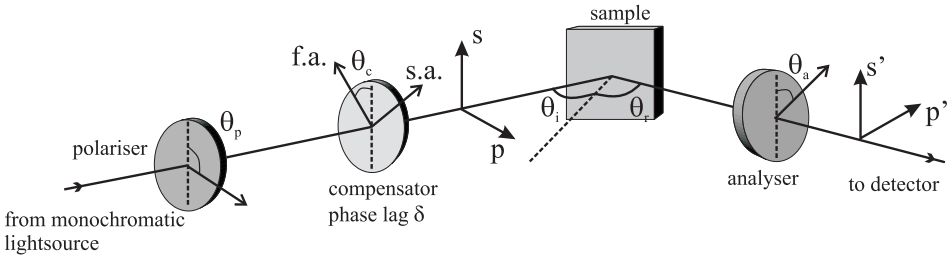


Fig. 3.1: The basic set-up of any ellipsometry experiment. It is important to note the orientation of the s-p axes in the incident and reflected light path. It is conventional to choose these in accordance with the fixed laboratory frame, as the Jones reflection matrix is almost always defined in this frame. It is important to pay attention to this. The use of the "wrong" combination of Jones coefficients and s-p frame will obviously cause polarization rotation and ellipticities of the wrong sign, causing for instance a circular polarized wave to change its handedness upon reflection off an isotropic sample. The optical compensator is nothing else but a retarder, introducing a phase lag δ for the polarization component parallel to its slow axis. The optical elements in this set-up are all assumed to be ideal. The azimuths of all optical components are measured with respect to the s-axis and are positive when rotated towards to the positive p-axis.

The complex polarization variable χ, thus becomes

$$\chi = \frac{E_p}{E_s} = \frac{\tan\theta_c - \cos\delta\tan(\theta_c-\theta_p) + j\sin\delta\tan(\theta_c-\theta_p)}{1+\cos\delta\tan\theta_c\tan(\theta_c-\theta_p) - j\sin\delta\tan\theta_c\tan(\theta_c-\theta_p)}, \tag{3.2}$$

which leads, with the formulae mentioned in a footnote in section 1.2.3 and after some straightforward trigonometry, to the following azimuth and ellipticity,

$$\tan 2\theta = 2\frac{\tan\theta_c(1-\tan^2(\theta_c-\theta_p)) - \cos\delta\tan(\theta_c-\theta_p)(1-\tan^2\theta_c)}{(1-\tan^2(\theta_c-\theta_p))(1-\tan^2\theta_c) + 4\cos\delta\tan\theta_c\tan(\theta_c-\theta_p)} \tag{3.3}$$

$$\sin 2\epsilon = 2\frac{\sin\delta\tan(\theta_c-\theta_p)}{1+\tan^2(\theta_c-\theta_p)} \tag{3.4}$$

(3.4) and (3.3) prove that the combination of a polarizer and a compensator with an arbitrary phase lag δ can produce any elliptic polarization, much like the combination of a half-wave and a quarter-wave plate[4]. Actually, the compensator most often encountered in ellipsometry experiments is one with $\delta = \pi/2$, i.e. a quarter-wave plate. On top of that it is clear from (3.4) and (3.3) that with a fixed polarizer azimuth setting all polarization states can be obtained by only changing the azimuth of the compensator, θ_c.

The phase lag can be constant (with any value up to 2π) or time-dependent. The two techniques described below will make use of a compensator with a harmonic time-varying phase lag. As a final note about the action of the compensator,

[4]with that important difference that the latter combination will in the ideal case not cause power loss, as opposed to the polarizer, who will in the worst case block all incoming power.

it has to be mentioned that for any ellipsometric experiment it makes no difference whether the compensator is placed between the polarizer and the sample or between the sample and the analyzer. In the former case one achieves complete control over the incoming polarizations, while in the latter case one can transform any reflected polarization into any other at the detector.
The sample in the experiment can be any kind of reflecting (mono- or multilayer) stack consisting of any material. Therefore, rather than making assumptions about this reflective stack, we will treat it in the most general way by assuming a full (not necessarily symmetric or antisymmetric) 2×2 Jones matrix. Finally, the idealness of the optical elements implies the following properties. An ideal polariser and analyser will only transmit the polarisation component parallel with their optical axis without changing the amplitude of this component. An ideal compensator will only introduce a phase lag for the component along the slow axis without power loss. And the idealness of the sample implies a perfectly flat interface with the ambient medium.

Ellipsometry – principles In ellipsometry there are basically two classes of techniques: *null* ellipsometry or *photometric* ellipsometry [9]. On top of that photometric ellipsometry can be performed in a static regime and a dynamic regime. In order to understand their difference it is important to recapitulate a few basic concepts about ellipsometry. The basic idea behind ellipsometry is to precisely measure a polarization state of light that undergoes a polarization change by the system under study. This system is characterized by a 2×2 Jones matrix. As a result it has two eigenvalues and two eigenpolarizations. The eigenpolarizations of the system are those input polarizations that retain their polarization state upon interaction with the system under study. Their corresponding eigenvalues simply express by what global[5] complex factor their Jones vector is multiplied. If one can determine the eigenvalues and eigenpolarizations of the system (in other words, diagonalize the Jones matrix) then this system is completely characterized and all its observable physical information (optical thicknesses, optical parameters, ...) is extractable. However, this is a vicious circle, as the diagonalization of the Jones matrix can only be performed if the Jones matrix is known, which in turn would mean that the system was already characterized[6]! In general the eigenpolarizations (let alone the eigenvalues) are not known beforehand, and on top of that they can be dependent on the angle of incidence. In this case, one first has to determine the unknown Jones matrix.
In Jones calculus it is proven that the function describing polarization change

[5] As it is a global factor in the Jones vector, it will obviously not influence the polarization state.

[6] There is one important exception to this general remark. If it is known by physical intuition what the eigenpolarizations should look like, then it is only a matter of determining the eigenvalues. This is much simpler in this case, as it is already known in what reference frame the Jones matrix becomes diagonal. Only the values of those diagonal entries must be determined. This is for instance the case for all isotropic systems, for which it is known that the s- and p-polarizations, retain their character. A more exotic example is given by the polar Kerr effect at normal incidence. As was seen in chapter 1, the eigenpolarizations in this case are the left and right circular polarized waves. Note also that this is only true at perfect normal incidence!

through interaction with a general optical system is always a complex bilinear conformal one-to-one mapping constructed with the four elements of the Jones matrix, $\overline{\overline{J}}$. Namely,

$$\chi_{\text{out}} = \frac{J_{22}\chi_{\text{in}} + J_{21}}{J_{12}\chi_{\text{in}} + J_{11}} \tag{3.5}$$

It is known from mathematics that every conformal bilinear mapping has an unique representation (up to a global complex multiplier) if its response to three different inputs is known [10]. In other words, if the output polarization corresponding to three known input polarizations can be measured, then the Jones matrix is characterized. The most pragmatical technique of measuring the output polarization is by searching for a combined setting of θ_p, θ_c and θ_a which nullifies the detected light power[7]. It is clear that there will be certain conditions under which this nullifying is achievable and if so, how many nulls can be found. This is extensively discussed in [9], where it is proven that it suffices, certain exceptions notwithstanding, to vary the azimuths of two of the three optical components. This is the basis of *null* ellipsometry. All information about the system under measurement is contained in those values of the azimuthal settings (θ_p, θ_c and θ_a) of the optical elements (and in the value of the phase lag δ and the angle of incidence) that reduce the detected light flux to zero. This is because the θ_p and θ_c (and δ) settings give χ_{in} according to (3.4) and (3.3). And χ_{out} is known to be linearly polarised with an azimuth $\theta_{\text{out}} = \theta_a \pm \pi/2$, because of the nulling of the light flux.
In *photometric* ellipsometry, one proceeds in a different way. Instead of collecting the raw data by looking for nulling settings, the information about the optical system is extracted from the *variation* of the light flux as a function of the *variation* of one or more of the above parameters (azimuth angle of optical elements, phase lag, ...). This variation can be static, i.e. for instance consecutively setting the azimuths at the different values and measuring the changes in the light flux. But most often the settings are dynamically varied. And the variations of the light flux are then analysed through Fourier analysis. A very-well known example of this dynamic photometric ellipsometry and of the first in its kind is the Rotating Analyser Ellipsometer (RAE) technique [11]. Nowadays, photometric ellipsometry and dynamic photometric ellipsometry in particular is the most widespread technique. There are two very simple reasons. First of all, present day advanced lock-in amplifiers allow harmonic analysis of modulated signals with great accuracy and with less sensitivity to all sources of noise. Secondly, and more importantly, there is a multitude of dynamic photometric ellipsometry schemes conceivable which do not require any mechanical moving optical elements for the data analysis. In these schemes the azimuth angles of the optical elements remain fixed during the experiment. This is a major advantage over null ellipsometry, where the accuracy of the measurements greatly depends on the accuracy (and the calibration) of the read-out of the azimuth settings of the optical elements during the search for their nulling settings. Next to that, these techniques will also be considerably faster, as no mechanical manipulation is needed to perform the measurement.

[7] actually the phase lag of the compensator is also a degree of freedom, but this is never used because it is very wavelength dependent and difficult to control with sufficient accuracy

In the next two sections, a static and a dynamic photometric technique, custom-conceived for the optical characterization of magneto-optical interfaces, will be elaborated. The actual experimental measurement setup is basically that of Fig. 3.1 with a few technique-dependent instrumentation details, that will be indicated in the next paragraphs. Anyhow, the two techniques have in common that they both make use of the longitudinal magneto-optic Kerr effect (MOKE) to probe for the magneto-optical parameter. The sample is placed between the pole shoes of an electromagnet, and the magnet oriented in such a way that its uniform field will be parallel with both the sample interface and the plane of incidence. As explained in Chapter 1 the Kerr effect manifests itself through the occurrence of non-diagonal elements in the Jones reflection matrix of the MO system under study. It was noted there that these off-diagonal elements are an odd function of the MO Voigt parameter. In the case of a single MO interface (a semi-infinite MO material or an optically opaque MO film), they have equal magnitude but opposite sign [12]. If the MO layer is included in a multi-layer reflective stack, their mutual symmetry property is lost, but they remain off course odd in Q. As the main goal of ellipsometry is to measure the normalized Jones reflection matrix coefficients, r_{pp}/r_{ss}, r_{sp}/r_{ss} and r_{ps}/r_{ss}, this is in this case equivalent to measuring the standard isotropic rotation and ellipticity (described by r_{pp}/r_{ss}), and the s- and p-MOKE rotation and ellipticity (described by r_{ps}/r_{ss} and r_{sp}/r_{ss}). Remark that this requires a minimum of three different incident polarizations, in accordance with the remarks made above.

3.1.1.1 Dynamic photometric magneto-optic ellipsometry in a PSMA configuration

Dynamic photometric ellipsometry – the modulator This dynamic photometric technique has the following specific experimental details. The compensator is placed in the reflected light path. This does not influence the capabilities of the measurement technique, as was explained above. All optical elements have a fixed azimuth setting, their value depending on the precise quantity to be measured. This leaves us only with the phase lag of the compensator as a degree of freedom[8] for the generation of data. This phase lag is harmonically sinusoidally modulated at an angular frequency ω. The compensator in this experiment thus acts as a polarization modulator (hence the M in PSMA).
There are several techniques available to achieve this phase modulation. The basic idea is of course to modulate the physical quantity that is responsible for creating the phase lag in the first place. The optical phase difference between the two eigenpolarizations of the compensator is the result of an optical anisotropy. In order to be modulative this optical anisotropy must be induced by an external "source" that can be modulated itself, such as a quasi-static[9] electric field (the *Pockels* or *linear electro-optic effect*) or a stress tensor (the *piezo-electric effect*).

[8] Of course the angle of incidence is also a degree of freedom. But this latter will only be used to generate the data points needed for nonlinear least-squares numerical fitting of the optical model of the system under study.

[9] with respect to the frequency of the propagating light wave

The use of the Pockels effect is wide-spread for the realization of all kinds of optical phase and amplitude modulators. However, it is only present in materials that have a non-centrosymmetric structure [13]. As such, it is inevitably only employed in a crystalline material, and this will by itself already posses a certain natural birefringence. This is an sich not a drawback, but achieving precise control over the modulated phase difference will require precise knowledge of the crystallographic planes along which the compensator material is cut. On top of that the Pockels coefficients are for most materials not that large, resulting in rather high voltages needed to generate reasonable phase modulation (in the order of kV)[10] [14].
The situation becomes different when the optical anisotropy is induced by a modulated elastic deformation in the compensator material, in other words a strain or acoustic wave. Without going into too deep detail, it can be understood by physical intuition that the harmonic density variations created by such a wave generate harmonic changes of the refractive index in the direction of the material "displacement". This can be understood by the simple intuitive logic that links the strain with an elongation or a compression of the electric dipoles, which in turn causes the electric polarization and thus the dielectric susceptibility to vary. Note that the argument can also be inverted, and that a modulated electric field can cause the generation of an acoustic wave. The major advantage of the piezo-electric effect (also known as the photo-elastic or -acoustic effect[11]) is that it can be generated in almost any solid material, in particular in pure isotropic amorphous materials, which will not exhibit natural anisotropy. As a result the induced phase changes will only be caused by the photo-elastic effect itself and not by some natural material birefringence. This allows a better control of the modulated phase lag. Fused quartz is a common PEM material. Next to that, the power consumptions of a PEM is considerably smaller than that of a Pockels cell. The actual instrumentation details are of course slightly more complicated than this. The generation of the acoustic waves requires more than just placing 2 acoustic frequency electrodes on a fused quartz block. This would only cause the acoustic wave to be very strongly damped. Instead a crystalline transducer material (such as crystalline quartz) has to be used which can be operated at its eigenfrequency and has a high Q factor. More instrumentation operation details of a PEM can be found in the original paper of Jasperson [15], who was the first to suggest the use of the photo-elastic effect for ellipsometric purposes. An in-depth review of the theoretical mechanisms of the photo-elastic effect requires a combination of tensor calculus, elasticity theory, and electromagnetic theory. This is beyond the scope of this text. The reader is referred to [16] for an excellent treatment of the theory of a PEM. For our purposes it is sufficient to take the pragmatical approach and treat the PEM as an optical element described in its natural frame ($x \parallel$ F.A. and $y \parallel$ S.A.) by the following Jones matrix,

$$\overline{\overline{J}}_m^{fs} = \begin{pmatrix} e^{j\delta(t)/2} & 0 \\ 0 & e^{-j\delta(t)/2} \end{pmatrix} \qquad \text{with } \delta(t) = \delta_0 \sin \omega t \tag{3.6}$$

[10] This can be decreased by using a transversely applied electric field.
[11] because its nature is basically characterized by an interaction between sound (material vibrations) and light (via the refractive index).

Magneto-optic ellipsometry – Jones matrices Returning to the magneto-optic ellipsometric experiment, for the evolution of the Jones vector through the optical sequence PSMA, the following Jones matrices are needed (remember that all azimuth angles are referenced to the locally valid *sp* frame – i.e. in the incoming or reflected path),

$$\overline{\overline{J}}_p = \begin{pmatrix} \cos^2\theta_p & \cos\theta_p \sin\theta_p \\ \cos\theta_p \sin\theta_p & \sin^2\theta_p \end{pmatrix} \tag{3.7a}$$

$$\overline{\overline{J}}_s = \begin{pmatrix} r_{ss} & r_{sp} \\ r_{ps} & r_{pp} \end{pmatrix} \tag{3.7b}$$

$$\begin{aligned} \overline{\overline{J}}_m &= \begin{pmatrix} \cos\theta_m & -\sin\theta_m \\ \sin\theta_m & \cos\theta_m \end{pmatrix} \begin{pmatrix} e^{j\delta(t)/2} & 0 \\ 0 & e^{-j\delta(t)/2} \end{pmatrix} \begin{pmatrix} \cos\theta_m & \sin\theta_m \\ -\sin\theta_m & \cos\theta_m \end{pmatrix} \\ &= \begin{pmatrix} \cos\delta/2 + j\sin\delta/2\cos 2\theta_m & j\sin\delta/2\sin 2\theta_m \\ j\sin\delta/2\sin 2\theta_m & cos\delta/2 - j\sin\delta/2\cos 2\theta_m \end{pmatrix} \end{aligned} \tag{3.7c}$$

$$\overline{\overline{J}}_a = \begin{pmatrix} \cos^2\theta_a & \cos\theta_a \sin\theta_a \\ \cos\theta_a \sin\theta_a & \sin^2\theta_a \end{pmatrix} \tag{3.7d}$$

It is not explicitly indicated, but it has to be noted that the Jones matrices of the polarizer, PEM and analyzer can have a non-negligible wavelength dependence. As they are essentially transmission components their operation range will crucially depend on their optical thickness. This is in particular the case for the PEM. The technical spec's of the PEM will tabulate the calibration tables of δ_0 as a function of λ and the applied voltage. Most polarizers and analysers have a quasi-flat wavelength transmission over several 100 nm's. For later use, it is also interesting to derive the Jones matrix of the combination PEM–analyzer with the latter at an azimuth of $\pm\pi/4$, and the whole assembly rotated over an angle ξ.

$$\begin{aligned} &\begin{pmatrix} \cos\xi & -\sin\xi \\ \sin\xi & \cos\xi \end{pmatrix} \overline{\overline{J}}_a(\theta_a = \pm\pi/4) \overline{\overline{J}}_m \begin{pmatrix} \cos\xi & \sin\xi \\ -\sin\xi & \cos\xi \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \cos\xi & -\sin\xi \\ \sin\xi & \cos\xi \end{pmatrix} \begin{pmatrix} e^{j\delta(t)/2} & \pm e^{-j\delta(t)/2} \\ \pm e^{j\delta(t)/2} & e^{-j\delta(t)/2} \end{pmatrix} \begin{pmatrix} \cos\xi & \sin\xi \\ -\sin\xi & \cos\xi \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \cos\delta/2[1 \mp \sin 2\xi] + j\sin\delta/2\cos 2\xi & \pm\cos\delta/2\cos 2\xi \mp j\sin\delta/2[1 \mp \sin 2\xi] \\ \pm\cos\delta/2\cos 2\xi \pm j\sin\delta/2[1 \pm \sin 2\xi] & \cos\delta/2[1 \pm \sin 2\xi] - j\sin\delta/2\cos 2\xi \end{pmatrix} \end{aligned} \tag{3.8}$$

Isotropic versus magneto-optic complex polarization rotation In order to determine the three normalized Jones ratios (r_{pp}/r_{ss}, r_{sp}/r_{ss} and r_{ps}/r_{ss}) in this experiment, we first need to conceive a way to separate the standard isotropic polarization rotation and elipticity from the MOKE rotations and ellipticity. Indeed, in order to observe the effects of the isotropic rotation and ellipticity, one inevitably has to work with an incident linear polarization which has both s- and p-components. This will then always be accompanied by MOKE effects due to the off-diagonal elements in $\overline{\overline{J}}_s$. In order to achieve this separation, one can advantageously make use of the oddness of the MOKE. The following reasoning will prove this.

Assume that the polarizer is set at a very small angle ($\theta_{\text{in}} = \theta_p$ order a few °), the incoming polarization variable is then given by $\chi_{\text{in}} = \tan\theta_p \approx \theta_p$, and the polarization variable after reflection off the sample is

$$\chi_{\text{out}} = \frac{r_{pp}\chi_{\text{in}} + r_{ps}}{r_{sp}\chi_{\text{in}} + r_{ss}} \approx \frac{r_{pp}}{r_{ss}}\theta_p + \frac{r_{ps}}{r_{ss}}, \tag{3.9}$$

where $r_{sp}\theta_p$ has been neglected as it is of second order ($r_{sp} \sim Q$ and is thus of first order).
(3.9) shows that, under the condition of an incoming linear polarization with a small p-component, the polarization at the output of the sample is the sum of an isotropic and a MOKE contribution. Both terms in (3.9) are small first order components. As a result, $|\chi_{\text{out}}|^2$ is of second order, and the output polarization azimuth and ellipticity are according to the formulae mentioned in a footnote in section 1.2.3, given by

$$\tan 2\theta_{\text{out}} = 2\Re\left(\frac{r_{pp}}{r_{ss}}\right)\theta_p + 2\Re\left(\frac{r_{ps}}{r_{ss}}\right) \tag{3.10a}$$

$$\sin 2\epsilon_{\text{out}} = 2\Im\left(\frac{r_{pp}}{r_{ss}}\right)\theta_p + 2\Im\left(\frac{r_{ps}}{r_{ss}}\right), \tag{3.10b}$$

or, due to the first order character of the terms in the above formulae,

$$\theta_{\text{out}} = \Re\left(\frac{r_{pp}}{r_{ss}}\right)\theta_p + \Re\left(\frac{r_{ps}}{r_{ss}}\right) \tag{3.10c}$$

$$\epsilon_{\text{out}} = \Im\left(\frac{r_{pp}}{r_{ss}}\right)\theta_p + \Im\left(\frac{r_{ps}}{r_{ss}}\right) \tag{3.10d}$$

Both contributions to the azimuth and the ellipticity have an odd behaviour. The isotropic contribution is odd with respect to the polarizer angle, while the MOKE term is of course odd in Q, which is in turn an odd function of the magnetization in the sample. This suggests an elegant method to separate both contributions. Measuring the average light intensity under magnetization reversal in the sample and at a fixed polarizer setting, will yield the isotropic contribution, while the average light intensity with a mirrored polarizer setting and a fixed magnetization will produce the MOKE contribution. Note that this separation trick only works because of the first order character of the contribution. Only then it is possible to express the total azimuth rotation and ellipticity as a sum of two terms. It can also be proven that the trick works independently of the setting of the precise setting of the modulator and the analyser in the reflected light path. Again this is a consequence of the first order smallness of both contributions.

The above qualitative reasoning is confirmed by a quantitative calculation using the Jones matrices of the optical sequence. With the modulator azimuth set at 0°, the polarizer at a small angle $\theta_p (\approx \sin\theta_p \approx \tan\theta_p)$, and the analyzer at an arbitrary angle θ_a, the Jones vector at the detector is given by (up to a global

complex factor)

$$\begin{aligned}\begin{pmatrix}E_s\\E_p\end{pmatrix} &= \overline{\overline{J}}_a(\theta_a)\cdot\overline{\overline{J}}_m(\theta_m=0^\circ)\cdot\overline{\overline{J}}_s\cdot\begin{pmatrix}1\\\theta_p\end{pmatrix}\\ &= \begin{pmatrix}\cos\theta_a(\cos\theta_a e^{j\delta(t)/2}r_{ss}+\sin\theta_a e^{-j\delta(t)/2}(r_{ps}+r_{pp}\theta_p))\\ \sin\theta_a(\cos\theta_a e^{j\delta(t)/2}r_{ss}+\sin\theta_a e^{-j\delta(t)/2}(r_{ps}+r_{pp}\theta_p))\end{pmatrix}\end{aligned} \tag{3.11}$$

As a result the detected light intensity becomes[12]

$$\begin{aligned}I(t) &\sim |E_s|^2+|E_p|^2\\ &\sim \cos^2\theta_a r_{ss}r_{ss}^*+\\ &\quad \sin 2\theta_a\Big(\underbrace{\theta_p\big(\cos\delta(t)\Re(r_{ss}^*r_{pp})+\sin\delta(t)\Im(r_{ss}^*r_{pp})\big)}_{\text{isotropic rotation:odd in }\theta_p}+\\ &\quad \underbrace{\cos\delta(t)\Re(r_{ss}^*r_{ps})+\sin\delta(t)\Im(r_{ss}^*r_{ps})\big)}_{\text{MO rotation:odd in }Q}\end{aligned} \tag{3.12}$$

This expression for the detected light intensity proves that both contributions to the complex polarization rotation can indeed be separated using the above described technique. Detecting the average intensity $I_{\text{iso}}(t) = 1/2(I(t;\mathbf{M}\uparrow) + I(t;\mathbf{M}\downarrow))$ will then be independent of MOKE contributions.

$$I_{\text{iso}}(t)\sim\cos^2\theta_a r_{ss}r_{ss}^*+\theta_p\sin 2\theta_a\big(\cos\delta(t)\Re(r_{ss}^*r_{pp})+\sin\delta(t)\Im(r_{ss}^*r_{pp})\big) \tag{3.13}$$

The MOKE contributions can be detected by measuring the average intensity at a fixed magnetization and with switched polarizer settings. However, it is much easier to simply set θ_p at either 0 or $\pi/2$. For the polarizer p-setting, it is better to set the PEM also at p-orientation (in the lab frame). This simply interchanges the fast and slow axis of the modulator and leads to a symmetrical formula for the complex s- and p-MOKE rotation. (3.12) then specializes to

$$I_{\text{s-MOKE}}(t)\sim\cos^2\theta_a r_{ss}r_{ss}^*+\sin 2\theta_a\big(\cos\delta(t)\Re(r_{ss}^*r_{ps})+\sin\delta(t)\Im(r_{ss}^*r_{ps})\big) \tag{3.14}$$

$$I_{\text{p-MOKE}}(t)\sim\sin^2\theta_a r_{pp}r_{pp}^*+\sin 2\theta_a\big(\cos\delta(t)\Re(r_{sp}r_{pp}^*)+\sin\delta(t)\Im(r_{sp}r_{pp}^*)\big) \tag{3.15}$$

Equations (3.13), (3.14) and (3.15) are specific versions of the basic equation of photometric ellipsometry, which is the detected light intensity as a function of the most general settings for the optical elements. However, in practical experiments it suffices to work with these specific equations. Remember, that only three input polarizations are needed to completely characterize the Jones matrix of the sample, and this property is convincingly expressed by the form of these three equations. This can be seen by factoring out the dominant reflectance ($R_s = |r_{ss}|^2$ or $R_p = |r_{pp}|^2$) in each of the three expressions, which will introduce in each of the three expressions its corresponding ellipsometric ratio. Note that of course the intensity

[12] Proportionality factors expressing the wavelength detection efficiency, the fraction of the emitted laser light that is passed through the polarizer, and other global set-up dependent factors have been omitted. The same obviously applies to second order terms.

is a time-periodic function through $\delta(t)$. This is the essential characteristic of dynamic photometric ellipsometry. In the next paragraph it will be shown how the basic equations of magneto-optic ellipsometry can be obtained out of (3.13), (3.14) and (3.15).

Basic magneto-optic ellipsometry equations In order to be able to use the previous equations for parameter extraction, their dependence on absolute experimental parameters has to be circumvented. This is can be done by Fourier analysis of the periodic intensity signal. Decomposition of $I(t)$ in its harmonics and separate detection of these harmonic components with e.g. a lock-in amplifier, allows to obtain relative (set-up independent) quantities by measuring the ratios of the harmonic components with respect to the DC intensity.
To identify the harmonic components of $I(t)$, the Fourier decomposition of $\cos(\delta_0 \sin(\omega t))$ and $\sin(\delta_0 \sin(\omega t))$ is required. These can be expanded in their Bessel-function series as follows [17]

$$\begin{aligned} \cos(\delta_0 \sin(\omega t)) &= J_0(\delta_0) + 2\sum_{m=1}^{\infty} J_{2m}(\delta_0)\cos(2m\omega t), \\ \sin(\delta_0 \sin(\omega t)) &= 2\sum_{m=0}^{\infty} J_{2m+1}(\delta_0)\sin((2m+1)\omega t), \end{aligned} \tag{3.16}$$

with $J_m(\delta_0)$ the Bessel function of the first kind of order m.
All elements are now available to calculate the three ellipsometric ratios characterizing the optical system under study. The settings that lead to the simplest formulae for that purpose are with the PEM along the s-direction (zero azimuth) and the analyser at $\pi/4$ relative to the PEM. The polariser setting is determined by the ratio to be measured. For the s-MOKE ratio, $\frac{r_{ps}}{r_{ss}}$, the polarizer is set along the s-axis, for the p-MOKE rotation, $\frac{r_{sp}}{r_{pp}}$, along the p-axis, and for the isotropic rotation, $\frac{r_{pp}}{r_{ss}}$, at a first-order small angle θ_p. Substituting the Bessel expansions of (3.16) into the three intensity expressions (3.13), (3.14) and (3.15), the ratios of the fundamental and the second-harmonic component to the DC component lead to a set of two equations in the real and imaginary parts of the ellipsometric ratios. The canonical form of this set of equations is[13]

$$\boxed{\begin{cases} \dfrac{I_\omega}{I_{\mathrm{DC}}} = \dfrac{4J_1(\delta_0)\Im(\mathcal{S})}{1+2J_0(\delta_0)\Re(\mathcal{S})} \\ \dfrac{I_{2\omega}}{I_{\mathrm{DC}}} = \dfrac{4J_2(\delta_0)\Re(\mathcal{S})}{1+2J_0(\delta_0)\Re(\mathcal{S})} \end{cases} \quad \text{with } \mathcal{S} = \frac{r_{ps}}{r_{ss}},\ \frac{r_{sp}}{r_{ss}},\ \text{or } \theta_p\frac{r_{pp}}{r_{ss}}} \tag{3.17}$$

[13]Note that the set of equations for the ellipsometric ratio $\frac{r_{sp}}{r_{pp}}$ will only assume this canonical form if the polarizer and the modulator are set at $\pi/2$. On top of that this set will act as a consistency check for the equations determining $\frac{r_{ps}}{r_{ss}}$ in the case of a single magneto-optic interface. Because then it is known from [12] that both ratios are exact opposites. If the MO layer is part of a multilayer stack this symmetry property of the sample Jones matrix is lost and the last set of equations is necessary to completely determine the system under study.

Calibration issues The set of equations (3.17) in principle solves the ellipsometry problem of magneto-optic systems. However, there are still some practical implementation issues. The most important one concerns the accuracy of the here sketched measurement technique. Accurate parameter extraction will greatly depend on the accuracy with which the different settings of the optical elements can be determined. Present-day motorized rotary stages easily achieve angular positioning with a resolution of the order of 10^{-4} rad $\approx$ 20arcsec. This makes clear that if the azimuth zero setting of all optical elements is acceptably calibrated, the accuracy of the measurement method will be mainly determined by the precision with which the retardation amplitude δ_0 is known. This retardation can deviate from the nominal setting by variations in the operation wavelength or in the optical length of the PEM. It is clear that this will lead to small deviations in the coefficients in the set of linear equations (3.17) (via the Bessel functions) determining the ellipsometric ratio, which will in turn translate into uncertainties in the optical parameters of the sample. Deviations of 1% in the phase retardation can easily add up to 10% uncertainty in the optical parameters [9]. Another source of errors in the experimental determination of the ellipsometric ratio results from frequency dependent detection of the different harmonic components of $I(t)$. It is known that a lock-in amplifier will have frequency-dependent amplification and sensitivities. As a result the intensity ratios in (3.17) will contain extra factors q_1/q_0, and q_2/q_0, with q_i the detection efficiency of the i$^{\text{th}}$ harmonic component[14].

All of the above makes it clear that it is desirable to have a procedure at one's disposal allowing a calibrated measurement of the ellipsometric ratios. This can be basically achieved in two manners: (1) through calibration of the nominal-retardation-dependent Bessel coefficients of the set of ellipsometric equations, or (2) through direct calibration of the polarization rotation and ellipticity.

1. Calibration of the Bessel coefficients is an ingenious method originally devised by Jasperson [18]. The basic idea behind it, is to devise an ellipsometric optical sequence for which the intensity ratios I_ω/I_{DC} and $I_{2\omega}/I_{\text{DC}}$ will be exactly equal to respectively $2J_1(\delta_0)$ and $2J_2(\delta_0)$, thus only depending on the retardation setting of the PEM. If we label these two calibration ratios as $\mathcal{R}_\omega(\text{cal};\delta_0)$ and $\mathcal{R}_{2\omega}(\text{cal};\delta_0)$, and measure them at the same nominal retardation amplitude as used in the magneto-optic ellipsometric experiment, then we have devised a way to calibrate the set of ellipsometric equations (3.17), i.e. to avoid precise knowledge of the retardation amplitude δ_0 and the corresponding calculation of the Bessel functions. Indeed,

$$\begin{aligned} \frac{1}{\mathcal{R}_\omega(\text{cal};\delta_0)}\frac{I_\omega}{I_{\text{DC}}} &= \frac{2\Im(\mathcal{S})}{1+2J_0(\delta_0)\Re(\mathcal{S})} \\ \frac{1}{\mathcal{R}_{2\omega}(\text{cal};\delta_0)}\frac{I_{2\omega}}{I_{\text{DC}}} &= \frac{2\Re(\mathcal{S})}{1+2J_0(\delta_0)\Re(\mathcal{S})} \end{aligned} \tag{3.18}$$

[14]Remark that it is common practice to include a low-frequency chopper in the optical path to reduce drift noise. If the chopper is operated at a much lower frequency than the PEM, this will not influence the ω and the 2ω components of $I(t)$, and the DC term can be detected at the chopper frequency.

(3.18) still depends on the 0^{th} order Bessel function. But the calculation of this one can also be avoided. It results from the DC contribution of $\cos\delta(t)$ to the detected light intensity. (3.15) and (3.14) show that the DC consists of a term independent of the magnetization and of J_0, and a term linear in J_0 and odd in the magnetization. Thus, locating the δ_0 of the PEM where the DC intensity difference (upon switching the magnetization) changes sign, will correspond to locating the first zero of J_0, without calculating it! After having found this PEM retardation setting δ_0^*, the calibration ratios $\mathcal{R}_\omega(\text{cal};\delta_0^*)$ and $\mathcal{R}_{2\omega}(\text{cal},\delta_0^*)$ can be measured for this setting and the full Jones matrix is determined in a calibrated manner!

$$\boxed{\frac{1}{\mathcal{R}_\omega(\text{cal};\delta_0^*)}\frac{I_\omega}{I_{\text{DC}}} = 2\Im(\mathcal{S}) \qquad \frac{1}{\mathcal{R}_{2\omega}(\text{cal};\delta_0^*)}\frac{I_{2\omega}}{I_{\text{DC}}} = 2\Re(\mathcal{S})} \tag{3.19}$$

The optical sequence needed to measure the calibration ratios can be found in [18]. For both calibration ratios an extra optical element must be inserted in the optical path and the sample must be replaced by an arbitrary, isotropic sample. Sato [19] later also reported another extended variant of this calibration method to include frequency dependent detection. This is important, as Jasperson's calibration method assumes that all frequency component will be detected with the same effciency. The method of Sato also allows a direct calibration of J_0 without having to operate at a zero of this Bessel function. A major drawback of Sato's approach is that it is only valid at near-normal incidence, while Jasperson's approach is independent of the angle of incidence. More details can be found in [19].

2. The calibration methods of both Jasperson and Sato are very laborious. For each sample and each wavelength, one has to recalibrate the measurements. And this calibration must be performed on a different sample, because both calibration methods need an isotropic sample. Therefore, it would desirable to have a method that allows direct calibration of the ellipsometric ratios without having to change anything in the set-up. We will show here that this is straightforwardly achieved for the real parts of the ellipsometric ratios (in other words the rotation), but a little more difficult for the ellipticity (the imaginary parts). The idea is to slightly alter the common azimuth of the whole optical assembly formed by the modulator and the analyzer, and rotate it over a small angle ξ. By doing this, the expressions (3.17) will change and now also contain the angle ξ. This introduces an extra degree of freedom in the ellipsometry equations, which can be used to nullify one of the measured intensity ratios[15]. Looking at the form of the expressions in (3.17), the nulling condition ξ_{null} will be a function of the real or imaginary parts of the ellipsometric ratios and of $J_2(\delta_0)$ or $J_1(\delta_0)$. If this null condition is independent of the Bessel functions, then one has devised again a calibrated way to determine the ellipsometric ratios. We prove here that this is possible for the real parts of $\mathcal{S}$ (with $\mathcal{S}$ as defined in (3.17)). Consider for that

[15] Viewed in this way, this calibration procedure is a hybrid combination of photometric ellipsometry and null ellipsometry.

purpose the experimental settings normally used to derive the equations for $\Im(r_{ps}/r_{ss})$ and $\Re(r_{ps}/r_{ss})$, i.e. the polarizer at 0°, the analyzer at $\pi/4$ and the PEM at 0°. The Jones vector at the detector is then found by multiplication of (3.8) (with ξ of first order smallness) and the first column of the sample Jones matrix.

$$\begin{aligned} E_s \sim r_{ss}\big((1-2\xi)\cos\delta(t)/2 + j\sin\delta(t)/2\big) + \\ r_{ps}\big(\cos\delta(t)/2 - j(1-2\xi)\sin\delta(t)/2\big) \end{aligned} \tag{3.20}$$

$$\begin{aligned} E_p \sim r_{ss}\big(\cos\delta(t)/2 + j(1+2\xi)\sin\delta(t)/2\big) + \\ r_{ps}\big((1+2\xi)\cos\delta(t)/2 - j\sin\delta(t)/2\big) \end{aligned} \tag{3.21}$$

Up to first order in ξ and r_{ps}, the squared amplitudes of the s- and p-components is given by

$$\begin{aligned} |E_s|^2 \sim r_{ss}r_{ss}^*\big((1-4\xi+\cancel{4\xi^2})\cos^2\delta(t)/2 + \sin^2\delta(t)/2\big) + \\ 2\Re\Big(r_{ss}r_{ps}^*\big((1-\cancel{2\xi})\cos\delta(t) + j\sin\delta(t) - j\cancel{2\xi(1-\xi)\sin\delta(t)}\big)\Big) \end{aligned} \tag{3.22}$$

$$\begin{aligned} |E_p|^2 \sim r_{ss}r_{ss}^*\big(\cos^2\delta(t)/2 + (1+4\xi+\cancel{4\xi^2})\sin^2\delta(t)/2\big) + \\ 2\Re\Big(r_{ss}r_{ps}^*\big((1+\cancel{2\xi})\cos\delta(t) + j\sin\delta(t) + j\cancel{2\xi(1+\xi)\sin\delta(t)}\big)\Big) \end{aligned} \tag{3.23}$$

The detected intensity is then found as the sum of (3.22) and (3.23)

$$\begin{aligned} I &\sim 2r_{ss}r_{ss}^*(1-2\xi\cos\delta(t)) + 4\Re\big(r_{ss}r_{ps}^*e^{j\delta(t)}\big) \\ &\sim 1 - 2\xi\cos\delta(t) + 2\Re\left(\frac{r_{ps}^*}{r_{ss}^*}e^{j\delta(t)}\right) \\ &\sim 1 - 2\xi\cos\delta(t) + 2\Re\left(\frac{r_{ps}}{r_{ss}}\right)\cos\delta(t) + 2\Im\left(\frac{r_{ps}}{r_{ss}}\right)\sin\delta(t) \end{aligned} \tag{3.24}$$

This result is in perfect agreement with the basic set of equations obtained earlier, as can be seen by putting $\xi = 0$. What is striking is that the second harmonic component of the intensity (which is contained within $\cos\delta(t)$) can be nullified by adjusting ξ. The nulling equation for the second harmonic component is simply $\xi = \Re(r_{ps}/r_{ss})$, which is independent of the Bessel function. As a result, this nulling scheme provides indeed a calibrated way to measure the real part of the here considered ellipsometric ratio. This procedure can be repeated for the real parts of the two other ellipsometric ratios. In other words, nullifying the second harmonic of the detected intensity by slightly adjusting the azimuth of the assembly PEM+analyzer allows a calibrated measurement of the different polarization rotations! The most important advantage of this method compared to that of Sato or Jasperson, is that it allows to perform the calibrated measurement without the need for changing anything in the original set-up (except for a joint rotation of the PEM and the analyzer).

A similar procedure for the calibration of the different induced ellipticities (i.e. the imaginary parts of the ellipsometric ratios) is not so straightforwardly found. There seems to be no simple method to achieve this without

having to insert extra optical elements in the set-up. However solely calibrating the rotations is often sufficient for complete parameter extraction, provided there are enough other data points for least-squares fitting. This can be done for instance by measuring the calibrated rotation as a function of incidence angle.

The calibration of the set-up is not only a matter of eliminating the precise PEM retardation settings out of the basic equations (3.17). It is obvious that important systematic errors can perturb the measured quantities or even make them meaningless if the reference azimuth settings of the different optical elements are not well calibrated. Just to convince the reader of the importance to calibrate the reference azimuth setting of the polarizer, the modulator and the analyzer with an accuracy of less than an arcmin (with respect to the plane of incidence), it suffices to realize that standard values for the MOKE rotation are generally in the order of tens of arcmin. The calibration methods of these optical elements are well known. The easiest way to achieve this is to calibrate the polarizer and the analyzer on an isotropic sample, with the compensator removed. For such a sample it is known that p- or s-polarized light will remain linearly p- or s-polarized after reflection. By setting the polarizer at its *presumed* s-azimuth and varying the analyzer around its *presumed* p-azimuth the detected intensity will show an intensity minimum. However, this is still not the calibrated setting for the analyzer, as it is not sure that the used polarizer setting transmitted purely s-polarized light! Therefore, this procedure must be repeated for a range of *presumed* s-settings for the polarizer. Plotting the obtained minimum settings will result in a straight line in the (θ_p, θ_a) plane. If now the procedure is inversed, thus setting the analyzer set at subsequent *presumed* p-settings and varying the polarizer to a minimum, a second straight line in the (θ_p, θ_a) plane is obtained. The intersection of both lines will logically yield the global minimum and the correctly calibrated polarizer and analyzer settings. Once the polarizer and analyzer settings are calibrated, it is easy to calibrate the modulator. Indeed by setting the setting the polarizer and the analyzer at their perfectly calibrated crossed s- and p-settings, and inserting the modulator, the detected ligt will only be extinguished if the fast axis (or the slow axis at $\pi/2$) is perfectly aligned with either the polarizer or the analyzer.
The only drawback of this method is that it works only with an isotropic sample. A magneto-optic sample will not retain the linearly s- (or p-) polarization of the incident light. As a result, this calibration method can only be used, if one is 100% certain that a MO sample can be replaced by an isotropic sample (and vice versa after the calibration has been performed) without changing the orientation of the plane of incidence! This is obviously nigh impossible. An alternative but very laborious calibration method of the azimuths without the need for an isotropic sample is described in [20]

3.1.1.2 Static photometric magneto-optic ellipsometry in an alternative PSA configuration

The measurement setup of the previous section is in many ways the most complete and general version of an ellipsometric setup capable of measuring the Jones matrix

of any reflective stack of general anisotropic materials. It is clear that there must be certain simplifications possible to this set-up and to the measurement technique if certain assumptions can be made. This is most certainly the case if the optical system under study consists of a single MO interface between an isotropic ambient medium and a MO material in longitudinal configuration, with the latter having no optical anisotropy nor magneto-optical anisotropy. This was suggested for the first time by Berger [21], [22], but is in many ways actually a very late specialization of the RAE method of Aspnes [11].

Static magneto-optic photometry – the set-up As explained in the paragraph about ellipsometric principles in section 3.1.1, in static ellipsometry the data is extracted by measuring changes in intensity as a function of varying predetermined settings of the optical elements. This important observation implies that for static (or dynamic) photometry, in principle, one doesn't need a retarder (or a modulator). This is because, if one is not pursuing investigation of the system by nulling schemes, then there is no need to be able to produce any possible polarization state. In a way this means that in the method of section 3.1.1.1, the modulator is redundant. However, there it was advantageously used to provide a time-varying parameter, allowing to leave all other parameters constant. Thus, the set-up we are considering now consists solely of a polarizer and an analyzer (and of course a magneto-optic sample meeting the above requirements). The information will be obtained by varying the settings of these two elements in a "sensible manner" and tracking relative intensity changes, i.e. independent of global set-up dependent parameters. We will now show what measurement procedure can be labelled "sensible".

Static magneto-optic photometry – the Jones matrices The above requirements for the magneto-optic sample under study considerably simplifies the unknown sample Jones matrix, since we are basically considering the reflection of a polarized plane wave off a semi-infinite MO medium having no other anisotropy effects except for the longitudinal gyrotropy, and with an isotropic entrant medium. This implies that $r_{ps} = -r_{ps}$ with a known analytical expression and that the diagonal elements are the standard s- and p-Fresnel coefficients [12]. In other words, there is an exact analytical expression available for all the elements of the sample Jones matrix, and this matrix is anti-symmetric. This reduces the number of unknown ellipsometric ratios to two, r_{pp}/r_{ss} and r_{sp}/r_{ss}, and these two ratios can be expressed in an analytical manner with only the complex refractive index and the complex magneto-optical constant of the MO layer as unknowns. The Jones matrix of the sample can be written as

$$\overline{\overline{J}}_s = r_{ss} \begin{pmatrix} 1 & \tilde{\alpha} \\ -\tilde{\alpha} & \tilde{r}_{pp}, \end{pmatrix} \tag{3.25}$$

with $\tilde{\alpha} = r_{sp}/r_{ss}$. Those of the polarizer and the analyzer are given by (3.7a) and (3.7d). And the Jones vector at the detector is given by

$$\begin{pmatrix} E_s \\ E_p \end{pmatrix} = \overline{\overline{J}}_a(\theta_a) \cdot \overline{\overline{J}}_s \cdot \begin{pmatrix} \cos^2\theta_p \\ \cos\theta_p \sin\theta_p \end{pmatrix}, \tag{3.26}$$

if we assume s-polarized light incident on the polarizer.

Static magneto-optic photometry – principles Using (3.26), after tedious calculations the intensity at the detector can be found to be proportional to (ignoring the MO second-order terms)

$$\begin{aligned} I &\sim |E_s|^2 + |E_p|^2 \\ &\sim \cos^2\theta_a(\cos^2\theta_p + \cancel{|\tilde{\alpha}|^2 \sin^2\theta_p} + 2\Re(\tilde{\alpha})\cos\theta_p \sin\theta_p) + \dots \\ &\quad \sin^2\theta_a(\cancel{|\tilde{\alpha}|^2 \cos^2\theta_p} + |\tilde{r}_{pp}|^2 \sin^2\theta_p - 2\Re(\tilde{\alpha}\tilde{r}_{pp}^*)\cos\theta_p \sin\theta_p) + \dots \\ &\quad 2\sin\theta_a \cos\theta_a(-\Re(\tilde{\alpha})\cos^2\theta_p + \Re(\tilde{r}_{pp})\sin\theta_p\cos\theta_p - \dots \\ &\qquad \cancel{|\tilde{\alpha}|^2 \sin\theta_p \cos\theta_p} + \Re(\tilde{\alpha}\tilde{r}_{pp}^*)\sin^2\theta_p) \end{aligned} \tag{3.27}$$

There are several ways to derive out of this intensity expression a relative measure which can be used for photometric ellipsometry. In any case, (3.27) contains four independent unknowns ($|\tilde{r}_{pp}|^2$, $\Re(\tilde{r}_{pp})$, $\Re(\tilde{\alpha})$ and $\Re(\tilde{\alpha}\tilde{r}_{pp}^*)$), in agreement with the fact that there are two complex ellipsometric ratios to be determined. This shows that at least four relative measurements need to be performed. One way to achieve this is to measure the intensity $I^{(i)}$ $(i = 1, \dots, 5)$ for <u>five</u> different, arbitrary combinations of the settings $(\theta_p^{(i)}, \theta_a^{(i)})$, and divide four of these measurements by the fifth to obtain relative quantities. This then results in four equations for the four unknowns.

This might seem in contradiction with the conclusions of the paragraph about ellipsometric principles in section 3.1.1. There it was explained how, for the full ellipsometric characterization of the Jones matrix of any optical system, it suffices to determine the response to only three different input polarizations, while here it is noticed that five different measurements are needed. On top of that nothing forbids to take for that purpose for instance five times the same polarizer setting and five different analyzer settings. However, the above described hypothetic measurement method is by no means equivalent to determining the response of an optical system to an arbitrary input polarization. Because in this statement the word "response" means the polarization transfer function of the optical system. In other words, the precise determination of the output polarization (rotation and ellipticity) for a known input polarization. The here described technique simply measures changes in the detected intensity, and relates these in purely mathematical manner to the system unknowns. This is far removed from analyzing the output polarization state. Therefore, even though the above technique might be viable theoretically spoken, there is reason to suspect that it will not lead to very accurate characterization.

One might question if it is at all possible to do accurate ellipsometric characterization with just a polarizer and an analyzer in the setup, given the fact that at first sight a single analyzer in the output path is not able to resolve the exact polarization state of the reflected light. However, this is not entirely correct. It is easy to see by looking at (3.27) that at a fixed polarizer setting the intensity is a periodic function of the analyzer azimuth θ_a with a period of π. And that this function has exactly two Fourier components, namely one at $\sin 2\theta_a$ and one at

$\cos 2\theta_a$. In [9] it is proven that this specific periodicity is a general property of the detection of an arbitrary polarized light beam with an arbitrary analyzer. It is also proven there how the determination of the two (normalized) Fourier components allows the exact determination of the polarization state of the light entering the analyzer. In other words, by "scanning" a reflected light beam by rotation of the analyzer over half a circle the polarization state of this light beam can be detected. Thus, if the procedure is repeated for at least three different polarizer settings (inputs) the exact polarization transfer function of any sample can be found. This procedure is nothing else but the RAE technique first described by Aspnes [11]. The normalized Fourier analysis of the periodic intensity (as a function of θ_a) is simply a matter of detecting the average signal over a period of π and calculating two Fourier integrals. Doing this for three different polarizer settings ($\chi_{\rm in}^{(i)} = \tan\theta_p^{(i)}$ $(i = 1, \ldots, 3)$) then allows, using the mathematical treatment of Aspnes, to find three pairs $(\chi_{\rm in}^{(i)}, \chi_{\rm out}^{(i)})$, which allow to construct the unique bilinear conformal function completely describing the Jones matrix of the sample. Indeed, the following implicit function,

$$\frac{(\chi_{\rm out} - \chi_{\rm out}^{(1)})(\chi_{\rm out}^{(3)} - \chi_{\rm out}^{(2)})}{(\chi_{\rm out} - \chi_{\rm out}^{(2)})(\chi_{\rm out}^{(3)} - \chi_{\rm out}^{(1)})} = \frac{(\chi_{\rm in} - \chi_{\rm in}^{(1)})(\chi_{\rm in}^{(3)} - \chi_{\rm in}^{(2)})}{(\chi_{\rm in} - \chi_{\rm in}^{(2)})(\chi_{\rm in}^{(3)} - \chi_{\rm in}^{(1)})} \tag{3.28a}$$

determines, after writing it in the standard form

$$\chi_{\rm out} = \frac{J_{22}\chi_{\rm in} + J_{21}}{J_{12}\chi_{\rm in} + J_{11}}, \tag{3.28b}$$

the Jones sample matrix by term to term comparison.
Berger proposes a slightly different approach, avoiding the need for Fourier analysis. Instead of Fourier analyzing the global periodic intensity, it is suggested to construct out of (3.27) a relative intensity function, and to measure this over a single period. The difference is that an explicit fitting function can be constructed which is directly related to the unknowns, without first having to calculate discrete Fourier integrals[16]. The relative intensity function is obtained by constructing the expression for the fractional intensity change after reversal of the magnetization (causing $\tilde{\alpha} \to -\tilde{\alpha}$).

$$\begin{aligned} \frac{\Delta I}{I_{\rm av}} &= 2\frac{I(\tilde{\alpha}) - I(-\tilde{\alpha})}{I(\tilde{\alpha}) + I(-\tilde{\alpha})} \\ &= 4\frac{f_1(\theta_p,\theta_a)\Re(\tilde{\alpha}) + f_2(\theta_p,\theta_a)\Re(\tilde{\alpha}\tilde{r}_{pp}^*)}{f_3(\theta_p,\theta_a) + f_4(\theta_p,\theta_a)\left|\tilde{r}_{pp}\right|^2 + 2f_5(\theta_p,\theta_a)\Re(\tilde{r}_{pp})} \end{aligned} \tag{3.29}$$

[16] Of course, it can be questioned whether this really guarantees an increase in accuracy. Indeed, (3.27) can also be used to write out theoretical expressions for the normalized Fourier coefficients in terms of the unknowns. In essence, there is no difference between Berger's approach and Aspnes's approach, because the fitting procedure of Berger is in fact also a generalized "Fourier decomposition".

with

$$\begin{aligned} f_1(\theta_p,\theta_a) &= \cos^2\theta_a \sin\theta_p \cos\theta_p - \cos^2\theta_p \sin\theta_a \cos\theta_a \\ f_2(\theta_p,\theta_a) &= \sin^2\theta_p \sin\theta_a \cos\theta_a - \sin^2\theta_a \sin\theta_p \cos\theta_p \\ f_3(\theta_p,\theta_a) &= \cos^2\theta_a \cos^2\theta_p \\ f_4(\theta_p,\theta_a) &= \sin^2\theta_a \sin^2\theta_p \\ f_5(\theta_p,\theta_a) &= \sin\theta_a \cos\theta_a \sin\theta_p \cos\theta_p \end{aligned} \tag{3.30}$$

The four unknowns can now be experimentally determined by measuring $\Delta I/I_{\text{av}}$ for at least three polarizer azimuths and analyzing these data according to (3.29). With the knowledge of these unknowns ($|\tilde{r}_{pp}|^2$, $\Re(\tilde{r}_{pp})$, $\Re(\tilde{\alpha})$ and $\Re(\tilde{\alpha}\tilde{r}^*_{pp})$), we are then able to determine the reflection coefficients $\tilde{r}_{ss}$ and $\tilde{\alpha}$, which in turn determine the complex optical and magneto-optical parameters n and Q (via known analytical expressions, because the incidence angle is fixed and there is only a single interface as explained.)

As a final remark, we would like to point out that the here described static photometric ellipsometry technique is at a first sight easier (and cheaper) than the dynamic technique of the previous section. However, it is obviously more sensitive to errors, as it requires several fitting procedures to determine the ellipsometric ratios, whereas the dynamic method achieves a direct calibrated measurement of these ratios. Finally its application is limited to the strongly simplified situation of a single MO interface. In theory, the expressions (3.27) and (3.29) could also be worked out for a multilayer in longitudinal configuration. The only difference would be that in (3.25) $-\tilde{\alpha}$ needs to be replaced by a new unknown $\tilde{\beta}$. However this would imply two extra fitting parameters for the determination of the Jones sample matrix. And then still the elements of this matrix need to be related to the optical parameters of the multilayer with a second fitting procedure. In the dynamic photometric method, the Jones matrix is always directly determined independent of whether the sample is a multilayer or not. Therefore the static approach can only be considered a serious alternative if the complexity of the experimental setup is of higher importance than the accuracy of the experimental results. Nevertheless, due to its simplicity and more importantly its speed, it is an interesting technique to get a first idea of the MO strength when comparing several materials.

3.1.1.3 "Designing" the magneto-optic material – a poor man's approach ?

Up till now, this section has focussed on the theoretical aspects of ellipsometric characterization of the magneto-optical strength of magneto-optical materials. This may seem far removed from the originally set out task of designing the optimal MO metal for the studied integrated optical SOA/isolator. Such an optimal material would be one that provides the ideal combination of low optical absorption and strong gyrotropy. In the strictest sense of the word, such a magneto-optical design

of the material is equivalent to performing ab initio quantum-mechanical calculations of the permittivity tensor (or the conductivity tensor) of the metal. This can be done using Heisenberg-Kramer dispersion relationship [23]. This requires band structure calculations of ferromagnetic metals (thus taking into account the important spin-orbit coupling and exchange coupling terms in the Hamiltonian), a calculation of the electron occupation probabilities (and thus of the density of states), and, finally, calculation of the matrix elements of the electric multipole operator within an as complete as possible set of Bloch wave functions. The mathematical effort and techniques needed to achieve some degree of accuracy in these calculations is enormous. As a result, despite the extensive work that has been done to achieve these ab initio calculations, only the elementary 3d ferromagnets Fe, Co and Ni have been characterized in this way [24], [25]. In any case it is clear that it makes no sense to design the magneto-optic contact material purely on the basis of such an approach. A more pragmatical approach would be through the availability of a semi-classical, phenomenological model relating the conductivity tensor to macroscopical observable physical quantities, such as lifetime or relaxation times of the carriers, density of states, effective mass, magnetization, plasma frequency, ... Such an approach has been tried by some authors [26], [27] using an extended Drude-Lorentz model for electronic conduction in magnetized ionic plasmas (as a model for a ferromagnetic metal). This is basically an intra-band model that describes the magneto-optical effects through a relaxed harmonic oscillator model for the electrons near the Fermi surface and subjected to the effective axial magnetic field of the spin-orbit coupling of these d-electrons [28]. Fitting this model to experimentally determined curves for $\Re[Q(\omega)]$ and $\Im[Q(\omega)]$ of Ni (via magneto-optic ellipsometry), yielded values for the relaxation times, the densities and the effective masses (through the Larmor frequency associated with the spin-orbit magnetic field) of the spin-up and spin-down electrons. Despite the simplicity of this model (and its flagrant contradiction of modern quantum-mechanical concepts) surprising agreement with known intraband electron relaxation times was observed, at least in the low-frequency region ($h\nu < 2$eV, i.e. up to visible wavelengths). Busschow subsequently used this method to measure the fitting parameters for Co and Fe, and got similar agreement [30]. However, he was unable to relate the fitting parameters in this simple model (and thus the magneto-optical properties) to other magnetic material properties for which "general" models do exist (such as the saturation magnetization). As a result, Busschow concluded that there is no such thing as a "rule of thumb" that could be used in the search for new materials having a large Q Voigt parameter. Every new material has to be characterized experimentally. This "poor man's approach" is well illustrated by the effort of Busschow's group in the early 80's, where they characterized close to 200 metallic ferromagnetic systems in the visible range in search for an optimal magneto-optic recording medium [29].

Based on the above conclusions, it is clear that the *a priori* "design" of an optimal magneto-optical ferromagnetic metal contact, is an impossible task. On the one hand, there doesn't seem to exist a general model relating the magneto-optical properties to the (well-known) magnetic properties of ferromagnetic materials. And on the other hand, theoretical *a priori* calculations of the magneto-optic

strengths is only possible after complicated band structure calculations[17]. Therefore, the pragmatical approach has been followed to use that metal which emerged out of the cross-section of two conditions: its technological availability, and the magnitude of its Kerr rotation (even if this one is only known in the visible spectrum). The only candidate left was the $Co_{1-x}Fe_x$ alloy system [30]. Table 3.1 summarizes the results of Busschow extensive study of the Kerr rotation (and the saturation magnetization) of this system. This table seems to suggest that the

Alloy	σ_r (A m^2 kg^{-1})	ϕ_K (deg) $\lambda = 633$ nm	ϕ_K (deg) $\lambda = 830$ nm
Fe	213	–0.41	–0.53
$Fe_{75}Co_{25}$	234	–0.42	–0.46
$Fe_{50}Co_{50}$	230	–0.51	–0.60
$Fe_{25}Co_{75}$	200	–0.48	–0.58
Co	156	–0.30	–0.36

TABLE. 3.1: The specific saturation magnetization and the polar Kerr rotations (at normal incidence) for two wavelengths as measured by Busschow [29] for a few members of the $Co_{1-x}Fe_x$ alloy system.

intermediate composition is the strongest, but one has to be careful in interpreting these data. The measured magneto-optic characteristics are dependent on the crystal quality of the studied samples. The values reported by Busschow were obtained on polycrystalline samples. It is also interesting to note that comparison of the concentration dependence of the magneto-optical properties of the various CoFe alloys with the corresponding saturation magnetization shows that the popular view of the magneto-optical strength as being proportional to the magnetization is a misconception. Magneto-optic ellipsometric characterization results obtained at IR telecom wavelengths are given in Chapter 4.

3.1.2 Magnetic anisotropy

3.1.2.1 Introduction

Once the ferromagnetic contact material has been chosen on the basis of its optical and magneto-optical properties, the second design task consists in finding an appropriate layout for this contact so that it can behave as a good permanent magnet in the direction transverse to the optical waveguide[18]. Studying this aspect brings us into the research area of magnetism of ferromagnetic thin films and more in particular their magnetic anisotropy. It must be said, that in view of the scientific background both of the author and of the research group where

[17] Even when using Voloshinskaya's simplified model, as the effective mass, the average density of states and the spin-orbit strength of the d-electrons must be known.

[18] Note that the terminology is somewhat poorly chosen, as obviously the materials that are being considered (transition metal alloys), are soft magnetic materials and would as such never make a good permanent magnet in the common sense of the word.

this work has been carried out — a background that is topically very far removed from the world of magnetism research — and also partially in view of the specific application-oriented focus of this work, limited research has been conducted on this magnetic aspect of the subject. However, for our particular application the magnetic anisotropy of the ferromagnetic contact of the SOA/isolator can be approached from the more down-to-earth viewpoint of magnetostatics. By this it is meant that the observed magnetic anisotropy in the device will be mainly caused by shape anisotropy considerations and not by crystal symmetry, either because the deposited ferromagetic film is polycrystalline with a very short range order or because the magnetocrystalline anisotropy is negligible.

3.1.2.2 Phenomenology of magnetic anisotropy

The term anisotropy is generally used to express that a physical quantity has a behaviour that is a function of direction. In most physical disciplines anisotropic behaviour is only exceptionally encountered. In (ferro-)magnetism however, it is an indispensable concept, often playing a very central role in the understanding of many magnetic devices. Magnetic anisotropy expresses the ease with which a magnetic material is magnetized in certain spatial directions. In this context, one speaks of *easy* and *hard* axes in the material. The direction of these axes with respect to a coordinate system "attached" to the ferromagnetic sample, is determined by a combination of the shape of the sample (*shape or magnetostatic anisotropy*), its crystalline state (*magneto-crystalline anisotropy*), its stress (*magneto-elastic anisotropy*), and its pre- and/or post-fabrication conditions (*field- or texture induced anisotropy*) [31]. It is clear that not all four mechanisms will have the same relative importance. Depending on the deposition conditions, the type of material, the aspect ratios of the sample, ..., one or the other anisotropy mechanism will be dominant.
As magnetic anisotropy is basically a matter of direction-dependent energy needed to magnetize a sample, the most straightforward manner to characterize it is by considering the magnetization curves along the different hard and easy axes in the sample. This allows to calculate the anisotropic energy density as a function of magnetization direction. A subsequent fit to a Legendre polynomial expansion of the energy density allows extraction of the effective anisotropy constants [32]. In practice though, one will more often measure the magnetic hysteresis of a ferromagnetic sample rather than its magnetization curve, as the latter can only be measured on a perfectly demagnetized sample. Such a hysteresis measurement will in principle not allow the determination of magnetization energy, as it is a manifestation of an irreversible process[19]. However, the magnetic hysteresis curves measured along different directions have features that are strongly dependent on the magnetic anisotropy. Fig. 3.2 show typical magnetic normalized hysteresis curves of a hypothetic ferromagnetic sample along its easy and hard axes. The qualitative features of magnetic anisotropy in real samples are easily distinguishable: very low coercivity and remanence along the hard axis, while along the easy axis, a remanence close to the saturation magnetization and coercivity close to

[19] One exception is of course when a perfect hard axis is considered.

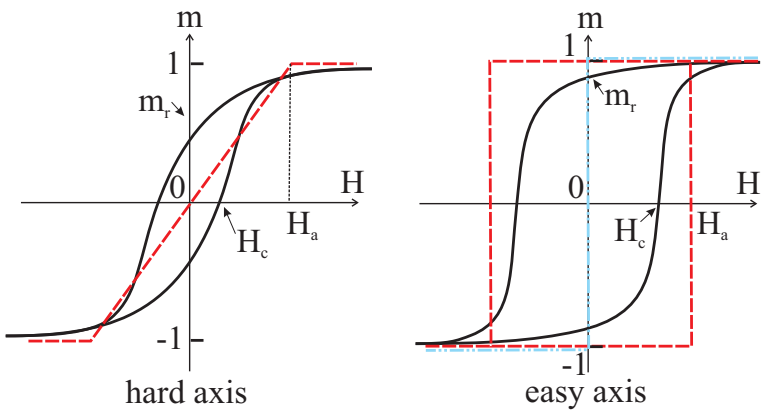


Fig. 3.2: Realistic and idealized normalized ($m = M/M_s$) hysteresis loops along orthogonal easy and hard axes in a hypothetical ferromagnetic sample showing canonical uniaxial anisotropy. The idealized curves depict the limiting situation of a single magnetic domain sample (or a sample with perfectly pinned domain walls) with pure rotational hysteresis (dashed curves). For the easy axis a second limiting situation is also depicted (dash-dotted curve), namely the one where the domain walls are perfectly free. In the ideal situation there is no coercivity along the hard axis (or along the easy axis if the walls are free), and the coercive field along the easy axis is usually called the anisotropy field H_a.

theoretical maximum anisotropy field is observed. The physical mechanisms that cause the deviations from the ideal curves are numerous. But in a way they are all related to phenomena taking place at a micro-level, be it crystalline imperfections (e.g. arbitrary oriented polycrystalline grains and their boundaries) averaging out the easy and hard axes directions, or magnetic structure imperfections (e.g. magnetic order-disorder, non-magnetic inclusions) pinning free domain wall motion. We will not expand on these mechanisms, but rather elaborate on the precise form of the idealized loops, without making any assumptions about the precise nature of the magnetic anisotropy at play. In this way an idea will be given about the magnitude of the effects at play. The realistic hysteresis curves can be thought of as an average of the idealized curves. For simplicity, it will only be assumed that the anisotropy is uniaxial. In a second step it will be investigated how magnetostatic shape effects influence the hysteresis loops [20]. It will be seen how shape effects can be very easily modelled by a shearing transformation.
Consider an idealized ferromagnetic sample consisting of a single domain (or with perfectly pinned domain walls) and where all magnetic moments will change their magnetization by coherent rotation in unison under the influence of an applied field[21]. Thus, we consider only coherent reversible and irreversible magnetization

[20] Note that this order of reasoning presupposes that the idealized loops, taken as starting points, are not yet influenced by shape anisotropy.

[21] In other words, we neglect non-uniform magnetization rotation mechanisms such as curling and buckling, which can lower the coercive field considerably.

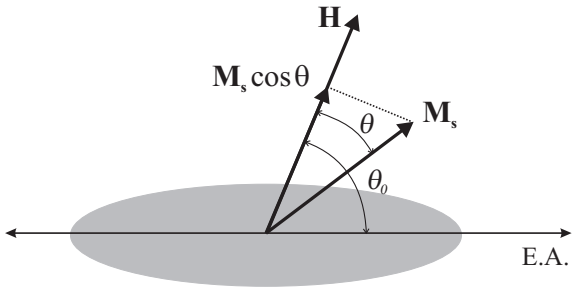


Fig. 3.3: The Stoner-Wohlfarth problem for the canonical case of uniaxial anisotropy with anisotropy energy constant K_u (in J/m^3). The ellipsoidal shape of the sample is only chosen to schematically indicate that a single-domain sample is considered. In principle we are not yet considering shape anisotropy effects (even though these can be described by an uniaxial phenomenological model as well and could be very easily included in the uniaxial anisotropy constant because here the ellipsoid axis coincides with the anisotropy axis).

rotation as possible mechanisms for magnetization re-orientation. The sample is supposed to have uniaxial magnetic anisotropy. We will derive the equilibrium position of the magnetization for various values of the magnetic field. Fig. 3.3 depicts the considered situation. This is referred to as the Stoner-Wohlfarth problem. Here we will only consider the limiting case of a magnetic field applied either perpendicular (i.e. along the hard axis) to the uniaxial easy axis or along it ($\theta_0 = \pi/2, 0$). Intuitively it is clear that, in view of the assumption of pure rotational hysteresis as only magnetization re-orientation mechanism, the easy axis magnetization will be 100% irreversible, while hard axis magnetization will consist of pure reversible magnetization rotation. We will now make these conclusions quantitative.
The magnetostatic energy density of the ferromagnetic sample in Fig. 3.3 is composed of two parts, an anisotropy term describing the equilibrium position of the magnetization in the absence of an external field (and without considering shape effects) and the standard Zeeman energy density due to the magnetization orientation in the applied magnetic field. If the anisotropy of whatever origin is supposed to be uniaxial, meaning there is a single preferred direction and the directions in the plane perpendicular to this are equivalent, then the free energy of the sample can be written as

$$u = \begin{cases} -K_u \cos^2\theta - \mu_0 H M_s \cos\theta & \text{for } H \parallel \text{easy axis (or plane)} \\ -K_u \sin^2\theta - \mu_0 H M_s \cos\theta & \text{for } H \perp \text{easy axis (or plane)} \end{cases} \tag{3.31}$$

If the uniaxial anisotropy energy constant K_u is positive, one has an easy axis. A negative K_u implies an easy plane perpendicular to the uniaxial reference direction. The equilibrium position for the magnetization at an arbitrary value for the magnetic field[22] is found by locating the stable extrema of the energy function

[22] with arbitrary it is meant of course below the saturation field at which strength M_s points in the direction of H.

u. These are the zeroes of the first derivative $\frac{\partial u}{\partial \theta}$ (i.e. the angular positions for which the field exerts no torque on the magnetization), and that have a positive second derivative $\frac{\partial^2 u}{\partial \theta^2}$ (i.e. with a positive restoring torque for deviations from the equilibrium). This gives

$$\left.\begin{aligned}(\mu_0 H M_s + 2K_u \cos\theta)\sin\theta = 0\\ \mu_0 H M_s \cos\theta + 2K_u \cos 2\theta > 0\end{aligned}\right\} \quad \text{for } H \parallel \text{easy axis (or plane)} \qquad (3.32a)$$

and,

$$\left.\begin{aligned}(\mu_0 H M_s - 2K_u \cos\theta)\sin\theta = 0\\ \mu_0 H M_s \cos\theta - 2K_u \cos 2\theta > 0\end{aligned}\right\} \quad \text{for } H \perp \text{easy axis (or plane)} \qquad (3.32b)$$

The solutions for both cases are now easily found.

Hard axis magnetization has a solution at $\theta = 0, \pi$, and one at $\cos\theta = \mu_0 H M_s / 2K_u$. According to the stability condition of (3.32b), the perfectly parallel and perfectly "anti-parallel"[23] solutions are stable for either $H > 2K_u/\mu_0 M_s$ and $H < -2K_u/\mu_0 M_s$. In other words at $|H| = H_a = 2K_u/\mu_0 M_s$ the sample is completely saturated along the hard axis. This is the definition of the anisotropy field H_a. The other solution describes the magnetization below saturation (i.e. at fields weaker than the anisotropy field), and is always stable[24]. Multiplying both sides of this solution by M_s and using the definition of the anisotropy field, this solution can be written as $M_s \cos\theta = \frac{M_s}{H_a} H$. The left-hand side is the component of the magnetization along H, and this equation states that (for fields weaker than the saturation field H_a), the magnetization increases linearly up to M_s with a slope $M_s/H_a = \mu_0 M_s^2/2K_u$. This is exactly what is depicted in Fig. 3.2. We clearly see here pure reversible magnetization rotation. Starting off with a completely saturated sample along the hard axis, the anisotropy forces the magnetization to gradually rotate towards the easy axis (along the equation of "motion" $M_s \cos\theta = H M_s/H_a$) as soon as the field is decreased below the anisotropy field. At zero applied field the magnetization is perfectly aligned with the easy axis (zero projection on the H-field). A further (now negative) decrease of the field rotates the magnetization gradually away from the easy axis towards "negative" values until it is again perfectly saturated at $-H_a$ along the hard axis in the opposite direction. Remark that this process is perfectly reversible.

Easy axis magnetization has a solution at $\theta = 0, \pi$, and one at $\cos\theta = -\mu_0 H M_s / 2K_u$. However this latter solution ($\cos\theta = -H/H_a$) is never stable. The perfect "parallel" and "anti-parallel" solutions are found to be stable (according to (3.32a)) for $H > -2K_u/\mu_0 M_s = -H_a$ and $H < 2K_u/\mu_0 M_s = H_a$. In other

[23] this is in the sense of the positive H-direction as defined in Fig. 3.3)

[24] This is easy to see by rewriting the solution as $\cos\theta = H/H_a$, and the stability as $H/H_a \cos\theta - \cos 2\theta > 0$. If the solution is is substituted in this latter condition, we get $\cos^2\theta - \cos 2\theta > 0$ which is always true.

words, the sample will always be saturated along the easy axis and will switch discontinuously its orientation only if an opposed field of at least the strength of the anisotropy field H_a is applied. This is pure irreversible magnetization rotation, and corresponds to the pure square hysteresis loop of Fig. 3.2. One may wonder why switching doesn't occur at a field strength of only $H_a/2$? After all, that is the height of the anisotropy barrier $-K_u \cos^2\theta$ between the two equivalent positions at 0 and π. This is because only rotational magnetization re-orientation is allowed. Therefore, while "climbing" the anisotropy barrier the magnetization increases its anisotropy energy faster than it lowers its potential Zeeman energy in the opposite oriented $-H_a/2$ field. As a result the solution at $\theta = 0$ remains stable, until the torque due to the Zeeman energy for small deviations away from $\theta = 0$ can overcome the opposite torque due to the anisotropy. From that point on, the Zeeman torque can overcome the anisotropy torque for every orientation between $\theta = [0, \pi/2]$, and the magnetization is instantaneously "accelerated" towards π. This is just a qualitative assessment of the mathematical result that $\theta = 0$ remains a stable extremum of the energy function (3.31) down to $-H_a$. If magnetization re-orientation due to domain wall motion was allowed, and if these walls were not hindered by magnetic defects, then at remanence a infinitesimal negative field would immediately nucleate walls and accelerate these walls throughout the entire sample, resulting in the perfectly reversible magnetization (dash-dotted) loop of Fig. 3.2. In the hard direction the rotation is always reversible because the Zeeman torque is at its maximum at $\theta = \pi/2$ (i.e. the easy axis direction), while the counteracting anisotropy torque is at its minimum for the easy axis orientation. As a result the magnetization can be reoriented for any value of the applied field.

