Diss ETH No. 17333

Instrument Systems for Imaging Spectro-Polarimetry

Alex J. Feller

Cuvillier Verlag Göttingen

Diss ETH No. 17333

Instrument Systems for Imaging Spectro-Polarimetry

Diss ETH No. 17333

Instrument Systems for Imaging Spectro-Polarimetry

A dissertation submitted to the

ETH Zürich

for the degree of
Doktor der Naturwissenschaften

Presented by

Alex J. Feller

Dipl. Phys. ETH
born September 18, 1977
from Luxembourg

accepted on the recommendation of

Prof. Dr. J. O. Stenflo, examiner
Dr. S. V. Berdyugina and
Dr. A. Gandorfer, co-examiners

Zurich, 2007

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.
1. Aufl. - Göttingen : Cuvillier, 2008
Zugl.: (ETH) Zürich, Univ., Diss., 2007

978-3-86727-554-5

© CUVILLIER VERLAG, Göttingen 2008
Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung
des Verlages ist es nicht gestattet, das Buch oder Teile
daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie)
zu vervielfältigen.
1. Auflage, 2008
Gedruckt auf säurefreiem Papier

978-3-86727-554-5

— *To my parents* —

Abstract

This thesis is devoted to the development and the application of novel instrument systems and measuring techniques in the field of solar polarimetry. The work is split into three parts, each part addressing an independent project in terms of scientific objectives and implementation.

In the first part I describe the development and implementation of a tunable narrow-band filter system to be used in combination with the Zurich Imaging Polarimeter for quasi-monochromatic imaging vector polarimetry. Its working range covers the visible spectrum between 3900 Å and 6600 Å. The bandwidth varies between 20 mÅ and 120 mÅ, depending on wavelength and optical setup. The main filter components are two lithium-niobate Fabry-Pérot interferometers which can be tuned by applying a voltage or by changing the temperature. In the current version of the instrument the interferometers are used in series in a single-pass tandem configuration. We have designed and implemented two complementary optical setups for which the filter system can be optionally employed. The telecentric setup aims at observations with high spatial resolution. It has been constructed to be easily transportable and to be used at foreign telescopes like the Swedish Solar Telescope (SST), one of the world's leading telescopes in terms of imaging performance. The collimated setup has been designed to have a large field of view and higher throughput. It is permanently installed at the Gregory-Coudé telescope of the Istituto Ricerche Solari Locarno (IRSOL). The filter system will allow us to explore the spatial structuring of the polarized solar spectrum, including vector mapping of the Hanle and Zeeman effects in any spectral line. First polarimetric images recorded at the SST and at IRSOL in the spectral lines of Ca I (4227 Å), Sr I (4607 Å) and Hα (6563 Å) show that the instrumentation and its operation is technically mature. The filter system is now available for scientific observations.

The second part deals with a laboratory experiment linked to a solar polarization anomaly that has remained an unresolved enigma for many years. While scattering processes usually produce polarized radiation, certain spectral lines, in particular the Fraunhofer D_1 line of neutral sodium at 5896 Å, which has electronic angular momentum quantum number $J = 1/2$ for both its lower and upper states and a nuclear spin of 3/2, should be intrinsically unpolarizable according to the standard quantum mechanical scattering theory, even when accounting for the hyperfine structure splitting. In contrast, observations in quiet regions on the Sun reveal a polarization peak centered in the cores of both the Na I D_1 line and the Ba II 4934 Å line, which has the identical quantum-numbers (including nuclear spin) as the sodium D_1 line. As the solar atmosphere is a complex environment with an optically thick medium permeated by a tangled and fractal-like magnetic field, we have set up a laboratory experiment with a sodium vapor cell to obtain an answer to the question: Is the D_1 anomaly a problem for atomic or for solar physics? The experimental results are consistent with the theoretical null prediction of atomic physics. The measurement uncertainty gives an upper limit for the polarization ratio between the D_1 and D_2 lines which is smaller than the ratio observed on the Sun. Although this result does not provide the definitive answer to the problem, it permits us to conclude that the D_1 anomaly has to involve the effect of magnetic fields.

The third project included in this thesis is the development of a dedicated instrument to capture the polarization of the chromospheric emission spectrum (flash spectrum) during the total solar eclipse of March 29, 2006. The whole visible range and small overlaps with near UV and near IR have been covered with a spectral resolution of 3000. The minimum sampling period was chosen to be 20 ms, which corresponds to a height resolution of some 10 km in the solar atmosphere, along the direction of the lunar movement. The instrument is composed of an 8-inch Dall-Kirkham type telescope and a slitless spectropolarimeter. The measurements are only feasible during an eclipse, since seeing and stray light prohibit a clean isolation of the chromospheric light with a coronagraph. The flash spectrum has been successfully recorded during the above-mentioned eclipse in the Sahara desert in southern Libya. For several spectral lines the noise level in the degree of polarization is of order 1% which allows us to compare with theoretical predictions. However, it is tricky to correct for the systematic errors because of optical aberrations in the observed spectra.

Zusammenfassung

Die vorliegende Doktorarbeit ist der Entwicklung und Anwendung von neuartiger Instrumentierung und neuen Messtechniken auf dem Gebiet der solaren Polarimetrie gewidmet. Die Arbeit ist in drei Teile gegliedert. Jeder Teil befasst sich mit einem unabhängigen Projekt was die wissenschaftliche Zielsetzung und deren Umsetzung angeht, auf instrumenteller Ebene gibt es aber viele Überschneidungen.

Im ersten Teil beschreibe ich die Entwicklung und die Implementierung eines durchstimmbaren optischen Schmalbandfilters. Verknüpft mit dem “Zurich Imaging Polarimeter”, einem hochpräzisen Instrument für Polarisationsmessungen, ist damit quasi-monochromatische bildgebende Polarimetrie von der Sonne möglich. Der spektrale Arbeitsbereich des Filtersystems deckt das gesamte sichtbare Spektrum zwischen 390 nm und 660 nm ab. Die Bandbreite beträgt zwischen 2 und 12 pm, je nach Wellenlänge und optischem Aufbau. Die Hauptkomponenten des Filtersystems sind zwei Fabry-Pérot Interferometer aus Lithiumniobat welche durch Anlegen einer elektrischen Spannung oder über die Temperatur abgestimmt werden können. In der aktuellen Version des Instruments werden die Interferometer in Serie verwendet und im sogenannten “single-pass”, d.h. das Licht geht einmal durch jedes Interferometer hindurch. Wir haben zwei sich ergänzende optische Aufbauten realisiert, in denen das Filtersystem wahlweise eingesetzt werden kann. Der telezentrische Aufbau ist für hohe räumliche Auflösung ausgelegt. Er wurde so entworfen, dass er leicht zu transportieren ist und daher an ausländischen Teleskopen eingesetzt werden kann, wie z.B. dem Schwedischen Sonnenteleskop (SST), einem der weltweit führenden Teleskope in punkto Abbildungsleistung. Der kollimierte Aufbau dagegen wurde für ein grosses Gesichtsfeld und eine höhere Lichtstärke ausgelegt und ist dauerhaft am Gregory-Coudé Teleskop des Istituto Ricerche Solari Locarno (IRSOL) installiert. Das Filtersystem erlaubt uns die räumliche Verteilung von Streupolarisation im sogenannten zweiten Sonnenspektrum zu studieren insbesondere auch in magnetisch aktiven Regionen mittels des Hanle Effekts. Darüber hinaus erlaubt es auch Zeeman-Diagnostik. Erste Polarisationsbil-

der, aufgenommen am SST und am IRSOL in den Spektrallinien von Ca I (4227 Å) und Sr I (4607 Å) und Hα (6563 Å), belegen, dass das Instrument in seiner ersten Version technisch ausgereift und für den wissenschaftlichen Betrieb bereit ist.

Der zweite Teil dieser Arbeit beschäftigt sich mit einem Laborexperiment das durchgeführt wurde um einer seit Jahrzehnten unverstandenen Anomalie im zweiten Sonnenspektrum auf die Spur zu kommen. Kohärente Streuprozesse erzeugen im Normalfall polarisiertes Licht. In einigen Fällen aber, wie z.B. beim Natrium D_1 Übergang mit Drehimpulsquantenzahlen $J = 1/2$ für oberen und unteren Zustand und Kernspin 3/2, sagt die quantenmechanische Streutheorie null Polarisation voraus, und zwar auch wenn man die im gemessenen Sonnenspektrum sichtbare Hyperfeinstruktur-Aufspaltung mit hinzurechnet. Im Gegensatz zur Theorie zeigen Beobachtungen auf der ruhigen Sonne ein deutliches Polarisationssignal im Zentrum der D_1 Linie. Einen weiteren Hinweis auf eine tatsächlich vorhandene Anomalie liefert die Beobachtung eines ähnlichen Signals in der Ba II 4934 Å Linie, ein Übergang mit genau den gleichen Quantenzahlen wie Natrium D_1. Die Sonnenatmosphäre ist allerdings eine komplexe Umgebung, optisch dick und durchdrungen von einem verworrenen fraktalartigen Magnetfeld. Darum haben wir ein besser kontrollierbares Laborexperiment mit einer Natriumdampfzelle aufgebaut, welches dabei helfen soll die folgende Frage zu klären: ist die Natrium D_1 Anomalie ein Problem der Atom- oder der Sonnenphysik? Es zeigt sich, dass die Messungen im Labor vereinbar sind mit dem theoretisch vorhergesagten Nullresultat. Desweiteren ist die Obergrenze für die Polarisierbarkeit von D_1, gegeben durch die Messgenauigkeit, kleiner als der aus den Sonnenbeobachtungen abgeleitete Wert. Obwohl diese Aussage das Problem selbst nicht zu lösen vermag, so lässt sie doch den Schluss zu, dass die Lösung etwas mit solaren Magnetfeldern zu tun haben muss.

Das dritte in dieser Arbeit behandelte Projekt ist die Messung der Polarisation des chromosphärischen Emissionsspektrums (Flash Spektrum) während der totalen Sonnenfinsternis vom 29.3.2006. Das eigens für diesen Zweck entwickelte Instrument deckt etwas mehr als den gesamten visuellen Wellenlängenbereich ab, mit einer spektralen Auflösung von etwa 3000. Die kleinste Integrationszeit beträgt 20 ms was in Richtung der Mondbewegung einer Auflösung von etwa 10 km in der Sonnenatmosphäre entspricht. Das Instrument besteht im Wesentlichen aus einem 8-Zoll Spiegelteleskop vom Dall-Kirkham Typus und einem spaltlosen Spektropolarimeter. Die Messungen sind nur während einer totalen Sonnenfinsternis machbar weil Seeing und Streulicht keine saubere Isolation des gesamten chromosphärischen Spektrums in einem Koronographen erlauben. Das Flash Spektrum wurde von einem Beobachtungsstandort in der Libyschen Sahara aus erfolgreich aufgezeichnet. Für einige Spektrallinien ist das Rauschen im Polarisationsgrad mit etwa 1% prinzipiell tief genug um die Daten mit theoretischen Vorhersagen vergleichen zu können. Optische Aberrationen im gemessenen Spektrum erschweren jedoch die saubere Abtrennung der systematischen Fehler sehr und die vorläufigen Resultate lassen noch keine eindeutige Aussage zu.

List of Symbols

The symbols are alphabetically ordered with latin letters before greek and capitals before lower-case letters. The same symbol is listed more than once if it appears with more than one meaning. The symbols denote a general physical quantity or concept. Often they are modified by superscripts or subscripts, in order to specify what they pertain to in a certain context. All contexts are not explicitely listed here, instead the page number indicates where in the book the symbol is first defined or used.

A	intensity absorption	10
A	area	42
A	parameter of the lens Mueller matrix	80
a	initial quantum substate	115
a	continuum opacity	124
B	Planck function	42
B	parameter of the lens Mueller matrix	80
b	intermediate quantum substate	115
b	continuum polarization	124
$\mathscr{C}$	calibration matrix	68
C	parameter of the lens Mueller matrix	80
C	polarizability coefficient function	123
c	speed of light	42
c	polarizability coefficient	117
c	reduced calibration matrix	69
d	geometrical distance or thickness	10
d	piezoelectric coefficient	30
D	pupil diameter	21
D	parameter of the lens Mueller matrix	80
E	electrical field amplitude	10
E	atmospheric extinction	43
E	parameter of the lens Mueller matrix	80
$\mathscr{E}_{11}$	Mueller matrix with first component equal to unity	117

$\boldsymbol{e}$	unit vector	115
F	finesse	11
F	total angular momentum quantum number	109
F	integrated flux	151
$f/\#$	f-ratio	13
f	focal length	21
f	wave phase error amplification function	61
f	final quantum substate	115
f	flux	151
f_{ab}	absorption oscillator strength	118
g_a	statistical weight of quantum state a	119
H	Helmholtz-Lagrange invariant	21
h	image size or height	21
h	Planck constant	42
I	intensity	10
I	nuclear spin quantum number	118
i	normalized intensity	69
J	electronic angular momentum quantum number	108
K	multipole index	117
k	collisional depolarization factor	125
$\mathscr{L}$	Mueller matrix of a lens	80
$\mathscr{M}$	Mueller matrix of a mirror	80
$\mathscr{M}$	general Mueller matrix (combined with subscript)	116
n	refractive index	9
n	integer used for numbering	53
$\mathscr{O}$	modulation matrix	67
o	normalized modulation matrix	69
P	polychromatic point-spread function	64
$\mathscr{P}$	scattering phase matrix	116
p	model parameter of Fabry-Pérot transmission	33
p	monochromatic point-spread function	61
p	degree of polarization	123
Q	Stokes Q parameter	48
q	short form for Q/I	48
q_{eff}	quantum efficiency of a detector	151
R	intensity reflectivity	10
R	parameter of free mirror model	80
$\mathscr{R}$	Mueller matrix of a rotation	80
r	electrooptic coefficient	30

r	radius	62
$\hat{\boldsymbol{r}}$	position vector operator	115
S	surface	9
$\boldsymbol{S}$	Stokes vector	67
$\boldsymbol{s}$	normalized Stokes vector	69
T	intensity transmission	10
T	temperature	30
$\mathscr{T}$	telescope matrix	67
t	time	69
$\mathscr{U}$	component matrix of singular value decomposition	70
u	short form for U/I	48
V	voltage	30
$\mathscr{V}$	component matrix of singular value decomposition	70
v	short form for V/I	48
v	wavelength in Doppler units	125
W	equivalent width	42
$\mathscr{W}$	component matrix of singular value decomposition	70
W_2	polarizability	117
x	coordinate value	61
y	coordinate value	61
γ	damping constant	115
Δ	finite difference (combined with another symbol)	11
δ	infinitesimal difference (combined with another symbol)	13
δ	parameter of free mirror model	80
ϵ_{opt}	total transmission of atmosphere and optics	151
θ	angle	9
κ	dispersion factor	11
λ	wavelength	9
μ	limb distance	43
μ	calibration scaling factor	70
ν	frequency	121
ρ	spectral resolution	12
σ	standard deviation	154
τ	amplitude transmission	10
Φ	phase of a wave	61
Φ	frequency profile function	116
ϕ	phase of Fabry-Pérot etalon transmission	10
Ω	solid angle	151
ω	angular frequency	115

List of Abbreviations

CCD	Charge Coupled Device
CTE	Cavity Thickness Errors
cw, ccw	clockwise, counter-clockwise
EW	Equivalent Width
FWHM	Full Width at Half Maximum
FOV	Field Of View
FP	Fabry-Pérot
FSR	Free Spectral Range
IBIS	Interferometric BIdimensional Spectrometer
IMaX	Imaging Magnetograph eXperiment
IP	Imaging Polarimeter
IPM	Italian Panoramic Monochromator
IRSOL	Istituto Ricerche SOlari Locarno
MTF	Modulation Transfer Function
PSF	Point-Spread Function
S/N	Signal-to-Noise ratio
SST	Swedish Solar Telescope
SVD	Singular Value Decomposition
TESOS	Triple (or Telecentric) Etalon SOlar Spectrometer
THEMIS	Téléscope Héliographique pour l'Etude du Magnétisme et des Instabilités de l'atmosphère Solaire
TNF	Tunable Narrow-band Filter
USAF	United States Air Force
VIM	Visible Imager and Magnetograph
VTT	Vacuum Tower Telescope
WPE	Wave Phase Error
ZIMPOL	Zurich IMaging POLarimeter

Contents

Abstract ix

Zusammenfassung xi

List of Symbols xv

List of Abbreviations xix

Part I Narrow-band imaging polarimetry 1

1. Introduction 3

2. The basics of Fabry-Pérot etalons 9
  2.1 Multiple beam interference 9
  2.2 Characteristics of the transmission spectrum 10
  2.3 Transmission for different ray geometries 12
    2.3.1 Parallel ray bundle 12
    2.3.2 Converging ray bundle 12
  2.4 Cavity thickness errors 13
  2.5 Characteristics of our etalons and prefilters 14

3. Optical setups 19
  3.1 Collimated versus telecentric 19
  3.2 Field of view versus spectral performance 21
  3.3 Optical setups chosen 23

4. Tuning and spectral calibration 29
  4.1 Tuning parameters and etalon control 29
  4.2 Voltage and temperature calibration 32
  4.3 Fine tuning and flatfield scanning 34