Solutions to the Stoner-Wohlfarth problem for general directions of the applied magnetic field, can be intuitively derived out of the above two limiting situations. The resulting loops will be a combination of both reversible rotation and and irreversible jumps. Irreversible behaviour and the corresponding coercive fields will increase with decreasing angles between the field and the easy axis. H_a being the fundamental limit for the coercive field[25]. The above reasoning can be extended for the more general case of cubic anisotropy. This is obviously more complicated as the hard and easy axes are not orthogonal anymore, and there can exist axes of intermediate "hardness". However, once again it is found that the order of magnitude of the coercive fields is $K/\mu_0 M_s$ with K an effective cubic anisotropy constant. This means that for most traditional transition metals with $K \approx 10^{+5}$ J/m^3 and $\mu_0 M_s \approx 1 - 2$ Wb/m^2, the coercive fields are of the order of $5 \cdot 10^4 - 10^5$ A/m or in the range 100-1000 Oe. This is very strong! Which would imply that most materials would make good permanent magnets, if only rotational magnetization re-orientation was considered. However, the irreversible magnetization switching is almost always due to irreversible domain wall motion, the so-called Barkhausen jumps. And this can occur at considerably lower field strengths than the ones calculated above. The modelling of these Barkhausen

[25] A rigorous treatment can be found in [33].

jumps requires complicated micromagnetic models of the interaction of a domain wall with defects in magnetic materials (see for instance [34]). It is found that, depending on the size and the strength of the defects, irreversible rotation jumps in general never get the opportunity to occur. The switching in the easy direction almost entirely consists of irreversible giant Barkhausen jumps near coercive field strengths lower than those calculated for rotational switching. The coercive Barkhausen field gradually increases up to the fundamental limit H_a as the applied field makes an increasing angle with the easy direction. And the magnetization reversal gradually becomes a combination of first reversible rotation (and small Barkhausen jumps) followed, at the (lowered) coercive field, by a giant irreversible Barkhausen jump[26]. In other words as the applied field points more and more in the hard axis direction, rotation slowly takes over again until, for pure hard axis magnetization, the magnetization again becomes a pure reversible rotation. This is a logical result as a field applied perpendicular to a 180^o domain wall does not exert any pressure on this wall.

All of the above observations were for a perfect sample with an unique direction of the easy axis. Real samples are polycrystalline with a random distribution of the easy direction among the grains. The magnetization loop will then in general be an average of all possible loops with H applied at an angle with the easy axis between 0 and $\pi/2$. On a macroscopic scale this averaging would thus lead to magnetic "isotropic" hysteresis loops independent of the direction of the applied magnetic field, and each direction exhibiting coercive fields in the order of the domain wall switching field[27]. In other words the magnetic sample looses its distinct easy direction. Therefore, if a certain magnetic anisotropic behaviour is desired in a polycrystalline sample, it has to be induced by other means that cause a majority of the grains to have their easy axis pointing in a particular direction. It is clear that this can only be achieved by a subtle control of the "fabrication" of the sample, either by taking pre-deposition measures, by controlled deposition or by post-deposition treatment of the magnetic sample.

By pre-deposition measures it is usually meant that a host layer for deposition is used which is "suitably" textured so that it can induce a preferred crystallographic orientation for the grains, and thus for the easy axes[28]. It is clear that for our purposes, this is not an option. Any extra layer between the magnetic contact and the underlying waveguide might do more harm than good to the operation of the SOA/isolator. Most often such a suitable host material is metallic and as such partially screens the guided light from the magneto-optic material, and if it is not absorbing, it will most certainly influence the electrical behaviour of the ferromagnetic contact.

Another option to achieve a high uniformity of the easy axis direction between the grains is by controlling the deposition conditions. In the ideal case, one would

[26]...sometimes accompanied by an irreversible rotation if the applied field is pointing more and more in the hard direction.

[27]which is for standard transition metal alloys in the order of tens of Oe.

[28]a specific example of this is of course the search for a materials that have good degree of lattice matching with the deposited magnetic material, so as to achieve a long range epitaxial degree in the deposited material (with known magneto-crystalline anisotropy). For the CoFe alloy system Ta is often chosen for this purpose.

strive to achieve quasi-perfect epitaxy, so as to have a maximal degree of crystallinity and to recover in this way the pure magneto-crystalline anisotropy of the deposited material. This is of course only possible if the magnetic material is lattice-matched to the host material, and this is certainly not the case for CoFe alloys on a InP-lattice. The former has a bcc structure with a lattice parameter of $\approx 3\mathring{A}$, while the latter is well-known to have the zincblende structure with a lattice constant of 5.868 Å. A better way to proceed is to deposit the material under applied magnetic field bias. When deposited in a magnetic field a uniaxial, in-plane magnetocrystalline anisotropy is induced in the polycrystalline ferromagnetic material [35]. The easy axis is set by the magnetic field direction. This can be qualitatively understood as a forced arrangement of the magnetic dipoles making up the magnetic material during deposition. This technique has been used for all fabricated devices throughout this work, and results in the qualitative hysteresis loops depicted in Fig. 3.2. Precise characterization of this induced anisotropy as a function of ferromagnetic film thickness (and stripe width) will be detailed in the next Chapter.
Finally, the most encountered technique to induce a certain magnetic anisotropy in a polycrystalline ferromagnetic sample is through so-called magnetic annealing. It is a post-treatment technique, which combines the basic idea of field-biased deposition with thermally induced mobility for the needed magnetic dipole rearrangement. Magnetic annealing is well-documented and modelled (see [36] for an excellent overview). However, the needed temperatures to achieve reasonable mobilities for the transition metal atoms are of the order of 500^oC. It will be seen in the next Chapter that these temperatures conflict with the ideal post-treatment temperatures for achieving a good Ohmic contact.
Even though both techniques of magnetic annealing and field-biased deposition are well-established, it is clear that their mechanisms are very hard to model theoretically. Some phenomenological models do exist, but it is by no means obvious to calculate a priori what the associated anisotropy constant is going to be. In any case, reported measurements on permalloy indicate that the resulting uniaxial anisotropy can be an order of magnitude lower ($K_u = 10^3 \mathrm{J/m}^3$) than the ideal crystalline anisotropy.

3.1.2.3 Shape effects – demagnetization fields

Up till now all that has been said applied to the case of semi-infinite blanket films[29]. When a ferromagnetic sample has finite dimensions, the flux lines must close by crossing an interface between two dissimilar materials (most often a magnetic and non-magnetic). As a result, free magnetic poles are created at this interface[30], which can be quantified by a magnetic surface charge density, $\sigma_m = \mu_0 \mathbf{n}_{12} \cdot (\mathbf{M}_1 - \mathbf{M}_2)$, with $\mathbf{n}_{12}$ the unit vector normal to the interface between medium 1 and 2 and pointing from 1 to 2. Note that the magnetic charges have

[29] For obvious reasons we are treating the magnetization problem of the deposited films as a 2D planar problem. The magnetization is not expected to vary a great deal in the vertical direction ... as will be shown in this section.

[30] that is if the Maxwell equations are formulated in the Chu model.

the same unit as the magnetic induction, namely Wb. In the Chu model for the Maxwell equations in matter, the fundamental electromagnetic field components are **E** and **H**, the former generated by magnetic current densities $\mathbf{J}_m$ and electric charge densities ρ, the latter by magnetic charge densities ρ_m and electrical current densities **J**. The generalization of the Gauss equation towards magnetic fields caused by magnetic charges reads $\mu_0 \boldsymbol{\nabla} \cdot \mathbf{H} = \rho_m$. If it is accepted that there can be no free magnetic monopoles (and thus no free magnetic charge density nor magnetic current density), then this magnetic charge density must be given by the divergence of the material magnetization $\rho_m = -\mu_0 \boldsymbol{\nabla} \cdot \mathbf{M}$. Only then can the divergence-freeness of the magnetic induction be guaranteed. This explains the above form for the magnetic surface charge density.

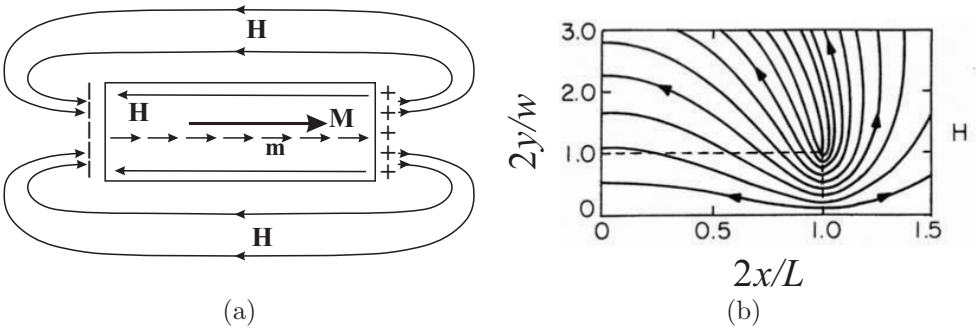


Fig. 3.4: (a): Schematic illustration of the creation of demagnetizing fields due to the magnetic surface charges created by the finite dimensions of the sample; (b): Magnetostatic simulation of the magnetic field in the neighbourhood of a uniformly magnetized bar (only a quarter of the bar is sketched because of the symmetry of the problem), creating an uniform magnetic charge density at both ends of the bar. Note that this is a hypothetical problem since only ellipsoids can be be uniformly magnetized!

These observations make clear that in magnetostatics[31] the magnetic field is generated by magnetic charges, which are themselves generated in any volume for which the magnetization has a net incoming or outgoing component. As a result, in any magnetized sample of finite dimensions[32], magnetic charges are created on the boundary surfaces, which in turn create inside the sample a counteracting magnetic field, according to $\boldsymbol{\nabla} \cdot \mathbf{H} = -\boldsymbol{\nabla} \cdot \mathbf{M}$. This is schematically illustrated in Fig. 3.4.

This figure serves to illustrate how a magnetized sample of finite dimensions will always be subjected to a demagnetizing field. This is a logical consequence of the fact that it costs energy to align magnetic dipoles in a finite volume. It it is intuitively clear that this demagnetizing field is proportional to the magnetization, and will have an amplitude that cannot exceed the strength of the magnetization itself. The latter observation can be justified by considering the energetically

[31] and assuming no free electric currents to be present.
[32] we assume for simplicity a defect-free material

least favourable situation, namely that of a infinitesimal thin sheet of parallel aligned elementary magnetic dipoles. Using Gauss' law, it is straightforwardly found that each of the magnetically charged sheets contribute $M/2$ to the internal magnetic field. Thus, right at the sheet (or better in the interior of the elementary magnetic dipole sheet) the demagnetizing field equal the magnetization. Therefore, for arbitrary magnetized shapes one usually writes the demagnetizing field as

$$H_{\text{dem}} = -NM \qquad \text{with } 0 \leq N \leq 1 \text{ the so-called demagnetizing factor} \tag{3.33}$$

The demagnetizing factor is a dimensionless number between 0 and 1 that expresses the difficulty to magnetize a shape along the considered direction. It is an anisotropic function of the geometry of the sample depending on its aspect ratios along the considered magnetization direction. It has no relationship with the material parameters. Note that in the above expression the demagnetizing factor N is only constant if the shape is an ellipsoid. In general, the demagnetization factor is a tensor function of the sample shape (reducing to a double- or triple valued diagonal tensor for ellipsoids), and from Fig. 3.4 it is clear that it is also a function of position in the sample (except for ellipsoids). If a very thin rectangular ferromagnetic film is considered[33], then Fig. 3.4(b) shows, by appropriately stretching, that the field inside the sample is essentially uniform in the direction of the thickness. Because Fig. 3.4(b) applies for a rectangular shape with one infinite dimension (i.e. either infinitely wide or infinitely long), the uniformity of the field is not guaranteed in the second direction, i.e. along the width when magnetized along the length or along the length when magnetized along the width. Therefore, the following formulae are strictly speaking only valid in the centerline of the sample width or length depending on its magnetization direction. But if the aspect ratio of the width versus the length is not too small or too large, this is deviations of the demagnetization factor are not too large. On top of that, it is sufficient to have only a semi-quantitative idea of N. After all, it is of course nigh impossible to calculate the exact effect of the sample shape on the hysteresis loops everywhere in the sample.
In [37] it is proven that for such thin rectangular ferromagnetic films the demagnetization factors are given by

$$N_{\text{dem,l}} \approx \frac{2t}{\pi l}\sqrt{\frac{2w}{l}} \tag{3.34a}$$

$$N_{\text{dem,w}} \approx \frac{2t}{\pi w}\sqrt{\frac{2l}{w}} \tag{3.34b}$$

The variables are illustrated in Fig. 3.5. As an example, a standard stripe with a thickness of 100 nm (=0,1 μm), a width of 25 μm, and a length of 1000 μm, will have demagnetization factors of resp. $N_{\text{dem,l}} \approx 1.42 \cdot 10^{-5}$ and $N_{\text{dem,w}} = 0.0228$. Or otherwise put, it is about three orders of magnitude more difficult to magnetize such a stripe along its width than along its length!

In order to visualize the demagnetization effect on the hysteresis loops of a ferromagnetic sample, it suffices to realize that the magnetic field on the abscissa

[33] with a thickness that is at least an order of magnitude lower than its width.

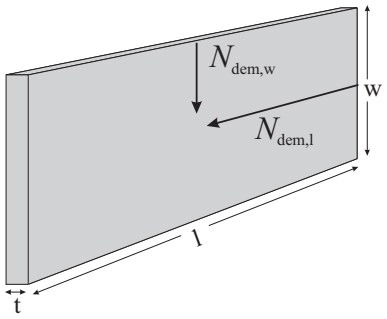


Fig. 3.5: Demagnetization factors of thin rectangular ferromagnetic film (see (3.34)*)*

of a hysteresis plot is the internal magnetic field in the sample. In other words, the true magnetic field in a sample taking its shape into account is given by

$$H_{\text{int}} = H_{\text{applied}} - N_{\text{dem}} M \tag{3.35}$$

The applied field is reduced by the demagnetization field. The transformation of the hysteresis curve characteristic of a material (with "infinite extension") to the hysteresis curve of that same material with a certain shape is a shearing transformation over an angle cotan$\alpha = \frac{1}{N_{\text{dem}}}$. This is straightforwardly found as follows. The vertical lines of constant $H_{\text{int}} = H^*$ in the (M, H_{int})-plane are according to (3.35) transformed into straight lines in the (M, H_{applied})-plane with an equation $M = \frac{1}{N_{\text{dem}}}(H_{\text{applied}} - H^*)$. Thus the vertical lines are sheared over the above angle α. This is illustrated in Fig. 3.6. Shape effects will cause a shearing of the

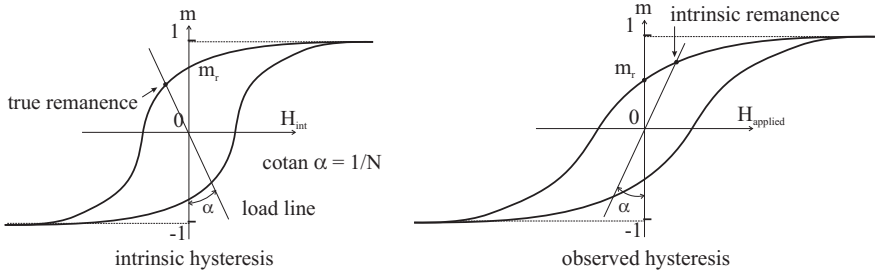


Fig. 3.6: The effect of sample size on the fundamental hysteresis loop of a ferromagnetic sample. Given the fundamental hystersis curve the remanent magnetization of a sample magnetized along a direction with demagnetization factor N is straightforwardly found as the intersection of the so-called "load line" with the loop.

intrinsic hysteresis loops. The main effect of shape anisotropy is thus a reduction of the sample magnetization for a given applied field, or equivalently an increase of the needed magnetic field to obtain a certain magnetization. Magnetostatic shape

anisotropy will in principle not change the magnitudes of the coercive field.
In the treatment above the effect of the demagnetization factor in one particular direction was taken into account. When considering magnetization processes in samples of finite size, demagnetization effects appear in all spatial directions. Therefore, it is better to describe the shape anisotropy by a difference in demagnetization factors. If the magnetostatic problem can be described by a 2-dimensional phenomenon, the shape anisotropy will be given by $\Delta N = N_{\mathrm{dem},1} - N_{\mathrm{dem},2}$ where $N_{\mathrm{dem},1}$ resp. $N_{\mathrm{dem},2}$ are the demagnetization factors along the "long" and the "short" axes[34].

3.1.2.4 Measurement of magnetic anisotropy

In order to experimentally characterize magnetic anisotropy in a quantitative manner, one ideally would have to measure the saturation magnetization energy in different directions (provided the sample is demagnetized first). This could be done using a magnetometer. The only problem with magnetometers is that the sample dimensions have to be precisely known in order to calculate the magnetization. This might be a problem with very small samples. A more precise method that doesn't require the knowledge of the volume of the sample, uses torque magnetometry. This allows to measure the anisotropy torque as a function of angle with the easy axis. This anisotropy torque is nothing else but the angular derivative of the anisotropy energy. Just as with measuring the magnetization energy, a fitting procedure can then give the anisotropy constants (without having to know the precise magnetization) [31]. In any case very accurate determination of the anisotropy constants is not so easy.
Magnetic anisotropy is therefore most often characterized in a qualitative manner by measurement of the relative characteristics of the hysteresis curves in different directions, without the need for an absolute determination of the magnetization[35]. For that purpose a relative dimensionless measure is needed that describes in a semi-quantitative fashion the direction-dependent "ease" of magnetizability. The observations of this section have proven that such a measure of anisotropy should not only be an expression of the tendency of high remanent magnetization but also of the relative stability of the remanent magnetization at opposite applied fields. This latter is an expression of the distinct magnetization bistability which is so characteristic of a sample with a well-defined easy axis. Both effects are "measured" by the squareness of the hysteresis loop. The traditionally encountered definition of squareness is $S = m_r = M_r/M_s$. However this doesn't say anything about the stability, which is related to H_c. This stability is expressed by the differential susceptibility at the coercive field $\left[\frac{\partial M}{\partial H}\right]_{H_c}$. The more this susceptibility exceeds the average susceptibility M_r/H_c in the second quadrant of the hysteresis curve, the squarer the hysteresis curves becomes. Obviously, the differential

[34] Note that this form for the shape anisotropy should only be used in energy calculations (e.g. in the calculations of magnetization process 'dynamics'). Indeed, a spherical sample will have no preferred axis due to shape aniosotropy, but still in any direction $H_{\mathrm{dem}} \neq 0$!

[35] For the measurement of the hysteresis one can in principle use any effect that is proportional to the magnetization. Any magnetometer will do. But for this purpose, one could also use the MO Kerr effect.

susceptibility at the coercive field is always greater than or equal to the average susceptibility. Therefore a better (dimensionless) definition of the squareness of the hysteresis curve with a value between 0 (not square) and 1 (most square) is given by

$$S^* = 1 - \frac{M_r/H_c}{[\partial M/\partial H]_{H_c}} = 1 - \frac{m_r/H_c}{[\partial m/\partial H]_{H_c}} = 1 - \frac{S}{H_c[\partial m/\partial H]_{H_c}} \tag{3.36}$$

This definition does contain the coercive field as opposed to m_r, and will as such also express the decrease of stability (or thus the increase in hardness) when the coercive field decreases in certain directions even when S doesn't change significantly. Fig. 3.7 illustrates these two measures of the loop squareness. Of course all

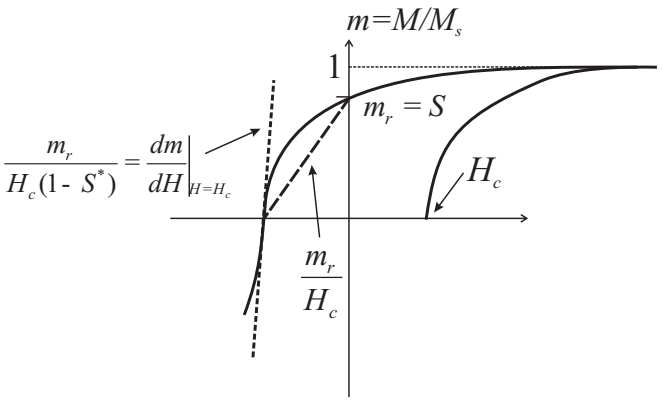


Fig. 3.7: Part of the typical measured hysteresis curve showing the various characterization parameters.

of the above should be treated with some caution when characterizing magnetic anisotropy. Measuring the squareness of the hysteresis loops for different directions in the sample will in the first place give an idea of the relative hardness of the different axes, or in other words about the distinctness of an easy axis in the sample. However, it obviously does not allow a comparison of the absolute strength of the anisotropy for different materials. For that purpose at least the absolute strengths of the coercive fields have to be considered also. And to be completely correct even the absolute values of the saturation magnetization would then have to be known also, because an increased coercivity could be accompanied by a decrease in saturation magnetization, which would imply a negligible change in the anisotropy constant. Therefore the measurement of the squareness is in the first place useful to detect how distinct the easy axis behaviour is and how this distinctness is influenced by deposition conditions, post treatments, geometric sample parameters, ... For our purposes, this is sufficient, as we are in the first place looking for a ferromagnetic metal contact with a distinct easy axis in the direction transverse to the waveguide (i.e. perpendicular to the ridge waveguide) and a reasonably hard axis in the longitudinal direction. It is by no means certain that an increased

crystalline anisotropy would also increase this distinction. It would however increase the coercive fields in the easy direction, which is by itself also an important parameter. However, the coercive field can always be easily measured without determining the anisotropy constants. Therefore, the magnetic characterization of the metal contact purely on the basis of hysteresis measurements provides all the necessary information.

3.1.2.5 Magnetic contact design

The discussion of this paragraph has shown that there are sufficient theoretical and phenomenological models available to study magnetic anisotropy. However all of these models inevitably make simplifying assumptions about the crystalline structure of the material making it almost impossible to predict/design the anisotropy of a polycrystalline realistic ferromagnetic film. It was indicated how for our application anisotropy in polycrystalline films can be induced by controlled deposition under the presence of a magnetic bias field. The phenomena responsible for this are well understood but are again difficult to model exactly. Therefore there is not much design to be expected from a material choice point of view. All that can be done is taking the pragmatical approach, and choose out of a series of ferromagnetic samples fabricated under varying conditions and/or with varying parameters the ones with the best anisotropic behaviour. These experimental results will be elaborated in the next chapter.
In any case, the section 3.1.2.3 on magnetic shape anisotropy has taught us that too large shearing effects (and the accompanying drop in remanence) can be avoided if the demagnetization factor is kept as small as realistically possible within the boundary conditions set by other design aspects. Using (3.34) it is seen that the detrimental small contact width-to-length aspect ratio w/l inherent of the layout of the SOA/isolator device only contributes as a square root in the expression for the inverse demagnetization factor (i.e. the cotangens of the shearing angle[36]), while the width-to-thickness aspect ratio w/t contributes linearly. This is implies that the negative impact of a small length aspect ratio (usually a factor 10^{-3}, namely a width feature in the order of μm's with respect to a length feature in the order of mm's) is decreased by the square root and can be further reduced by an increase of the (usually) large w/t. This is illustrated in Fig. 3.8, plotting (3.34) as a function of both mentioned aspect ratios. This plot illustrates how the shearing effect will decrease with increasing w/t, and how at large l/w a further increase in length of the device will have an increasingly smaller influence on the shearing of the hysteresis. In this way a reduction of squareness of the hysteresis due to the magnetic shape anisotropy is always possible by an increase of the width-to-thickness ratio. It is clear that for the SOA/isolator this is best achieved by increasing the width of the contact, rather than by decreasing the contact thickness. Indeed, the latter is closely related to strength of the non-reciprocal absorption shift (recall Fig. 2.22). From the viewpoint of magnetic anisotropy it is therefore advisable to keep the magnetic contact thickness as thin as allowed by the waveguide design!

[36] A very low shearing angle requires a very small demagnetization factor, or a very large inverse demagnetization factor

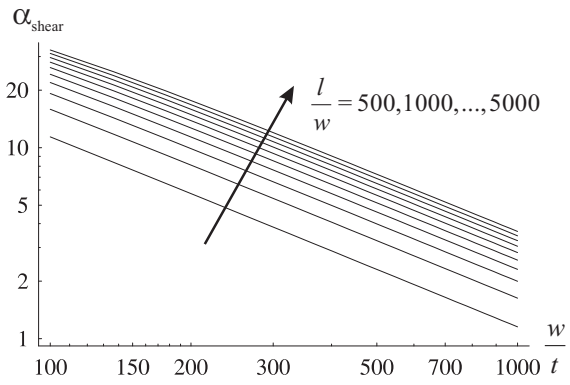


Fig. 3.8: The shearing angle as a function of w/t for a various values of l/w.

With contact thicknesses of the order of 50-100 nm, it is thus advisable to have a contact width of about 25-50 μm. In the next chapter, experimental results will be given that confirm these observations. Note also that the formulae (3.34) have been indicated to be only valid in a twodimensional approximation of the magnetostatic problem of the magnetized contact and this approximation was only justified for situations were the magnetic field was not varying significantly over the thickness of the rectangular shape. Increasing w/t will therefore not violate this approximation.

3.1.3 Ferromagnetic p-type Ohmic contacts

3.1.3.1 Introduction

The ferromagnetic metal contact in the SOA/isolator device fulfills two basic functions. It is the source of the non-reciprocal magneto-optic effect on which the whole operation principle of the device relies. Next to that, its metallic properties are used to bias the underlying amplifying waveguide. In this context it should be noted that this electrical function of the ferromagnetic metal is not an essential feature of the concept for non-reciprocity integration. Indeed, the basic idea is only to use the large non-reciprocal loss shift introduced by a magneto-optic ferromagnetic metal and to compensate the residual optical losses introduced by this metal by an optical amplifying structure. Viewed in this way, it is clear that the needed electrical interface for the amplifier is fully independent of the source of the non-reciprocal loss effect. Using the magneto-optic metal also as the electrical contact for the amplifier is clearly not the "conditio sine qua non". One can formulate both pro- and counter-arguments for this choice.

- **Pro** From a geometric viewpoint the symmetrical placing of the magneto-optic region with respect to the TM mode will maximize the non-reciprocal absorption shift. Next to that the fabrication scheme is at its simplest. It is basically a SOA processing scheme with just a more "exotic" choice for the

metal electrical interface.

- **Contra** The need to achieve a low-resistivity ohmic contact inevitably requires a strongly doped low-bandgap semiconductor layer (as will be seen later on). In the standard InP material system the lowest-bandgap alloy that is lattice-matched to InP is the ternary $In_{0.53}Ga_{0.47}As$ with a gap of 0.75 eV. The optical absorption in this layer for the telecom wavelength ($h\nu > 0.78$) will not be negligible (see for instance [38]). As a result the magneto-optic ferromagnetic contact will be partially shielded. Next to that, if the electrical contact also provides the magneto-optic effect, the semiconductor cladding layer of the amplifier must be reduced to obtain a sufficient magneto-optic perturbation of the TM mode profile. However due to the plasmonic enhancement near the metal of the evanescent TM waves in the cladding, the transverse (vertical) modal profile of the TM mode will be strongly asymmetric. It is clear that this will cause a big transition loss at the interface between the integrated SOA/isolator and e.g. the DFB laser diode (or any active photonic component for that matter). In Chapter 6 alternative schemes will be indicated where the MO function and the electrical contact interface can be separated, resulting in much lower modal asymmetry and avoiding the contact layer shielding of the MO region.

In any case, it is clear that an eventual electrical ohmic function for the ferromagnetic metal has to be of subordinate importance as compared to its primary non-reciprocal function. If it should become clear that using the ferromagnetic metal as an electrical contact requires a compromise choice for the magneto-optic material, one must consider the separation of the non-reciprocity and the electrical function. Therefore, instead of looking for the "best" Ohmic contact material with also ferromagnetic, magneto-optic properties, it was chosen to simply characterize the contact resistivity of the chosen CoFe[37] alloy system. The remaining "tuning" parameters to minimize the contact resistivity are then the contact processing parameters (more in particular the annealing conditions) and the properties of the semiconductor contact layer (that is compatible with the boundary conditions imposed by the non-reciprocal waveguide design).

3.1.3.2 Short review of the theory of metal-semiconductor (Ohmic) contacts

It is well known that the junction between a metal and a semiconductor will in general form a (unipolar) Schottky diode, whose saturation current greatly depends on the barrier height of the Schottky barrier formed at the interface between metal and semiconductor. Ideally, an Ohmic contact is a Schottky diode whose operation does not perturb the performance of the device for which it used as a contact. In other words, it has to supply the required operation current to the "underlying device" with a voltage drop that is small compared with the drop across the

[37] on the basis of its MO strength

active region of the device[38]. This can be achieved by maximizing the saturation current of the Schottky diode, as the specific contact resistivity is given by[39] $\rho_c \equiv \left(\frac{\partial J}{\partial V}\right)^{-1}_{V=0} = \frac{kT}{qJ_S}$. In order to maximize the saturation current density J_S one has to consider the fact that current transport through a metal-semiconductor contact can have a large contribution of carriers tunnelling through the barrier [39]. This is opposed to the standard pn-junction, where the tunnelling mechanism is only important in the reverse direction near junction breakdown. The reason for the higher relative importance of the quantum-mechanical tunnelling component in Shottky diodes as compared with pn-diodes is obviously the extremely short lifetime of the carriers in the metal due to thermalisation. As a result a majority carrier that can tunnel through the barrier doesn't need to diffund over a large distance in the metal (which is precisely the current limiting process in a pn-junction!)

It is easy to see that tunnelling becomes more important as the dopant concentration in the semiconductor increases. Increasing the dopant concentration will decrease the thickness of the space charge layer, and hence the "thickness" of the barrier, making it increasingly easier for carriers at the base of the barrier to tunnel through it. This important extra current component will make the saturation current rise considerably and hence make the specific contact resistance drop. On top of that, the quantummechanical tunnelling probability is independent of temperature [40], so that the saturation current density will become increasingly temperature-independent as the barrier gets narrower[40]. A high, temperature-independent saturation current density is an obvious expression of increased ohmic behaviour of the contact. This transition from pure thermionic diode behaviour to tunneling-dominated Ohmic behaviour is illustrated in Fig. 3.9(a).

The relative importance of the tunnelling versus the thermionic current as a function of the dopant concentration and the temperature is illustrated in Fig. 3.9(b). This proves how the relative importance of the tunnelling component increases very quickly at low doping concentration and at higher concentrations (when it becomes the dominant conduction mechanism) it gradually saturates. Actually, in [41] it is proven that the tunneling current increases exponentially as $\sim \exp(-C/\sqrt{N})$, with C a constant depending on the barrier height[41] and on other material parameters such as the effective mass of the majority carriers, and with N the dopant concentration.

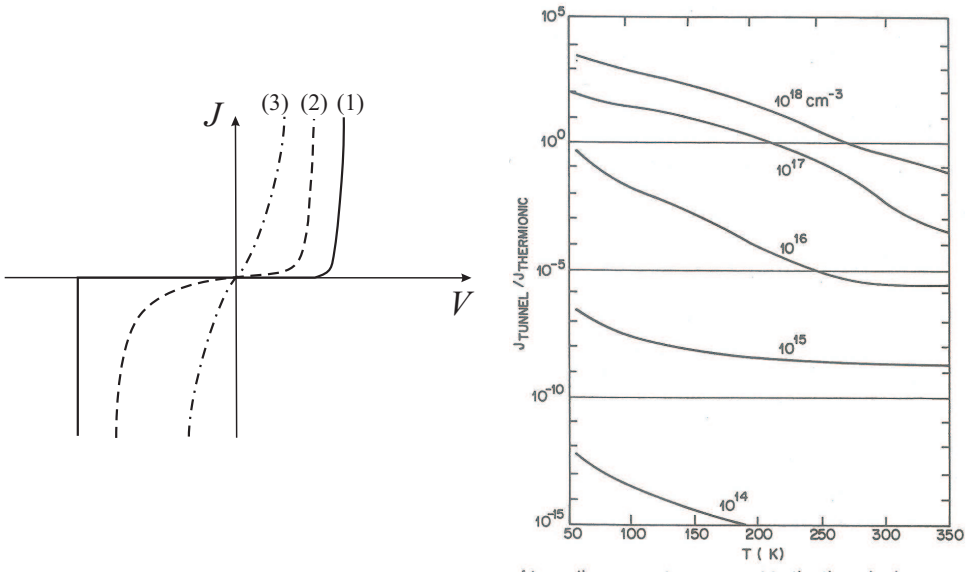
Maximizing the saturation current density (and hence minimizing the specific contact resistivity) is thus mainly achieved by increasing the dopant concentration to levels of 10^{+19} cm^{-3} or higher. One might remark that lowering the Schottky barrier height as much as possible is another way of achieving good Ohmic properties. The saturation current density at zero applied voltage through a Schottky diode is

[38] It is a misconception that an Ohmic contact must have a linear current-voltage relationship. All that is required is an approximate linearity over a voltage region spanning a large enough current region. It is not difficult to see that this can always be achieved with the intrinsic nonlinear diode relationship provided the diode has a very large saturation current density

[39] if we neglect the voltage dependence of the saturation current

[40] because then all thermic excited carriers in the conduction or the valence are tunnelled with the same probability.

[41] which has only a very small dependence on the dopant concentration



(a) (b)

Fig. 3.9: (a) Schematic illustration of the current-voltage relationship of a Schottky barrier contact for progressively higher dopant concentrations.(1): Bethe-Shottky thermionic-diffusion conduction regime, (2): thermionic-field-emission conduction regime (i.e. tunnelling from carriers at the top of the barrier), (3): field-emission conduction regime (temperature-independent tunnelling throughout the whole barrier).(b) Ratio of tunnelling current component to the thermionic current component of a Au-Si barrier (after Chang and Sze [41]). It is seen how the tunnelling current gradually starts to dominate as the doping concentration increases.

after all inversely exponential proportional to this barrier height (no matter what conduction mechanism is dominant). Indeed, in [39] it is proven that the specific contact resistivity in the tunneling regime is given by

$$\rho_c \sim \exp\left[\frac{2\sqrt{\epsilon_s m^*}}{\hbar}\left(\frac{\phi_B}{\sqrt{N}}\right)\right] \tag{3.37}$$

with ϕ_B the barrier height (in V), m^* the effective mass of the majority carriers, ϵ_s the relative permittivity of the semiconductor and N the dopant concentration. The barrier height in general is given by [42]

$$\phi_m - \Delta = \chi + \phi_{Bn} \qquad \text{for n-type contacts} \tag{3.38a}$$

$$= \chi + E_g - \phi_{Bp} \qquad \text{for p-type contacts} \tag{3.38b}$$

with χ the electron affinity of the semiconductor, ϕ_m the metal work function, E_g the semiconductor band gap and Δ the electron energy increase when crossing the

interfacial layer from the semiconductor into the metal due to the induced charges at the metal surface (caused by the surface charges and the space charges in the depletion region of the semiconductor). (3.38) proves that for n-type contacts the barrier is minimized when the metal work function is minimal and when the donor concentration is very high (inducing a very high negative metal surface charge and thus a strong energy "jump" Δ). While for p-type contacts the Schottky barrier is minimized when the metal work function is maximized, the electron energy jump at the surface is minimized (preferably becoming negative), implying a strong acceptor concentration in the semiconductor leading to a strongly positive metal surface charge density, and when the band gap of the semiconductor is minimal. For these p-type contacts in particular this proves that it is far from easy to make a good Ohmic contact on a large band gap semiconductor, as this would require according to (3.38b) a metal with a work function that must counteract both the electron affinity of the semiconductor ánd its band gap. While (3.38a) shows that for n-type contacts the metal work function only needs to counteract the electron affinity of the semiconductor. With metal work functions reaching a maximum of 4-5 eV (Pt, W, Au) and electron affinities in the same order of magnitude for most III-V semiconductor compound, this explains why it is much more difficult to make p-type contacts than n-type Ohmic contacts. However, all of the above observations have been made without considering the effect of the interfacial potential jump Δ (caused by the induced metal surface charges). It was remarked that a strong dopant concentration[42] with the accompanying space charge density in the semiconductor will induce a high metal surface charge density and will thus increase Δ. An other and often much stronger cause for the induced charges at the metal surface are the occurrence of semiconductor surface states. These surface traps are caused by the interruption of the perfect crystal structure at the surface of the semiconductor and can have such a large density in the band gap of the semiconductor that they pin the Fermi surface of the metal-semiconductor contact[43]. Bardeen was the first to suggest this Fermi level pinning, when he observed that the Schottky barrier of some metal-semiconductor contacts was almost independent of the used metal [43]. Later it was proven by Kurtin [44] that the amount of Fermi level pinning is a function of the covalent character of the semiconductor material. Perfectly covalently bound semiconductors exhibit a pinned p-barrier height of one third of their band gap (independent of the work function of the metal used), while perfectly ionic semiconductors have an unpinned Fermi level and thus a barrier height that is fully determined by the metal work function. This is nowadays commonly referred to as the 1/3 rule of Bardeen.

All of the above makes clear that for the standard III-V semiconductor systems (with their high covalent character), the barrier height is pinned and cannot be "engineered" to lower the specific contact resistivity (along (3.37)). The Ohmic character has to be obtained solely by achieving a dopant concentration as high

[42] for which it has already been proven above that it causes tunneling to become the dominant current mechanism, and thus helpful in achieving an Ohmic contact.

[43] What happens in fact is that the high density of acceptor surface states has such a large associated capacitance that it completely determines the charge neutrality conditions at the interface, causing the band bending to become independent of the used metal.

as technologically possible. Within this context it has to be remarked that high acceptor (p-type) dopant concentrations are much more difficult to achieve due to their tendency to occupy interstitial positions. These interstitials have a high mobility, hence causing the dopants to easily diffund and consequently making it difficult to achieve large concentrations. Having indicated what the theoretical and practical subtleties of p-type Ohmic contacts are, we now turn our attention to the practical details of the characterization of the specific contact resistivity.

3.1.3.3 Characterization of contact resistivity

The measurement of a resistance requires a voltage/current ratio. This is not different for the characterization of an Ohmic contact. However, the resistance of any resistor structure such as a processed metal Ohmic contact is geometry dependent. The specific contact resistivity is obviously a geometry-independent measure. Therefore, a good characterization method must be able to model the voltage distribution beneath the contact, taking the specific geometry of the contact into account and relate this distribution to the geometry-independent specific contact resistivity. In general, such a model should be three-dimensional and be based on the traditional Poisson equation, describing the potential variation both in the metal and the semiconductor, and two carrier continuity equations, one for the minority and one for the majority carriers. It can be proven that this will lead to a Helmholtzian partial differential equation for the potential, subjected to a boundary condition at the metal-semiconductor interface expressing the linear relationship between the injected current density and the local potential difference at the contact [45]. It is clear that this most general description is too detailed for the modelling of practical contact test structures, for which the contacts are most often shallow, planar structures in which the current flow is in good approximation two-dimensional. In the following we describe how and under which reasonable assumptions a sufficiently accurate two-dimensional contact model can be derived. Subsequently, it will be indicated how this model can be used to extract the geometry independent contact resistivity.

Model for the static current flow in finite-size contacts The contact structure is assumed to be planar. Fig. 3.10 gives a schematic representation of the shallow "slab"-like contact characterization structure. The current density J flowing through the Ohmic contact is linearized through a Taylor expansion and given by

$$J(v_{ms}) = \left.\frac{\partial J}{\partial v_{ms}}\right|_{v_{ms}=0} v_{ms} = \frac{v_{ms}}{\rho_c} \tag{3.39}$$

with ρ_c the specific contact resistivity (usually expressed in Ωcm^2), and v_{ms} the potential difference between the Fermi levels in the metal and the semiconductor. It is assumed that the contribution of the minority carriers to the current flow can be neglected. This is equivalent to neglecting minority recombination effects in the semiconductor diffusion layer. In other words, this is equivalent to neglecting the extension of the depletion layer and assuming the diffusion length of the minority carriers to be much larger than the thickness of the semiconductor diffusion

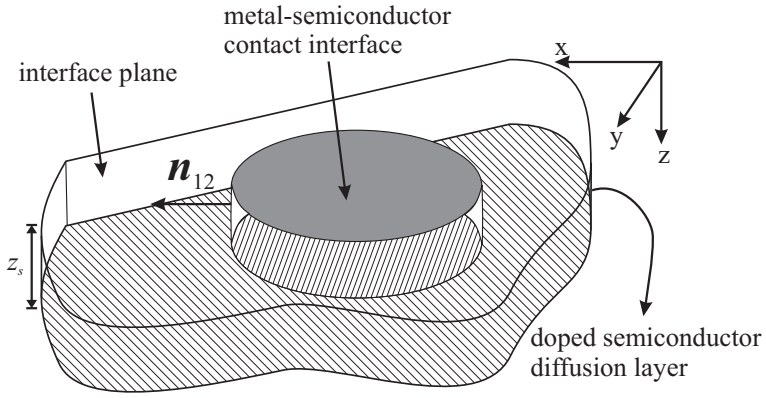


Fig. 3.10: Planar Ohmic contact test structure. The strongly doped semiconductor has a thickness z_s. The plane at $z = 0$ is the interface plane containing both the metal-semiconductor interface and the interface between the semiconductor and the isolating material next to the metal, usually polyimide or SiO_2. The metal geometry is arbitrary. Usual test structures will however be square or rectangular. The unit vector $\boldsymbol{n}$ is the normal vector between the two regions with possibly different conductivities. This is for instance the case if the Ohmic contact is selectively alloyed, having only large doping concentration below the metal.

layer in the test structure. This is not a strong limiting factor for the validity of the following model, as for Ohmic contacts the depletion is always very thin due to the need for very high doping levels (order a few nm). And the thickness of the semiconductor contact layer used in the test structure is of the order of 100 nm while the diffusion length of n- and p-type carriers in III-V semiconductors is usually several μm. This implies that all electrostatic effects due to the space charge region of the Schottky-like contact (and their influence on the semiconductor potential) are artificially bundled in a discontinuous step at the interface itself. A second implication of this assumption is that quasi-neutrality is fullfilled throughout the semiconductor diffusion layer and that there are no recombination terms in the majority carrier continuity equation. The basic equations that govern the current flow in the planar test structure are thus given by

$$\boldsymbol{\nabla} \cdot \mathbf{J} = 0, \tag{3.40a}$$

$$\mathbf{J} = -\sigma \boldsymbol{\nabla} v, \quad \text{and} \tag{3.40b}$$

$$\boldsymbol{\nabla}^2 v = 0 \tag{3.40c}$$

With conductivities that are stepwise constant, the last equation follows straightforwardly from the first two. The boundary conditions for the above complete 3D model are, at the contact interface using (3.39),

$$-\mathbf{u}_z \cdot \boldsymbol{\nabla} v = -\frac{v}{\rho_c} \tag{3.41}$$

(using the coordinate system in Fig 3.10). Here it was assumed that the metal has such a large conductivity that it will be an equipotential volume for which the potential is set at 0. At the boundaries between two regions with different conductivities, such as the boundary between semiconductor regions with different dopant concentration depicted in Fig. 3.10, the conditions are the continuity of the potential and of the normal current density.

$$v_1 = v_2 \tag{3.42a}$$

$$-\sigma_1 \mathbf{n}_{12} \cdot \boldsymbol{\nabla} v_1 = -\sigma_2 \mathbf{n}_{12} \cdot \boldsymbol{\nabla} v_2 \tag{3.42b}$$

The last boundary condition can in good approximation be simplified to the disappearance of the normal current density for boundaries between regions with strong discontinuities in the conductivity, as will be the case at the isolation interface and the interface with the semi-insulating or moderately doped substrate.

Two-dimensional model This rigorous 3D description of the current flow and the potential distribution in the contact characterization structures is only necessary when the semiconductor diffusion region has a considerable thickness, such as in solar cells, or when the contact interface and/or the diffusion layers are not planar. In practice, the diffusion layer will be shallow and the current flow will be mainly a planar 2D phenomenon (except for the discontinuous "vertical" injection at the metal itself). The transition towards a 2D model is then traditionally done by averaging out the z-dependent effects and lumping them together in a single resistance parameter, known as the sheet resistance (also known as the "resistance-per-plane-square") R_s. Assuming a conductivity for the semiconductor independent of the spatial variables x and y, $\sigma = \sigma(z)$, the constant sheet resistance is defined as[44]

$$R_s = \frac{1}{\int_0^{z_s} \sigma(z) \mathrm{d}z} \tag{3.43}$$

Next to a 2D equivalent for the conductivity, a new variable is needed to effectively describe the potential in 2 dimensions. This is the conductivity weighted average potential, defined as

$$V = V(x,y) = R_s \int_z \sigma(z) v(x,y,z) \mathrm{d}z \tag{3.44}$$

[44]Note that if the conductivity were not independent of the planar variables x and y, the sheet resistance could still be defined in the same manner. This would then lead to a sheet resistance that is dependent on the position in the diffusion layer. This is hard to imagine (unless of course something went terribly wrong in the dopant diffusion leading to very large non-uniformities in the dopant profile). It has been suggested by Reeves and Harrison [46] that even with perfect uniformity, the sheet resistance beneath the metal can differ from the sheet resistance next to the metal contact, due processes such as sintering and alloying of the contact. However, the accuracy with which this change in sheet resistance can be measured is questionable, moreover Loh proves in his seminal paper on contact modelling [45] that a change of e.g. 200%, will only induce a relative change in the measured contact resistance (not to be confused with ρ_c) of less than 10% in the worst case. The question remains of course what implication a change of 10% in contact resistance would have on the actual extracted specific contact resistivity. This is strongly dependent on the used test structure and the actual value for ρ_c. Loh proves that the "best" (i.e.) the most tolerant, is the Cross Bridge Kelvin Resistor. This is another argument for using CBKR's instead of TLTR's, on top of the arguments that will be given later.

With these two definitions the majority carrier continuity (3.40a) equation becomes

$$-\boldsymbol{\nabla}_t^2 \sigma(z) v(x,y,z) + \frac{\partial J_z}{\partial z} = 0 \tag{3.45}$$

Integrating along the z-axis and using the definitions (3.44) and (3.39), we obtain

$$-R_s^{-1} \boldsymbol{\nabla}_t^2 V(x,y) = J_z(0) - J_z(z_s) \tag{3.46}$$

$$\boldsymbol{\nabla}_t^2 V(x,y) = -R_s J_z(0) = \frac{R_s}{\rho_c} v(x,y,0) = \frac{R_s}{\rho_c} V(x,y)\Gamma \tag{3.47}$$

with $\Gamma = \frac{v(x,y,0)}{V(x,y)}$ the ratio of the actual potential jump at the interface at the position (x,y) to the weighted potential. Assuming a shallow junction, Γ will be close to 1. It was also assumed in (3.46) that the normal (z-)current density, $J_z(z_s)$ at the lower interface of the semiconductor diffusion layer is negligible (due to the large difference in conductivity between the strongly doped diffusion layer and the moderately doped or semi-insulating substrate). Defining the transfer length of the majority carriers as $l_t = \sqrt{\rho_c / R_s}$ we obtain the basic equation describing the 2D current flow

$$\boxed{\boldsymbol{\nabla}_t^2 V = \frac{V}{l_t^2}} \tag{3.48}$$

This is the celebrated Helmholtz equation describing the (averaged) 2D potential distribution beneath the metal contact. The transfer length l_t appearing in this equation is the paramount parameter in contact characterization, and can be seen as a measure of the distance which the current travels in the diffusion layer before entering the contact. In the other regions of the diffusion layer (i.e. next to the metal contact), the right-hand side of (3.46) will, after integration, contain two zero current densities, as there is no vertical current injected at $z = 0$ for these regions and there is as before no vertical current flow into the substrate. The Helmholtz equation in those regions reduces to the standard sourceless Poisson equation.

The boundary conditions for this 2D model are the "dual" translations of (3.42) in the case of shallow diffusion layers[45]. Thus, the weigthed potential is continuous everywhere, and the 2D current is continuous at the boundaries between regions with different sheet resistance. This can be proven by using the definition of the weighted potential (3.44)[46].

$$V_1 = V_2 \tag{3.49a}$$

$$\frac{\mathbf{n} \cdot \boldsymbol{\nabla}_t V_1}{R_{s1}} = \frac{\mathbf{n} \cdot \boldsymbol{\nabla}_t V_2}{R_{s2}} \tag{3.49b}$$

[45] Of course there are no more boundary conditions at the interface plane itself! Since the z-dependence has been integrated away this boundary conditions is already implicitly enclosed in the right-hand side of (3.46)

[46] The first of these boundary equations is clearly not exact, as the value of the weighted potential strongly depends on the thickness of the diffusion region. Therefore, imposing the continuity of the weighted potential between two regions with a different conductivity is only allowed for regions of limited thickness.

The total current flowing in the contact characterization structure can easily be calculated as

$$I_{\text{tot}} = \frac{1}{R_s} \int \boldsymbol{\nabla}_t V \cdot \mathrm{d}\mathbf{l}_n \tag{3.50}$$

where $\mathrm{d}\mathbf{l}_n$ is the infinitesimal line element of any line in the $x-y$ plane across which all the current must pass, and whose vector is pointing perpendicular to this line (in the direction opposite to the current flow). This formula is easily derived by considering the complete 3D formula ($\iint \sigma(\boldsymbol{\nabla} v) \cdot \mathrm{d}\mathbf{A}$) expressing the total current through a surface. If this surface is taken parallel to the $z-$axis (i.e. $\mathrm{d}\mathbf{A}$ has no $z-$component) and the integration along the $z-$axis is performed first, then the expression of (3.50) for I_{tot} is obtained (using also (3.44)) This concludes the 2D model.

One-dimensional model In certain cases where the contact has an elongated (rectangular) geometry, the potential variation along the $y-$direction can be neglected and the Helmholtz equation reduces to its 1D form

$$\frac{\partial^2 V}{\partial x^2} = \frac{V}{l_t^2} \tag{3.51}$$

This is straightforwardly integrated to

$$V(x) = V_i \frac{\cosh\left(\frac{l-x}{l_t}\right)}{\cosh\left(\frac{l}{l_t}\right)} \tag{3.52}$$

with l the length (i.e. the short(!) dimension) of the contact, and V_i the potential at the leading edge ($x = 0$) of the current collecting contact. The boundary condition at the trailing edge is of course the absence of a planar current. The total current in this model is the straightforward integration of (3.50)

$$I_{\text{tot}} = -\frac{W}{R_s} \left.\frac{\partial V}{\partial x}\right|_{x=0} \tag{3.53}$$

with W the width (i.e. the long dimension) of the contact. If the total current collected by the contact is known, then the above boundary condition can be used in a Neumann sense to express the potential in terms of I_{tot}. This leads to $V_i = \frac{l_t}{\tanh(l/l_t)} \frac{R_s I_{\text{tot}}}{W}$ in (3.52). Outside the contact the one-dimensional equation for the potential again simply reduces to $\frac{\partial^2 V}{\partial x^2} = 0$, i.e. Ohm's law for static currents, namely a linearly increasing (or decreasing) potential.

Specific contact resistivity extraction The above model for the current flow in contact characterization structures makes clear that the extraction of the contact resistivity gets down to measuring the transfer length l_t, and the semiconductor diffusion sheet resistance R_s (in $\Omega/\square$). Indeed, if these 2 parameters can be measured then ρ_c is found as $l_t^2 R_s$. The measurement of R_s is not that difficult. It

can be derived straightforwardly out of the slope of a measured resistance between electrical contacts of varying spacing (see Fig. 3.12). However, one should be aware that this gives the sheet resistance between the contacts, while the sheet resistance in the Helmholtz equation (3.48) is the one that applies beneath the contact. As pointed out in a footnote above, we are going to neglect a possible change in this sheet resistance.

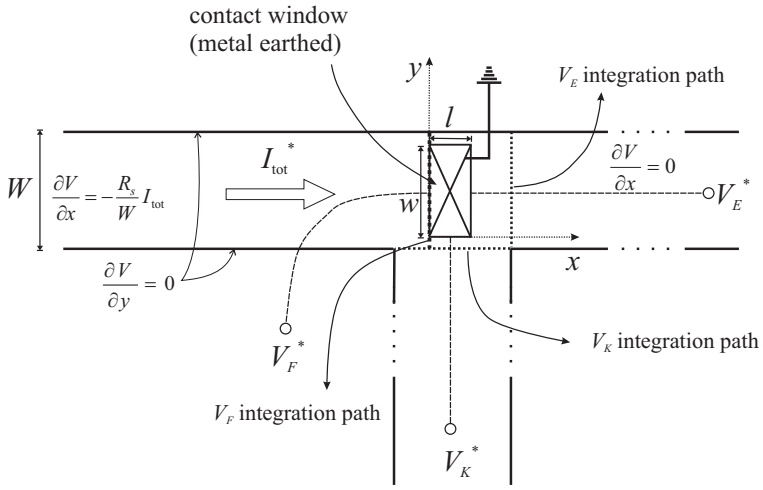


Fig. 3.11: Schematic illustration of the measurement of the transfer length l_t with the boundary conditions needed for the 2D numerical analysis. The three possible configurations for the contact resistance measurements are schematically indicated, these are respectively the front resistance of the contact, the end resistance and the side resistance (commonly known as the Kelvin resistance [47]). In practice these resistances can only be measured if the diffusion layer is "extended" in the measurement direction as indicated. This also implies that the front resistance cannot be measured directly, but has to be extracted in an implicit manner (see Fig. 3.12).

The measurement of the transfer length is a bit more complicated and strongly depends on the model used for the current flow in the characterization structure. The basic measurement principle is sketched in Fig. 3.11. A rectangular (or square) contact geometry is assumed, and a known static current $I^*_{\rm tot}$ is collected by this contact. The potential distribution accompanying this current flow is either described by (3.51) or (3.48) (and the accompanying boundary equations) depending on whether it can be assumed that the current flow around (or into) the contact is essentially a 1D or a 2D phenomenon. The idea is to measure the contact potential jump V^* either at the leading, the trailing or the side edge of the contact, and measure in this way "some" contact resistance, $R^* = V^*/I^*_{\rm tot}$. This same resistance can be calculated with the 1D or 2D current model and this simulated resistance, $R_{\rm sim}$, will obviously be a function of the transfer length l_t

(and the known geometric parameters of the characterization structure). Varying l_t as a fitting parameter R_{sim} can be made to converge towards the measured R^*. This procedure results finally in an extraction of the contact resistivity via the definition of the transfer length, $\rho_c = l_t^2 R_s$.

The main difference between the 1D and 2D model in this extraction procedure is that the 1D model allows analytic expressions for the different contact resistances. As such it allows an analytic expression for the transfer length as a function of the geometric parameters, the sheet resistance and of the measured resistance.
In the 2D model, one inevitably has to resort to numerical techniques to solve the two-dimensional boundary value problem described by (3.48) and the boundary conditions (3.49). In these boundary equations, the possibility for different sheet resistances next to and beneath the contact is neglected. These means that the boundary conditions at the four sides of the contact window will simply be the continuity of the potential and its normal derivative. Note that at the far-left side of the structure depicted in Fig. 3.11, one can use a Neumann boundary condition $\frac{\partial V}{\partial x} = -\frac{R_s}{W} I^*_{\text{tot}}$, expressing that the unknown potential distribution has to sustain the known DC current I^*_{tot} in the direction of the contact. At the far-right side there should be no more current flowing in the diffusion layer, leading to the boundary condition[47] $\frac{\partial V}{\partial x} = 0$. The last boundary condition is of course the absence of any current perpendicular to the etched sidewalls of the highly doped semiconductor diffusion layer. With these boundary conditions the potential problem of Fig. 3.11 can be solved completely and the simulated resistance (which is a function of l_t) can be fitted to the corresponding measured resistance $\frac{V_i^*}{I^*_{\text{tot}}}$ (with $i = F, E, K$)

$$R_i^* \longleftrightarrow R_{i,\text{sim}}(l_t) = \frac{V_{i,\text{sim}}(l_t)}{I^*_{\text{tot}}} \qquad \text{with } i = F, E, K \tag{3.54}$$

The simulated potential value to be used in the above formula is the average value over the width of the measurement arm taken along the integration paths as indicated in Fig. 3.11[48].
If a 1D model applies (for instance because the contact (almost) covers the entire width of the diffusion arm or because the contact resistivity is expected to be so large that there will be no spreading of the current around the contact), then $R_{i,\text{sim}}(l_t)$ can be calculated analytically. (3.52) gives the following results for the

[47] Note that this boundary condition must be imposed "far" enough from the contact, in order to allow the possibility for current crowding around the contact (occurring when the contact resistivity is really low).

[48] Note that the exact position of the integration paths for the Kelvin voltage and the end voltage is not that crucial, as long as they enclose the entire width of the sensing arm, since for these two there will be no current flow in the sensing arm. For the front potential the situation is obviously different because the simulation sensing arm and current feeding arm coincide. If the integration path is not taken exactly at the leading edge of the contact window, there will be an extra contribution to the simulated potential from the potential drop in the feeding arm

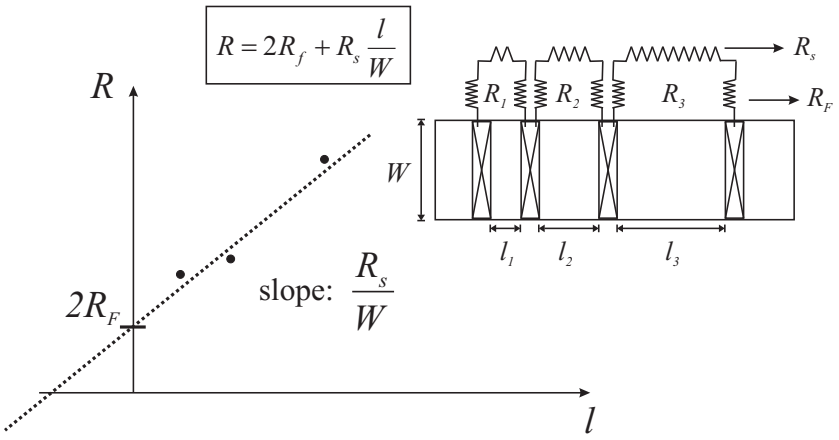


Fig. 3.12: Schematic representation of the Transmission Line Method.

three different configurations

$$R_F = \frac{V(0)}{I^*_{\text{tot}}} = \frac{R_s}{W}\frac{l_t}{\tanh(l/l_t)} \tag{3.55a}$$

$$R_E = \frac{V(l)}{I^*_{\text{tot}}} = \frac{R_s}{W}\frac{l_t}{\sinh(l/l_t)} \tag{3.55b}$$

$$R_K = \frac{\int V(x)\mathrm{d}x}{lI^*_{\text{tot}}} = \frac{R_s}{Wl}l_t^2 = \frac{\rho_c}{Wl} \tag{3.55c}$$

The last equation is obtained by averaging (3.52) over the width of the contact (i.e. assuming that the Kelvin voltage arm has the same width as the contact). The first two expressions are obviously obtained without an averaging integration as the 1D model assumes a y-independent potential. These formulae can be used directly to calculate l_t once a contact resistance measurement has been made. Of course their validity has to be looked upon with caution.

A final remark has to be made regarding the measurement of the front potential. It is clear from Fig 3.11 that this potential can not be measured directly, as the sensing arm would coincide with the current feeding arm. The only way to measure the front potential (and the corresponding front resistance) is through a linear extrapolation procedure in which the total resistance between contacts with a varying spacing is plotted as a function of this spacing. This is the so-called Transmission Line Method (TLM) [48]. It is historically the first method devised to measure contact resistivities. Its principle is sketched in Fig. 3.12. It is clear that each measured resistance is a series circuit of twice the unknown front resistance and the diffusion resistance obtained as the diffusion sheet resistance multiplied by the number of squares between two consecutive contacts, l/W. This linear relationship between the measured resistances and the contact spacing allows the extraction of (twice) the front resistance as the resistance that hypothetically

would be measured if the consecutive contacts were touching. For this purpose one needs at least three contacts, however it is clear that the accuracy of the needed linear regression will increase with an increasing number of contacts. This is not a vain remark. One of the drawbacks of the TLM is that it is based on the addition of a large (R_s) and a small resistance (R_F). As a result the the line in Fig. 3.12 will have a rather steep slope and a small value for the intersection with the R-axis. Small errors in the knowledge of the contact spacing will result in a large uncertainty on the front resistance.

Loh has made a thorough comparison of the three basic planar contact characterization structures [45], observing that the front resistance characterization structure (TLM) is the least influenced by 2D parasitic current crowding effects, while the end resistance is for obvious reasons the worst. That is because, contrary to the 1D model, the 2D current crowding around the contact shifts the potential minimum inside inside the contact. This results in the small end potential being dominated by much larger parasitic parasitic effects and by high sensitivities to geometrical parameters. Sadly enough the end resistance structure is the simplest from a technological viewpoint. The TLM on the other hand, even though much more robust, is also not ideal as it depends on an indirect measurement technique where a small number needs to be filtered out of a data set of large numbers. The Kelvin structure is the best compromise between moderate sensitivity to parasitics and ease of measurement, at least when it is combined with numerical tool allowing 2D simulations. We have developed a simple MATLAB tool that solves the 2D boundary value problem (3.48) using finite differences, and have used it to extract the contact resistivities of the processed CoFe contacts. The experimental results of these contact resistivity measurements and the details of the investigated contacts will be given in the next Chapter.

3.1.3.4 P-type Ohmic contact – reported performances

Before concluding this section on Ohmic contact characterization, it is worthwhile to give a very short review of existing p-contact technologies and their performances. This will allow an assessment of the quality of the processed CoFe contacts.

Unlike Schottky barriers, the subject of metal-semiconductor Ohmic contacts has experienced limited theoretical scientific interest. This is probably, at least in part, due to the unexciting performance of a satisfactory Ohmic contact. This doesn't make these contacts less important of course. But it is probably a fair assessment that the Ohmic contact technology has developed more as a technical art than as a theoretical science. An art which is characterized by continuous trial and error attempts to lower contact resistivities by changing the composition (and thus the band gap) of the semiconductor contact layer, changing its doping levels and dopant types, changing the choice of metals or metal alloys, changing deposition conditions of these metals, varying annealing conditions of the contacts, ... All of this has led over the last twenty years to a multitude of publications reporting on experimental realization of Ohmic p-type contacts with an enormous variety of metallic materials, semiconductor dopants, deposition and annealing

conditions, ...

However when it comes to the traditional InP/InGaAsP material system for 1.3/1.55 optical telecom applications, a certain favorite contact technology can be distinguished throughout all of these publications, with only slight and subtle differences between the precise processing techniques and the metals used. Even more important is that the best performing ohmic contacts all show contact resistivities with values in the region between several $10^{-5}\Omega\text{cm}^2$ and down to a lower record value of $10^{-8}\Omega\text{cm}^2$ [49].
We mention the following results. The best performing contacts all make use of doping levels higher than 10^{+19}cm^{-3}. Furthermore, by making use of the 1/3 rule of Bardeen [43], the contact resistivity can be lowered by using a semiconductor contact layer with a band gap as low as possible. For the InP/InGaAsP material system InAs has the lowest band gap (0.35 eV), while GaAs has the highest band gap (1.42 eV). InAs is however not lattice-matched to InP and can therefore not be used as a contact material as it would contain an unreasonably high number of crystal defects. As a result, one normally makes use of lattice-matched $In_{0.53}Ga_{0.47}As$, and exceptionally of quaternary lattice-matched InGaAsP [50]. The popular choice of metals for the contact is most often a combination of a noble metal (Au, Pt and Ti), both as a protective measure and a stabilizing layer, and an underlying metal showing proper p-dopant properties for InGaAs (such as Zn, Be, Sn and Cr), so that possible indiffusion of this dopant through annealing can increase the doping levels of the contact layer. However, advances in epitaxial growth over the last decade have made very high doping levels of $> 10^{19}$ cm^{-3} readily available in epitaxial growth. Therefore nowadays there is no more need for using p-dopant-like metals, and most often one resorts to using those metals which show the lowest work function for InGaAs. Ti and W are a very popular choice here.
Finally, the best results for the contact resistivity are obtained through a subsequent rapid thermal annealing procedure (RTP) of the contacts. Katz et al. [51] have postulated that this is a result of thermally stimulated metallurgical conversions from wide-gap into small-gap interfacial compounds.

Table 3.2 gives an overview of the performance of what can be considered as the different "standard"ohmic contacts used in III-V photonic devices. The last two references in Table 3.2 are an example of how self-aligned p-type ohmic contacts can be formed by thermally assisted local diffusion of p-dopants into very low-doped semiconductor layers. From a historical point of view, this technique was the first to be used to form Ohmic contacts using almost always Au-based alloys. Nowadays Au is commonly replaced by metals like Pd, W, ... because of their higher thermal stability, as Au is known to easily diffuse already at moderate temperatures.

Bearing in mind that a well-behaving active InP opto-electronic device usually exhibits series resistance values of the order of 10-20 Ω, a resistance value of maximum 1 Ω for the contact can be considered acceptable. In a zero-order approximation this is equivalent to specific contact resistivities of $10^{-5}\Omega\text{cm}^2$ (considering devices of 1000 μm^2 dimensions).

Ref.	Metal (nm)	Contact layer	Dopant (10^{19}cm^{-3})	ρ_c ($10^{-8}\Omega$cm^2)	RTP (°C)
[50]	Pt/Ti (75/50)	InGaAsP ($\lambda_g = 1.3\mu$m)	Zn (0.5)	36000	450
[51]	Pt/Ti (60/50)	InGaAs	Zn (0.5)	300	450
[49]	Pt/Ti/Au (50/100/200)	InGaAs	Zn (4)	23	400
[52]	W/Mn/Ti/Mn (12/15/50/100)	InGaAs	Be (2)	270	
[53]	PdZn alloy	InGaAsP ($\lambda_g = 1.14\mu$m)	Zn (0.2)	30	400
[54]	WZn alloy	InGaAs	Zn (< 0.1)	500	500

TABLE. 3.2: Selection of reported "standard" Ohmic p-contacts to the InP/InGaAsP material system.

3.1.4 Conclusion

This concludes the extensive section on the subtleties of the ferromagnetic metal contact for use in the SOA/isolator. Each of the three key functions (non-reciprocal anisotropic material, permanent thin film magnet and electrical Ohmic contact) that this metal layer has to fulfill has been studied both from a phenomenological (and where possible from a purely theoretical) viewpoint. Next to that, the characterization procedure for each of these three properties has also been detailed. It has been indicated how the magneto-optical properties and the magnetic anisotropy of a ferromagnetic metal both strongly depend on the structural material quality of the deposited metal. The former via the relationship between the permittivity tensor and the conductivity tensor of a metal, with the conductivity being evidently dependent on the crystallinity of the metal through parameters such as carrier lifetime and effective mass. Needless to say that also magnetic anisotropy depends on the crystal quality of the metal, as anisotropy is pre-eminently a crystallographic material property. And polycrystallinity will strongly influence the magnetic properties through mechanisms such as domain wall pinning. The metals used in this research are all sputter-deposited and have a polycrystalline character with a grain size of several tens of nm. In view of the above remarks, it is clear that there is no easy way to design and/or predict the "best" candidate material from a magneto-optic and a magnetic viewpoint. All one can do is either somewhat naively rely on predictions made in other wavelength regions for the magneto-optic strength and on technological expertise to achieve reasonable magnetic anisotropy. In this way the "design" of the ferromagnetic contact might seem more like a trial-and-error task, characterizing several candidate metals with the procedures here described until minimal criteria are fulfilled. In this context it is worthwhile to recall that, based on our analysis in Chapter 2, it is the real part of the off-diagonal permittivity element that needs to be maximized, and that this

cannot be simply related to either the Kerr rotation or the Kerr ellipticity (but is a function of both!). In other words, while an easy criterion for the magnetic anisotropy can be formulated (maximum coercive field, maximum remanence, in short maximum squareness), it is not that straightforward to do the same for the magneto-optical strength. The hypothetical benchmark example of the previous chapter however suggested that this off-diagonal element shouldn't be much less than of the order 1. Of course, things are a bit more complicated, as also the absorption of the metal contributes to the whole design. A possible material figure of merit could therefore be the maximization of $\frac{\Re(\epsilon_{xz})}{\Im(\epsilon_{xx})}$.

For the electrical Ohmic function, things are a bit easier, as we have seen that for p-type contacts on covalent III-V semiconductors the Ohmic behaviour is more a function of the dopant level in the semiconductor than of the metal. A good Ohmic contact is therefore expected if large dopant levels are achievable. As a requirement a maximal contact resistivity of $10^{-5}\Omega\text{cm}^2$ was put forward. It has been only slightly touched upon in the introduction of section 3.1.3, but ideally one would prefer a wide band gap high-doped semiconductor contact layer in order to avoid shielding of the MO metal through absorption in the low band gap ($E_g = 0.75$ eV) ternary InGaAs[49]. Table 3.2 mentions a few wide band gap InGaAsP semiconductor contact layers, however as a rule it is more difficult to get high dopant concentrations in materials with increasing band gap. In the next chapter we will show the experimental results of our investigation whether it is possible to work with quaternary contact materials.

As a final remark we would like to emphasize that in this section the three key properties of the ferromagnetic metal/semiconductor contact interface were each approached from the idealized behavior of two touching unperturbed bulk materials. In other words, possible surface effects arising from the interface between the metal and the semiconductor have been neglected. One exception is of course the electrical properties where it was pointed out that precisely these surface effects pin the Fermi surface and give rise to a Schottky barrier height that is quasi-metal-independent. For the magnetic anisotropy on the other hand it is clear that as long as there is no degree of epitaxy (and this is not expected given the large lattice mismatch between CoFe and InP) there will be little influence from this surface on the hysteresis properties. On top of that even if there were a degree of epitaxy present, it has been investigated by Bayreuther et al. [55] that the magnetic spin reorientation due to the interface magnetic anisotropy constant will only show up if the film thickness is below 6 ML. It is clear that for our purposes these thin-film effects are of no importance. Recall Fig. 2.22 of the benchmark example in chapter 2, where it was shown that there is a "minimal" thickness for onset of the non-reciprocity due to interference effects in thin films.
There has up till now been no experimental evidence of changed optical and magneto-optical constants due to interfacial effects. However, since the density of states is strongly perturbed at a surface, it can be expected that also the conductivity tensor will undergo a perturbation. But again it can be expected, as in the case of magnetic anisotropy, that these perturbations will only be consider-

[49] as the device under study is designed to operate at 1300nm a quaternary contact material with a band gap wavelength well below $\lambda_g < 1.3\mu$m would be preferrable

able in the first monolayers of epitaxial ferromagnetic films. In other words the first few nm of the ferromagnetic film might behave differently (magneto-)optically spoken. In the worst case scenario this difference is expressed by the formation of a so-called "magnetic dead layer". This is a layer that has lost its ferromagnetic properties (and a fortiori its magneto-optic properties). Such a layer would only constitute an extra shielding absorber for the TM light with an absorption coefficient of the same order of magnitude of the diagonal elements in the permittivity tensor of CoFe. A magnetic dead layer can be detected with an accuracy of about 1nm by doing absolute magnetization measurements as a function of thickness and extrapolating to zero magnetization. In the next chapter experimental evidence will be given that for the used sputtered CoFe films, the dead layer will be at least thinner than 2 nm, i.e. can be neglected. A possible change of the values of the (magneto-)optical constants in the first few nm's of the metal remains. However, it is safe to assume that this change in a region of say 3nm (which is less than 5% of the normally used film thickness) will influence the device characteristics by not more than a few %. These interfacial effects could be measured by doing ellipsometry on very thin films and taking the multiple reflections into account, but the oxidation would require to cover the very thin CoFe films with a passivation layer of precisely known thickness. This doesn't make the measurement of surface effects any easier. Therefore interfacial magneto-optic effects have been neglected in the design.

3.2 Bandgap engineering: Strained Quantum Wells

3.2.1 Introduction

After the in-depth study of the ferromagnetic metal(/semiconductor interface), we now turn our attention to the other key aspect of the studied SOA/isolator device: the optical amplifying active region of the layer structure. It has been pointed out in Chapter 1 and in the benchmark example of section 2.3.3 in Chapter 2 that the residual forward TM modal loss to be compensated might reach levels of 200/cm or more. Roughly speaking, this implies material optical gain levels of several 1000/cm if optical confinement in the gain region is of the order of 10-20%. Such material gain levels can only be obtained in QW's. On top of this the gain provided by the wells must be polarization selective. Since the transverse magneto-optic Kerr effect only works for the TM polarization, any present TE light in the device will, if not suppressed, add noise to the signal and undermine the operation principle of the device through possible gain saturation.

Tailoring the polarization-dependent gain of a QW to such an amount that one polarization is left unamplified, can be achieved by artificially engineering the valence band structure. If the degeneracy of the heavy-hole and light hole valence band can be lifted, then TE gain can be discriminated versus TM gain or vice versa. Adams (and Yablonovitch independently almost at the same time) was the first to suggest the idea of valence band engineering by growing the QW's with a slight lattice mismatch, thus introducing strain in the crystal structure [56, 57]. Their original intention was to improve the operation characteristics of long-

wavelength semiconductor InGaAsP/InP quantum well lasers by reducing the in-plane hole effective mass through the strain-induced band structure modification of the valence band. However, the year was 1986, and unstrained QW lasers were just starting to mature and enter the market [58]. Understandably, there was some reservedness in applying strained-layer quantum wells due to the feared negative effects on the reliability of these semiconductor lasers. The impact of the suggestions of Adams and Yablonovitch and the first laboratory experimental successes by Thijs in 1989 [59] was nevertheless that immense that nowadays (just over a decade later) strained-layer semiconductor QW lasers are the only type of semiconductor lasers that are used in most advanced commercial telecom systems.

The study of strained-layer QW lasers is a research domain in itself. It would take us to far to elaborate on the quantum physics determining their characteristic properties, let alone to describe the modern crystal growth technology needed to fabricate these devices. In this section, we therefore limit ourselves to a review of the principles of valence band engineering through strain and of the existing material systems to produce 1.3-1.55 μm TM-selective semiconductor active gain regions.

3.2.2 Strained-layer quantum wells: principles

3.2.2.1 Optical gain

Incorporating strain in the active layer of an active optical device fits in the continuous search for improvements in the gain characteristics of semiconductor lasers and amplifiers. This evolution was started in 1970 with the first demonstration of room temperature CW operation of double heterostructure lasers [60], continued with the first report of MQW lasers with threshold current densities and internal quantum efficiencies outperforming bulk lasers (in 1981 by Tsang [61]), evolved through the first uses of compressively strained wells (1984 [62]) and the first genuine room-temperature demonstration of a 1.55 μm strained-layer MQW laser (by Thijs in 1989 [59]), and is presently seeing the commercialization of ultralow threshold quantum dot lasers.

The above sketched evolution is the logical sequence towards achieving the same levels of optical gain for a continuously decreasing carrier density. This is best explained by considering the density of states of the carriers in the amplifying semiconductor medium. It is well-known that the optical gain is proportional to the reduced density of states (DOS), ρ_{red}, and the "level-of-inversion" in the occupation of the band states of the semiconductor, $f_c - f_v$, (the difference in the Fermi-Dirac occupation probabilities of the interacting conduction (sub-)band and valence (sub-)band) [63].

$$g(\hbar\omega) = \left(\frac{1}{\hbar\omega}\right) \frac{\pi e^2 \hbar}{\epsilon_0 c m_0^2} \frac{n_g}{n^2} |M_T|^2 \rho_{\mathrm{red}}(\hbar\omega - E_g)(f_c - f_v) \tag{3.56}$$

where n_g is the group index of the guided optical mode[50], and $|M_T|^2$ the squared

[50] for unguided modes the group index is just equal to the semiconductor refractive index n

absolute value of the transition matrix element of the perturbation term in the Hamiltonian, describing the interaction between the radiation field and the matter wavefunctions. This latter will be studied in more detail in the section on polarization dependence 3.2.2.3, as it is strongly dependent on the polarization of the amplified light. E_g is to be seen as the energy separation at the band edges of the recombining bands. It is thus only equal to the semiconductor band gap if bulk material is considered. In quantum-confined structures E_g is increased because of the quantization of the conduction bands (CB) and valence bands (VB), and also varies for different recombining subbands. The population inversion, $\rho_c - \rho_v$, is a function of the injected carrier density (via the quasi-Fermi functions) and of the photon energy $\hbar\omega$ (via the energies of the recombining states[51]). It is trivial to prove that population inversion can only be achieved if the separation between the quasi Fermi levels exceeds at least E_g (defined as above). The reduced density of states (also known as the joint DOS) is a DOS function that describes how many states in the recombining semiconductor bands have the correct energy separation ($\hbar\omega$) to contribute to the optical gain. Under the assumption of **k**-selection, every CB state has an unique VB state with which it recombines, and thus an unique photon emission energy. The transition rate (which is related to the optical gain) of CB-VB recombinations at $\hbar\omega$ will thus be proportional to the density of electron CB states at a wavevector k that has the correct magnitude to lead to emission of a $\hbar\omega$ photon. It is however more convenient to express this density of states as a function of the emitted energy instead of wavevector. The energy difference E_{eh} between the recombining states can be written as

$$E_{eh} = E_g + E_c(k) + E_v(k), \tag{3.57}$$

with the CB and VB energies measured from the band edge. The (volume) density of (electron) CB states $\rho_c(k)\mathrm{d}k$ (with a wavevector leading to emission at $\hbar\omega$) can then be written as function of energy through the change of variable $k \rightarrow E_{eh}$, leading to

$$\rho_c(k)\mathrm{d}k \longrightarrow \left[\frac{\rho_c(k)}{\mathrm{d}E_{eh}/\mathrm{dk}}\right]\mathrm{d}E_{eh}, \tag{3.58}$$

where (3.57) is derived with respect to k and the whole right hand side is expressed in terms of E_{eh} by inversion of (3.57). The term between brackets is known as the reduced DOS, as it correctly describes via (3.57) how many states in the CB and VB are available in the correct combination for emission at a certain $\hbar\omega$. In general, this is a very complicated function, because E_c and E_v are not guaranteed to be simple polynomial functions of k. As a result, after derivation of (3.57) it is not always straightforward to express k as a function of E_{eh}. If a parabolic band model is assumed for both the CB and the VB then $\frac{\mathrm{d}E_{eh}}{\mathrm{d}k}$ simply reduces to $\hbar^2 k(\frac{1}{m_c} + \frac{1}{m_v}) \equiv \frac{\hbar^2 k}{m_r}$ with $k = \sqrt{\frac{2m_r}{\hbar^2}(E_{eh} - E_g)}$. The reduced DOS then assumes the same form as the conduction (or valence) band DOS but with the CB (or VB) effective mass replaced by the reduced effective mass m_r and with the electron or

[51] Remark in this context that **k**-selection is assumed, meaning that we are neglecting recombination via localized states. As a result there is a one-to-one correspondence between every conduction band state and every valence band state.

hole energy relative to the band edge now replaced by photon energy relative to the band gap! This result is independent of whether the amplifying semiconductor is a bulk or a quantum-confined structure. Therefore, if parabolic bands can be assumed then ρ_{red} will increase in the same manner as the band DOS, i.e. as the square root of the photon energy for bulk active layers, or as a stepwise constant staircase function for QW's[52].
Inspecting (3.56), the above observations make clear that in order to achieve a high gain it is necessary to obtain population inversion at levels with a high reduced density of states. In a bulk active layer, this can only be achieved after first filling the lower lying energy levels (which have a very low DOS). In a QW, this is not the case, as ρ_{red} has a step-like discontinuous onset and remains constant for higher energies. As a result the maximum gain will always occur at the band edge and the CB quasi-Fermi level mustn't be "pushed" deep into the CB to reach high population inversion at this band edge. On the other hand, because of the steep onset of the band DOS in a QW the transparency carrier density, i.e. when the quasi-Fermi levels separation equals the band gap, is higher in a QW than in an bulk active layer. All of this is illustrated in Fig. 3.13.

Fig. 3.13 makes it clear that in a hypothetical bulk medium with $m_c = m_v$ (for which population inversion occurs right at the band edge), when one increases the carrier densities above transparency, the maximum gain will "smoothly" increase ánd shift towards higher photon energies, since the gain at the band gap itself is always zero due to DOS being zero. On the other hand, in an idealized hypothetical QW (thus again having equal effective masses and parabolic bands), the step-like onset of the DOS will cause a "fast" increase in gain when the carrier density starts exceeding the transparency density, and the maximum gain will always occur at the gap energy, since the DOS remains constant for higher energies and the population inversion is always maximal at the band edge. In other words, due to the step-like DOS, the differential gain of a QW, defined as $\left.\frac{\mathrm{d}g}{\mathrm{d}N}\right|_{N=N_{\text{tr}}}$ at transparency, will greatly exceed that of a bulk active layer. The differential gain of an active layer is a very important parameter, as it is directly related to speed with which the optical gain can be modulated. A high differential gain allows the optical gain to be varied over the same amount with a much smaller variation of carriers. A situation which is clearly beneficial for direct current modulation of for instance laser diodes.
Next to that, while the maximum gain in a bulk active layer only very gradually increases, this increase will at the same time only slowly tend to saturate. This is because, even if the population inversion factor $f_c - f_v$ saturates to 1, the DOS keeps on increasing as the square root. In a QW however, the maximum gain will increase very strongly above transparency (as can be seen in Fig. 3.13 when the Fermi function is pushed into the CB) but this increase will quickly saturate because the population inversion at the band edge will quickly approach its maximum value of 1.
Fig. 3.13(c) shows what happens in a more realistic QW structure with a strong

[52]This band DOS behaviour is proven in any basic text book on solid-state physics.

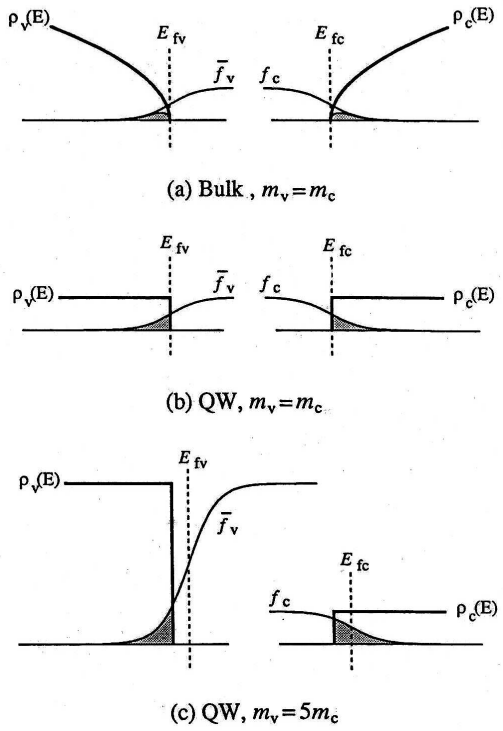


Fig. 3.13: CB and VB DOS (under transparency conditions) in three idealized cases assuming parabolic bands. For clarity the hole Fermi function $\overline{f}_v = 1 - f_v$ has been used in the valence band. Note that the Fermi functions have been normalized to the DOS levels. The shaded areas correspond to transparency carrier density.

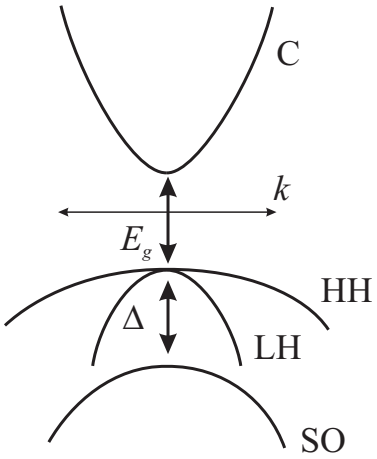
mismatch between the effective masses of the conduction band and the valence band. Of course it is still a very idealized picture of the band structure as it still assumes isotropic parabolic bands (and thus stepwise constant DOS). In order to maintain charge neutrality the quasi-Fermi functions are shifted towards the band with the smallest effective mass[53]. Therefore, at transparency ($E_{fc} - E_{fv} = E_g$) the CB quasi-Fermi function lies well into the CB, and the transparency carrier density is considerably higher than in the case of symmetric band effective masses. On top of that the reduced mass m_r will approach the mass of the lightest band, and thus the reduced DOS will approach the CB DOS! As a result the differential gain will be dominated by the CB, and since it is clear from Fig. 3.13 that already at transparency the population inversion at the band edge is closer to its maximum value of 1 than in the symmetric case, the differential gain will be much smaller

[53]Remember that in the case of parabolic bands the DOS is directly proportional to the effective mass, $\rho = \frac{m}{2\pi\hbar^2}\frac{1}{L_z}$ with L_z the thickness of the well.

(but still considerably higher than for a bulk layer).
Summarizing, in order to minimize the transparency carrier density ánd to maximize the differential gain it is found that matching the CB DOS ρ_c and the VB DOS ρ_v, and reducing their overall magnitude is highly desirable. We will now prove that this desirable engineering of the band structure can be achieved by introducing strain in the active layer. However, before we can consider the engineering of the effective masses of the CB and the VB we must understand what influences the characteristics of these bands and how their behaviour can be predicted. It is therefore necessary to consider some band structure basics, focussing specifically on QW band structures. Incorporating strain in a material without introducing crystal lattice defects, can after all only be achieved if the strained layer has quantum size features.

3.2.2.2 Conduction and valence bands in quantum wells

Bulk band structure It is well known that in a bulk semiconductor crystal having a zincblende structure (which most often leads to a direct gap band structure) the conduction band can be seen as a remnant of a s atomic orbital, while the valence band is threefold degenerate and each of the valence bands are related to the remnants of the p_x, p_y and p_z atomic orbitals (see Fig. 3.14).



$$|u_c\rangle = |u_s\rangle$$

$$|u_{\mathrm{hh}}\rangle = -\frac{1}{\sqrt{2}}|u_x\rangle - j\frac{1}{\sqrt{2}}|u_y\rangle$$

$$|u_{\mathrm{lh}}\rangle = -\frac{1}{\sqrt{6}}|\bar{u}_x\rangle - j\frac{1}{\sqrt{6}}|\bar{u}_y\rangle + \frac{2}{\sqrt{6}}|u_z\rangle$$

$$|u_{\mathrm{SO}}\rangle = -\frac{1}{\sqrt{3}}|\bar{u}_x\rangle - j\frac{1}{\sqrt{3}}|\bar{u}_y\rangle - \frac{1}{\sqrt{3}}|u_z\rangle$$

Fig. 3.14: Schematic representation of the band structure of a zincblende-like direct gap semiconductor; and the Bloch functions near the Brillouin zone center. Note that the overlined Dirac ket notation indicates the Bloch function with the opposed spin orientation. Each band Bloch function $|u_i\rangle$ is degenerate with its spin conjugated $|\overline{u_i}\rangle$, which can be found by Hermitian conjugation of the given Bloch functions.

For the conduction band, things are quite simple. If spin degeneracy is not included, then the CB is a non-degenerate band, that furthermore is relatively far removed from any other bands in the material. Therefore the CB will not mix up

with other bands (having orbital character of higher symmetry) and it will mainly maintain its s-like symmetry, i.e. isotropic (spherical symmetric) and parabolic in a sufficiently wide **k**-space around the Brillouin zone center. Its interaction with the periodic potential (and thus the other energy bands) is purely restricted to its curvature and thus its effective mass.
For the valence band it is immediately clear that the threefold degeneracy complicates matters. First of all, due to the spin-orbit term in the Hamiltonian, one of the three bands, the split-off band, is shifted downwards over an energy Δ, leaving only a degeneracy between two bands, the heavy and light-hole band. Just as the conduction band, the three valence bands will all bear the original p orbital character in them, but intermixed in such a way that the spin-orbit interaction term in the valence band Hamiltonian is diagonalized. This is indicated in Fig. 3.14. Actually, the valence band Bloch functions as indicated in Fig. 3.14 are those that diagonalize the spin-orbit interaction term with the electron **k**-vector directed along a $\langle 100 \rangle$, here in particular the z-direction[54]. However, as opposed to the conduction band, the valence bands are far from being isotropic. In other words as the electron **k**-vector, tilts away from any $\langle 100 \rangle$ crystal direction the curvature of the band changes. The theoretical formalism that describes this band anisotropy has been developed by Luttinger and Kohn [64, 65]. It would take us too far to elaborate this formalism here. It suffices to point out the general conclusions of such calculations. As the electron **k**-vector takes on a direction away from a $\langle 100 \rangle$, the heavy and light hole Bloch functions start interacting, and band mixing occurs. On top of this, the band mixing is dependent on the plane in which the electron **k**-vector rotates, being for instance considerably different in the $\{100\}$ family of planes than in $\{110\}$-planes. As a result of this band anisotropy the constant energy surfaces are not spheroids or ellipsoids, and only along $\langle 100 \rangle$ the hole energy increases parabolically. The effective mass along other directions can only be numerically evaluated. The most important feature of the Luttinger-Kohn theory however is that the Bloch functions of Fig. 3.14 start intermixing. As a result the heavy hole band acquires some light hole band characteristics (more in particular some $|u_z\rangle$-character, as can be seen from the $|u_{\mathrm{lh}}\rangle$ expression in Fig. 3.14). This is an important point when it comes to considering the polarization of the emitted light as we shall see in section 3.2.2.3. One could question if the intermixing of the HH and LH band preserves the "heaviness" character of the bands. In bulk semiconductors this is still the case, because along any crystal direction the band energy $E_h(\mathbf{k})$ is the solution of a quadratic eigenvalue equation [66]. By definition, the smallest root (measured along the hole energy axis– i.e. opposite to the electron energy axis) can then be associated with the heavy hole band. In a quantum-confined structure and a fortiori in a strained quantum confined structure, things change dramatically.

QW band structure The band structure in a quantum well[55] is found by solving a square potential well problem. This is the most standard of quantum-mechanical quantization problems. Its solution is treated in every standard Quan-

[54] The correct Bloch functions for **k** directed along x or y are found through cyclic permutation.
[55] throughout this section the QW confinement direction is chosen along the z-axis.

tum Mechanics textbook. The quantization of the conduction band levels is indeed pretty straightforward. For the valence band, however, there are a few extra subtleties due to the band mixing.

We start by considering the simplest case first, namely that of the non-degenerate conduction band. Because interaction with other energy bands can be neglected the solution to the Schrödinger equation will be a $|u_s\rangle$ Bloch function multiplied by an envelope function[56] $F(\mathbf{r}_t, z) = F(z)\exp[j(k_x x + k_y y)]$, having plane wave character in the plane of the well and bound character transverse to the well. The $z-$dependence of the envelope function is easily found to satisfy

$$-\frac{\hbar^2}{2m_c(z)}\frac{\mathrm{d}^2F}{\mathrm{d}z^2} + \left(V(z) + \frac{\hbar^2}{2m_c(z)}(k_x^2 + k_y^2)\right)F = E_c F \tag{3.59}$$

where the $m_c(z)$ underlines the fact that the CB mass may be entirely different between the QW and the barrier material. The potential function $V(z)$ describes the square-well potential formed by the difference in CB band edges between the QW and the barrier. It is important to note that the potential profile has an extra contribution from the "kinetic-energy-like" term originating from the plane wave behaviour in the plane of the well. Because the CB mass of the QW and the barrier can differ, this extra term is not simply a global increase of the potential! At the band edge ($k_{xy} = 0$) this kinetic energy term disappears, and the solution of this Schrödinger eigenvalue equation poses no problems. The discrete, quantized energy levels $E_{c,n}(0)$ for which non-trivial solutions of (3.59) exist, follow straightforwardly from the application of the boundary continuity conditions (F and $\frac{1}{m_c}\frac{\mathrm{d}F}{\mathrm{d}z}$) at the QW interfaces. Away from the band edge ($k_{xy} \neq 0$) the kinetic energy term can be introduced as a perturbation, which is only necessary if $m_{c,\mathrm{QW}} \neq m_{c,\mathrm{b}}$. In [63] it is proven that the final result for the conduction subband energies then becomes

$$E_{c,n}(k_x, k_y) = E_{c,n}(0) + \frac{\hbar^2(k_x^2 + k_y^2)}{2m_{\mathrm{eff}}} \qquad \text{with } \frac{1}{m_{\mathrm{eff}}} = \frac{\Gamma}{m_{c,\mathrm{QW}}} + \frac{1-\Gamma}{m_{c,\mathrm{b}}}, \tag{3.60}$$

where Γ is the envelope function confinement factor.
Thus, the conduction subbands in a QW remain well-behaved parabolic bands but with an in-plane effective mass that is the weighted average of the bulk effective masses inside and outside the well.
The simplicty of the band structure in the CB of a QW is actually a direct consequence of the assumption that the interaction with other energy bands is weak enough so that we can treat it perturbatively by replacing that interaction with a conduction band effective mass. This is precisely what is expressed by the Schrödinger-like effective mass equation of (3.59). For the strongly degenerate valence band, it is necessary to work with coupled effective mass equations, as

[56] The envelope function approximation [67] for the electron wavefunction is a frequently used construct for localized states (such as in QW's). The key assumption is that within a given energy band, the Bloch function is not a strong function of $\mathbf{k}$ and can thus be approximately represented by the band edge Bloch function. Thus a localized approximate wavefunction solution in a given energy band consists of two components: the band edge Bloch function multiplied by a slowly varying envelope function $F(\mathbf{r})$.

already indicated in the discussion of the valence band structure in bulk. There it was mentioned that within {100} and {110} crystal planes the four coupled equations (heavy and light hole including spin degeneracy) could be decoupled into two identical sets of coupled equations. This diagonalization requires the use of a new set of Bloch functions for the heavy and light holes [68]

$$|u^*_{\rm hh}\rangle = \frac{1}{\sqrt{2}}(|u_{\rm hh}\rangle - |\overline{u}_{\rm hh}\rangle) \tag{3.61a}$$

$$|u^*_{\rm lh}\rangle = \frac{1}{\sqrt{2}}(|\overline{u}_{\rm lh}\rangle - |u_{\rm lh}\rangle), \tag{3.61b}$$

with $|u_{\rm hh}\rangle$ and $|u_{\rm lh}\rangle$ the heavy and light hole band edge Bloch functions as defined in Fig. 3.14. Along any crystal direction in the {100} or {110} plane, each of these two sets can again be diagonalized into two independent equations, one for the heavy hole and for the light hole. This last diagonalization mixes the Bloch functions of (3.61) (but in bulk their energy dispersion retains the heaviness character of each solution[57])

$$\psi_{\rm hh}(\mathbf{k}_t, k_z, \mathbf{r}) = e^{j\mathbf{k}\cdot\mathbf{r}}(F_{{\rm hh},h}(\mathbf{k}_t, k_z)|u^*_{\rm hh}\rangle + F_{{\rm hh},l}(\mathbf{k}_t, k_z)|u^*_{\rm lh}\rangle) \tag{3.62a}$$

$$\psi_{\rm lh}(\mathbf{k}_{,t}, k_z, \mathbf{r}) = e^{j\mathbf{k}\cdot\mathbf{r}}(F_{{\rm lh},h}(\mathbf{k}_t, k_z)|u^*_{\rm hh}\rangle + F_{{\rm lh},l}(\mathbf{k}_t, k_z)|u^*_{\rm lh}\rangle), \tag{3.62b}$$

with $F_{{\rm hh},h}, F_{{\rm hh},l}, F_{{\rm lh},h}$ and $F_{{\rm lh},l}$ the elements of the matrix that diagonalize the two coupled equations (in the $(|u^*_{\rm hh}\rangle, |u^*_{\rm lh}\rangle)$-basis). The solution in a QW can now be constructed out of the bulk solutions by taking in each region (well and barrier) the plane wave superposition of the bulk solutions that exist at a certain but still unknown energy and at the considered value for the in-plane k_t-vector[58]. In view of the fact that the hole eigenenergies of the eigenfunctions in (3.62) are found as the solutions to a quadratic equation in k_z^2 and k_t^2, this procedure will lead to a superposition of 4 plane waves [66].

$$\Psi = e^{j\mathbf{k}_t\cdot\mathbf{r}_t}(A_\pm\psi_{\rm hh}(\mathbf{k}_t, \pm k^*_{\rm hh}, \mathbf{r})e^{\pm jk^*_{\rm hh}z} + B_\pm\psi_{\rm lh}(\mathbf{k}_t, \pm k^*_{\rm lh}, \mathbf{r})e^{\pm jk^*_{\rm lh}z}), \tag{3.63}$$

with $k^*_{\rm hh}$ and $k^*_{\rm lh}$ the k_z values which combine with the given $\mathbf{k}_t$ to the sought energy respectively along the bulk heavy hole energy dispersion equation, and along the bulk light hole energy dispersion equation. This implies a total of 12 unknown expansion coefficients, out of which 4 can be eliminated due to the boundary conditions at $\pm\infty$. The eight remaining coefficients are subjected to 4 boundary conditions at each interface of the QW leading to an 8×8 homogeneous set of equations for which non-trivial solutions only exist a certain discrete values of the unknown valence band energy $E_{v,n}(\mathbf{k}_t)$. This procedure is generally started at the band edge ($\mathbf{k}_t = 0$) where the heavy hole and light hole function are still uncoupled, i.e. $F_{{\rm hh},l} = 0$ and $F_{{\rm lh},h} = 0$ in (3.62). In this way the band edge

[57] meaning that for every direction the lowest energy solution has a distinct low curvature (heavy) and the highest energy solution a distinct large curvature (light).

[58] This is actually the same procedure as was applied for the solution of the CB QW, but there the bulk solution is very simple and the k_z wavevector is straightforwardly related with the sought energy by $k_z^2 = \frac{2m_c}{\hbar^2}E - (k_x^2 + k_y^2)$.

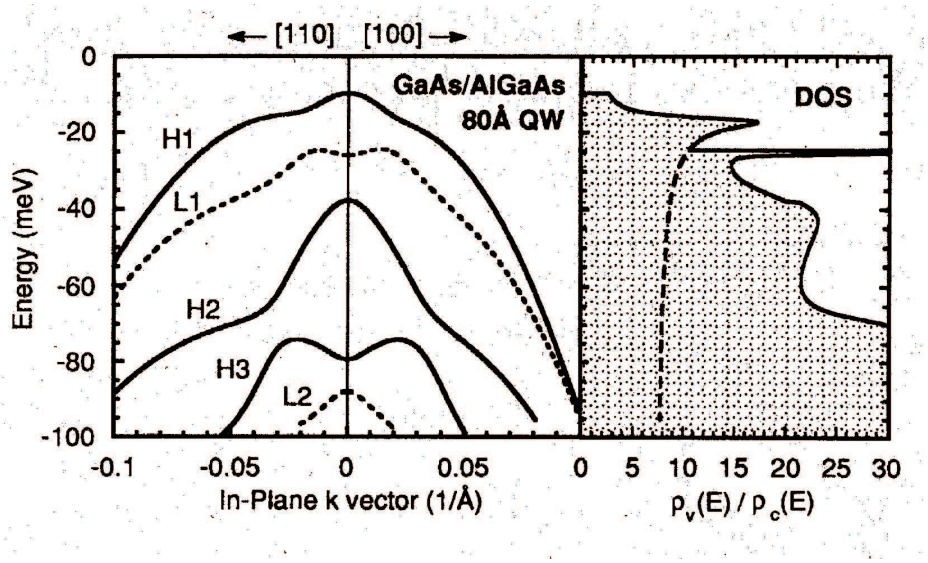


Fig. 3.15: Valence subband structure of an 80 Å $GaAs/Al_{0.2}Ga_{0.8}As$ quantum well (after [69]). The coupling between the several heavy (solid curves) and light (dashed curves) hole subbands is very dramatic, as can be seen for instance from the occurrence of a "local" negative effective light-hole mass near the band edge for the first LH subband. The highest lying HH subband also changes its effective mass from light to negative to explicitly heavy. At right the impact of the band coupling on the DOS is plotted. The total VB density of states relative to the first conduction band DOS is plotted as a solid curve. Ideally, this should be a step-like curve! The dashed DOS curve is the ratio of the HH1 DOS to the C1 DOS.

energies of the different valence subbands are found. Subsequently, in each of the sub-bands $\mathbf{k}_t$ is gradually increased, and the band dispersion in each subband is calculated using the above described method. As opposed to the situation of a bulk semiconductor where the valence band mixing was moderate (expressed by the $F_{\mathrm{hh},l}$ and $F_{\mathrm{lh},h}$ elements), and hardly influenced curvature of the bands, the plane wave expansion of (3.63) makes it clear that in a QW the resonance in the confinement direction will come as a result of a strong interplay between the heavy and light hole bulk wavefunctions. Not only will this strongly modify the Bloch functions of the different valence subbands but more importantly this is expected to have a profound impact on the effective mass and the curvature of the bands. Fig. 3.15 shows a theoretical example of such a QW valence band structure calculation. Here it can be clearly seen how the heavy-light hole valence band mixing dramatically influences the valence band structure. First of all it can be noted how at the band edge the quantum confinement effect has lifted the degeneracy between the HH and LH bands. As expected, the highest lying band has heavy-hole character, because in a potential well heavy "particles" are

easier bound than light "particles". Also note how the HH1 band reduces its mass right at the band edge, and how this mass reduction is quickly counteracted when interaction with the LH1 band starts. In this context it should be noted that in view of the assessments of section 3.2.2.1 this mass reduction should be advantageous. However, in an ordinary unstrained QW band coupling sets in so quickly that this advantage is immediately lost. This is overwhelmingly clear by looking at the DOS plot at the right-hand side of Fig. 3.15. The ideal staircase DOS with steps of unit magnitude, is far removed from the DOS plotted in this figure. As soon as band coupling sets in the HH1 band is "pushed away" by the LH1 band, thereby increasing its effective mass, and making the DOS rapidly increase (until it falls back at its asymptotic mass value that is about three times bigger than at the band edge, as can be seen in the dotted curve). Note how the band-coupling even introduces a singularity spike in the DOS owing to the zero slope in the LH1 band! Thus, if band coupling weren't present, the behaviour of HH1 in the neighbourhood of the band edge could be retained over a considerably larger energy range leading to a DOS curve closer to the one depicted in Fig. 3.13(b). In order to "eliminate" the band-mixing one has to devise a way to "push" either the HH or the LH subbands out of the potential well. This can be achieved by growing the QW under a state of biaxial stress.

Strained QW band structure When a sufficiently thin epitaxial layer is coherently grown lattice-mismatched between two thick host layers, a tetragonal deformation of the cubic symmetry of the crystal lattice of this layer will occur. If the relaxed lattice parameter a_s of the thin layer is larger than that of the surrounding material (a_0), its lattice will be elongated in the growth direction and biaxially compressed in the growth plane, and vice versa if the lattice parameter of the thin layer is smaller than that of the surrounding host layers. The strain is described by a diagonal strain tensor $\underline{\underline{\epsilon}} = \mathrm{diag}(\epsilon_{\mathrm{xx}}, \epsilon_{\mathrm{yy}}, \epsilon_{\mathrm{zz}})$, with $\epsilon_{xx} = \epsilon_{yy}$. The in-plane strain is given by

$$\epsilon_{\parallel} = \epsilon_{xx} = \epsilon_{yy} = \frac{a_0 - a_s}{a_s}, \tag{3.64}$$

and the out-of-plane strain ϵ_{zz} is related to $\epsilon_{\parallel}$ by Poisson's equation.

$$\epsilon_{\perp} = \epsilon_{zz} = -\frac{2\sigma}{1-\sigma}\epsilon_{\parallel}, \tag{3.65}$$

where σ is Poisson's ratio. For tetrahedral semiconductors σ is approximately $1/3$ so that $\epsilon_{\parallel} \simeq -\epsilon_{\perp}$. It is known from continuum mechanics [70] that the total strain in the lattice-mismatched layer can be decomposed into two components, a hydrostatic component that expresses how the volume of the unit cell is changed, and an axial component that describes how its form is changed.

$$\epsilon_{ax} = \epsilon_{\perp} - \epsilon_{\parallel} \simeq -2\epsilon_{\parallel} \tag{3.66a}$$

$$\epsilon_{hyd} = \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} \simeq \epsilon_{\parallel} \tag{3.66b}$$

The hydrostatic component does not lower the symmetry of the structure. Its main effect is therefore to change the overall band gap by $\Delta E_g = a\epsilon_{hyd}$, with a

$a_s \lessgtr a_0$	$\epsilon_\parallel$	$\epsilon_\perp$	ΔE_g	S
$>$ (compressive)	$-$	$+$	$+$	$-$
$<$ (tensile)	$+$	$-$	$-$	$+$

TABLE. 3.3: Influence of strain on band edge characteristics. Fig 3.16 serves as an illustration of the here tabulated trends.

the hydrostatic deformation potential. The axial strain is responsible for the shape deformation and thus for the cubic symmetry breaking of the lattice, introducing a tetragonal distortion which can be shown to split the degeneracy of the light and heavy hole states at the band edge. The relative shift varies approximately linear with the strain[59], $E_{\rm hh} - E_{\rm lh} = S = 2b\epsilon_{ax}$, with b the axial deformation potential and the energy measured along the electron energy axis. For the common III-V semiconductor compounds a and b are negative numbers. Using the above formulae, Table 3.3 summarizes the different band edge features of strained layers.

The splitting of the heavy and light hole band in strained layers leads one to suspect that the band mixing will strongly decrease. Next to that owing to the reduced symmetry of the lattice it is reasonable to expect a strong increase in the band anisotropy leading to a large difference in band energy dispersion between in-plane and out-of-plane orientations for the electron **k**-vector. This can be rigorously proven by doing 6×6 Luttinger-Kohn like quantum-mechanical calculations [71]. Without going into detail about the precise quantum mechanics associated with incorporating strain in the valence band Hamiltonian, we present schematically the results in Fig. 3.16 It is clearly seen how the strain causes the bands to become strongly anisotropic. While it can be roughly stated that out-of-plane the highest lying band will more or less maintain its characteristics, the in-plane situation changes drastically. The effective mass of the highest lying band is strongly influenced by the strain. For compression the heavy hole band sees a large reduction of its effective mass, while in tensile strained layers the effective mass of the LH band increases, but remains below the unstrained in-plane HH effective mass! In other words, strain causes an overall reduction of the VB DOS as compared to the unstrained material. In view of the conclusions formulated in section 3.2.2.1, the transparency carrier is expected to be considerably decreased in strained material.

The above assessments are for bulk material. Of course it is intuitively clear that the incorporation of strain is only practical in layers with quantum features. We had seen that in unstrained material the valence band mixing is so drastic (see Fig. 3.15) that the theoretically expected improvement due to a step-like DOS is only very partially reached. In a strained QW however, it is now expected that this band mixing will be strongly reduced owing to the strain-induced splitting

[59] if the strain remains small. At higher strains the light-hole band starts interacting with the SO band, and the shift between LH and HH saturates.

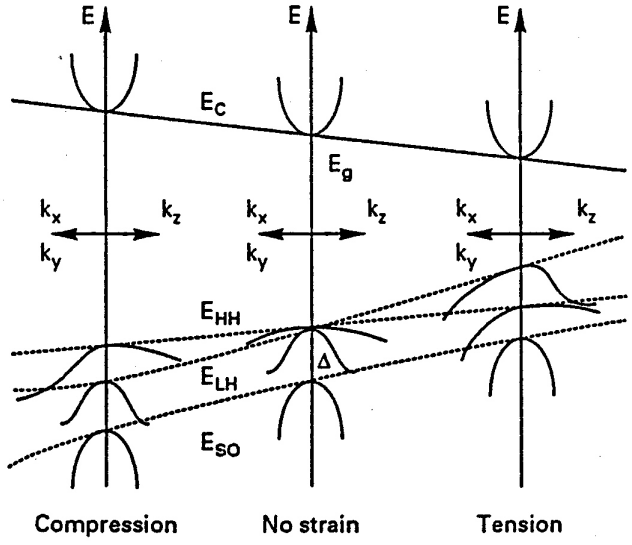


Fig. 3.16: Schematic representation of the (bulk) band dispersion of a strained III-V semiconductor under compression, lattice matching, and tension.

of the HH and LH bands. Qualitatively this can be understood by correctly interpreting the effects of the different in- and out-of-plane effective masses and the different band gaps for the heavy and light hole bands. In compression the light holes experience a seemingly higher band gap and will thus see a less deep potential well. On top of that, the spacing of its subband will still be large owing to its quasi-unchanged (small) out-of-plane mass. As a result one expects a QW valence band structure with hardly any band mixing (thus a staircase DOS) and LH subbands almost pushed out of the well. On top of that the reduced in-plane HH effective mass will lead to a VB DOS that is much better matched to the CB DOS. In tension, the situation is slightly more complicated. Even though the heavy hole experiences a shallower potential well due to the strain-induced splitting, the small out-of-plane mass of the light holes leads to a high subband spacing (and a high zero point energy), while the heavy hole subbands remain more closely spaced. In this way, the effective splitting of the LH and HH subbands is reduced in a tensile-strained QW, and band-mixing will again increase. The only way to solve this is by increasing the well width in order to decrease the quantum confinement effect. All of this is illustrated in Fig. 3.17. For both strain cases it can be seen how the band mixing is reduced, leading to much more parabolic bands. Next to that, the reduction of the in-plane effective mass as compared to the unstrained case is also obvious. As a consequence the VB DOS will not only be much more staircase-like, it will also be better matched to the CB DOS. As explained, it is also noticed how reduction of band mixing is easier to achieve in a

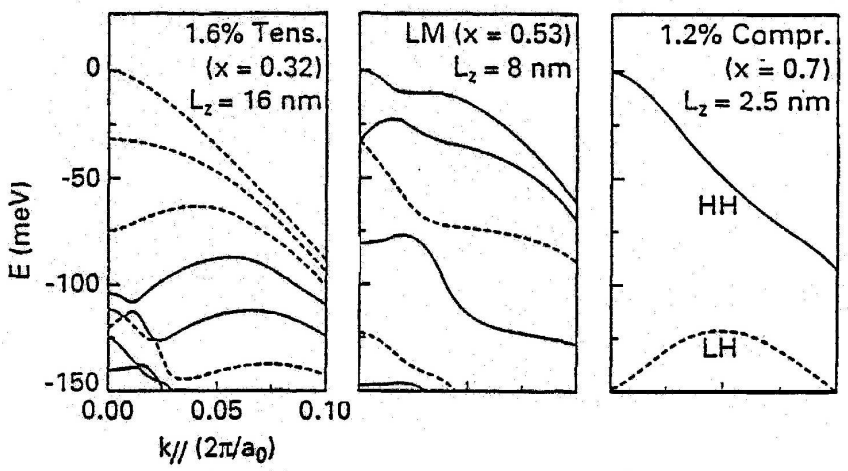


Fig. 3.17: Illustration of the valence band structure of strained and lattice matched $In_xGa_{1-x}As$ wells (with lattice-matched InGaAsP Q1.25 barriers for the compressive strained well, and Q1.15 for the tensile strained well). The width of the well and the strain is adjusted to have emission at $1.5\mu m$ *(after [72]). The dashed curves are the LH bands.*

compressive strained well than in a tensile strained one.

3.2.2.3 Polarization-dependent effects

We have seen how incorporating strain in a QW improves the parabolicity of the valence band, hence the "shape" and magnitude of the VB DOS function density of states, and thus basically improves the gain characteristics of the semiconductor by reducing the transparency carrier density and increasing differential gain. However, the greatest contribution to the improved characteristics in strained QW's arises from the variation of the character of the Bloch function in the highest lying valence states (that contribute the most to the optical gain). In order to understand this, it is necessary to turn our attention to the characteristic properties of the transition matrix element in (3.56). It is this part that contains the information about the wavefunctions.
In a dipole approximation for the light-matter interaction Hamiltonian (see for instance [73]), this matrix element can be written out for a transition between a certain conduction subband n_c and a certain valence subband n_v in a QW as follows

$$|M_T|^2_{n_c,n_v} = 2 \sum_{u_c,\overline{u}_c} |\langle u^*_{\rm hh}|\hat{\mathbf{e}}\cdot\mathbf{p}|u_c\rangle\langle F_h^{n_v}|F_c^{n_c}\rangle + \langle u^*_{\rm lh}|\hat{\mathbf{e}}\cdot\mathbf{p}|u_c\rangle\langle F_l^{n_v}|F_c^{n_c}\rangle|^2, \quad (3.67)$$

where the factor 2 accounts for the spin degeneracy in the considered valence subband. The summation is over the spin degenerate conduction subband. And the two terms inside the expectation value "operator" are a result of the band-mixing discussed in detail in (3.63), just as the occurrence of the light and heavy hole envelope functions. The valence band Bloch function are those mentioned in (3.61) that are needed to decouple the 4 coupled effective mass equations for the valence band. $\hat{\mathbf{e}}$ is the unit polarization vector of the photon, and $\mathbf{p}$ is the Hermitian electron momentum operator. (3.67) is the most complete version of the transition matrix element taking both quantum effects and band-mixing effects into account[60]. A rigorous treatment of this expression can be found in [71]. We are here mainly interested in the qualitative features of this element with relation to its behaviour towards the TE and TM polarization of the photon field. For that purpose it suffices to consider the characteristics of (3.67) in the direct neighbourhood of the band edge, i.e. where the band-mixing can be neglected.
In that case one of the valence envelope functions will vanish: $F_l = 0$ for a HH subband and $F_h = 0$ for a LH subband. And only one of the valence band Bloch functions is retained: $|u^*_{\rm hh}\rangle$ for a HH subband and $|u^*_{\rm lh}\rangle$ for a LH subband. Using (3.61) and the fact that the spin parts of these Dirac kets fulfill the orthogonality condition $\langle\uparrow \mid \downarrow\rangle = 0$, leaves us with the following possible $\hat{\mathbf{e}}\cdot\mathbf{p}$ matrix elements to be determined, $\langle u_{\rm hh}|\hat{\mathbf{e}}\cdot\mathbf{p}|u_c\rangle$ and $\langle u_{\rm lh}|\hat{\mathbf{e}}\cdot\mathbf{p}|u_c\rangle$ and their spin-flipped counterparts. Using the expressions for the band edge light hole and heavy hole Bloch functions as given in Fig. 3.14, this actually implies the following type of $\hat{\mathbf{e}}\cdot\mathbf{p}$ matrix elements

$$\langle u_s|\hat{e}_x p_x|u_x\rangle, \qquad \langle u_s|\hat{e}_x p_x|u_y\rangle, \qquad \langle u_s|\hat{e}_x p_x|u_z\rangle, \tag{3.68a}$$
$$\langle u_s|\hat{e}_y p_y|u_x\rangle, \qquad \langle u_s|\hat{e}_y p_y|u_y\rangle, \qquad \langle u_s|\hat{e}_y p_y|u_z\rangle, \tag{3.68b}$$
$$\langle u_s|\hat{e}_z p_z|u_x\rangle, \qquad \langle u_s|\hat{e}_z p_z|u_y\rangle, \qquad \langle u_s|\hat{e}_z p_z|u_z\rangle, \tag{3.68c}$$

and spin-flipped counterparts, evaluating to the same values. In view of the p-orbital origin of $|u_x\rangle$, $|u_y\rangle$ and $|u_z\rangle$, elementary quantum mechanics learns us that

$$\begin{aligned} \langle u_s|p_i|u_j\rangle &= 0 \qquad \text{for } i \neq j, \\ \langle u_s|\hat{\mathbf{e}}\cdot\mathbf{p}|u_i\rangle &= M\hat{e}_i, \end{aligned} \tag{3.69}$$

where the basis momentum matrix element M has been introduced[61]. It is not necessary for our purposes to evaluate the this element. In the above treatment we tacitly neglected the evaluation of the envelope function overlaps $\langle F_c|F_h\rangle$ and $\langle F_c|F_l\rangle$. That is because at the band edge these overlaps evaluate to a value close to 1 (and exactly to 1 if the in-plane effective masses of the valence band and conduction band are equal[62]). Using the above results for the momentum matrix

[60] Actually this is not entirely correct. In order to complete it needs an extra averaging in the electron in-plane k-space, in order to include in-plane band-dispersion in the QW.

[61] It suffices to use the symmetry properties of the p-like Bloch function to see that $\langle u_s|p_x|u_x\rangle = \langle u_s|p_y|u_y\rangle = \ldots$.

[62] which is because the square well problem in that case is exactly the same for the electrons and the holes. Away from the band edge band-mixing occurs and F_h and F_l are simultaneously present in any one wavefunction, and the envelope function overlaps strongly deviate from 1, and redistributes the transition strengths.

elements, and under the assumption of unity envelope overlaps, and using the form of the Bloch functions in Fig. 3.14, the transition matrix element in (3.67) at (and very close) to the band edges evaluates to

$$\frac{|M_T|^2_{c,v}}{|M|^2} = \begin{cases} \hat{e}_x^2 + \hat{e}_y^2 = 1 - |\hat{\mathbf{e}} \cdot \hat{\mathbf{k}}|^2 & \text{for a HH band} \\ \frac{1}{3}(\hat{e}_x^2 + \hat{e}_y^2 + 4\hat{e}_z^2) = \frac{1}{3} + |\hat{\mathbf{e}} \cdot \hat{\mathbf{k}}|^2 & \text{for a LH band} \end{cases}, \tag{3.70}$$

with the transition strength expressed in units relative to the basis momentum matrix element.
Of course the above formulae shouldn't depend on the precise choice of the coordinate axes. Remembering that the form of the Bloch functions of Fig. 3.14 is that which is valid if the **k**-vector of the electron is pointing in the z-direction, the expressions in (3.70) can be made reference frame-independent by introducing the unit vector along the electron's **k**-vector (as is done in the right-hand parts). In order to interpret the implications of (3.70), we have to be aware of the fact that a photon will interact with electrons with **k**-vectors pointing in all directions. Therefore, $\hat{\mathbf{k}}$ in the above formula must be seen in the sense of an *average* unit vector. In a bulk material the k-states of the carriers are isotropically distributed and $\hat{\mathbf{k}}_{\text{ave,bulk}} = \frac{1}{\sqrt{3}}(\mathbf{1}_x + \mathbf{1}_y + \mathbf{1}_z)$ while in a QW the average **k**-vector near the band edge is along the confinement direction of the QW, say $\hat{\mathbf{k}}_{\text{ave,QW}} = \mathbf{1}_z$. With these results we can compose the following table with the transition strengths for light polarized along any one Cartesian direction. [We choose the z-direction along the growth direction of the device, and the x-direction along the growth plane and perpendicular to the facets[63].] From Table 3.4 it is seen that in a bulk active layer

$\frac{\lvert M_T\rvert^2}{\lvert M\rvert^2}$	TE $\hat{\mathbf{e}} = \mathbf{1}_y$		TM $\hat{\mathbf{e}} = \mathbf{1}_z$		S.E. $\hat{\mathbf{e}} = \mathbf{1}_x$	
	C–HH	C–LH	C–HH	C–LH	C–HH	C–LH
bulk	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$
QW	1	$\frac{1}{3}$	0	$\frac{4}{3}$	1	$\frac{1}{3}$

*TABLE. 3.4: Transition strengths for the three basic light polarization directions for bulk active layers and QW's with a $z-$axis growth direction. These results are only valid for transitions near the band edge. Away from the band edge band-mixing occurs, and the valence band wavefunction used to evaluate M_T contain both HH and LH contributions. On top of that in a QW the electron **k**-vector then tilts away from the confinement axis, changing the value of $\boldsymbol{k}_{\text{ave}}$ to be used in* (3.70)*. Therefore there is a redistribution of the total transition strength (= 2) between the different polarization. However, this redistribution only gets important when the carrier densities get high enough so as to fill up the states away from the band edge.*

each transition has the same probability to occur. Therefore it can be stated that

[63] Light that is polarized along this direction is not guided by the amplifier and is tagged as spontaneous emission.

only one out of three injected carriers contribute to the desired polarization amplification! In a QW things changes dramatically. In an unstrained well the HH1 band is shifted upwards with respect to the LH1 (see Fig. 3.15), owing to the higher mass of heavy holes, and one would expect the TM amplification to be completely suppressed. However it was seen above that in an unstrained well band-mixing occurs relatively quickly, which increases the TM transition strength. If the well is either tensile or compressive strained the HH-LH shift is much stronger and band-mixing is heavily suppressed as discussed above. In the case of compressive strain, the suppression of TM gain will be much more effective, and for transitions near the band edge Table 3.4 shows that one out of two carriers contribute to TE amplification, which is an increase of 50% with respect to the bulk case. This immediately implies an important increase in differential gain. For tensile strained wells the increase in differential gain is even more drastic, as it can be seen that in theory near the band edge two out of three carriers contribute to TM amplification (i.e. an increase of 100%!), while only one out of six contribute to TE amplification[64].

Now that it has been clearly explained why and how the band structure can be engineered and how this valence band engineering leads to a considerable increase in material gain of the active layer, we turn our attention to the limits of strain incorporation.

3.2.2.4 Elastic properties – critical thickness and critical strain.

The most important restriction to the use of strain in an epitaxial layer is the limited elasticity of the material. The stored elastic energy in a strained layer will increase linearly with the thickness of that layer (with a slope that is proportional to the amount of strain). It is clear that there must be a certain critical thickness above which the elastic energy in the lattice is released by the formation of misfit dislocations. These misfit dislocations will not only act as very effective non-radiative recombination centers, but will also be a source for further multiplication of crystal defects. The study of the critical thickness t_c has been a source of many controversies, mostly due to a disagreement on when a layer has lost its epitaxial coherence. Is it when its luminescence properties degrade, or when its morphology changes, or when the carrier mobility is reduced? The most conservative approach[65] is that of Matthews and Blakeslee [74–76]. They have derived a transcendental equation for the calculation of the critical thickness as a function of lattice strain $\epsilon_{\|}$, crystallographic properties (dislocation line direction, slip plane, Poisson modulus, ...), and the length of the Burgers vector b of the misfit dislocation. The Matthews formula is only useful if a lot of crystallographic information

[64] Actually, it can proven that the fractional $|u_z\rangle$-character, which is responsible for the TM amplification in a LH subband, will increase with increasing strain, even at the band edge (where the HH and LH are after all already decoupled). That is because the character of the LH Bloch function also changes at the band edge due to an interaction with SO band [71]. This increases the relative amount of $|u_z\rangle$-character in the LH band. It might even approach 90%, which would imply an increase in TM differential gain of about 200%!

[65] and the one mostly used in opto-electronics since luminescence properties are usually the first seen to degrade.

is known about the materials. Therefore, Andersson has derived a more practical rule of thumb based on the Matthews theory and valid for a whole range of III-V compounds [77]. He demonstrated that high-quality, "defect-free" strained epitaxial layers can be grown if the thickness-strain product $t \times \epsilon_{\|} < 100 - 200\mathring{A}\%$. In other words, a critical thickness t_c of about 10nm for a 1% lattice mismatch.
Finally, it must be mentioned that the above treatment is only valid for a single strained layer. In a multi-quantum well active layer the total strain in the device accumulates along $\epsilon_{\text{tot}} = \frac{N_w t_w \epsilon_w + N_b t_b \epsilon_b}{t_{\text{tot}}}$, making the dislocation generation still possible even with each layer below the critical thickness. Defining the average lattice constant for the active MQW layer as the averaged summation over the number of monolayers in each material [79]

$$a_{\text{ave}} = \frac{\sum a_w n_w + \sum a_b n_b}{\sum n_w + \sum n_b}, \tag{3.71}$$

it is possible to achieve an averaged lattice parameter that matches that of the substrate, e.g. by taking $a_w > a_{\text{sub}}$ and $a_b < a_{\text{sub}}$. This so-called strain compensation does not lift the critical thickness limitation for each individual layer, however. Another aspect to be taken into account in strain-compensated devices, is that now also the valence band structure of the barriers is modified. And this modifies the quantum-mechanical calculation of the hole envelope functions and Bloch functions. This in turn will alter the gain characteristics of the MQW active layer. Theoretical calculations have indicated that this might deteriorate the gain characteristics [80]

Another restriction for strained layers is the possible formation of type II QW's, for which the electron carriers and hole carriers are confined in spatially different regions of the active layer. This is often mistakenly thought to be a result of a difference in hydrostatic deformation potential between the conduction band and the valence band. Actually, the change in band gap due to the hydrostatic lattice strain, $\Delta E_g = a\epsilon_{\text{hyd}}$, can only come about if the deformation potential for the conduction band and the valence band are different, since obviously $\Delta E_g = (a_c - a_v)\epsilon_{\text{hyd}}$. There has been some controversy about the precise sign and magnitude of the conduction and valence band hydrostatic deformation potentials [81, 82]. The experiments of Cavicchi [78], who was the first to measure the CB offsets in $In_xGa_{1-x}As$ as a function of strain, seem to suggest that a_c must be rather large and positive. This is in complete contradiction with the tabulated values. However, it mustn't be overlooked that a large part of the variation in the band offsets is due to the compositional variation of the band gap. As a result the strain-induced decrease (or increase) of the band gap might well be counteracted by the compositional increase (or decrease) of the band gap. However, the compositional changes of the band gap are generally speaking several times higher than those induced by strain. Therefore, the occurrence of type II behaviour is typically purely caused by the compositional changes of the material and the associated changes in band gap. Therefore, departing from the precise lattice-matched band offset ratio $\frac{\Delta E_c}{\Delta E_v}$, either tension or compression can lead to type II behaviour, depending on the precise material system under consideration. As an example, for $In_xGa_{1-x}As$ on InP, the tensile region will have a band gap increase, and taking

into account the $\Delta E_c \approx 0.2$ eV (and assuming no bowing of the CB offset with In content), a negative ΔE_c will be found as soon as the tensile strain is 2% or more.

3.2.3 Material systems

It was pointed out in Chapter 1 that the studied integrated isolator concept is in the first place aimed at application in 1300 nm ($E_g \approx 0.95$ eV) modules. Strained active layers in this wavelength region can be found in a number of III-V semiconductor material systems. Fig. 3.18 plots the energy gap as a function of lattice constant for the best known binary zincblende-like III-V compounds.

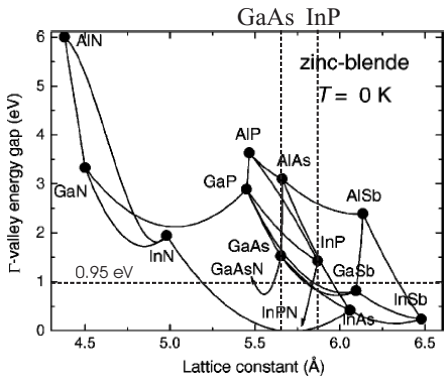


Fig. 3.18: Energy gap versus lattice constant for several zincblende-like binary III-V semiconductor compounds (and curves for some ternary alloys). The dashed line indicates the desired operation band gap, and the lattice position of the two standard substrates, InP and GaAs.

By inspection of Fig. 3.18, it is clear that tensile strained active layers for 1300 nm application, will be very hard to realize on GaAs substrates, unless one is able to work with nitride-based III-V alloys. This InGaAsN:GaAs material system, however, is a very recent advent on the opto-electronic research scene. Its technology is still maturing. When these become available, any opto-electronic active telecom component will be able to benefit from the increased temperature stability of GaAs-substrate-based devices. Presently, this nitride material system seems most likely to emerge as the toughest competitor for InP-based active materials. The expected superior temperature performance of GaAs-based devices is mainly due to the significantly larger conduction band offset [83], allowing a better electron confinement and reducing thus radiative carrier leakage due to current overflow.

For InP substrates Fig. 3.18 shows that there are several material systems available to realize a tensile-strained QW with a 0.95 eV band gap. The simplest system seems to be the ternary alloy $In_xGa_{1-x}As$. However, Thijs has shown that it is theoretically quasi-impossible to grow tensile-strained InGaAs wells [72]. Fig. 3.19 explains why this is. In a ternary material system there is only one degree of

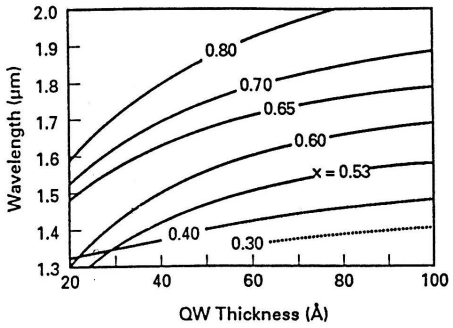
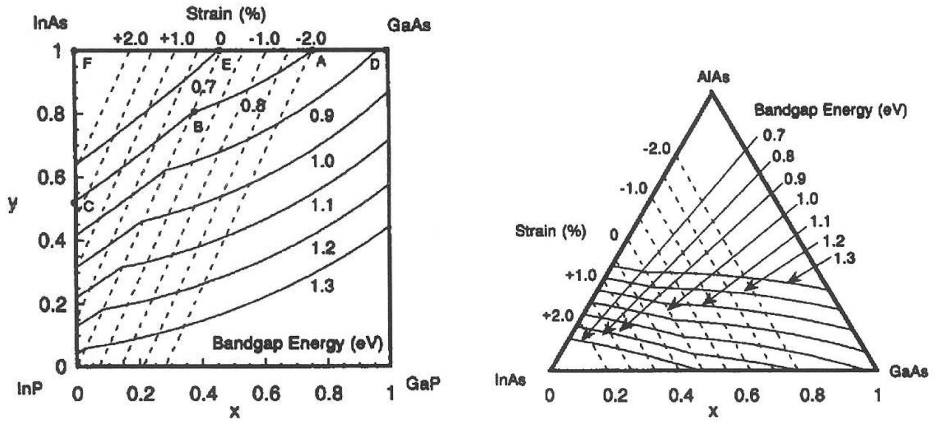


Fig. 3.19: Emission wavelength of $In_xGa_{1-x}As/InGaAsP$ QW's as a function of QW thickness and strain. The solid lines are C-HH transitions (thus mainly TE polarized), while the dashed lines are C-LH transitions. Below 30% In fraction ($\approx$ 1.6% tensile strain) the CB offset approaches zero and type II behaviour occurs (after [72]).

freedom. As a result the amount of strain also fixes at the same time the band gap energy (and thus the emission wavelength if the barrier material is fixed). If one wants tensile strain, and still an acceptable conduction band offset, the QW has to be unrealistically thin. The larger band gap quaternaries $In_{1-x}Ga_xAs_yP_{1-y}$ and $Ga_xAl_yIn_{1-x-y}As$ are better suited, because the strain and the quantum well thickness can be adjusted independently (within certain limits). This is illustrated in Fig. 3.20. If the band offsets are known with respect to the barrier material and combined with the results of this figure, the potential well problem is completely determined and can be solved along the lines of section 3.2.2.2, resulting in the wavelengths of the different subband transitions (as a function of the chosen well width). A thorough treatment of band line-up together with analytical expressions for these two material systems, can be found in [84]. Because Fig. 3.20 proves that it is possible to vary within certain limits the bulk band gap energy independently of the strain and because the formulae of Bowers show that the same applies for the CBO and the VBO, there is an extra degree of freedom with respect to InGaAs:InP. Of both quaternary systems however, the Al-based material is expected to have a considerably better gain performance in view of its generally larger CBO's. Table 3.5 compares the average CBO's[66] of the different material systems that can be used for 1300nm tensile strained applications.

[66] these values have to be used with great care. These are just averages of the trends observed within one type of material system. Of course the CBO varies as a function of both strain and composition as mentioned before. The treatment of Bowers is better to determine the conduction and valence band positions for arbitrary alloy compositions.



(a): $In_{1-x}Ga_xAs_yP_{1-y}$ (b): $Ga_xAl_yIn_{1-x-y}As$

Fig. 3.20: Contours of the strain and band gap energy of "bulk" (a) InGaAsP and (b) GaAlInAs (after [84]). The strain is defined with respect to InP and is taken negative for tension.

material system	$\frac{\Delta E_c}{\Delta E_g}$
InGaAsN:GaAs	0.80
$Ga_xAl_yIn_{1-x-y}As$:InP	0.72
$In_{1-x}Ga_xAs_yP_{1-y}$:InP	0.40

TABLE. 3.5: Averaged conduction band offsets (in lattice-matched conditions) for the three candidate TM material systems (after [85]).

3.2.4 Design and characterization methods of strained-layer MQW active layers

3.2.4.1 Design

Based on all of the previous observations it should in principle be possible to theoretically design the optimum active layer structure for our purposes. That is find the tensile-strained MQW layer structure that will give an as high as possible TM material gain in combination with a maximum TM/TE gain suppression for a minimum current density. However, the amount of parameters that has to be taken into account for this optimization problem is immense. Even if one limits oneself to the problem of optimizing a single strained (quaternary) QW, the number of parameters is not less than five (2 compositional parameters for the well and 2 for the barrier, and the well thickness), and, more importantly, these parameters are all strongly interlaced by the boundary conditions on the critical thickness (Matthews and Blakeslee), on the type II behaviour, and on the emission wavelength.

Next to that it mustn't be forgotten that any gain calculation will always result in a relationship between (peak) gain and well carrier density. The accessible device variable is not the carrier density but the current density into the QW. Even if we ignore carrier loss through vertical and lateral carrier leakage, there are still the two important non-radiative carrier recombination processes, Schockley-Read-Hall (SRH) and Auger, that consume important fractions of the injected current. While the radiative current density can relatively "easily" be modelled as the integrated spontaneous emission rate[67], the same doesn't hold for the Auger recombination or the SRH recombination. For the latter there are some semi-classical models based on Boltzmann statistics (see for instance [86]), but the capture coefficients of the crystal defect levels are not known with good accuracy. The Auger process is a three-body-quantum-mechanical problem, making its modelling even harder [87]. Because of the fundamental momentum and energy conservation, any Auger process heavily depends on band structure, and thus on the strain. Actually, it is easy to see that owing to the splitting of the HH and LH states and to the reduction of the hole effective mass, there is considerable decrease of the number of holes able to contribute to the Auger processes. This indicates how important it is to incorporate strain into a theoretical model of Auger recombination. In any case all of the above makes clear that it is far from straightforward to model the optical gain of a strained QW as a function of current density. If the SRH and Auger rates are introduced as empirical parameters in the model (with all the accompanying uncertainties), one may wonder what the added value of a theoretical model still is. After all if the non-radiative recombination rates are experimentally determined, one can at the same time measure the gain-current relation of the active layer.
For a number of wells higher than 1, things get even more complicated, as then the total modal gain is also a function of the possible non-uniform carrier distribution between the wells and the possible non-uniformity of the optical modal overlap with the wells. There have been some attempts at analytically modelling the gain in a single strained QW [88]. The authors were strangely enough admitting themselves within the same publication that the practicality of these formulas is very limited. The model merely confirmed in a theoretical manner the widespread use of the empirically found logarithmic (peak) gain-current relationship for a quantum well [89, 90],

$$g = G_0 \frac{J_0}{\eta_i} \ln(\frac{\eta_i J}{J_0}), \tag{3.72}$$

with G_0 the differential gain at transparency, J_0 the transparency current density, and η_i the internal quantum efficiency. The method described by the authors is very laborious to use for the calculation of G_0 and J_0 as a function of the device parameters[68].

For all of the above reasons, it was decided to take a very pragmatic approach towards the design of the active layer. Based on the well-founded assessments of the extensive research by Thijs into InGaAsP/InP 1.3 μm tensile-strained devices [91],

[67] and in this way be related to the Fermi level separation or the carrier density

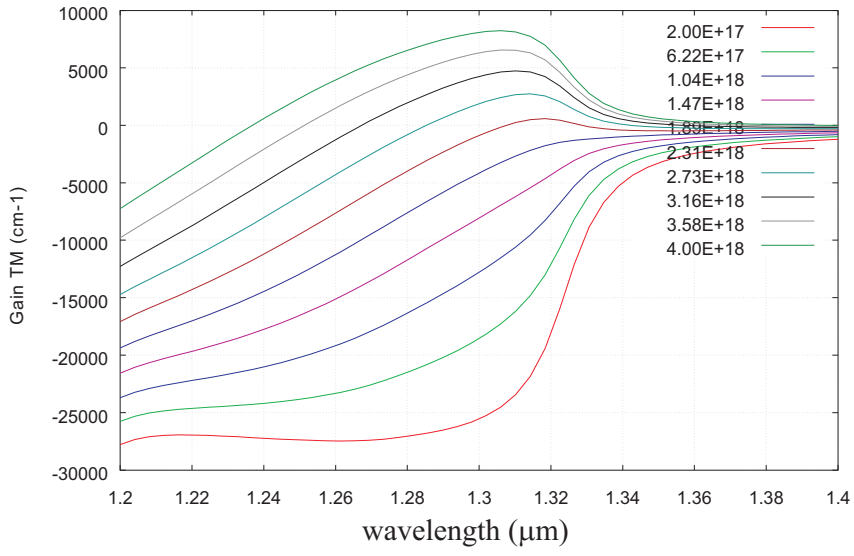
[68] and it doesn't allow a calculation of η_i, only a calculation of the gain as a function of the radiative current density.

a whole range of tensile-strained MQW InGaAsP material was grown with varying strain, varying number of wells, and varying strain compensation in the barriers[69] By experimentally measuring the gain parameters of (3.72) of these materials, trends in the gain performance as a function of the accessible material parameters can be observed, and the material combination optimized for the application can be empirically chosen. In view of the theoretical knowledge of the possibly better gain performance of InAlGaAs, a similar empirical procedure has been followed for that material system. The results of this experimental design procedure are extensively treated in the next chapter.

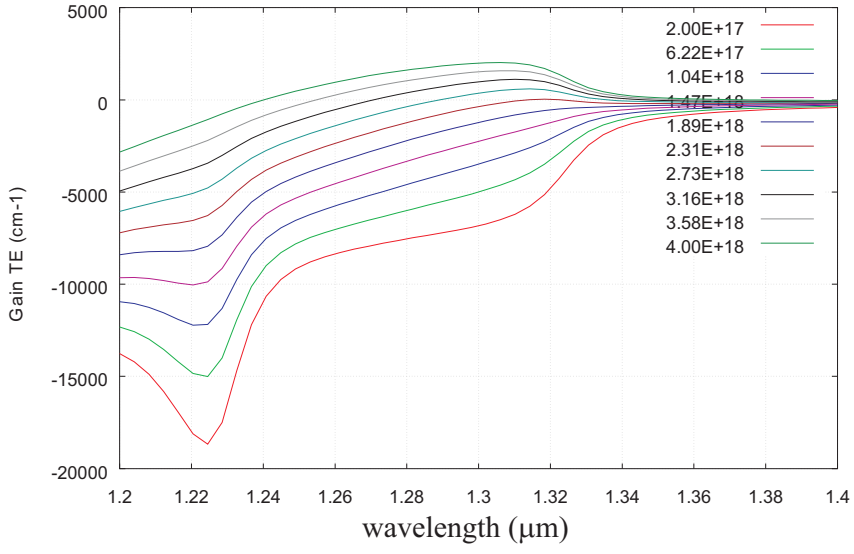
It is not entirely correct to state that the gain performance of the tensile strained material has been purely determined on an empirical basis. A few theoretical 6×6 Luttinger-Kohn $\mathbf{k} \cdot \mathbf{p}$ band structure calculations were performed, combined with optical transition matrix calculation along (3.56). This was however not done with the intention of thoroughly studying the tensile-strained material system, but merely to get an idea of what would at least be theoretically possible as material gain. These calculations were nevertheless important as they served as a guideline for the design of the first generation of SOA/isolator devices (see section 3.3.2). Fig. 3.21 gives an example of these calculations.

The details of the layer structure are described in section 3.3.2.1. The well thickness is 11nm, and has a tensile strain of -1.1%. It was calculated to have essentially a single LH subband. In Fig. 3.21(a) it can be seen how the TM gain peak barely shifts (the small shifts are a result of the gain broadening due the intraband carrier scattering with a relaxtion constant of $\tau_s = 0.1$ps). The absence of any other resonances in the spectral gain proves the presence of only a single LH. The TE gain plotted in Fig. 3.21(b) proves the strong suppression of the TE C-LH1 transition. The other resonance at a much shorter wavelength indicates the location of the C-HH1 subband. As expected based on the analysis of Table 3.4 this transition is not found in the TM material gain, though at higher carrier densities the band mixing effects might start to slightly enhance the TM strength of this transition.
The observed splitting of the LH-HH subbands is about 85 nm, or in other words $\Delta E_{hh-lh} \approx 50$ meV. With the strain at 1.1% this would imply an axial deformation potential b of about 2 eV. This is sligtly above the range of the values tabulated in literature [81], which is perfectly logical in view of the extra splitting due to the quantum confinement effect. Using an approximated linear interpolation of the TM gain peak, the TM differential gain at transparency can be roughly estimated to be about 10^{-15}cm^2, which is about a factor of two better than in a lattice-matched QW [71], i.e. the enhancement predicted by Table 3.4. Finally, it is worth to make a rough estimate of the radiative current density corresponding to the here plotted gain levels, as this would give some idea of how much current density is required per well to achieve this gain. If it is assumed that the spontaneous recombination rate can be approximated by a parabolic model BN^2, and a value for the radiative coefficient $B = 1{-}2{\times}10^{-10}$ is taken [92], then for $N = 4{\times}10^{18}$cm^{-3} an (optimistic)

[69] A constant composition has been chosen for the barrier material with a sufficiently high band gap wavelength ($\lambda_g = 1.05\mu$m) in view of the above mentioned possible problems with type II behaviour for wells under tension.



(a)



(b)

Fig. 3.21: Theoretical 6×6 $\boldsymbol{k}\cdot\boldsymbol{p}$ *Luttinger-Kohn based optical gain calculations for the tensile strained InGaAsP QW used in the first generation SOA/isolator demonstrator, as a function of wavelength and QW carrier density (in* cm^{-1}*) for (a) TM polarization and for (b) TE polarization.*

estimate of $J_{\mathrm{rad}} = 300 - 400\mathrm{Acm}^{-2}$ per well is found, which corresponds to a total

current density per well of about $J_{\text{tot}} = 500 - 650\text{Acm}^{-2}$ (assuming $\eta_i = 60\%$) for a gain of more than 5000 cm^{-1}!

3.2.4.2 Characterization Methods

The gain parameters of (3.72) can be measured in several ways. The most direct method is to measure the gain/loss spectrum of a laser biased below his threshold current. This can be achieved by detecting the Fabry-Perot resonances undergone by the spontaneous emission of a waveguide semiconductor laser cavity with a monomodal cross-section. The spectral cavity gain $g_{\text{net}}(\lambda)$ of the spontaneous emission resonating in a cavity of length L with a gain source $g_{\text{mod}}(\lambda)$ and an internal loss source $\alpha_{\text{int}}(\lambda)$ can be described by the following relationship

$$g_{\text{net}}(\lambda) = \frac{1}{L}\left(\ln\left(\frac{1}{R}\right) + \ln\left(\frac{\sqrt{r(\lambda)}-1}{\sqrt{r(\lambda)}+1}\right)\right), \tag{3.73}$$

with $r(\lambda)$ the peak-to-valley ratio of the measured Fabry-Perot resonances and R the power reflection coefficient of the facets of the (symmetrical) cavity. This net cavity gain is related to the internal loss by $g_{\text{net}}(\lambda) = g_{\text{mod}}(\lambda) - \alpha_{\text{int}}(\lambda)$. Its peak value will increase with increasing injected current density until, at threshold, it equals the distributed mirror loss of the cavity, $\frac{1}{L}\ln\left(\frac{1}{R}\right)$. The modal gain $g_{\text{mod}}(\lambda)$ of the laser cavity is related to the material optical gain of the active layer by $g_{\text{mod}} = \Gamma g_{\text{mat}}$, with Γ the modal confinement factor in the QW's[70]. Thus, by measuring the peak net gain in the above manner for different values of the injected current density, the QW material gain as a function of total injected current density can be plotted, provided the internal loss and the confinement factor are known. This procedure is known as the Hakki-Paoli method [93].
Its main advantage lies in its simplicity. However, there are a few important drawbacks, the most obvious one being the need for a separate determination of the internal optical loss. In the net gain spectrum this loss term can only be unambiguously separated from the modal optical gain (or stimulated absorption) in wavelength regions beyond the bandgap wavelength (where the stimulated recombination of carriers, and thus g_{mod}, disappears). Mechanisms like intraband scattering and band gap renormalization broaden the spontaneous emission spectrum of a QW to that wavelength region. In other words, in the long wavelength tail $g_{\text{net}} \xrightarrow{\lambda\nearrow\nearrow} -\alpha_{\text{int}}$. If the internal loss can be assumed wavelength-independent, the modal gain throughout the entire spectrum can then be determined as $g_{\text{mod}}(\lambda) = g_{\text{net}}(\lambda) + \alpha_{\text{int}}$. This assumption however is questionable, since one of the dominating loss mechanism in this long wavelength limit is intervalence band absorption and this is known to be both very carrier-dependent and linearly dependent on the wavelength [94]. There seems no easy way out of this. All one can do is accept the assumption of wavelength-independent internal loss and look

[70] Note that this is only an approximation in the case of ridge waveguide lasers. Due to carrier diffusion, the gain in the QW's will not be homogeneously distributed and the perturbation formula that led to the confinement formalism for modal gain calculation is not valid anymore. The material gain will be slightly overestimated.

upon the measured modal gain spectrum as a lower limit. Another minor drawback of the Hakki-Paoli method is its limited sensitivity at high net gain levels. This is because the peak-to-valley ratio will not vary a great deal at increasing high gains (as can be seen from (3.73)). This is known as the high Q-factor limitation of the HP method [95]. This is a tiny bit annoying, because it limits the current range for which the (peak) gain-current relationship can be accurately measured. At the same time however, it is easier to perform HP measurements on short cavities, thus having higher mirror losses and requiring higher net gains before lasing threshold is reached, as this will increase the Fabry-Perot fringe separation. Finally, the HP method does in no way allow the measurement of the internal quantum efficiency.

A second method, that avoids some of the problems of the HP method, consists of measuring the differential quantum efficiency η_d of a broad-area laser above threshold as a function of length. If temperature effects are neglected (for instance because the laser is operated in pulsed regime), the LI curve has a linear behaviour above threshold with a slope that is given by

$$\frac{\mathrm{d}P}{\mathrm{d}I} = \frac{\hbar\omega}{2q}\eta_d \tag{3.74}$$

assuming that all light emitted by a single facet is 100% collected and that the facets are symmetric (hence the factor 2). The differential quantum efficiency thus describes how many photons are emitted out of the laser per number of injected carriers (above threshold). This is equal to the product of the "probability" that an injected carrier effectively recombines to generate a photon and the "probability" that a generated photon escapes the laser cavity. Or, in other words

$$\eta_d = \eta_{\mathrm{int}}\frac{\alpha_m}{\alpha_{\mathrm{int}} + \alpha_m}, \tag{3.75}$$

with the internal quantum efficiency expressing the carrier radiative recombination "probability" and with the ratio of the distributed mirror photon loss to the total photon loss expressing the photon escape "probability". If the above equation is rewritten as

$$\frac{1}{\eta_d} = \frac{1}{\eta_{\mathrm{int}}}\left(1 + \frac{\alpha_{\mathrm{int}}}{\alpha_m}\right) = \frac{1}{\eta_{\mathrm{int}}}\left(1 + \frac{\alpha_{\mathrm{int}}L}{\ln(1/R)}\right), \tag{3.76}$$

the internal loss and the internal quantum efficiency are found by linear regression of the measured inverse differential quantum efficiency as a function of length. The power reflection coefficient of an uncoated facet is with good approximation given by 28% for TM polarization 32% for TE polarization. Using the extracted loss value in

$$g_{\mathrm{mat,th}}(\lambda_{\mathrm{peak}}) = \frac{1}{\Gamma}\left(\alpha_{\mathrm{int}} + \frac{\ln(1/R)}{L}\right), \tag{3.77}$$

the peak material gain can be plotted for the different threshold current densities measured at the different lengths. With the knowledge of the internal quantum efficiency, the logarithmic gain-current relationship of (3.72) can then be fitted to these data using the gain parameters J_0 and G_0 as fitting parameters.
This method avoids the problems of the wavelength dependency of the internal

loss, as this loss is measured precisely at the wavelength for which the peak gain occurs, namely the lasing wavelength. However, it suffers from another drawback! It assumes, that the internal loss and the internal quantum efficiency are independent of the length of the laser cavity. This assumption is questionable for the following reasons. Especially in long-wavelength lasers, internal optical losses are mainly caused by free carriers and IVBA. In both cases, the absorption coefficient rises proportionally to the carrier concentration. The injection efficiency η_{int} is influenced by the lateral spreading of carriers, carrier escape from the active region, and non-radiative recombination losses within the active layers, which all depend on the carrier concentration. The mirror loss coefficient α_m rises with a shorter cavity length, requiring higher gain and more carriers in the quantum wells. The higher carrier concentration causes the internal absorption to increase and the differential internal quantum efficiency to decrease with cavity length. As a result, if α_{int} and η_{int} are assumed to be constant and equal to their long cavity limit values, the peak gain at higher values is underestimated, while the needed current to achieve this gain is underestimated (as can be seen by inverting (3.72)). The only way out of this dilemma is either to numerically simulate the LI curve of the laser as a function of length, and in this way determine the exact dependencies $\alpha_{\text{int}}(L)$, $\eta_{\text{int}}(L)$ that can be used to plot the correct (g, J)-pairs. This is a numerical gigantesque task that can only be tackled by specialized software packages that self-consistently combine carrier transport throughout the layered heterostructure, heat flux, optical gain computation and wave guiding. Or we can resort to restricting ourselves to measuring the (g, J)-pairs for long cavities only.