4.4 Tuning through the prefilter 35

5. Performance 41
5.1 Photometric sensitivity 41
5.1.1 Wavelength dependent elements 42
5.1.2 Wavelength-independent elements 44
5.1.3 Comparison with measured photon flux 46
5.2 Spectral and polarimetric performance 48
5.2.1 Performance parameters 48
5.2.2 Reflective parasitic light 49
5.2.3 Model of the tunable narrow-band filter 52
5.2.4 Results 56
5.3 Imaging performance 61

6. Polarimetric calibration 67
6.1 Polarimeter 68
6.1.1 General procedure 68
6.1.2 Different modulator types 70
6.1.3 PEM calibration 73
6.2 Telescope 77
6.2.1 Model of the Swedish Solar Telescope 78
6.2.2 Measurements and model fitting 81
6.2.3 Results 85

7. First observations 89

8. Conclusions and outlook 93

Bibliography 99

Part II Polarized scattering at sodium vapor 103

9. Introduction 105

10. Solar evidence for a fundamental problem 107
10.1 The Na I D_1 evidence 107
10.2 Comparison with related multiplets 110

11. Why the D_1 line is expected to be unpolarized 115
- 11.1 Dependence of polarizability on angular momentum 117
- 11.2 Hyperfine structure of D_2 and D_1 118
- 11.3 Exact cancellations in the D_1 system 120
- 11.4 Relation between W_2 and the polarized spectrum 121
- 11.5 Influence of collisions and magnetic fields 124

12. Laboratory experiment 127
- 12.1 Sodium cell 127
- 12.2 Experimental setup 129
  - 12.2.1 Light source 129
  - 12.2.2 Imaging optics 130
  - 12.2.3 Fabry-Pérot etalon 130
  - 12.2.4 Polarization optics and detector 131
- 12.3 Calibration of the light source 131
- 12.4 Absence of spurious signals 133
  - 12.4.1 Reflections off the glass walls 134
  - 12.4.2 Null polarization with horizontal entrance polarizer . . 134
  - 12.4.3 Temporal instabilities 135
- 12.5 Results 135
  - 12.5.1 Interpretation model 135
  - 12.5.2 Upper limit for the D_1 polarization 137
  - 12.5.3 Upper limit for the asymmetry of the D_2 phase matrix 139

13. Conclusions and outlook 141

Bibliography 143

Part III Polarimetry of the solar flash spectrum 147

14. Introduction 149

15. Instrument design 151
- 15.1 Photon budget 151
- 15.2 Telescope and spectropolarimeter unit 155
- 15.3 Instrument control and observing procedure 158

16. Conclusions and outlook 161

Bibliography 163

Acknowledgments 165

Curriculum Vitae 167

Part I

Narrow-band imaging polarimetry

Chapter 1

Introduction

During the the last decade the Institute of Astronomy of ETH Zürich has developed the powerful CCD-based Zurich Imaging Polarimeter (ZIMPOL), described in detail by Povel (1995), Gandorfer et al. (2004) and Gisler (2005).

Until now ZIMPOL has mainly been used in combination with a spectrograph (cf. the review of Stenflo, 2004). Figs. 1.1, 1.2 and 1.3 show some results obtained with the current ZIMPOL II system at the spectrograph of the McMath-Pierce Solar Telescope on Kitt Peak (Arizona, USA).

The new tunable narrow-band filter (TNF) will allow us to combine high-precision polarimetry with monochromatic imaging, thus providing a new scientific dimension to our work with ZIMPOL. Prior to the development of the TNF a first promising exploratory attempt in this direction has been made with the Universal Birefringent Filter at the Richard B. Dunn Solar Telescope on Sacramento Peak (New Mexico, USA). These observations have provided a first glimpse of the type of data that may be obtained with ZIMPOL in monochromatic imaging mode (Stenflo et al., 2002).

The multi-beam interference theory has been derived by the French physicist Charles Fabry in 1890-1892. During the period 1896-1898 Fabry and his colleague Alfred Pérot developed the first type of interferometer which is named after them today. Their most important paper describing the interferometer has been published in 1897 (Pérot & Fabry, 1897) and was followed by many joint publications describing its applications in metrology, spectroscopy and astrophysics (e.g. Pérot & Fabry, 1899). The Fabry-Pérot (FP) interferometer, also frequently called *etalon*[1], exists nowadays in solar physics in two different versions. The *air-spaced* version, the original version of Fabry and Pérot, consists of two separate reflective plates and is tuned by

---

[1] The expression comes from the French word *étalon* which can be translated with "gauge" or "standard". We adopt here the English spelling *etalon* found in most publications. Sometimes the word etalon is only used for the crystal version of the interferometer. We will however use it for both versions, air-spaced and crystal.

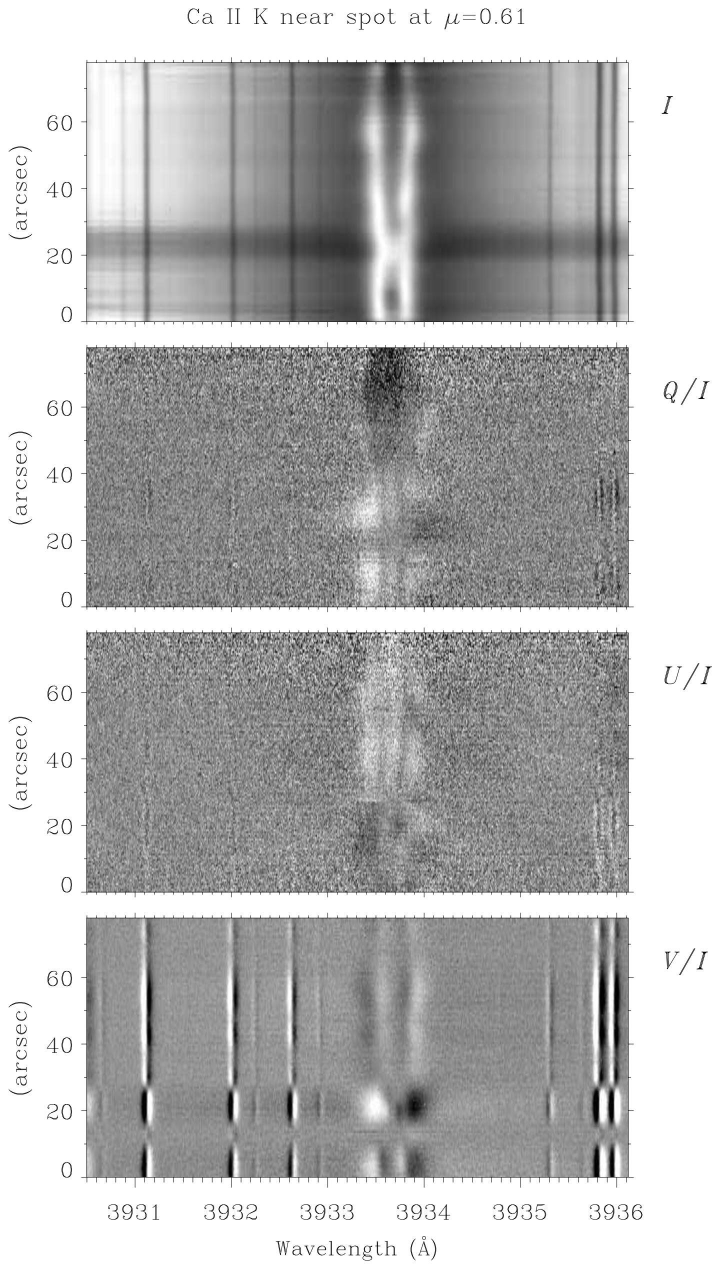


Fig. 1.1: Ca II K near spot (dark band in the I image, NOAA 10743). Limb distance $\mu = 0.61$. ZIMPOL, McMath-Pierce Telescope, March 11, 2005. Grayscales are ± 0.003 for Q/I and U/I, and ± 0.01 for V/I.

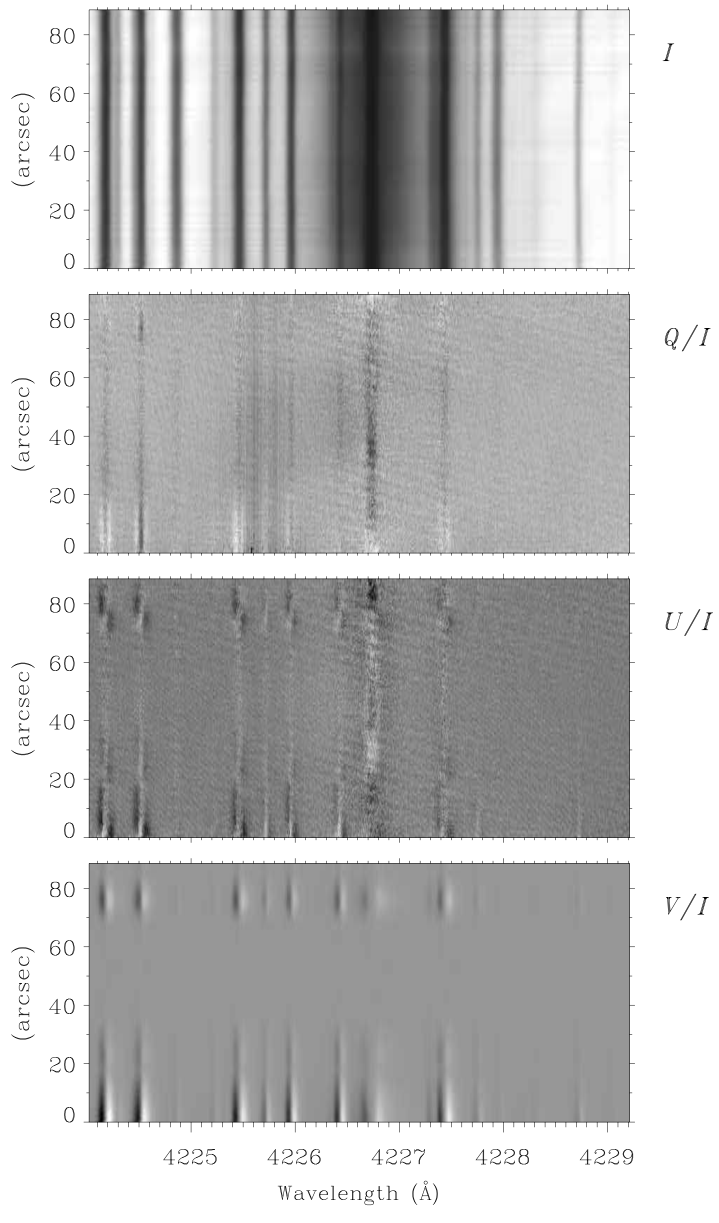


Fig. 1.2: Forward scattering and Hanle effect in Ca I. Measurement just below active region NOAA 10296, about 5 arcmin from disk center. ZIMPOL, McMath-Pierce Telescope, March 7, 2003.

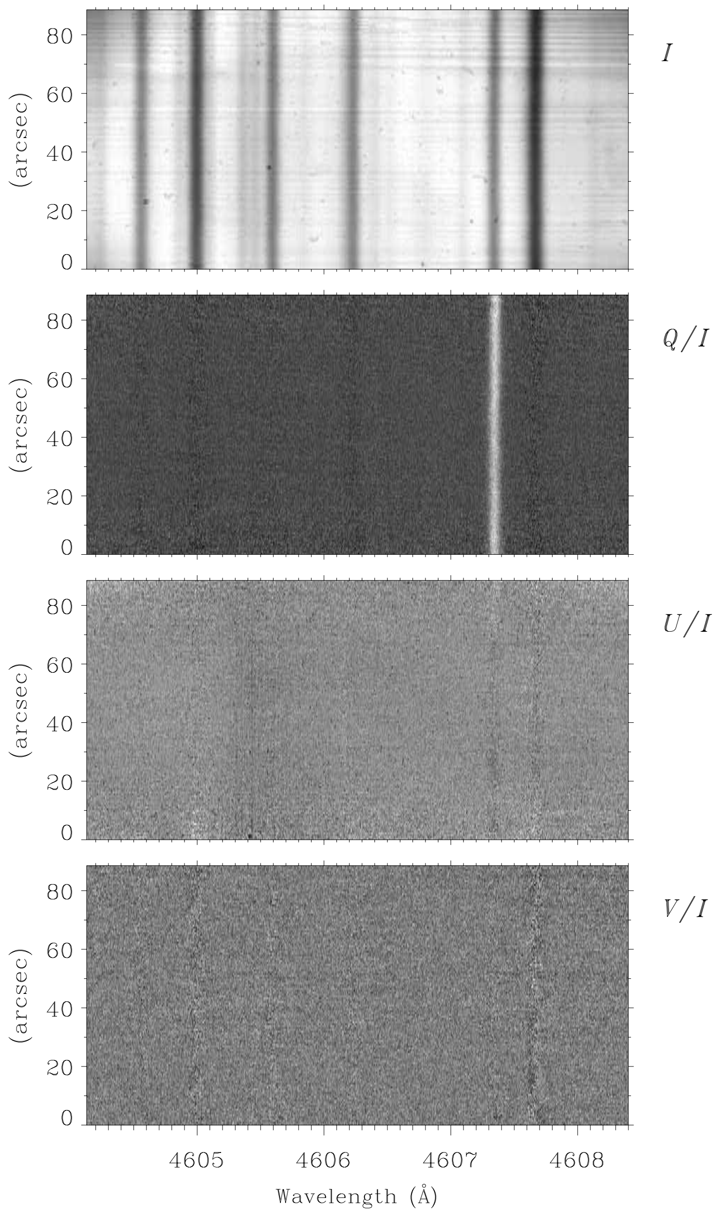


Fig. 1.3: Sr I, limb scattering at $\mu = 0.12$. ZIMPOL, McMath-Pierce Telescope, March 13, 2003. Q/I shows no spatial variations along the slit. One scientific goal for the new imaging polarimeter described in this part is to look for variations at the spatial scale of the granulation (1 arcsec) which is not resolved here.

varying the plate distance with the help of piezoelectric actuators. The *crystal* version consists of one piece of lithium niobate crystal, with polished and coated reflective surfaces. The tuning is done by varying the optical thickness of the crystal, either via thermal expansion or by altering the refractive index via an electrooptic effect.

During the last years, tunable narrow-band filter systems based on air-spaced Fabry-Pérot etalons have come into use at most major ground-based solar telescopes. In the following some references to current systems are given. Since 1998 the Kiepenheuer Institut is operating the TESOS instrument (Triple Etalon SOlar Spectrometer) at the German Vacuum Tower Telescope (VTT) at the Observatorio del Teide, Tenerife (Kentischer et al., 1998; von der Lühe & Kentischer, 2000; Tritschler et al., 2002), composed of three etalons in series in a telecentric optical setup. A two-etalon system in a collimated setup, the IBIS (Interferometric BIdimensional Spectrometer) from the Arcetri group (Cavallini et al., 2000; Cavallini, 2002; Cavallini et al., 2003) is in operation since 2003 at the Richard B. Dunn Solar Telescope on Sacramento Peak. In November 2006 it has been combined with a full-Stokes polarimeter and since then first Stokes images of active regions have been obtained with a spatial resolution close to the telescopic diffraction limit in Fe I, 6302 Å and Ca II, 8542 Å (Kleint, 2007). Another instrument developed by the Arcetri group is the IPM (Italian Panoramic Monochromator, Cavallini, 1998) at the French-Italian solar telescope THEMIS (Téléscope Héliographique pour l'Etude du Magnétisme et des Instabilités de l'atmosphère Solaire) at the Observatorio del Teide. In 2005 the institute of Astrophysics in Göttingen has upgraded (essentially renewed) their Göttingen FPI (Fabry-Pérot Interferometer). This system, based on two etalons in a collimated setup will become a post-focus instrument for the new German 1.5 m GREGOR solar telescope presently under construction at the Observatorio del Teide. Polarimetry with an earlier version of this instrument was performed by Volkmer et al. (1995), among others, using a Stokes V polarimeter.

Apart from ours, the only current system based on a lithium-niobate etalon is the IMaX (Imaging MAgnetograph eXperiment, Martinez Pillet & al., 2004) on-board the balloon-borne SUNRISE solar telescope which is foreseen to fly in 2008. Unlike our etalons the IMaX etalon is z-cut i.e. not birefringent. Moreover, IMaX consists of one single etalon used in double pass, in contrast to the serial combination of different etalons used in all other systems.

Chapter 2

The basics of Fabry-Pérot etalons

Here we compile the basic etalon equations which are needed in the next chapters. Moreover, we present the specific characteristics of our lithium-niobate etalons.

2.1 Multiple beam interference

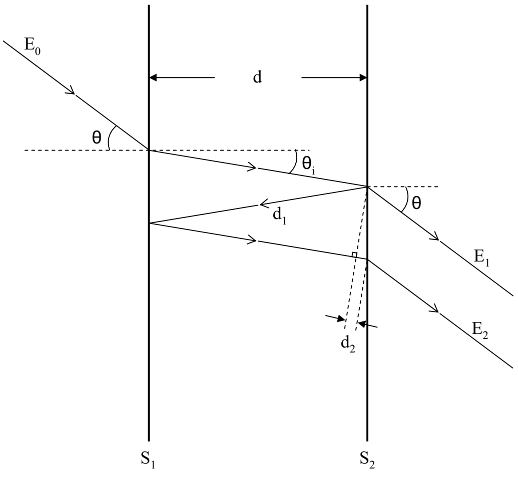


Fig. 2.1: Illustration of the beam interference in a Fabry-Pérot etalon.