Summarizing, the characterization of the gain performance of tensile-strained MQW devices is subtle task that needs to be conducted with care. The above assessments make clear that only in the neigbourhood of the transparency current density a fairly reliable guess of the differential gain can be made. So that it must be at least possible to get a detailed idea about the influences of strain and strain compensation on the gain performance of the active layer. However, one has to be careful in extrapolating the logarithmic gain-current relationship towards higher gain levels. The main reason is that with increasing carrier densities the internal quantum efficiency can get seriously overestimated, leading in reality to a faster saturation of the material gain than what would be predicted on the basis of (3.72) with a constant η_{int}. As a final remark, it has to be pointed out that the current density in (3.72) is that of a single well. By relating the threshold modal gain to the threshold material gain using the confinement factor in (3.77), one assumes essentially that the carriers (and hence the total current density) are uniformly distributed over the wells. This is an assumption that can be questioned at increasing carrier concentrations and for active layers with a high number of wells. This serves as another indication that using the extracted gain parameters for high material gains has to be done with great care.

3.3 MQW ferromagnetic metal clad SOA/isolator: layer design and optimization

3.3.1 Introduction

After having assessed the design criteria for the building blocks of the SOA/isolator, the main task is to check whether a non-reciprocal waveguide layer structure can be designed within the boundary conditions set by these criteria, and with an acceptable isolation ratio and loss-compensating bias current injection. For this purpose the modelling tools developed in Chapter 2 are at our disposal. A few introductory remarks have to be made.
First of all, the here discussed non-reciprocity simulations are all performed on a 1D slab model of the waveguide. The reason for this is quite simple. It is mainly the transversal modal behavior that will determine the optical confinement near the metal. Two-dimensional modelling of the device, including the precise layout of the guiding profile in the lateral direction will cause only a slight refinement of the model. It is indeed obvious that for this particular device the design of the lateral dimension will not add specific non-reciprocal effects that are not already included in the pure transversal modelling. The inclusion of the lateral dimension will only lead to a general decrease of the effect due to a decreased confinement near the MO metal. The lateral design is thus purely a question of finding a compromise between achieving maximal non-reciprocity (slab waveguide) and obtaining appropriate other 2D waveguide characteristics such as monomodality, surface-recombination, current spreading, ... Within this research only limited attention has been given to the modelling of the lateral dimension. This task has been limited to finding the etch depths for which the device is monomodal up to ridge widths of 2.5μm.
Secondly, the benchmark simulations of the previous chapter (section 2.3.3) have convincingly demonstrated that in all realistic design situations there is a negligible difference between the rigorous approach and the perturbation theory model for the calculation of the isolation ratio. Therefore, all simulations have been performed using the considerably faster perturbation model.
Finally, the designs presented here are organized in a "historical" and a "pragmatical" manner. By this it is meant that they are entirely based on what was experimentally known and achievable at that precise moment in time during the research. The experimental characterization results on the sub-components are all detailed in the next chapter. Within the timeframe of this research two generations of SOA/isolators have been fabricated. While at the time of the first prototype there was little to no knowledge about the gain behaviour of tensile-strained InGaAsP wells and no characterization of the (magneto-)optical constants of CoFe had been achieved yet, the same doesn't apply for the second generation of devices. By then, it was experimentally assessed that strained InAlGaAs wells outperform InGaAsP ones and the (magneto-)optical constants of $Co_{90}Fe_{10}$ had been determined with sufficient accuracy. The details of both generations are detailed in the next two sections.

3.3.2 Slab optimization

3.3.2.1 Generation 1: $Co_{90}Fe_{10}$/InGaAsP

Input parameters As explained above, the only input information available at the time of design of the first prototype was the vague knowledge of the (magneto-)optical constants of pure Co (based on the 60's publication of [6]) and the theoretically calculated gain curves of the previous section. It was assumed that $Co_{90}Fe_{10}$ wouldn't have constants very different from pure Co, especially as in [6] it was proven that Fe and Co have the same relative signs for the real and imaginary parts of the gyrotropy parameter in the neigbourhood of 0.95eV and within the same order of magnitude.
All that had been assessed about the gain region is that a 3QW region with -1% strain in the wells (12nm) and +0.2% compressively strained barriers (Q1.17, 22nm) led to TM lasing with an acceptable though slightly high threshold current density (see section 4.2.1.1). For the first generation of demonstrators it was therefore decided to increase the strain compensation slightly (+0.3%) and to increase the band gap of the barriers up to $E_g = 1.18$eV ($\lambda_g = 1.05\mu$m) (to improve the electron carrier confinement). This led to the gain simulations of the previous section. Next to this it had been assessed that the ideal electric contacting scheme consists of a hybrid combination of a 100nm transparent quaternary InGaAsP (Q1.17) and a thin 15nm absorptive ternary layer (see section 4.1.3). This scheme has been included in all simulations with a realistic extinction coefficient $k = +0.2$ for the ternary layer [38]. Finally, for the MO metal the benchmark simulations of the previous chapter have proven that a thickness of 100 nm would be sufficient to maximize the non-reciprocal effect.

Design strategy With these fixed input parameters (strains, compositions and thicknesses for the MQW region, contacting layer and MO metal thickness), the design parameter space is seriously reduced. If for the moment the uncertainty on the gyrotropy constant is not taken into account, the optimization of the layer structure is a question of finding the right compromise between optimal TM mode confinement in the amplifying QW's (for loss compensation) and confinement in the non-reciprocal (lossy) MO metal (for optimal optical isolation). This leaves us with three layer parameters to be optimized: both guiding Separate Confinement Heterostructure (SCH) layers[71] beneath and above the MQW stack and the InP spacer thickness between the MQW+SCH combination and the MO metal. In principle, it should be possible to find a single layer structure that has the optimal combination of needed gain in the wells and non-reciprocity caused by the MO metal by scanning through the entire 3D parameter space and using a proper figure

[71] These lattice-matched guiding layers are quaternary InGaAsP layers with a band gap wavelength of 1.03μm ($E_g = 1.03$eV). This is above but very close to the band gap of the barriers. Because of the relatively large band gap of the barriers, there is obvious not much "room to play" as on the one hand from an electrical viewpoint the SCH layers must have a higher band gap than the barrier layers and on the other hand they must be sufficiently lower in band gap than InP in order to maintain a sufficient refractive index difference with this cladding material (and hence a good guiding effect).

of merit to properly compare each parameter space point. An obvious choice for the SOA/isolator figure of merit (FOM) would be

$$\text{FOM} = \frac{g_{\text{QW}}}{\Delta\alpha^m}, \tag{3.78}$$

with g_{QW} the material gain needed in the wells to achieve forward transparency and $\Delta\alpha$ the non-reciprocal absorption shift, and m a well-chosen power expressing the relative importance of the needed amplification versus the obtained isolation. The use of a $m \neq 1$ might seem somewhat unclear, but its purpose is obvious. If material gain is not really an issue then it makes sense to use a $m > 1$ to accentuate the increases in isolation ratio. If on the other hand the needed gain in the QW is the limiting factor for the operation of the device, then a $m < 1$ will stress any change in g_{QW}. For the moment we will use $m = 1$. In the next section we will see how it can make sense to use a FoM with a m different of 1. It is also clear that the use of this FoM does not put any limitations on the absolute value of the QW gain. It may well be that the FoM keeps on decreasing because $\Delta\alpha$ increases faster than g_{QW} or that (even with a $m < 1$) the (global) minimum for FoM in the 3D parameter space is reached at a point where the gain is unreasonably high. Therefore, in any case it is necessary to supplement the FoM with an inequality boundary condition on the QW gain. This is a bit of a weak point, as the previous section 3.2.4.2 has indicated that it is not straightforward to determine the effective material gain. From a theoretical point of view it is possible to use the gain calculations of the previous sections. But these merely give the gain as a function of the experimentally non-accessible carrier density. Therefore without the knowledge of a clear gain vs. total current relationship, the optimization of the layer structure is for a large part based on assumptions concerning the achievable material gain in the QW's.

With all of the above in mind, it was deemed more appropriate to use a slightly different approach towards the optimization of the layer than by simply calculating the FoM by brute force throughout the entire parameter space. The following design strategy has been employed.

The parameter space is two-dimensional and is formed by the thicknesses of the two SCH layers. First of all, a certain thickness for the InP spacer is chosen. For this fixed thickness, the optical mode loss and the optical QW confinement of the unperturbed zero order guided TM mode are calculated within a realistic range of the 2-dimensional parameter space. Subsequently, for the same 2D-range of SCH thicknesses the imaginary part of the non-reciprocal transverse phase shift on the zero order guided TM mode is calculated via the first order perturbation formula (2.120). This allows to plot contours of the isolation ratio for the considered range of the parameter space. Then, in a first order approach the needed material gain for transparency is calculated as the ratio of the forward mode loss[72] to the confinement factor. Choosing now a certain level of attainable material gain and plotting the contours of needed TM material gain for forward transparency will define a 2D region of acceptable SCH thicknesses. Finally, the optimum combina-

[72] obviously found as the zero order mode loss corrected with the imaginary part of the transverse magneto-optic phase shift.

Layer	Composition (dopant, strain ε)	Refractive index (bandgap λ_g)	Thickness
MO layer	$Co_{90}Fe_{10}$	$n = 3.29 - i\,5.54$ $Q = (0.02, 0.03)$ $+i\,(0.02, 0.03)$	100 nm
absorbing contact layer	$In_{0.54}Ga_{0.46}As$ Be: $3 \cdot 10^{19}$ cm^{-3}	$n = 3.6 - i\,0.2$ 1.62 μm	15 nm
transparent contact layer	$In_{0.81}Ga_{0.19}As_{0.41}P_{0.59}$ Be: $1 \cdot 10^{19}$ cm^{-3}	$n = 3.37$ 1.17 μm	100 nm
spacer layer	InP	$n = 3.203$ 0.9 μm	250/300/450 nm
upper SCH	$In_{0.9}Ga_{0.1}As_{0.21}P_{0.79}$	$n = 3.28$ 1.03 μm	optimize
barrier ($\times$4)	$In_{0.93}Ga_{0.07}As_{0.22}P_{0.78}$ $\varepsilon = +0.3$ %	$n = 3.29$ 1.05 μm	22 nm
QW ($\times$3)	$In_{0.48}Ga_{0.52}As_{0.78}P_{0.22}$ $\varepsilon = -1.1$ %	$n = 3.34 + i\,k$ 1.3 μm	12 nm
lower SCH	$In_{0.9}Ga_{0.1}As_{0.21}P_{0.79}$	$n = 3.28$ 1.03 μm	optimize
substrate	InP	$n = 3.203$ 0.9 μm	

TABLE. 3.6: Slab layer structure of the first generation InGaAsP SOA/isolator. Refractive indexes are taken from Refs. [96]. For the optical constants of the CoFe layer the values of pure Co have been used [98]. The uncertainty in the magneto-optical Voigt parameter of CoFe is reflected by the variation considered in the real and imaginary parts of Q. These have been obtained by considering a 50% uncertainty on both real and imaginary parts of the constant for pure Co (found in [6]).

tion of SCH thicknesses is found by locating the maximum of the non-reciprocal phase shift within this region.

In this design strategy, the choice of the InP spacer thickness seems to be arbitrary. This is not the case. Choosing a certain level of attainable material gain[73] implicitly implies a minimum thickness for the InP spacer in order to be able to operate the device in a forward transparent regime (or with a minimal forward insertion loss). In this way, it is clear that it is useless to try to artificially increase the non-reciprocal effect by decreasing the InP spacer thickness, unless very strong TM material gain can be reached in the QW's. For this reason, it was decided to design three different layer structures for the first isolator prototypes. These designs differed in the thickness of the InP spacer layer and hence in the required level of needed material gain for transparent operation. Increasing levels of re-

[73] based on the theoretical gain calculations presented in the previous section.

quired QW material gain in the three designs can be considered as an increasingly higher risk on the transparent operation of the device. For all three designs it was decided to work with a MQW stack having only three wells. That is because at the time of design, only 3QW devices had been successfully grown and characterized. However, it must be noted that this characterization had at that time not yet produced an accurate gain-current relationship. In a later stage, after it became clear that it was possible to grow a MQW strained stack with a higher number of wells, two designs with 6 resp. 9 wells were added. For these only one InP spacer thickness was considered, making a total of five designs.

A final remark should be made concerning the strength of the magneto-optic effect of the used CoFe ferromagnetic metal film. As pointed out, at the time of design, there was only a rough knowledge of the value of the gyrotropy constant g_{CoFe} at the used wavelength of 1300nm. For this reason, the perturbational non-reciprocity calculations have actually been mapped in a 4-dimensional parameter space! That is: on top of the 2D-parameter space formed by the SCH thicknesses, the real and imaginary parts of g_{CoFe} were varied within reasonable ranges.

Simulation results – Designs The layer structure to be optimized is detailed in Table 3.6. Fig 3.22 shows an example of an optimization simulation. These figures deserve some explanation. The needed TM material QW gain to achieve forward transparency is a function of both the modal loss and the optical confinement in the QW's via $g_{\mathrm{QW}} = \alpha_{\mathrm{mod,TM}}/\Gamma_{\mathrm{QW}}$. At high SCH thicknesses the dependence of the needed material gain on these thicknesses is quasi-identical, meaning that a decrease in either $t_{\mathrm{SCH_u}}$ or $t_{\mathrm{SCH_l}}$ causes a similar increase in g_{QW}. Because for these high SCH thicknesses the modal loss will vary little, this is an indication that in this case the performance is confinement-limited. For a decreasing $t_{\mathrm{SCH_u}}$ the dependence of g_{QW} on the lower SCH gradually decreases. This indicates that the increasing modal loss (due to the gradually decreasing distance between the guiding region and the metal) dominates the performance of the device. In other words, in this situation the variations in the optical confinement due to variations of $t_{\mathrm{SCH_l}}$ are not "strong" enough to counteract the quick increase in modal loss.
An intuitive reasoning would suggest that the non-reciprocal absorption shift has the same functional behaviour in the 2D parameter space as the needed QW transparency TM gain. After all, in the lowest order approximation, in the confinement-dominated regime any change in the SCH thicknesses causing a variation in the QW confinement will cause a similar but opposite change near the metal. Thus in this regime, $\Delta\alpha$ will follow the same trend as g_{QW}. In the loss-dominated regime (at decreasing upper SCH thicknesses), a similar reasoning leads to false conclusions. One would expect that at low upper SCH thicknesses the closeness of the MO metal would lead to continuously increasing $\Delta\alpha$, only slightly "tempered" by an increase of the lower SCH thickness, just as was observed for g_{QW}. However a radical different behaviour is seen in Fig. 3.22(b). As the upper SCH decreases (for a certain fixed $t_{\mathrm{SCH_l}}$), at first $\Delta\alpha$ indeed increases as expected, however only to reach a maximum, followed by a rapid decrease after a further reduction of $t_{\mathrm{SCH_u}}$!

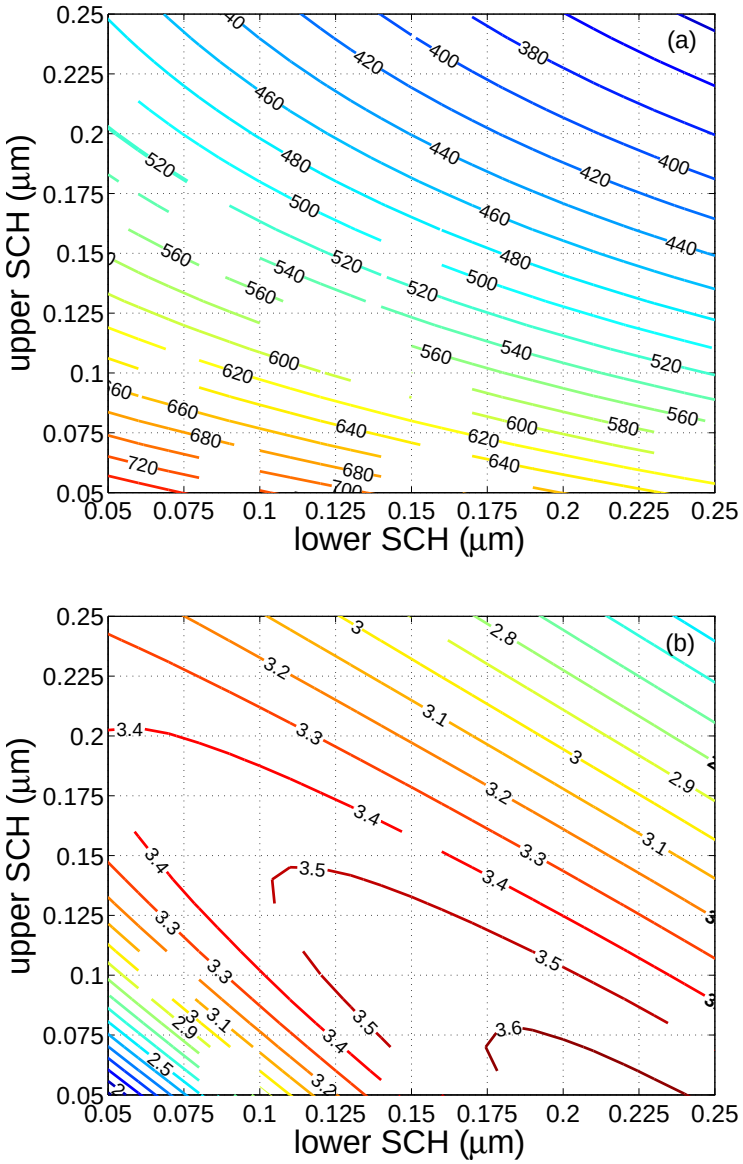


Fig. 3.22: Typical optimization simulations in the 2D parameter space formed by the two SCH thicknesses ($t_{\mathrm{SCH_u}}, t_{\mathrm{SCH_l}}$). The InP spacer thickness is 300nm. The layer structure is that of Table 3.6 at a worst case value for $Q_{\mathrm{CoFe}} = 0.02{+}i0.02$. (a): needed TM material QW gain g_{QW} (in 1/mm) for forward transparency; (b): optical absorption difference $\Delta\alpha$ (in dB/mm) between forward and backward guided TM ground mode.

Moreover this behaviour is observed to be dependent on the precise choice of the lower SCH. For increasing thicknesses of the lower SCH, the final decrease of $\Delta\alpha$ as a function of $t_{\mathrm{SCH_u}}$ occurs at continuously smaller values of $t_{\mathrm{SCH_u}}$. It is not so difficult to understand what is going on here. A similar behaviour of $\Delta\alpha$ has already been observed in the benchmark calculations of the previous chapter. In Fig. 2.20 it was seen how for a fixed InP spacer thickness (as is the case here) the isolation as a function of the guiding core thickness increases to a maximum with decreasing core thickness, and subsequently drops rapidly. There it was explained how this behaviour is a result of the approach towards cut-off of the guided TM mode due to the asymmetry of the waveguide layer structure. As in the here considered MQW layer structure the guiding layer is predominantly determined by the thicknesses of the SCH layers, the observed $\Delta\alpha$ behaviour in Fig. 3.22 has exactly the same origin as in Fig. 2.20, namely an approach towards cut-off at decreasing thicknesses for either one of the SCH layers. A very convincing argument proving that it is indeed this mechanism that causes the decrease of $\Delta\alpha$ is found by the observation that the points of onset of the $\Delta\alpha$-decrease all lie on the straight line in the $(t_{\mathrm{SCH_u}}, t_{\mathrm{SCH_l}})$-space given by $t_{\mathrm{SCH_u}} + t_{\mathrm{SCH_l}} \approx 0.25\mu$m. In other words, the decrease starts as soon as the total thickness of the guiding region drops below $\approx 370\mu$m (the thickness of the QW's + buffers is $\approx 120\mu$m). This is about the same order of magnitude as observed in Fig. 2.20.

In order to choose the optimum combination of SCH thicknesses in the above simulation example, one should in principle choose that point which minimizes $\frac{g_{\mathrm{QW}}}{\Delta\alpha}$. In this point the needed QW material gain will be most effectively used. It is not difficult to see that this point will lie on the above mentioned "straight" line $t_{\mathrm{SCH_u}} + t_{\mathrm{SCH_l}} \approx 0.25\mu$m, i.e. the locus connecting the positions of $t_{\mathrm{SCH_u}}$ which maximize $\Delta\alpha$ as $t_{\mathrm{SCH_l}}$ varies (or vice versa). At the same time it is clear that along this line $\Delta\alpha$ varies by less than 5%. Actually this indicates that the layer structure is particularly tolerant towards the design FoM. The ideal point obtained in this manner would be (0.075μm,0.175μm) with an isolation of about 3.45dB/mm at a needed material gain of 527/mm. It is known that this is a very high material gain. Because of the uncertainty at the time of design about the experimentally achievable gain, and because the above calculations have revealed the rather limited sensitivity of $\Delta\alpha$ towards the SCH thicknesses, it was decided to assume a maximum attainable material gain of 450/mm (for this particular InP spacer thickness). The design point is then found by choosing the SCH combination on this 450/mm contour corresponding to the highest $\Delta\alpha$. This procedure might seem arbitrary. After all, we could have chosen a completely different value for maximum achievable material gain, and as a result gotten an entirely different SCH combination. That is of course an appropriate remark. However, given the low sensitivity of $\Delta\alpha$ towards SCH variations (order of 10% for variations up to 50nm around the optimum point), a not perfectly optimized choice would predominantly result in a non-optimal use of the available material gain and not in a gross miscalculation of the isolation behaviour. In any case, the choice of the attainable maximum material gain remains indeed somewhat arbitrary. That is also why three different designs have been made. These differed in the assumptions of

maximum attainable gain. Should for instance a higher gain be attainable (than the here assumed 450/mm), then it is obvious that it is advisable to also at the same decrease the thickness of the InP spacer and repeat the above procedure. By how much is again rather arbitrary. In any case all of the above is a direct consequence of the many experimental uncertainties at the time of design.

Table 3.7 summarizes the design choices made and the corresponding assumed QW material gain, InP spacer thicknesses and predictions for the isolation ratios (within the limits of the uncertainty on Q_{CoFe}). The 6 and 9 QW designs were added at a later stage. As a final remark it should be added that, if the needed gain for forward transparency in the chosen design can for any reason not be obtained, then the device will operate as a non-reciprocal absorber with the same non-reciprocity as designed but with a forward modal loss that equals $\Gamma^*_{\mathrm{TM}}(g - g^*_{\mathrm{QW}})$, with g, g^*_{QW} and Γ^*_{TM}, the actual obtained QW material gain, the design transparency gain and the QW TM confinement at the design point.

design	t_{InP}	$t_{\mathrm{SCH_l}}$ (nm)	$t_{\mathrm{SCH_u}}$ (nm)	$\Delta\alpha$ (dB/mm)	g_{QW} (1/mm)
safe	450	120	100	1.9–2.8	300
normal	300	150	150	3–5	450
risk	250	155	170	4–6	600
6QW	300	100	50	4–5.5	300
9QW	300	50	50	3.5–5.3	170
CoFe	t_{CoFe} (nm)	Q_{CoFe} worst case		Q_{CoFe} best case	
	100	0.02+i0.02		0.03+i0.03	

TABLE. 3.7: Design values of the 5 first-generation demonstrators. The epitheta used to describe the 3QW designs are an expression of what was thought at the time of design as being a safe (etc. ...) amount of needed TM QW material gain for transparency based on the gain calculations of Fig. 3.21.

FoM optimization on 6QW device It has been pointed out in detail that the above design procedure is in a way arbitrary due to the numerous uncertainties present. In order to illustrate what a "correct design" (based on minimizing $g_{\mathrm{QW}}/\Delta\alpha$) would predict as the optimum SCH values, consider Fig. 3.23. Here, the evolution of the optimum SCH points and the corresponding transparency QW material gain and isolation ratio are plotted as a function of the InP spacer thickness for a 6QW device with the same layer structure as in Table 3.7 (except for the number of wells and bariers). The main difference with the calculations of the previous paragraph is, next to the different optimization procedure, the fact that gain in the QW's has been taken rigorously into account. Thus, instead of calculating g_{QW} as the ratio of the modal loss and the modal TM QW confinement in an unpumped device, the imaginary part of the QW index itself has been varied until transparency is obtained in the forward direction. And this has been done for every point in a sufficiently large portion of the SCH-space (and this again

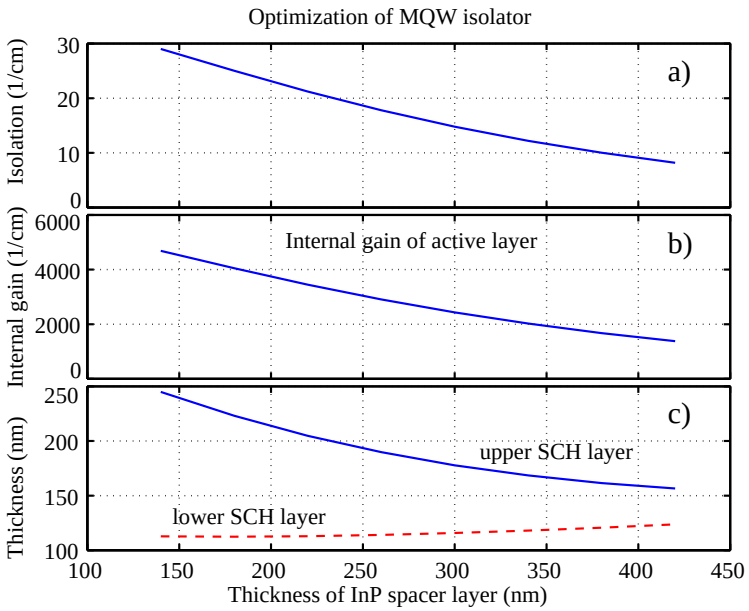


Fig. 3.23: $\frac{g_{QW}}{\Delta\alpha}$ minimization on a 6QW device.(a): isolation ratio at optimum point;(b): forward transparency gain at optimum point;(c): optimum SCH thicknesses

for every considered InP thickness) in order to locate the minimum of $g_{QW}/\Delta\alpha$ at each considered t_{InP}. This is numerically a very intensive task. In principle, for every simulation point (i.e. each combination of SCH and InP spacer thicknesses) a vectorial modal calculation and a MO perturbation calculation[74] has to be performed for every gain value until transparency is reached. On top of that a sufficiently large SCH space must be scanned in order to track the evolution of the optimum point[75]. Without any "accelerating" tricks such a calculation as the one depicted in Fig. 3.23 took easily four full days (≈100 hours) on a Pentium III machine.

It can be seen from this figure that the optimization procedure used successfully leads to a trend where the needed QW gain increases less quickly than the isolation ratio for decreasing InP spacer thickness. This is achieved by appropriately varying the SCH thicknesses, so that for each InP thickness the needed gain is optimally used. Looking at the obtained optimum values for the SCH thicknesses, it is immediately obvious that this is mainly achieved by exponentially increasing the upper SCH thickness, while the optimum lower SCH remains quasi-constant. This indicates, as already observed, that at a low InP spacer thickness (with the

[74] assuming that a change in imaginary part of the QW index might influence the confinement near the metal.

[75] In principle this procedure could be alleviated by using every found optimum point as a starting point to find the next one.

accompanying high modal losses to be compensated), the lower SCH hardly "lends a helping hand" to keep the confinement in the well at reasonable levels, and the upper SCH has to do the trick all by itself. Its exponential increase achieves this by a combined effect of at the same time increasing the distance to the metal and keeping the confinement to "normal" levels.
Comparing the values obtained here with those tabulated in Table 3.7 for the 6QW design (for a 300nm InP spacer thickness) demonstrates that in this calculation the needed gain is more efficiently used, but the designed isolation ratio is not necessarily a lot higher (15/cm or thus 6dB/mm versus 4–5.5dB/mm). Finally, it should be remarked, that in a later stage after these design calculations had been done, an approximate gain-current relationship has been experimentally obtained on these 6QW -1% strained InGaAsP MQW gain layers (albeit with symmetrical SCH layers). These are detailed in section 4.2.1.2 of the following chapter. The main conclusion of these experiments is that it is very hard to achieve a material gain much higher than 1500/cm and more importantly that this situation doesn't change significantly with increasing wells! This indicates that unless a more performant active layer structure is used the limiting value for the isolation ratio of this SOA/isolator device fabricated in InGaAsP will be in the neigbourhood of 4dB/mm at maximum, or a 6 mm long device to obtain 25dB of isolation.

3.3.2.2 Generation 2: $Co_{90}Fe_{10}$/InAlGaAs

Input parameters By the time of design of the second generation devices an in-depth experimental characterization of the gain performance of both InGaAsP and InAlGaAs tensile-strained MQW regions had been performed, and the (magneto-)optical constants of $Co_{90}Fe_{10}$ had been accurately assessed at 1300nm (see sections 4.2.1 and 4.1.1 of the following chapter). It was found that the InAlGaAs material system clearly outperforms the phosphide-based system when it comes to gain performance, allowing a thinner InP spacer thickness (and thus higher isolation ratios) for quasi-unchanged injection current into the device. The characterization of the equiatomic CoFe alloy was at the time of this design still ongoing. But there were indications that the relevant imaginary part of the gyrotropy constant might be a factor 2 stronger (this was proven to be correct in a later stage see section 4.1.1).

As opposed to the first generation design, this input allows a much more accurate and practical design. First of all there is no more uncertainty on the magneto-optical constant of CoFe. Secondly and most importantly, we have access to an experimentally determined gain-current relationship. This latter allows to express the gain in the wells directly in terms of the injected current into the device, or, in other words, to derive the forward transparency condition as a function of an accessible device operational parameter. Some cautionary remarks have to be made though. The gain-current relationship obtained on -1.25% strained InAlGaAs wells and described in detail in the next chapter, expresses the modal gain of a 6QW laser structure as a function of current, and not the material gain. But this is not the main problem, as in principle it is perfectly viable to obtain

the material gain in the wells by dividing the obtained modal gain by the QW confinement factor in the characterized 6QW laser structure. The main problem is the fact that the obtained logarithmic gain-current relationship is inherently dependent on the internal quantum efficiency of the device. As the internal quantum efficiency can vary greatly with the precise device geometry (and number of wells) it is not so logical to use it for the modelling of a device that doesn't have exactly the same layer structure (i.e. thicknesses of guiding layers and number of wells) as the characterized laser. These remarks are however only minor drawbacks and it can be safely assumed that a scaling transformation will correctly (or with acceptable accuracy) transform the gain-current relationship from a 6 wells structure to that of a structure with a differing number of wells. The most significant problem is that, as explained at length in section 3.2.4.2, the internal quantum efficiency is not constant with current. At very high modal gains and thus high injected current densities the internal quantum efficiency drops significantly. It is almost impossible to measure this variation of the internal quantum efficiency with current. In any case this implies that the gain-current relationship cannot be described by the simple (3.72) with a constant internal quantum efficiency. As a result we have to be aware of the fact that the extrapolated use of the experimentally assessed gain-current relationship at material gains which greatly exceed the ones in the laser measurements, will most definitely underestimate the needed current density. There is no way out of this.

Design Strategy With these input parameters, the most logical equivalent of the FoM of (3.78) seems to be

$$FoM = \frac{J_{\mathrm{tr}}}{\Delta\alpha}, \tag{3.79}$$

with J_{tr} the needed total current density to achieve forward transparency. This FoM is clearly a better indication of the operational characteristics of the device. Indeed, it is not difficult to see that this definition is nothing else but the total absolute current needed to achieve a certain fixed amount of isolation (at a fixed width of the waveguide). However, it doesn't seem logical to take only the total operational current of the device as a design criterion. This could for instance cause to favour a certain design point that leads to a rather large total device length. A less optimal design point (according to (3.79)), with a larger $\Delta\alpha$, but needing relatively more J_{tr} to achieve this, may well seriously reduce the device length without considerably increasing the total current. Therefore, there are many good reasons to include also the total device length in the design criterion, the most obvious being the strive towards miniaturization, the increased stability, the larger component count per wafer (implying a higher yield), and in particular for this device the improved magnetic stability, due to the lower magnetic shape anisotropy. In view of this, a better design criterion is the minimization of the $I_{\mathrm{tot,tr}} \times L$ product of the device. This is equivalent to the redefinition of the FoM as

$$FoM = I_{\mathrm{tot,tr}} \times L \sim J_{\mathrm{tr}} L^2 = \frac{J_{\mathrm{tr}}}{(\Delta\alpha)^2} \tag{3.80}$$

This an example where it makes sense, owing to more practical device considerations, to use an exponent different from 1 for $\Delta\alpha$ in the FoM definition.

For the design optimization itself we could again try and minimize this latter FoM over the entire parameter space formed by the InP spacer thickness, and both InAlGaAs guiding layers. But as indicated in the 6QW FoM optimization of the previous section, such a brute force approach is very time consuming. Next to that, one important drawback of such a design procedure is that it yields a different SCH combination for every different thickness of the InP spacer. This is quite logical, as the FoM optimization procedure naturally seeks the most optimal combination of the three thicknesses $t_{\mathrm{SCH_l}}, t_{\mathrm{SCH_u}}$ and t_{InP} that use the available gain in the most advantageous way. However, from a technological viewpoint it is interesting to be able to work with an identical active region (i.e. MQW+SCH's) for all processed structures. In this way, one can directly experimentally asses the influence of the most important device parameter, namely the thickness of the InP spacer, by a single active layer growth and a series of different overgrown InP spacer layers. One can object that this will not yield the optimum layer structure. However, simulations, like the one depicted in Fig. 3.22(b), have shown that purely from a non-reciprocal viewpoint $\Delta\alpha$ is rather tolerant towards variations in the SCH thicknesses[76]. Therefore, it was chosen to focus the design on finding an unique layer structure for the active region which could be used for a whole range of InP spacer thicknesses. For that purpose the following design strategy was devised

1. Choose the start value of the InP thickness to be 500nm, which is a definite upper limit for this parameter.
2. Choose a relatively "low" desired current density value $J_{\mathrm{tr,t_{InP}=500nm}}$ for which forward transparency should occur at the starting thickness of $t_{\mathrm{InP}} = 500\mathrm{nm}$.
3. Look for the SCH thicknesses that actually yield forward transparency at the chosen starting value of J_{tr} and t_{InP}.
4. On this contour in SCH space pick the SCH combination that results in the highest non-reciprocal absorption difference $\Delta\alpha$.
5. Keep from here on these SCH values fixed, and decrease the value of the InP layer thickness while tracking how the transparency current density J_{tr} and the non-reciprocal absorption difference $\Delta\alpha$ change.
6. To examine the dependence of these results on the starting conditions, repeat this procedure for different starting values of $J_{\mathrm{tr,t_{InP}=500nm}}$.
7. Finally, compare the results obtained in this way using the FoM definition of (3.80)

[76] It will be seen in the next paragraph that this device is in general very tolerant towards the thickness of the guiding layers.

Layer	Composition (dopant, strain ε)	Refractive index (bandgap λ_g)	Thickness
MO layer	$Co_{90}Fe_{10}$	$n = 4.35 - i\,4.76$ $Q = 0.0214 + i\,0.0442$	100 nm
absorbing contact layer	$In_{0.54}Ga_{0.46}As$ Be: $3 \cdot 10^{19}$ cm^{-3}	$n = 3.6 - i\,0.2$ 1.62 μm	15 nm
transparent contact layer	$In_{0.81}Ga_{0.19}As_{0.41}P_{0.59}$ Be: $1 \cdot 10^{19}$ cm^{-3}	$n = 3.37$ 1.17 μm	100 nm
spacer layer	InP	$n = 3.203$ 0.9 μm	optimize
upper SCH	$In_{0.86}Ga_{0.14}As_{0.31}P_{0.69}$	$n = 3.34$ 1.05 μm	optimize
barrier (×7–9)	$In_{0.62}Al_{0.3}Ga_{0.08}As$ $\varepsilon = +0.6$ %	$n = 3.46$ 1.05 μm	20 nm
QW (×6–10)	$In_{0.34}Al_{0.14}Ga_{0.52}As$ $\varepsilon = -1.25$ %	$n = 3.57 + i\,k$ $k = 0.00483 \ln\left(\frac{6}{N_w}\frac{J_{\text{tot}}}{352}\right)$ 1.3 μm	10 nm
lower SCH	$In_{0.86}Ga_{0.14}As_{0.31}P_{0.69}$	$n = 3.34$ 1.05 μm	optimize
substrate	InP	$n = 3.203$ 0.9 μm	

TABLE. 3.8: Slab layer structure of the second generation InAlGaAs SOA/isolator. Refractive indexes are taken from Refs. [96]. The (magneto-)optical constants of the CoFe layer have been obtained through magneto-optic generalized ellipsometry(see section 4.1.1 and [97]). The gain-current relationship used to determine the material gain of a single strained well as a function of the <u>total</u> inject current density has been obtained through laser measurements on a 6QW BA laser (see section 4.2.2) with the same composition of the guiding layers. It is assumed that the same gain-current relationship can be used for a structure with a higher number of wells through a simple scaling procedure $J_{\text{tot}} \rightarrow \frac{6}{N_w} J_{\text{tot}}$, with N_w the number of wells. In other words it is assumed that the internal quantum efficiency doesn't change with increasing wells. In the next chapter, this is assumption is experimentally proven to be valid up to at least 9 wells.

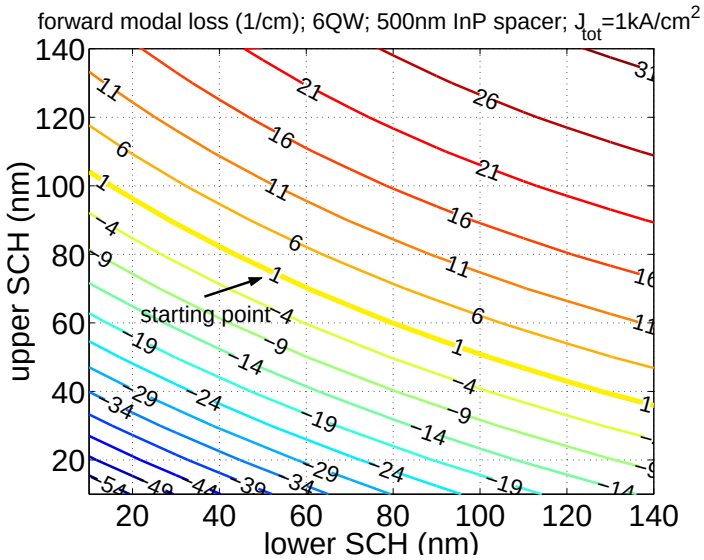
This procedure might seem laborious, but in fact it is an intuitive economic equivalent of scanning the entire parameter space for a minimization of the FoM. Indeed, this procedure samples the SCH parameter space by choosing physical feasible starting points and tracking their optimality. The feasibility of the sampled starting points lies in the fact that at a spacer thickness which is relatively high for the considered device, these starting points all correspond to SCH combinations with both a relatively low forward transparency current density (at that "thick" 500nm of InP) and a high non-reciprocal absorption shift. Because the needed current density for forward transparency will always exponentially increase with

decreasing InP thickness, it is only by choosing a starting point with a relatively low value for $J_{\rm tr}$ that one can expect to obtain a considerable improvement in FoM with decreasing InP thickness. The simulation examples of the next paragraph will illustrate this.

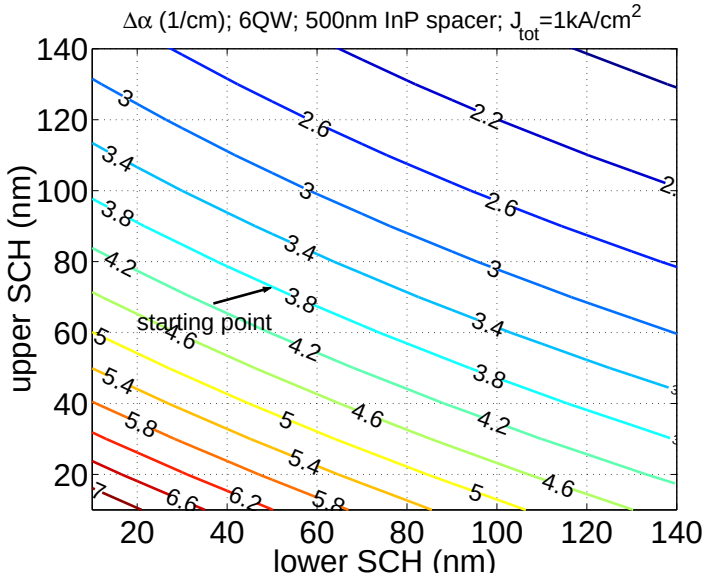
Simulation results – Designs The simulated layer structure is detailed in Table 3.8. The gain in the wells has obviously been incorporated using the known gain-current relationship. The needed current density for forward transparency has been obtained by rigorously using the corresponding imaginary index for the wells while doing the modal calculations, and not by a confinement calculation of the needed material gain and a posteriori inversion of the gain-current relationship.

Fig. 3.24 gives an example of the choice of a feasible set of starting points for a 6QW design. Fig. 3.24(a) shows the forward modal loss in the SCH space for an injected current density of 1kA/cm^2, while in Fig. 3.24(b) the non-reciprocal absorption difference under the same conditions is plotted. It can be seen that on the SCH-contour close to transparency, the point (50nm,70nm) has maximum $\Delta\alpha \approx 4.0$/cm. According to the procedure outlined above this SCH combination is chosen as starting point.

The next step is to track how $J_{\rm tr}$ and $\Delta\alpha$ change as the thickness of the InP spacer decreases, while keeping this combination of thicknesses fixed. This is plotted in Fig. 3.25.
Of course $J_{\rm tr}$ and $\Delta\alpha$ increase quasi-exponentially with decreasing InP thickness. However, what is really striking is that the rate of this increase is fundamentally different between $\Delta\alpha$ and $J_{\rm tr}$. Actually, $J_{\rm tr}$ is not well-fitted by a simple exponential function of $t_{\rm InP}$. While the non-reciprocal phase shift has a well behaved exponential increase over the whole range of plotted InP thicknesses, the transparency current density shows a sudden steep onset from about 330nm. As a result, for this specific design point, from about 10kA/cm^2 $\Delta\alpha$ starts saturating as a function of the needed transparency current density, as can be seen in Fig. 3.25(b). This is nothing else but a patent indication that the available QW gain is increasingly less optimally utilized. Or in other words that the FoM of the device has passed a minimum! By sampling the SCH parameter space for other starting points and repeating the above procedure, one may find other combinations where the saturating behaviour sets in at higher values of $\Delta\alpha$ and/or higher values of $J_{\rm tr}$. An example of such sampling, is illustrated in Fig. 3.26.

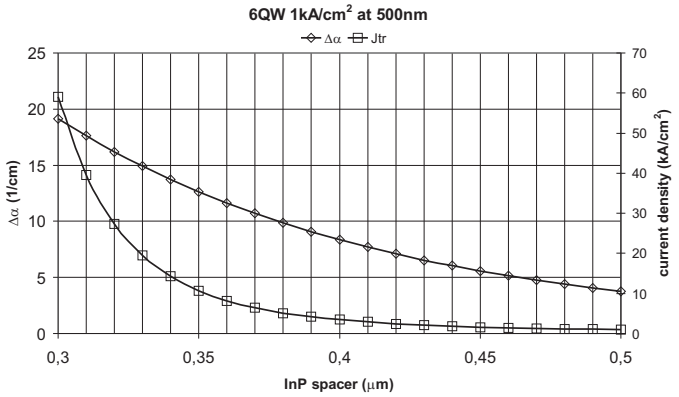


(a)

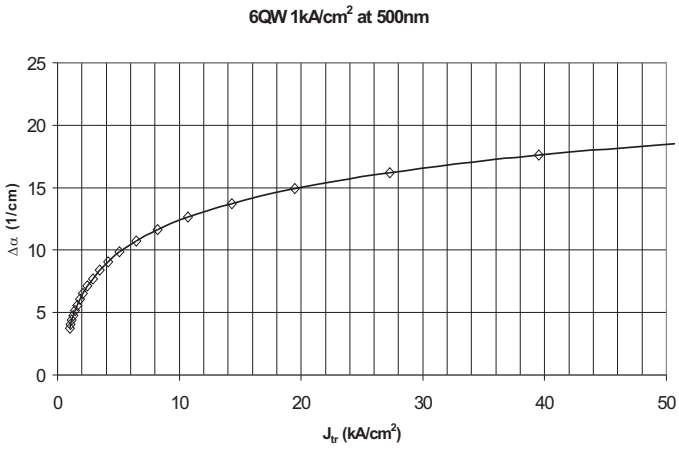


(b)

Fig. 3.24: Sampling the SCH space for a feasible starting point. After choosing first of all a "low" value for $J_{\mathrm{tr},t_{\mathrm{InP}}=500\mathrm{nm}}$ (here 1kA/cm^2) (a): first the "forward transparency" contour is located and then (b) the point with maximum $\Delta\alpha$. This procedure leads to a starting point near (50nm,75nm).



(a)



(b)

Fig. 3.25: Tracking the evolution of J_{tr} and $\Delta\alpha$ for the chosen starting point of Fig. 3.24 (a): as a function of InP spacer thickness and (b): plotted as a function of eachother after elimination of the InP thickness.

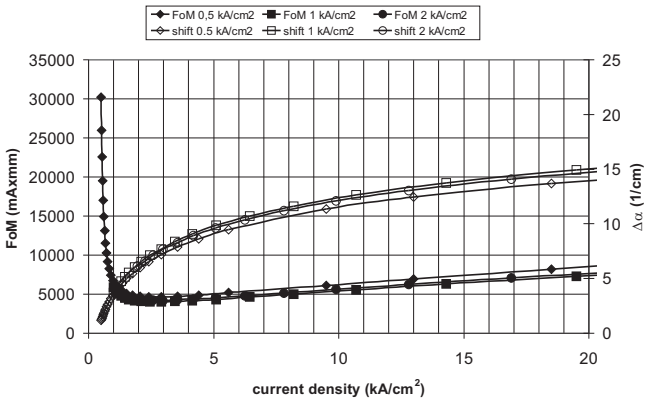
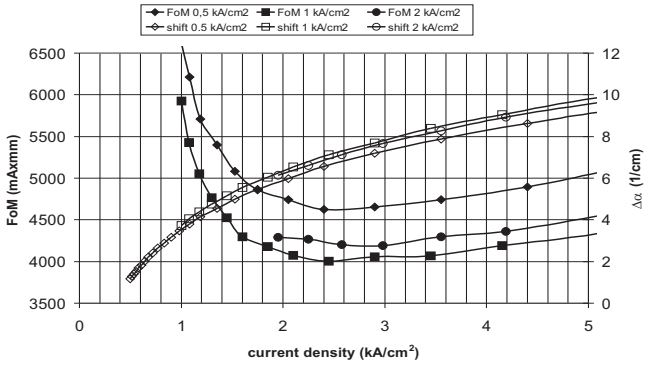
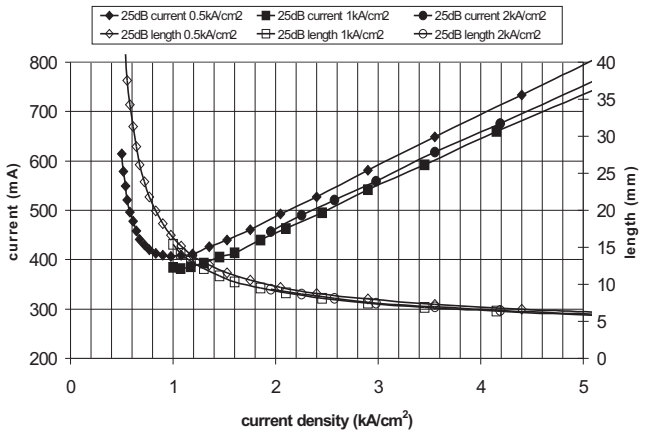


Fig. 3.26: Comparison of the influence of the choice of starting point. For clarity only three choices are plotted each obtained by considering a different value for the needed current density for forward transparency at 500nm. These are resp. 0.5kA/cm^2, 1.0kA/cm^2 and 2.0kA/cm^2 and the corresponding starting points are (180nm,175nm), (50nm,75nm) and (15nm,30nm). The FoM ((3.80)) is calculated as the absolute total current-length product (in mA×mm) to obtain 25dB of isolation in a 2,5 μm wide device. Note that it actually is not important what width or what total isolation value is pursued for the device. These are just global factorial constants in the FoM calculation. Actually, it doesn't make sense to use a certain value for the width as the simulations are done in a slab model.

The starting values all belong to entirely different regions of the SCH space. Still, it is very striking that the performance of the device doesn't change significantly with the chosen starting point. For all three starting values the FoM's have a minimum value close to eachother and all three reach this minimum at approximately 2.5 kA/cm^2. This indicates that the SOA/isolator is in general very tolerant towards the SCH guiding layers. Still there is some optimum detectable. This is better visible by considering the detail of Fig. 3.26 depicted in Fig. 3.27(a). By choosing a relatively safe SCH starting point (needing only 500A/cm^2 current density at 500nm), the use of the available gain can still be drastically improved by decreasing the InP thickness (or thus increasing the needed transparency current density), as can be seen from the drastic decrease of the FoM. However as soon as the needed current density approaches 3kA/cm^2, the increasing saturation of the gain-current performance of the wells prevents a further decrease of the FoM. This is perfectly logical, because as soon as gain saturation of the wells is setting in the increase of the needed transparency current density at each next lower value of the InP thickness is so huge that also the corresponding non-reciprocal absorption loss starts saturating as a function of the needed current density. In a way, the above explanation indicates that the logarithmic behaviour of the $(\Delta\alpha, J_{\mathrm{tr}})$-relationship is nothing else but a direct reflection of that same behaviour of the gain-current relationship of the wells. This implies that, since $\Delta\alpha$ will always start saturat-



(a)



(b)

Fig. 3.27: (a): Detail of Fig.3.26; (b): Total length (mm) and current (mA) needed for the three considered design points, in order to achieve 25dB of isolation in a 2.5 μm wide device (under the hypothetical assumption that the confinement in the wells and near the MO metal doesn't change in going from a slab to a ridge waveguide).

ing around 2–3 kA/cm^2, a too "safe" choice for the starting point may prevent $\Delta\alpha$ to reach to higher values. On the other hand, a too enthousiastic choice of the starting point (for instance needing at 500nm already a transparency current density close to the saturation value of 2–3kA/cm^2, like the 2kA/cm^2 simulation example), will very quickly limit the increase of $\Delta\alpha$. This is also visible in the FoM of such a starting point. It hardly improves when departing from its starting value.

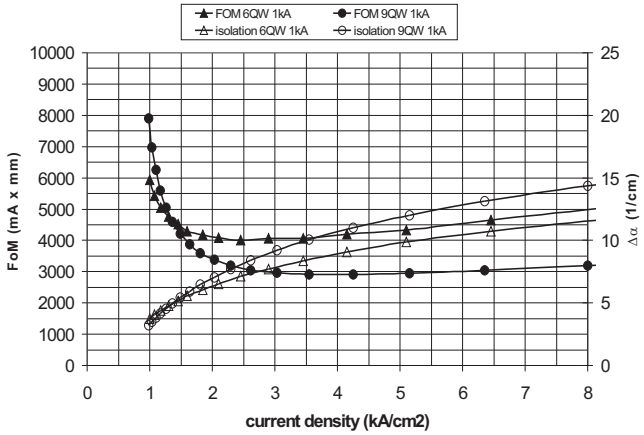
Fig. 3.27(b) shows what all of the above implies for the total current and length needed to achieve an isolation ratio of 25dB. It the neigbourhood of the optimum

point, we can expect an isolator length of about 8mm and a total current of 500mA. A shorter device can be realized, but our FoM consideration indicate that such a device will not optimally use the injected current. Therefore a shorter device will come at the expense of heavily increased current. The same argument holds for a lower operation current. When compared to the results of the InGaAsP design (see Fig. 3.23), one may be surprised. There a 25dB (6QW) isolator was predicted with a length of only 6mm, while it was clearly set out in the beginning of this section that InAlGaAs would behave much better. However, it mustn't be forgotten that there was no gain-current relationship available for the simulation of Fig. 3.23. The prediction was based on the fact that the needed InGaAsP 1500cm^{-1} material gain was the maximum that could be expected based on the measurements presented in the next chapter (see section 4.2.1.2). Actually, in these measurements it is found that 5.5kA/cm^2 is needed for that amount of gain, which corresponds to a total current of almost 800mA! If that same current would be used in this design, the device length would only be 5mm (not visibly on the graph of Fig. 3.27).

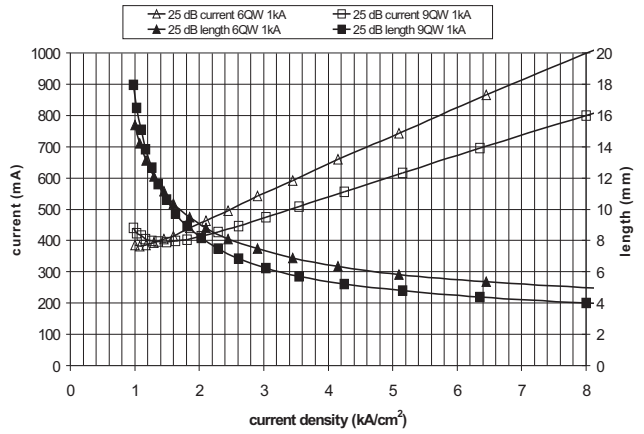
Still, the results of this design are somewhat disappointing, therefore it was checked whether a big improvement could be achieved by increasing the number of wells up to 9, in the knowledge that an increased confinement in the wells will allow a thinner InP cladding and thus a higher non-reciprocity. Fig. 3.28 presents the results of these simulations. The optimum SCH combination (using the above procedure) for a 9QW design is compared to the best 6QW design. It can be seen how the FoM minimum (where the gain-current starts saturating) shifts to a current density exactly 50% higher than in the 6QW case, which is a logical consequence from the fact that a scaled version of the gain-current relationship is used. Because the 9QW has a much higher TM optical confinement in the wells, at this (scaled) position of the minimum FoM a much higher total forward loss can be compensated and thus a higher non-reciprocity can be achieved. Next to that, the FoM also doesn't increase as quickly as in the 6QW case once the minimum has been reached, suggesting that the design point is much more tolerant. This is of course because the higher number of wells scales the current density per well to a lower value, "slowing down" the gain saturation. The results show a considerable increase when switching to 9 wells. And the values for the optimum SCH combination (0nm(!),15nm) suggest that the increased confinement almost makes the role of the guiding layers superfluous. In the neighbourhood of the optimum FoM a 25dB isolator is obtained in a length of just 5mm and needing a total operation current of about 500mA. A clear improvement both over the InGaAsP design and the 6QW InAlGaAs design.

Table 3.9 summarizes the design choices made. It was decided to grow three versions of the optimum 9QW structure, one near the optimum FoM and two relatively far beyond the optimum FoM. The latter two were added to check on the one hand whether the used gain-current relationship could indeed be extrapolated towards these high values, and on the other hand to simply obtain a higher $\Delta\alpha$ even at the cost of a much increased and not optimally utilized J_{tr}.

In conclusion, the InAlGaAs SOA/isolator design predicts as expected a considerable increased performance. However, the theoretical promise of a very compact



(a)



(b)

Fig. 3.28: Comparison of the optimum 6QW (50nm,75nm) and 9QW designs (0nm,15nm) (a): FoM and $\Delta\alpha$; (b): total length and current for a 25dB isolator.

integrated SOA/isolator operating at a relatively low current as originally theoretically suggested by Nakano seems somehow far-fetched (see Chapter1). At this time, in an idealized slab model of the device supplemented with experimentally assessed QW gain performance measurements and relatively accurate data for the magneto-optical constants of the MO metal, the best one can hope for is a 25dB isolator with a length of 5mm and a total operation current of at least 500mA. However, as will be shown in the next chapter, there is still a lot of room for improvement of the gain behaviour of strained InAlGaAs wells. At the time of writing an improvement of about 50% of the differential gain of tensile-strained

t_{InP} (nm)	$t_{\mathrm{SCH_l}}$ (nm)	$t_{\mathrm{SCH_u}}$ (nm)	$\Delta\alpha$ (dB/mm)	J_{tr} (kA/cm^2)
345	0	15	5.6	6
295	0	15	8.25	20
280	0	15	9.55	30

TABLE. 3.9: Design values of the 3 second-generation 9QW InAlGaAs demonstrators. The other layer structure parameters are summarized in Table 3.8.

InAlGaAs wells (with respect to the wells used in this design) has been measured. A very rough approximation based on the here presented results would predict a further decrease of the length down to 3.5mm and of the operation current down to 350mA. Next to this, it has also been assessed that the equiatomic $Co_{50}Fe_{50}$ has an approximately 60% stronger MO behaviour. In any case, it seems that the ultimate performance of this device will not do better than an operation current of about 200mA in a device length of 2–3 mm.

3.4 Conclusion

This concludes the extensive chapter on all the theoretical aspects of the design and the characterization of the building blocks of the SOA/isolator.

The subtleties of the measurement of the very small magneto-optical constant of magneto-optic materials have been reviewed. Two measurement methods have been explained in depth. The characterization measurements using both techniques will be treated in the next chapter. It has been explained why an ab initio prediction of the magneto-optic strength of magnetic materials is quasi-impossible, and why there is reason to believe that CoFe is probably potentially the best candidate for our application.

The aspect of magnetic anisotropy plays a very important role in the studied device. It has been explained how the mechanisms of anisotropy determine the permanent magnetic properties of the thin-film magnetic films, and how these properties can be somehow tuned by pre- and/or post-treatment. For this device the anisotropy introduced by the high aspect ratio rectangular shape of the magnetic metal contact might be a severe problem, and some design criteria have been given to minimize the influence of this shape anisotropy. The characterization methods used to asses the strength of the anisotropy have been reviewed, introducing the "improved" squareness parameter. The results of such characterization measurements will be detailed in the next chapter.

It has been explained why and how the realization of a good Ohmic contact on III-V semiconductors is often more a question of achieving extreme dopant concentrations in the semiconductor than of finding a suitable metal, certain exceptions notwithstanding. For this reason it is believed that the realization of an Ohmic contact with a CoFe alloy will not pose a great problem. Its resistance may vary though anywhere between several Ohms and hundredths of Ohms. From a thermal viewpoint this implies a great variation of power consumption. Therefore,

considerable attention has been paid to an accurate characterization of the contact resistivity. Experience has taught that this aspect is often overlooked and that the traditional characterization methods tend to underestimate the resistivity by using characterization structures not fitted for the contacts under test. It has been explained how and why a Cross Kelvin Resistor is the most ideal characterization structure.

The second key aspect of the studied device, being the need for a very strong TM-polarization selective active gain region, has been studied in depth. It has been extensively explained why band structure engineering through the incorporation of tensile strain in the QW's can greatly enhance their gain performance. Still nowadays the use of tensile strain instead of compressive strain is very little seen, in spite of the utterly convincing theoretical predictions of a much better gain behaviour. It has been explained what material systems are at our disposal for realizing good tensile-strained QW's with an emission wavelength of 1300nm. As long as the InGaNAs material system is not fully mature, it seems that there is no competitor for strained InAlGaAs QW's. From a theoretical viewpoint the band offsets in this system are almost the mirror image of those in InGaAsP, namely a high CBO and a low VBO. Especially for tensile-strained wells this would imply record improvements in the gain behaviour due to the complete absence of heavy hole states in the valence band. All these theoretical predictions are experimentally assessed in the next chapter using the described characterization methods.

Finally, the design and optimization study of the layer structure for this SOA/isolator has revealed that the limiting factor is not so much the magneto-optic strength of the MO metal itself, but the fundamental gain saturation of the QW's. This cannot be avoided, and the only way to somewhat relax the design due to the gain limitations of the wells is by increasing the number of wells. For the moment the limit on the number of wells is not clear yet. The measurements detailed in the next chapter indicate that up to 9 wells, there is no decrease in quantum efficiency nor in material quality.

References

[1] A. K. Zvezdin, and V. A. Kotov, Chapter 2 in *Modern Magnetooptics and Magnetooptical Materials.* (Institute of Physics Publishing, Philadelphia, USA, 1997).

[2] R. Atkinson, and P. H. Lissberger, "Sign conventions in magneto-optical calculations and measurements." *Appl. Opt.,* vol. 31(28), pp. 6076–6081, 1992.

[3] J. T. Chang, J. F. Dillon, Jr., and U. F. Gianola, "Magneto-Optical Variable Memory Based Upon the Properties of a Transparent Ferrimagnetic Garnet at Its Compensation Temperature." *J. Appl. Phys.,* vol. 36(3), pp. 1110–1111, 1965.

[4] J.-P. Krumme, J. Verweel, J. Haberkamp, W. Tolksdorf, G. Bartels, and G. P. Espinosa, "Thermomagnetic Recording in Thin Garnet Layers." *Appl. Phys. Lett.,* vol. 23(11), pp. 451–453, 1972.

[5] G. S. Krinchik, and V. A. Artemjev, "Magneto-optic Properties of Nickel, Iron, and Cobalt." *J. Appl. Phys.,* vol. 39(2), pp. 1276–1278, 1968.

[6] G. S. Krinchik, and V. A. Artemjev, "Magneto-optical properties of Ni, Co, and Fe in the ultraviolet visible, and infrared parts of the spectrum." *Sov. Phys.-JETP,* vol. 26(6), pp. 1080–1085, 1968.

[7] M. Mansuripur, "Analysis of multilayer thin-film structures containing magneto-optic and anisotropic media at oblique incidence using 2×2 matrices." *J. Appl. Phys.,* vol. 67(10), pp. 6466–6475, 1990.

[8] Š. Višňovský, R. Lopuňík, M. Bauer, J. Bok, J. Fassbender, and B. Hillebrands, "Magnetooptic ellipsometry in multilayers at arbitrary magnetization." *Optics Express,* vol. 9(3), pp. 121–135, 2001.

[9] R. M. A. Azzam, and N.M. Bashara, *Ellipsometry and Polarized Light.* (North-Holland Physics Publishing, Elsevier Science Publishers, Amsterdam, The Netherlands, 1987).

[10] H. A. Priestley, *Introduction to Complex Analysis.* (Oxford University Press, Oxford, England, 1985).

[11] D. E. Aspnes, "Fourier transform detection system for rotating-analyzer ellipsometers." *Opt. Commun.,* vol. 8(3), pp. 222–225, 1973.

[12] C.-Y. You, and S.-C. Shin, "Generalized analytic formulae for magneto-optical Kerr effects." *J. Appl. Phys.,* vol. 84(1), 541–546, 1998.

[13] A. Yariv, Chapter 14 in *Quantum Electronics,* 3^{rd} *ed.* (John Wiley & Sons, New York, USA, 1989).

[14] I. P. Kaminow, and E. Turner, "Linear Electrooptic Materials," Chapter 15 in *Handbook of Lasers,* (Chemical Rubber Co., Cleveland, USA, 1971).

[15] S. N. Jasperson, and S. E.Schnatterly, "An Improved Method for High Reflectivity Ellipsometry Based on a New Polarization Modulation Technique." *Rev. Sci. Instr.,* vol. 40(6), pp. 761–767, 1969.

[16] J. Badoz, M. P. Silverman, and J.-C. Canit, "Wave propagation through a medium with static and dynamic birefringence: theory of the photoelastic modulator." *J. Opt. Soc. Am. A,* vol.7(4), pp. 672–682, 1990.

[17] M, Abramowitz, and I. A. Stegun (Eds.), "Bessel Functions J and Y." in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, ,* 9^{th} *printing* (New York: Dover, pp. 358-364, 1972.)

[18] S. N. Jasperson, D. K. Bürge, and R. C. O'Handley, "A modulated ellipsometer for studying thin film optical properties and surface dynamics." *Surface Sci.,* vol. 37, pp. 548–558, 1973.

[19] K. Sato, "Measurement of Magneto-Optical Kerr Effect Using Piezo-Birefringent Modulator." *Jpn. J. Appl. Phys.,* vol.20(12), pp. 2403–2409, 1981.

[20] M. J. Dignam, and M. Moskovits, "Azimuthal misalignment and surface anisotropy as sources of error in ellipsometry." *Appl. Opt.,* vol.9(8), pp. 1868–1873, 1970.

[21] A. Berger, and M. R. Pufall, "Generalized magneto-optical ellipsometry." *Appl. Phys. Lett.,* vol. 71(7), pp. 965–967, 1997.

[22] A. Berger, and M. R. Pufall, "Quantitative vector magnetometry using generalized magneto-optical ellipsometry." *J. Appl. Phys.,* vol. 85(8), pp. 4583–4585, 1999.

[23] A. K. Zvezdin, and V. A. Kotov, Chapter 5 in *Modern Magnetooptics and Magnetooptical Materials.* (Institute of Physics Publishing, Philadelphia, USA, 1997).

[24] C. S. Wang, and J. Callaway, "Band structure of nickel: Spin-orbit coupling, the Fermi surface, and the optical conductivity." *Phys. Rev. B,* vol. 9(11), pp. 4897-4907, 1974.

[25] P. M. Oppeneer, T. Maurer, J. Sticht, and J. Kübler, "*Ab initio* calculated magneto-optical Kerr effect of ferromagnetic metals: Fe and Ni." *Phys. Rev. B,* vol. 45(19), pp. 1092410933, 1992.

[26] J. L. Erskine, and E. A. Stern, "Magneto-Optic Kerr Effects in Gadolinium." *Phys. Rev. B,* vol. 8(3), pp. 1239–1255, 1973.

[27] N. M. Voloshinskaya, and G. A. Bolotin "Question of nature of magneto-optical effects in nickel." *Fiz. Met. Metalloved.,* vol. 38(5), pp. 975–984, 1974.

[28] H. S. Bennett, and E. A. Stern, "Faraday Effect in Solids." *Phys. Rev.,* vol. 137(2A), pp. A448-A461, 1965.

[29] K. H. J. Buschow, P. G. van Engen, and R. Jongebreur, "Magneto-optical properties of metallic ferromagnetic materials." *J. Magn. Magn. Mater.,* vol. 38(1), pp. 1–22, 1983.

[30] K. H. J. Buschow, *Ferromagnetic Materials vol. 4,* ed. E. P. Wohlfarth, and K. H. J. Buschow, pp. 493–595, (Elsevier, Amsterdam, The Netherlands, 1988).

[31] R. C. O'Handley, Chapter 6 "Magnetic Anisotropy" in *Modern Magnetic Materials : Principles and Applications,* (Wiley-Interscience, New York, USA, 2000).

[32] E. R. Callen, and H. B. Callen, "Anisotropic magnetization." *J. Phys. Chem. Sol.,* vol. 16(3-4), pp. 310–328,1960.

[33] H. Pauwels, Chapter 2 in "Elektrische en Magnetische Eigenschappen van Materialen (Part 2)." (in Dutch) Master Course, Ghent University, Belgium, 1998.

[34] B. Edwards, and D. I. Paul, "Ferromagnetic coercivity and applied field orientation." *J. Magn. Magn. Mater.,* vol. 147(3), pp.331–340, 1995.

[35] Th.G.S.M. Rijks, R.F.O. Reneerkens, R. Coehoorn, J.C.S. Kools, M.F. Gillies, J.N. Chapman, and W.J.M. de Jonge, "Switching field interval of the sensitive magnetic layer in exchange-biased spin valves." *J. Appl. Phys.,* vol. 82(7), pp. 3442–3451, 1997.

[36] R. C. O'Handley, Chapter 14 "Magnetic Annealing and Directional Order" in *Modern Magnetic Materials : Principles and Applications,* (Wiley-Interscience, New York, USA, 2000).

[37] R. C. O'Handley, Chapter 2 "Magnetostatics" in *Modern Magnetic Materials : Principles and Applications,* (Wiley-Interscience, New York, USA, 2000).

[38] D. A. Humphreys, R. J. King, D. Jenkins, and A. J. Moseley, "Measurement of absorption coeficients of $Ga_{0.47}In_{0.53}As$ over the wavelength range 1.0-1.7 μm." *Electron. Lett.,* vol. 21(25-26), pp.1187–1189, 1985.

[39] S. M. Sze, Chapter 5 "Metal-Semiconductor Contacts" in *Physics of Semiconductor Devices, 2^{nd} edition,* (Wiley-Interscience, New York, USA, 1981).

[40] V. L. Rideout, "A Review of the Theory and Technology for Ohmic Contacts to Group III-V Compound Semiconductors." *Solid-St. Electron.,* vol. 18(6), pp. 541–550, 1975.

[41] C. Y. Chang, and S. M. Sze, "Carrier Transport across Metal-Semiconductor Barriers," *Solid-St. Electron.,* vol. 13(6), pp. 727–740, 1970.

[42] A. M. Cowley, and S. M. Sze, "Surface States and Barrier Height of Metal-Semiconductor Systems." *J. Appl. Phys.,* vol. 36(10), pp. 3212–3220, 1965.

[43] J. Bardeen, "Surface States and Rectification at a Metal Semi-Conductor Contact." *Phys. Rev.,* vol. 71(10), pp. 717–727, 1947.

[44] S. Kurtin, T. C. McGill, and C. A. Mead, "Fundamental Tranistion in the Electronic Nature of Solids." *Phys. Rev. Lett.,* vol. 22(26), pp. 1433–1436, 1969.

[45] W.M. Loh, S.E. Swirhun, T.A. Schreyer, R.M. Swanson, and K.C. Saraswat, "Modeling and Measurement of Contact Resistances." *IEEE Trans. on Electron Devices,* vol. ED-34(3), pp. 512–523, 1987.

[46] G. K. Reeves, and H. B. Harrison, "Obtaining the Specific Contact Resistance from Transmission Line Model Measurements." *IEEE Electron Device Lett.,* vol. EDL-3(5), pp. 111–113, 1982.

[47] S. J. Proctor, L. W. Linholm, and J. A. Mazer, "Direct Measurement of Interfacial Contact Resistance, End Contact Resistance, and Interfacial Contact Layer Uniformity." *IEEE Trans. Electron Devices,* vol. ED-30(11), pp. 1535–1542, 1983.

[48] H. H. Berger, "Contact Resistance and contact resistivity." *J. Electrochem. Soc.,* vol. 119, pp 507–513, 1972.

[49] G. Stareev, H. Knzel, and G. Dortmann, "A Controllable Mechanism of Forming Extremely Low-resistance Ohmic Contacts to Group III-V Compound Semiconductors." *J. Appl. Phys.,* vol. 74(12), pp. 7344–7356, 1993.

[50] A. Katz, W. C. Dautremont-Smith, P. M. Thomas, L. A. Koszi, J. W. Lee, V. G. Riggs, R. L. Brown, J. L. Zilko, and A. Lahav "Pt/Ti/p-InGaAsP Nonalloyed Ohmic Contact Formed by Rapid Thermal Processing." *J. Appl. Phys.* vol. 65(11), pp. 4319–4323, 1989.

[51] A. Katz, S. N. G. Chu, B. E. Weir, C. R. Abernathy, W. S. Hobson, S. J. Pearton, and W. Savin, "Rapid Isothermal Processing of Pt/Ti Contacts to p-Type III-V Binary and Related Ternary Materials." *IEEE Trans. on Electron Devices,* vol. 39(1), pp. 184–192, 1992.

[52] T. Y. Chang, Y. He, N. J. Sauer, E. H. Westerwick, F. G. Storz, M. D. Williams, "Shallow p-Type Ohmic Contact to $Ga_{0.47}In_{0.53}As$ Using Au/Ti/Mn/W." *Electron. Lett.,* vol. 29(12), pp. 1139–1141, 1993.

[53] M. H. Park, L. C. Wang, and C. J. Palmstrom "Low resistance Pd/Zn/Pd Ohmic Contact to p-$In_{0.82}Ga_{0.18}As_{0.39}P_{0.61}$." *J. Appl. Phys.,* vol. 81(6), pp. 2720–2724, 1997.

[54] A. Katz, A. El-Roy, A. Feingold, M. Geva, N. Moriya, S. J. Pearton, E. Lane, T. Keel, and C. R. Abernathy, "W(Zn) Selectively Deposited and Locally Diffused Ohmic Contacts to p-InGaAs/InP Formed by Rapid Thermal Low Pressure Metalorganic Chemical Vapor Deposition." *Appl. Phys. Lett.,* vol. 62(21), pp. 2652–2654, 1993.

[55] G. Bayreuther, M. Dumm, B. Uhl, R. Meier, and W. Kipferl, "Magnetocrystalline volume and interface anisotropies in epitaxial films: Universal relation and Néel's model.(invited)" *J. Appl. Phys.,* vol. 93(11), pp. 8230–8235, 2003.

[56] A. R. Adams, "Band structure engineering for low-threshold high effciency semiconductor lasers." *Electron. Lett.,* vol. 22(), pp. 249–250, 1986.

[57] E. Yablonovitch, and E. O. Kane, "Reduction of lasing threshold current density by the lowering of valence band effective mass." *J. Lightwave Technol.,* vol. 4(5), pp. 504–506, 1986.

[58] P. J. A. Thijs, Chapter 1 in "Strained-layer InGaAs(P)/InP Quantum Well semiconductor Lasers Grown by Organometallic Vapour Phase Epitaxy." Ph. D. Thesis, Technische Universiteit Delft, Delft, The Netherlands, 1994.

[59] P. J. A. Thijs, and T. Van Dongen, "High quantum efficiency, high power, modulation doped GaInAs strained-layer quantum well laser diodes emitting at 1.5 μm." *Electron. Lett.,* vol. 25(25), pp.1735–1737, 1989.

[60] I. Hayashi, M. B. Panish, P. W. Foy, and S. Sumski, "Junction lasers which operate continuously at room temperature." *Appl. Phys. Lett.,* vol. 17(3), pp. 109–111, 1970.

[61] W. T. Tsang, "Extremely low threshold (AlGa)As modified multiquantum well heterostructure lasers grown by molecular-beam epitaxy." *Appl. Phys. Lett.,* vol. 39(10), pp. 786–788, 1981.

[62] W. D. Laidig, P. J. Caldwell, Y. F. Lin, and C. K. Peng, "Strained-layer quantum-well injection laser." *Appl. Phys. Lett.,* vol. 44(7), pp. 653–655, 1984.

[63] S. W. Corzine, R.-H. Yan, and L. A. Coldren, Chapter 2 "Optical Gain in III-V Bulk and Quantum Well Semiconductors." in *Quantum Well Lasers,* (ed. P. S. Zory, Jr.) (Academic Press, San Diego, USA, 1993).

[64] J. M. Luttinger, and W. Kohn, "Motion of Electrons and Holes in Perturbed Periodic Fields." *Phys. Rev.,* vol. 97(4), pp. 869–883, 1955.

[65] J. M. Luttinger, "Quantum Theory of Cyclotron Resonance in Semiconductors: General Theory." *Phys. Rev.,* vol. 102(4), pp. 1030–1041, 1956.

[66] G. Dresselhaus, A. F. Kip, and C. Kittel, "Cyclotron Resonance of Electrons and Holes in Silicon and Germanium Crystals." *Phys. Rev.,* vol. 98(2), pp. 368-384, 1955.

[67] J. Callaway, Chapter 1 "Electronic Structure." in *Quantum Theory of the Solid State,* (2^{nd} ed.) (Academic Press, San Diego, USA, 1991)

[68] D. A. Broido, and L. J. Sham, "Effective masses of holes at GaAs-AlGaAs heterojunctions." *Phys. Rev. B,* vol. 31(2), pp. 888–892, 1985.

[69] S. W. Corzine, R. H. Yan, and L. A. Coldren, "Theoretical gain in strained InGaAs/AlGaAs quantum wells including valence-band mixing effects." *Appl. Phys. Lett.,* vol. 57(26), pp. 2835–2837, 1990.

[70] F. Verheest, *Continuum Mechanics,* Master Course, Ghent University, Belgium, 1999.

[71] G. Jones, and E. P. O'Reilly, "Improved performance of long-wavelength strained bulk-like semiconductor lasers." *IEEE J. Quantum Electron.,* vol. 29(5), pp. 1344–1354, 1993.

[72] P. J. A. Thijs, L. F. Tiemeijer, J. J. M. Binsma, and T. Van Dongen, "Progress in long-wavelength strained-layer InGaAs(P) quantum-well semiconductor lasers and amplifiers." *IEEE J. Quantum Electron.,* vol. 30(2), pp. 477–499, 1994.

[73] W. Greiner, Chapter 2 "Interaction of Electromagnetic Fields with Matter." in *Quantum Mechanics – Special Chapters,* (Springer Verlag, Berlin, Germany, 1998).

[74] J. W. Matthews, and A. E. Blakeslee, "Defects in epitaxial multilayers: I. Misfit dislocations." *J. Crystal Growth,* vol. 27, pp. 118–125, 1974.

[75] J. W. Matthews, and A. E. Blakeslee, "Defects in epitaxial multilayers : II. Dislocation pile-ups, threading dislocations, slip lines and cracks." *J. Crystal Growth,* vol. 29(3), pp. 273–280, 1975.

[76] J. W. Matthews, and A. E. Blakeslee, "Defects in epitaxial multilayers : III. Preparation of almost perfect multilayers." *J. Crystal Growth,* vol. 32(2), pp. 265–273, 1976.

[77] T. G. Andersson, Z. G. Chen, V. D. Kulakovskii, A. Uddin, and J. T. Vallin, "Variation of the critical layer thickness with In content in strained $In_xGa_{1x}As$-GaAs quantum wells grown by molecular beam epitaxy." *Appl. Phys. Lett.,* vol. 51(10), pp. 752–754, 1987.

[78] R. E. Cavicchi, D. V. Lang, D. Gershoni, A. M. Sergent, J. M. Vandenberg, S. N. G. Chu, and M. B. Panish, "Admittance spectroscopy measurement of band offsets in strained layers of $In_xGa_{1x}As$ grown on InP." *Appl. Phys. Lett.,* vol. 54(8), pp. 739–741, 1989.

[79] B. I. Miller, U. Koren, M. G. Young, and M. D. Chien, "Strain-compensated strained-layer superlattices for 1.5 μm wavelength lasers." *Appl. Phys. Lett.,* vol. 58(18), pp. 1952–1954, 1991.

[80] Y. Seko, and A. Sakamoto, "Valence subband structures and optical properties of strain-compensated quantum wells." *Jpn. J. Appl. Phys.,* vol. 40 - Part1(1), pp.34–39, 2001.

[81] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for IIIV compound semiconductors and their alloys." *J. Appl. Phys.,* vol. 89(11), pp. 5815–5875, 2001.

[82] S.-H. Wei, and A. Zunger, "Predicted band-gap pressure coefficients of all diamond and zinc-blende semiconductors: Chemical trends." *Phys. Rev. B,* vol. 60(8), pp. 5404–5411, 1999.

[83] M. Kondow, K. Uomi, A. Niwa, T. Kitatani, S. Watahiki, Y. Yazawa, "GaInNAs: A novel material for long-wavelength-range laser diodes with excellent high-temperature performance." *Jpn. J. Appl. Phys.,* vol. 35 - Part1(2B), pp. 1273–1275, 1996.

[84] T. Ishikawa, and J. E. Bowers, "Band lineup and in-plane effective mass of InGaAsP or InGaAlAs on InP strained-layer quantum well." *IEEE J. Quantum Electron.,* vol. 30(2), pp. 562–570,1994.

[85] J. Piprek, *Semiconductor Optoelectronic Devices – Introduction to Physics and Simulation*, pp. 43–48, (Academic Press, San Diego, USA, 1993).

[86] S. M. Sze, Chapter 1 "Physics and Properties of Semiconductors – a R'esumé," pp. 35–38 in *Physics of Semiconductor Devices, 2^{nd} edition,* (Wiley-Interscience, New York, USA, 1981).

[87] A. D. Andreev, and G. G. Zegrya, "Theoretical study of thresholdless Auger recombination in compressively strained InAlAsSb/GaSb quantum wells." *Appl. Phys. Lett.,* vol. 70(5), pp. 601–603, 1997.

[88] T. A. Ma, Z. M. Li, T. Makino, and M. S. Wartak, "Aprroximate Optical Gain Formulas for 1.55-μm Strained Quantum-Well Lasers." *IEEE J. Quantum Electron.,* vol. 31(1), pp. 29–34, 1995.

[89] T. A. DeTemple, and C. M. Herzinger, "On the semiconductor laser logarithmic gain-current density relation." *IEEE J. Quantum Electron.,* vol. 29(5), pp. 1246–1252, 1993.

[90] J. E. A. Whiteaway, G. H. B. Thompson, P. D. Greene, and R. W. Glew, "Logarithmic gain/current-density characteristic of InGaAs/InGaAlAs/InP multi-quantum-well separate-confinement-heterostructure lasers." *Electron. Lett.,* vol. 27(4), pp 340–342.

[91] P. J. A. Thijs, T. van Dongen, L. F. Tiemeijer, and J. J.M. Binsma, "High-performance $\lambda = 1.3\mu$m InGaAsP-InP strained-layer quantum well lasers." *J. Lightwave Technol.,* vol. 12(1), pp. 28–37, 1994.

[92] M. C. Wang, K. Kash, C. E. Zah, R. Bhat, and S. L. Chuang "Measurement of nonradiative Auger and radiative recombination rates in strained-layer quantum-well systems." *Appl. Phys. Lett.,* vol. 62(2), pp. 166–168, 1993.

[93] B. W. Hakki and T. L. Paoli, "Gain spectra in GaAs double-heterostructure injection lasers." *J. Appl. Phys.,* vol. 46(3), pp. 1299–1306, 1975.

[94] J. Taylora and V. Tolstikhin, "Intervalence band absorption in InP and related materials for optoelectronic device modeling." *J. Appl. Phys.,* vol. 87(3), pp. 1054–1059, 2000.

[95] D. Hofstetter and J. Faist, "Measurement of semiconductor laser gain and dispersion curves utilizing Fourier transforms of the emission spectra." *IEEE Photon. Technol. Lett.,* vol. 11(11), pp. 1372–1374, 1999.

[96] B. Jensen, in *Handbook of Optical Constants of Solids II*, E. D. Palik, ed., Chap. 6. Calculation of the refractive index of compound semiconductors below the band gap, pp. 125–149. (Academic Press, 1991).

[97] M. Compin, B. Bartenlian, P. Beauvillain, P. Gogol, J. Hamrle, L. Lagae, J. Pistora, K. Postava, S. Visnovsky, and R. Wirix-Speetjens, "Détermination des indices optiques et magnéto-optiques de films minces de CoFe constituant un isolateur intégré à 1.3 μm." In *Colloque Luis Neél CMNM 2004*, (France, 2004).

[98] P. B. Johnson and R. W. Christy, "Optical constants of transition metals: Ti, V, Cr, Mn, Fe, Co, Ni, and Pd." *Phys. Rev. B,* vol. 9(12), pp. 5056–5070, 1973.

4

Experimental characterization of the building blocks of the SOA/isolator

IN this chapter the characterization results of the subcomponents of the SOA/isolator are presented. The requirements for the magnetic, magneto-optic and electric properties of the MO/semiconductor interface and for the gain performance of the tensile-strained MQW active region, discussed in Chapter 3, are one by one experimentally assessed. The results of these experiments will serve as a direct feasibility proof for the novel integrated isolator concept.
This chapter is organized in a complementary fashion to Chapter 3. This means that in the first part of this chapter we successively present the results on the optical and magneto-optical, magnetic anisotropy and electrical Ohmic properties of the MO metal/semiconductor interface. The second part will then be devoted to the in-depth gain performance characterization of the tensile-strained MQW active regions. The theoretical details of the measurement techniques are extensively treated in the corresponding sections of the previous chapter, and will obviously not be repeated here. The reader will at appropriate instances be referred to the relevant sections of Chapter 3.

4.1 Characterization of ferromagnetic CoFe metal alloys

It was indicated in section 3.1.1.3 how the pragmatical choice was made to use the CoFe alloy system as the magneto-optical metal for the studied integrated isolator concept. There are a few MO metal alternatives that have been considered. But their magneto-optical characterization at the relevant wavelength (1300nm) is not yet fully performed at the time of writing. The alternatives under study are the Mn-based alloys MnAs and MnSb. Therefore little can be stated about their MO strength in the near infrared, however, what is clear is that they appear to have a very beneficial in-plane magnetic anisotropy, with very strong coercivity and good squareness.
Next to these Mn-based alternatives, recent work by other researchers [1] seem to indicate that pure Fe might after all serve as a MO material alternative, in contrast to what the results of Busschow predict [2]. However it is not clear whether these authors are reporting mono- or polycrystalline Fe. After all, as Busschow indicated, the MO strength relies on the crystal quality of the material. Therefore, even though monocrystalline Fe might be stronger[1], it is very doubtful that this can be grown on InP-based semiconductors (or GaAs-based for that matter). Finally, it is worth mentioning that recently another alternative "source" for the MO effect has been theoretically suggested in literature [3, 4]. It concerns so-called ferromagnetic-semiconductor composite materials. These composites consist of nanometer-size ferromagnetic particles distributed throughout a semiconductor host. The MaxwellGarnet theory seems to suggest that the optical loss of such a composite is reduced compared to a bulk ferromagnetic layer, while the Faraday rotation and, hence, magneto-optic Kerr effects remain the same [5]. Growth of such materials for magneto-electronic purposes has already been reported earlier [6]. Experimental verification of the MO effect assertion however is still awaited. On top of that, the remanent properties of these materials are really low and the saturation fields very high (close to 0.5 T), indicating that a very strong integrated thin film magnet would be required to keep these layers saturated [7].

Irrespective of the feasibility of these recently suggested alternatives[2], we will focus in this section on the characterized properties of the two CoFe alloys that are technologically available to us, $Co_{90}Fe_{10}$ and $Co_{50}Fe_{50}$. The CoFe films of all the samples reported throughout this section have been deposited by our colleagues of the ME research group at the Inter-Universitair Micro-Elektronica Centrum (IMEC), using an optimized custom-made magnetron sputtering system [8]. This sputter system achieves a thickness control of better than 3% on a wafer-scale. The CoFe films deposited in this way are polycrystalline with a grain size of some 10 nm. During deposition the sample can be subjected to an uniform magnetic field, helping to induce uniaxial, in-plane magnetic anisotropy. Most of the realized samples consisted of CoFe films deposited on InP samples with a layer structure

[1] Results to prove this are to our knowledge not published yet.
[2] which in the first place seem to indicate to us that our work has stirred the attention of the international research community.

closely resembling the actual contact structure of the final SOA device, i.e. a highly Be-doped InGaAs(P) layer with a thickness of about 100nm. This implies a clear (1 0 0)-texture for the deposition surface[3]! Never has a precursing buffer layer such as Ta or Ta/Cu been used for the deposition. Good adhesion was observed, and processing of sputtered-deposited CoFe films into stripe patterns using standard lift-off techniques proved to be no problem.

In this section we will give an overview of the experimental characterization of the (magneto-)optical, magnetic and electric performance of these thin film polycrystalline CoFe alloys.

4.1.1 Optical and magneto-optical parameters

The optical and magneto-optical characterization of the CoFe alloys was done using both the dynamic and static generalized magneto-optic photometry techniques described in 3.1.1. The dynamic photometry experiments have been performed by our colleagues of the MMS department at the Institut d'Electronique Fondamentale at the Université Paris-Sud on a custom-made optimized set-up, while the alternative static photometry measurements were performed by us at IMEC. As explained in section 3.1.1.2 this alternative photometric scheme has limited possibilities, due to the fact that it is only suited for MO stacks with a single reflective interface, but it is very easy to build, and allows a rather fast characterization. The main purpose of these experiments was to check whether such a low-cost and fast photometry bench could be created out of an (non-dedicated) existing MOKE measurement set-up at IMEC by simply adding the needed high-resolution rotary stage for the optical analyser. In the next two sections the experimental results of both techniques will be presented. While the accuracy of the custom-made dynamic setup is hard to compete with, it will be seen that surprisingly enough the rudimentary static set-up achieves good results for the optical characterization of the materials. However, it suffers severely from its lack of calibration in order to obtain a good magneto-optic characterization.

4.1.1.1 Dynamic photometric characterization of CoFe alloys

Fig. 4.1 shows two pictures of the dynamic photometry setup at IEF. The different optical elements can be easily distinguished. Also note how the electromagnet allows the applied magnetic field to be arbitrarily oriented within the plane of the sample. The sample itself is mounted on a rotary stage allowing this setup to be used in a simplified form as a tool for MOKE-based in-plane magnetic anisotropy measurements. The polarizer is kept at a fixed azimuth setting, and is followed by a motorized $\lambda/2$-plate to adjust the incident linear polarization to the desired setting. This stage also allows for the azimuth calibration of all optical elements (as explained in 3.1.1.1). The light source of the setup is modular. Up till now all measurements have been performed at a wavelength of 1310 nm.

[3]Remember that the Co-rich alloys have a hcp crystal structure while the other compositions are bcc

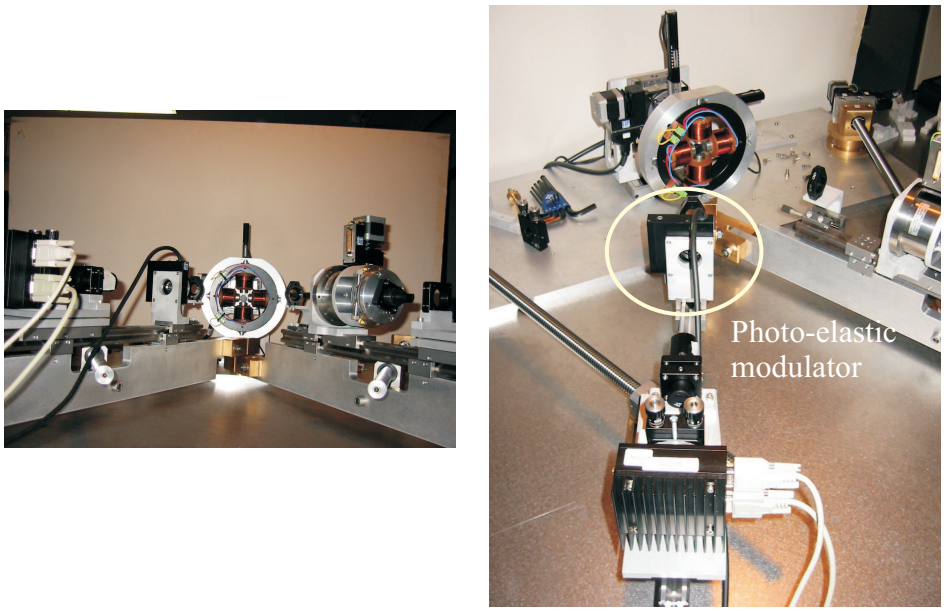


Fig. 4.1: Dynamic photometric set-up. Clearly visible are the motorized arms allowing scanning of the incidence angle with a resolution better than 1 arcmin. Note the presence of the PEM in the outgoing light path. It can also be seen how the electromagnet allows the application of an arbitrarily directed magnetic field within the sample plane. The sample holder is mounted on a rotary stage allowing azimuthal rotation for detailed investigation of the in-plane magnetic anisotropy.

For all measurements the procedure described in 3.1.1.1 has been used, with a direct calibration of the MOKE rotation and ellipticity by simultaneous nulling rotation of the PEM and the analyser. The magnetic field was applied in the longitudinal direction. However for the isotropic ellipsometric ratio a less elaborate procedure was used than the one described.
Instead of measuring the normalized isotropic Jones matrix element $\frac{r_{pp}}{r_{ss}}$ using a polarization with a small but known p-content, it was opted to directly measure the s- and p- power reflectances, R_s and R_p, neglecting any MO contribution. The reasoning behind this approach is quite logical. If the MO sample is not subjected to a magnetic field, it will be in a remanent state which will, depending on its previous "treatment" and its anisotropy, consist of magnetic domains adding up to either a net zero magnetization or a net magnetization pointing mainly in a direction of an easy axis. Not caring about the precise magnetic anisotropy of the sample, this means that there will be either no MO effect or a MO effect that can be either longitudinal, polar or transverse. If pure s- or p-polarized light is incident on this sample, and the reflectance is measured (thus without an optical analyzer), the detected intensity will be proportional to $|r_{ss}|^2 + |r_{ps,\mathrm{MO}}^{\mathrm{pol|long}}|^2$ or to

$|r_{pp} + \delta r^{\text{tranv}}_{pp,\text{MO}}|^2 + |r^{\text{pol|long}}_{sp,\text{MO}}|^2$. It has been shown in Chapter 1 how the MO Kerr reflection coefficients are all of first order in the MO Voigt parameter Q. Therefore only for p-incident light ánd with transverse magnetized domains there will be a first-order MO contribution to the reflected light. For all the other cases the MO contribution will be of negligible second order. Therefore, it can be assumed that the reflectances depend purely on the isotropic optical parameters, provided there is only a negligible amount of transversely magnetized domains. This is confirmed by experimental reflectance measurements for both s- and p-polarization as a function of angle of incidence. Both curves can be fitted simultaneously using only isotropic parameters, which serves to prove that the presence of the transverse MO effect is negligible.

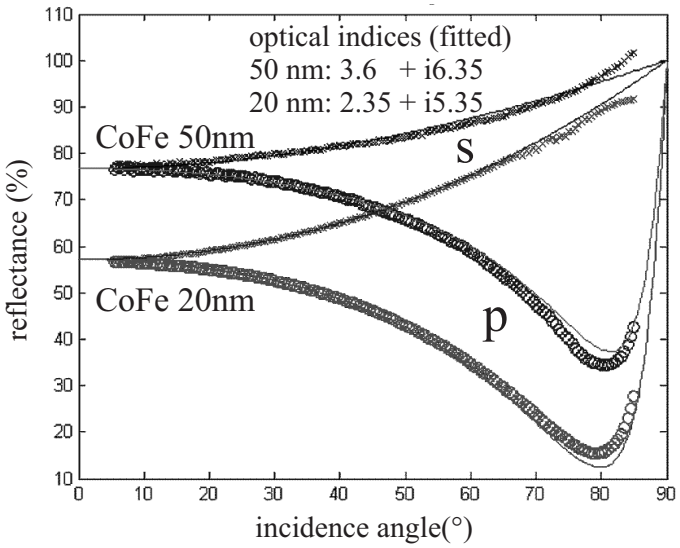


Fig. 4.2: Optical parameter extraction at 1310nm through s- and p-reflectance measurements on two unprotected CoFe films of different thickness. The differences in the extracted values for the optical indices serve as an illustration of the need for protected and passivated CoFe layers.

Fig. 4.2 shows an example of one of the earlier performed reflectance measurements. The s- and p-reflectances measured on two unprotected thin $Co_{90}Fe_{10}$ film (20 and 50 nm) sputtered on a InGaAs substrate as a function of incidence angle are plotted together with their corresponding least-squares fit. The optical parameters are extracted using a numerical MO multilayer reflection tool [9]. The divergences in the fitting procedure at incidence angles above 70˚are not caused by some spurious neglected MO effect, but are to be attributed to the limited entrance aperture of the electromagnet as can be seen in the picture of Fig. 4.1. These preliminary experiments on unprotected CoFe samples serve mainly to illustrate the importance of having a precise knowledge of the kinds of layers (and their thicknesses) present in the multilayer stack. Indeed, even though the fitting

is quite acceptable for both samples the extracted values for the optical indices are significantly different. This is mainly thought to be a result of deviations in the precise thickness of the CoFe layer due to surface oxidation. Accurate determination of the optical constants of the CoFe films therefore inevitably requires the presence of a protection passivation layer on top of the ferromagnetic metal. Any modern ellipsometric setup should achieve an accuracy of at least 5% in the determination of the optical constants. The importance of this precision in the extracted isotropic optical parameters is even enhanced when it concerns magneto-optical ellipsometry. This is illustrated by the general formulae introduced in [10] for both the polar and longitudinal MO Kerr reflection effects at the interface between an isotropic and a MO medium[4]. It is shown there that these general formulae can all be written as

$$\Psi^i \left(= \frac{r_{ps(sp)}}{r_{ss(pp)}} \right) = f^i(\theta, \epsilon_{\mathrm{MO}}, \epsilon_{\mathrm{inc}}) \frac{g_{\mathrm{MO}}}{\epsilon_{\mathrm{MO}} - \epsilon_{\mathrm{inc}}}, \text{ with } i = \text{pol, long} \tag{4.1}$$

and with ϵ_{inc} the isotropic permittivity of the (isotropic) incidence medium, ϵ_{MO} and g_{MO} the isotropic permittivity and the gyrotropic parameter of the MO medium, and θ the angle of incidence. As the real and imaginary parts of the left-hand side of (4.1) are nothing else but the MO Kerr rotation and Kerr ellipticity, it is clear that an error in determination of the isotropic permittivity of the medium will inflate the error on the gyrotropic parameter of the MO medium. In a multilayer stack things get even worse as the relevant ellipsometric ratios $r_{ps(sp)}/r_{ss(pp)}$ are then incorporated into a concatenated product of Jones matrices. As a result the total relative error gets additional contributions from the uncertainties on the different layer thicknesses.

Protecting the ferromagnetic films from natural oxidation can be done by covering them with a passivation oxide layer with a controlled thickness. This can be achieved in many ways. The sputter system at IMEC allows for a reactive sputtering of metal oxides by sputtering a metal target in a mixed Ar/O_2 ambient. In this way deposition of Al_2O_3 and NiO can be achieved. However for Al_2O_3 an even better control of the oxide thickness can be achieved by using in-situ natural oxidation of a deposited Al layer. It is known that the natural oxidation process in Al is self-limiting for a depth around 2 nm. In situ natural oxidation is performed in the load lock of the sputter system, which can be completely isolated by extra gate valves so that the required O_2 ambient can be created and controlled. In this way protected blanket samples were prepared for both the $Co_{90}Fe_{10}$ and $Co_{50}Fe_{50}$ alloys, by sputter depositing a 10 nm Al layer, and subsequently subjecting them to in-situ natural oxidation. While the described ellipsometric setup is in principle capable of multi-variable parameter extraction on reflective blankets with several unknown layers, superior accuracy can be expected when extra samples are available to determine the optical (and magneto-optical) constants of the constituent materials separately. Therefore the following set of blanket samples has been fabricated.

[4] we do not consider the transverse Kerr effect as this will not cause a polarization rotation

	passivation characterization		CoFe characterization
Al_2O_3	2.5	2.5	2.5
Al	8	50	8
CoFe	0	0	50
GaAs substrate			

TABLE. 4.1: Thin film thicknesses (in nm) of the multilayer blanket films prepared for optical and magneto-optical parameter extraction. The aluminium oxide is obtained through in-situ self-limiting natural oxidation of the deposited Al layer. IMEC has developed over the years a good expertise of this process. Therefore, the tabulated layer thicknesses of Al and Al_2O_3 can be considered to be very accurate.

Reflectance measurements at 1300nm on the two passivation characterization samples yielded the following values for the complex optical constants of Al and Al_2O_3.

	n	k
Al_2O_3	1.75	0
Al	1.7	13.6

For GaAs an optical index of 3.45 was used. All extracted optical constants agree very well with values reported in literature. There is however some inaccuracy in the real part of the Al index, which might be caused by the formation of a thin layer AlGaAs at the Al/GaAs interface. However, in view of the very high extinction coefficient of Al, this small deviation in the real part of Al is not expected to have a big influence on the optical parameter extraction of CoFe. This is illustrated in Fig. 4.3, which plots the measured s- and p-reflectivities and corresponding fit. The inset shows the uncertainties on the extracted values by plotting the least squares variance of the fit within the (n, k)-plane. It is seen that the minimum is well-defined and that the correlation between the refractive index and the extinction coefficient is as required very low.

After extraction of the optical isotropic indices of $Co_{90}Fe_{10}$, the MO Kerr rotation and ellipticity can be measured by inserting the PEM and the analyzer in the reflected light path. As explained in detail in section 3.1.1.1 this is achieved by setting the PEM[5] at 0˚azimuth and the analyzer at 45˚. The basic set of ellipsometry equations (3.17) yield this rotation and ellipticity as a function of the measured normalized intensity ratios (with respect to the detected DC intensity) at the 2^{nd} and the 1^{st} harmonic component of the detected intensity. By measuring the Kerr rotation (or the ellipticity) at two different incident angles using the method described in the paragraph on **calibration issues** of section 3.1.1.1, the Bessel functions in the canonical ellipsometry equations (3.17) can be calibrated. A calibrated measurement of the Kerr rotation and the Kerr ellipticity can then be performed throughout the entire range of incidence angles. With these measurements the (normalized) Jones matrix of the entire reflective multilayer stack is

[5]The PEM is operated at a frequency of 50kHz.

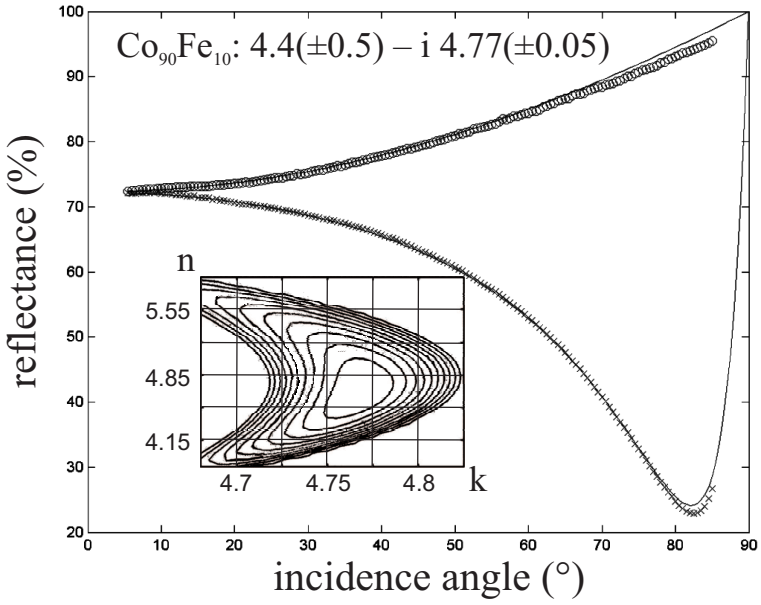


Fig. 4.3: Measured s- and p-reflectances (s:o; p:×) on the $Al_2O_3/Al/Co_{90}Fe_{10}$ sample of Table 4.1. The inset shows the least-squares variance within the (n, k)-plane of fitting parameters. It is clear that the deviation of the real part of Al has limited influence on the accuracy of the extracted optical parameters of CoFe.

known throughout the range of available incidence angles. Combined with a multilayer MO reflection algorithm and using the earlier determined isotropic optical constants of the constituent materials in the multilayer, the off-diagonal element of the dielectric permittivity tensor of the CoFe film is found through nonlinear least-squares-fitting of the measured calibrated rotation and ellipticity. Fig. 4.4 plots the measured Kerr rotations and ellipticities of the $Al_2O_3/Al/Co_{90}Fe_{10}/GaAs$ sample as a function of the incidence angle, shows the corresponding fit and gives the extracted complex value for the off-diagonal element of the MO layer.

Note that owing to the fact that only two parameters have to be fitted it is in principle sufficient to use any two of the four measured dependencies, e.g. the s- and p-Kerr rotation. The two remaining dependencies can then be used as a consistency check of the fitting procedure. The good fit for both the rotations and the ellipticities in Fig. 4.4 illustrates the consistency of the performed MO parameter extraction. The apparent lower accuracy with which g is determined is not a result of poor experimental conditions, but can be mainly attributed to error propagation from the uncertainties on the isotropic optical constants, as explained above.

The above procedure can be repeated for the $Co_{50}Fe_{50}$ alloy with similar accuracy. The following table summarizes the optical and magneto-optical extraction results. In the last column the ratio $|\Im(g)/\kappa|$ is given as some FoM for the here

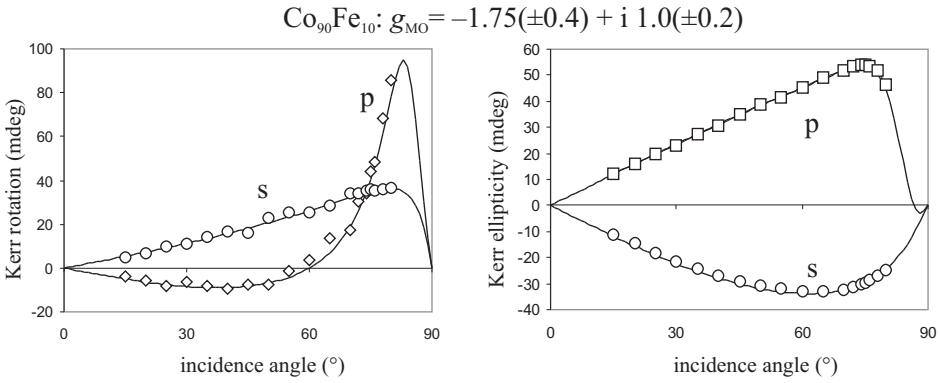


Fig. 4.4: Measured s- and p-Kerr rotation and ellipticity (s: ∘, p: □) of the $Al_2O_3/Al/Co_{90}Fe_{10}/GaAs$ multilayer and corresponding nonlinear least-squares fit. The sign convention of Atkinson [11], introduced in section 1.2.2.1, has been used for the signs of the real and imaginary parts of g_{MO}.

alloy	n	κ	$\Re(g)$	$\Im(g)$	FoM = $\left\lvert\frac{\Im(g)}{\kappa}\right\rvert$
$Co_{90}Fe_{10}$	4.4	4.77	-1.75	1.0	0.21
$Co_{50}Fe_{50}$	3.15	4.55	-1.72	-1.65	0.36

TABLE. 4.2: The extracted optical and magneto-optical constants (at 1300nm) for both available CoFe compositions. The FoM seems to suggest that the $Co_{50}Fe_{50}$ is almost twice better than the one with a higher Co content. It is safe to state this as the extinction coefficient hardly changes (and even decreases slightly), while $\Im(g)$ increases with almost 70%.

considered application of the magneto-optic metal. Recall that the non-reciprocal absorption loss is "proportional" to the imaginary part of g. As with any FoM one has to be careful in drawing too enthusiastic conclusions. Only a precise simulation of the non-reciprocal waveguide behaviour combined with the gain-current performance of the MQW region can give a clear idea about the performance of the device. Indeed, it is clear that the same proportional increase for both $\Im(g)$ and κ does not necessarily imply a same performance of the device, as a increase in κ will obviously imply an (impractical) increase in current. Therefore an exponent in the denominator of the FoM might be necessary. In this way the FoM of Table 4.2 can be seen as an exaggerated upper limit[6].

[6]Next to that we also have to be aware of the influence of n of the metal, as it will clearly have its impact on the confinement of the guided light near the MO metal. However, as long as the real part of the permittivity n^2 remains negative, it can be shown that a TM mode will always have a strong plasmonic enhancement near the interface [12].

4.1.1.2 Static photometric characterization of CoFe alloys

As stated earlier in this section, the main purpose of the static photometric setup was to check whether it was possible to build a low-cost alternative to the "Rolls-Royce" setup of the previous section. In section 3.1.1.2 the static photometric setup of Berger [13] was shown to be able to achieve this, at least in theory and for a range of simplifying assumptions. The most important of the assumptions of Berger is the requirement that the sample under study has only a single MO interface. By now we know that this will not be the case, as Fig. 4.2 has taught us that an unprotected CoFe layer will most probably be covered by some non-negligible thickness of natural oxide. However, it is still worth trying to check whether this setup might be used as a fast and low-cost alternative that can at least give a first indication of the strength of a MO material, even when it will certainly not achieve "world-record" ellipsometric accuracies. The method of Berger was straightforwardly implemented on an existing basic MOKE measurement setup at IMEC that was mainly used for azimuthal in-plane magnetic hysteresis and anisotropy measurements. It has a fixed incidence angle of about 30°, with a small degree of freedom necessary for alignment purposes. The sample is mounted between the pole shoes of an electromagnet applying a magnetic field in longitudinal configuration. As explained in section 3.1.1.2, Berger's method doesn't require a modulator as it gets all its information from analyzing a periodic intensity function with a rotating analyzer. The rotary stage introduced for this purpose has an angular resolution of about 0.2 mrad (which is about 1 cdeg) and is motorized in order to automate the measurement procedure. Actually, this motorized fine-resolution analyzer rotary stage is the only extension to the existing MOKE setup! So that at least it can be safely stated that the setup is cheap and easy to build.

Berger's method requires the measurement of the fractional intensity change upon magnetization reversal as a function of the analyzer azimuth for at least three different incident polarizations[7]. A simultaneous nonlinear least-squares fit to these 3 sets of intensity data using (3.29) yields values for the four ellipsometric fitting parameters (normalized to r_{ss}) $|\tilde{r}_{pp}|^2$, $\Re(\tilde{r}_{pp})$, $\Re(\tilde{\alpha})$ and $\Re(\tilde{\alpha}\tilde{r}^*_{pp})$ (with $\tilde{\alpha} = r_{sp}/r_{ss}$). With these four values we can calculate the values of the normalized Jones elements $\tilde{\alpha}$ and $\tilde{r}_{pp}$ of the single air/MO metal interface at the used setting of the incidence angle[8]. And these 2 complex values lead to a set of four higly nonlinear equations in the unknown n and g, via the longitudinal MOKE formulae of [10]. It is clear that this procedure will be even more sensitive to error propagation than the dynamic method of the previous section. It can now already be seen that this is mainly caused by the fact that Berger's method intrinsically

[7] Actually this is not typical of Berger's method but is a general basic principle of static photometric ellipsometry as proven in [14]

[8] Actually this is not entirely correct. It can be easily seen that if, out of the four fitting values a certain pair of complex numbers $\tilde{\alpha}_1$ and $\tilde{r}_{pp,1}$ is calculated then the complex conjugate pair $\tilde{\alpha}^*_1$ and $\tilde{r}^*_{pp,1}$ is obviously also a solution. However it can be proven that the additional constraints of a positive real part and a negative imaginary part for the extracted refractive index always cancels one of the solutions as physically unfeasible. In any case, this expresses in fact that Berger's method is underdetermined. This is already one indication that this method is not the most ideal to use.

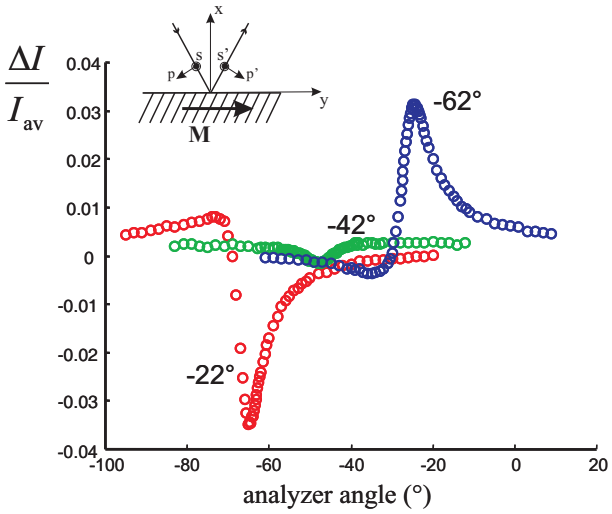


Fig. 4.5: Static photometric ellipsometry on an unprotected opaque (100nm) CoFe sample at a wavelength of 633nm. The polarizer angles and analyzer angles are each defined in their natural frame and not with respect to the global lab frame.

determines the isotropic ánd gyrotropic constants in a single fitting. As a result the four fitting parameters will all have about the same amount of inaccuracy, and is not possible to separately increase the accuracy on either the isotropic or the MO constants (as was done in the dynamic method).

First experiments with this setup on unprotected, thick (100nm) $Co_{90}Fe_{10}$ sample are illustrated in Fig. 4.5. The resonances observed in these curves are in perfect agreement with what is expected. Indeed, when the magnetization switches back and forth the Kerr contribution to the polarization rotation and ellipticity changes sign, while the isotropic complex polarization rotation is unaffected. This means that depending on the direction of the magnetization the isotropic polarization rotation and ellipticity gets either "counteracted" or "amplified" by the Kerr rotation and ellipticity. As a result when the analyzer scans through the reflected polarization ellipses the detected periodic intensity functions (with period π) are very slightly shifted with respect to eachother (both "horizontally", i.e. a change in position of the major axis and "vertically", i.e. a change in ellipticity) when the magnetization is switched. The absolute difference of these two intensity function will again be a smooth sine-like function with period π. However, the average intensity as a function of analyzer angle hardly changes from the pure isotropic case. As a result the fractional intensity change, $\Delta I/I_{\text{ave}}$, will be "amplified" in the regions where the average intensity is small. This is at analyzer settings that are quasi-crossed with the isotropic rotation. This leads to a resonance of the MOKE induced difference of the intensity at an analyzer setting corresponding to the isotropic rotation $\pm\pi/2$. This explains the resonances and their positions in Fig. 4.5.

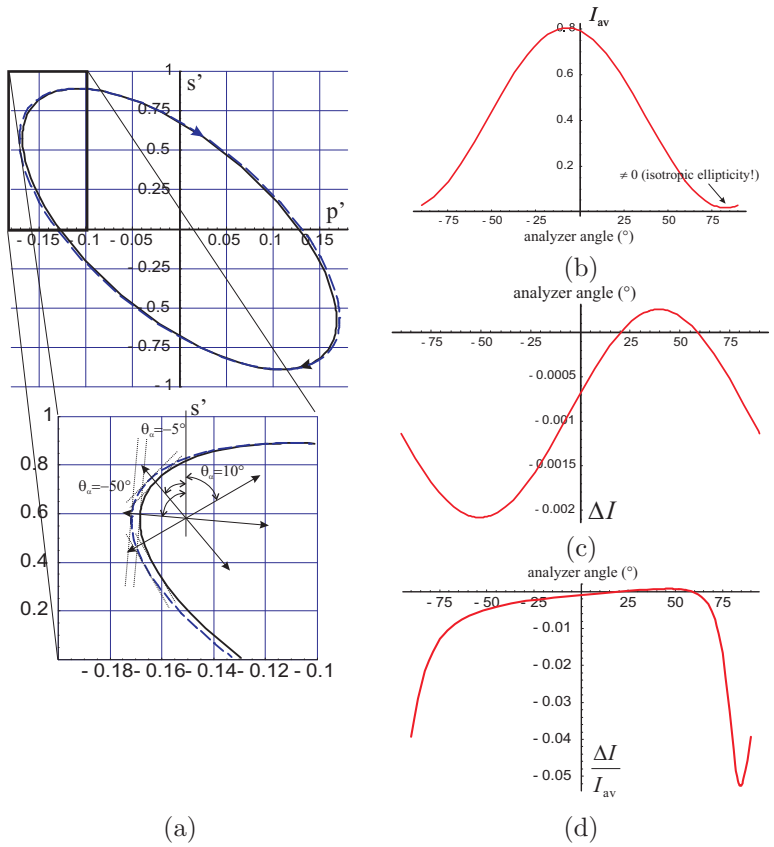


Fig. 4.6: Hypothetical example of Berger's measurement method. The incidence linear polarization makes an angle of 10˚with the s-axis. The incidence angle is set at 75˚. The magneto-optical and optical parameters for the bulk hypothetical MO material are set at $n = 4.4 - i4.77$ *and* $g = -5.0 + i2.0$. *(a): The reflected polarization ellipses in the natural s-p frame of the reflected light beam. Note the difference in scale! The dashed curve is the polarization ellipse with the magnetization switched to* $-y$ *in the lab frame (see Fig. 4.5). Expanded view of these ellipses illustrating the origin of the method (note again that the difference in scales distorts the angles); (b): The average intensity is approximately equal to the pure isotropic reflection. Therefore, the maximum is found near -10˚and the minimum near 80˚in the natural s-p frame of the reflected light. Note also that the two intensities (magnetization along* $+y$ *and along* $-y$*) could not be resolved if these were plotted in this figure; (c): The difference between the two intensities. The MOKE effect is "hidden" in the amplitude and the "vertical" offset of this sine-like function; (d): The fractional intensity change obtained by dividing (b) and (c). As a result the MOKE-induced changes in the polarization ellipses get projected onto the minimum of the average (≈ isotropic) intensity function.*

Fig. 4.6 shows a hypothetical example illustrating the above qualitative assessments. In order to make the effect better "visible" the polarizer is set at a small angle of 10°(to keep the isotropic rotation[9] and ellipticity limited), and a hypothetically strong value for the gyrotropy parameter is taken, $g = -5.0 + 2.0i$. For the refractive index the extracted value for $Co_{90}Fe_{10}$ of the previous section is used, $n = 4.4 - 4.77i$. Finally, the incidence angle is taken to be 75°, as it is known that the longitudinal MOKE increases with increasing incidence angle.

Even though the measurements of Fig. 4.5 have a behaviour that is in accordance with the above assessments, it was not possible to obtain a reasonable fit for the optical and magneto-optical parameters. As it was previously noticed that the presence of a natural oxide layer on top of CoFe is very probable (see Fig. 4.2), it was chosen to check whether the fitting problems were mainly due to this possible rather thick oxide layer (causing multiple interfering reflections which can not be included in Berger's model) or to some other experimental limitations of the method. For that purpose, it was decided to try and calibrate the setup with a well-characterized magneto-optic metal. A very thick (200nm) Ni layer was sputter-deposited. The optical and magneto-optical constants for Ni are known throughout the infra-red, visible and ultra-violet spectral range [15]. On top of that it is known that the natural oxidation of Ni is just like for Al self-limiting. Thus, the eventual presence of an eventual oxide layer on this Ni layer shouldn't have a strong influence[10]. Fig. 4.7 presents the results of these calibration tests on Ni both for visible light (632nm) and IR light (1340nm).
It can be seen that the measurements at 632nm are "noisier" than at 1340nm, but this is probably due to the worse stability of the used HeNe laser source compared to the much more stable semiconductor IR laser diode. The nonlinear least-squares fit in both cases is acceptable and the convergence of the used optimization algorithm[11] was very robust. Table 4.3 summarizes the results of the fitting procedure. As already mentioned these parameters are found through the solution of a set of four coupled nonlinear equations in $\Re(n), \Im(n)$, $\Re(g)$ and $\Im(g)$ with the fitting parameters $|\tilde{r}_{pp}|^2$, $\Re(\tilde{r}_{pp})$, $\Re(\tilde{\alpha})$ and $\Re(\tilde{\alpha}\tilde{r}^*_{pp})$ and the angle of incidence as coefficients. Obviously this requires a second optimization routine. The results of these calibration tests are very striking. This ellipsometric setup achieves an acceptable accuracy of about 10% for the optical constants despite its apparent basicness. However, it completely fails in its main purpose. The deviations on the magneto-optical gyrotropy parameter are sometimes no less than 400%! While the literature values in infrared are subjected to some degree of uncertainty [15], they are certainly not uncertain over a factor 4. It is also intuitively clear that these false results are not caused by an oxide layer. Indeed, if that were the case, a much higher deviation would also have been observed in the isotropic parameters. It is

[9] Remember that upon external reflection off a metal s-polarized light undergoes a phase shift close to π, while p-polarized light has only a small positive phase shift, hence the change of the major axis of the polarization to negative values in the s-p frame of the reflected light beam.
[10] It was actually proven by Atkinson [16, 17] that any magneto-optic multilayer containing highly absorbing materials can under certain conditions be replaced a single equivalent layer.
[11] A Broyden-Fletcher- Goldfarb-Shanno method was used [18].

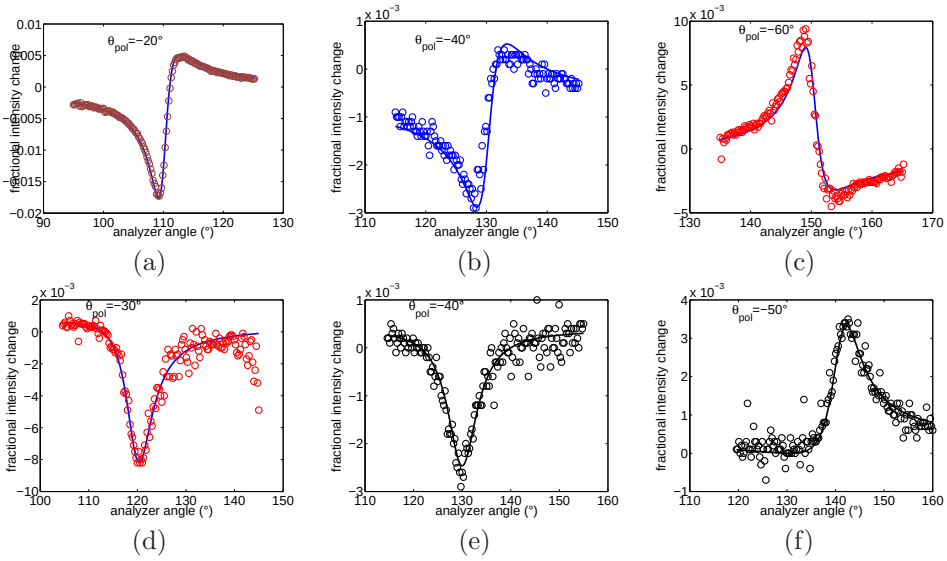


Fig. 4.7: Calibration tests of the setup at 1340nm [(a), (b) and (c)] and 632nm [(d), (e) and (f)] on a thick Ni sample, together with the simultaneous nonlinear least-squares fit of (3.29)*. The incidence angle in both cases was 30˚.*

	633nm		1340nm	
	exp.	lit.	exp.	lit.
n	1.78	1.92	3.228	3.18
κ	3.91	3.75	5.885	6.23
$\Re(g)$	-0.057	-0.023	-3.33	~ 0.6
$\Im(g)$	-0.52	-0.105	-0.797	~ 0.3

TABLE. 4.3: The extracted values for Ni compared to the values in literature.

our conviction that this failure to accurately measure the magneto-optical constants with the here described setup can be entirely attributed to an inadequate calibration of the setup, more in particular a good calibration of the azimuthal settings of the polarizer and the analyzer.

In order to illustrate what a faulty calibration can do to the extracted values for the MO parameters, consider Fig. 4.8, where an example is plotted illustrating the influence of the real and imaginary parts of the magneto-optical parameter on the here described measurement method. In this simulation the extracted parameters for $Co_{90}Fe_{10}$ of Table 4.2 are used. The curves corresponding to the extracted values are plotted in dashed style. The incident polarization was set at -20˚, and an incidence angle of 30˚was used. The real and imaginary parts of the gyrotropy parameter are varied by approximately a factor 2 in both senses.

The imaginary part of g mainly influences the angular position of the resonance, while the real part mainly influences the amplitude of the resonance. This is in qualitative agreement with the fact that the polarization rotation of the MOKE is principally caused by the imaginary part of g and that the real part of g causes the MOKE ellipticity. These simulations make clear that a faulty calibration of the azimuth of the analyzer of just a few tenths of a degree can cause a deviation of close to 100% in the extracted value for $\Im(g)$. On the other hand it is quintessential that the fractional intensity change must be measured with an accuracy of better than 1%. This last remark is not in vain, as it was noticed that the limited available light power at the detector in our experiments fluctuated easily over more than this amount. This is not surprising in view of the fact that the resonance in this measurement method is always located at analyzer angles close to the minimum average intensity. One way to solve this problem is by using a more intense incident light beam. Next to that, the azimuth of the analyzer must be calibrated with a precision of at least a 10 cdeg. This latter is not a trivial task as it is close to the resolution of the used rotary stage, but can in principle be done using the calibration methods described in section 3.1.1.1.

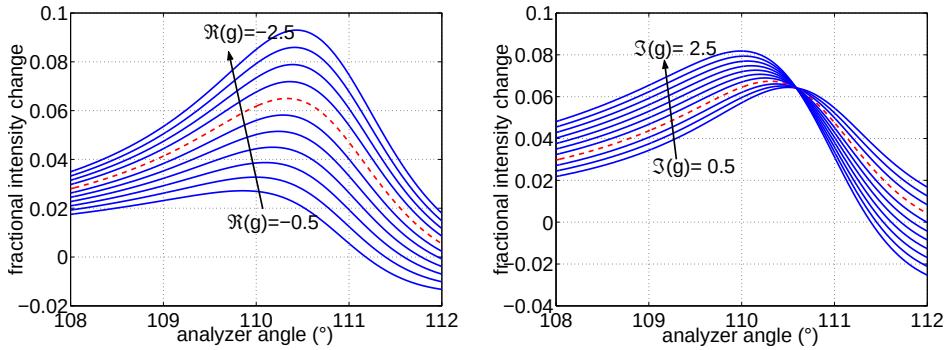


Fig. 4.8: Influence of variations in the MO parameter on the measurement method of Berger. Simulations are for $Co_{90}Fe_{10}$ at an incidence angle of 30°and with an incidence polarization of -20°. The dashed curves correspond to the measured values of the dynamic method.

4.1.1.3 Conclusion

In this section the experimental results of both a dynamic and a static ellipsometric technique for the optical and magneto-optical characterization of the used CoFe alloys have been presented. The very specialized dynamic setup achieves a very good accuracy and the extracted values seem to confirm the popular belief that also in the infrared the symmetrical composition is indeed magneto-optical stronger. The design simulations of section 3.3.2 prove that with the here measured strengths

the isolator concept is perfectly feasible, at least from a non-reciprocal viewpoint. The attempt at a basic, low-cost and fast static ellipsometric measurement method failed in a first instance. However, the shortcomings have been indicated, and are not insurmountable. Future optimizations of this setup might improve its accuracy. Already it achieves a good accuracy for the extraction of the optical constants. One important drawback however is that it is probably not possible to accurately characterize a magneto-optical multilayer (such as an oxidized or a passivated sample) with this method without having to sacrifice numerical accuracy.

As a final remark, we would like to point out that all materials have been characterized at the ambient interface and not at the substrate interface. The reason is obvious. Characterization through the substrate would require a very precise knowledge on a subwavelength scale of the thickness of this layer in order to account for the multiple reflections. This lack of precise information of the magneto-optical strength at the interface between the MO metal and the semiconductor might be seen as a drawback. It is indeed conceivable that at this interface a layer is formed (during deposition or posterior during heat treatments) that has lost its magneto-optical properties. However, if such a magnetic "dead" layer were present its thickness would not extend over more than a few monolayers and its effect on the performance of the isolator can be neglected. Future sections will indeed prove experimentally that the formation of a magnetic dead layer is negligible.

4.1.2 Magnetic anisotropy of CoFe on InP-based semiconductors

It has been explained in section 3.1.2.4 how for the magnetic characterization of the thin film MO metal contact it is not necessary to extract the specific values of the anisotropy constants. Sufficient information about the in-plane anisotropic behaviour can be obtained by investigating the squareness of the hysteresis curves of the magnetic films. Next to that, it is questionable what meaning can be attributed to an anisotropy constant in non-epitaxial polycrystalline films. The measurement of the hysteresis curves is done using longitudinal MOKE or an Alternating Gradient Field Magnetometer (AGFM)[12]. The latter also allows to measure the absolute magnetization of the film when its precise dimensions are known. In these hysteresis curves we will be looking for indications of the distinctness of the easy axis in the plane of the film and its stability by investigating the values of the coercive field in the different in-plane directions. As was explained in the previous chapter a commonly used measure collecting both properties is the squareness ratio S^*, defined in (3.36).

The discussion of section 3.1.2 has led to the conclusion that for the here considered polycrystalline films and in view of the specific application of magneto-

[12] Actually, any magnetic measurement method that produces a signal that is linearly proportional to the magnetization allows the measurement of the hysteresis curve, let this be the amplitude of the periodical displacement of a sample in a harmonic varying magnetic force field, the detected light intensity along a polarization direction that is quasi-orthogonal to the incident polarization of light reflecting off a magneto-optic surface, or even the amplified spontaneous emission of a MO metal-clad seminnconductor optical amplifier,...

optic contact layer for an integrated isolator, the desired easy axis behaviour can only be induced by applying a magnetic bias field during sputter deposition. Texturizing the subtrate by deposition of thin metal host films with a more suitable crystallographic texture is out of the question and post-treatment through magnetic annealing requires temperatures that destroy the electric contact (as will be seen later). All CoFe films have thus been deposited under the presence of a magnetic bias film of about 100 Oe by placing the samples between small permanent magnets during deposition. The mechanisms that control the direction of the easy axis during such biased sputter deposition are very hard to model. Therefore, the pragmatical approach was taken to investigate a number of films whose only varying parameters were their geometrical dimensions. All films have been deposited on a epitaxial highly doped ternary InGaAs layer, this being the semiconductor material that will be used as semiconductor contacting layer in the final device (see further in section 4.1.3).
In a first run the hysteresis properties of blanket films have been investigated as a function of thickness. This serves as a first indication whether the intended in-plane uniaxial anisotropy, induced by the sputter magnetic bias field, is present and, if so, whether it will be strong enough to counteract the shearing of the hysteresis curve due to shape anisotropy in the final device layout. In a second run the effects of shape anisotropy have been specifically experimentally characterized by hysteresis measurements on patterned CoFe stripes with varying thickness and varying width. As a final remark, it has to be pointed out that all measurements have been performed on the $Co_{90}Fe_{10}$ alloy. The reason for this is very pragmatical. At the time of these experiments a $Co_{50}Fe_{50}$ target was not available yet in the sputter system. When it became available and after it had been established that it was better candidate from a pure magneto-optical viewpoint, it was decided to check its anisotropy directly on the non-reciprocal SOA waveguide setup (see Chapter 5). There is reason to believe that the magnetic anisotropy of $Co_{90}Fe_{10}$ might differ significantly from that of $Co_{50}Fe_{50}$. It has been noted earlier that Co-rich alloys have a hcp crystal structure, while the equiatomic CoFe has a bcc crystal structure. This has profound implications for the crystalline anisotropy if there is any degree of epitaxy. A hcp crystal has a perpendicular easy axis along its c-axis while it is isotropic and relatively hard in-plane. Bcc crystals on the other hand have a relative hard axis behavior along $\langle 111 \rangle$-directions, while the $\langle 100 \rangle$- and $\langle 110 \rangle$-directions are relatively easy axes.

Fig. 4.9 shows the hysteresis curves measured on four blanket $Co_{90}Fe_{10}$ films of increasing thickness. These measurements have been performed on a MOKE setup with visible HeNe light. With the input polarizer set to pure s-polarization and the sample saturated, the detected light intensity is minimized by varying the azimuth of the analyzer. This analyzer setting will roughly correspond to the minor axis of the polarization ellipse induced by the s-Kerr rotation. During the magnetization reversal process the s-Kerr rotation (and ellipticity) will change sign and the detected intensity will emulate the hysteresis of the magnetic film. The sample is mounted on a rotation stage allowing to vary the direction of the film plane axes with respect to the applied longitudinal magnetic field through an azimuthal rotation. This allows for the measurement of the magnetic hysteresis

along the different in-plane directions.

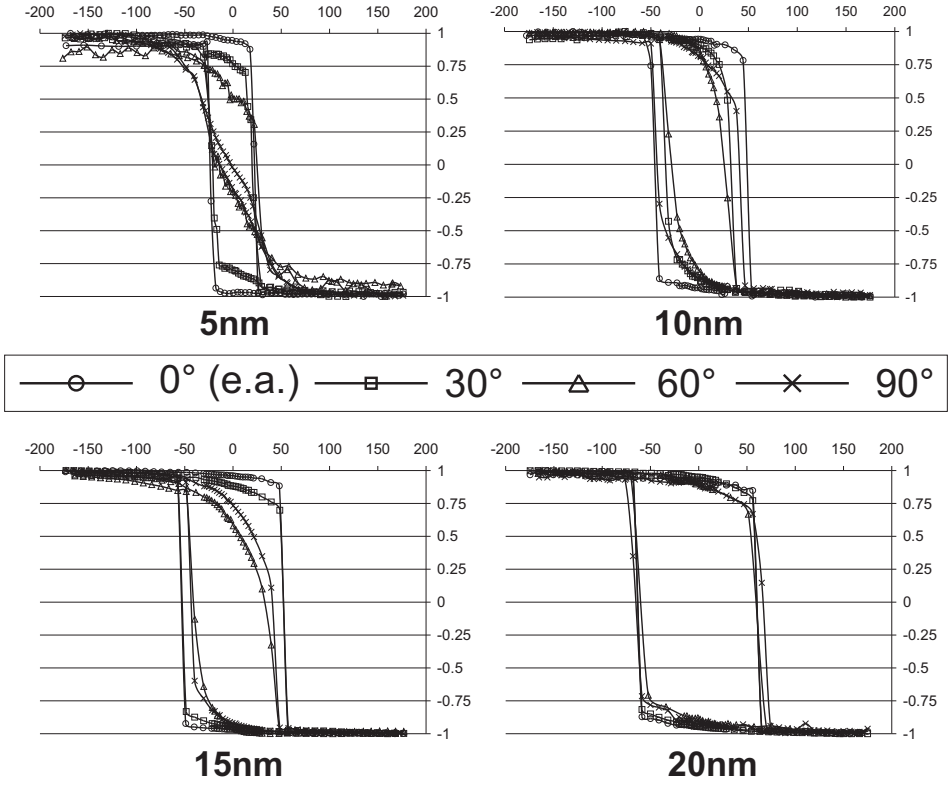


Fig. 4.9: In-plane MOKE hysteresis curves of blanket $Co_{90}Fe_{10}$ films with varying thickness. The reference in-plane direction (0°) corresponds to the direction along which the bias field was applied during sputtering, and thus to the desired easy axis direction. This direction corresponds to a $\langle 110 \rangle$ of the zincblende crystal structure of InGaAs. The MOKE intensity curves are normalized to the saturation values of the detected intensity.

The thickness dependence of the in-plane magnetic anisotropy depicted in these curves is, to say the least, somewhat surprising. Only at very thin thicknesses the applied magnetic field during sputtering succeeds in creating a strong enough anisotropy. However, the coercive field is low ($H_c \approx 25$ Oe). As soon as the thickness increases above 10nm, the induced uniaxial anisotropy is gradually decreasing. At 20nm the film behaves perfectly isotropic with coercive fields in all directions of the order of 50 Oe. There is no clear understanding of what is precisely counteracting the magnetic bias, but a possible explanation for this disconcerting result might be the occurrence of some degree of hcp epitaxy. Co-rich alloys have this crystal structure which is characterized by an uniaxial anisotropy with an easy

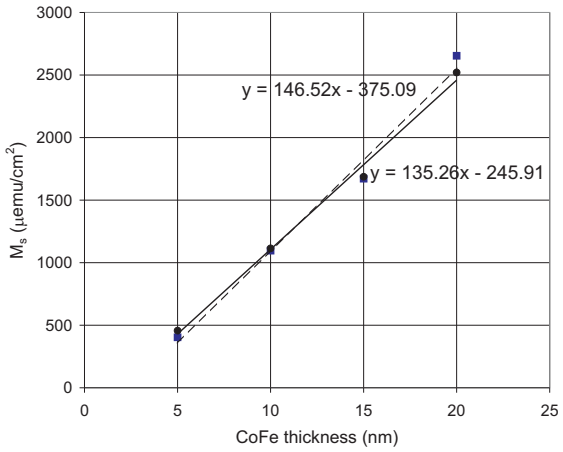


Fig. 4.10: AGFM measured surface magnetization as a function of blanket film thickness on two series of samples. The slope of about 140 $\mu emu/cm^2 nm$ is in agreement with the value known for $Co_{90}Fe_{10}$. The intersection with the thickness axis of the linear trendlines predict dead layers of a maximum thickness of about 2nm.

axis perpendicular to the hexagonally close-packed planes. On top of that it is known that the anisotropy constant of Co is very high, $K_u = 4.1 \times 10^5$ J/m^3. It is not sure whether this suspected epitaxy is present. This has to be checked by microscopic material evaluation or by direct anisotropy measurements using torque magnetometry. In any case, these results prove that it is far from clear whether patterned CoFe stripes will exhibit sufficient anisotropy in the in-plane direction perpendicular to the stripe. The good news is that these experiments teach us that the presence of a magnetic dead layer at the interface between CoFe and InGaAs is suspected to be negligible. This suspicion is dictated by the fact that CoFe layers as thin as 5nm still produce a strong MO signal and is confirmed by AGFM measurements on these blanket samples. By measuring the absolute magnetization of these samples per cm^2 as a function of thickness of the blanket films a linear curve is expected with a slope that should approach the known magnetization density 1460 emu/cm^3 of $Co_{90}Fe_{10}$. Linear extrapolation of this curve allows an estimate for the thickness of the "magnetic dead layer". This is plotted in Fig. 4.10

The apparent in-plane magnetic isotropy of the CoFe blanket films combined with the relatively low coercive fields leads to suspect poor anisotropy properties in patterned CoFe stripes. After all, recalling the discussion on shape anisotropy of section 3.1.2.3, a demagnetization factor of merely[13] $N_{\mathrm{dem}} = \tan\alpha_{\mathrm{shear}} \approx H_c/M_r \approx 10^{-6}$ would cause a serious drop of the remanent magnetization. There-

[13] As a reminder: 1 emu/cm^3 = 1000 A/m and 1 Oe $\approx$ 0.0125 A/m.

fore, an extensive experimental study of shape demagnetization effects has been carried out. For that purpose, sputter-deposited $Co_{90}Fe_{10}$ stripes with respective widths 7μm, 20μm and 50μm have been fabricated using standard lift-off techniques. The CoFe thicknesses used in these experiments corresponded to the three different degrees of observed anisotropy in blanket films, namely 5 (good anisotropy), 10 (decreasing anisotropy) and 20 nm (quasi-perfect isotropy). For the here studied application however, the anisotropy behaviour of "thick" stripes is of primordial concern. It was impossible to measure the hysteresis curves via MOKE, as the rather small surface coverage of the samples by magnetic material[14] resulted in only a small fraction of the laser beam cross-section being subjected to the Kerr effect. An AGFM doesn't suffer from this drawback, as its signal is proportional to the total volume of magnetic material present on the sample under study. However, even in the AGFM the 5nm samples produced signals close to the detection limit of 100nemu, and only the widest 10nm stripes led to detectable AGFM signals. Therefore, it is not really possible to give a comparison of shape anisotropy effects as a function of thickness. But after all, as said, we are mostly interested in the behaviour of thick stripes. Fig. 4.11 shows the AGFM hysteresis measurements on 20 nm stripes along directions parallel and orthogonal to the length axis.

These hysteresis plots should in principle allow for a study of the demagnetization effects. A qualitative study indeed reveals that with decreasing stripe width the squareness of the hysteresis loops along the in-plane direction perpendicular to the stripe length decreases and the remanence drops. What is very striking about these measurements though is the relatively low degree of squareness of the hysteresis curves measured along the stripes. The samples used in these experiments were cleaved into rectangular pieces of a length of several mm. In other words, creating stripes with a l/w aspect ratio of the order of 10^2. Under these circumstances one should expect a squareness along the length of the stripes comparable to the one of the blanket film depicted in Fig. 4.9. These results seem to suggest that the magnetically induced anisotropy field at sputtering is more effective in micron-size patterns than in blanket films. In other words, if the explanation of the existence of some degree of epitaxy for the behaviour of the blanket hysteresis curves is accepted, then the hysteresis curves of the patterned structures seem to indicate that the epitaxial tendency vanishes in micron-sized patterns, and a low degree of uniaxial anisotropy is indeed introduced by applying a magnetic bias field during sputtering.

Another striking property of the depicted hysteresis of patterned structures is the strong reduction of the coercive fields to levels of only 20-25 Oe as compared to coercive field of 50 Oe for blanket films of the same thickness. Strangely enough the same reduction is observed in the longitudinal direction for which, as explained above, there is no reason to believe that the behaviour should differ from the blanket case. It is not clear what causes this, since it cannot be explained by a shearing effect due to shape anisotropy. Indeed shape anisotropy alone would in principle not affect the value of the coercive field, as the magnetization disappears for those fields and the internal and external applied magnetic fields would

[14] The used lift-off mask has a stripe spacing about 100μm.

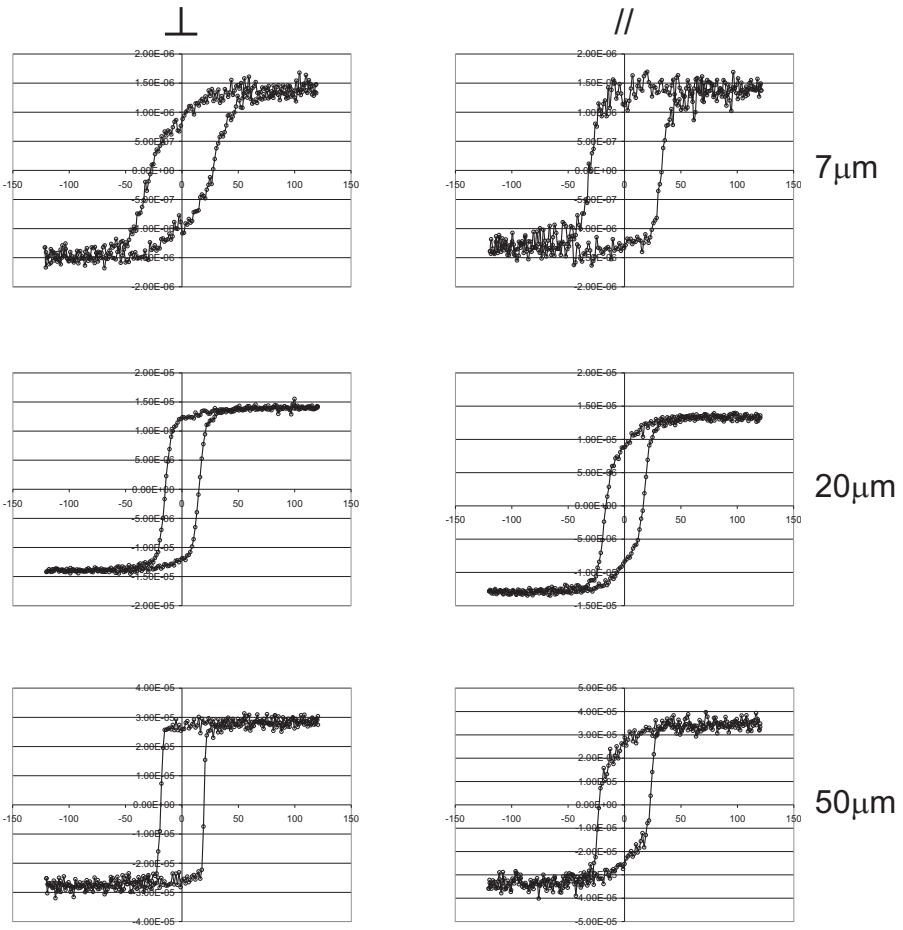


Fig. 4.11: In-plane hysteresis of 20nm thick CoFe stripes with a width of 7, 20 and 50 μm measured by AGFM along and perpendicular to the stripe length. The magnetic moments are measured in emu (=1000 Acm^2), while the field is varied up to 150 Oe (1 Oe ≈ 1/80 A/m). The precise value for the magnetic moment is not so relevant as the precise volume of magnetic material on the samples was not known. This is not too important as all relevant information can be extracted by normalizing the curves to the measured saturation moment. However, it should be noted that due to the low coverage of the measured samples by the stripes, the measured moments are only of the order of 1-10 μemu, which is particularly for the thinnest samples close to the precision of the AGFM.

be equal in that case (recall (3.35)). This latter observation makes it impossible to relate the measured hysteresis curves to the corresponding blanket hysteresis curve by the straightforward shearing transformation $M = \frac{1}{N_{\text{dem}}}(H_{\text{applied}} - H^*)$ deduced in section 3.1.2.3. All one can conclude from these experiments is that

the desired uniaxial magnetic anisotropy (induced by the magnetic bias field at sputtering) is somewhat recovered as soon as the stripe width exceeds values of 50 μm, as opposed to the observations on blanket films. This is confirmed by the trends of the measured squareness values S* of these curves[15] (see Table 4.4).

S*	$\perp$	$\parallel$
7μm	0.65	0.75
20μm	0.57	0.625
50μm	0.86	0.762

TABLE. 4.4: Squareness (calculated along (3.36)*) for the CoFe stripes of Fig. 4.11*

In this section the magnetic anisotropy properties of $Co_{90}Fe_{10}$ sputtered on epitaxial InGaAs layers under the presence of a magnetic bias field have been studied. It has been noted that in blanket films the desired in-plane uniaxial anisotropy can only be realized in very thin films. As soon as the thickness increases above 10nm a magnetic isotropic behaviour reminiscent of the magnetocrystalline anisotropy of epitaxial Co is observed. Surprisingly enough, the desired magnetic anisotropy was partly recovered in patterned CoFe stripes, despite the experimentally confirmed presence of shape demagnetizing effects. Still the degree of anisotropy remains low and for a stripe length of several mm's a width of at least 50 μm is necessary to keep the squareness of the perpendicular direction close to 0.9. Even then care has to be taken, as the coercive fields remain as low as 25 Oe. Therefore, even though it is very unlikely that the device will come close to other magnets, a stray field of about this strength can alter the remanent magnetization direction in the magneto-optic contact, and ruin the operation principle of the integrated device. Another profound implication of the above conclusions is that in the processing of the SOA/isolator the metal contact width has to be kept as large as possible and as planar as possible in order to avoid any physically breaking of the metal stripe in the perpendicular direction. This implies that serious attention has to be paid to planarization of the current isolation layer on which the CoFe will be deposited. This will be illustrated in the next chapter by SEM pictures of the facets of the processed prototype integrated isolators.

4.1.3 P-type CoFe Ohmic contacts to InP-based semiconductor compounds

4.1.3.1 Semiconductor contact layers

It has been explained at length in section 3.1.3.2 that owing to Bardeen's 1/3 rule (Fermi level pinning) for covalent semiconductors, like the InGaAsP compounds, the Shottky barrier height can only be lowered by using a semiconductor

[15]The squareness is defined in (3.36) and can be derived out of the measured hysteresis curves after normalizing them by the measured saturation moment. This is necessary as the precise magnetization (i.e. magnetic moment/volume) is not exactly known, as the precise volume of magnetic material has not been determined.

material with a minimal band gap, and depends only very weakly on the metal work function. Therefore, the realization of a p-type Ohmic contact on InP-based semiconductors is mainly achieved by maximizing the saturation current density by making quantummechanical tunnelling the dominant current mechanism of the metal/low-band-gap semiconductor Schottky diode. This can be basically achieved by maximizing the p^{++} dopant concentration in the semiconductor contacting layer.