In this section we derive general expressions for the amplitude and intensity transmission of a FP etalon. The beam interference is illustrated in Fig. 2.1. Let us track a monochromatic plane wave with wavelength λ, associated with the complex electrical field amplitude E_0, which is incident at outer angle θ to the surface S_1 of the etalon. Inside the etalon cavity with refractive index n the wave propagates at the inner angle θ_i with $\sin\theta_i = \sin\theta/n$

(Snellius). A fraction of the light exits the etalon at surface S_2 without any internal reflections and propagates as wave with amplitude

$$E_1 = T_s \sqrt{A}\, e^{i\phi_0} \tag{2.1}$$

T_s is the intensity transmission of the surface coatings[1] and A is the intensity absorption of the etalon cavity. ϕ_0 is the phase change induced by the etalon.

Another fraction of the light is reflected at surfaces S_2 and S_1 before exiting the etalon as wave with amplitude

$$\begin{aligned} E_2 &= A\, R\, e^{i\phi} \cdot E_1 \quad \text{where} && (2.2) \\ \phi &= \frac{2\pi}{\lambda} n\,(2d_1 - d_2) = \frac{4\pi n d \cos\theta_i}{\lambda}, && (2.3) \end{aligned}$$

R is the intensity reflectivity and d the cavity thickness. For the definition of d_1 and d_2 see Fig. 2.1. At this point the game repeats itself with a fraction of the light reflected a second time at S_2 and S_1 giving rise to an exiting wave with amplitude $E_3 = A\, R\, e^{i\phi} \cdot E_2$ and so on.

All these amplitudes add up *coherently* to produce a wave with total amplitude

$$E = \sum_{j=0}^{\infty} E_j = E_1 \sum_{j=0}^{\infty} \left(A\, R\, e^{i\phi}\right)^j . \tag{2.4}$$

This is a well-known geometrical progression. Together with (2.1) we finally obtain

$$E = \frac{T_s \sqrt{A}\, e^{i\phi_0}}{1 - A\, R\, e^{i\phi}}\, E_0 =: \tau E_0 \tag{2.5}$$

where τ is the amplitude transmission of the etalon. The transmission T for the intensity I can be directly found from the above equation with

$$\begin{aligned} I &= E\, E^* = \tau\, \tau^* E_0\, {E_0}^* = T\, I_0 \quad \text{with} && (2.6) \\ T &= \tau\, \tau^* = \underbrace{\frac{{T_s}^2\, A}{(1-AR)^2}}_{=:T_0} \; \frac{1}{1 + \frac{4AR}{(1-AR)^2} \sin^2 \frac{\phi}{2}} && (2.7) \end{aligned}$$

2.2 Characteristics of the transmission spectrum

The absorption of lithium-niobate is negligible for our purposes, so we assume $A = 1$ from now on. As can be seen from the above equation, T is periodic in

---

[1] Both surfaces, S_1 and S_2 are assumed to have the same transmission T.

ϕ. For each ϕ multiple of π the transmission is at its maximum T_0. Translated to wavelength this means that a transmission peak is reached for each

$$\lambda_k = \frac{2nd\cos\theta_i}{k} \tag{2.8}$$

with k integer.

To simplify the notation we assume normal incidence ($\cos\theta_i = 1$) in the following, unless otherwise stated. For oblique incidence replace nd with $nd\cos\theta_i$.

When we are using the etalon in a wavelength interval which is very small compared to the actual wavelength λ_0, the phase ϕ can be approximated to first order:

$$\frac{\phi}{2} \approx \frac{2\pi nd}{\lambda_0} - \frac{2\pi d}{\lambda_0}\left(\frac{n}{\lambda_0} - \frac{\partial n}{\partial\lambda}(\lambda_0)\right)(\lambda - \lambda_0) \tag{2.9}$$

The transmission spectrum appears to be periodic in λ, with period

$$\Delta\lambda_{\mathrm{FSR}} = \frac{{\lambda_0}^2}{2nd}\underbrace{\left(1 - \frac{\lambda}{n}\frac{\partial n}{\partial\lambda}\right)^{-1}}_{=:\kappa} \tag{2.10}$$

which is also called *free spectral range* (FSR). The factor κ takes into account the dispersion of the cavity material. Its contribution to the FSR is substantial for lithium-niobate etalons ($\kappa \approx 1.1$) whereas it can be neglected for air-spaced etalons.

Close to a transmission peak, T versus the peak distance $\Delta\lambda$ can be approximated as

$$T \approx \frac{T_0}{1 + \frac{4R}{(1-R)^2}\left(\frac{\pi}{\Delta\lambda_{\mathrm{FSR}}}\Delta\lambda\right)^2} = \frac{T_0}{1 + \left(\frac{2}{\Delta\lambda_{\mathrm{FWHM}}}\Delta\lambda\right)^2} \tag{2.11}$$

The FSR $\Delta\lambda_{\mathrm{FSR}}$ and the full-width of half maximum $\Delta\lambda_{\mathrm{FWHM}}$ are thus proportional and the proportionality factor

$$F = \frac{\Delta\lambda_{\mathrm{FSR}}}{\Delta\lambda_{\mathrm{FWHM}}} = \frac{\pi\sqrt{R}}{1-R} \tag{2.12}$$

called *finesse* only depends on the reflectivity R. This is an essential property of a FP etalon. The higher the reflectivity, the better the separation of the transmission peaks.

In fact, when R is close to 1, which is true for all etalons used in solar physics, the approximation (2.11) is good all over the free spectral range.

(2.11) is therefore a very convenient model for an isolated transmission peak, which is suitable to study various effects on the etalon spectrum, like, for instance, the effect of the ray geometry or the cavity thickness errors (see below).

The spectral resolution of a spectrograph is defined as the ratio $\lambda/\Delta\lambda$ of the wavelength and the distance $\Delta\lambda$ between to adjacent spectral features which can still be resolved. Different criteria defining $\Delta\lambda$ are actually in use, the most prominent one being the Rayleigh criterion. For purposes of comparison of different instruments the particular choice of criterion is of little importance. In this sense the spectral resolution of a FP etalon is commonly defined as

$$\rho = \frac{\lambda}{\Delta\lambda_{\mathrm{FWHM}}} \tag{2.13}$$

When the dispersion effects are left out one can say that ρ roughly scales with λ^{-1}.

2.3 Transmission for different ray geometries

In this section we will present the basic relations that are used later on to describe the etalon passband in different optical setups.

2.3.1 Parallel ray bundle

A parallel raybundle corresponds to a collimated optical setup which will be dealt with in chapter 3. For this ray geometry (2.12) exactly holds and the spectral resolution is at its maximum

$$\rho_0 = \frac{2ndF}{\kappa\lambda}. \tag{2.14}$$

When the parallel raybundle is incident at an oblique angle to the etalon surface the transmission spectrum is blueshifted by the amount

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta(nd\cos\theta_i)}{nd\cos\theta_i} \approx -\frac{1}{2}\theta_i^2 \tag{2.15}$$

2.3.2 Converging ray bundle

A converging raybundle with a symmetric angle distribution around the etalon surface normal corresponds to the telecentric setup discussed in chapter 3. According to (2.11) and (2.15) an isolated transmission peak produced by a ray at inner angle θ_i can be modeled as

$$T(\Delta\lambda, \theta_i) \approx \frac{T_0}{1 + \left(\frac{2}{\Delta\lambda_{\mathrm{FWHM}}}\right)^2 (\Delta\lambda + \lambda\theta_i{}^2/2)^2} \tag{2.16}$$

The complete transmission spectrum is obtained by integrating $T(\Delta\lambda, \theta_i)$ over the ray bundle i.e. between 0 and the principle ray angle θ_p (in air)

$$T(\Delta\lambda) = \frac{2n^2}{\theta_p{}^2} \int_0^{\theta_p/n} T(\Delta\lambda, \theta_i)\theta_i \mathrm{d}\theta_i \tag{2.17}$$

This integral has an explicit primitive:

$$T(\Delta\lambda) = \frac{T_0}{\Delta\phi} \cdot [\arctan(\phi + \Delta\phi) - \arctan\phi] \quad \text{with} \tag{2.18}$$

$$\phi = \frac{2\Delta\lambda}{\Delta\lambda_{\mathrm{FWHM}}} \quad \text{and} \tag{2.19}$$

$$\Delta\phi = \frac{\rho}{(2nf/\#)^2} \tag{2.20}$$

Here we have replaced θ_p by the more convenient f-ratio $f/\# = 1/(2\theta_p)$. The above expression describes a transmission peak which is broadened and slightly blueshifted.

2.4 Cavity thickness errors (CTE)

The above relations assume an ideal etalon in the sense that the optical thickness of the cavity is constant across the aperture. In reality though the cavity thickness slightly fluctuates because of inhomogeneities in the crystal and polishing errors. According to (2.8) these CTE δnd produce local wavelength shifts

$$\delta\lambda = \lambda\frac{\delta nd}{nd} \tag{2.21}$$

of the transmission spectrum over the etalon aperture. This results in performance losses as well in terms of imaging performance (cf. section 5.3) as in terms of spectral resolution. Here we take a look at the latter problem using the model

$$T = \frac{T_0}{1 + \left(\frac{2F}{\Delta\lambda_{\mathrm{FSR}}}\right)^2 \left(\Delta\lambda - \lambda\frac{\delta nd}{nd}\right)^2} \tag{2.22}$$

for a single transmission peak, derived from (2.11), (2.12) and (2.21).

For small CTE δnd, randomly distributed around 0, the transmission averaged over the etalon aperture is

$$\begin{aligned}\langle T\rangle &\approx \frac{T_0}{1+\left(\frac{2F}{\Delta\lambda_{\mathrm{FSR}}}\right)^2\left(\Delta\lambda^2+\left(\frac{\lambda}{nd}\right)^2\langle\delta nd^2\rangle\right)}\\ &= \frac{T_0\,(F_{\mathrm{eff}}/F)}{1+\left(\frac{2F_{\mathrm{eff}}}{\Delta\lambda_{\mathrm{FSR}}}\Delta\lambda\right)^2} \qquad (2.23)\end{aligned}$$

The average transmission is conveniently expressed in terms of an *effective finesse* F_{eff} which is a function of F, also called *reflective finesse* in this context, and the variance $\langle\delta nd^2\rangle$ of the CTE.

$$\frac{1}{{F_{\mathrm{eff}}}^2}=\frac{1}{F^2}+\underbrace{\frac{16}{\kappa}\,\frac{\langle\delta nd^2\rangle}{\lambda^2}}_{=:1/{F_d}^2} \qquad (2.24)$$

F_d in the above equation is called *defect finesse.*

The estimate (2.24) is good to get an idea of the magnitude of the CTE effect on the FP transmission. The true spectral profile of the transmission peak is a crucial property that we need to know very accurately for a reliable performance evaluation of the instrument. Atherton et al. (1981) have done an exhaustive analytical study of the CTE including lower order defects like the departure from parallelism and a spherical bowing of the cavity, which permit to refine the value for the effective finesse. As we dispose of thickness maps for our etalons we follow a more direct approach. We insert the maps into the model (2.22) and numerically compute the spectral profile. Fig. 2.2 shows the results.

At the end of this section we also want to point out that the equivalent width (EW) of the transmission peaks is not affected by the CTE. This can be easily understood by recalling that the aperture-averaged peak is obtained by a sum of local peaks with identical spectral profile, namely the profile corresponding to the ideal etalon.

2.5 Characteristics of our etalons and prefilters

Our TNF is composed of two lithium niobate FP etalons and a prefilter. The etalons have been manufactured by the Australian Center for Precision

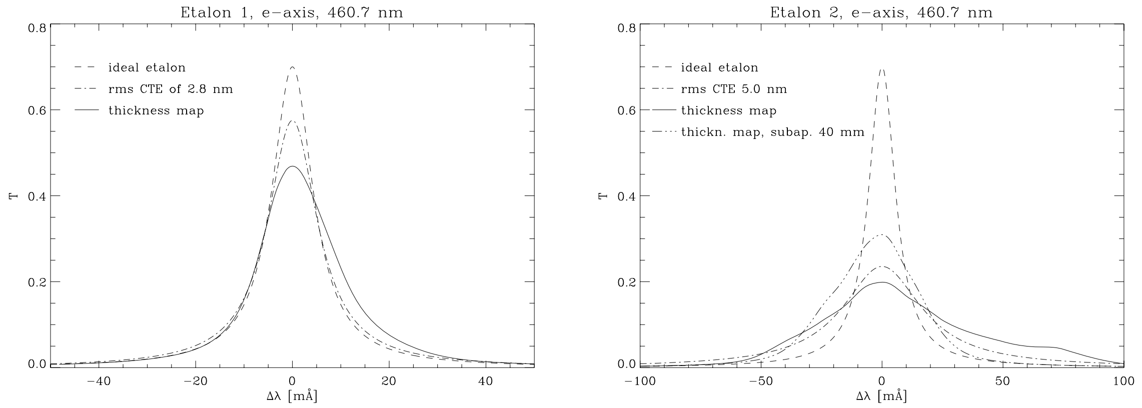


Fig. 2.2: Etalon transmission peaks, averaged over the aperture. Example for the e-axis and $\lambda = 460.7$ nm. The profiles denoted with "rms CTE" are computed from (2.23) and (2.24) using the square of the rms of the thickness maps for $\langle \delta n d^2 \rangle$. The transmission peak of etalon 2 (right panel) shows a very significant broadening when the full aperture is used. By choosing an adequate sub-aperture the profile can be improved. A sub-aperture of 40 mm is used in our collimated setup.

Optics. In the current version of the instrument the etalons are used in series in a single-pass tandem configuration. The requirements on the synchronization of the etalons and the performance of this configuration as well as the calibration and tuning are discussed in chapter 5. In this section we will restrict ourselves to a presentation of the individual etalons.

	Etalon 1	Etalon 2
Cavity material	LiNbO$_3$, Y-cut	
Cavity thickness	0.985 mm $\pm 0.6\vert 17$ nm	0.757 mm $\pm 0.5\vert 6$ nm
Spectral range	390 … 660 nm	
Peak transmission	30...70 %	
Average reflectivity	92.5 %	
Effective finesse	17 … 40	

Tab. 2.1: Main characteristics of our FP etalons.

Tab. 2.1 compiles the main etalon characteristics. Both etalons are made of Y-cut lithium niobate crystals which means that they are birefringent. The transmission spectra and tuning parameters are thus different for the two orthogonal states of polarization (ordinary and extraordinary rays). This allows the finesse to be enhanced by a special double-pass configuration where the polarization of the light is turned by 90° between the passes (Netterfield et al., 1997), a possibility that is not available to other FP systems in the world. It will be explored in the future (cf. chapter 8) but is not used in the present initial version of the instrument. The ratio of the cavity thicknesses between etalon 1 and 2 has been chosen such that the least common multiple and therewith the free spectral range of the combined etalons is large. The exact numbers have been determined by means of an optimization process following Darvann & Owner-Petersen (1994) which also takes into a account the behavior of parasitic light. The two error values denote the rms and peak-to-valley polishing errors. More detailed information on the polishing errors is available to us in terms of thickness maps for each etalon. The peak transmission of the etalons depends on the optical setup and the wavelength. The transmission tends to the maximum value when the diameter of the ray bundle drops below the spatial scale of the CTE. This is the case for our telecentric setup at the highest magnification (cf. chapter 3). From (2.23) it can be seen that the effective finesse is proportional to the peak transmission of an individual etalon. Nevertheless, We indicate the finesse for easy comparison with other FP filter systems.

Tab. 2.2 shows the free spectral range and the tuning characteristics of the etalons corresponding to the extraordinary axis at wavelength 460.7 nm. The idea is to communicate the magnitude of the voltage and temperature sensitivity. The origin of the voltage and temperature dependence and the etalon tuning based on this dependence will be discussed in chapter 4.

In addition to the etalons a prefilter is needed to complete the TNF. The prefilter isolates one peak of the periodic FP transmission spectrum. Tab. 2.3 lists all the prefilters that are currently available. The requirements on the prefilters are discussed in section 5.2. The prefilters have to be quite narrow which degrades their peak transmission and drives the costs. This is a disadvantage all current filter systems based on FP etalons are afflicted with. One therefore tries to increase the free spectral range as much as possible by combining several etalons.

	Etalon 1	Etalon 2
Free spec. range	413.7 ±0.3 mÅ	538.2 ±0.4 mÅ
Temp. response	169 mÅ/°C	136 mÅ/°C
Temp. range	20 …	40 °C
Voltage response	76 mÅ/kV	102 mÅ/kV
Voltage range	−3.5 …	3.5 kV

Tab. 2.2: Free spectral range, voltage and temperature dependence of our FP etalons. The values correspond to the extraordinary axis, which is the only axis used in the current version of the filter system.

Center wavel. (Å)	FWHM (Å)	Peak transm. (%)
3933.7	1.2	30
4226.7	1.3	30
4607.3	1.6	30
5141.0	2.2	30
6563	0.5	5

Tab. 2.3: Characteristics of the currently available prefilters for the TNF. The first four filters are custom-made. They have 3 cavities and their out-of-band transmission is $\leq 10^{-4}$ which is required to sufficiently suppress all unwanted transmission peaks within the sensitivity range of the camera. The last filter is a DayStar filter. It has a low peak transmission but its small bandwidth allows a combination with a single etalon.