Next to these general guidelines, it has been remarked that the specific application studied here should preferably have a semiconductor contacting layer with a band gap larger than 0.96 eV, being the band gap corresponding to an absorption edge of 1300nm. In this way extra optical absorption in the semiconductor layer beneath the MO metal, which would shield the guided light from the region inducing the non-reciprocal effect, can be minimized. This slightly complicates matters, as it is known that it is technologically more difficult to achieve high doping levels in large band gap semiconductors. Even though it is expected beforehand that the best CoFe Ohmic contact would be achieved on ternary $In_{0.53}Ga_{0.47}As$, which is the alloy composition with the lowest band gap within the InGaAsP alloys lattice-matched to InP, the optical absorption in such a layer would easily exceed levels of $2–5\times10^4 cm^{-1}$. With a standard thickness of some 100nm for this layer this can shield about 20% of the guided "incoming" light away from the MO contact[16].

Therefore, it was decided to characterize three types of contacting layers: "standard" (absorptive) ternary InGaAs (as a reference), transparent quaternary InGaAsP (with a $E_g = 1.05$ eV), and a hybrid quaternary/ternary contact. The latter contacting structure was tried to check whether it was possible to combine the best of both worlds: the transparency of the quaternary layers and the low band gap and high dopant solubility of ternary layers. Table 4.5 summarizes the grown test contacting layers.

As a contact layer is more of a necessary evil for the studied integrated isolator

label	composition	thickness (nm)	dopant concentr. (cm^{-3})
T	InGaAs	100	3×10^{19}
Q1	InGaAsP ($\lambda_g = 1.17\mu$m)	100	2×10^{19}
Q2	InGaAsP ($\lambda_g = 1.17\mu$m)	50	2×10^{19}
H1	InGaAsP / InGaAs	100 / 15	2×10^{19}/ 3×10^{19}
H2	InGaAsP / InGaAs	100 / 5	2×10^{19} / 3×10^{19}

TABLE. 4.5: Semiconductor characterization contact layers.

than a benefit (as it after all acts as an extra separation between the guiding layer and the MO metal) it was chosen to study the possibility of reducing the thickness of this layer, hence the two different thicknesses for the transparent quaternary layer (labelled Q1 and Q2). The hybrid structures were grown in a later stage, after experiments made it clear that if quaternary material was going to be used

[16]this very rough estimate is obtained by calculating the imaginary part of the transverse wavevector in the contacting layer using values for the effective index obtained in the benchmark calculation of Chapter 2.

in the contact it had to be at least 100nm thick. Therefore, in these two (labelled H1 and H2) it was only checked how thin the absorptive ternary layer could be made without sacrificing too much of its good Ohmic qualities. All layers were epitaxially grown by MBE on a non-intentionally doped InP substrate on which first an extra undoped 100nm InP buffer layer was grown. The used dopant in all 5 structures was Be. It can be seen from Table 4.5 that despite the mentioned limited solubility of p-dopants in high band gap semiconductors, careful optimization of the growth temperature and of the ion flux made it possible to achieve record concentrations of Be in the quaternary layers. The used dopant concentrations should in principle be sufficient to activate the field-emission (i.e. tunnelling) mode of current transport.

4.1.3.2 Characterization structures

The epitaxial grown contact layer structures were processed into contact characterization structures using a dedicated mask. Fig. 4.12 shows a unit cell of this mask. In this figure the used Cross Kelvin Resistor (CKR) and Transmission Line Tap Resistor (TLTR) characterization structures can be clearly distinguished. There are 4 CKR structures with a square contact window (side widths 3, 5, 10 and 20 μm), each of them symmetrically placed with respect to their diffusion and sensing arm which is 3 μm wider (thus 6, 8, 13 and 23 μm). There are also 4 types of TLTR structures. As opposed to the CKR structures, these are designed to be described by a 1D current model, i.e. with a metal contact covering the entire diffusion arm. The varied parameter between these 4 TLTR's is the width of the metal contact, resp. 2, 4, 10 and 20 μm. Each TLTR has 6 contacts for which the subsequent spacing are repeatedly doubled, starting at an intercontact spacing of 5 μm.
Just as in the previous section, only $Co_{90}Fe_{10}$ has been used in these contact experiments, for the very pragmatic reason that the equiatomic sputter target was not available yet at the time of these experiments. Again, a later characterization of the Ohmic properties of $Co_{50}Fe_{50}$ could be checked afterwards if this would be deemed necessary upon eventual poor IV-characteristics of the integrated SOA/isolator. The thickness used for the metal was 50nm. And to protect it from damage, the probing squares were covered with an extra protective Ti/Au bilayer (40/150 nm). The dimensions of the unit cell are approximatively 2mm×1.5mm, meaning that a processed quarter of a 2" wafer will contain well over 100 copies of the unit cell. The processed pieces are cleaved into 4 pieces, of which one is left unalloyed and the three others are subjected to Rapid Thermal Processing[17] (RTP) at resp. 250°C, 350°C and 450°C.

4.1.3.3 Specific contact resistivity characterization

CKR versus TLTR It has been extensively explained in section 3.1.3.3 how the CKR and TLTR structures allow the extraction of the specific contact resis-

[17]The temperature is increased to the setpoint in a time span of about 20–30sec, after which the heater is switched off and the sample is left to cool down.

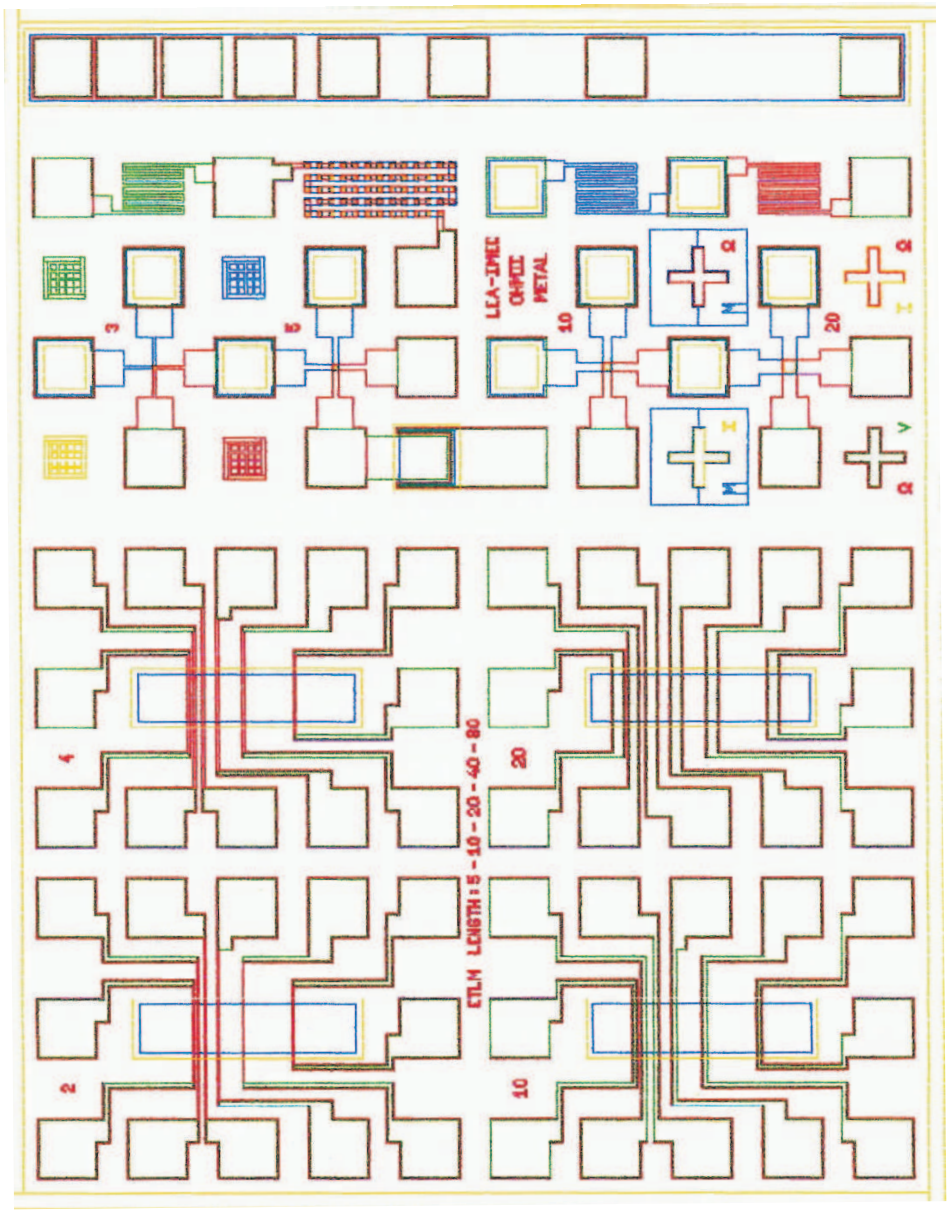


Fig. 4.12: One unit cell of the contact mask.

tivity, ρ_c. It was indicated that for this purpose, even though the TLTR is the least influenced by parasitic effects due to current crowding, this resistor structure suffers from inaccuracies due to the linear regression needed for the extraction of the small front resistance out of a set of resistance data points which are usually

two or three orders of magnitude higher (see Fig. 3.12 and the accompanying explanation). Even though the processed TLTR structures in our experiments are essentially 1D (as can be seen from Fig. 4.12 where the metal is seen to cover the entire diffusion mesa), so that the parasitic effects due to the current crowding are not present, this doesn't change the need for linear regression for the extraction of the front resistance and the accompanying inaccuracies. In short, as was already remarked by Loh [19], the TLTR is not the ideal characterization structure for contact resistivity experiments (unless the expected contact resistivity is very high). This was confirmed by our experiments [20]. Relative errors of up to 300% have been observed on the intercept values obtained through linear regression! Next to that, another drawback of the specific 1D TLTR structures on the mask of Fig. 4.12 is the fact that the metal contacts are not square at all but rectangular with rather high aspect ratios (up to $w/d = 40$). And because the current flow in the metal contact is in a direction perpendicular to the current flow in the semiconductor diffusion mesas underneath the contact, the sheet resistance per square in the metal can unexpectedly cause an extra important "geometric" contribution to the total measured contact resistance, even though the sheet resistance of the metal will be much lower than the sheet resistance of the semiconductor diffusion mesa[18]. Such an extra contribution to the resistance is almost impossible to model without ressorting to a full 3D model of the current flow through the structure. For these reasons the TLTR structures have only been used to extract the sheet resistance of the semiconductor diffusion regions, and the specific contact resistivity was extracted using the CKR structures, combined with a developed 2D current modelling tool. This is in accordance with the conclusions of section 3.1.3.3, where it was noted that a square 2D CKR structure is the best compromise between (limited) current crowding and extraction accuracy.

CKR measurements Fig. 4.13 shows a detailed picture of a typical processed CKR structure. This figure serves to illustrate the possible imperfections during processing. Due to misalignments of the several mask layers and the underetching of the semiconductor mesa, the dimensions of the contact structure can deviate from its design values. This is clearly the case in this example. Underetching has made the mesa less wide, decreasing w to almost the same width as d, the width of the square contact. Secondly, the contact window is not a perfect square, but at the corners it is a little bit rounded. This makes the contact area smaller which is equivalent to a smaller value for d. Thirdly, the inner corner of the L-shaped mesa is rounded, which causes the feeding and sensing arms to be somewhat wider right at the contact window, so that 2-D current crowding around the contact window is still occurring. Finally, a small misalignment of the isolation mask has resulted in a contact window not entirely symmetrically placed with respect to the semiconductor mesa.

[18]This is very easy to understand by comparing the number of "squares" in the metal that the carriers have to travel, to the number of squares in the semiconductor diffusion layer beneath the contact. In the worst case this is $\square_{\text{met.}}/\square_{\text{s.c.}} = w^2/d^2 = 1600$, with w the width of the mesa and d the width of the metal.

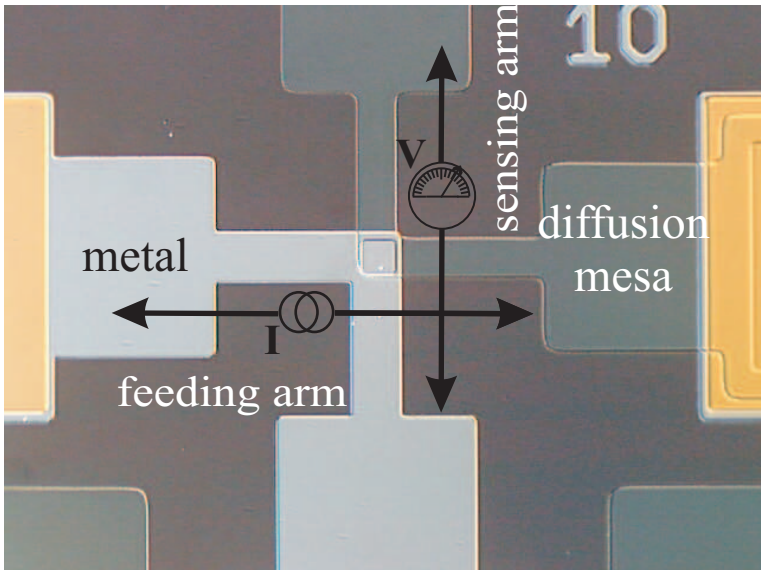


Fig. 4.13: Detail of a CKR structure with a square contact window of $10^2 \mu m^2$*. The isolation material (60nm* Al_2O_3*) itself is transparent, but it can be clearly seen where this isolation has been etched away to form the square contact between the CoFe and the doped semiconductor etched mesa.*

The reason why these imperfections are stressed here is that they cause a systematic error on the design parameter d/w, which plays a very important role in the 2D current model for the contact structure. It is directly related to the relative importance of current crowding in the characterization structure (at a given value for the transfer length of the contact). A systematic error on this parameter could cause the extracted value for the contact resistivity to become apparently geometry dependent. To understand why this is, it suffices to remember that with a known value for the sheet resistance, the transfer length and thus the specific contact resistivity of the contact is uniquely determined by measuring the Kelvin resistance on a single square contact with known d and w. This measurement will then uniquely fix the (absolute) Kelvin resistances on any other contact with different geometric parameters. Therefore, if CKR structures of different geometries are measured, the obtained resistances can only come in certain combinations in order to obtain the same contact resistivity. If there is a certain systematic error present (for instance a systematic decrease of $w - d$ due to underetching of the semiconductor diffusion mesa) it is clear that the accompanying under- or overestimation of the relative contribution of the current crowding mechanism for the different measured contact structures will lead to conflicting values of the measured Kelvin resistances and a falsely geometry dependent contact resistivity. Obviously, this effect will be all the more drastic at low values for the contact resistivity and for small contacts.

The general measurement procedure is as follows. First the sheet resistance is determined via the slope of the resistance of the TLTR structures. After that, the Kelvin resistance (or voltage) is measured over a large number of CKR's with equal geometry in order to minimize statistical errors. And finally a finite difference solution (implemented in a Matlab algorithm) of the 2D contact voltage model (3.48) is fitted to the measured Kelvin voltage using l_t as a fitting parameter and using the known sheet resistance R_s, the contact mask geometric parameters d and w and the feeding current I_{in} (i.e. the Neumann boundary condition at the current feeding arm $\frac{\partial V}{\partial x} = -\frac{R_s}{W} I^*_{\text{in}}$) as input parameters. The contact resistivity is then found as $\rho_c = l_t^2 R_s$. As a consistency check, this procedure is repeated for contacts with different geometries. If this leads to a discrepancy in ρ_c which cannot be explained within the margins set by the statistical errors, then a study of the processed structure is done in order to reveal any systematical errors on d and/or w. With these eventually corrected geometric parameters the fitting procedure is repeated to obtain a consistent geometry independent contact resistivity.
Fig. 4.15 gives an example illustrating typical simulated voltage distributions in the used contacts. This figure also serves as an example of the need for correction of the geometric dimensions of the processed contact structure in order to get geometry independent convergence. Note that in view of the results presented in this figure, it can be questioned why all CKR structures aren't essentially made 1D by reducing the dimension of the overlap region, $w-d$, to values much smaller than l_t. In this way, current crowding is always avoided, and the Kelvin resistance will equal $R_s l_t^2/d^2$, i.e. the parallel resistance circuit of the number of squares of side l_t (each with a resistance R_s). However, it is technologically almost impossible to make the overlap region narrower than 200nm, and for very good contacts this is still above the typical l_t value. Current crowding will therefore always contribute in a CKR, and it is inevitable to use a 2D tool to calculate its effect.

Characterization results Fig 4.14 summarizes the results of the CKR measurements on all five contact structures as a function of the different thermal treatments. The statistical relative error on these contact resistivities is in the range of 10–20%, and is caused both by the error on the sheet resistance R_s determined via the TLTR structures and by the error on the transfer length. The latter is much lower, but contributes quadratically via $\rho_c = l_t^2 R_s$.

The results presented in this figure immediately make clear that the performance of the "standard" low band gap ternary contact can hardly be beaten, as was expected. However, the hybrid contact with a 15nm ternary layer on top comes close. There is negligible difference between the as-deposited behaviour of the pure quaternary contacts Q1 and Q2 and the hybrid contact H2 (with only 5nm of ternary on top). This seems to suggest that the interfacial reaction between the metal and the semiconductor at sputtering has an extension of several nm's (note that as-deposited the H2 is still relatively better than pure quaternary). So far everything is as expected, except that amazingly enough it suffices to add only a very thin capping of low band gap ternary material (15nm) to high band gap material to obtain contact properties close to pure ternary contacts. When the

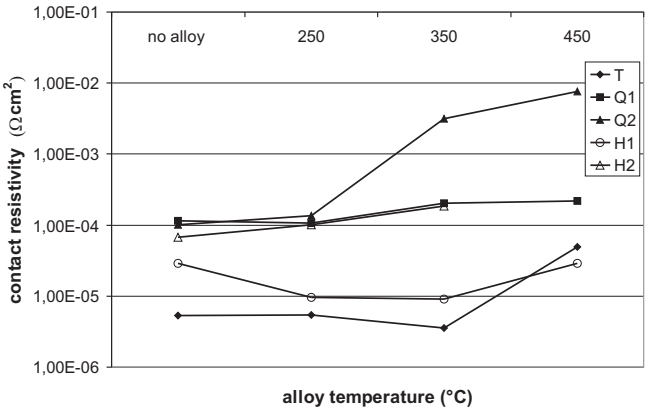


Fig. 4.14: CKR extracted specific contact resistivity of 50nm sputter-deposited $Co_{90}Fe_{10}$ on the five different InP-based semiconductor contact layers of Table 4.5 as a function of RTP temperature.

contact resistivity is studied as a function of alloy temperature, it immediately catches the eye that only the T and H1 contacts profit from a thermal treatment. However, it must be noted that it could well be that there is some decrease of ρ_c for Q1, Q2 and H2 at temperatures lower than 250°C. The same remark applies to the T and H1: the minimum ρ_c might be between 250 and 350°C. In any case, the observation that the H1 contact follows the same trend as the T contact as a function of temperature is another confirmation of the fact that a small ternary capping of the unfavourable high band gap material will make the contact behaviour of this hybrid layer be dominated by its ternary content (and thus that, even after alloying, the interfacial reaction between the metal and the semiconductor will not extend much beyond 5nm). For the group of "poorly" performing contacts (Q1, Q2 and H2) there is no beneficial "reaction" or diffusion from the constituents to improve the contact. This might be related to the lower InAs mole fraction in these contact layers. The formation of minimal band gap InAs layers at the interface due to thermal treatment has indeed been shown to be one of the mechanisms for ρ_c lowering [21]. The steep increase of the resistivity of the thin quaternary contact as a function of temperature is rather logical. The relative importance of the apparently deteriorating thermal treatment will be much higher in a thin quaternary layer, strongly reducing, among other things, the band-bending of the doped layer and hence the strength of the quantum-mechanical field-emission.
Table 4.6 summarizes the lowest observed contact resistivity for each contact layer with the corresponding alloy temperature and the relative error.

In conclusion it can be stated that even though the observed contact resistivities of these CoFe contacts are not world class when compared to the reported performances of Table 3.2, the performance of the H1 contact is very acceptable in view of the use of non-traditional contact metals and very thin low band gap

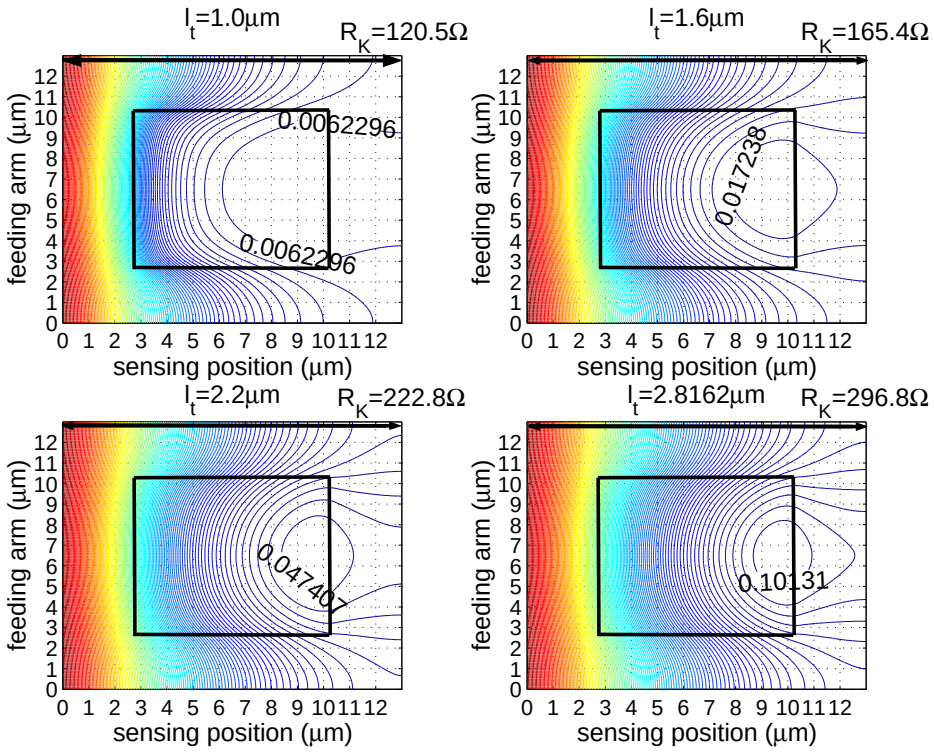


Fig. 4.15: Successive 2D voltage simulations for the extraction of the contact resistivity of the Q1 (see Table 4.5) contact layer alloyed at 250°C. The contour spacing is 5mV and the square indicates the position of the contact window. Inspection of the processed structures revealed that the actual contact window for all different geometries is systematically about 25% smaller than the mask design value, and this correction on d is necessary to obtain geometry-independent convergence of ρ_c. Here, the fitting simulations on a $7.5^2\mu m^2$ contact window on a mesa of $13^2\mu m^2$ are shown. The simulated resistance, obtained as the mean voltage along the sensing side of the mesa divided by 1mA (i.e. the injected current), converges to the measured Kelvin resistance ($\approx 296\Omega$) for a transfer length $l_t = 2.82\mu m$, corresponding to a $\rho_c = 1.07 \times 10^{-4}\Omega cm^2$. The sheet resistance measured via TLTR is 1348Ω/□. It can be clearly seen how at too low values for l_t, i.e. for an overestimation of the contact conductivity, the current crowding around the front side of the contact causes the simulated Kelvin resistance to be seriously underestimated. With increasing transfer length, the carriers have to travel an increasing distance in the diffusion layer beneath the contact, the crowding decreases, and hence the average voltage along the sensing arm increases towards the measured value.

material. In view of the simulations of section 3.3.2 the total contact area of the integrated device will be in the range of 5000–8000 μm^2. Therefore, while a thick

layer	ρ_c ($\times 10^{-8}\Omega\text{cm}^2$)	error (%)	RTP (°C)
T	417.2	45	350
Q1	11270	10	250
Q2	13140	11	no alloy
H1	967.8	14	350
H2	8193	18	no alloy

TABLE. 4.6: "Best" observed contact resistivity for each contact layer and apparent temperature at which this is obtained. Note that due to the limited number of temperatures used, an even better performance might be found at an intermediate temperature. The rather high error on the ternary contact resistivity is of course a consequence of its very low value.

transparent pure quaternary contact might still be acceptable causing "only" 1 à 2 Ω of total contact resistance, the results here and the design simulations done with a H1 hybrid contact make clear that the big decrease of the resistivity of at least an order of magnitude outweighs the negligible increase of the already high modal absorption. This makes the H1 contact structure the preferred and satisfactory behaving contacting scheme for the studied application.

4.2 Tensile-strained MQW active layer structures

As explained in section 3.2.3 for the realization of tensile-strained active layers with a band gap of 0.95 eV there are basically two material systems available, InAlGaAs and InGaAsP. While only the InGaAsP system allows for a flexible extension of the wavelength window towards the full C band (as can be seen from Fig. 3.20), it cannot rival with the InAlGaAs system when it comes to conduction band offset and electron confinement. Therefore, if we limit ourselves to the target application of an integrated isolator at the fibre dispersion minimum, the InAlGaAs material is expected to have far better gain performance than InGaAsP material. However, it was also pointed out that in view of the many interplaying parameters of tensile-strained active layers (strain of wells and barriers, well and barrier widths, barrier composition, number of wells,...) that it is quasi impossible to analytically model the performance of strained-layer amplifiers/lasers a priori. For these reasons, it has been chosen to take a pragmatical approach to the design of a suitable amplifying layer structure for the integrated SOA/isolator, namely by an ad hoc empirical characterization of the gain performance of a whole range of epitaxial grown layer structures. This section presents the results of this study. The material presented here is both "historically" and "pragmatically" organised. Work started on InGaAsP materials and only in a later stage, after the limitations of this material became clear, the growth of InAlGaAs material was considered. Next to this, it has to be remarked that the InGaAsP structures have been grown by GSMBE, while the InAlGaAs active layers were grown by MOVPE.

4.2.1 Tensile-strained InGaAsP active layers

4.2.1.1 Run 1: TM lasing

The purpose of the first run was to check at what strain levels selective TM amplification could be obtained. As extensively explained in section 3.2.2.2 the modification of the valence band structure in tensile-strained QW's is counteracted by the quantization effect of the valence subbands. As a result, at moderate tensile strains there is still a large amount of disadvantageous band mixing between the heavy and light hole subbands and possibly even a dominant heavy-hole character for the topmost valence subband. The strain level at which the first HH and LH subbands cross (and thus the dominant amplification changes from TE to TM polarization) is dependent on the QW material, on the thickness of the wells, and on the barrier material (as it determines the potential well problem). Thicker wells will obviously have a much lower strain at which the LH-HH crossing occurs. This is logical as in the limit of bulk material there is no quantum effect and the band crossing occurs at zero strain. This seems to suggest that for a given material system of well and barrier, there is a unique level of tensile strain one has to "cross" in order to get TM amplification. This is not correct. The emission wavelength at which this TM amplification should occur, also plays a role. The emission wavelength is determined by the well thickness and by the amount of strain one seeks to incorporate in the wells. Therefore, for a given desired wavelength and a fixed well thickness[19] there is a certain range of strains for which no TM amplification can occur, and this range changes with wavelength.

In a first run three MQW InGaAsP structures, with varying tensile strain and designed to emit at 1300nm, were grown by GSMBE. Table 4.7 summarizes the details of the MQW region.

Neither X-ray diffraction studies nor topoluminiscence or TEM studies revealed the presence of dislocations in these structures. After MOVPE regrowth of a thick (3μm) p→p$^+$ Zn-doped ($5 \times 10^{17} \rightarrow 1.4 \times 10^{18}$cm^{-3}) InP cladding and a 300nm p^{++} Zn-doped (3×10^{19}cm^{-3}) InGaAs contact layer, all three samples were processed into broad area Fabry-Perot lasers with a Ti/Au bilayer metal contact

[19] Obviously, things are slightly more complicated than that. To keep the emission wavelength constant at varying strains the well thickness has to change along. The thickness that is needed at zero strain for emission at a certain wavelength is determined by the lattice-matched composition that is chosen as a starting point. Fig. 3.20 shows that there is a degree of freedom here. Next to that there is a degree of freedom in how the strain is incorporated. Suppose that the constant band gap contour 0.8 eV is followed (between B and A in Fig.3.20). This implies that if bulk material was considered the energy separation between the CB and topmost VB remains 0.8eV along this line. In order to get to an emission wavelength of 1300nm (0.95eV) the material has to be given QW features. At zero strain this puts the HH band on top. But with increasing tensile strain, the HH band gets shifted downwards in the bulk material (while the LH band remains at 0.8eV because we stay on the 0.8eV contour of Fig. 3.20). As a result, even though a QW would again lift the HH band above the LH band for these moderate strains, the quantization effect mustn't be that strong anymore and the needed well thickness gets thicker than in the zero strain case. This trend continues for increasing strain, thus increasing well thickness for a constant wavelength emission, leading to a situation where gradually the HH-LH crossing is approached.

	composition	strain	thickness (nm)	number
S1	well (Q1.3)	-0.33%	8	6
	barrier (Q1.17)	+0.05%	10	5
S2	well (Q1.3)	-0.7%	10	6
	barrier (Q1.17)	+0.05%	10	5
S3	well (Q1.3)	-1.0%	12	3
	barrier (Q1.17)	+0.2%	22	2

TABLE. 4.7: Details of the MQW regions of the first run InGaAsP tensile-strained structures. All barriers were grown under a slight strain-compensating compressive strain.

(100μm) and cleaved into laser samples of varying lengths. Table 4.8 summarizes the laser characterization results obtained on 600 μm as-cleaved FP lasers. The lasers were measured at room temperature under pulsed current conditions.

sample	J_{th} (kA/cm^2)	efficiency= $\frac{\hbar\omega}{2q}\eta_d$ (W/A)	polarization
S1	1.5	0.141	TE
S2	1.946	0.151	TM
S3	0.880	0.16	TM

TABLE. 4.8: 600 μm FP laser characterization at 20°C. The very high threshold current density of the -0.7% strained lasers seems to suggest that this strain level is in the neighbourhood of the HH-LH crossing. Note however that this S2 sample has double the number of wells as the S3 sample. Still even in the very rough approximation (and slight overestimation) of a doubling of the current density at a doubling of the number of wells, the S2 sample still has a considerably larger threshold current density than the S3.

The increase of the threshold current density and the change of the polarization of the emitted laser light in going from strain levels of -0.33% to levels of -0.7%, followed by the reduction of the threshold at even higher tensile strains of -1.0%, is in accordance with the W-shaped J_{th} vs. ϵ-curve first observed by Thijs on strained InGaAs MQW lasers [22], and seems to indicate that the crossing of the LH-HH subbands for 1300nm tensile strained InGaAsP layers occurs between -0.33% and -0.7 % tensile strain (and rather close to the latter value). For a first run these results are already very promising. But there is room for improvement. As already indicated in section 3.2.3 the conduction band offset decreases drastically with strain in the InGaAsP material system. Here, for all layer structures a rather low band gap material has been used as a barrier. Bowers has calculated that with Q1.2 InGaAsP material the CBO to VBO ratio is about $\frac{\Delta E_c}{\Delta E_v} \approx 0.2 : 0.8$ [23]. While a low value for this ratio is typical of the InGaAsP system, a slight enhancement for the electron confinement can be expected by using higher band gap material for the barriers. This is examined in the next subsection.

In any case, the good results of the S3 samples obtained with a non-optimized structure motivated a more detailed study of their gain performance. For that purpose, both characterization methods described in section 3.2.4 have been carried out. Weakly index-guided Fabry-Perot ridge lasers were processed with widths varying between 1.5 and 9 μm. This was on the one hand necessary to decrease the absolute threshold current to levels sufficiently covered by the range of the used current source (so that sufficient measurement points could be obtained for efficiency vs. length measurements), and on the other hand to obtain transversally monomodal guiding structures to perform Hakki-Paoli measurements.
Internal losses of these active ridge waveguides structures have been determined by efficiency vs. length measurements. Very low optical losses of 6-7/cm have been obtained (see Fig. 4.16). However, the efficiency vs. length method is known to underestimate the internal optical loss. As explained in section 3.2.4 this is a result of the neglected length dependence of $\eta_{\text{int}}(L)$ and $\alpha_{\text{int}}(L)$ in this method, causing a linear fit to produce too optimistic values [24]. One could argue that obviously the extracted unrealistic value for the internal quantum efficiency ($\sim$60% where one normally expects a value close to 100%) artificially decreases the value obtained for the internal optical loss, as the slope of the fitted straight line must be multiplied with $\eta_{\text{int}}/\ln(1/R)$ to obtain α_{int}. The higher the value for η_{int}, the higher the optical loss. However, it was also shown by Piprek that in MQW devices carrier losses in the QW's due to non-radiative recombination mechanisms can have a large influence on the internal differential quantum efficiency, because of carrier non-uniformities among the QW's [25]. This non-uniformity makes it not unrealistic to observe internal efficiencies way below 100%, contrary to previous popular common assumption.
While the discussion of section 3.2.4 has made clear that it is in principle perfectly viable to use the length versus efficiency measurements to calculate the gain-current relationship, the assumption of a constant η_{int} and α_{int} will result in an underestimation of the modal gain and the associated current density, especially at higher gain levels. Therefore, polarization resolved spontaneous emission spectra below threshold have been measured and related to the modal gain via the Hakki-Paoli method. Fig. 4.17 shows the results of these measurements on a 3μm wide and 300μm long ridge for both polarizations and for a range of injected current densities. The facet reflectivity has been set at 28% for the TM mode and at 32% for the TE mode. It is clearly seen that the TE gain is strongly suppressed, indicating that the LH-HH splitting is considerable and band mixing is negligible.
If we take the approximative assumption that the long wavelength loss is wavelength-independent (recall the discussion of section 3.2.4.2) then the modal gain is found as $g_{\text{mod}}(\lambda) = g_{\text{net}}(\lambda) - \alpha_{\text{int}}$, with $\alpha_{\text{int}} = 15/cm$ for both TE and TM. With these values, the TM peak gain is indeed found to be approximatively a factor of 4 higher than the TE gain, as predicted by Table 3.4. Plotting the peak modal gain as a function of injected current density leads to gain-current data pairs to which a logarithmic function can be fitted. This is illustrated in Fig. 4.18.

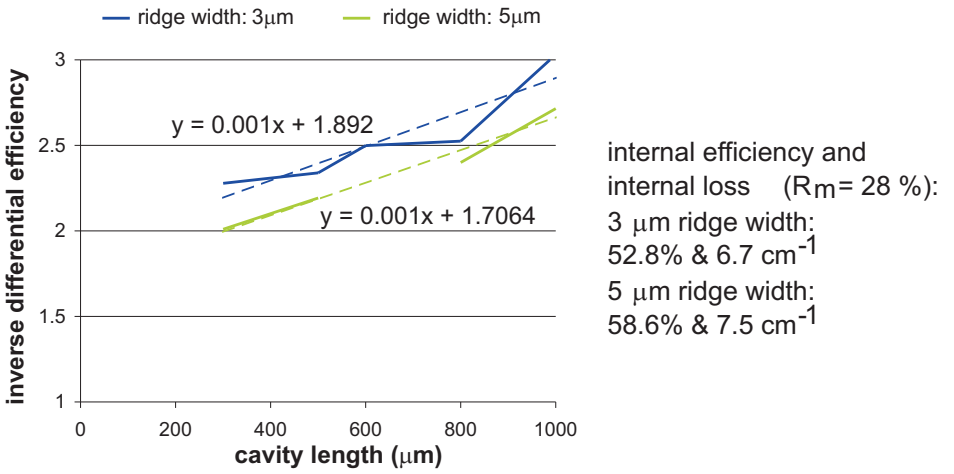


Fig. 4.16: Length versus inverse differential efficiency measurements on the S3 sample on stripes of different widths. Remark that even though the slope of the curves is apparently equal for both samples, the extracted internal loss is lower for the 3μm ridge due to the apparent lower internal efficiency. This result is counterintuitive and serves as an indication that these $(1/\eta_d, L)$-measurements should be used with caution, as they do not take the length dependence of the loss nor of the internal quantum efficiency into account.

The logarithmic gain-current relationship fitted to these data is

$$\begin{aligned} g_{\mathrm{mod,peak}} &= \beta J_0 \ln(\frac{J}{J_0}) \\ &= 51.48 \ln(\frac{J}{0.958}) \qquad (\beta = 53.71 \frac{\mathrm{cm}}{\mathrm{kA}}), \end{aligned} \tag{4.2}$$

with the current density expressed in units of $\mathrm{kA/cm^2}$. It was chosen to use a gain-current relationship in terms of the total current density and the total transparency current density instead of one in terms of the radiative currents. This gets down to incorporating the effects of the internal quantum efficiency into the transparency current density, $J_0 \rightarrow J_{\mathrm{rad.transp.}}/\eta_{\mathrm{int}}$. The reason for this is quite simple. As was explained above the internal quantum efficiency is strongly dependent on the precise geometry of the tested FP laser. It makes therefore little sense to extract out of these measurements a value for the transparency current density and treat it as a material parameter. Moreover, the HP method doesn't allow a measurement of η_{int}. As a very cautious estimate, one could use the (overestimated) η_{int} obtained in the $(L, 1/\eta_{\mathrm{int}})$-measurements. This would give a radiative transparency current density of about $450\mathrm{A/cm^2}$. However, it is very dangerous to extrapolate this transparency current density of $150\mathrm{A/cm^2}$ for a single well, as it is by no means certain that all three wells are fed by an equal current (throughout the whole range of injected current densities). The most acceptable prediction one can make about the gain of this -1% tensile-strained material is that as a whole it has a

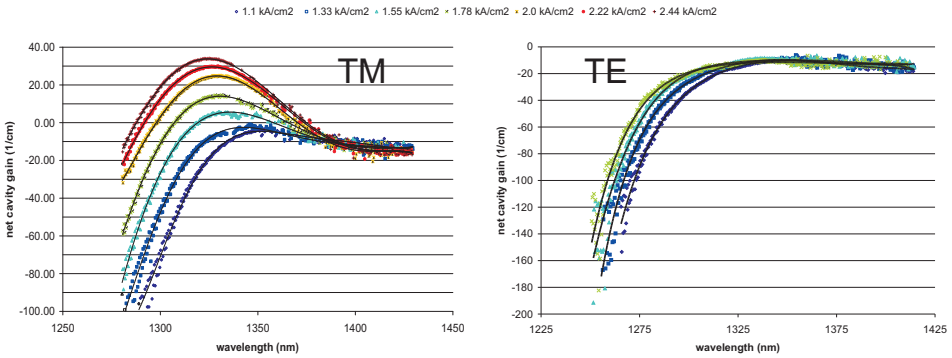


Fig. 4.17: Hakki-Paoli measurements on 300μm long S3 FP ridge laser ($w = 3\mu m$).

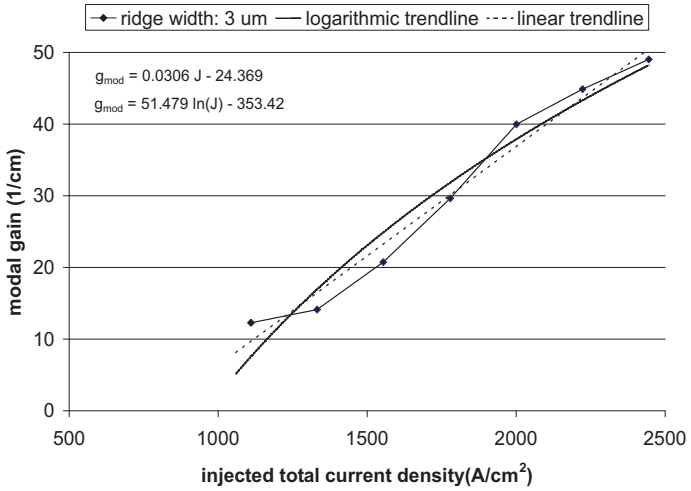


Fig. 4.18: Peak TM modal gain as a function of total injected current density obtained from Fig. 4.17. It is clear that only near transparency the linear trendline is acceptable.

transparency in the range of 400-500A/cm^2. What is more important is the value for the differential gain at transparency. While again this value can also not be considered as the intrinsic differential gain of the strained material, at least it is independent of the internal efficiency [20]. The value obtained here expresses that in the neigbourhood of transparency per kA/cm^2 increase of the injected current density the modal gain increases by approximately 50/cm. A typical value for standard compressive material for 1300nm operation is in the neigbourhood of 90/cm [26]. In view of the fact the here studied material is only a first run test,

[20] that is if it is accepted that the internal efficiency near transparency doesn't change considerably with increasing current.

this already indicates the potential of tensile strain and confirms the very general theoretical conclusions of the previous chapter.

4.2.1.2 Run 2: Material optimization

First batch The objectives of the second GSMBE run were to optimize the growth conditions and the band structure of the gain region by increasing the band gap of the barrier material and to explore the limits of strain by increasing the number of QW's. For that purpose two batches of tensile-strained MQW were grown. The first batch concentrated on the proven successful -1% strained wells, but with higher band gap barriers and an increasing number of wells. Table 4.9 gives the details.

	composition	strain	thickness (nm)
well	Q1.3	-1.1%	10
barrier	Q1.05	+0.3%	22

TABLE. 4.9: Details of the MQW regions of the first batch of the second run of InGaAsP tensile-strained structures.

As can be seen from comparison with Table 4.7 the thickness of the wells is decreased as it was noticed from the HP curves that the S3 sample led to a slightly too high emission wavelength. X-ray diffraction also revealed that the actual strain was a little higher than -1%. The most important difference is the composition of the barrier material, with a much higher band gap (Q1.05 as compared to Q1.17), which is expected to lead to much better electron confinement and higher values for the internal quantum efficiency. Also the strain compensation in these barriers is somewhat increased. With this buffer-well combination MQW's were grown with up to 15 wells (3, 6, 9, 12 and 15). X-ray diffraction didn't reveal any structural defects. Again, these structures were subjected to the same MOVPE regrowth as in the first run for FP laser processing.

Fig. 4.19 summarizes the laser characterization results obtained on 600μm long lasers as a function of number of wells. First of all it must be noted that the extreme high threshold current for 15 QW's is directly caused by a degradation of the material due to the high total strain of this sample[21]. Note that for this sample the total strain×thickness product is about 650Å–% (which is about a factor 3 higher than O'Reilly's rule of thumb).
As expected the increased optical confinement decreases the threshold current density per well for an increasing number of wells. However, the overall threshold current density already starts gradually increasing again from 6 wells on. It is

[21]Actually it was observed through TEM studies that it is not the strained material that relaxes but rather the InP cladding in which dislocations are generated. This has been found to be the result of the growth of InP on an undulating interface [28]. This undulating interface has been observed on the top of each of the strained wells, and is probably caused by a complex Ga/In exchange during the growth interruption used for changing the V-elements flow.

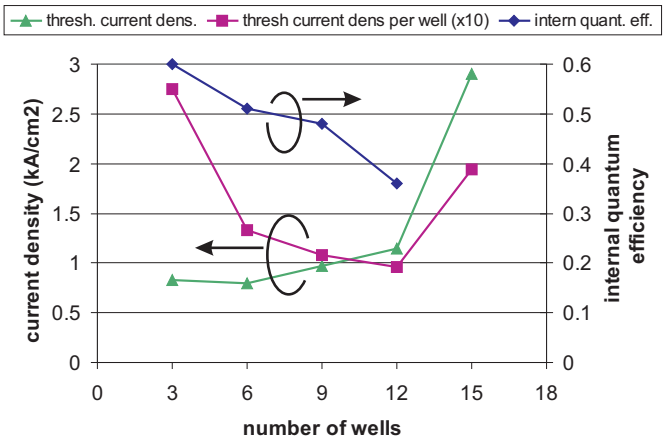


Fig. 4.19: Threshold current density on a 600µm long FP lasers as a function QW number (see Table 4.9). Note that the density per well is multiplied by a factor 10. Also plotted is the internal quantum efficiency determined via $(L, 1/\eta_{\text{int}})$*-measurements.*

unrealistic that the optical confinement is already saturating at this limited number of wells. More probably, for this specific laser length the required material gain per well is too low for an increasing well number and most of the provided current is lost in "pumping" all the wells to transparency. This argument is only correct if the internal loss α_{int} is effectively independent of the number of wells and if the wells are uniformly fed [27]. Inverse efficiency vs. length measurements confirmed that the optical loss is indeed quasi-independent of the number of wells, and equals approximately 7/cm. However, these measurements also revealed that the internal quantum efficiency decreases with increasing well number as can be seen from Fig. 4.19. This is rather disturbing, as it makes it difficult to pinpoint the cause of the increasing total threshold current density as a function of number of wells: it could as well be caused by the reduced internal efficiency as by the too high total transparency current density.

In any case, it is clear that for 3 QW's η_{int} isn't much increased with respect to the sample S3 of the first run, and that for 6 QW's it is already at the same level as for S3. This seems to indicate that increasing the band gap of the barrier didn't entirely achieve the desired effect. This might be understood as follows. If it is assumed that $\frac{\Delta E_c}{\Delta E_v}$ is quasi constant (recall that 0.2:0.8 is a good approximation), then an increase of the band gap of the barrier material will also further increase the already high VBO. In view of the lower mobility of the holes as compared to the electrons, this might cause an increasing asymmetric well filling by the holes for an increasing number of wells, causing a reduced efficiency of radiative recombination.

In any case it is clear that it makes little sense to use a number of wells much

beyond 9. Based on the evaluation of the threshold current density as a function of length (and in combination with the extracted optical loss), the gain current relationship for the samples with 3 and 6 QW's has been deduced. This is plotted in Fig. 4.20. It can be seen that within the accuracy of the measurement method the transparency current density and the differential gain parameters are in accordance. The transparency current density of the 6QW sample is only 34% higher instead of 100%, but this can probably be attributed to an overestimation of the internal efficiency of the 6QW sample. Most notable is that the differential gain parameter is almost twice the value obtained in the first run for -1% strained wells, and that the transparency current density is reduced by approximately the same amount. Thus even though the internal efficiency isn't significantly improved, the gain performance of the wells is much better, indicating that the material quality of the second run is enhanced. Next to that, with these values the tensile strained material is competitive with standard compressively strained material.

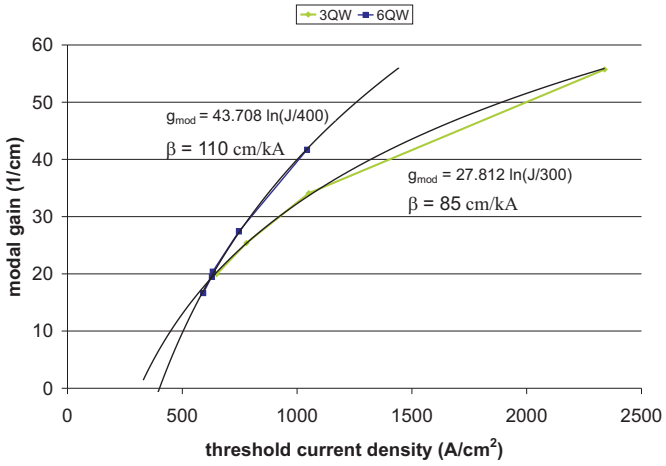


Fig. 4.20: Modal gain-current relationship obtained via threshold measurements of BA lasers with different lengths for the 3QW and 6QW samples (at room temperature).

Second batch The goal set forward for the second batch of material grown in the second GSMBE run was to investigate whether the so-far obtained best results for tensile-strained InGaAsP material [28] could be further improved by increasing the strain in the wells and by using an even stronger strain compensation in the barriers. Especially, the limited value for the internal quantum efficiency is some source of concern. Table 4.10 summarizes the details and the results of these optimization structures. They have been labelled O1 through O4.

Compared to the study of the first batch it seems that with increasing strain the MQW hardly improves its behaviour. On top of that, the O3 sample indicates that

	O1	O2	O3	O4
$\epsilon_{\text{well}}(\%)$ (Q1.3 10nm)	-1.25	-1.18	-1.5	-1.25
$\epsilon_{\text{barrier}}(\%)$ (Q1.05 22nm)	+0.3	+0.3	+0.6	+0.6
LI characteristics of 600μm BA FP lasers (at 20°C)				
$J_{\text{th}}(\text{kA/cm}^2)$	1.130	1.130	3.5	1.110
$\frac{h\nu}{2q}\eta_d(\text{W/A})$	0.160	0.160	0.110	0.180
gain characteristics (at 20°C) obtained through (J_{th}, L)-measurements and $(1/\eta_d, L)$-measurements				
α_{int}	7.3	9.0	43.6	8.8
η_{int}	0.505	0.569	0.761	0.515
$J_0(\text{kA/cm}^2)$	0.553	0.390	1.380	0.520
$\beta(\text{cm/kA})$	82	100	49	72

TABLE. 4.10: Layer structure details and room temperature characterization. All samples have 6 wells. The behaviour of the O3 sample is disastrous. But TEM inspection revealed that this highly strained sample contains dislocations in the InP cladding with the same origin as for the 15QW sample of the first batch.

a further increase of the strain leads to severe material degradation. Therefore it is safe to conclude that without an improvement of the interface quality between the wells and the barriers, the InGaAsP material system has temporarily reached its limits. Nevertheless, the reported values prove that tensile strain is at least competitive with standard compressive strain. However, the predictions of the previous chapter are not observed, namely that with increasing tensile strain such structures could outperform compressively strained structures due to the much stronger enhancement of the TM differential gain (as opposed to the enhancement of TE differential gain at compressive strain). But because the undulations of the well-barrier interface are not seen to cause dislocations in the MQW gain region itself, it seems that the "saturating" gain behaviour of this material as a function of strain indicates that the bad band offset ratio inherent of the InGaAsP material system is the limiting factor. In other words, within the limits posed by the critical thickness it seems that it is not possible to induce a type II behaviour for the heavy holes. As explained in section 3.2.2.3 such a behaviour would greatly increase the transition strengths of the CB1-LH1 transitions and thus the TM differential gain.

4.2.2 Tensile-strained InAlGaAs active layers

As mentioned already in section 3.2.3, to realise high-performance tensile-strained QW lasers, the InGaAlAs material system should be regarded as an interesting alternative to the GaInAsP system. This "aluminium system" has the advantage of a larger conduction band offset, which improves electron confinement and leads to higher T_0 temperature and a lower valence band offset, which improves carrier density uniformity in the quantum well structure. A typical value for the band

offset ratio $\frac{\Delta E_c}{\Delta E_v} = 0.72 : 0.28$ (as compared to $\frac{\Delta E_c}{\Delta E_v} = 0.4 : 0.6$ for InGaAsP), and this can vary considerably with strain. Most of the offset is thus in the conduction band. On top of that the axial deformation potential b is slightly larger than 2eV [29]. This is about 10–20% larger than for the InGaAsP system. The combination of both properties, a larger intrinsic LH-HH splitting and a band offset that is for 3/4 in the conduction band, make it clear that it will be very likely to induce type II behaviour for the HH holes! It is thus theoretically expected that tensile-strained InAlGaAs MQW's will have a much lower transparency current density, and a large improvement of the differential gain, because the negative influence of HH1 can be completely eliminated by the type II behaviour. As a result, the in-plane effective masses of the LH carriers will be much lower than in tensile-strained InGaAsP wells (where at the band edge the presence of the HH1 lowers the curvature of the LH1).

The first run was aimed at checking the above, while in the second run a further study of the gain performance as a function of increasing strain was performed. The next sections will detail the results obtained from both runs. It should be mentioned that this "aluminum"-system is grown by MOVPE.

4.2.2.1 Run 1: "pushing out" HH1...

In the first run three 6QW InAlGaAs structures were grown. The wells (10nm) have a tensile strain of -1.25% while the barriers (E_g = 1.12eV and 20nm) are compensating compressively strained over +0.6%. With these values all three structures should in principle be type II for the HH. They differed in the guiding layers around the MQW region. In view of the mentioned difference in band offsets between the InAlGaAs and the InGaAsP system, it is intuitively clear that a better hole injection into the wells can be expected from InAlGaAs guiding layers, while the lower CBO of the wells with respect to an InGaAsP guiding layer would ensure a better electron capture. As a direct consequence this would lead too a much better internal quantum efficiency. The three grown structures reflected three possible combinations[22]: symmetric guiding layers InGaAsP/InGaAsP and InAlGaAs/InAlGaAs or the possible electrically improved InAlGaAs/InGaAsP. Table 4.11 summarizes the characterization results obtained on 600μm BA as-cleaved FP lasers.

	P/P	Al/P	Al/Al
J_{th}(kA/cm^2)	0.56	0.53	0.70
$\frac{h\nu}{2q}\eta_d$(W/A)	0.22	0.22	0.18

TABLE. 4.11: The LI characterization results (at room temperature) of the first InAlGaAs run obtained on 600μm BA FP lasers.

[22]The guiding layers are of course lattice matched to InP. They are 40nm thick and have a band gap of 1.18eV

This first run immediately outperforms the best InGaAsP structures with the same number of wells (see Table 4.10). The threshold current density is reduced by a factor 2 and the external efficiency (optical power per current per facet) is increased by almost 40%! This is a clear indication that the type II behaviour for the HH decreases the transparency current density, and that it, in combination with the increased CBO, should increase the radiative recombination efficiency. $(J_{\rm th}, L)$-measurements and $(1/\eta_d, L)$-measurements on the Al/P sample confirmed this (Fig. 4.21). FP cavities down to 95μm have been cleaved for this purpose. The very short cavities were investigated to look for the limits of the gain behaviour. At 95μm the mirror losses by themselves would rise up to no less than 130/cm.

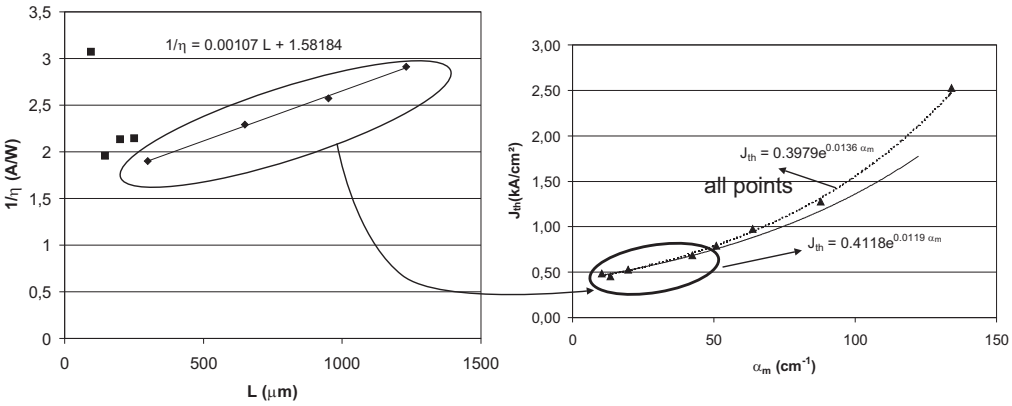


Fig. 4.21: $(J_{\rm th}, L)$-measurements and $(1/\eta_d, L)$-measurements at room temperature on a 600μm long BA FP laser processed out of the Al/P sample.

These very high losses are obviously investigated in view of the very high modal loss compensation that is needed in the integrated SOA/isolator device.
The $(1/\eta_d, L)$-measurements clearly reveal the observations made at the end of section 3.2.4.2. The linear fit is only valid for laser lengths down to 300μm. Shorter cavities see a sharp decrease of the efficiency. According to Piprek this is the combined result of the increasing internal loss and of the decreasing internal quantum efficiency as a result of the very high carrier densities required to obtain lasing [25]. The optimistic fit to solely the "long" cavities would predict a $\eta_{\rm int} = 0.7$ and a $\alpha_{\rm int} = 8.5{\rm cm}^{-1}$. But the $(J_{\rm th}, \alpha_{\rm m})$-graph makes clear that by only taking these points into account the actual threshold current density would be underestimated by almost 20% at higher mirror losses. It also indicates that in reality for all gain structures the theoretical logarithmic gain-current relationship of (4.2) is not valid throughout the entire current range. It would be actually better to use a relationship with a transparency current density that is gradually increasing as a function of the injected current density $J_0(J)$. This makes it difficult to extract a globally valid gain-current relationship out of <u>all</u> the measurement points

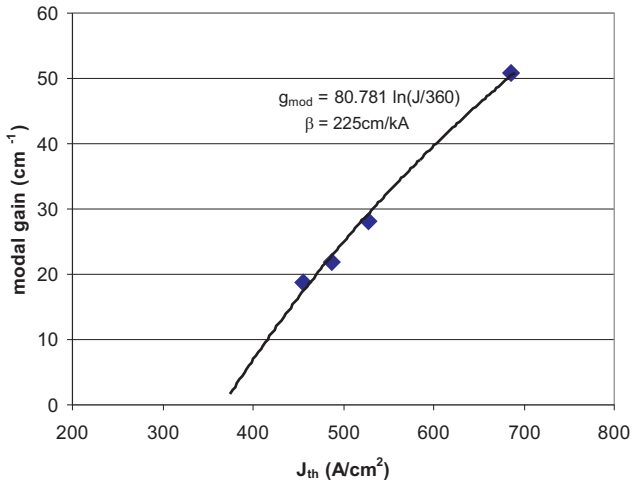


Fig. 4.22: "Optimistic" (room-temperature) gain-current relationship for the Al/P structure.

of Fig. 4.21. Note that the fit to all the points underestimates the threshold density in infinitely long lasers. Besides, it is by no means clear what value for the internal loss should be used for the shorter cavities. All one can say is that up to modal gains of 50cm^{-1} (i.e. mirror losses of a 300μm laser plus an internal loss of 8.5cm^{-1}), a constant internal loss of 8.5cm^{-1} can be used. The obtained gain-current relationship is plotted in Fig. 4.22. Even though a very optimistic attitude is taken, these gain characteristics are greatly improved in comparison to the best InGaAsP structures. Especially the differential gain factor is increased by more than a factor 2. The transparency current density is also significantly decreased, to a value around 60A/cm^2 per well (as compared to almost 90A/cm^2 per well for InGaAsP structures).
In any case, even without a precise knowledge of the internal losses at higher carrier densities the above results reveal that at a current density of merely 2.5kA/cm^2 a modal gain of approximately 150/cm is obtained (if an $\alpha_{\rm int}$ of about 20cm^{-1} is assumed at short cavity lengths).

A similar experiment on the best InGaAsP sample (O2) cleaved into a BA FP laser of 100μm length gave a threshold current density of 5.45kA/cm^2. Still, one has to be careful in overenthusiastically transposing this result to any other arbitrary optical gain device. The result obtained is greatly dependent on the internal efficiency and thus among others on the precise geometry. It should for instance not be forgotten that in these gain-guided BA lasers the lateral carrier leakage can in general be neglected, while in the typically used weakly index-guided ridge waveguide the relative importance of lateral carrier diffusion and the accompanying carrier losses is much stronger. In strongly index-guided structures (such as

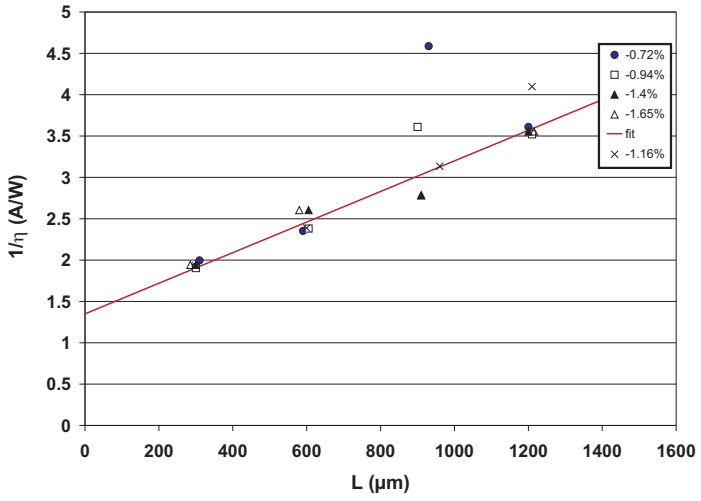
deeply etched ridge waveguides and buried heterostructures with current blocking layers) there are other efficiency "killer" mechanisms that are not present in BA lasers. Surface recombination being the most important of these. Nevertheless, a modal gain at room temperature of almost 150cm^{-1} at only 2.5kA/cm^2 is of course a very positive result, even in a gain-guided BA laser.

4.2.2.2 Run 2: strain study

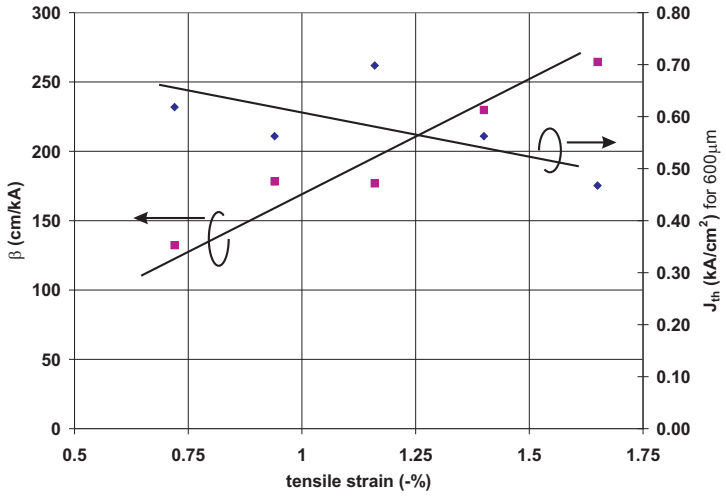
The success of the first InAlGaAs run spurs the motivation to look for a further increase of the gain performance by a more in-depth study of the influence of the well tensile strain. For that purpose 5 structures were grown with well strains increasing as -0.72%, -0.94%, -1.16%, -1.40%, and -1.65%. The other geometric and compositional properties of the MQW and guiding layers are left untouched. No material degradation has been observed as opposed to the -1.5%-strained InGaAsP structures where there was a clear interface problem. The critical thickness "barrier" has for the moment not been reached in the InAlGaAs material. Note that in the highest strained structures the strain–thickness product is almost 300Å–% (i.e. almost twice as high as the O'Reilly rule of thumb for InGaAsP). The results on these structures are summarized in Fig. 4.23.

It is first of all observed that all structures have a comparable inverse efficiency versus length behaviour, indicating that a possible gain improvement as a function of well strain is not coming from an increase in internal quantum efficiency or a decrease in internal optical loss. Actually, quasi strain-independent values of approximately 0.75 and 15cm^{-1} are found for resp. η_{int} and α_{int}. The high value for the internal quantum efficiency confirms that the increased CBO indeed improves the radiative recombination process, but its quasi-independence on the strain indicates that the theoretically predicted reduction of Auger recombination with strain is weaker than expected. The same argument applies for the internal loss. Even though it is known that the intervalence absorption should drop with increasing strain, that behaviour is not observed in Fig. 4.23. However it is likely that the observed internal loss is dominated by another mechanism that is not influenced by strain, such as free carrier absorption in the cladding. After all a loss of only 8cm^{-1} is very low.
The most profound impact of the increased strain is observed in Fig. 4.23(b). The threshold current density for 600 μm long lasers slightly but gradually decreases with increasing strain, indicating that there is both a reduction of the transparency carrier density due the decreased hole mass and a reduction of the non-radiative carrier losses. In parallel, the differential gain is enormously enhanced (almost by a factor 2) when moving from a -0.72% to a -1.65% strain level. Such an increase can't be solely attributed to a decrease of the transparency current density. Besides, the transparency current density for all five structures was measured to be quasi-constant at around 350 A/cm^2.[23] Therefore the only mechanism left that

[23]This is not so illogical, as it was calculated that already at the lowest strain level used (-0.72%) the HH carriers are very close to cut-off (type II). As the LH subbands do not couple among themselves and as there is no HH band left to perturb the LH1 in-plane effective mass,



(a)

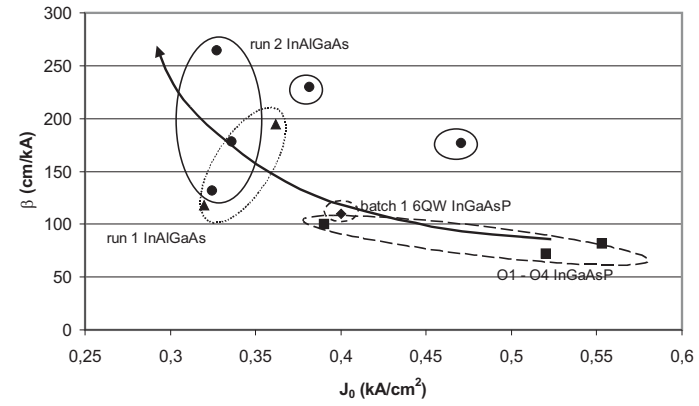


(b)

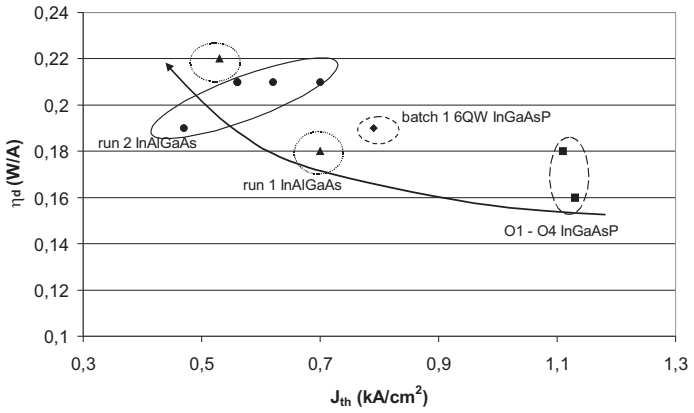
Fig. 4.23: Results of the in-depth study of the gain performance as a function of strain for the 1.3µm InAlGaAs material system.(a): collected $(1/\eta_d, L)$-measurements for all five structures;(b): Threshold current density for 600µm BA lasers and differential gain parameter as a function of strain.

can explain this β-increase is an increase of the TM C1-LH1 transition strength as a function of strain. This is precisely what has been theoretically assessed in section 3.2.2.3. The experimental observation of the increase of the TM differential

this mass immediately "saturates" at a beneficial low value.



(a)



(b)

Fig. 4.24: Room temperature comparison of the performance of all studied 6QW devices. (a): Differential gain versus transparency current density; (b): External efficiency versus threshold current density on 600µm long BA FP lasers.

gain with strain serves as the most convincing evidence that tensile strain increases the fractional p_z orbital-like character of the LH Bloch function to levels of almost of 80-85% [30].

4.2.3 Conclusion

In conclusion, the move from MBE grown InGaAsP tensile strain materials towards InGaAlAs MOVPE grown materials is shown to lead to very promising and exciting results. It is clear that the InAlGaAs system is without a doubt the pref-

erence material for the integrated SOA/isolator. It outperforms the traditional InGaAsP system in all areas (threshold current density, transparency current density, internal quantum efficiency, and differential gain), and this sometimes by more than 100%! Actually, tensile strained InAlGaAs MQW's are still relatively unexplored, but a comparison with the more conventional TE gain compresively strained InAlGaAs reveals that the here presented results are already competitive with this more developed approach. Moreover, the trendlines in Fig. 4.23 indicate that it is not unrealistic to expect further improvements with even higher strains. The observed progress in the developed TM gain material is summarized Fig. 4.24. The positive trend towards lower transparency current density and higher differential TM gain, and lower threshold current density and higher external efficiency (for 600 μm long lasers) is clearly observed. Note that this comparison of course only makes sense if the devices have the same number of wells (and the same lengths). The reader will have noticed that throughout this section the reported results have been mainly obtained on 6QW devices and comparison of absolute laser characteristics was on 600 μm long lasers.

4.3 Conclusion

In this chapter the experimental behaviour of the building blocks of the integrated SOA/isolator has been assessed. The obtained results prove the feasibility of the novel isolator concept, assuming that the technological assembly of the building blocks into a fully integrated device does not cause "interference" between the observed properties. There is no reason to believe that this will be the case, though as pointed out the magnetic anisotropy of the ferromagnetic contact stripes and the internal quantum efficiency of the tensile strained MQW region are dependent on the precise device geometry.

The measured magneto-optical gyrotropy of the ferromagnetic CoFe alloy system at 1300nm is observed to be sufficiently strong, as confirmed by the design simulations of the previous chapter. And it has been indicated that optimization of the alloy composition can increase the non-reciprocal behaviour by almost 75%. From a magneto-optic viewpoint there seem to be no competitors for the equi-atomic $Co_{50}Fe_{50}$. The recently theoretically suggested use of ferromagnetic-semiconductor composites, having comparable magneto-optic strength as the bulk MO metal but with strongly reduced optical absorption, might prove to be a revolutionary novel approach to inducing a non-reciprocal absorption shift in SOA's. However, experimental proof is still awaited.
The magnetic anisotropy of patterned CoFe stripes with very high aspect ratios on InP-based semiconductors is limited but acceptable, though care has to be taken in controlling the physical width of the metal contacts. And still it is advisable to magnetically shield the final device from any stray fields in view of the limited coercive field strength of only 25 Oe.
It has been proven possible to achieve good Ohmic contacts with CoFe on quasi-transparent high band gap InGaAsP semiconductors by using a hybrid layout

for the semiconductor contacting layer (with only a small fraction of absorptive ternary material). This contacting scheme is novel, and its development is of particular importance for the studied application, as it allows to largely avoid shielding of the non-reciprocity region. A good though not-world record contact resistivity in the order of $10^{-5}\Omega\text{cm}^2$ is obtained.
Finally, magnetic AGFM measurements indicated a quasi-negligible magnetic dead layer formation at the interface. This seems to suggest that the magneto-optic strength is largely retained at the interface. However, precise MO measurements to actually prove this were very difficult to perform.

Very exciting results have been obtained on InAlGaAs tensile-strained TM gain material. Indications are that this material might even outperform the best standard compressively strained TE gain materials, and it definitely beats tensile-strained InGaAsP. A TM differential gain parameter up to 250 cm/kA at transparency current densities as low as 60A/cm^2 per well have been obtained at room temperature. Though, the measurements also seem to suggest that it makes little sense to increase the optical confinement by increasing the number of wells much beyond 6–9 wells. The focus has been on the determination of the modal gain in BA lasers and not on the actual material gain per well. It has been explained earlier that the simple division of the modal gain by the confinement factor in order to obtain the material gain, only makes sense at moderate current injection and gain levels. As this assumes a homogeneous filling of the wells one has to be careful in extrapolating such predictions to higher current densities in view of the probable carrier redistribution among the wells. In any case, a rough estimate seems to indicate material gains of the order of 1500/cm at a current density of 500A/cm^2 per well. Next to that, measurements on very short lasers also prove that at these gain levels the internal efficiency starts decreasing very sharply. Therefore, one has to be very careful in using the obtained logarithmic gain-current relationships at even higher current densities and gains. The simulations of the previous chapter prove that in view of the obtained simulated values for the forward modal loss a material gain of 1500/cm is somewhat on the low side but still acceptable. Nevertheless, already now it is safe to state that it will be very difficult to operate the SOA/isolator at current levels below 500mA and still obtain isolation ratios of the order of 20-25dB.

References

[1] H. Shimizu and Y. Nakano, "First Demonstration of TE Mode Nonreciprocal Propagation in a Semiconductor Active Waveguide for an Integratable Optical Isolator." Post-Deadline paper PD1 at *19th IEEE International Semiconductor Laser Conference,* Matsue-shi, Japan, 21–25 September 2004.

[2] K. H. J. Buschow, P. G. van Engen, and R. Jongebreur, "Magneto-optical properties of metallic ferromagnetic materials." *J. Magn. Magn. Mater.,* vol. 38(1), pp. 1–22, 1983.

[3] H. Shimizu and M. Tanaka, "Design of semiconductor-waveguide-type optical isolators using the nonreciprocal loss/gain in the magneto-optical waveguides having MnAs nanoclusters." *Appl. Phys. Lett.,* vol. 81(27), pp. 5246-5248, 2002.

[4] J. M. Hammer, G. A. Evans, G. Ozgur, and J. K. Butler, "Isolators, Polarizers, and Other Optical Waveguide Devices Using a Resonant-Layer Effect." *J. Lightwave Technol.,* vol. 22(7), pp. 1754–1763, 2004.

[5] K. Baba, F. Takase, and M. Miyagi, "Ferromagnetic particle composite polymer films for glass and semiconductor substrates." *Opt. Commun.,* vol. 139(1), pp. 3538, 1997.

[6] J. De Boeck, R. Oesterholt, H. Bender, A. Van Esch, C. Bruynseraede, C. Van Hoof, and S. Borghs, "Controlled formation of nanoscale MnAs magnetic clusters in GaAs." *J. Magn. Magn. Mater.,* vol. 156(1-3), pp. 148–150, 1996.

[7] T. Zaman, X. Guo, and R. J. Ram, "Faraday rotation in semiconductors for photonic integration." in the Proceedings of *the 2004 Conference on Lasers and Electro-Optics (CLEO 2004),* p. CFH1, San Francisco, USA, 2004.

[8] H. Boeve, Ph. D. Thesis, Katholieke Universiteit Leuven, Belgium, 2001.

[9] Š. Višňovský, R. Lopuňík, M. Bauer, J. Bok, J. Fassbender, and B. Hillebrands, "Magnetooptic ellipsometry in multilayers at arbitrary magnetization." *Optics Express,* vol. 9(3), pp. 121–135, 2001.

[10] C.-Y. You, and S.-C. Shin, "Generalized analytic formulae for magneto-optical Kerr effects." *J. Appl. Phys.,* vol. 84(1), 541–546, 1998.

[11] R. Atkinson, and P. H. Lissberger, "Sign conventions in magneto-optical calculations and measurements." *Appl. Opt.,* vol. 31(28), pp. 6076–6081, 1992.

[12] H. Raether, *Surface Plasmons – on smooth and rough surfaces and on gratings.* Springer Tracts in Modern Physics, vol. 111 (Springer-Verlag, Berlin, Germany, 1988).

[13] A. Berger, and M. R. Pufall, "Generalized magneto-optical ellipsometry." *Appl. Phys. Lett.,* vol. 71(7), pp. 965–967, 1997.

[14] R. M. A. Azzam, and N.M. Bashara, *Ellipsometry and Polarized Light.* (North-Holland Physics Publishing, Elsevier Science Publishers, Amsterdam, The Netherlands, 1987).

[15] G. S. Krinchik, and V. A. Artemjev, "Magneto-optical properties of Ni, Co, and Fe in the ultraviolet visible, and infrared parts of the spectrum." *Sov. Phys.-JETP,* vol. 26(6), pp. 1080–1085, 1968.

[16] R. Atkinson, "Single layer equivalence of magneto-optic multilayers for normal incidence (I)." *J. Magn. Magn. Mat.,* vol. 95(1), pp. 61–68, 1991.

[17] R. Atkinson, "Single layer equivalence of magneto-optic multilayers for oblique incidence (II)." *J. Magn. Magn. Mat.,* vol. 95(1), pp. 69–75, 1991.

[18] R. Fletcher Chapter 4 "Conjugate Direction Methods" in *Practical Methods of Optimization,* (John Wiley & Sons, USA, New York, 2000).

[19] W.M. Loh, S.E. Swirhun, T.A. Schreyer, R.M. Swanson, and K.C. Saraswat, "Modeling and Measurement of Contact Resistances." *IEEE Trans. on Electron Devices,* vol. ED-34(3), pp. 512–523, 1987.

[20] W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "Development of a ferromagnetic transition metal Ohmic p-type contact for a novel integrated optical isolator conept." submitted to *Electron. Lett.,* (February 2005).

[21] A. Katz, W. C. Dautremont-Smith, S. N. G. Chu, P. M. Thomas, L. A. Koszi, J. W. Lee, V. G. Riggs, R. L. Brown, S. G. Napholtz, J. L. Zilko, and A. Lahav "Pt/Ti/p-$In_{0.53}Ga_{0.47}As$ low-resistance nonalloyed ohmic contact formed by rapid thermal processing." *Appl. Phys. Lett.,* vol. 54(23), pp. 2306–2308, 1989.

[22] P. J. A. Thijs, L. F. Tiemeijer, J. J. M. Binsma, and T. Van Dongen, "Improved performance of 1.5 μm wavelength tensile and compressivele strained InGaAs-InGaAsP quantum well lasers." in *Tech. Digest ECOC/IOOC,* Paris, France, Sept. 1991, vol. 2, pp. 31–38.

[23] T. Ishikawa, and J. E. Bowers, "Band lineup and in-plane effective mass of InGaAsP or InGaAlAs on InP strained-layer quantum well." *IEEE J. Quantum Electron.,* vol. 30(2), pp. 562–570,1994.

[24] J. Piprek, P. Abraham and J. E. Bowers, "Cavity Length Effects on Internal Loss and Quantum Efficiency of Multi-Quantum-Well Lasers." *IEEE J. Sel. Topics Quantum Electron.,* vol. 5(3), pp. 643–647, 1999.

[25] J. Piprek, P. Abraham and J. E. Bowers, "Carrier Non-Uniformity Effects on the Internal Efficiency of Multi-Quantum Well Lasers." *Appl. Phys. Lett.,* vol. 74(4), pp. 489–491, 1999.

[26] O. Gauthier-Lafaye, V. Colson, J. Py, B. Thedrez, and J.-L. Gentner, "High temperature 10 Gbit/s directly modulated 1.3 μm DFB lasers using InAsP/InGaAsP materials." *Electron. Lett.,* vol. 38(6), pp. 275–277, 2002.

[27] P. W. A. Mc Ilroy, A. Kurobe, and Y. Uematsu, "Analysis and application of theoretical gain curves to the design of multi-quantum-well lasers." *IEEE J. Quantum Electron.,* vol. QE-21(12), pp. 1958–1963, 1985.

[28] F. Lelarge, B. Dagens, C. Cuisin, O. Le Gouezigou, G. Patriarche, W. Van Parys, M. Vanwolleghem, R. Baets, and J.-L.Gentner, "GSMBE growth of GaInAsP/InP 1.3 μm-TM-lasers for monolithic integration with optical waveguide isolator." *J. Cryst. Growth,* vol. 278(1–4), pp. 709–713, 2005.

[29] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for IIIV compound semiconductors and their alloys." *J. Appl. Phys.,* vol. 89(11), pp. 5815–5875, 2001.

[30] G. Jones, and E. P. O'Reilly, "Improved performance of long-wavelength strained bulk-like semiconductor lasers." *IEEE J. Quantum Electron.,* vol. 29(5), pp. 1344–1354, 1993.

5

Experimental characterization of the SOA/isolator

THIS chapter is the culmination of the theoretical and experimental results of the previous chapters. Experimental proof will be presented here of the first worldwide assessment of the novel integrated isolator concept introduced in Chapter 1 and originally proposed by Nakano [1].

This chapter is divided into two sections. In the first section, the experimental techniques developed for the characterization of the non-reciprocity will be elaborated. The focus will be mainly on the analysis of the non-reciprocal amplified spontaneous emission generated by the ferromagnetic-metal-contacted-SOA. The observation of non-reciprocal guiding of externally injected light is after all only possible if the SOA can be biased to reach transparency. And the theoretical and experimental observations of the previous chapter have made clear that this is not evident beforehand. If the SOA can be biased up to and above transparency so that not only the cavity losses but also the facet mirror losses can be compensated, then non-reciprocal laser emission can be observed. It will be shown that in such a case the analysis of the non-reciprocity is very straightforward and can be derived directly as a function of the difference in differential efficiency. Finally, with a transparent device non-reciprocal transmission of guided light can be observed. These latter measurements are the ultimate goal, as they allow to assess the device characteristics that are of the utmost importance for integration of the isolator with a semiconductor laser. These are among others the wavelength transparency, the influence on the extinction ratio and the optical-signal-to-noise ratio (OSNR), and the extension of the coherence collapse threshold of the laser under optical feedback.

In the second section these methods will be applied to two developed generations of SOA/isolator demonstrators. Throughout this section attention will be paid to the technological challenges and difficulties encountered in “assembling” the building blocks of the SOA/isolator.

5.1 Characterization methods of the non-reciprocal transverse MO Kerr absorption in a SOA

5.1.1 Experimental bench

An experimental optical non-reciprocal bench for the characterization of the transverse Kerr effect-induced absorption shift must be able to detect optical power differences depending on the propagation direction through the device under test. A very naive way of achieving this, would be by trying to detect these differences by measuring the optical power at both sides of the device. This obviously makes no sense as the detected light power greatly depends on the detection efficiency. By measuring at two physically different locations it will be very hard or even quasi-impossible to realize the same experimental measurement conditions at both detection sides. Such an approach would fail any accuracy test for the simple reason that one is trying to measure a relative effect (*the non-reciprocal optical power change*) in an absolute manner (*by measuring the absolute power in opposite directions*).
As the non-reciprocity is entirely determined by the magneto-optic effect, and because, as explained in Chapter 1, the magneto-optic effects are directly proportional to the gyration vector $\mathbf{g}$, changing the direction of $\mathbf{g}$ while monitoring the same observation side of the device, is equivalent to switching between the observation of the properties of forward and backward guided light. Now, because of Onsager's principle $\epsilon_{ik}(\mathbf{M}) = \epsilon_{ki}(-\mathbf{M})$ and because the gyrotropic contributions to the dielectric tensor are by definition proportional to the gyration vector and are fundamentally antisymmetric[1], the gyration vector $\mathbf{g}$ changes sign when the magnetization vector $\mathbf{M}$ changes sign. Hence, setup-independent detection of optical non-reciprocity can be achieved by monitoring relative changes of the optical properties of the guided light while changing the direction of the magnetization. Note that this is of course also clear intuitively, but still the above argument is the only fundamental physical proof.

With this argumentation, the basic configuration for an experimental bench for the characterization of the non-reciprocal transverse MO Kerr effect is the one depicted in Fig. 5.1. The SOA is mounted between the pole shoes of an electromagnet, and the power emitted by one facet is sent through a polarizer, aligned along the TM polarization, and detected by an optical power meter. It is not important whether the SOA itself is processed with a tilted or a straight facet, or if its facets are coated with an AR coating. The characterization methods of the subsequent sections will treat both situations where the emission generated inside the cavity is or is not subjected to resonant reflections. The current probe needles allow both DC and pulsed current biasing of the device. All equipment, including the electromagnet, has been automated in a Labview environment. The

[1]This is not difficult to prove. Group theory predicts that every tensor can be reduced in a symmetric and an anti-symmetric part. Because gyrotropy implies non-reciprocity and because non-reciprocity implies a non-symmetric dielectric tensor, it follows that the gyrotropic contribution to $\underline{\epsilon}$ cannot have a symmetric part and hence must be entirely antisymmetric!

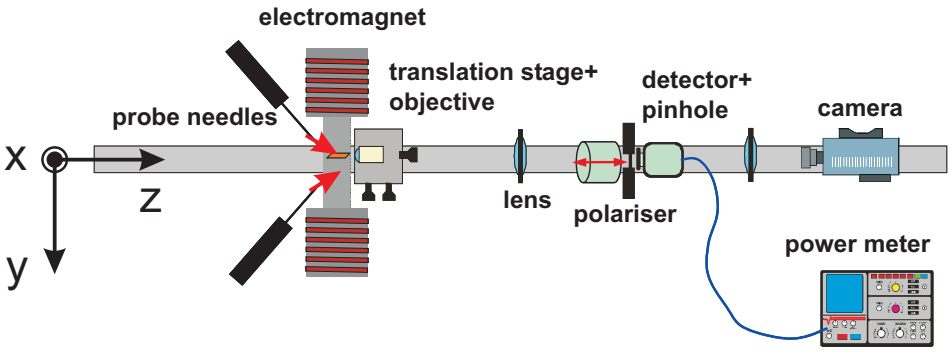


Fig. 5.1: Basic setup for the characterization of the transverse Kerr effect induced optical non-reciprocity in a SOA. The lab coordinate frame is used to relate forward and backward propagating light to the direction of the magnetization. With this lab frame as a reference, forward propagating light will be detected when the magnetic field is applied along $+y$, and vice versa. Note that this doesn't imply that forward propagating light will have by definition a lower loss. That depends on the precise value of the real and imaginary parts of $\boldsymbol{g}$ when characterized in this precise frame (with $\boldsymbol{M}$ pointing along $+y$).

electromagnet itself is capable of delivering DC magnetic inductions of up to 1T (which is equivalent in the old CGS units to a magnetic field of 10000 Oe) by varying the coil current up to 5A.[2] Of course this all depends on the width of the air gap between the pole shoes. Therefore, after fixing the pole shoes at a practical distance (for mounting the sample holder), the electromagnet has been calibrated by absolute magnetic field measurements using a calibrated Gaussmeter. This calibration lookup table has been implemented in the Labview environment, allowing a sweep of the magnetic field by setting the ranges of the desired magnetic field instead of the needed coil voltage. Switching the polarity of the electromagnet is possible by the use of a self-made relais switchbox, which is RS232 controllable. In this way there was no need for an expensive dual polarity voltage power source. Finally, the SOA temperature is controlled by a 10W Peltier TEC cooler beneath the sample holder, and a thermistor is fitted into a drilled hole within the sample holder and very close to the sample.

With this setup the most direct qualitative proof of the non-reciprocal absorption shift is obtained by measuring the ASE emission as a function of the applied magnetic field. As the above argumentation proves that the gyration vector is an odd, linear function[3] of the magnetization, detecting the TM optical power while varying the applied magnetic field from $-H_{\mathrm{sat}}$ to $+H_{\mathrm{sat}}$ and back (with H_{sat} the field needed to saturate ferromagnetic stripe contact in the transverse direction), must result in an emulation of the magnetic hysteresis of the contact stripe. The

[2] Actually in this setup the applied coil voltage is varied, but that is of course equivalent to varying the coil current using the knowledge of the coil resistance.

[3] third order nonlinear magnetooptic effects have never been observed up till now, and can be considered to be if not nonexistant, undetectably small.

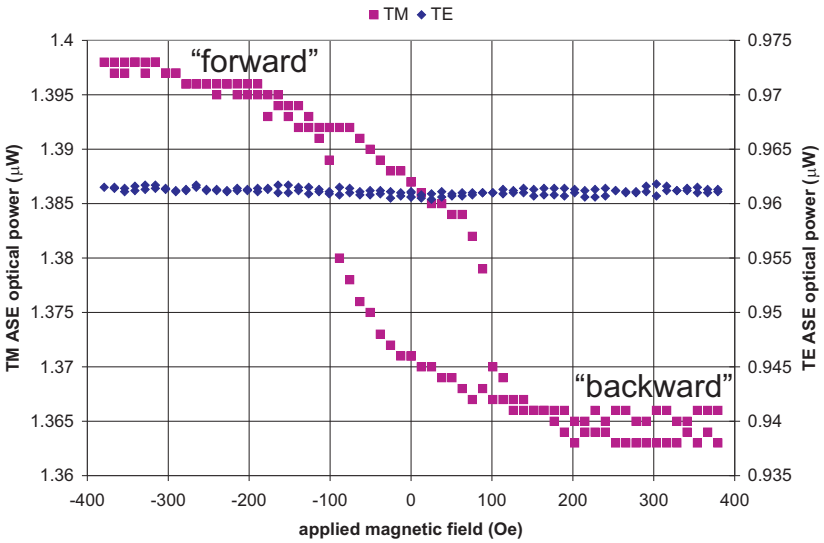


Fig. 5.2: Example of ASE hysteresis measurement on a first generation $Co_{90}Fe_{10}/InGaAsP$ device (3QW, $t_{\mathrm{InP}} = 250nm$, $L = 800\mu m$, $w = 2.5\mu m$, $J = 5kA/cm^2$) [2].

absence of any spurious magneto-mechanical effects, possibly caused by the presence of magnetic materials in the various mechanical parts of the setup close the magnetized space, can be checked by measuring the TE ASE output of the SOA as a function of the magnetic field. As the transverse Kerr effect has no influence on the TE polarized mode of the optical amplifier the detected optical power should be flat. Fig. 5.2 shows an example of such an ASE hysteresis measurement. Actually, this measurement was the first ever observation of the presence of the desired non-reciprocal effect in the novel integrated SOA/isolator.

Such a qualitative measurement is the first and most straightforward way of proving the presence of the desired non-reciprocal effect. Its consistency can be checked by comparing the obtained hysteresis curves with the AGFM magnetic anisotropy measurements of section 4.1.2. But this is not where its applicability ends. At first sight, such an ASE hysteresis measurement doesn't seem to allow to extract any quantitative information about the non-reciprocal absorption shift. This is not the case. Comparing the levels of the "forward" TM ASE power to that of the "backward" TM ASE power yields a relative measure of the nonreciprocal effect, independent of the precise measurement conditions. The understanding that this ratio could be used for a quantitative characterization of $\Delta\alpha$ came after having performed another qualitative experimental proof of the presence of the non-reciprocal absorption loss. This was the measurement of the TM ASE hysteresis curves as a function of the injection current density into the device. This

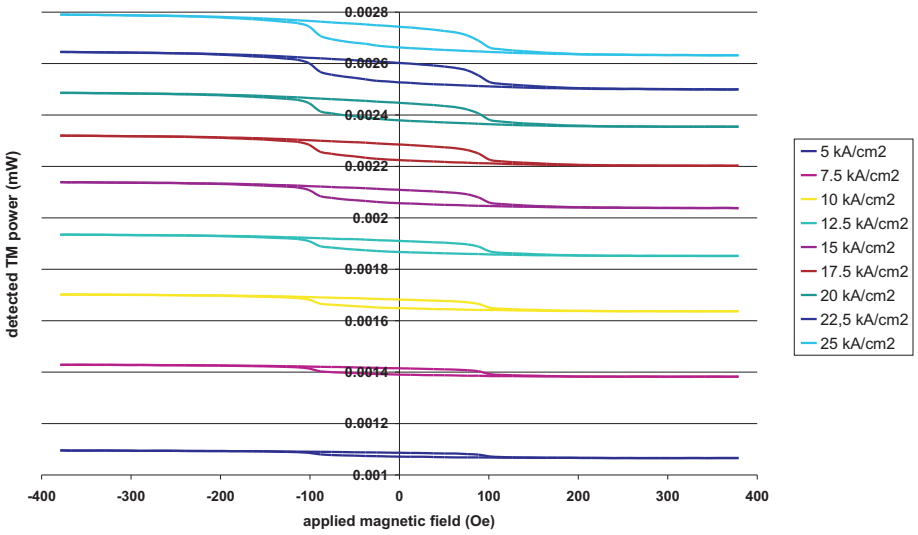


Fig. 5.3: TM ASE hysteresis measurements on the same device as in Fig. 5.2 for varying injection current densities [3].

is plotted in Fig. 5.3. The observation that the global intensity increases (in a saturating manner) with current density was not the big surprise of course. However, the fact that also the "amplitude" of the hysteresis increases with current density was less expected. It took not long to realize that this is a direct consequence of the increased contribution to the total detected power of ASE generated deeper inside the SOA cavity. This light has travelled a larger distance inside the non-reciprocal SOA and hence increases the total observed optical non-reciprocity. With these measurements one has several measurement points for the same effect at one's disposal. Therefore, measuring the ASE ratio as a function of injected current density allows, when combined with an appropriate model for the emitted ASE, a fitting procedure to extract the non-reciprocal absorption shift [4]. The next subsection will discuss these models.

5.1.2 Non-reciprocal amplified spontaneous emission

The treatment of the coupling of spontaneous emission to the waveguide modes and the propagation of these inside an amplifying cavity can be conducted on several levels of complexity depending on the effects one wishes to study. The treatment of ASE inside a SOA has received lots of attention. The seminal work was by Henry [5] deriving a general theory on spontaneous emission noise in both open and closed resonators, and by Haus [6] treating the coupling of spontaneous emission in multimode gain-guided cavities (but neglecting possible gain saturation effects due to spatial longitudinal hole burning). De Merlier [7] extended this work by collecting both models in an encompassing theory for both gain-guided and index-guided waveguide structures and also including gain saturation. It is

clear that for this work a very simple model will suffice.
First of all, it is not necessary to take gain saturation effects into account, as these will only come into play at large photon densities. Due to the fundamental operation principle of the studied device it is clear that high photon levels will never build up inside the SOA cavity. Indeed, the desired bias current results in QW carrier densities near the point where the gain strictly compensates the loss. Therefore, up to this operation current it can be assumed that there is no exponential increase of the optical power and that consequently the QW gain will be longitudinally homogeneous. On top of that it has been found in section 3.3.2 that the transparency operation current densities are close or even slightly beyond the point where the fundamental QW material gain starts saturating. Therefore, in view of the experimental observation that the internal quantum efficiency severely drops at these high current densities, it is very probable that the SOA/isolator may never be able to reach the theoretical operation regime of a non-reciprocal amplifier. It is only in this latter regime that possible longitudinal gain saturation might occur.
Secondly, it has been shown by De Merlier (in [7, 8]) how cross-correlation between the spontaneous emission generated in different guided modes of a multimode waveguide may seriously influence the propagation of optical power and ASE through an active device. This cross-correlation comes about through lateral non-uniform gain distribution in the active layer of the device, caused by a non-uniform carrier profile. The lateral carrier profile is determined both by the local power density (*via stimulated recombination*) and the local carrier density itself (*via nonradiative, radiative and Auger recombination*). Cross-correlation between spontaneous emission in different guided modes is thus "self-feeding" as it will further influence the lateral carrier profile. The question is whether all of these effects will have an influence on the behaviour of our studied application. The effects described by De Merlier are only relevant in highly multimodal devices such as MMI's. Next to that, even though the carrier generation through the injected current density is very high in our device and will thus lead to non-negligible carrier-dependent diffusion terms, the arguments of the previous paragraph make clear that the same certainly does not hold for the optical power-dependent diffusion terms. Therefore, while the QW's in the integrated SOA/isolator will certainly exhibit a non-uniform lateral carrier distribution due to the carrier-dependent diffusion terms, its effect will mainly be a reduction of the effective gain of the MQW region for all guided waveguide modes and not a cross-correlated coupling of these modes, nor a longitudinal variation of the gain.
In view of both conclusions, the model that will be used for the propagated ASE is that of incoherently summed and exponentially growing (or decreasing, depending on the modal gain or loss) modes. Next to that, because it is assumed that longitudinally the carrier distribution is homogeneous, the amount of spontaneous emission generated in every infinitesimal segment along the propagation direction is taken to be constant.

The question remains however whether the detected ASE power has a monomodal origin or is the incoherent sum of the power of several guided modes. While the SOA/isolators have all been designed to be monomodal up to 2.5μm ridge

widths, a numerical study by Van Parys[4] [9] has revealed that the first order TM mode is even at rather low-etch depths just beyond cut-off. The most probable explanation for this is that the CoFe metal contacts have a rather high real part for the refractive index (see table 4.2), causing a large lateral refractive index profile. The main problem is that slight variations in the processing might cause this first order TM mode to become cut-on. On top of that, it is not so difficult to see that a TM mode in the neigbourhood of cut-off (or better cut-on) will experience a much lower metal optical absorption. Finally, it will be explained later on that it was far from easy to process devices in which the ridge is entirely contacted by the metal. An accidentally cut-on first-order TM mode will in this case experience even less absorption, and might even get better amplified than the ground TM mode[5]. Next to that, there have also been measurement on devices with a larger ridge width (than the assumed monomodal width of 2.5μm).
Still, the model used to extract the nonreciprocal absorption shift will assume a monomodal waveguide. It will be indicated in the next sections how this assumption when falsely applied on a multimodal waveguide, will lead to an underestimation of the nonreciprocal effect. In any case, it is clear that it is far from evident to correctly analyze the observed non-reciprocity of the ASE when the device is multimodal. If one does not want to resort to a precise modelling of the spontaneous emission factor for higher order modes[6], then all one can do is to include extra fitting parameters in the models of the next sections. This is clearly not the path which will lead to high accuracy.

5.1.2.1 Model 1: no optical feedback

The simplest model for the detected ASE emitted by one facet of the nonreciprocal SOA cavity, is the one which assumes that the back facet does not contribute. By this it is meant that ASE which is generated in the backward direction (relative to the observation direction, thus away from the observation facet), and which subsequently reaches the back facet and finally gets reflected of this facet towards the front facet, will not contribute to the detected power. The reason for this can be multifold. Either the SOA is processed with tilted facets and it can be assumed that facet reflectivity is negligible (and/)or the bias current is not high enough for backward generated ASE to "survive" just above half a round trip in the cavity.
In this model, TM ASE generated in a portion dz of the amplifier cavity at position z and coupled into a certain mode will grow exponentially proportional to

$$\begin{aligned} e^{(\Gamma g-\alpha_{f,b})(L-z)}dz &= e^{(\Gamma g-\alpha_{\text{int}}\pm\frac{\Delta\alpha}{2})(L-z)}dz \\ &= e^{G_{f,b}(L-z)}dz, \end{aligned} \tag{5.1}$$

the plus sign being applicable for forward propagating light and the minus sign for backward propagating light, and assuming that the origin of the x-axis is

[4] which will certainly be an in-depth discussed topic in his Ph.D. work [12].

[5] This latter behaviour is quite annoying in non-reciprocal lasing devices as will be shown later.

[6] which can only be achieved if first the lateral carrier diffusion equation is solved in the QW's (see for instance [5, 10]).

chosen at the back facet. Do not confuse the meaning of the term "forward" (and "backward") in this context with that same term used in the context of "ASE generated in the backward direction". In the latter case, a distinct physical direction is meant, while here both terms refer to the detection of ASE light propagated in the same direction but with opposite direction for the magnetization! L is of course the cavity length, $\Gamma g - \alpha_{\text{int}}$ is the net unperturbed modal gain/loss of the considered TM mode (with α_{int} collecting all modal loss mechanisms, such as IVBA, metal losses, scattering losses, free carrier absorption, ...), and $\Delta\alpha$ is the nonreciprocal absorption shift for the considered mode. Relating $\alpha_{f,b} = \alpha_{\text{int}} \mp \Delta\alpha/2$ as in (5.1) implies that a positive $\Delta\alpha$ leads to a lower absorption loss for what is assumed to be forward propagating light. However, this is not a priori certain. In this context, "forward propagating light" merely implies the light that is detected when the magnetization is switched along the positive y-axis (as in Fig. 5.1). Whether this light has indeed a negative nonreciprocal loss shift depends on the MO properties of the metal. The isolation ratio is then found as $IS[\text{in dB/cm}] = 10/\ln(10)|\Delta\alpha| \approx 4.34|\Delta\alpha|$ (with the absorption shift expressed in 1/cm). Integrating (5.1) over the length of the amplifier once for forward propagating light and once for backward propagating light and dividing both integrals yields

$$\begin{aligned} S = \frac{P_{f,\text{ASE}}}{P_{b,\text{ASE}}} &= \frac{\int_0^L e^{G_f(L-z)}\mathrm{d}z}{\int_0^L e^{G_b(L-z)}\mathrm{d}z} = \frac{e^{G_f L}}{e^{G_b L}}\frac{\int_0^L e^{-G_f z}\mathrm{d}z}{\int_0^L e^{-G_b z}\mathrm{d}z} \\ &= \frac{G_b}{G_f}\frac{1-e^{G_f L}}{1-e^{G_b L}} = \frac{G^{(0)} - \frac{\Delta\alpha}{2}}{G^{(0)} + \frac{\Delta\alpha}{2}}\frac{1-e^{(G^{(0)}+\frac{\Delta\alpha}{2})L}}{1-e^{(G^{(0)}-\frac{\Delta\alpha}{2})L}}, \end{aligned} \tag{5.2}$$

where $G^{(0)}$ is introduced as a short hand notation for the unperturbed modal gain. This simple ASE model thus leads to an analytical expression for $S = \frac{P_{f,\text{ASE}}}{P_{b,\text{ASE}}}$, expressing the ratio of the detected TM ASE levels with $\mathbf{M}$ switched between $+\mathbf{M}_{\text{sat}}$ and $-\mathbf{M}_{\text{sat}}$. As explained, this quantity is a relative measure for the nonreciprocal effect and (5.2) can be used as a fitting model. The independent variables in this model are the length of the amplifier L and the injected current density J. The latter is hidden in $G^{(0)}$. By assuming that the unperturbed modal gain (or loss) is logarithmic dependent on the injected current[7] with the form $A\ln(J) - B$ (with J usually expressed in kA/cm^2). Written in this form, the fitting parameter B collects both the effects of the QW transparency current density and the internal losses α_{int}.

This model thus introduces three fitting parameters, A, B and $\Delta\alpha$, and two inde-

[7] In principle one could assume an extended modal gain/loss versus current relationship with the form $A\ln(J/B) - CJ$, taking into account carrier concentration dependent internal losses. However, the experiments of Chapter 4 have revealed that even at very high current densities the pure logarithmic relationship is still a very good approximation. In any case, the extra fitting parameter C would be very small and it has been our experience that its use more often than not leads to convergence problems in the fitting procedure.

pendent variables L and J.

$$\boxed{S(J,L|A,B,\Delta\alpha) = \frac{A\ln(J) - B - \frac{\Delta\alpha}{2}}{A\ln(J) - B + \frac{\Delta\alpha}{2}} \frac{1 - e^{(A\ln(J)-B+\frac{\Delta\alpha}{2})L}}{1 - e^{(A\ln(J)-B-\frac{\Delta\alpha}{2})L}}} \tag{5.3}$$

It has to be remarked that this model is only useful if $\frac{\partial\Delta\alpha}{\partial J} = 0$. It could indeed be that higher current densities cause the unperturbed modal profile to reduce the confinement near the MO metal, and thus a reduction of $\Delta\alpha$. However Fig. 2.18 of the benchmark calculations in Chapter 2 proves that this is not the case for a large range of material gain levels in the active layer. Looking at the form of (5.2), it is clear that its limit behaviour is correct. For $\Delta\alpha \to 0$, $S \to 1$, just as for $G^{(0)} \to \pm\infty$. For $\Delta\alpha \to \pm\infty$, $S \to +\infty, 0$. Finally, if $\Delta\alpha$ is positive, implying that the "forward" mode has a lower loss (or a higher gain), S is correctly always larger than 1, as can be found by a McLaurin expansion of the last expression of (5.2).
The characterization procedure with this model has been implemented in the `Labview` environment of the setup and proceeds as follows

1. Determine through ASE hysteresis measurements what value for the external magnetic field is needed for saturation of the CoFe thin film.

2. Set the applied magnetic field at $+H_{\text{sat}}$, and sweep the injected current from 0 up to 500 mA, while detecting the emitted TM ASE power.

3. Subsequently switch the magnetic field to $-H_{\text{sat}}$ and again measure the TM ASE power for the same range of currents.

4. repeat steps 2. and 3. several times on the same device, and in order to average out statistical fluctuations also for different devices on the same cleaved bar.

5. Finally repeat the whole above procedure on other amplifiers of a different length, BUT with the same ridge width for the amplifiers. This last remark is rather important, as $\Delta\alpha$ is of course dependent on the width of the ridge.

6. Having collected all the data, the ratios of the forward versus backward TM ASE power are calculated together with their standard deviations and plotted as a function of length and injected current density.

7. Finally, fit these data (with error flags) in a least-squares sense to the model (5.3) with as fitting parameters A, B and $\Delta\alpha$ (all expressed in 1/cm).

A final remark should be made about this procedure when applied to a multi-modal waveguide. It was already indicated above that for the studied device the detected TM ASE power is the incoherent sum of the ASE of all guided TM modes in the device. Each of these modes will have its own set of fitting parameters in the model of (5.3), because for every mode the unperturbed modal gain, internal

loss and nonreciprocal absorption shift will be different. Obviously, we are mainly interested in the nonreciprocal absorption shift of the ground mode. The question is thus how using a monomodal model will over- or underestimate this $\Delta\alpha_0$ if the detected ASE power originates from more than one TM mode. Let's assume that the waveguide is bimodal, and that the ground and the first order mode respectively have a forward and backward modal gain noted as $G_{f,0}(J|A_0, B_0, \Delta\alpha_0)$ and $G_{b,0}(J|A_0, B_0, \Delta\alpha_0)$, and $G_{f,1}(J|A_1, B_1, \Delta\alpha_1)$ and $G_{b,1}(J|A_1, B_1, \Delta\alpha_1)$. Here we have explicitly indicated that these modal gains (or losses) depend on the variable J and their specific fitting parameters. The extension of (5.2) to an expression for the TM ASE ratio resulting from two modes now reads

$$\begin{aligned} S &= \frac{1-e^{G_{f,0}L}}{G_{f,0}} + P\frac{1-e^{G_{f,1}L}}{G_{f,1}} \Big/ \frac{1-e^{G_{b,0}L}}{G_{b,0}} + P\frac{1-e^{G_{b,1}L}}{G_{b,1}} \\ &= \frac{G_{b,0}}{G_{f,0}}\frac{1-e^{G_{f,0}L}}{1-e^{G_{b,0}L}}\left(1 + P\frac{G_{f,0}}{G_{f,1}}\frac{1-e^{G_{f,1}L}}{1-e^{G_{f,0}L}} \Big/ 1 + P\frac{G_{b,0}}{G_{b,1}}\frac{1-e^{G_{b,1}L}}{1-e^{G_{b,0}L}}\right) \end{aligned} \tag{5.4}$$

Here we have introduced the constant factor P which expresses the relative amount of spontaneous emission coupled into the first order mode. Note that in principle this relative spontaneous emission ratio is $z-$dependent, but that for the here studied device due to the absence of longitudinal gain saturation the spontaneous emission generated in the different guided modes along the length of the cavity can be assumed to be longitudinally homogeneous. A fortiori the same applies for their relative strength. The prefactor in the last expression in (5.4) is exactly the expression that would normally be used if the ASE was assumed to have a monomodal (ground mode) origin. The magnitude of the factor between brackets in this expression thus determines whether an under- or an overestimation will occur. If the factor between brackets is larger than 1, then the actual ground mode TM ASE ratio will be smaller than the measured ASE ratio S. As a result, using the monomodal model of (5.3) on the measured S (with a bimodal origin) will result in an overestimation of the nonreciprocal effect for the ground mode. If the factor between brackets is smaller than 1, then it is the other way round. It is not very difficult to prove that the bracketed factor will always be smaller than 1.[8]

Indeed, it can be safely assumed that the nonreciprocal effect for the first-order mode is several times smaller than for the ground mode. As a result $G_{f,1}$ and $G_{b,1}$ will not differ much and the last expression in (5.4) can be approximatively rewritten as

$$S = \frac{\mathcal{A}}{\mathcal{B}}\left(\frac{1+\mathcal{C}/\mathcal{A}}{1+\mathcal{C}/\mathcal{B}}\right) \tag{5.5}$$

[8] at least if it is assumed that $\Delta\alpha > 0$.

where

$$\mathcal{A} = \frac{1 - e^{G_{f,0}L}}{G_{f,0}} \tag{5.6}$$

$$\mathcal{B} = \frac{1 - e^{G_{b,0}L}}{G_{b,0}} \tag{5.7}$$

$$\mathcal{C} = P\frac{1 - e^{G_{f,1}L}}{G_{f,1}} \approx P\frac{1 - e^{G_{b,1}L}}{G_{b,1}} \tag{5.8}$$

This last approximation holds because for a higher order mode the nonreciprocal effect will be much smaller than for the ground mode. Under the assumption that $\Delta\alpha > 0$, it is found that $\mathcal{A} > \mathcal{B}$, and as result the bracketed factor will be smaller than 1. Or, in other words the measured total TM ASE ratio S for a bimodal amplifier will be an underestimation of the TM ASE ratio of the ground mode. As a result using the monomodal fitting model (5.3) on ASE ratios measured on a multimodal amplifier will always underestimate the nonreciprocal effect for the ground mode. If $\Delta\alpha < 0$, then a similar argumentation can be made to prove that again the nonreciprocal effect will be underestimated.

5.1.2.2 Model 2: optical feedback

While the previous model has its applicability, it remains questionable whether there is no contribution to the measured ASE power from light that has first "travelled" some distance in the backward direction (away from the detection facet) and subsequently is propagated in the forward direction (after reflection off the back facet). This situation can occur if the SOA is not processed with tilted facets and if the operation current is high enough to make the device quasi-transparent in the forward direction (which is after all the ultimate goal). In order to study how this will alter the previous ASE fitting model, consider Fig. 5.4. The spontaneous emission generated at an arbitrary position z in the cavity couples both to forward and backward propagating modes. In principle these amounts are different because the modal profiles in the forward and backward direction are different. But it was explained in Chapter 2 (section 2.2.3.4) how the transverse magneto-optical effect on the modal profiles is only of second order. Therefore it can be safely assumed that at each point in the cavity an equal amount of spontaneous emission is coupled into forward and backward propagating modes, and that this amount is longitudinally homogeneous due to the absence of gain saturation. At each facet both of these ASE components add incoherently, as they get coupled with random phase variations into the backward and forward guided modes. This means that for both components the Fabry-Perot effects can be calculated separately. Looking at Fig. 5.4 it is obvious that both Fabry-Perot summations will differ only in their leading term (for detection at the front end facet). For spontaneous emission generated in the infinitesimal section dz at z these leading terms are

$$\text{forward: } t\mathrm{d}z e^{G_f(L-z)/2} e^{-jk_0 n_{\text{eff},f}(L-z)} \tag{5.9a}$$

$$\text{backward: } tr\mathrm{d}z e^{G_b z/2} e^{+jk_0 n_{\text{eff},b} z} e^{G_f L/2} e^{-jk_0 n_{\text{eff},f} L} \tag{5.9b}$$

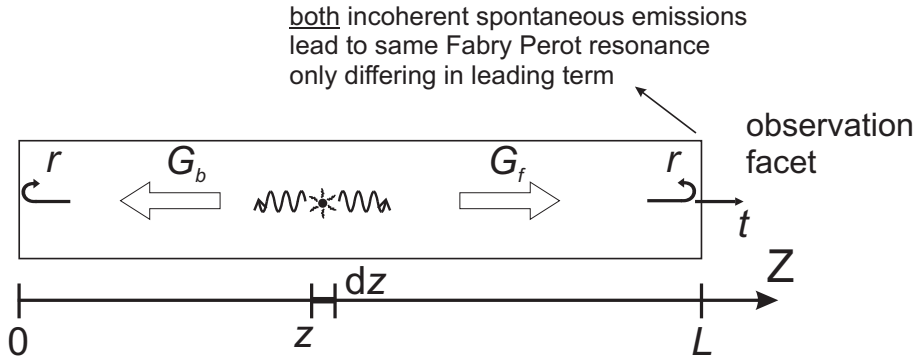


Fig. 5.4: Incoherent ASE contributions at the observation facet in a non-reciprocal cavity close to or above transparency with the magnetization assumed to be pointing in the $+y$ direction. When the magnetization is switched to $-y$, the modal gains G_f and G_b switch.

Factoring out this leading term in both summations, taking the modulus square of both complex amplitude terms to calculate the optical power and adding them gives

$$(|t|^2 e^{G_f(L-z)} R_{FP} + |t|^2|r|^2 e^{G_b z} e^{G_f L} R_{FP})\mathrm{d}z = |t|^2 R_{FP}(e^{G_f(L-z)} + |r|^2 e^{G_b z} e^{G_f L})\mathrm{d}z, \tag{5.10}$$

with R_{FP} the Fabry-Perot power factor. Integrating this expression over the length of the amplifier cavity, and then taking the ratio of the emitted TM ASE powers for opposite directions of the applied magnetic field[9] leads to the following theoretical model for the measurement points

$$\begin{aligned} S = \frac{P_{\mathbf{M}\|+y,ASE}}{P_{\mathbf{M}\|-y,ASE}} &= \frac{e^{G_f L}}{e^{G_b L}} \frac{\int_0^L \mathrm{d}z (e^{-G_f z} + |r|^2 e^{G_b z})}{\int_0^L \mathrm{d}z (e^{-G_b z} + |r|^2 e^{G_f z})} \\ &= \frac{e^{G_f L}}{e^{G_b L}} \left(\frac{\frac{1-e^{-G_f L}}{G_f} - |r|^2 \frac{1-e^{+G_b L}}{G_b}}{\frac{1-e^{-G_b L}}{G_b} - |r|^2 \frac{1-e^{+G_f L}}{G_f}} \right) \end{aligned} \tag{5.11}$$

Replacing in (5.11) $G_{f,b} = A\ln(J) - B \pm \Delta\alpha/2$ finally leads to the extended fitting function taking resonant Fabry-Perot effects into account.

$$\begin{aligned} S(J, L|A, B, \Delta\alpha) = \frac{e^{(A\ln(J)-B+\Delta\alpha/2)L}}{e^{(A\ln(J)-B-\Delta\alpha/2)L}} \cdots \\ \left(\frac{\frac{1-e^{-(A\ln(J)-B+\Delta\alpha/2)L}}{(A\ln(J)-B+\Delta\alpha/2)} - |r|^2 \frac{1-e^{+(A\ln(J)-B-\Delta\alpha/2)L}}{(A\ln(J)-B-\Delta\alpha/2)}}{\frac{1-e^{-(A\ln(J)-B-\Delta\alpha/2)L}}{(A\ln(J)-B-\Delta\alpha/2)} - |r|^2 \frac{1-e^{+(A\ln(J)-B+\Delta\alpha/2)L}}{(A\ln(J)-B+\Delta\alpha/2)}} \right) \end{aligned} \tag{5.12}$$

[9] which will switch the indices f and b in the above expression.

The measurement procedure now proceeds along the same lines as in the previous section. In this context it must be indicated that we explicitly did not include the power reflection coefficient in (5.12) as a fitting parameter. Even though it should in principle be perfectly possible to do this, it has been our experience that the optimization problem posed by the nonlinear least-squares fitting procedure tends to get stuck at local minima with nonphysical values for $R = |r|^2$. Therefore R is treated as a known parameter (with a value ≈ 0.3). It has been checked afterwards what the influence of changes in R are on the fitting values. These seem to be negligible. As a final remark, it is obvious that the same argumentation as in the previous section can be used to prove that this model will again lead to an underestimation of the nonreciprocal effect of the ground TM mode in the case of a multimodal nonreciprocal cavity.

5.1.2.3 Intermezzo: alternative approach via Hakki-Paoli measurements

Looking at the above models, it is clear that the use of a fitting procedure with no less than three fitting parameters is not the most ideal characterization procedure. As with any fitting procedure the rate of convergence decreases with the number of fitting parameters and the risk of getting stuck at a local minimum increases. On top of that it is far from certain that the simple logarithmic gain model is valid throughout the entire current density range. It was indeed noticed in the characterization of the tensile-strained MQW lasers (see section 4.2.1.2) that at high gains, it is better to use a a model with a current-density dependent transparency current density. In other words, to use in the above models a $B(J)$. However, it is not entirely clear what functional form should be used. Anyway it is not relevant, as the extra fitting parameters would only further burden the convergence rate of the nonlinear fitting procedure.
Therefore, an alternative approach in determining the optical non-reciprocity $\Delta\alpha$ consists of directly measuring the unperturbed modal gain/loss, $G^{(0)}$, instead of introducing it via 2 fitting parameters. What is left then is a nonlinear optimization problem in a single parameter, $\Delta\alpha$, for which eventual concentration dependent losses, thermal effects, quantum efficiencies drops, ... are automatically taken into account, since these are implicitly included by directly measuring $G^{(0)}$. Measuring $G^{(0)}$ at different current densities is normally done by the Hakki-Paoli method [11].

For these devices however, there are a few subtleties associated with the Hakki-Paoli method. The very first measurements on the devices (see section 5.2.2) already indicated that the internal losses are very high. This was immediately clear from the rather low power levels of the emitted TM ASE. And it was also afterwards confirmed by the extracted values for B. The problem with high losses in the Hakki-Paoli method is obvious. As the HP method is based on the analysis of Fabry-Perot resonances in the optical spectrum of the emitted (TM) ASE power, it is clear that for the measurement of low power levels very high sensitivities will be needed, and there is a real danger of "hitting the sensitivity floor" of the optical spectrum analyzer (OSA). Powers of a few 10 nW per wavelength interval

of 1 nm are to be analyzed. Moreover, this presupposes that one is able to couple close to 100% of the available power into the spectrum analyzer. In reality this will be closer to 20% or less. All of this leads to the conclusion that spectral power densities of the order of about -50dBm/nm need to be analysed, for which it would be advisable to have a sensitivity of -80 or more dBm. Also, in order to limit the uncertainties in determining the peaks and valleys of the resonances high wavelength resolutions of the order of 100 pm or better are needed. The combination of high wavelength sensitivity and high amplitude sensitivity leads to long sweep times, risking decreasing power levels during the spectral sweep. All of the above remarks are of course important but the biggest difficulty of HP measurements on low-level signals is the problem of aligning a very weak free space invisible infrared beam with a GRIN lensed fiber coupled to the OSA. Working with free space beams is of course required for polarization selective determination of the modal gain/loss . Therefore an alignment trick was devised as shown in Fig. 5.5.
Using a removable 45˚ mirror in the light path of the optical beam, the beam can be steered towards a sensitive CCD camera, whose image is visible on a monitor. Without the 45˚ mirror the light beam is steered towards the GRIN lens terminated fibre. Aligning a high power TM infrared laser beam with the help of infrared detection cards into the GRIN lens terminated fibre, and subsequently putting the mirror back in place provides us with a reference position on the monitor screen. This position on the monitor is used as an alignment mark. Replacing the alignment laser with the non-reciprocal amplifier to be tested, and once again using the 45˚ mirror (in its fixed position) the weak infrared beam will easily be detected by the sensitive CCD camera. The translation stage bearing the optical objective to produce the parallel beam can now be positioned to “align” the weak beam with the alignment mark on the monitor. As a direct consequence this will lead to a first rough alignment of the weak beam with the GRIN lens (after removing again the 45˚ mirror). Afterwards, the piezo-electric actuators on the translation stage can be used to optimize this alignment.

This alignment procedure is quite laborious but is the only way of doing HP measurements on weak beams. In section 5.2.2 a few examples of attempts at measurements with this extended setup will be given. But it can already be remarked here that in order to compete with the ASE fitting characterization procedure of the previous sections, a lot of attention has to be paid to (thermal) stabilization of the measurements. Due to the high requirements on the wavelength resolution and the amplitude sensitivity, a spectral sweep of 10 nm can take up to 1 minute. It has been noticed that it is very difficult to keep the alignment constant during this time interval (probably due to thermal drift). Reducing the wavelength range of the sweep limits the number of valleys and peaks in the Fabry-Perot resonances, and will greatly reduce the accuracy of the determined modal gain/loss. In general, we would like to have a spectral sweep spanning some 50nm and for at least 10 different current densities in order to achieve a reasonable fitting for $\Delta\alpha$. It is clear that this might easily take up to one hour of measurement time. The speed of the ASE characterization methods of the previous sections will be

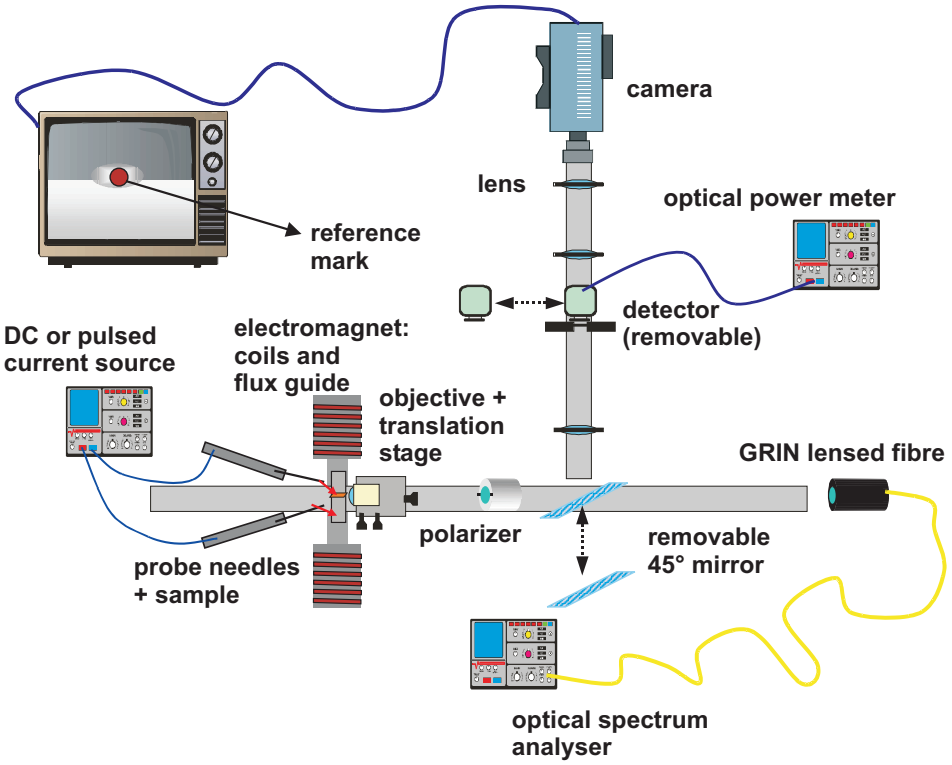


Fig. 5.5: Alternative measurement set-up allowing to extract the unperturbed modal gain/loss directly via Hakki-Paoli measurements (instead of treating it via two fitting parameters). Only $\Delta\alpha$ is then left as a fitting parameter, and no assumptions have to be made about the precise analytical model for the unperturbed gain/loss.

difficult to beat. But this might be compensated for by the increased accuracy

5.1.3 Non-reciprocal stimulated emission

The use of the extended ASE model taking resonant reflections into account is primordial when the device under test approaches transparency in either one direction. Though one might argue that in this case, the isolator design has failed as obviously to be a good isolator the other direction has to be strongly absorbing. Of course everything scales with length, and because of practical experimental considerations, especially the amount of needed absolute current, the lengths of the characterized devices more often than not lead to a total backward loss that is not that drastic and still might lead to the occurrence of resonances. Remember that the conclusions of the section on design predicted for the strongest nonreciprocal designs maximal backward losses of 100 dB/cm. With cleaved test devices of 500μm this leads to "only" 5dB cavity loss. With a roundtrip reflection loss of 10dB ($R \approx 0.3$), this implies that it is not unthinkable to observe onset of lasing action even in a nonreciprocal cavity. For the above example, lasing could be observed if an additional single pass unperturbed modal gain (i.e. on top of the gain needed to bring the device in forward transparency) of 7.5dB could be provided. In the 500μm device of this example, this means an extra unperturbed modal gain of about 35/cm. This is not a small amount, especially since it has to be provided on top of the gain needed for forward transparency but it is not unrealistic.

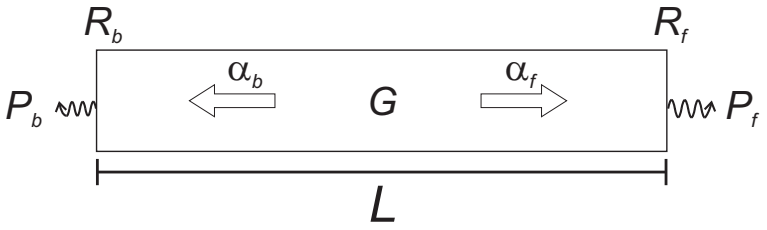


Fig. 5.6: Nonreciprocal active cavity with an unperturbed modal gain G, and direction dependent losses α_f and α_b. These latter include the unperturbed modal loss $\alpha_{\rm int}$ For completeness the possibility of different facet reflectivities has been allowed. However even with $R_f = R_b$, the forward and backward emitted power under lasing conditions will differ due to the nonreciprocity of the cavity.

When a non-reciprocal cavity starts lasing, the characterization of the nonreciprocal absorption loss gets considerably less complicated. For that purpose consider the cavity depicted in Fig. 5.6.
Lasing (reciprocal or non-reciprocal) occurs when the roundtrip amplitude gain is unity.

$$\sqrt{R_f R_b} e^{(2G-\alpha_f-\alpha_b)L/2} = 1, \tag{5.13}$$

or equivalently,

$$e^{GL} = \frac{1}{\sqrt{R_f R_b}} e^{(\alpha_f + \alpha_b)L/2} \tag{5.14}$$

Next to that, denoting the optical powers at the facets just inside the cavity with a prime, we have

$$P_b = (1 - R_b)P_b^{'} = (1 - R_b)e^{(G-\alpha_b)L} R_f P_f^{'} \tag{5.15a}$$

$$P_f = (1 - R_f)P_f^{'} \tag{5.15b}$$

Dividing both expressions and using (5.14), the ratio of the emitted forward to backward laser power becomes

$$\begin{aligned} \frac{P_f}{P_b} &= \frac{1 - R_f}{1 - R_b} \frac{1}{R_f} e^{-(G-\alpha_b)L} \\ &= \frac{1 - R_f}{1 - R_b} \frac{1}{R_f} \sqrt{R_f R_b} e^{-(\alpha_f + \alpha_b)L/2} e^{\alpha_b L} \\ &= \frac{1 - R_f}{1 - R_b} \sqrt{\frac{R_b}{R_f}} e^{(\alpha_b - \alpha_f)L/2} \end{aligned} \tag{5.16}$$

With $\alpha_b - \alpha_f = \Delta\alpha$ and in good approximation[10] $R_f \approx R_b$, we finally find as a simple nonreciprocity characterization formula for a nonreciprocal laser above threshold

$$\boxed{\Delta\alpha = \frac{2}{L} \ln\left(\frac{P_f}{P_b}\right)} \tag{5.17}$$

This last equation could of course also have been obtained by deriving the direction-dependent external differential efficiencies of the nonreciprocal cavity. In any case, it is clear that if the SOA/isolator happens to lase the characterization of the non-reciprocal absorption shift gets much simpler. On top of that, in such cases one shouldn't worry about higher order effects, as the lasing condition clamps the resonances in the cavity to the mode needing the lowest roundtrip gain. Still, it could be that the cavity is lasing in a higher order mode (as was suggested by Van Parys [9]). In any case, whether lasing occurs at a higher order mode or not, the observed nonreciprocity will have a single-mode origin.

5.1.4 Non-reciprocal transmission of TM guided light

It could be argued that the ASE and non-reciprocal lasing characterization methods of the previous sections are not entirely capable of a correct assessment of the device performance. There is nothing wrong with these methods when it comes to a pure determination of the MO non-reciprocal absorption shift. But one of the most critical aspects of the studied design is the fact that a purely passive optical function is realized in an active layout. This is not a vain remark. First

[10] the nonreciprocity of the facet reflection coefficients is a second-order effect, as it mainly results from the difference in the forward and backward modal profiles.

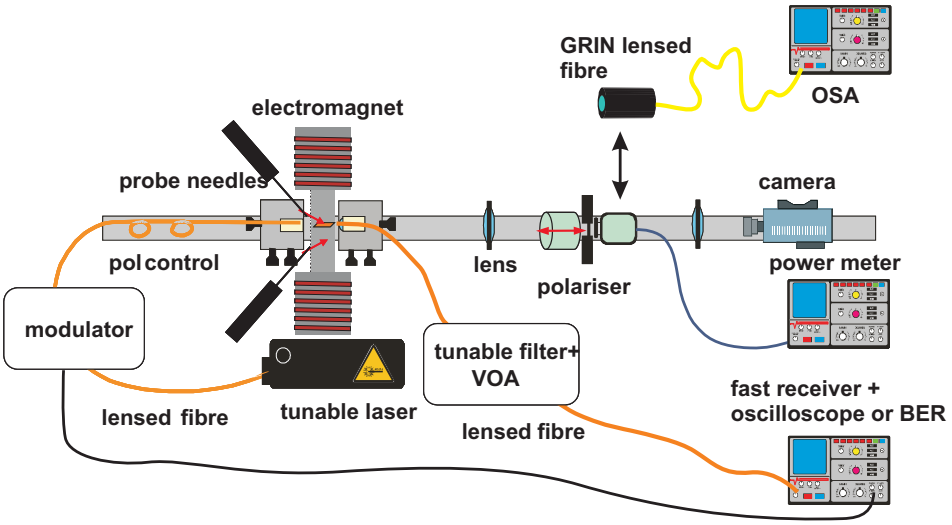


Fig. 5.7: Ideal measurement setup for a complete characterization of all the implications of an active realization of an intrinsically passive function.

of all, it is obvious that the SOA layout of the studied isolator will inevitably add spontaneous emission noise to the output signal. But an even more important consequence of the active layout of the device is that, due the intrinsic light-carrier interaction in an optical amplifier, the transfer function of the studied isolator is not "power-transparent"! It is not difficult to realize that as a consequence the extinction ratio of a modulated laser signal will be decreased by the SOA/isolator. This is a nasty drawback, as it intrinsically limits the output extinction ratio of an integrated laser/isolator using this novel concept. Finally, the carrier dynamics of the SOA also has implications for the "bit-rate transparency". It is by no means clear up to what speeds the carrier concentrations in the SOA will be able to "follow" the injected modulated laser light.
Within the time frame of this research, which was after all intended in the first place to demonstrate for the first time ever the novel concept of Nakano, these important device characteristics have not been studied yet. A theoretical analysis combined with a characterization will be the subject of a further optimization study of the device [12]. A possible layout for an experimental bench studying these effects is depicted in Fig. 5.7. This setup has been partially implemented, allowing at the time of writing only an observation of the nonreciprocal absorption of externally injected CW TM light. An example of such a measurement will be given in section 5.2.3.

The polarization of the injected light (modulated or CW) is controlled by polarization wheels. This is necessary as the optical fibre will not maintain the polarization state of the emitted laser light. CW transmitted light can be directly analyzed by an optical power meter or a spectrum analyzer as a function

of the magnetization direction, the injected current, the injected power, and the wavelength. The use of an optical spectrum analyzer is necessary to separate the non-reciprocity undergone by the generated ASE noise from the non-reciprocity undergone by the injected CW TM light. Next to this, it gives an indication of the noise figure of the device. For a modulated signal, the degradation of the extinction ratio and of the optical-signal-to-noise ratio (OSNR) can be analyzed using an oscilloscope preceded by a tunable bandpass filter (in order to eliminate superfluous noise). When combined with a variable optical attenuator (VOA) just before the receiver, error rate measurements can be done, in order to check what receiver sensitivity penalty the SOA layout of the integrated isolator introduces.

Finally, we would like to indicate that the ultimate experimental setup will need to characterize the fully integrated laser/isolator and more in particular the efficiency of the protection provided by the SOA/isolator when this fully integrated device is subjected to deliberate optical feedback. For this purpose it suffices to add a kind of fibre-based back reflector apparatus at the output facet of the device in the setup of Fig. 5.7. A simple version of such an apparatus would be a fibre-based 2×2 splitter one of whose outputs is provided with a mirror. The other output is connected to the measurement instruments as in the above sketched set-up. By controlling the reflectivity of the mirror, the optical feedback can be controlled (and monitored at the unconnected input of the splitter). Observation of the optical spectrum of the laser light will reveal at what levels of feedback the coherence of the laser light collapses. BER measurements can be performed to investigate up to which feedback levels the laser can be operated quasi error-free. Within the time frame of this research, there have been no integrated modules fabricated. It makes no sense to start this work before the SOA/isolator itself is capable of providing sufficient isolation.

5.2 Experimental results

5.2.1 Processing and mask layout

As explained in section 3.3.2, within the time frame of this research two generations of SOA/isolator demonstrators have been developed. The next two sections will detail the characterization results for both generations, using the variety of methods discussed in the previous sections. Before presenting these results, it is worthwhile to spend a few words on the processing of the devices, and indicate important points of concern during processing.

The main trump of the novel integrated isolator concept is its straightforward monolithic integration with standard III-V semiconductor laser diodes. To a large extent this is not only because both the laser and the SOA/isolator can be grown in the same epitaxial growth sequence on the same substrate, but also because the processing sequence of the integrated isolator is in essence not different from that of standard SOA's and lasers, and hence can profit of the extensive know-how on this processing. In principle the only extra processing step as compared to a standard laser processing consists of the extra metallization needed to form the ferromagnetic CoFe contact. Incidentally however, this processing step is the most

critical in the entire scheme, especially when a ridge-type layout is used for the devices. This is because the current isolation layer (usually polyimide or Al_2O_3) needed in these cases, requires a very critical alignment of a photo-resist mask with ridges of just a few μm. Of course this issue also arises in the processing sequence of ordinary ridge type lasers, but there it is less critical as the aspect ratio of the ridges is usually very high. This is illustrated in Fig. 5.8

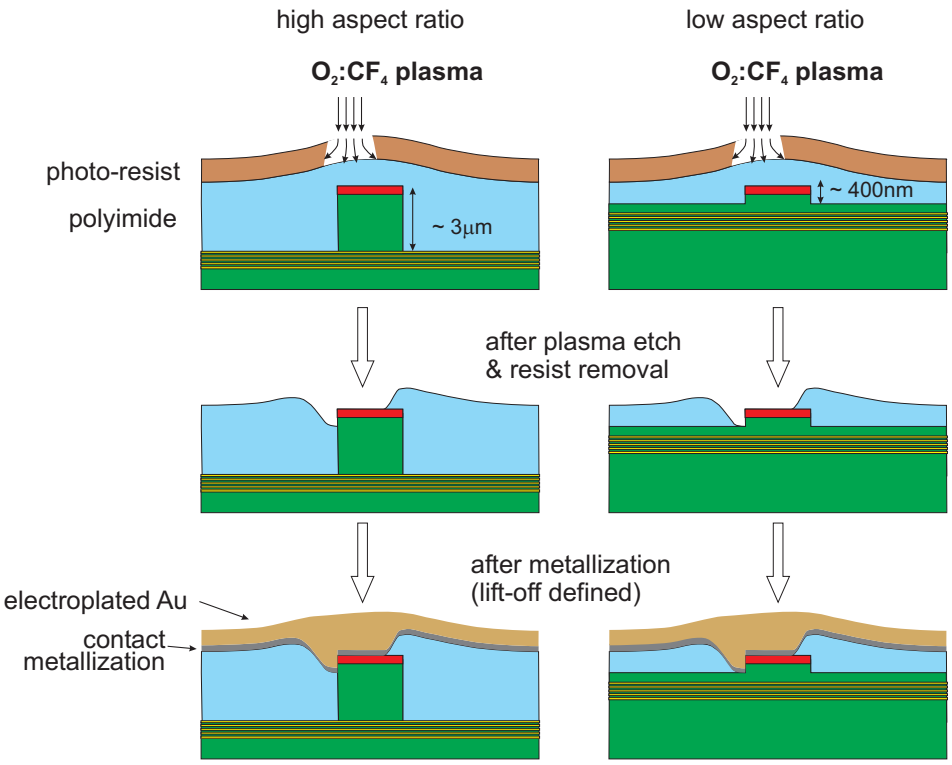


Fig. 5.8: Illustration of the effects of a misalignment of the isolation mask on high and low aspect ratio ridges. Due the isotropy of the plasma etch (and the etching of the resist itself, a misalignment might lead to part of the contact window opening up next to the ridge. For a low aspect ratio ridge this can have dramatic consequences.

The plasma etch process, used to open the isolating layer, is isotropic and also etches the photo-resist (defined through image reversal). As a result the current windows defined in this step will always be slightly wider than defined by the dimensions of the isolation mask. Therefore, it is usually aimed to deposit an isolating layer of a thickness that only slightly exceeds the etch depth of the ridges. As a result the thickness of the isolating layer above the ridge is thin enough so that the effects of the isotropic character of the plasma etch remain limited. However

when the ridge wave guides have a very small aspect ratio, the thickness of the isolation layer above and next to the ridge won't differ a lot. A small misalignment of the isolation mask will inevitably result in the formation of a current contact window next to the ridge. Such a situation is highly unfavorable as an important fraction of the injected carriers get injected into a region where there is no optical guiding. For the studied SOA/isolator concept, this is in particular important, as its design relies entirely upon the limited thickness of the InP cladding. In a ridge type layout (with a current contact on top of the ridge, i.e. without a mechanism such as lateral current injection) there is not much that can be done to avoid this[11]. Next to a very precise mask alignment, one can think of using an isolation mask with significantly smaller widths than the ridge widths (in order to relax the alignment) or changing the processing scheme and etching the ridges by using the metal contact as an etch mask. However, the former "trick" lowers the partial ferromagnetic metal coverage of the ridge and hence reduces the overall non-reciprocity, while the latter "trick" would indeed result in a quasi 100% coverage of the ridges by the ferromagnetic metal (without the risks depicted in Fig. 5.8) but with a width-to-length aspect ratio that is very disadvantageous for the magnetic shape anisotropy of the contact. In other words, the latter processing scheme would never allow the device to obtain reasonable remanent magnetization in the transverse direction. It will be shown later on why this critical alignment processing step needs extra attention.

The following sequence summarizes the processing scheme of the SOA/isolators. Except for the extra metallization step, there is no difference with standard ridge type laser processing.

a. wafer preparation (natural oxide removal)

b. deposition SiO_x (150 nm)

c. lithography (ridge structures)

d. etching of SiO_x: plasma etch (O_2:CF_4)

e. removal of resist

f. RIE etching

g. removal of oxide: wet etch (HF)

h. spinning of polyimide (or sputter deposition of Al_2O_3)

i. lithography: definition of current windows (image reversal)

j. etching of current windows: plasma etch

[11] A buried layout is not confronted with this very critical alignment as the injected carriers are automatically channelled to the low band-gap active region, and there is no need for a current isolation layer in order to prevent carrier injection in the regions next to where the optical mode is confined. However, all of the devices in this work have a ridge-type layout. Of course, the fully integrated module will most likely have a buried layout. Future work will be devoted to this.

k. cleaning of exposed semiconductor contact interface (sulfuric acid dip)

l. lithography: definition of 50μm wide lift-off patterns for CoFe metallization (image reversal)

m. sulfuric acid dip (diluted 1:4). This extra step is needed because the sample is transferred to the sputter machine in IMEC.

n. deposition of 50nm CoFe under magnetic bias ($\perp$ stripes)

o. lift-off in aceton of CoFe

p. lithography: definition of extra metallization lift-off patterns for protection of CoFe (image reversal)

q. sputter deposition and Joule evaporation of Ti/Au bilayer (40/150 nm)

r. aceton lift-off

s. optional step: Au plating

t. thinning of the substrate (down to $\sim$ 150μm)

u. deposition back contact AuGe: no anneal (avoiding detrimental damage to the Ohmic properties of the CoFe contacts – see the results of section 4.1.3), no Ni (to avoid any guiding of the magnetic flux lines through the back contact)

In a later stage the processed wafers can be submitted to a RTP treatment in order to enhance the Ohmic quality of the CoFe contacts or left unannealed. After processing, the structures are cleaved and mounted as-cleaved with silver epoxy on a copper or aluminum carrier, with dimensions dedicated for the mechanical design of the sample holder in the measurement setup. Ideally, the SOA/isolator facets should be coated with AR coatings. This would greatly simplify the analysis of the non-reciprocal transmission of externally injected TM light, as it reduces the occurrence of FP resonances inside the cavity. However, within the time frame of this research this effort has not been made. Next to that, it wasn't even considered to process the SOA/isolator ridges with tilted facets, which is another way of avoiding cavity resonances. In other words, the processed structures are really nothing else but FP ridge type lasers with a non-reciprocal contact. The reason for this choice is very pragmatical. First of all, it is far from obvious beforehand whether the devices will be able to reach transparency, which is of course the conditio sine qua non to be hindered by FP resonances. Secondly the treatment of the previous section has made it abundantly clear that even with a resonant cavity, there are sufficient characterization tools available to analyze the performance of the SOA/isolator. And finally, it is of course considerably simpler from an experimental point of view to measure simple straight facet waveguides. The laser mask used to process these non-reciprocal structures therefore consists of simple straight waveguides whose only varying parameter is the ridge width. The widths on the mask are 1.5, 2, 2.5, 3, 5, 7 and 9 μm.

5.2.2 Generation 1: CoFe/InGaAsP SOA/isolator

As explained in section 3.3.2.1, 5 InGaAsP MQW layer structures have been designed. Their details are summarized in Tables 3.6 and 3.7. Just as a reminder: there are three 3QW designs, labelled as 'safe', 'normal' and 'risk' referring to the thickness of the InP cladding, and a single 6QW and 9QW design. The etch depths of the structures were calculated to guarantee TM monomodality up to a ridge width of 3μm. Using a commercial 2D waveguide solver it was found that most of the structures can be etched through the active MQW layer without making them TM multimodal, which is a good thing bearing in mind that current spreading is nihil in this case. However to avoid surface recombination and still keep the current spreading low enough, it was opted to etch just through the upper SCH layer. Only for the 9 QW structures this was not possible. All etch depths are simulated to have a 20 nm tolerance. This is within the ranges of the technological possibilities of RIE etching.

5.2.2.1 TM ASE hysteresis measurements

On all devices TM ASE hysteresis measurements have been performed. As explained above these measurements serve in the first place as a qualitative indication of the presence of the non-reciprocal optical absorption shift. Nevertheless, they convey also important information about the magnetic characteristics of the device, more in particular the performance as a permanent magnet of the processed thin CoFe stripes. It was demonstrated in section 4.1.2 that sputter-deposited CoFe stripes of thicknesses of 50nm need at least a $w/l \approx 0.015$ aspect ratio to ensure the desired in-plane transverse magnetic anisotropy. The above processing scheme indicates a 50μm width for the CoFe lift-off patterns. This implies maximal device lengths of about 3–4mm, which is in the expected range of needed device lengths according to the simulations of section 3.3.2.1. Moreover, for characterization purposes, the devices are usually cleaved in bars of only 1mm or less, which further increases w/l. Therefore, one expects to observe sufficiently square hysteresis curves similar to the bottom left curve in Fig. 4.11.
Fig. 5.9 shows some typical TM ASE measured hysteresis curves. The absolute values of the detected TM ASE powers are irrelevant for the magnetic properties of the contact. All curves have therefore been normalized. The results depicted here are somewhat disturbing. Despite the 50μm pattern width for the CoFe lift-off contacts and the particularly "short" device lengths (leading to contact aspect ratios as high as $\approx$0.08), the observed hysteresis curves are far from square! Remanence ratios in the order of only 50–70% are observed (and for the 9QW device even as low as 40%). Strangely enough at the same time the coercive fields are close to 100 Oe. This is not even closely related to what is observed for the highest aspect ratio CoFe stripes in Fig. 4.11. Comparing these results to results obtained in Fig. 4.11 suggests that the actual physical width of the CoFe contact is considerably lower than the desired 50μm. SEM evaluation of the facets of the processed devices revealed that the CoFe contact material was indeed physically broken along its width, and that these defects reduced the effective contact width to that of the ridges themselves. It will be shown further on that this is the result

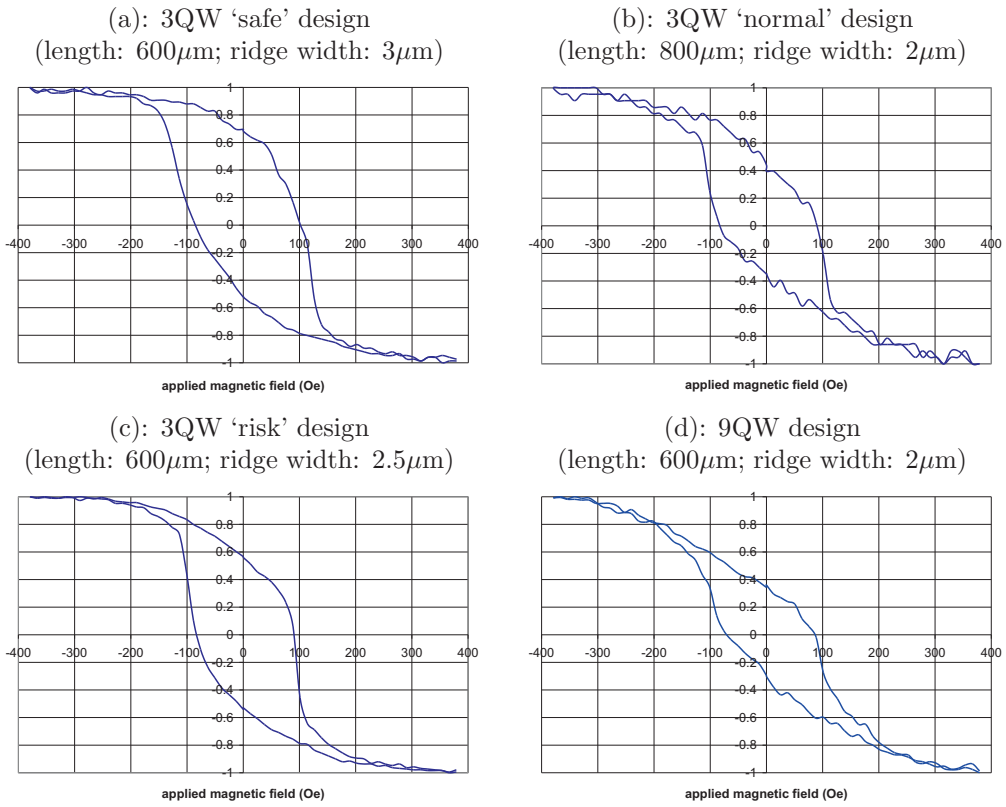


Fig. 5.9: Normalized TM ASE hysteresis curves measured on a variety of devices.

of both a processing error and a "non-optimal" choice of current isolation material. In any case, the successful observation of TM ASE hysteresis curves on all five processed structures was the very first experimental verification of the novel concept for integrated non-reciprocity.

5.2.2.2 Quantitative ASE characterization

The measurement procedure described in sections 5.1.2.1 and 5.1.2.2 is applied on all five layer structures. For each layer structure two or three different lengths have been cleaved. The following table (Table 5.1) summarizes the geometrical details of the cleaved devices. There have been no characterization efforts on devices with a larger width. The main reason was not their multimodality, but rather the current limit of the current source preventing to reach sufficiently high current densities through the samples. After all, it was very quickly noticed on the monomodal samples that the output power levels were rather low. In order to be not too much hindered by noise rather high current densities were required (in the order of 20kA/cm^2). For the largest ridge widths of 5, 7 and 9μm these levels

structure	width (μm)	length (μm)
3QW safe	3	400, 600, 750
3QW normal	2.5	600, 800
3QW risk	2.5	300, 500, 800
6QW	2.5	450, 850
9QW	2.5	300, 580, 750

TABLE. 5.1: Geometry of the tested first generation demonstrator devices.

were not attainable with the available current source.

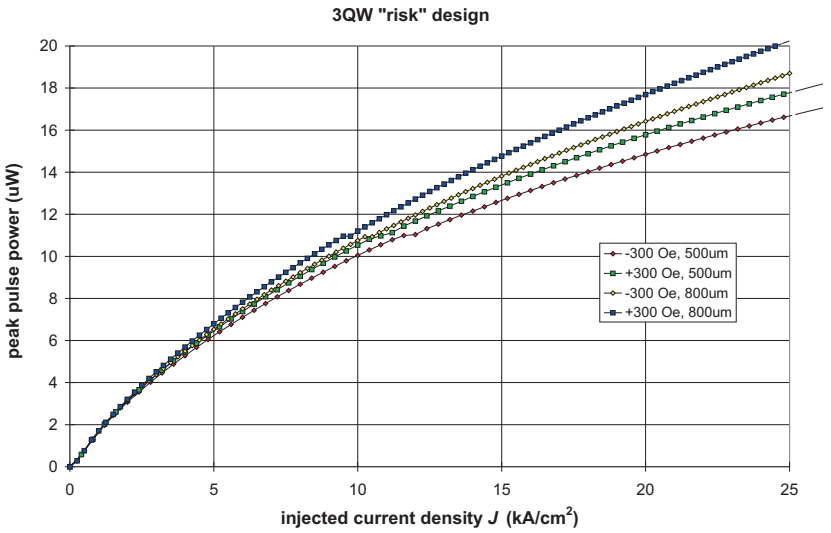


Fig. 5.10: Typical TM ASE measurements on 3QW "risk" devices with the applied magnetic field switched between $\pm H_{\mathrm{sat}}$.