Chapter 3

Optical setups

3.1 Collimated versus telecentric

Filter systems, based on FP etalons, are basically used in two types of optical setups: *collimated* and *telecentric*. Before discussing these setups in the next two paragraphs, it is useful to distinguish between a real and an ideal etalon i.e. an etalon with and without CTE. In this way it becomes clearer which effects are due to the ray geometry alone, and which effects are caused by the etalon imperfections.

Collimated setup In a collimated setup the FP etalons are in a section of the beam path where the image is at infinity. They are typically positioned close to a pupil. Each image point in the FOV is “drawn” by a parallel ray bundle through the etalon. The width of the transmission peaks of an ideal etalon is therefore minimal, limited only by the etalon properties (FSR and reflective finesse). In a real etalon, different subsections of the ray bundle produce transmission spectra which are slightly shifted in wavelength with respect to each other, resulting in a broadening of the overall transmission peaks. If the etalon position coincides with the pupil the ray bundles of every image point illuminate exactly the same etalon aperture, which means that the broadening is constant across the FOV. This type of broadening is further described in section 2.

The angle between a ray bundle and the etalon surface normal is equal to the field angle[1] i.e. it depends on the position of the image point on the detector. As a consequence the position of the transmission spectrum is shifted to the blue as we go radially away from the image center (assumed to coincide with the optical axis). The amount of this blueshift and the resulting trade-offs for the design of the collimated setup are investigated in the next section.

---

[1] Assuming that the surface normal is aligned with the optical axis, which is normally the case. If the etalon is slightly tilted e.g. for ghost elimination (cf. section 5.2.2) the angle of the bundle is equal to the sum of field angle and tilt.

An ideal etalon does not affect the image quality at all in a collimated setup. For a real etalon the effect on the image quality is twofold. First, the CTE of a single etalon introduce wave phase errors i.e. a degradation of the wavefront. Secondly, the same fluctuations produce spectral shifts of the main transmission peak across the etalon aperture, as mentioned before. When two or more etalons are combined in series, these spectral shifts result in local transmission variations causing pupil apodization, which in turn degrades the wavefront. The pupil apodization in a collimated setup is difficult to correct within the optical layout because of its small-scale random nature. The image degradation specific to the FP etalons is further discussed in section 5.3 where some quantitative statements are made for our TNF. In this chapter we restrict ourselves to the optics enwrapping the etalons.

Telecentric setup In a telecentric setup the FP etalons are installed in a section of the beam path where the pupil is at infinity. Each image point in the FOV is drawn by a conic ray bundle whose axis (= chief ray) is parallel to the optical axis. For an ideal etalon the transmission spectrum is then the same for every image point but the peaks are broadened because of the angular distribution of the rays in the bundle.

The image quality is affected in a different way than in the collimated setup. The angular distribution of the rays produces a radial variation of the phase ϕ with respect to the optical axis which can be partially compensated by refocusing (Scharmer, 2006). Again we refer the reader to section 5.3 for more details on FP specific image effects.

Each ray bundle illuminates a different etalon sub-aperture. For a real etalon the shape and the position of the transmission spectrum can therefore vary across the FOV. When two or more etalons are combined this can result in significant local transmission variations. When working at high f-ratios such that the diameter of the ray bundle in the etalon is comparable to the scale of the CTE, their effect on imaging quality is negligible. At smaller f-ratios the CTE can produce wave phase errors and pupil apodization as in the collimated case, adding to the image degradation caused by the ray geometry.

3.2 Field of view versus spectral performance

The design of the optical setup for a FP based filter system, whether collimated or telecentric, is governed by a trade-off between blueshift or spectral resolution on the one side and FOV on the other side.

In this section we will work out trade-off relations for both types of setups. For the collimated case a similar derivation has already been done by Darvann & Owner-Petersen (1994). In the telecentric case we renounce the approximation of Darvann & Owner-Petersen (1994) which is limited to a large FOV or high spectral resolution and therefore unsatisfying for our purpose, and come up with a more general solution.

For an ideal etalon the blueshift in the collimated setup or the broadening in the telecentric setup is determined by the angle dependence of the transmission spectrum. To quantify these effects we therefore need to know the angles of the rays passing through the etalon. The problem is tackled with the Helmholtz-Lagrange invariant

$$H = \frac{h}{f/\#} = \theta_{\mathrm{FOV}} D \tag{3.1}$$

of a paraxial optical system, i.e. the ratio between the image size h and the f-ratio $f/\#$ of the beam projecting the image or, equivalently, the product θ_{FOV} of the FOV spanned by the image and the entrance pupil diameter D of the optical setup, typically the telescope aperture. In the following we stick with the paraxial approximation which is adequate to study the trade-off relations.

In a collimated setup, the angle of a parallel ray bundle with respect to the etalon surface normal is the field (or chief-ray) angle

$$\theta_{\mathrm{f}} = \frac{h}{2f} = \frac{H}{2D_{\mathrm{FP}}} \tag{3.2}$$

where f is the focal length of the imaging optics and D_{FP} is the illuminated aperture of the etalon. The minimum aperture is equal to the pupil diameter, because the parallel ray bundles of all image points cross each other at the pupil.

In a telecentric setup, the maximum angle of a ray in the conic ray bundle with respect to the etalon surface normal is the angle of the principle ray

$$\theta_{\mathrm{p}} = \frac{1}{2f/\#} = \frac{H}{2h} \tag{3.3}$$

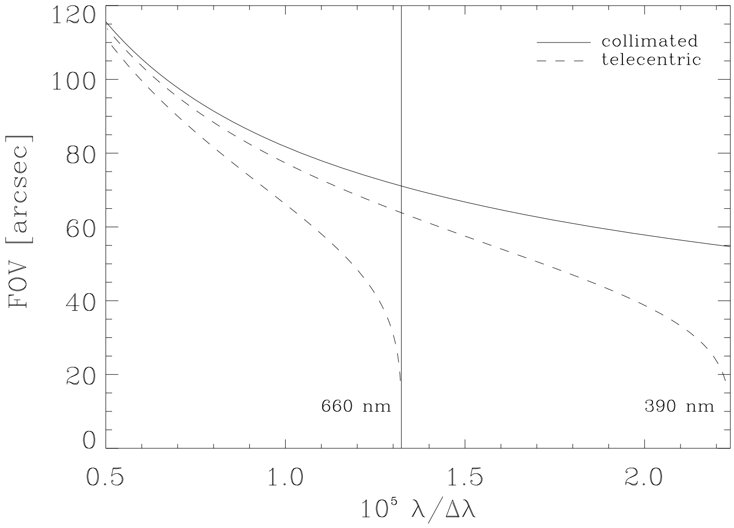


Fig. 3.1: Relation between spectral resolution/blueshift and FOV for a telecentric/collimated optical setup. The FOV is computed for a 1m telescope, a lithium-niobate etalon cavity with a refractive index of 2.2 and an etalon aperture (collimated setup) or detector diagonal (telecentric setup) of 20 mm. The telecentric spectral resolution saturates at the resolution for a parallel beam, normal to the etalon surfaces, which scales with $1/\lambda$. Two telecentric curves are plotted, corresponding to the boundaries of the spectral working range of our TNF (390-660 nm). The saturation values are denoted by vertical lines.

The equations (3.2) and (3.3) are very similar. In fact, the image size h in (3.3) can also be interpreted as the minimum clear aperture of the etalon in an optical layout where the etalon is exactly in an image plane. In our setup this is not possible as will be shown later in this section. Instead h is limited by the detector diagonal. By using Lagrange-Helmholtz (3.1) and by replacing the ray angles θ in air by the proper inner ray angles θ/n, we can finally resume the last two equations in the form

$$\text{Collimated}: \quad \theta_{\text{FOV}} = 2\frac{D_{\text{FP}}}{D}\theta_f = 2\sqrt{2}\,\frac{nD_{\text{FP}}}{D}\,\frac{1}{\sqrt{\rho}} \tag{3.4}$$

$$\text{Telecentric}: \quad \theta_{\text{FOV}} = 2\frac{nh}{D}\,\frac{1}{nf/\#(\rho_0,\rho)} \tag{3.5}$$

with $\rho = \lambda/\Delta\lambda$. In the collimated case, $\Delta\lambda$ is the blueshift. In the telecentric case, ρ is the effective spectral resolution.

The last expression in (3.4) has been obtained using the angle-dependence of the blueshift (2.15). The relation $nf/\#$ between the f-ratio (inside the etalon) and the telecentric spectral resolution is found by computing the FWHM of the broadened transmission peak described by (2.18). The solution does not have an explicit algebraic form but we note that it also depends on the maximum resolution ρ_0. The higher the maximum resolution of the etalon, the stronger the decrease in effective resolution with FOV (cf. Fig. 3.1).

Fig. 3.1 shows the FOV as a function of ρ in the collimated and in the telecentric case for a 1m telescope ($D = 1m$), a lithium-niobate etalon ($n \approx 2.2$) and an illuminated etalon aperture or detector diagonal of 20 mm ($D_{FP} = 20$ mm or $h = 20$ mm). The plots can be easily scaled to a different optical configuration with (3.4) and (3.5). Note that the telecentric resolution saturates at the value ρ_0 for a parallel normal beam which scales with λ^{-1}.

3.3 Optical setups chosen

We have realized both a collimated and a telecentric setup, on both ends of the trade-off between FOV and spectral performance. A schematic drawing of both setups is shown in Fig. 3.2.

Collimated version at IRSOL The collimated version is implemented as a stationary setup at IRSOL. The telescope at IRSOL, despite its future adaptive optics, will not be able to compete with other solar telescopes in terms of imaging performance. So, instead of high spatial resolution we aim at high spectral resolution and light gathering power or - which is equivalent - FOV.

We have therefore chosen a collimated setup which projects the maximum available FOV of 3 arcmin onto the ZIMPOL camera (Fig. 3.2). This first requirement fixes the demagnification (1:1.27) i.e. the ratio between the CCD diagonal and the size of the 3 arcmin-field in the telescopic focal plane. On the other hand we want to keep the blueshift as small as possible and still be able to shift the etalons inside the collimated beam section without vignetting. This second requirement fixes the focal length of the collimating optics (2.16 m).

Both, the collimating and the re-imaging optics are realized as spherical concave mirrors. We have chosen them instead of lenses for two main reasons: (1) Even with such a simple mirror shape and a slight off-axis alignment, the imaging is nearly diffraction limited. In fact the spatial resolution is limited above all by the camera sampling (Fig. 3.3). So, this type of mirrors is a cost-efficient alternative to achromatic lenses and we can have it without cutting back imaging performance. (2) It allows us to experiment with a double-pass configuration in a straightforward way, without the need of additional optics (cf. Fig. 3.2).

Telecentric version for the SST The telecentric version is designed for high spatial resolution. The low light throughput that comes with it requires long integrations. Even at the SST with an aperture of 1m, the integration time at a single spectral position can reach the lifetime of solar structures we are interested in, like the granulation for instance (cf. section 5.1). A collimated setup typically requires several exposures at different spectral positions covering the blueshift range, which are then combined to a single, spectrally homogeneous image. A luxury which we cannot allow ourselves in this case. This argument, together with tight constraints on the setup length, has let us decide in favor of a telecentric setup.

Normally a telecentric setup is realized as two separate stages. An example is the TESOS instrument installed in the German Vacuum Tower Telescope on Tenerife (Kentischer et al., 1998; von der Lühe & Kentischer, 2000). The first stage provides the telecentric beam for the FP etalons which are positioned close to the image plane. The second stage, typically also telecentric, projects the intermediary image on the camera at the desired magnification. The main advantage of this setup is that it permits to make use of the full etalon aperture and to minimize the spectral broadening therewith. Our setup does not have a first stage, instead the etalons are placed in the telecentric camera beam (Fig. 3.2). At the magnification for diffraction limited sampling ($\theta_{\mathrm{FOV}} = 36$ arcsec), the spectral resolution is still close to saturation (cf. Fig. 3.1), which means that we nearly perform as well as with an optimized two-stage setup. For the two lower magnifications available ($\theta_{\mathrm{FOV}} = 74$ arcsec and $\theta_{\mathrm{FOV}} = 110$ arcsec) the loss in spectral resolution is however substantial (cf. Fig. 3.1 and Tab. 3.1).

In our telecentric setup we use achromatic lenses. The lenses with $f = 97$ mm and $f = 470$ mm are custom-made by the manufacturer, based on a model that we made with the optical design software ZEMAX[2]. In this way we practically do not degrade the diffraction-limited MTF corresponding to the telescope aperture (cf. Fig. 3.3).

---

[2] ZEMAX Development Corporation, San Diego, USA (http://www.zemax.com).

	Collimated (IRSOL)	Telecentric (SST)
Cam. illum.	F/44	F/223, F/108, F/72
FOV [arcsec]	181 × 132	29 × 21 - 91 × 66
Spatial res. [arcsec]	0.5	0.1 - 0.3
Spectral res. [10^5]	1.67 - 2.82	0.52 - 2.47
Blueshift [mÅ]	15 - 25	0
Used FP aperture [mm]	40 - 60	24 - 27
Imaging optics	spherical mirrors	achromats (F2/N-BK7)
Setup length [mm]	9300	1400 - 1800

Tab. 3.1: Comparison of the collimated setup at IRSOL and the telecentric setup at the SST.

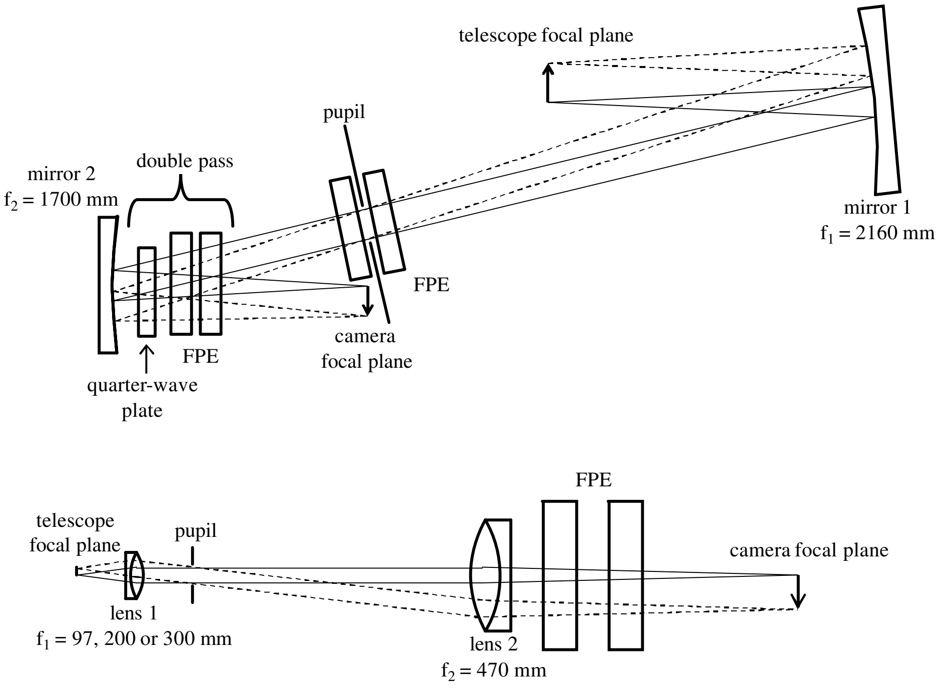

telescope focal plane
pupil
double pass
mirror 2
f_2 = 1700 mm
mirror 1
f_1 = 2160 mm
FPE
FPE
camera focal plane
quarter-wave plate
FPE
telescope focal plane
pupil
camera focal plane
lens 1
f_1 = 97, 200 or 300 mm
lens 2
f_2 = 470 mm

Fig. 3.2: Schematic drawing of our two optical setups (not to scale). **Top:** Collimated setup. This setup is designed for low spatial but high spectral resolution. It also permits to experiment with a double-pass configuration. **Bottom:** Telecentric setup. At the highest magnification ($f_1 = 97$ mm) the ZIMPOL camera works close to the diffraction limit of the SST. The compact design aims at portability.

Tab. 3.1 resumes the main properties of our collimated and telecentric setup.

A model of the spatial resolution of both setups, based on ZEMAX raytracings is shown in Fig. 3.3. Effects of the FP etalons are not included. To investigate the performance of the real setups, etalons included, we have carried out a simple test so far with a USAF target in the telescopic focal plane. This target features sets of horizontal and vertical bar patterns with increasing density. By visual inspection of the target image we determine the first pattern where the bars cannot be clearly distinguished anymore and compare the corresponding line pairs/mm with the modeled MTF. Note that, due to the masking, the ZIMPOL CCD has a different effective pixel size in the direction of modulation charge shifting (defined as vertical in Fig. 3.3) than in the orthogonal (horizontal) direction. The target measurements are not precise and give us only a rough idea of the resolution. In the telecentric case the result is close to the theoretical limit. This also tells us that the FP etalons do not have a major effect on the imaging performance. In the collimated case however we are rather far off the predicted performance. A further, more thorough investigation is needed but still owing.