Fig. 5.10 shows typical measured TM ASE output powers with the applied magnetic field set at either $+H_{\mathrm{sat}}$ or $-H_{\mathrm{sat}}$. All measurements have been performed under pulsed conditions (duty cycle 10% and pulse width of 100ns). The applied magnetic field is switched between ±300 Oe, enough to saturate the CoFe contacts as seen on the hysteresis measurements. It immediately strikes that the detected power levels are low. The saturating behaviour of the TM ASE has therefore nothing to see with nonlinear gain saturation inside the cavity, but is a direct emulation of the fundamental QW gain saturation. This is quite logical looking at the values for the injected current density. Moreover, it was observed that for all the layer structures the power levels remained this low. On top of that, no consistent varia-

tion of the output power as a function of InP cladding thickness was observed. In this context, it should be remarked that strangely enough the 3QW "safe" design, which has the thickest InP cladding and thus in principle the lowest unperturbed isotropic modal loss, led to ASE output levels which were hardly detectable! All of this indicates that not only the isotropic optical absorption of the CoFe film was seriously underestimated at the time of design, but that also the gain performance of the InGaAsP wells was overestimated[12].

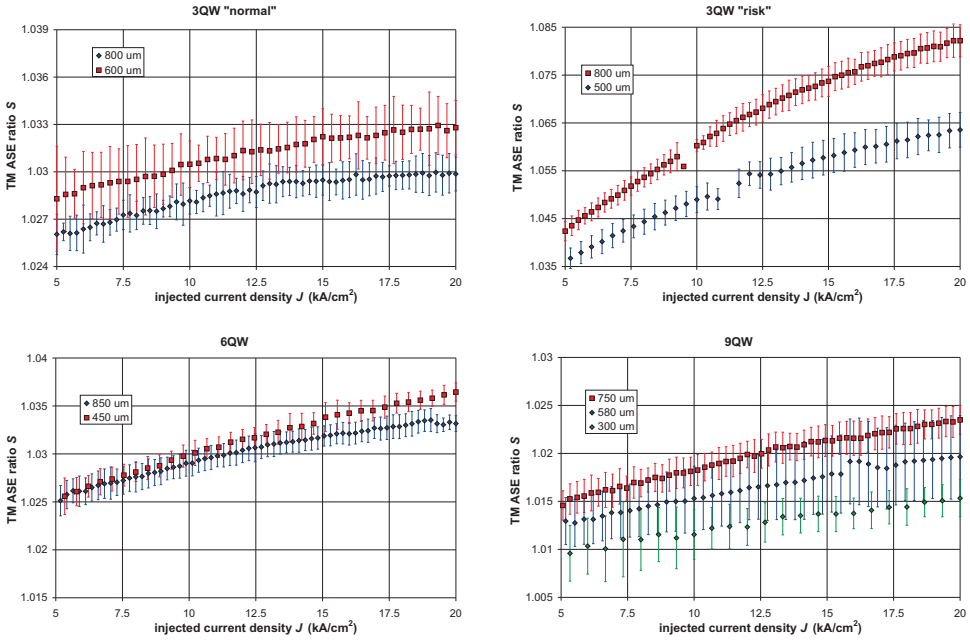


Fig. 5.11: TM ASE ratio measurements on the first generation demonstrators.

Even though the measurements as the one depicted in Fig. 5.10, indicate that all devices are operated far away from forward transparency, this obviously doesn't prevent to observe the non-reciprocal absorption effect in the SOA cavity. This is clearly observed in this figure, though it is not easy to distinguish the dependence of the non-reciprocal effect on the injected current and on the cavity length in these kind of plots. Plotting the ratio of the ASE levels, makes this clearer, and also produces the measurement data for the fitting function S of (5.3) and (5.12). Figure 5.11 plots this ratio for all layer structures (except the 3QW "safe"). All measurements have been averaged out over at least three different devices, and on each device the measurements have been repeated five times. The results depicted here make immediately clear that for all layer structures, except the 3QW "risk"

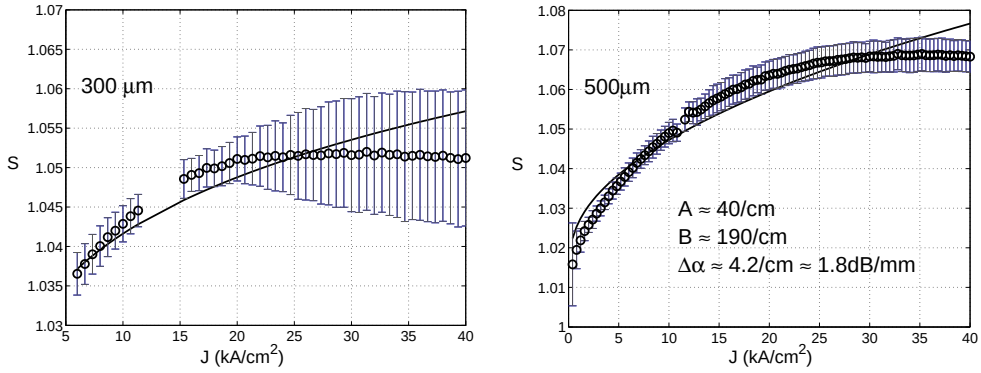
[12]Recall that for this first generation design there was no experimental input neither for the CoFe (magneto-)optical constant nor for the gain behaviour of the strained InGaAsP wells.

structure, either the unperturbed modal loss is very high or the gain performance of the MQW region is less strong than expected (or a combination of both). In view of the quasi independence of the TM ASE ratio S with respect to the length of the device, it is clear that the "normal" and 6QW devices definitely suffer from a high modal loss. The 9QW devices do show some degree of dependence on the length, but the values for S are very low and the relative flatness of the curves as function of J seems to suggest that the confinement near the CoFe contact is much lower than expected and that the unperturbed gain is also overestimated. In this context it should be remembered that this latter layer structure is the only one that was shallowly etched (for reasons of monomodality). Possibly, this led to considerable current spreading and an underestimated drop of the modal confinement near the MO metal.

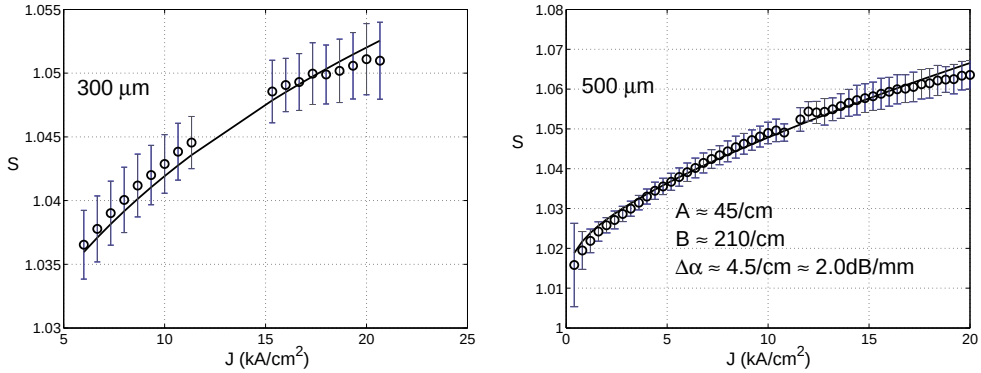
In order to quantify these assumptions, the data points for S in Fig. 5.11 are fitted to the ASE models of (5.3) and (5.12). As explained in the sections 5.1.2.1 and 5.1.2.2, this nonlinear fitting is done simultaneously to all available data points of a single family. This fitting procedure will only be illustrated for the 3QW risk device, being the most successful measurement. A good fit was nevertheless obtained for all four depicted layer structures. The fitting procedure takes the error flags into account as weighting factors in the least-squares sum, and the search algorithm itself used to minimize the least-squares sum is either a Levenberg-Marquardt algorithm or a Broyden-Fletcher-Golfarb(-Shanno) algorithm. Both seem to converge rather well. All of this has been implemented in a `MATLAB` program with a very rudimentary user interface.

Fig. 5.12 shows the result on the 3QW "risk" device. There are a few subtleties involved when fitting the data. If all measurement points are used for the fitting, then the used model shows a tendency to diverge at larger current densities (see Fig. 5.12(a)), a behaviour that is not observed at low current densities. By now it is known, that this behaviour is most likely to be caused by band-filling effects. As extensively indicated in the tensile-strained MQW characterization study of the previous chapter, this implies that the simple logarithmic model for the unperturbed modal gain doesn't hold anymore at very high current densities. A better model would use a QW transparency current density which slightly depends on the injected current density $J_0(J)$. The unperturbed gain/loss model $A\ln(J) - B$ of (5.3) and (5.12) assumes a constant J_0. On top of that it is probable that temperature effects start playing a role at these very high current densities. In any case, it is believed that the most reliable fitting values for this model are obtained by only taking the "low" current density measurements. Table 5.2 summarizes the fitting results obtained on all samples with both fitting models.

The obtained fitting parameters are to say the least somewhat surprising. First of all it can be noticed, that the "normal" and the 6QW devices suffer from a very high unperturbed modal loss (contained in the fitting parameter B), while the "risk" and the 9QW device show considerable lower loss. Next to that, the extracted value for the non-reciprocity absorption loss doesn't show any tendency in accordance with the design values. On the contrary, the "normal" design (with the thickest InP cladding) even has a non-reciprocal effect that is more than twice



(a)



(b)

Fig. 5.12: Nonlinear fitting of $S(J, L|A, B, \Delta\alpha)$ (see (5.3)) to the measurement points of the 3QW "risk" device, taking all measurement points into account (a), or only the measurements up to 20kA/cm^2 (b).

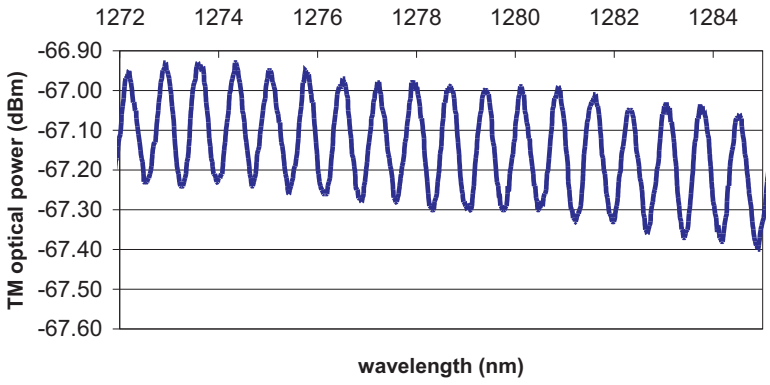
layer	(5.3)			(5.12) with $R = 0.3$		
	A (1/cm)	B (1/cm)	$\Delta\alpha$ (dB/mm)	A (1/cm)	B (1/cm)	$\Delta\alpha$ (dB/mm)
3QW (N)	40	420	4.4	40	450	4.5
3QW (R)	45	210	2.0	45	200	1.98
6QW	84	520	6.5	115	800	6.4
9QW	30	160	0.6	30	150	0.55

TABLE. 5.2: Extracted fitting parameters for the ASE ratio measurements of Fig. 5.11, using both ASE models.

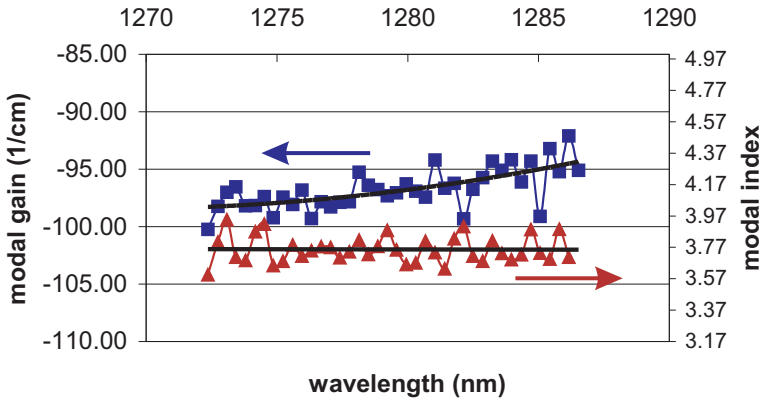
as strong as the "risk" design (with the thinnest InP cladding). Only the values for the fitted differential gain parameter A are more or less in accordance with previously obtained results (see Table 4.10 and Fig 4.20). It can be seen that the use of the extended ASE model (taking optical feedback into account) leads to a negligible change of the fitting parameters. This is not a surprise in view of the high modal losses observed for all the layer structures, making the occurrence of optical feedback and its influence on the output ASE power almost negligible.
The question remains however if the above extraction values are trustworthy. The lack of any consistency in the values for $\Delta\alpha$ and the very high values for B in the 6QW and "normal" designs are some reason of concern. Indeed, even though at the time of design of this first generation there was no experimental assessment of the (magneto-)optic constant of CoFe nor of the gain performance of the tensile-strained InGaAsP wells, it is bizarre to observe this complete lack of trends between the samples. In order to check the validity of the used model, it was decided to try to confirm the observed values for the cavity loss by Hakki-Paoli measurements.

5.2.2.3 Hakki-Paoli measurements

As explained in section 5.1.2.3, the tried and trusted HP method can also be used to characterize the optical non-reciprocity of the SOA/isolators. It has the obvious advantage of reducing the number of fitting parameters as the unperturbed modal gain/loss is directly measured. However as indicated in that section, its drawback is its laboriousness and its need for a very long measurement time interval when high loss samples are investigated. The measurement results of the previous section prove that this will be certainly the case for the here considered devices. The HP setup of Fig. 5.5 has therefore not been used to do non-reciprocity measurements, but only to validate the bizarre fitting values for the loss parameter. This has been tried for the "normal" and "risk" 3QW designs and for the 6QW designs.
Fig. 5.13 shows the results obtained on the "risk" device. For the 6QW and the "normal" 3QW device it was absolutely impossible (even after several alignment optimizations and measuring on all available samples) to observe any ripple whatsoever in the optical TM spectrum. This led to believe that a very strong absorption effect was at play, which would be a confirmation of the results of Table 5.2. On the other hand, the values obtained for the modal loss of the risk device are, within the accuracy of the HP method, in accordance with the predictions made by the ASE fitting procedure of the previous section! Indeed, with $A = 45/\text{cm}$, $B = 210/\text{cm}$ (from Table 5.2) and $J = 6.5\text{kA/cm}^2$, $A\ln(J) - B \approx 120/\text{cm}$. Similar accordance was obtained at current densities up to 10kA/cm^2. These observations seem to confirm the validity of the used ASE fitting model, that is: at least for the observed loss parameters B. Remains to find an explanation for the strange fitting values of $\Delta\alpha$. However, now that it is confirmed that the obtained B values are very likely to be correct, it is not difficult to understand what could be at the origin of the observed trends for $\Delta\alpha$ (or better lack thereof). Indeed, Table 5.2 shows that very high losses go hand in hand with (relatively) high values for $\Delta\alpha$. This can only be explained if the source for the high losses is the MO metal itself.



(a)

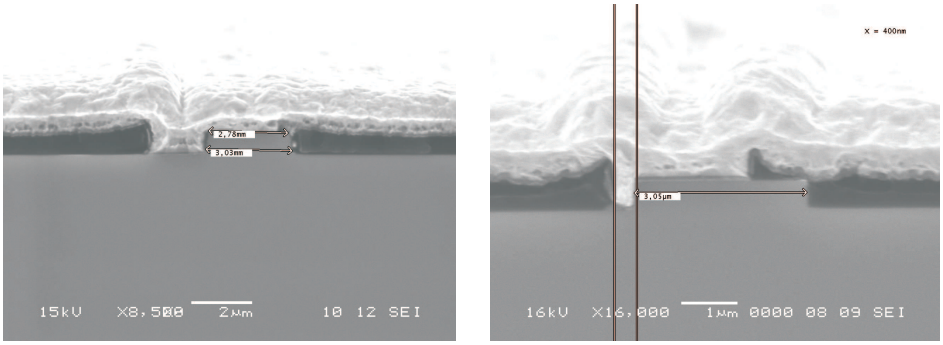


(b)

Fig. 5.13: (a): small fraction (15nm) of the Hakki-Paoli spectrum of a risk 3QW device (L= 300μm , w= 2.5μm , J= 6.5kA/cm^2) near the gain peak; (b): corresponding extraction of the Fabry-Perot ripple (it is clear that due to the low power levels only a limited accuracy of about 10% is obtained)

And if, for the particular devices of concern, this metal is much closer to the guided light than intended. Only then can it be explained that e.g. the "normal" design has a higher loss and a higher non-reciprocity than the "risk" design.

In order to check this assumption the facets of the "normal" and 6QW designs were inspected by SEM. Fig. 5.14 shows typical observations of these facets. These observations are the ultimate confirmation of the correctness of the ASE fittings. The processing error described in section 5.2.1, namely a misalignment of the isolation mask on a low aspect ratio device, has led to an opening of the polyimide



(a) (b)

Fig. 5.14: (a): SEM picture of the facet of a 2.5 µm wide ridge of a 3QW "normal" device; (b): SEM picture of the facet of a 3µm wide ridge of a 6QW device. In both cases a considerable amount of metal is deposited next to the ridges.

layer next to the ridge. This has resulted in the presence of CoFe very close to the guiding layer of the devices, explaining the origin of the very high modal loss ánd the relatively higher $\Delta\alpha$. SEM observation of the facets of the devices with a more reasonable modal loss was also performed. In Fig. 5.15 the facet of a risk device is shown.

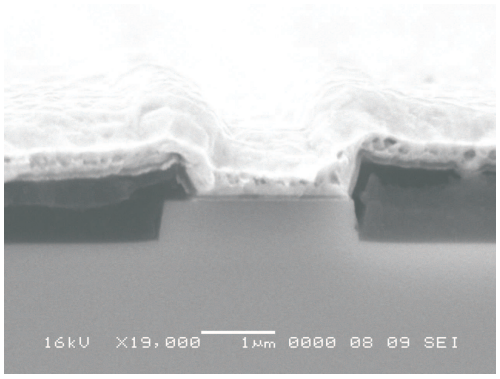


Fig. 5.15: SEM picture of a 3QW "risk" device.

It is clear that also here there is some slight misalignment of the isolation mask, though not with the same drastic consequences as in Fig. 5.14. As an extra, these SEM observations also explain the somewhat disappointing magnetic properties of the devices. It was remarked in section 5.2.2.1 that the observed squareness of the TM ASE hysteresis loops suggested that the CoFe were most probably considerably less wide than expected. Here we can see why this is. For the devices

with the processing error, it is obvious. The CoFe contact breaks at the edge of the ridge. For the other devices however, the SEM picture as in Fig. 5.15 reveals that the polyimide (PI) edges are much steeper and especially higher than expected. The height of the PI edges is almost 1μm! This is a result of a limited control of the thickness of the PI above the ridges[13]. It is obvious that a 50nm CoFe film will break at such a high edge. With a better control of this thickness, the height of the PI edge would be much lower and the edges less pronounced. As a result the sputter deposited CoFe film could smoothly follow this height difference. All of this explains the very low magnetic remanence observed in Fig. 5.9. The CoFe contact width that actually contributes to the emitted TM ASE is not 50μm wide but merely 2.5μm or even less. Finally, it is also observed that even with perfectly symmetric processing of the isolation mask the CoFe metal contact only covers about 70% of the ridge! This is not unimportant. As it reduces the amount of light subjected to the non-reciprocal effect.

5.2.2.4 Influence of annealing

Finally, to conclude the characterization of the first generation devices, it remains to assess whether the device performance changes after a RTP treatment of the contacts. It was not explicitly mentioned, but all the presented results have been obtained on devices with as-deposited CoFe contacts. It has been concluded in section 4.1.3 of the previous chapter that the as-deposited CoFe contacts in itself already prove to be of very good Ohmic quality. However, rapid thermal processing of the contacts at 300°C leads to a decrease of a factor 4 of the contact resistivity. Therefore, it is expected that the gain performance of the amplifier will increase if the samples are alloyed, since the series resistance of the devices will decrease and thermally induced spill-over of the carriers will decrease. However, it is not clear what the influence of such a temperature treatment will be on the magneto-optic properties of the ferromagnetic contact. It is evident that a decrease of the contact resistivity by annealing is in one way or another always the result of some intermixing or in/out-diffusion of the chemical elements at the interface between the metal and the semiconductor. Therefore, it is sound to assume that some kind of absorbing semi-metallic layer of a certain thickness will form at the CoFe/InGaAs(P) interface. Whether this layer will still show magneto-optic properties and how thick this layer will be, is very difficult to assess experimentally. Therefore, a pragmatic approach has been taken. Of all the characterized devices (with as-deposited contacts) the ones with the highest non-reciprocal effect[14] have been annealed at 300°C for about 30 s, and then measured again. An eventual decrease in non-reciprocity would be easiest to detect for the best devices. Should the decrease of the optical non-reciprocal effect be negligible or small, then there is no need for a deeper investigation of the thermal behaviour of the magneto-optic properties of the CoFe/InGaAs(P) interface. The measurements on the annealed

[13] Polyimide is spun onto the wafer to be processed and not deposited by some calibrated evaporation process.

[14] i.e. the "risk" devices, as the higher non-reciprocity of the "normal" and 6QW designs is more a result of the CoFe next to the actually intended electrical contact region.

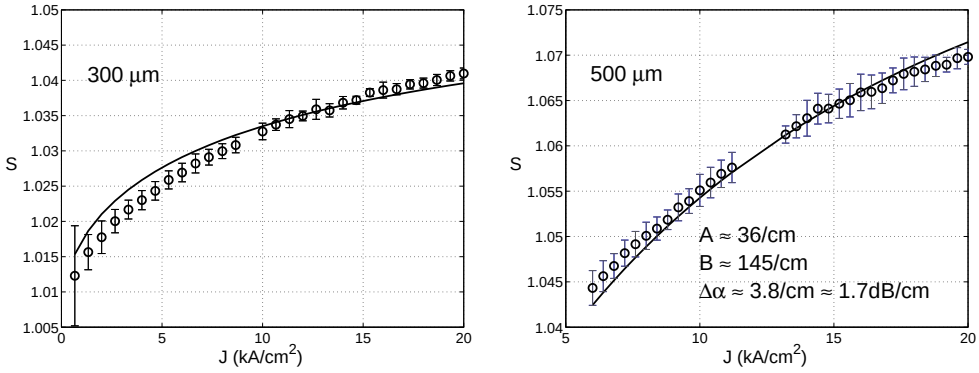


Fig. 5.16: Results of fitting procedure for an annealed 3QW risk device. The rapid thermal annealing temperature was 300˚C.

3QW "risk"device are shown in Fig. 5.16 along with the results of the fitting procedure. It is clear that main effect on the performance of the non-reciprocal amplifiers is indeed to decrease its internal loss. There is only a small decrease of the optical non-reciprocity itself.

It is safe to say that the magneto-optic properties of the CoFe interface are only slightly degraded by the annealing treatments needed for the optimization of the electrical contact.

5.2.2.5 Conclusion

This concludes the characterization of the first generation SOA/isolator demonstrators. In many ways the experiments described here are milestone results as they provide first ever quantitative assessment of the novel integrated isolator concept. Nevertheless, it is clear that the obtained results can only be partially called successful. Indeed, an integrated isolator with an extinction ratio of 20.0dB/cm and an insertion loss of 45/cm at a current density of 25kA/cm^2 is not a device which one would easily promote. However, in view of the very limited information at the time of design, the afterwards confirmed overestimation of the gain performance of the tensile-strained InGaAsP wells, and the underestimated challenges of technological assembly of the building blocks of the SOA/isolator, it is a quite a feat to have observed these isolation levels.
Moreover, towards the end of the characterization of the first generation devices, experimental evidence for the magneto-optic strength of CoFe was available and the gain performance of the tensile-strained InGaAsP MQW layers was determined. This allows a reassessment of the design of section 3.3.2.1. Table 5.3 compares the here obtained extraction results to the reassessed performance for three of the characterized layer structures.
The most interesting results in this table are of course those for the “risk” device,

layer	t_{InP}	A	B	$A\ln(J)-B$ (for $J=20\mathrm{kA/cm^2}$)	$\Delta\alpha_{\mathrm{meas}}$	$\Delta\alpha_{\mathrm{sim}}$
	(nm)	(1/mm)	(1/mm)	(1/mm)	(dB/mm)	(dB/mm)
norm.	300	4.0	42	-30	4.4	7.6
risk	250	4.5	21	-7.5	2.0	9.3
6QW	300	8.4	52	-26.8	6.5	8.7

TABLE. 5.3: Reassessment of the first generation design, comparing the measured non-reciprocity values to the 1D design results using the known magneto-optic constant of $Co_{90}Fe_{10}$ (see section 4.1.1).

as this device was quasi-free of any processing error. Apparently, the observed non-reciprocity is about a factor 4–5 off target. However, one has to be careful in comparing the numbers in the last two columns of the above table. The re-assessed design values are obtained on a 1D slab model of the device, while the measurements are done on 2D ridge-type structures (etched down to the SCH guiding layers). 2D simulations of the device indicate that with these etch depths the non-reciprocity easily decreases by 20%. On top of that, the SEM observation of the previous section indicate that only 70% of the ridge is effectively covered by the MO metal, implying a further decrease of the non-reciprocity. Finally, the measurements of section 4.1.1 show an uncertainty of about 20% on the relevant imaginary part of the gyrotropy constant of $Co_{90}Fe_{10}$.
The assessed gain performance of the tensile-strained InGaAsP MQW region (see Table 4.10 in section 4.2.1.2), also explains the failure to obtain transparency in the first generation devices. For the "normal" and 6QW devices the processing error depicted in Fig. 5.14 sufficiently explains these high losses. However, it is nice to observe that the extracted modal loss for the "risk" device is not only confirmed by the HP measurements but is also afterwards confirmed by the QW gain measurements of the previous Chapter. Indeed, assuming no decrease in internal quantum efficiency up to $20\mathrm{kA/cm^2}$ (which is a bit overenthusiastic), the gain parameters obtained in Table 4.10 would predict a material gain of about 2000/cm for the QW's at $20\mathrm{kA/cm^2}$. It is safe to say that the needed 6000/cm QW transparency gain obtained in the design simulations of section 3.3.2.1, which are not only 1D slab model simulations but have also used an overestimated value for the absorption loss of CoFe, can be roughly relaxed down to values of $4000/\mathrm{cm}$[15] for the processed devices. Using a confinement factor of about 5% for a 3QW device this results in a modal loss in the neigbourhood of 100/cm at $20\mathrm{kA/cm^2}$, which is close to the observed 75/cm.

Taking all of this into account, it is clear that the at first sight somewhat disappointing result of 2.0dB/mm is after all an exciting result for a device designed "in hazy conditions", and probably much closer to the real design value than the

[15] this value is actually the result of a rough guess based on the same arguments as the one used to explain the discrepancy in the non-reciprocity. Namely a decrease of twice 20%, once for the etch depth and once for the overestimated extinction coefficient of CoFe.

measurements would suggest [2–4].

5.2.3 Generation 2: CoFe/InAlGaAs SOA/isolator

By the time the experiments on the first generation were finished, experimental evidence had been gathered that the InAlGaAs system greatly outperforms the InGaAsP system, and that the equi-atomic alloy composition $Co_{50}Fe_{50}$ held the promise of a stronger gyrotropy parameter. On top of that, the experiments on the first generation devices had taught us a lot about the pitfalls of the processing of the SOA/isolator. It was therefore obvious to quit the efforts on the InGaAsP demonstrators, and start the design and processing of second generation devices using this material system with slight corrections to the processing scheme, rather than to try a second processing run on the InGaAsP material. After all, taking the above concluding remarks into account, the most important assessment of the previous section is that even with an improved processing the non-reciprocity might be acceptable (at almost 100dB/cm) but it would prove almost impossible to pump the device to forward transparency. And the experiments on InGaAsP MQW material didn't promise any remarkable improvement to justify a second design.

The design of this second generation has been described at length in section 3.3.2.2. The details of the designed layer structures and their simulated performance are summarized in the Tables 3.8 and 3.9. Three 9QW InAlGaAs layer structures have been designed differing only in the thickness of the InP cladding. The calculated (slab) needed forward transparency current density varied from 6kA/cm^2 to 20 and 30 kA/cm^2. The former design is purely based on a Figure-of-Merit type optimization of the layer structure while the latter two are hypothetical attempts at obtaining a larger $\Delta\alpha$ at the cost of current. We will label these designs as "O", and "R1" and "R2". All three layer structures have been processed using the same standard laser mask as described in section 5.2.1, once with $Co_{90}Fe_{10}$ and once with $Co_{50}Fe_{50}$ as MO metal contact. The specific modifications to the processing sequence will be detailed in the next subsections.

It has to be pointed out that the growth of and characterization work on this second generation of demonstrators was only initiated at the very end of this research. The characterization results will therefore not be discussed in detail. A more in-depth assessment will be part of the successoral PhD research of Van Parys [12].

5.2.3.1 Processing 1

Processed devices and corrections to processing scheme It was explained in sections 5.2.1 and 5.2.2.3 how the combination of the use of polyimide as an isolation layer and the very critical alignment of the isolation mask, can lead and has led to poor magnetic properties for the contact and even to faultily processed devices. In order to solve both problems it was decided for the second generation processing to alter steps h. and i. in the listed processing scheme of section 5.2.1. These should now read:

h. plasma deposition of SiO_2 as current isolation.

i. lithography: definition of current windows (image reversal) *with a shifted mask*!

The use of plasma-deposited SiO_2 allows to achieve a better control of the thickness of the isolation layer above the ridges than with polyimide, allowing a much thinner isolation layer to be etched open above the ridges. And the plasma process ensures a smoother surface for this isolation layer. As a result the lift-off of the CoFe contact runs less risk of breakage at the edges of the contact windows. However this still doesn't solve the dangers of a misaligned isolation mask. Therefore as an extra cautionary step, the isolation mask has been shifted with respect to the ridge mask (modified step i.). In practice this means that in this lithography step the isolation stripes of a certain width have been shifted two "positions". In this way the contact window is considerably less wide than the ridge, and the alignment requirements are relaxed. Remember that the stripe widths on this mask set are: 1.5, 2.0, 2.5, 3.0, 5.0, 7.0 and 9.0μm. In other words, the processed devices have the following characteristics[16].

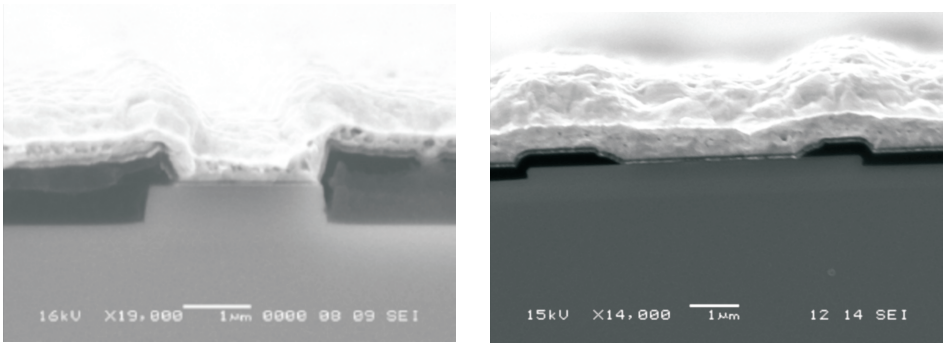
ridge width (μm)	contact width (μm)	coverage (%)
3	2	67
5	2.5	50
7	3	43
9	5	56

TABLE. 5.4: Ridge and contact characteristics of the second generation devices (run 1)

This makes a total of 24 families (3 layer structures, 2 CoFe compositions and 4 widths) of devices[17]. Fig. 5.17 illustrates the successful improvement of the processed devices. Though not perfectly visible, it can be seen how the first metal layer of the contact metal stack smoothly follows the etched SiO_2. As a result, the actual CoFe width contributing to the magnetic contact properties will most likely be the desired full 50μm as defined by the lift-off mask. This is indeed observed in the TM ASE hysteresis of these devices, a typical example of which is shown in Fig. 5.18. The observed hysteresis curve is in perfect agreement with the results depicted in the bottom left curve of Fig. 4.11. Actually the coercivity observed here is even considerably larger, suggesting that the induced in-plane anisotropy by the applied magnetic bias field during sputtering is even stronger than what

[16] the 1.5μm has not been used as a contact window in view of its expected rather high resistance.

[17] The final number of processed device families was not 24 but 16. During processing the $Co_{50}Fe_{50}$ "O"- and "R1"-families were accidently destroyed.



(a) (b)

Fig. 5.17: Comparison of the processed devices of the first and the second generation. (a): first generation 3QW "risk" device; (b): second generation "R1" device.

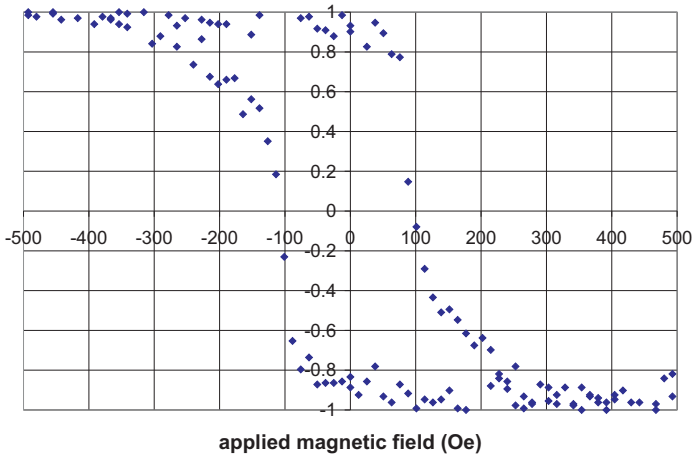


Fig. 5.18: TM ASE hysteresis measurement on a 7µm wide "R1" ridge (with $Co_{90}Fe_{10}$ contact). The hysteresis curve is almost perfectly square, and a remanence ratio of over 90% is observed, indicating that the magnetic flux lines within the 50µm wide CoFe contact are not broken.

was expected previously. The most important result is the remanence ratio of over 90%. The remanent isolation ratio therefore almost equals the isolation ratio at saturation, which is very important from a device point of view, as it allows to operate the isolator without the need for an external magnetic field. The first generation devices only achieved a remanent isolation of about 50–60%.
It can also be seen in Fig. 5.17(b) how the very difficult alignment of the isolation mask is now freed of the danger of a contact processing error. The contact is indeed

not placed perfectly symmetric on the ridge, but there is no risk of accidently opening the SiO_2 next to the ridge. Still, it is clear that this is not the ideal solution. The ridges are only partially covered by magneto-optic material. As a result, the processed device does not fully exploits its capabilities, because an important fraction of the guided TM light near the top of the ridge doesn't overlap with MO material.

Characterization – non-reciprocal lasing It was expected to observe much higher TM output power than in the first generation devices, in view of the experimentally assessed much stronger gain performance of the tensile-strained InAlGaAs MQW material. This was indeed the case. Whereas in the first generation devices the peak pulse power[18] was typically not more than 20μW, these second generation devices all show TM ASE output powers of not less than 150μW, while the widest ridges (7 and 9μm) even start lasing! This is illustrated in Fig. 5.19 for a range of $Co_{90}Fe_{10}$ "R1" devices.

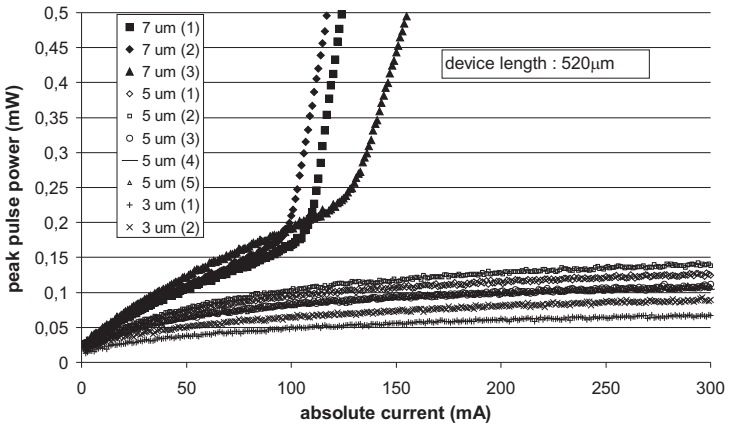


Fig. 5.19: LI curves for a range of $Co_{90}Fe_{10}$ "R1" devices.

It can be seen that within the same family of devices there is a considerable variation on the output powers, and on the differential efficiencies and threshold currents of the lasing devices. Most probably this is due to variations on the position of the metal contact with respect to the center of the ridge. It was illustrated in Fig. 5.17(b) that is far from evident to achieve a symmetric positioning of the contact. Small variations on this positions will obviously cause variations in the modal properties (more specifically the loss) of the guided TM modes, and hence on the output power. In any case, these high ASE output powers and lasing action

[18] In order to avoid thermal effects as much as possible all experiments were once again conducted in the same pulsed conditions as with the first generation devices (10% duty cycle, 100ns pulse width).

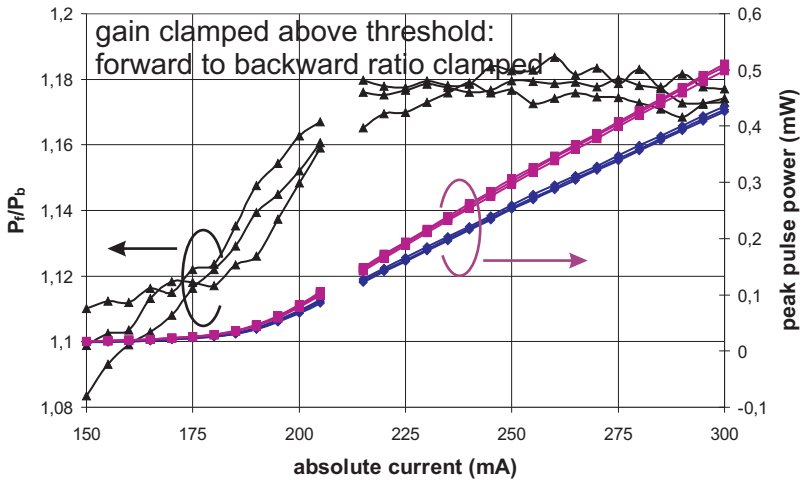


Fig. 5.20: Example of a nonreciprocal lasing measurement on a $Co_{50}Fe_{50}$ "R2" sample (ridge width = 9μm, length = 520μm). Using (5.17) *the observed nonreciprocity is 6.45/cm or 2.8dB/mm.*

serve as the umpteenth proof of the excellent gain performance of tensile-strained InAlGaAs material.
For reasons that are still not clear, the output power of most of the 3 and 5μm devices quickly degraded with time. This prevented performing non-reciprocity fitting TM ASE measurements on these devices. On the other hand, the observation of lasing action allows a very straightforward characterization of the non-reciprocity, as explained in section 5.1.3. Fig. 5.20 shows an example of observed non-reciprocal lasing. The magnetic field is switched between $\pm$500 Oe. Above threshold the gain provided by the MQW is clamped to the sum of the mirror loss and the net cavity internal loss. As a result the single pass gain is also clamped. This single pass gain relates the power emitted at the front and back facets of the device (assuming that the reflectivities of both facets are equal), and thus above threshold the ratio of these powers is clamped and directly related to the non-reciprocal absorption shift (see (5.17)).

This type of measurements has been performed on all lasing devices. However, it seemed almost impossible to observe any consistent trends between the observed $\Delta\alpha$ as a function of the varied device parameters. Remember that these parameters are the thickness of the InP cladding (and hence the $\Delta\alpha$), the CoFe alloy composition (or the gyrotropy constant and hence the $\Delta\alpha$), and the ridge width (or in view of Table 5.4, the partial MO metal coverage, and hence the $\Delta\alpha$). It is very likely that this lack of trends is due to the fact that the characterized lasing devices are in fact multimodal[19]. It is suspected that the observed lasing is

[19]The etch depth of the devices was simulated to lead to TM monomodality up to ridge widths

in fact occurring in a higher order TM mode. This would be in accordance with the remarks formulated in section 5.1.2 and the simulations performed by Van Parys [9]. He observed that already for a ridge of 2.5μm there is a risk of mode competition between the ground TM mode and the first order mode when the ridge is not fully covered by the MO metal. It is not difficult to see that with a partially covered ridge chances are very high that the first order mode will have the lowest modal loss and will reach threshold first. On top of that, because a first order mode has a nodal plane in the center of a ridge waveguide, its modal loss is much more sensitive to variations in the symmetric positioning of a partially covering metal contact than that of the ground mode. This could explain the considerable variations on the observed threshold currents and differential efficiencies. Limited far field experiments have confirmed these assumptions [12].

In a way, this higher-order lasing is both a gift and a drawback. It is a gift in the sense that the obtained non-reciprocity values are definitely very conservative lower limit guesses of the $\Delta\alpha$ for the ground mode. On the other hand, it is of course the ultimate goal to obtain forward transparency in a purely monomodal non-reciprocal device. In Table 5.5 a cautious overview of the observed non-reciprocities as a function of the different device parameters is given. It is clear that the here tabulated trends are more qualitative than quantitative. For reasons explained above, the characterization results often varied considerably within one family of devices. Therefore, the values in the following table are for devices where this variation was limited.

family	$\Delta\alpha$ (dB/mm)		
$Co_{90}Fe_{10}$, 7μm	O	R1	R2
	1.95	1.52	3.26
$Co_{90}Fe_{10}$, R1	5μm	7μm	9μm
	3.47	1.52	1.09
R2, 5μm	$Co_{90}Fe_{10}$	$Co_{50}Fe_{50}$	
	1.95	2.82	

TABLE. 5.5: Observed trends for the non-reciprocal absorption shift as a function of the device parameters. The first column entry indicates on which family of devices the trend was observed.

Except for the variation of $\Delta\alpha$ as a function of CoFe alloy composition, the trends tabulated here are at first sight counterintuitive. Especially the trend as a function of ridge width doesn't seem logical. But it can probably be explained, when it is taken into account that the observed $\Delta\alpha$ is most likely that of a higher order mode and that with increasing ridge width the overlap of the TM_1 mode with the metal decreases. Remember that for all devices only the central part of the ridge is covered by CoFe with a quasi-constant coverage of 50%. Thus for increasing ridge width the antisymmetric TM_1 modal overlap with the MO metal

of 2.5μm.

decreases. The variation of $\Delta\alpha$ as a function of alloy composition confirms more or less the observed increase of 70% of the strength of the imaginary part of the gyrotropy constant (see Table 4.2). Only the variation of $\Delta\alpha$ as a function of layer structure is mysterious. The "R2" structure (with the thinnest InP cladding) has indeed the highest non-reciprocal effect. But the effect is weakest in the "R1", while it should be the "O" structure that has the lowest $\Delta\alpha$. Of course, it is dangerous to extrapolate the predictions of the simulations of Chapter 3, which were for a TM ground mode in a slab waveguide, to a ridge waveguide lasing in a first order mode.
In any case, the best result obtained here, namely 3.47dB/mm is an improvement of almost 75% compared to the first generation devices. On top of that, it has been indicated that this is definitely a lower limit in view of the higher order effects.

Characterization – non-reciprocal TM transmission Having devices at our disposal which can provide sufficient gain to start lasing (albeit in a higher order mode), it is reasonable to try and observe non-reciprocal transmission of externally injected TM light through the device. Up till now, such an experiment has been successfully conducted on just a single device using a basic version of the setup depicted in Fig. 5.7. Light from a fixed wavelength FP laser ($\lambda = 1312$nm) is coupled into the device using a lensed fibre. The polarization of the injected light is controlled using polarization wheels. The transmitted light is picked up by an AR coated objective and detected by an optical power meter. The relative variation of the detected power when switching the magnetic field between $\pm$500Oe results in a straightforward extraction of $\Delta\alpha$ as

$$\Delta\alpha = \frac{1}{L} \ln \frac{P_f}{P_b} \tag{5.18}$$

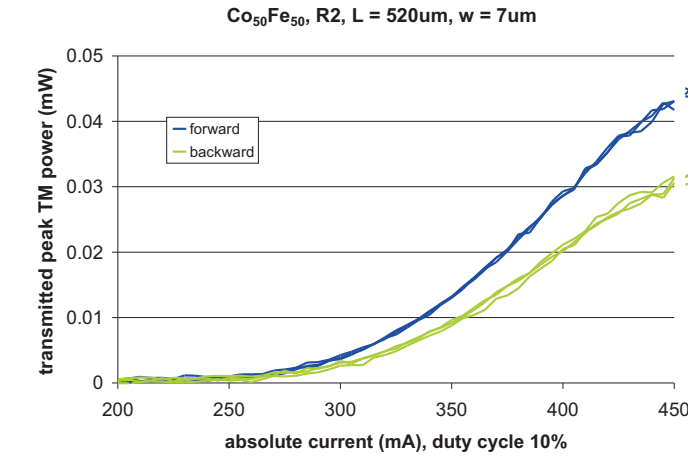
This type of experiment investigates the SOA/isolator device from an external viewpoint by directly measuring its non-reciprocal transmission. Even though this leads to a more fundamental characterization of the device performance, there are a few subtleties involved with this measurement procedure. These are a direct result of the active layout of the device. The light emitted by the device contains both the external signal ("data") and the ASE generated by the SOA itself ("noise"). If the detection is not spectrally resolved, the measured non-reciprocity will be some weighted average of the $\Delta\alpha$ undergone by the noise and by the data, making it very difficult to extract the "pure" data non-reciprocity. This is especially the case if the device is multimodal. The multimodal noise added to the output signal will have a non-reciprocal behaviour which is very difficult to analyze. On top of that it is even unsure whether the data signal will only excite the ground TM mode.
But even in a purely monomodal SOA/isolator, it is not so easy to correctly analyze the observed non-reciprocity. While the signal's non-reciprocity is entirely described by (5.18), the ASE's non-reciprocity rather follows one of the models (5.2) or (5.11). Separating the signal from the ASE can partly be done by measuring with an OSA or by adding a filter centered at the wavelength of the data signal, but that still doesn't allow to extract the longitudinal ASE mode that lies within the band of the filter or within the detection interval of the OSA. In any

case, the most accurate determination of the non-reciprocal transmission properties of a SOA/isolator requires both a monomodal device and a spectrally-resolved detection (or detection with a tunable passband filter).

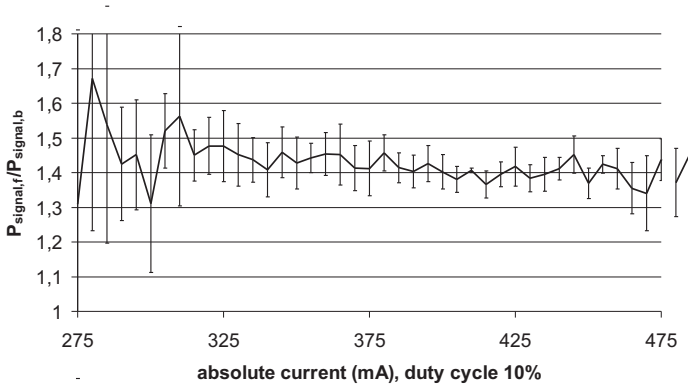
At the time of writing, a spectrally resolved transmission measurement had not been performed yet. On top of that, the above measurements make clear that the devices on which transmission measurement can be performed are inevitably multimodal. In order to get some idea of the non-reciprocity of the data signal alone without using an OSA, the ASE content of the output signal is estimated by first measuring the ASE separately, thus without an input signal. This method is of course not entirely correct as it assumes that the presence of an input signal will not influence the amount of generated noise. However, when only using absolute optical power detection this is the best attempt.
Fig. 5.21 shows the measured transmitted power for both propagation directions corrected by substraction of the separately measured ASE power for both propagation directions [13]. The measurements have been performed in pulsed conditions (10% duty cycle, 100ns pulse width), and the input TM power was 2.5mW. The device is a 520μm long $Co_{50}Fe_{50}$ R2 sample with a ridge width of 7μm. The lower figure shows the forward to backward ratio corresponding to the measured transmitted TM powers of Fig. 5.21(a).
The measured forward to backward ratio for the transmitted TM signal is between 1.4 and 1.5. This corresponds to an isolation extinction ratio of about 3.5dB/mm. This result is close to the one observed in the nonreciprocal lasing experiments (see the last entry in the first row of Table 5.5). This seems to suggest that the injected TM light mainly excites a first (or even higher order) TM mode. In a way this might be counterintuitive. However, in the measurement procedure the position of the lensed fibre at the input facet is optimized by maximizing the output power. In view of the first (or higher order) lasing behaviour of the devices, it is very probably that this optimization will chiefly lead to the excitation of a first order TM mode. In this way, the observed non-reciprocity is supposedly an average of the ground and a first order mode effect[20]. A deeper analysis of higher order non-reciprocal effects will be the subject of Van Parys' Ph.D. work.
The total insertion loss of the device (coupling and Fresnel losses included) is about -17dB at 30kA/cm^2. Conservative guesses of the coupling loss (-3dB) and the Fresnel losses (-7dB) give an on-chip loss of 7dB. Again, a spectrally resolved measurement is needed to correctly analyze the Fabry-Perot resonances and the corresponding on-chip loss of the data signal. This corresponds to a forward on-chip loss of about 3.0/mm. This is much better than the net loss observed on the first generation devices. However, again this result is somewhat blurred by the presence of the first-order mode.

In any case, this very first non-reciprocal transmission measurement is a milestone result in the development of the novel integrated isolator concept. The evolution from a device that could only be characterized on the basis of its nonreciprocal ASE emission (with very high internal loss) towards a device that effectively acts

[20]The same applies for that matter to the results of Table 5.5. There the total output power will certainly have important contributions of longitudinal cavity modes of the ground TM mode that are just below threshold.



(a)



(b)

Fig. 5.21: (a): "Transmitted" TM peak power as a function of pulsed injected current with the applied magnetic field switched between ±500Oe. The plotted curves are actually the observed total output powers corrected with the separately measured ASE output powers without input signal; (b): $\frac{P_f}{P_b}$ corresponding to the "measured" powers of (a). The ratio is quasi-constant as it should be, since $\Delta\alpha$ is not current-dependent. Below 300mA, the transmitted powers are too low, leading to high fluctuations on the observed ratio.

as a stand-alone isolating device (albeit with an insertion loss that could do with improvement), is notable.

5.2.3.2 Annex: Processing 2

At the time of writing, a second processing run was initiated on the same InAlGaAs demonstrator material. The main purpose of this run was to aim at a further improvement of the device performance by eliminating the higher order mode effects of the first run. The characterization of this run was still ongoing when this work was finished, but the great success of the first measurements deserves a mention here. Of course in-depth treatment of these results will be part of the PhD thesis of Van Parys.

Processing details In the first run the contact processing was modified in order to avoid disastrous contact processing errors as observed in the first generation. However this modification was very conservative, and has led to unexpected higher order mode effects as extensively described in the previous section. The main problem is that a contact processing scheme has to be devised that will lead to an entirely covered ridge without running the risk that the contact material gets deposited next to the ridge, thus without the risks associated with having to align a mask that will define the electrical contacting regions. A logical way to achieve this is by defining the electrical contacts in a self-aligned manner. This can be done by first processing the metal contact stripes through lift-off and subsequently using the metal stripes as an etch mask for the ridge waveguide definition. The details of this processing scheme will be discussed in-depth by Van Parys [12].
In any case it is already clear that even though this modification of the contact processing might avoid higher order effects, it will come at the cost of its magnetic properties. Indeed, the width-to-length aspect ratio of the contacts will now be unavoidably as low as 0.001, i.e. a ratio of a few μm to a few mm. As a result, the remanence isolation ratio will be negligible compared to the saturation isolation. This is illustrated in Fig. 5.22

Preliminary results With this new processing it was expected that the devices would not be hindered by first order TM mode lasing. For a fully covered ridge the difference between the modal loss of the first order and ground TM mode would in principle be big enough to avoid the modal competition that has led to this higher order lasing. As a result it is expected that the observed non-reciprocity is chiefly originating from the ground TM mode. In view of the fundamentally stronger non-reciprocal effect for a ground mode combined with the fact that the MO coverage of the ridge is now close to 100% (as opposed to only 50%), it is expected to observe an increase of $\Delta\alpha$ towards the original design values of Table 3.9.
Again the processed devices were observed to achieve lasing, and preliminary characterization of the non-reciprocal difference in external differential efficiency proved indeed the expected $\Delta\alpha$-increase (see Fig. 5.23) [14]. The observed forward-to-backward ratio of 1.625 in a 520μm long device corresponds to an isolation extinction ratio 8.10dB/mm. This is an amazing improvement of a factor 4 as compared to the first generation devices, and is almost spot on target when com-

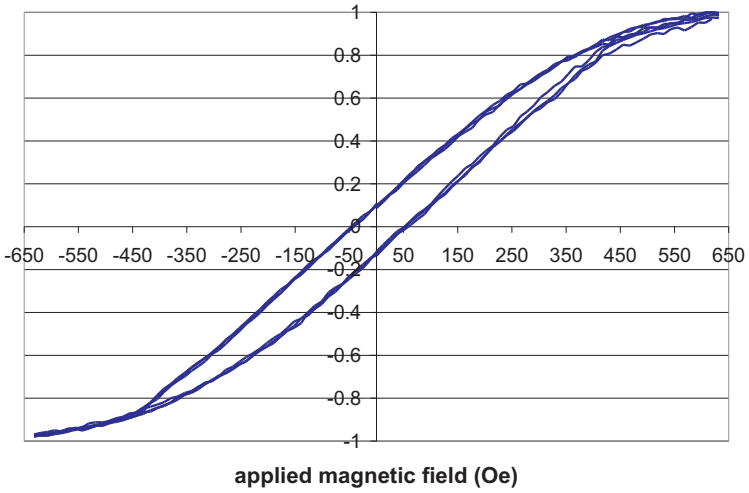


Fig. 5.22: Typical TM ASE hysteresis curve of the second self-aligned contact processing run. The remanence ratio is as low as 10%.

pared to the 1D slab design value of 9.55dB/mm.[21] A more in-depth study of this new optimization step of the device will be given in the Ph.D. work of Van Parys.

5.2.3.3 Conclusion

This concludes the characterization of the second generation devices. In every respect these devices have lived up to their expectations of an optimized performance for the SOA/isolator. The much stronger gain performance of the strained InAlGaAs quantum wells which was assessed prior to the design, has effectively led to an operation of the device close to forward transparency as confirmed by the observation of lasing and by the successful first-time transmission experiments. On top of that the more reliable design, precisely as a result of the better a priori insight into the gain performance of the InAlGaAs wells and of the MO strength of the CoFe alloys, has led to experimental results which are much better in accordance with the design values than was the case for the first generation devices. Admittedly, it is true that both statements paint maybe a too optimistic picture of the observations as described in the previous sections. Indeed, one might argue that the occurrence of lasing behaviour is precisely caused by the presence of low modal loss higher order TM modes and that the observed non-reciprocities are anything but close to the design values. However, that the $\Delta\alpha$ of a first order mode in a partially MO-metal-covered ridge waveguide is a factor of 2–3 lower than the value designed for ground TM slab modes, is anything but illogical. On the contrary, the recent observation of nonreciprocal ground mode TM

[21] ... were it not for the fact that that design value was for $Co_{90}Fe_{10}$...

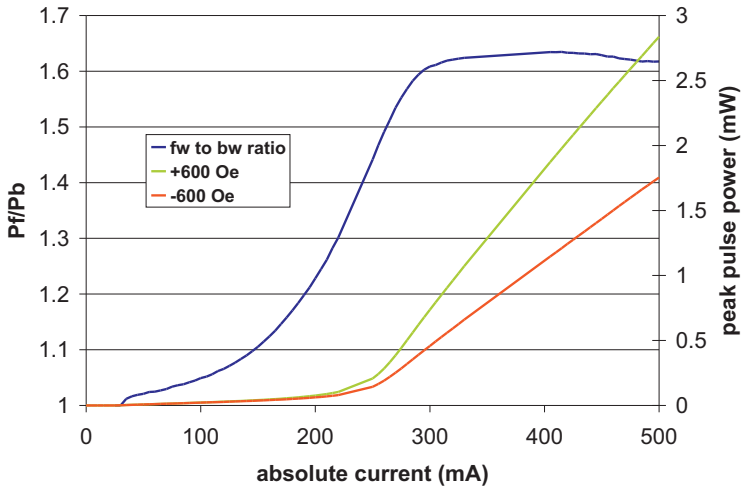


Fig. 5.23: Nonreciprocal lasing measurement on a 520μm long, R2 2.5μm SOA/isolator ridge waveguide with a self-aligned $Co_{50}Fe_{50}$ MO metal contact.

lasing with a nonreciprocal optical absorption difference of 8.1dB/mm obtained in the 2nd processing run, more or less confirms the above statements. Furthermore, that this latter result has been obtained after a second refinement of the contact processing scheme is a very convincing indication that the insight in the technological issues of these devices has matured and has efficiently tackled existing problems: from ill-processed 1st generation devices over improved but multimodal devices (with processing safety margins) to (almost) on-target monomodal devices with controlled contact processing.

5.3 Conclusion

This final chapter has described the assembly of the building blocks of the novel integrated isolator concept. It is the successful culmination of the theoretical framework of Chapter 2, the design work of Chapter 3, and the experimental assessment of the individual building blocks of Chapter 4.

A lot of attention has been devoted to the variety of characterization methods available to assess the performance of the SOA/isolator. From relatively simple fitting models for the description of the ASE emission of a nonreciprocal Fabry-Perot amplifier cavity well below or close to transparency, over the non-reciprocal lasing of such a cavity when it exceeds transparency, to a full-option dynamic measurement setup capable of characterizing the performance of the SOA/isolator from a transmission system viewpoint. All but the last method have been tried and tested on the processed devices and proven their usefulness. Especially the successful fitting of the ASE models to the measured curves was an important

observation in view of the relative simplicity of these models (in particular the simplifying assumptions made concerning the absence of saturation effects).

The experimental characterization of the two designed generations of SOA/isolators has revealed in the first place that, the technological realization of the SOA/isolator is not simply a matter of stacking a "strong" MO metal contact on top of a "strong" amplifying layer structure in a LEGO manner. The realization of a quasi-stratified, sufficiently wide metal contact (for good magnetic behaviour) that also entirely contacts a low-aspect-ratio-ridge (for optimal use of the available non-reciprocity and in order to avoid multimodality) is still an issue. For the moment the best device performance has been obtained using a sideway solution, namely processing self-aligned contacts by using the metal stripes as an etch mask for the ridge waveguides. The price that has to be paid is high: the quasi complete absence of a non-reciprocal effect in remanence. Of course all of the above might not be an issue if a buried layout is chosen for the SOA/isolator, or if the magnetization can be controlled by on-chip current paths...

A second important conclusion of the experiments is that the stumbling block of the concept is not so much the magneto-optic strength of the MO metal but rather its optical absorption or better the strength of the optical gain of the MQW region. After all re-assessment of the first generation devices revealed that their limited non-reciprocity could be more or less explained by a combination of poor processing and an overestimation of the TM confinement near the MO metal in a rather deeply etched ridge. However, their quite obvious failure to even slightly approach transparency was rather fundamental. The tensile-strained InGaAsP wells were simply not able of providing sufficient gain. That the main issue is indeed the MQW gain and not the MO metal is underlined by comparing the ideal theoretical design values of the strongest first and second generation designs:

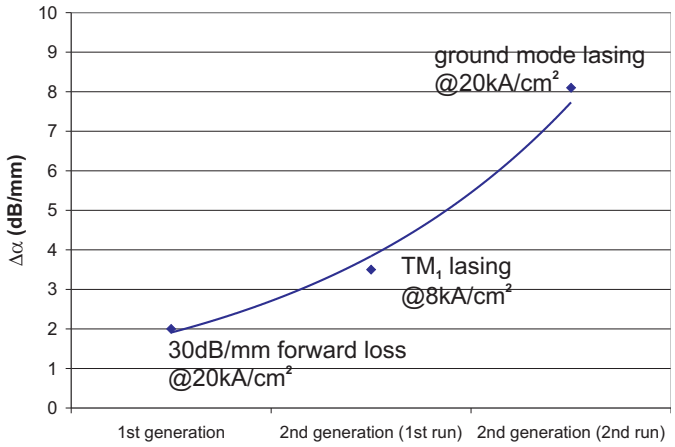


Fig. 5.24: Evolution of the characterized SOA/isolators.

9.3dB/mm for the reassessed "risk" 3QW InGaAsP design and 9.55dB/mm for the "R2" 9QW InAlGaAs design. The latter however was able to reach ground mode transparency (as confirmed by the latest measurements)!

While this work has up till now not yet led to the development of a competitive monolithically integrated InP-based waveguide isolator providing say 25dB isolation in a 2mm long device with an operation current not beyond 100mA, it is safe to say that the evolution from the first observation of a very hypothetical novel integrated non-reciprocity concept via TM ASE hysteresis measurements towards a nonreciprocal lasing device with a non-reciprocal absorption difference of 8.1dB/mm is remarkable. Figure 5.24 sketches the evolution of the best performing devices along the timeline of this research. While it is probably a bit overenthousiastic to use an exponential trendline, there are still several optimizations thinkable that might further improve the device. We will try to indicate a few in the following Chapter.

References

[1] M. Takenaka, and Y. Nakano, "Proposal of a novel semiconductor optical waveguide isolator." in the Proceedings of *11^{th} International Conference and Related Materials,* pp. 298–292, Davos, Switzerland, 1999.

[2] M. Vanwolleghem, W. Van Parys, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens and L. Lagae "Experimental demonstration of nonreciprocal amplified spontaneous emission in a CoFe clad semiconductor optical amplifier for use as an integrated optical isolator." *Appl. Phys. Lett.,* vol. 85(18), pp. 3980-3982, 2004.

[3] M. Vanwolleghem, W.F.H. Van Parys, D. Van Thourhout, R. Baets, F. Lelarge, O. Gauthier-Lafaye, B. Thedrez, R. Wirix-Speetjens, and J. De Boeck, "Experimental verification of a novel integrated isolator concept." in the Proceedings of *29^{th} European Conference on Optical Communication,* post-deadline papers 6, pp.78–79, Rimini, Italy, 2003.

[4] M. Vanwolleghem, W.F.H. Van Parys, D. Van Thourhout, R. Baets, O. Gauthier-Lafaye, F. Lelarge, B. Thedrez, R. Wirix-Speetjens, and J. De Boeck, "First Experimental Demonstration of a monolithically integrated InP-based waveguide isolator." in the Proceedings of the *2004 Optical Fibre Communication Conference (OFC 04),* p. TuE6, Los Angeles, USA, 2004.

[5] C. H. Henry, "Theory of spontaneous emission noise in open resonators and its application to lasers and optical amplifiers." *J. Lightwave Technol.,* vol. LT-4(3), pp. 288-297, 1986.

[6] H. A. Haus, and S. Kawakami, "On the excess spontaneous emission factor in gain guided laser amplifiers." *J. Quantum Electron.,* vol. QE-21(1), pp. 63-69, 1985.

[7] J. De Merlier, D. Van Thourhout, G. Morthier, and R. Baets, "Amplified spontaneous emission in index-guided multimodal waveguide structures." *J. Quantum Electron.,* vol. 39(9), pp. 1099–1105, 2003.

[8] J. De Merlier, G. Morthier, and R. Baets, "The influence of lateral carrier diffusion and surface recombination on the behavior of semiconductor optical amplifier (SOA)–based MMIs." *J. Quantum Electron.,* vol. 39(7), pp. 880–885, 2003.

[9] W. Van Parys, *private communication*, November 2004.

[10] K. Petermann, "Calculated spontaneous emission factor for double-heterostructure injection lasers with gain-induced waveguiding." *J. Quantum Electron.,* vol. QE–15(7), pp. 566–570, 1979.

[11] B. W. Hakki and T. L. Paoli, "Gain spectra in GaAs double-heterostructure injection lasers." *J. Appl. Phys.,* vol. 46(3), pp. 1299–1306, 1975.

[12] W. Van Parys, "Geïntegreerde Optische Isolatoren en Circulatoren Gebaseerd op de Interactie van Licht met Ferromagnetische, Gyrotrope Metalen." Ph. D. thesis (*in preparation*), Ghent University, Belgium, 2007.

[13] W. Van Parys, M. Vanwolleghem, D. Van Thourhout, R. Baets, J. Decobert, B. Dagens, B. Thedrez, R. Wirix-Speetjens and L. Lagae, "InP-Based Monolithically Integrated Optical Waveguide Isolator with 32 dB/cm Isolation." in the Proceedings of *17^{th} LEOS Annual Meeting,* p. TuY5, Puerto Rico, 2004.

[14] W. Van Parys, M. Vanwolleghem, D. Van Thourhout1, R. Baets, J. Decobert, B. Dagens, B. Thedrez, L. Lagae and R. Wirix-Speetjens "Demonstration of 81dB/cm Isolation on an InP-based Optical Waveguide Isolator." submitted to the *12^{th} European Conference on Integrated Optics,* Grenoble, France, 2005.

6

Conclusions and Outlook

IN this work a novel integrated optical waveguide isolator has been studied and successfully realized. The main conclusions of this research are summarized here, and some directions for further research are indicated.

6.1 Conclusions

The suggested optical isolator concept of Nakano and Ando may be able to solve most of the performance and integration problems that current state-of-the-art planar optical waveguide isolators are facing. Because this design is essentially a non-reciprocal SOA with a magneto-optic metal contact (operated in forward transparency), it allows a direct monolithical integration with a III-V based laser or amplifier, without the need for active-passive integration schemes nor for an interferometric layout. Nevertheless, this concept heavily relies on two previously unassessed experimental assumptions: a very strong TM-favouring gain region is needed and ferromagnetic metal/semiconductor contact must be available that has at the same time good Ohmic properties, strong magneto-optic gyrotropy in the IR spectrum, and strong in-plane magnetic anisotropy. Therefore, at the very outset of this work, it was far from clear whether a practical realization of this novel concept was feasible.

This work has sketched in a detailed manner the road we have followed from theoretical concept to experimental demonstrator. Starting by an in-depth study of the modelling of non-reciprocal optical waveguides, and magneto-optical waveguides in particular, we have developed the theory of two approaches to model this type of waveguides, and implemented it in two numerical modelling tools. An important conclusion of this work was that even with the very strong gyrotropy parameters of ferromagnetic metals, a perturbation-theory-based modelling of the device is still valid. On top of that, the benchmark calculations that led to this conclusion have made clear that at least from a numerically viewpoint the isolation performance predicted by Nakano is realistic.

An in-depth study of the subcomponents needed for a successful realization of this device has been performed. The subtle methodology necessary for the characterization of the gyrotropy constant of a magneto-optic material has been elaborated. It was indicated why it is quasi impossible to find a suitable MO material purely on a first-principles theoretical basis alone. This inevitably necessitates the magneto-optic ellipsometric characterization of each new possible candidate material. There are however some vague theoretical indications that the CoFe alloy system is the best suited.
The study of the different mechanisms influencing the magnetic anisotropy of the thin film ferromagnetic metal contact revealed that due to the specific function to be fulfilled by the metal in this precise device and due to its very high length-to-width aspect ratio, it is very difficult to counteract the fundamental shape anisotropy by pre-treatment measures. Only deposition under the presence of a magnetic bias field is available to somehow induce the desired in-plane magnetic anisotropy. A simple design rule has therefore been provided to minimize the influence of the shape anisotropy.
Finally, a precise methodology has been developed to accurately characterize the contact resistivity of a metal/semiconductor contact. It was found that a CKR structure should in principle always be preferred above any other characterization

structure for a reliable extraction of the contact resistivity.
An extensive review has been given of the theoretical aspects of incorporating strain (specifically tensile strain) in amplifying MQW regions. This revealed the many advantages of tensile strained QW's as compared to compressive strained wells. Purely on a theoretical basis it is expected that Al-based wells should behave better than the more traditional P-based III-V semiconductor compounds.
Two generations of layer structures for the MQW SOA/isolator have been designed using the developed modeling tools. While the input data were only partially available, an important conclusion was that it is very likely that this type of isolator design will invariably operate closely to the gain limit of the wells. The original predictions of Nakano have been cautiously adjusted: a 25dB isolator with a length of 2–3 mm and an operation current density of 5kA/cm^2 seems a more appropriate prediction.

The experimental behaviour of the subcomponents has been one-by-one assessed. The expected strong gyrotropy of CoFe alloys has been confirmed experimentally. The observed magnetic anisotropy of CoFe stripes with a high length-to-width aspect ratio and sputtered deposited in the presence of a bias field, is limited but acceptable if a sufficiently wide contact is used. As expected, a good Ohmic behaviour is rather easily obtained by using very high doping levels in the semiconductor contact layer. It was however remarkable to observe that by adding only a very thin low band gap ternary contact layer on top of a transparent thick quaternary layer, a resistivity comparable to that of standard (but highly absorptive) ternary contact layers could be obtained.
A thorough study of tensile-strained MQW material has revealed that somehow as theoretically expected the Al-based wells greatly outperform the P-based tensile strained wells. On top of that, the measurements seem to indicate that the strain limit has not been reached yet and that higher strains might further improve the gain performance. The best obtained results on InAlGaAs wells achieve levels of gain that are in accordance with what is needed for forward transparency according to the design calculations. Still again it is noted that in any case the wells will need to be operated at rather high carrier densities.

The final step in this work was the successful assembly of the individual building blocks into the first worldwide version of the novel SOA/isolator concept and its characterization. It is the successful culmination of the preparatory work: going from the development of the theoretical modelling framework, through the design of the device, to the experimental assessment of its subcomponents. A methodology for the characterization of this type of active non-reciprocal devices has been developed. The most important conclusion is the successful observation of the expected optical non-reciprocity. However, it was noted that the assembly of the building blocks is confronted with some technological issues, that are at this moment not entirely resolved, namely the realization of a quasi-stratified sufficiently wide metal contact that entirely covers a low-aspect ratio ridge waveguide. Another important conclusion is that the ultimate isolator performance is not so much hindered by the strength of the gyrotropic material, but more by the

gain performance of the strained MQW regions. Small increases in the gyrotropy constant of different metals might still be found, but it is not expected that the optical absorption caused by the metal will vary drastically. As a result, while it is expected that a well-processed device might always easily reach non-reciprocal absorption differences of more than 100dB/cm, the main issue is whether this difference can be "transferred" to a forward transparency operation point. In any case, the evolution observed in going from the very first observation of the effect on ill-processed InGaAsP-devices, that could not be operated anywhere close to transparency, to strong non-reciprocal monomodal lasing in InAlGaAs devices, not only proves that the concept of Nakano is feasible, but also underlines even more that there is a clear understanding of how to gradually improve its performance with each new generation.

6.2 Outlook

This work can be considered as the first step in the experimental realization of a novel revolutionary integrated optical isolator. The device is far from being optimized and the ultimate goal of integrating it with a DFB laser has even not been started yet. On top of that, there are several important device parameters that haven't been assessed yet. The most important of those is the dynamic behaviour of the SOA/isolator, more in particular its response towards a modulated signal (both in terms of bit error rate and decrease of extinction ratio). This are therefore the three most important paths along which future work will proceed (and that in the mean time is already ongoing by means of the successoral Ph.D. research of Van Parys): a characterization of the behaviour of the device when inserted in a realistic transmission link, a further improvement of the general performance, and the fabrication of an integrated DFB laser monolithically integrated isolator (with considerable attention towards minimization of the coupling losses to a fibre mode).

At this moment, the best reported device can achieve 25dB isolation in a length of 3mm but needing an unrealistically high transparency current close to 1A. In order to improve this we see several possibilities.

- The most logical improvement seems to be to aim at an increase in modal gain provided by the strained MQW region. Achieving this by increasing the number of wells is probably not the way to proceed. The measurements on the MQW lasers indicate that with an increasing number of wells, one runs the risk of decreasing internal quantum efficiencies. Next to that, the total current scales with the number of wells, while it is not certain that the MQW confinement will scale with the same proportionality (especially not at a high number of wells). Moreover, aiming at an increase in optical MQW confinement will go alongside of a decrease of confinement near the MO metal and thus a decrease of effect. A more interesting way to proceed, is to try MQW regions with a higher tensile strain.

- The use of a buried layout might solve many problems, both technologi-

cally (probably easier to process a good stratified wide MO contact) and theoretically (better lateral confinement and higher internal efficiency).

- It is always interesting to keep on looking for alternative MO materials. These latter do not necessarily have to be metallic. In a buried layout the current can always be injected in a lateral manner. The most important aspect is that the MO material can be easily deposited on top of III-V semiconductors (in other words that no techniques such as wafer bonding are needed).

- Think of alternative layouts that exploit the non-reciprocal absorption difference. One very interesting concept is that of a ring resonator, with a non-reciprocal absorbing part inserted. The resonance effect can lead to a total non-reciprocity which is much higher than the non-reciprocal absorption difference of one round-trip. Another resonant-like behaviour occurs in photonic crystals. The study of magneto-photonic crystal waveguides is an entirely new research field that has only very recently seen the light of day. The exciting possibilities that might crop up by combining magneto-optics and photonic crystals are still difficult to predict.

On a more theoretical and academic scale, interesting future work can also be found in the extension of the developed rigorous modelling tools towards full 3-D capabilities. This is of course very ambitious. On a less ambitious level, it is not a bad idea to spend some more attention to the modelling of the carrier diffusion inside the active device, and spend some more attention on a better understanding of multimodal effects. Of course these latter modelling tasks are less important when a buried layout is used. Finally, there is the work on extending the wavelength window towards 1.55μm and looking for ways to make the device also operational for the TE-polarization.