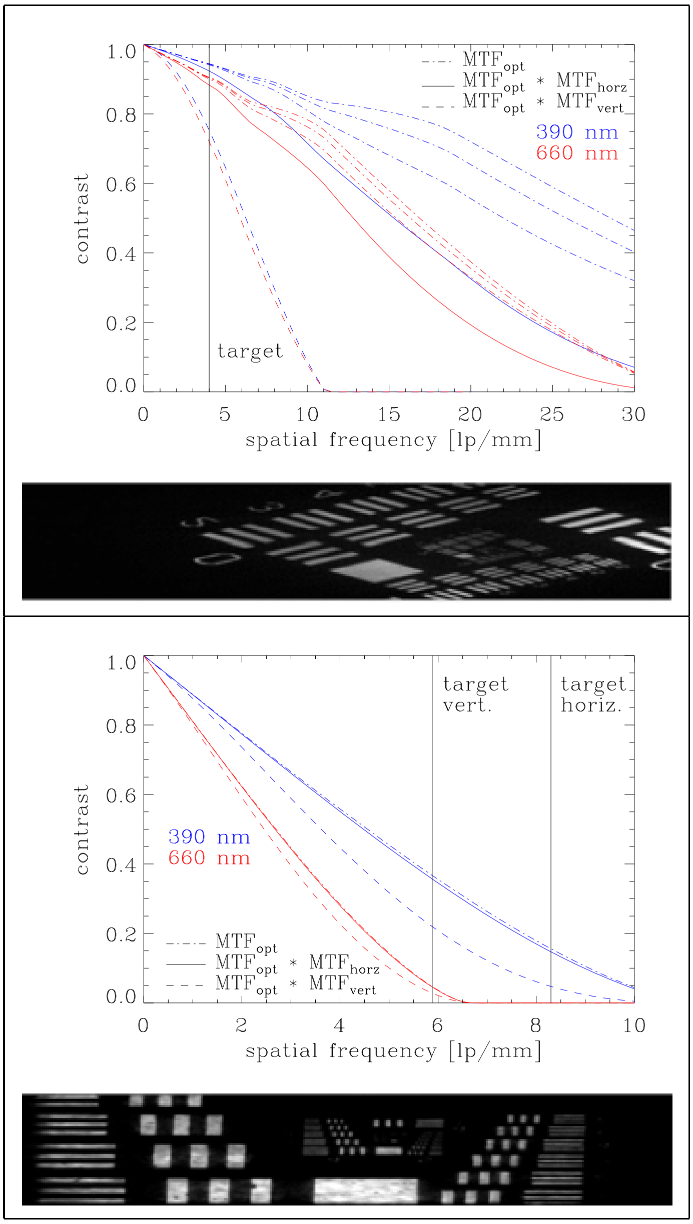


Fig. 3.3: Spatial resolution of the collimated and the telecentric setups (at the highest magnification). **Upper panel:** MTF model of the collimated setup at IRSOL for 390 nm and 660 nm (boundaries of the spectral working range). MTF$_{opt}$ includes the optics, without TNF. For each wavelength three MTF$_{opt}$ curves are plotted. The upper curve corresponds to an ideal aberration-free system. Due to field-dependent aberrations the real MTF$_{opt}$ covers a region which is delimited by the two lower curves. The MTFs corresponding to the detector sampling are denoted by MTF$_{vert}$ and MTF$_{horz}$, for vertical and horizontal sampling respectively. The vertical line marks the spatial frequency corresponding to the first pattern of the USAF target (picture below) that appears blurred. **Lower panel:** MTF model of the telecentric setup at the SST, at the highest magnification (4.8:1). The real system is practically diffraction limited, so only one MTF$_{opt}$ curve is plotted for each wavelength. Here we distinguish between a vertical and a horizontal target pattern that is visually resolved.

Chapter 4

Tuning and spectral calibration

4.1 Tuning parameters and etalon control

The transmission spectrum of a FP etalon can be tuned by varying the optical thickness of the cavity. As can be seen from (2.9) a small change Δnd in optical thickness provokes a wavelength displacement

$$\frac{\Delta\lambda_0}{\lambda_0} \approx \frac{\Delta nd}{nd} \tag{4.1}$$

of the etalon transmission spectrum. On the other hand the shape of the spectrum, given by the FSR and the finesse, is practically unaltered[1].

For air-spaced etalons a change in Δnd can be achieved by varying the geometrical distance d between the plates, with the help of piezoelectric actuators. Examples of such etalons, used in solar physics, are the IBIS (Cavallini et al., 2000) and TESOS (Kentischer et al., 1998) instruments. In contrast, for our lithium-niobate etalons, Δnd can be altered by applying either a voltage on the cavity or by changing its temperature. Another tuning method, which is common to both types of etalons, is tuning by tilting. In this case the angle θ_i of the incident rays is changed, and with it the effective cavity thickness $nd \cos\theta_i$ (cf. section 2).

In the next paragraphs we take a closer look at the three possible tuning methods and discuss their capabilities and disadvantages. The voltage and temperature effects on lithium niobate crystals have been thoroughly studied by (Bonaccini & Raymond, 1988). Here we recapitulate the results which are relevant for a qualitative understanding of the etalon tuning. The quantitative voltage and temperature behavior of our etalons is however not based on this results but determined from direct measurements of the transmission spectra (cf. section 4.2).

---

[1] Within the typical prefilter ranges of a few Å.

Voltage tuning The voltage dependence $\Delta nd(V)$ (at constant temperature) of the optical thickness of a lithium niobate crystal has two significant contributions:

$$\frac{\Delta nd(V)}{nd} \approx \frac{\Delta n(V)}{n} + \frac{\Delta d(V)}{d} \tag{4.2}$$

The change $\Delta n(V)$ of the refractive index is due to the Pockels effect. For the o-axis the voltage dependence is behaving linearly

$$\frac{\Delta n_o}{n_o} \approx \frac{n_o^2 \, r_{22}}{2\, d} V = O(10^{-8}) \cdot V[\text{Volt}] \tag{4.3}$$

where r_{22} is the corresponding unclamped electrooptic coefficient (cf. Bonaccini & Raymond, 1988). For the e-axis the voltage dependence is more quadratic and an order of magnitude smaller.

On the other hand the electrical field causes a mechanical strain in the crystal via the converse piezoelectric effect

$$\frac{\Delta d}{d} \approx \frac{d_{ii}}{d} \lambda = O(10^{-8}) \cdot V[\text{Volt}] \tag{4.4}$$

where d_{ii} is the coefficient of the so-called piezoelectric strain tensor corresponding to the chosen optical axis. The values for the e- and o-axes somewhat differ but they are of the same order of magnitude.

We have only indicated the order of magnitude of these effects because the values of r_{ij} and d_{ij} reported in various publications differ considerably. The manufacturer of our etalons also reports that they depend on the specific sample of lithium niobate.

The electrical field is applied perpendicular to the etalon surfaces by means of two transparent electrodes of indium-doped tin oxide. The applicable voltage range is ±3.5 kV which permits to tune over a complete FSR. The tuning speed is limited by the maximum stress variation the crystal can support. We do not exceed about 80 mÅ/s. This is much faster than temperature tuning (see next paragraph) and we therefore use it for spectral scans or to cycle between several discrete spectral positions, which is required for differential Zeeman and Hanle diagnostics.

Temperature tuning The temperature dependence (at constant voltage) of the optical thickness also has two dominating contributions

$$\frac{\Delta nd(T)}{nd} \approx \frac{\Delta n(T)}{n} + \frac{\Delta d(T)}{d} \approx O(10^{-5}) \cdot T[^\circ\text{C}]. \tag{4.5}$$

Both terms on the right side of the above equation, namely the temperature dependence of the refractive index and the thermal expansion, are of the same order. This high sensitivity allows to tune the etalons over 1 FSR with a temperature change of only a few degrees but it also calls for a thermal control with a precision of 10^{-2} °C. Without this precision the transmission spectrum cannot be sufficiently stabilized at a given temperature.

Temperature tuning is much slower than voltage tuning as the etalons need between 10 minutes and half an hour to stabilize at the new value (depending on the initial temperature difference). We therefore use temperature tuning only to park the etalons at an optimum home position within the spectral region of interest.

Crosstalk between temperature and voltage dependence The electrooptic and piezoelectric effects are also temperature dependent.

Bonaccini & Raymond (1988) conclude that the sensitivity of the electrooptic effect is mainly due to changes of the refractive index with temperature rather than to changes of the electrooptic coefficients themselves. From (4.3) we find that $\frac{\partial\lambda}{\partial V}$ only varies by some 0.01 mÅ/kV when the temperature is changed by 10 degrees, which is negligible.

The converse piezoelectric effect shows a more important sensitivity to temperature than the electrooptic effect. The thermal expansion of the crystal (at constant applied voltage) induces a decrease of the electric field which produces a variation of the piezoelectric strain. Bonaccini & Raymond (1988) state that this effect is far more important than the thermal variation of the piezoelectric coefficients d_{ij} themselves. From (4.4) we find that a temperature change of 10 degrees causes a variation in $\frac{\partial\lambda}{\partial V}$ of order 1 mÅ/kV which is not negligible in our case. We therefore fit the calibration measurements for different voltages and temperatures with a 2-dimensional function $f(T, V)$ as will be shown in section 4.2.

Tilt tuning Tilting the etalon with respect to the optical axis amplifies the blueshift in a collimated setup or further broadens the transmission peaks in a telecentric setup (cf. section 2.3). We therefore do not apply this tuning method. In the collimated setup tilting is only used for ghost elimination.

Etalon control All three tuning parameters can be controlled manually or via computer (Fig. 4.1). Under computer control one has the possibility to access these parameters within the framework of the ZIMPOL II software

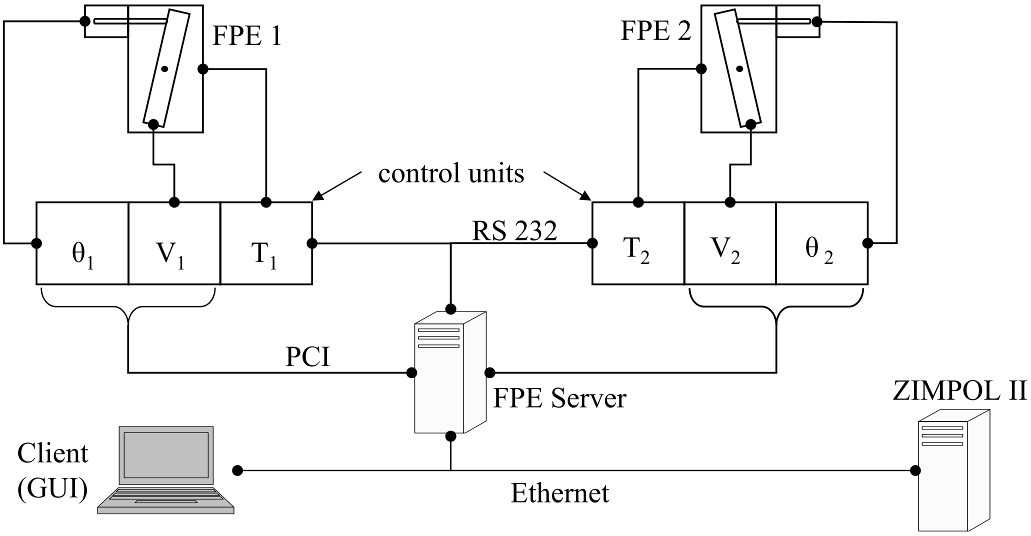


Fig. 4.1: Etalon control. All three tuning parameters, i.e. temperature (T), voltage (V) and tilt angle (θ) can be controlled manually or via computer. Under computer control one has the possibility to access these parameters within the framework of the ZIMPOL II software, which allows to script entire measuring procedures and to synchronize the etalons for tandem operation.

(cf. Gandorfer et al., 2004) which permits to script entire calibration, tuning and measurement procedures for the TNF and the polarimeter.

4.2 Voltage and temperature calibration

As opposed to air-spaced etalons it is difficult to model the optical cavity thickness with sufficient accuracy at once for the whole spectral working range. On the one hand the best dispersion model of lithium niobate known to us (Schlarb & Betzler, 1993) is not good enough to allow a prediction of the transmission peak positions with the required mÅ accuracy. On the other hand our knowledge of the effects playing in the temperature and voltage dependence is not sufficient for quantitative predictions with sufficient accuracy. We are therefore in the need of adequate calibration measurements for our etalons.

With the help of the high-resolution spectrograph available at IRSOL (Bianda et al., 2007) we record transmission spectra of the etalons, for each prefilter range separately and at different temperatures T_k and voltages V_l.

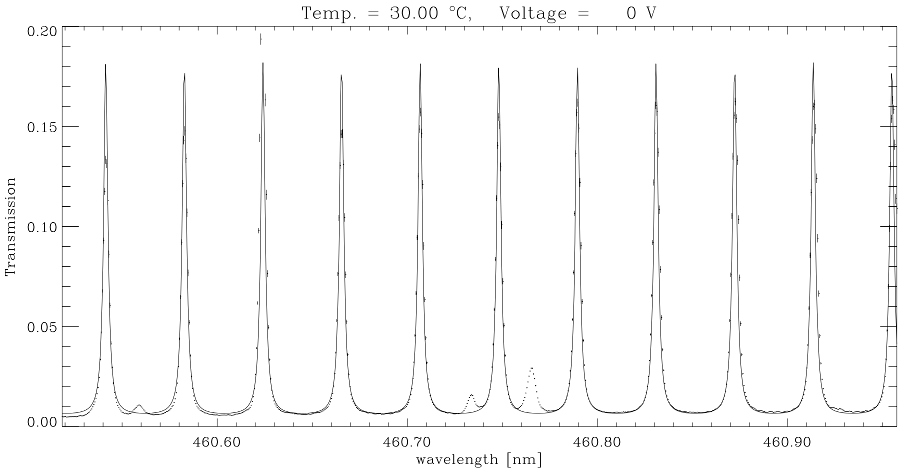


Fig. 4.2: The position of the etalon transmission peaks can be determined very accurately by fitting the etalon model (4.6) to the measured spectrum. This plot shows an example of such a fit (solid line) for the extraordinary axis of etalon no. 1. The data points are represented by their error bars. The uncertainty in the spectral position is typically below 1 mÅ. To calibrate the etalons for spectral tuning we repeat this procedure at different temperatures and voltages.

We then fit a simplified local etalon model with 5 free parameters $p_0 \ldots p_4$ to the data:

$$T = \frac{p_0}{1 + p_1 \sin^2(p_2(\lambda - p_4))} + p_3 \tag{4.6}$$

The model is local in the sense that within the narrow prefilter range the non-linearity of the phase ϕ of the true transmission function (2.6) is replaced by a linear approximation. In this way the stability of the fit is considerably improved. Fig. 4.2 shows an example. By doing so we end up with a set of model parameters $p_i(T_k, V_l)$. The two parameters of interest, namely the spectral position p_4 and the FSR π/p_2 are determined very precisely with this method: the typical uncertainties are 1 mÅ and 0.1 mÅ respectively.

To determine the temperature and voltage dependence of the transmission spectrum the 2-dimensional grid $p_4(T_k, V_l)$ is fitted with a polynomial of degree 2 or 3 (Fig. 4.3). After this procedure and with the help of the FSR π/p_2 we are able to predict the position of any transmission peak within the prefilter range with an accuracy of a few mÅ or about 1/10 FWHM. The residual errors in the etalon synchronization and in the positioning of the TNF can be corrected just before the actual scientific measurement with the fine-tuning and flatfield-scanning procedures described in the next sections.

4.3 Fine tuning and flatfield scanning

The fine tuning procedure helps to correct the residual synchronization error of the etalons. One etalon is kept fixed at its parking position at 0 Volt whereas the other etalon is tuned through one FWHM in steps of some 100 volts ($\sim$ 5 mÅ). At each step an intensity image is recorded. The result is a 3-D data cube with 2 spatial dimensions and one voltage or spectral dimension[2]. The spectrum of each pixel (i,j) is then fitted with a Gaussian to yield the wavelength displacement of maximum intensity $\Delta\lambda_{\mathrm{max}}(i,j)$. As the intensity is maximal for perfectly synchronized etalons, $\Delta\lambda_{\mathrm{max}}(i,j)$ is simply the synchronization error at the position (i,j) in the image. The mean value of the image is the correction to be applied to the scanning etalon, either by a voltage offset or by a temperature adjustment.

Fig. 4.4 shows examples of synchronization-error images obtained with this method. The vertical gradient in the collimated setup is due to differential blueshift and shows that the etalons are not properly aligned in this example (rel. tilt of 0.4°). Normally the differential blueshift reduces to about 2.5 mÅ which corresponds to the relative etalon tilt of 0.1° needed for ghost elimination (cf. section 5.2.2). In the telecentric setup the residual synchronization error is due to the CTE. This is an inherent property of the etalons which cannot be corrected. The rms value is 2.3 mÅ which is in accordance with the etalon thickness maps (cf. section 2) and still acceptable in terms of TNF performance (cf. section 5.2, in particular Fig. 5.7).

The flatfield-scanning procedure is mainly used to determine the local variations of the TNF transmission peak across the image. The telescope is operated in flatfield mode i.e. it is moved randomly around the Sun's center during the measurements to smear out any solar structures. In the meantime the TNF is scanned through a line of the solar spectrum in steps of some 10 mÅ. Like in fine-tuning we get a 3-D data cube and fit a Gaussian to the spectrum of each pixel.

Fig. 4.5 shows the images of two Gaussian parameters of interest. The scanned line depth (= amplitude of the Gaussian, normalized to 1) is an indicator of peak broadening or of spectral stray light (the lower the value the more broadening/stray light). The spectral shift (= center of the Gaussian) is the displacement between the theoretical and the scanned line center. The telecentric images in Fig. 4.5 show a bias of about 20 mÅ in spectral shift from left to right, which seems to be correlated with a slight decrease of line depth. This is consistent with a deviation of the chief ray angle from the optical axis of 0.17° across the FOV, but its cause is still eluding us. The non-concentric spectral-shift pattern in the collimated image reveals that the etalons are not well aligned with the optical axis. In addition the gradient in the line-depth image indicates a synchronization error due to a relative tilt of the etalons.

---

[2] The translation between voltage and wavelength displacement is given by the calibration described in the previous section

By means of the above examples we have shown that the fine-tuning and flatfield-scanning are two complementary procedures which permit to correct small adjustment errors of the etalons. We recapitulate the results here in form of step-by-step instructions.

1. In the collimated setup, adjust the tilt of one etalon to the minimum angle required for ghost elimination. The gradient in the fine-tuning image can be used as a verification. It should not exceed the differential blueshift corresponding to the relative tilt (2.5 mÅin our case), cf. (5.15).

   In the telecentric setup, adjust the tilt of one etalon until the average amplitude of the Gaussians fitted to the fine-tuning spectra is maximal.

2. Adjust the temperature of one etalon until the mean value of the fine-tuning image is zero.

3. Correct the temperature of both etalons with the respective value corresponding to the mean spectral shift found with the flatfield-scanning.

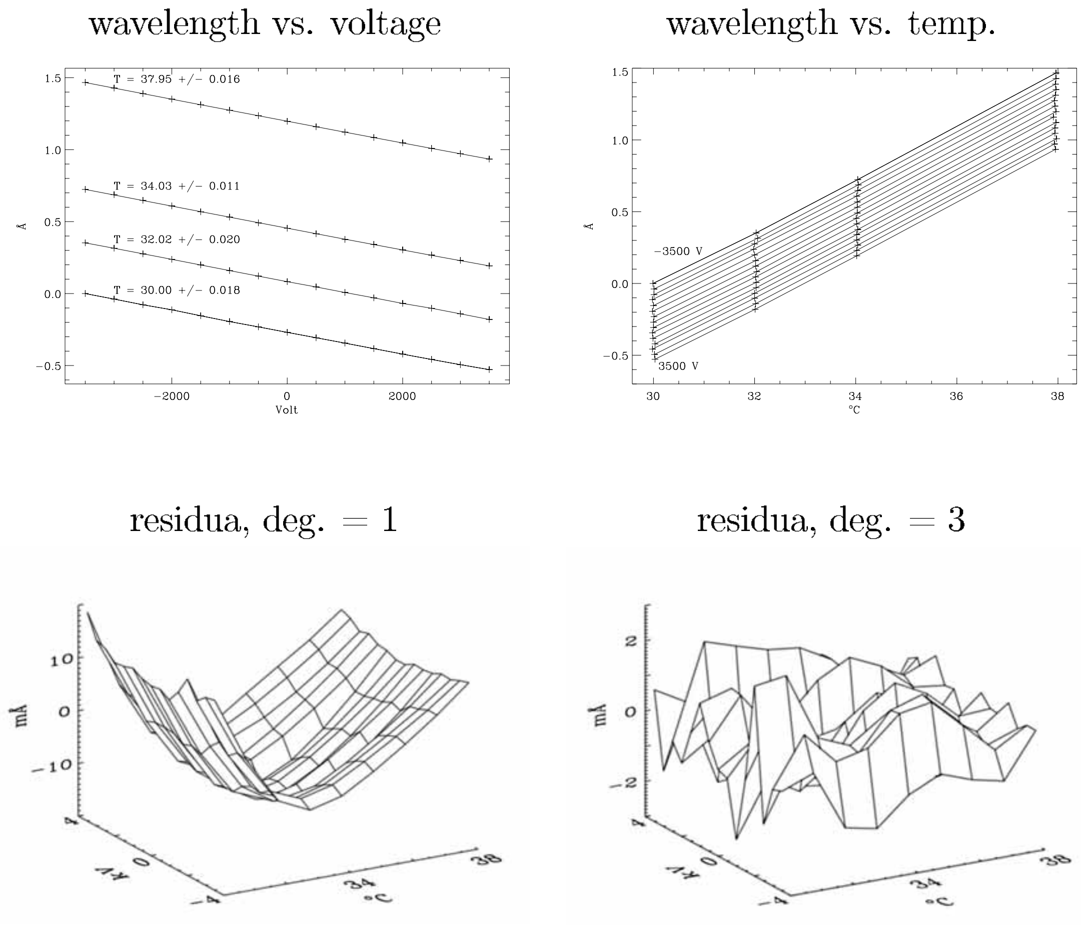
4. In the collimated setup, apply the same tilt adjustment to both etalons until the spectral shift is symmetric around the image center.

   In the telecentric setup adjust the tilt until the mean line depth is maximal.

4.4 Tuning through the prefilter

The voltage tuning range (± 3.5 kV) only permits to tune an etalon through one FSR (of order 0.5 Å) which is much less than the typical prefilter passband (some Å). Thus to access any wavelength within the prefilter range one has to switch the combination of etalon transmission peaks. This is illustrated in Fig. 4.6. The first two panels show the individual etalon spectra and the third panel the combined TNF spectrum (without prefilter). Each etalon transmission peak can be moved by voltage tuning within $\pm 1/2$ FSR (vertical lines) from its zero-voltage home position. In the example of Fig. 4.6 the solid spectra correspond to the TNF tuned at its home position of 460.73 nm. To reach the wavelength 460.77 nm, which is more than half a FSR away, a different combination of transmission peaks is chosen and then tuned by voltage to the correct position (dashed spectra).

For a given wavelength, the correct peak combination and etalon voltages are automatically computed by means of a ZIMPOL II script, using the results of the voltage and temperature calibration (cf. section 4.2). This procedure has been tested with the IRSOL spectrograph. The TNF is tuned through the Sr I prefilter and the predicted and measured positions of the main transmission peak are compared. Fig. 4.7 shows the results. Apart from the 40 mÅ offset, which can be easily corrected, the residuals are in the order of 5 mÅ which is acceptable in terms of performance (cf. section 5.2.4). At the positions marked by the vertical lines the combination of etalon transmission peaks is switched, which is more often than not accompanied by a small discontinuity.



wavelength vs. voltage

wavelength vs. temp.

residua, deg. = 1

residua, deg. = 3

Fig. 4.3: Spectral calibration for the Sr I prefilter range. The top panels show the results of the calibration measurements. Each data point represents the parameter $p_4(T_i, V_i)$, which results from the fit of the etalon model (4.6) to the measured etalon spectrum at temperature T_i and voltage V_i. The bottom panels show the residua of polynomial fits to $p_4(T_i, V_i)$. The voltage dependence of p_4 is practically linear whereas the temperature dependence is significantly nonlinear. In this case a polynomial fit of 3rd degree gives the best results, in other cases a polynomial fit of 2nd degree is sufficient.

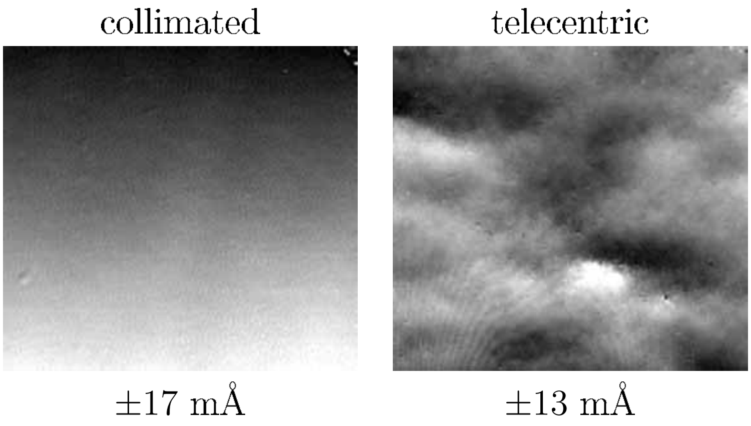


Fig. 4.4: Measured residual synchronization errors of the FP etalons at 4607 Å. Grayscales are noted below the images.

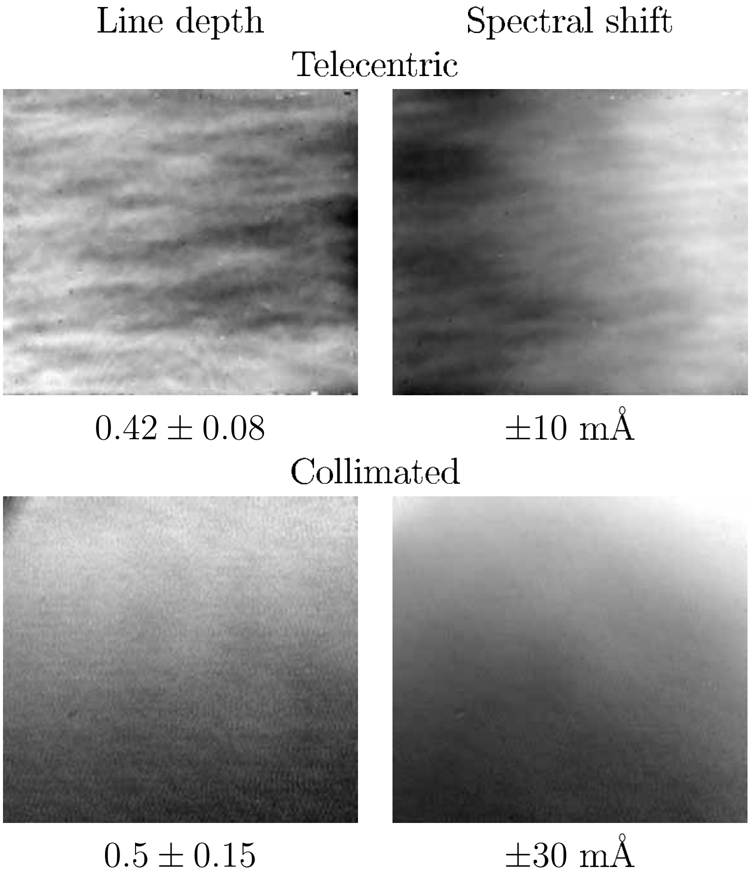


Fig. 4.5: Flatfield scans through the Sr I line with the telecentric setup and the collimated setup. Grayscales are noted below the respective images.

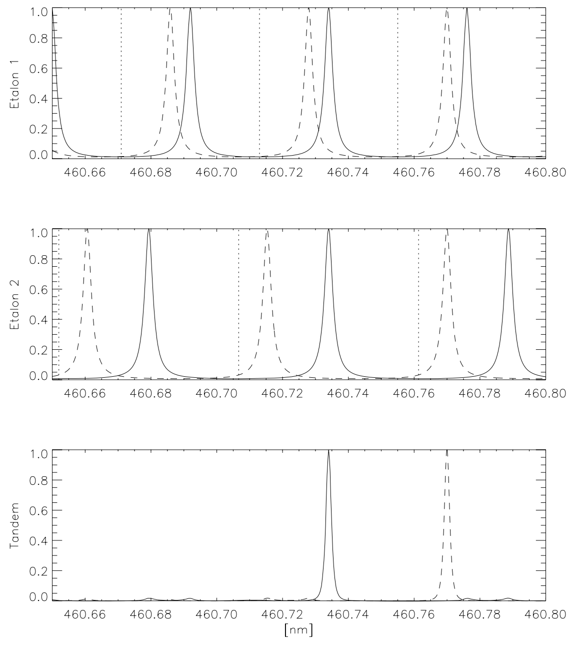


Fig. 4.6: Illustration of the peak-switching technique. The first two panels show the individual etalon spectra and the third panel the combined TNF spectrum. The vertical lines indicate the voltage-tuning limits. Two configurations are shown: home position at 460.73 nm (solid line) and position 460.77 nm (dashed line), which is reached by peak switching and voltage tuning.

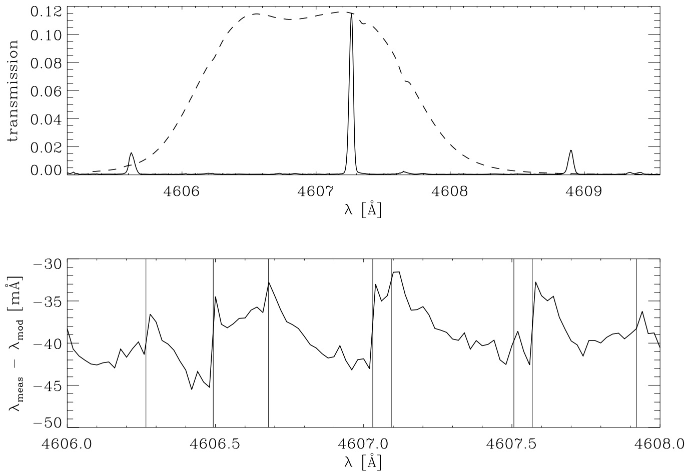


Fig. 4.7: Top panel: Example of the tandem transmission spectrum (solid curve) in the range of the Sr I prefilter (dashed curve). The prefilter curve is scaled to the tandem peak transmission. **Bottom panel:** Difference $\lambda_{\mathrm{meas}} - \lambda_{\mathrm{mod}}$ between measured and modeled peak positions during a tandem scan through the prefilter.

Chapter 5

Performance

Although it can be operated independently, the TNF has been mainly designed to work together with ZIMPOL. The combined system permits to record polarized quasi-monochromatic images of the Sun. To simplify matters we will refer to this system as Imaging Polarimeter (IP) in the following. As every instrumentation our IP has of course its deficiencies and its limits in terms of spatial and spectral resolution as well as polarimetric accuracy. In this chapter we will systematically explore these three performance criteria and define them more precisely. The goal is to clearly trace out the potential of the IP in its current state and to become aware of what can be measured and what not. In this way we are able to reasonably define first science targets that we can start with and to seek out the shortcomings in the instrumentation and measurement techniques where a future improvement is most promising in terms of investment and possible scientific return.

5.1 Photometric sensitivity

The current ZIMPOL II is limited by photon noise up to a S/N of 10^4 (Gisler, 2005, Fig. 3.4). Only at higher precision, instrumental effects start to play a role. Our imaging polarimeter (IP), i.e. the combined system ZIMPOL and TNF, is generally working in a photon-starving regime. Its high spectral and (in our telecentric setup) high spatial resolution, together with the required S/N of 10^4 or better, can require integration times corresponding to the timescales of solar structures of interest, like the granulation for instance. This imposes a severe limitation which we want to push as far as possible by saving photons from being lost by instrumental deficiencies. A photon budget helps to discover such deficiencies.

In this section we will estimate the expected photon flux received by the IP in different optical setups and compare the predictions to actual measurements. The goal is to identify the major photon drains with a potential for improvement.

5.1.1 Wavelength dependent elements

Fig. 5.1 shows the elements of the photon budget with a significant wavelength dependence.

Solar flux and TNF transmission Within the spectral working range of our TNF (390-660 nm) the radiation of the photospheric continuum is very close to black-body radiation with an effective temperature of $T = 5780$ degrees Kelvin, so we just use the Planck function $B(\lambda, T)$. The number of photons per time and wavelength, received at wavelength λ by one detector pixel of area A from an optical system with an effective f-ratio $f/\#$ is

$$\Phi = \frac{\pi}{4} B(\lambda, T)\, A \, \frac{1}{f/\#^{\;2}} \, \frac{\lambda}{hc} \,. \tag{5.1}$$

The last factor is the energy per photon.

The curve “Sun · TNF”, plotted in Fig. 5.1, shows the solar flux $\Phi \cdot W$ transmitted by our TNF if the etalons had no CTE. In this ideal case the EW W of the TNF can be computed starting with the equation (2.16) which describes the transmission peak of a single etalon. Neglecting higher-order terms in $\Delta\lambda$ a TNF transmission peak can be closely approximated with

$$T = \frac{T_1\, T_2}{1 + \left(2\left(\frac{1}{\Delta\lambda_{\mathrm{FWHM},1}} + \frac{1}{\Delta\lambda_{\mathrm{FWHM},2}}\right)\Delta\lambda\right)^2} \,, \tag{5.2}$$

where $\Delta\lambda_{\mathrm{FWHM},1}$ and $\Delta\lambda_{\mathrm{FWHM},2}$ is the respective FWHM of etalon 1 and 2. The EW of the ideal TNF is then

$$W = \int T \mathrm{d}\lambda = \frac{\pi}{2} T_1\, T_2 \left(\frac{1}{\Delta\lambda_{\mathrm{FWHM},1}} + \frac{1}{\Delta\lambda_{\mathrm{FWHM},2}}\right)^{-1} \tag{5.3}$$

In reality the optical thickness varies slightly across the aperture (cf. section 2). These variations cause local spectral shifts of the transmission peaks which leads to synchronization errors. As a result the EW of the TNF is somewhat degraded with respect to an ideal TNF. To have an idea of the significance of this effect we have computed the variations of the EW, based on the thickness maps from the manufacturer. An example for $\lambda = 390$ nm is shown in Fig. 5.3. The effect on the image depends on the optical setup. In the collimated setup, the variations of the EW are averaged over the aperture and therefore result in a general loss of intensity which is constant everywhere in the image. In the telecentric setup the variations in the EW show up as local intensity variations in the image, which can be flatfielded. However, if one aims for high spectral accuracy one has to be aware of two concomitant effects, namely local variations of the shape and the position of the transmission peaks. This can be corrected with flatfield scanning as described in section 4.3. In Fig. 5.1 we have plotted the ratio between the EW averaged over the TNF aperture and the EW of the ideal TNF given by (5.3). The corresponding curve is denoted with "Cavity thickness". The computation of the EW is done for a normal and parallel beam. In principle the EW depends on the f-ratio, but for our telecentric setups (minimum f-ratio of f/72) the difference in EW between converging and parallel beams is negligible.

Limb darkening The limb darkening in Fig. 5.1 is computed for a limb distance of $\mu = 0.1$, based on Neckel (1996).

Atmospheric extinction The wavelength dependence of the atmospheric extinction is adopted from the model of King (1985), which includes Rayleigh scattering and ozone absorption (mean annual value). The latter contribution is based on measurements. Water vapor and dust scattering are not included, as their contribution to the extinction can vary significantly. We use here the extinction factor E i.e. the ratio

$$E = \frac{I}{I_0} \tag{5.4}$$

between the intensity I_0 outside the atmosphere and the intensity I on the ground, and refer to it as extinction. It must not be confounded with the commonly used extinction E_{mag} in units of stellar magnitudes, related to E by

$$E = 10^{-\frac{E_{\mathrm{mag}}}{2.5}} \tag{5.5}$$

From comparisons between their model and observed mean extinction, the authors suggest that their calculated E_{mag} is probably accurate to within a few 0.01 mag within the spectral working range of our TNF.

The extinction curve for the SST can be roughly translated to IRSOL by considering the height-dependent Rayleigh scattering only:

$$E_{\mathrm{IRSOL}} \approx E_{\mathrm{SST}}{}^{1.27} \tag{5.6}$$

The scaling of the extinction with zenith angle θ is

$$E(\theta) = 10^{-\frac{E mag}{2.5 \cos\theta}} = E^{\frac{1}{\cos\theta}} \tag{5.7}$$

The curve displayed in Fig. 5.1 is computed for the SST site and $\theta = 0$.

Detector The quantum efficiency plotted in Fig. 5.1 includes the effect of the open-electrode structure of the ZIMPOL II cameras with enhanced UV sensitivity. A fraction of the light, falling on a pixel, directly hits the photo-sensitive layer whereas the rest of the light has to pass the electrode layer first. The resulting wavelength dependence of the pixel transparency is taken into account (cf. Gisler, 2005, chap. 9).

Since October 2006 one camera of the open-electrode type is equipped with a microlens array. The light falling on the masked pixel rows is focused into the unmasked pixel rows which results in an enhancement of about a factor 4 in sensitivity (Fig. 5.2). This camera was successfully used at the SST observing campaign of the same year and has become the standard camera of the IP since then. For the high f-ratios of the TNF optics the light is not only focused into the unmasked pixel row but into the open pixel section[1], which further enhances the sensitivity with respect to a classical ZIMPOL II camera.

As can be seen from Fig. 5.1 the photon flux "Sun · TNF" strongly decreases toward the blue side of the spectral working range, because the photon flux and the EW of the TNF both go down. Both effects are inherent to the Sun and to the working principle of an FP etalon. The only means to improve this flux is an increase of the telescope aperture, or by sacrificing spectral or spatial resolution. The same argument holds for limb darkening and atmospheric extinction. Concerning the latter, the SST site on La Palma is slightly better than IRSOL (some 10%). On the other hand the detector has a significant potential for improvement as 70% of the photons get lost here.

5.1.2 Wavelength-independent elements

Tabs. 5.1 and 5.2 keep an account of all the elements in the beam path with an insignificant wavelength dependence. Such elements include mirrors,

---

[1] The light is just focused along the charge-shifting direction. In the orthogonal direction, the cylindrical microlenses do not have any effect.

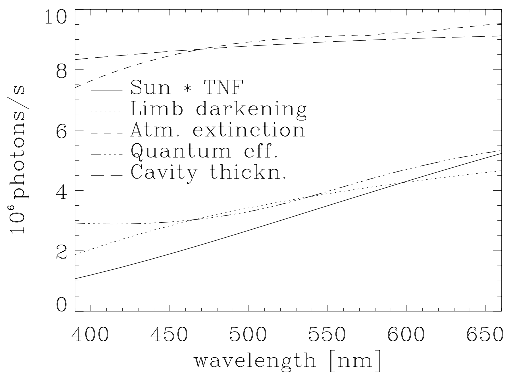


Fig. 5.1: Elements of the photon budget with a significant wavelength dependence. See text for details, in particular the scaling for other scenarios like e.g. a different f-ratio or zenith angle. "Sun * TNF" is the photon flux from the photospheric continuum received, through the TNF and from an F/50 beam, by a surface corresponding to one detector pixel. The limb-darkening factor corresponds to a limb distance of $\mu = 0.1$. The atmospheric extinction is for the SST site with the Sun in zenith and zero humidity. The y-scale is for the photon fluxes. All other curves are plotted relative to a (0,1) y-scale.

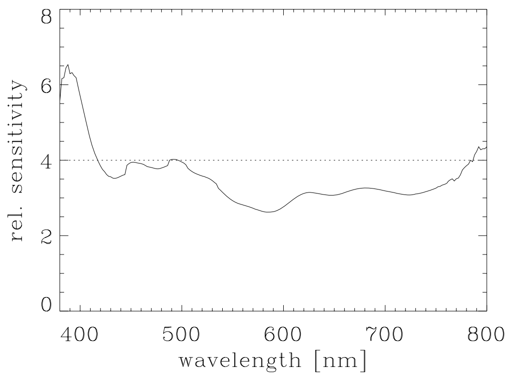


Fig. 5.2: Comparison of ZIMPOL II UV-enhanced cameras with and without microlenses. The cameras have been illuminated one after the other by the same stable light source in the same optical setup with a monochromator. The plot shows the ratio between the measured flux of the camera with microlenses divided by the flux of the other camera. The horizontal line denotes the ratio given by the pixel geometry alone. The steep rise in relative sensitivity below about 420 nm is due to the focusing into the open pixel section whereas in the camera without microlenses a larger part of the light is blocked by the electrode layer (measurements: D. Gisler).

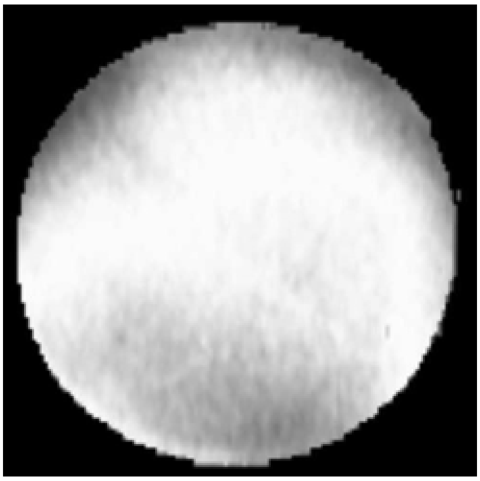


Fig. 5.3: Variations of the EW of the main transmission peak across the aperture of the TNF, tuned at 390 nm. The grayscale is defined as follows: black corresponds to 0 and white to the EW of an ideal TNF given by (5.3). Simulation based on the thickness maps of the individual etalons, provided by the manufacturer.

glasses (like e.g. lenses or windows), beamsplitters and polarizers. The polarizers are the single elements which produce the biggest losses. About half of the photons are thrown away. This can be avoided by using two detectors in combination with a polarizing beamsplitter, each detector recording one orthogonal linear polarization.

5.1.3 Comparison with measured photon flux

In Tab. 5.3 the predicted and measured photon flux are compared for two representative situations.

The column “SST, 4227 Å” represents a measurement with the telecentric setup during our 2006 campaign at the SST. The imaging was at the lowest magnification ($\theta_{\mathrm{FOV}} = 112$ arcsec, cf. Tab. 3.1), the sky was clear, no haze or clouds. The row I/I_c denotes the line intensity at the TNF position, divided by the continuum intensity. T_{fix} is the total attenuation factor from the wavelength-independent elements (Tab. 5.1).

The column “IRSOL, 4227 Å” represents a measurement with the collimated setup at IRSOL. Despite the lower altitude of IRSOL, the atmospheric extinction is not so strong as in the SST measurement because the Sun was at a higher elevation. The spectral position of the TNF was almost identical to the position of the SST measurement.

Prediction and measurement are of the same order. The remaining gap

SST

Elements	**Transm.**
7 mirrors, Al coating	$0.91^7 = 0.52$
2 periscope mirrors	$0.99^2 = 0.98$
9 coated transm. elements	$0.99^9 = 0.91$
AO beamsplitter	0.7
2 parallel pol. beamsplitters	$0.5 \cdot (0.48/0.5)^2 = 0.46$
Total	0.15

Tab. 5.1: Elements of the photon budget with insignificant wavelength dependence. List of all elements included in the telecentric setup at the SST.

IRSOL

Elements	**Transm.**
9 mirrors, Al coating	$0.91^9 = 0.43$
entrance window (BK7 uncoated)	0.92
3 coated transm. elements	$0.99^3 = 0.97$
derotator	0.7
2 parallel pol. beamsplitters	$0.5 \cdot (0.48/0.5)^2 = 0.46$
Total	0.12

Tab. 5.2: Elements of the photon budget with insignificant wavelength dependence. List of all elements included in the collimated setup at IRSOL.

	SST, 4227 Å	IRSOL, 4227 Å
Sun · TNF $[s^{-1}]$	2.9e+05	9.9e+05
Cavity thickn.	0.84	0.85
Limb darkening	0.41	0.52
Atm. extinction	0.64	0.64
Quantum eff.	0.29	0.29
Microlenses	3.70	1.00
I/I_c	0.70	0.75
T_{fix}	0.15	0.12
Total $[s^{-1}]$	7.2e+03	7.3e+03
Measured $[s^{-1}]$	3.1e+03	2.1e+03

Tab. 5.3: Comparison between photon budget and measured photon flux.

between theory and measurement (about a factor of 2 for SST and 3 for IRSOL) should be further investigated. Howsoever, by comparing the relative transmissions of the individual components it becomes clear that any efforts to improve the photometric sensitivity should start with the camera.

5.2 Spectral and polarimetric performance

5.2.1 Performance parameters

To illustrate the relation between the true and the measured degree of polarization, we pick out a q measurement[2] as an example. The arguments are the same for u and v measurements. So we have

$$q_{\text{meas}} = \frac{\int Q(\lambda)\,T(\lambda)\,\mathrm{d}\lambda}{\int I(\lambda)\,T(\lambda)\,\mathrm{d}\lambda} \tag{5.8}$$

where q_{meas} is the measured q and T the TNF transmission spectrum. The integral is evaluated in principle from 0 to ∞ but can be narrowed down at this point to the prefilter band. The contribution of the out-of-band transmission is estimated separately. We now break down (5.8) to isolate the different effects on q_{meas}. First of all the total TNF transmission

$$T = T_t + T_r \tag{5.9}$$

consists of two parts. The first part T_t is the product of the transmission spectra of the individual TNF components (FP etalons and prefilter), whereas the second part T_r has its origin in multiple reflections between these components. Both parts entail an unwanted contribution of *parasitic light*.

---

[2] q, u and v are short forms for Stokes Q/I, U/I and V/I respectively.

The *transmittive parasitic light* manifests itself in form of secondary peaks - transmission ghosts - and a flat background. Its contribution to the measured Q signal is

$$Q_{\mathrm{tp}} = \int Q(\lambda)\, T_t(\lambda)\, \mathrm{d}\lambda - \underbrace{\int_{\mathrm{mainpeak}} Q(\lambda)\, T_t(\lambda)\, \mathrm{d}\lambda}_{=:\, Q_{\mathrm{main}}} \tag{5.10}$$

In our performance simulation (described below) we do not distinguish between the transmission ghosts and the background. The transmission ghosts are further discussed in section 5.2.3, where we describe in particular a fast algorithm to compute their EW.

The spectrum of the *reflective parasitic light* is different in shape and amplitude from the transmittive one. Its contribution is

$$Q_{\mathrm{rp}} = \int Q(\lambda)\, T_r(\lambda)\, \mathrm{d}\lambda \tag{5.11}$$

A model for this type of parasitic light and techniques to eliminate or attenuate it are discussed in section 5.2.2. The final bookkeeping equation of the effects on q_{meas} is

$$\begin{aligned} q_{\mathrm{meas}} &= \left(\frac{Q_{\mathrm{main}}}{I_{\mathrm{main}}} + \frac{Q_{\mathrm{tp}}}{I_{\mathrm{main}}} + \frac{Q_{\mathrm{rp}}}{I_{\mathrm{main}}}\right) \frac{I_{\mathrm{main}}}{I} \quad \text{with} \qquad (5.12)\\ Q_{\mathrm{main}} &= \int_{\mathrm{mainpeak}} Q(\lambda)\, T_t(\lambda)\, \mathrm{d}\lambda, \\ I_{\mathrm{main}} &= \int_{\mathrm{mainpeak}} I(\lambda)\, T_t(\lambda)\, \mathrm{d}\lambda \quad \text{and} \\ I &= \int I(\lambda)\, T(\lambda)\, \mathrm{d}\lambda \end{aligned}$$

From the above analysis we extract three appropriate performance parameters, outlined in table 5.4.

5.2.2 Reflective parasitic light

In the current version of the TNF the two FP etalons are used in series in a single-pass tandem configuration. The transmission spectrum of this combination cannot be simply described as a product of the individual etalon spectra. One has to take into account multiple reflections between the etalons.

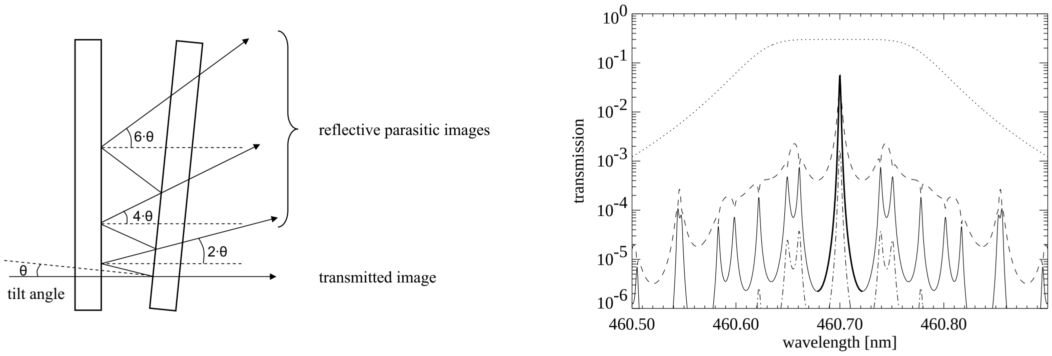


Fig. 5.4: Left: The reflective parasitic light can be eliminated by tilting one FP etalon if the distance between the TNF and the camera is large enough. The tilt angle needed to eliminate the ghost images is the angle that projects the first ghost image just on top of the CCD. In our collimated setup, where we use this method, the tilt angle is only about 0.1°. For this angle the differential blueshift is tolerable (cf. text). **Right:** Reflective and transmittive parasitic light compared (simulation). As an example the range of our Sr I prefilter has been chosen. The main transmission peak (thick solid line) is surrounded by transmission ghosts (solid line). Without prefilter attenuation, the reflective parasitic light (dashed line) contains almost 3 times the power of the main peak. With the prefilter inserted between the etalons (dash-dotted line) the power is reduced to 5% of the main-peak power.

Performance parameters

1. The square root of the EW of the main TNF transmission peak $\sqrt{W_{\mathrm{main}}}$, which is proportional to the S/N of q.
2. The ratio W_{main}/W between the EW of the main transmission peak and the total EW of the TNF. This parameter describes the dilution of the true polarization signal by the parasitic light.
3. The ratio W_{ghost}/W between the EW of the strongest transmission ghost and the total EW. This parameter is an indication of possible localized stray light that can sneak into the true polarization signal, for example when the ghost is accidentally aligned with a strongly polarizing spectral line that neighbors the main transmission peak.

Tab. 5.4: Performance parameters used to analyze the spectropolarimetric performance of the TNF.

Assuming that the reflections add up incoherently[3], the total contribution T_r of the reflective parasitic light to the TNF transmission is (Mack et al., 1963)

$$T_r = T_1\, T_{\mathrm{pf}}\, T_2\, \frac{(1 - T_1)\,(1 - T_2)}{1 - (1 - T_1)\,(1 - T_2)} \tag{5.13}$$

where T_1, T_2 and T_{pf} are the individual transmission spectra of the FP etalons and the prefilter respectively.

One can completely eliminate the reflective parasitic light by tilting one etalon with respect to the other one. For a given tilt angle θ the reflective parasitic light appears in form of an infinite series of ghost images, separated each by an angle 2θ (Fig. 5.4). If

$$\theta \geq \frac{\text{CCD size}}{2 \cdot \text{distance between CCD and TNF}} \tag{5.14}$$

all ghost images miss the CCD.

---

[3] Coherence across the TNF aperture imposes a tight condition on the parallelism of the two FP etalon surfaces facing each other: the optical distance covered by a ray bouncing forth and back between the etalons must vary much less than a wavelength. This condition is not met by our setups.

The tilt angle and therewith the practicability of this method is however limited by the fact that tilting one FP etalon introduces a differential blueshift and/or a peak broadening in the TNF transmission spectrum. The resulting synchronization error $\Delta\lambda$ in the collimated setup, corresponding to a field angle $\Delta\theta$ is

$$\frac{\Delta\lambda}{\lambda} \approx \theta\,\Delta\theta - \frac{\Delta\theta^2}{2} \tag{5.15}$$

where the first and second term describe the field-dependent blueshift at tilt angle $\theta \neq 0$ and $\theta = 0$ respectively. In our setup, where we apply this method, θ is only 0.1°, inducing a maximum synchronization error of about 2.5 mÅ or 1/10 of a FWHM at 4607 Å, which is tolerable in terms of TNF performance (cf. Fig. 5.8).

In our telecentric setup we cannot tilt one etalon because the resulting broadening of the main transmission peak is too large.

We therefore recur to the method of inserting the prefilter in between the FP etalons, proposed by Cavallini et al. (2000). After each reflection the beam is attenuated by the prefilter. In this way the reflective parasitic light is reduced to

$$T_r = T_1\,T_{\mathrm{pf}}\,T_2\,\frac{(1-T_1)\,(1-T_2)\,T_{\mathrm{pf}}^2}{1-(1-T_1)\,(1-T_2)T_{\mathrm{pf}}^2} \tag{5.16}$$

as compared to (5.13). Figure 5.4 demonstrates the effectiveness of this method on the example of our Sr I prefilter with a peak transmission of about 30%. However, to avoid reflections between the prefilter and the FP etalons we have to tilt the prefilter by 1° which shifts its transmission spectrum significantly: 0.6 to 1 Å depending on wavelength. This should be accounted for when specifying the prefilter for manufacturing.

5.2.3 Model of the tunable narrow-band filter

In this section we describe the TNF model used for the performance simulation as well as the algorithm for determining the peak position and the EW of the transmission ghosts.

The model provides the transmittive part T_t and - optionally - the reflective part T_r of the TNF spectrum at any given wavelength. T_t is simply the product of the transmission spectra of the individual TNF components:

$$T_t = T_1\,T_{\mathrm{pf}}\,T_2 \tag{5.17}$$

T_r is computed from (5.16) or (5.13), depending on the actual optical setup.

Prefilter The spectrum around the peak transmission λ_0 of a prefilter with n_{cav} cavities is modeled[4] as (Darvann & Owner-Petersen, 1994)

$$T_{\mathrm{pf}} = \frac{1}{1 + \left(2\frac{\lambda-\lambda_0}{\Delta\lambda_{\mathrm{FWHM}}}\right)^{2n_{cav}}} . \tag{5.18}$$

FP etalons For the FP etalon spectra we employ the model (4.6). The parameters p_2 and p_4 are adopted from the the spectral calibration (cf. section 4). The parameter p_3 is set to 0, assuming that the background is not produced by the etalons themselves. This assumption is backed up by the comparison between modeled and measured TNF transmission spectra (Fig. 5.6). The current model with $p_3 = 0$ for both etalons overestimates the type-2 ghosts (defined below) and a $p_3 \neq 0$ would enhance these ghosts even more.

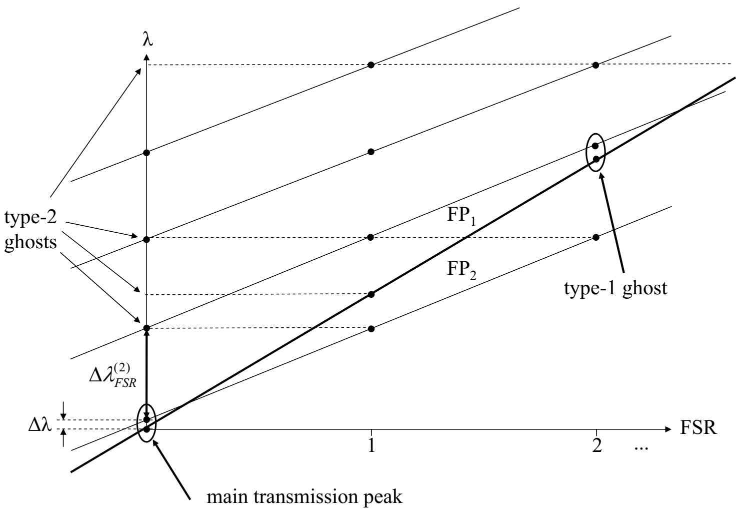


Fig. 5.5: Illustration of the algorithm used to identify the ghost peaks in the TNF transmission spectrum. See text for details.

Transmission ghosts The third performance parameter W_{ghost}/W is computed by integrating over the strongest transmission ghost i.e. the strongest peak in the transmittive parasitic light. How can we identify this particular ghost? The brute force method is to seek out all local maxima in the model TNF spectrum, integrate over a reasonable region around these maxima and then compare the results. This method is however numerically expensive and therefore slow.

---

[4] For the available prefilters one could replace the model with a measured spectrum. Nevertheless, we always resort to the model as it reproduces the real prefilters very well around λ_0.

We present here an algorithm that allows to determine the ghost positions in a much more efficient way and to find the strongest ghost without numerical integration. The procedure is illustrated in Fig. 5.5. The peaks of a single FP etalon spectrum are discrete points on a straight line[5] following each other in a regular interval of one FSR. This periodicity allows to represent the peak positions of etalon 2 with a set of parallel lines stacked on top of each other at intervals of $\Delta\lambda_{FSR}^{(2)}$ (wavelength difference corresponding to the FSR of etalon 2). The lines representing etalon 1 and 2 (denoted by FP_1 and FP_2 in Fig. 5.5) are not parallel since the etalons have different optical thicknesses. The origin of the coordinate frame sits at the main transmission peak which is determined by the (nearest) coincidence of two points belonging to different etalons. A possible synchronization error $\Delta\lambda$ between the etalons causes a small displacement of the line FP_2 on the vertical axis.

We distinguish 2 types of ghosts. Type 1 comprehends the strongest ghosts and is associated with the points that are closest to the intersection between the FP_1 and a FP_2 line. All other points belong to type 2. The position n_{FSR} on the abscissa (in Fig. 5.5) of type-1 ghost no. k is

$$n_{\mathrm{FSR}} = \left(k \, \frac{\Delta\lambda_{\mathrm{FSR}}^{(2)} + \Delta\lambda}{\Delta\lambda_{\mathrm{FSR}}^{(1)} - \Delta\lambda_{\mathrm{FSR}}^{(2)}} \right)_{\mathrm{round}} \tag{5.19}$$

The expression enclosed by $()_{\mathrm{round}}$ denotes the intersection point of the k-th FP_2 line and the FP_1 line. It is rounded to the closest integer. From n_{FSR} we can derive the wavelengths λ_1 and λ_2 of the individual etalon peaks belonging to the ghost:

$$\begin{aligned} \lambda_1 &= n_{\mathrm{FSR}} \, \Delta\lambda_{\mathrm{FSR}}^{(1)} + \lambda_0 \\ \lambda_2 &= n_{\mathrm{FSR}} \, \Delta\lambda_{\mathrm{FSR}}^{(2)} + \Delta\lambda + \lambda_0 \end{aligned} \tag{5.20}$$

where λ_0 is the wavelength of the main peak. The shorter the distance $|\lambda_1 - \lambda_2|$, the stronger the ghost.

---

[5] The lines are not straight strictly speaking as the FSR is wavelength dependent. But within the narrow prefilter range, the linear approximation (constant FSR) is adequate. Nevertheless, the procedure also works if the lines are not straight.

Now that we have determined the ghost position, the next step is to find out which type-1 ghost has the largest EW. Do do this without numerical integration we make use of the fact that a single ghost, generated by two etalons with a finesse well above 1, can be neatly approximated by

$$T \approx \frac{p_0^{(1)}\, p_0^{(2)}}{1 + \left(p_1^{(1)}(p_2^{(1)}\right)^2 (\lambda - \lambda_1)^2 + \left(p_1^{(2)}(p_2^{(2)}\right)^2 (\lambda - \lambda_2)^2} \tag{5.21}$$

The prefilter transmission can be approximated by a constant average value $T_{pf}(\frac{\lambda_1+\lambda_2}{2})$ as it varies slowly over the narrow wavelength interval spanned by the ghost. With these approximations the ghost EW is of the form

$$W_{\mathrm{ghost}} \approx T_{pf}\left(\frac{\lambda_1 + \lambda_2}{2}\right) \int_{-\infty}^{\infty} T \mathrm{d}\lambda \tag{5.22}$$

The integration is carried out over all wavelengths as (5.21) describes an isolated ghost peak with a spectral profile that drops rapidly with the distance from the center wavelength. In this way the integral can be expressed in terms of a simple analytical solution.

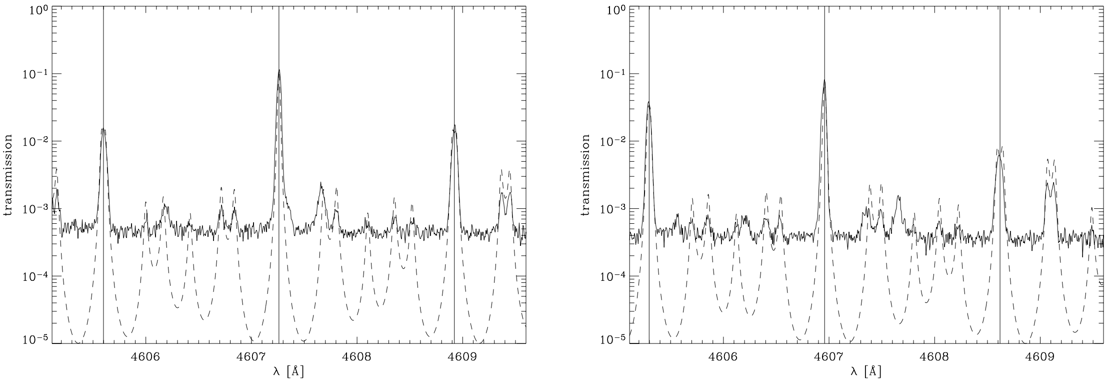


Fig. 5.6: Comparison between measured (solid) and modeled (dashed) TNF transmission spectra across the Sr I prefilter region. The measurements have been performed with the spectrograph at the Gregory-Coudé telescope of IRSOL. **Left panel:** Synchronized etalons. **Right panel:** Synchronization error of 35 mÅ. Note that the TNF spectrum becomes asymmetric. The left type-1 ghost is significantly enhanced.

Comparison with measurements Before using the TNF model with confidence for a performance analysis we have to be sure that it represents reality sufficiently well. We therefore compare the model with a high-resolution spectrum of the tandem FP (without prefilter), recorded with the spectrograph of the Gregory-Coudé telescope at IRSOL (Fig. 5.6). The model reproduces very well the main transmission peak and the type-1 ghosts (marked by the vertical lines). The peak transmission of the type-2 ghosts is generally overestimated which is not too disturbing as they are an order of magnitude fainter than the strongest ghost. Note also that the measured spectra are superimposed to a flat stray-light background of $4-5 \cdot 10^{-4}$. It is not clear if this background is inherent to the TNF or to the particular optical setup used to record the TNF spectrum. The consequence for the performance analysis is a possible overestimation of the dilution parameter W_{main}/W. The right panel shows a configuration where one etalon is tuned 35 mÅ out of phase. The positions of two ghosts (no. 3 counted from the main peak) are not reproduced correctly by the model, for yet unknown reasons. Again the consequences for the performance analysis are not too critical as they are an order of magnitude fainter than the strongest ghost.

5.2.4 Results

Using the TNF model and the ghost classification algorithm described in the previous section, we have run two kinds of performance simulations for every wavelength region of interest. In this section we will discuss the interpretation of these simulations by means of one example: the region around the Sr I line at 460.73 nm.

In all simulations we assume that the reflective parasitic light has either been completely eliminated or attenuated to a negligible level by the techniques described in section 5.2.2.

The first simulation (Fig. 5.7) shows the three performance parameters as a function of the prefilter FWHM and of the distance between the main TNF transmission peak and the center of the prefilter transmission profile. The goal is to find the prefilter FWHM yielding the best overall performance. The number of cavities can also be easily included as another free parameter in the simulation. We have however tied down to three cavities mainly because of manufacturing constraints. The first performance parameter (upper panel), the S/N, is not very useful in this type of simulation. Considering this parameter only one would simply opt for a prefilter with a FWHM as large as possible. The second parameter (middle panel) is more instructive. If the prefilter band is too large, the main peak is always significantly diluted by parasitic light, no matter at which spectral position we are observing. If we now require for instance that the main peak should at least contain

80% of the power (red region) the optimum FWHM would lie around 0.15 nm, allowing for a maximum scanning range of about ±0.07 nm around the prefilter center wavelength. The scanning range is sharply delimited by the third performance parameter, the strongest ghost (lower panel). For a prefilter FWHM below some 0.2 nm almost all the parasitic light that shows up in the dilution parameter is contained in this ghost.

Based on this analysis we have chosen a prefilter with a FWHM of 0.16 nm. By doing so we still have about the same performance close to the prefilter center wavelength than with a narrower prefilter passband. Nevertheless, we are able to extend the scanning range a bit further when the localized stray light from the strongest ghost is not a critical issue, for example when the ghost does not coincide with the neighboring Fe I line.

The second simulation (Fig. 5.8) permits to define requirements for the maximum synchronization error, for a given prefilter (our Sr I prefilter in this example). The larger the distance from the prefilter center wavelength, the tighter the requirements. The S/N is the most sensitive parameter. The y-axes represent the different physical parameters that affect the synchronization. Our etalons have an rms CTE in the order of a few nm over the whole aperture (cf. section 2) and the temperature can be stabilized to a few hundreds of a degree. This permits to keep the S/N within about 90% of its maximum possible value. To hold this performance level during scanning the voltage errors must not exceed ±100 Volt. Whereas the high-voltage drivers are stable enough to meet these requirements, the main concern is about the residual errors in the voltage dependence of the optical thickness. The scanning accuracy stated above can however be reached with a careful voltage calibration of the individual etalons as described in detail in section 4.

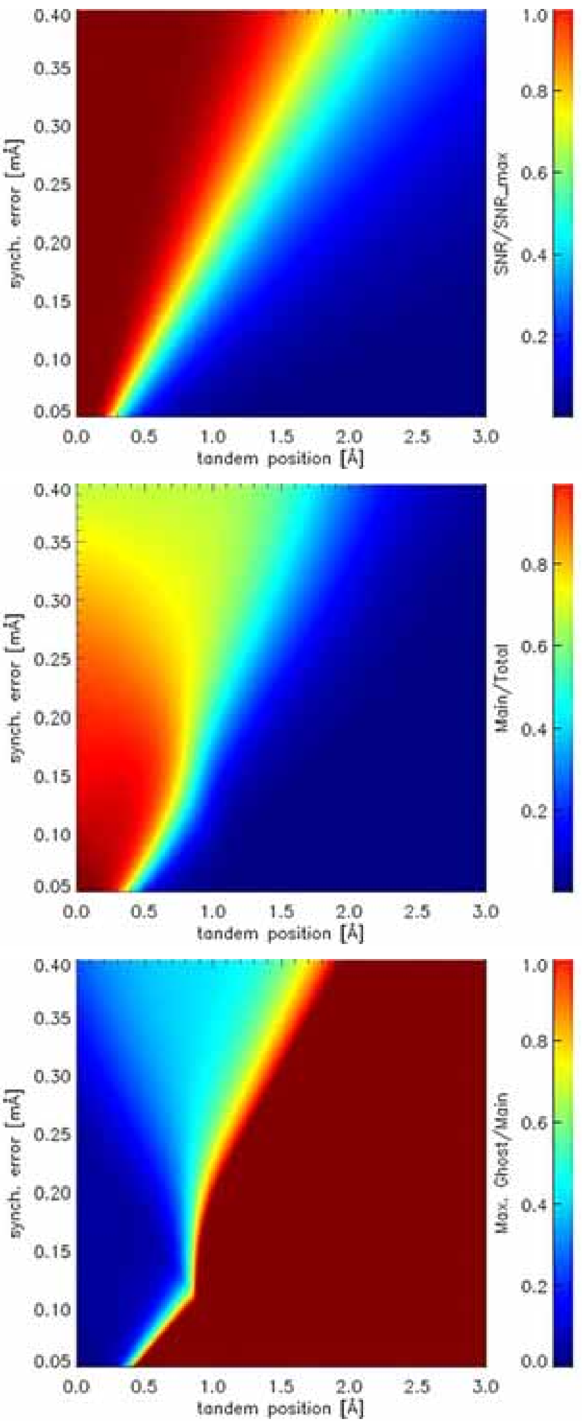


Fig. 5.7: Performance parameters $\sqrt{W_l}$ (top), W_l/W (middle), W_g/W_l (bottom) as a function of the prefilter FWHM and the distance of the tandem FP transmission peak from the center of the prefilter transmission profile. This example is computed for a prefilter centered at Sr I (460.7 nm).

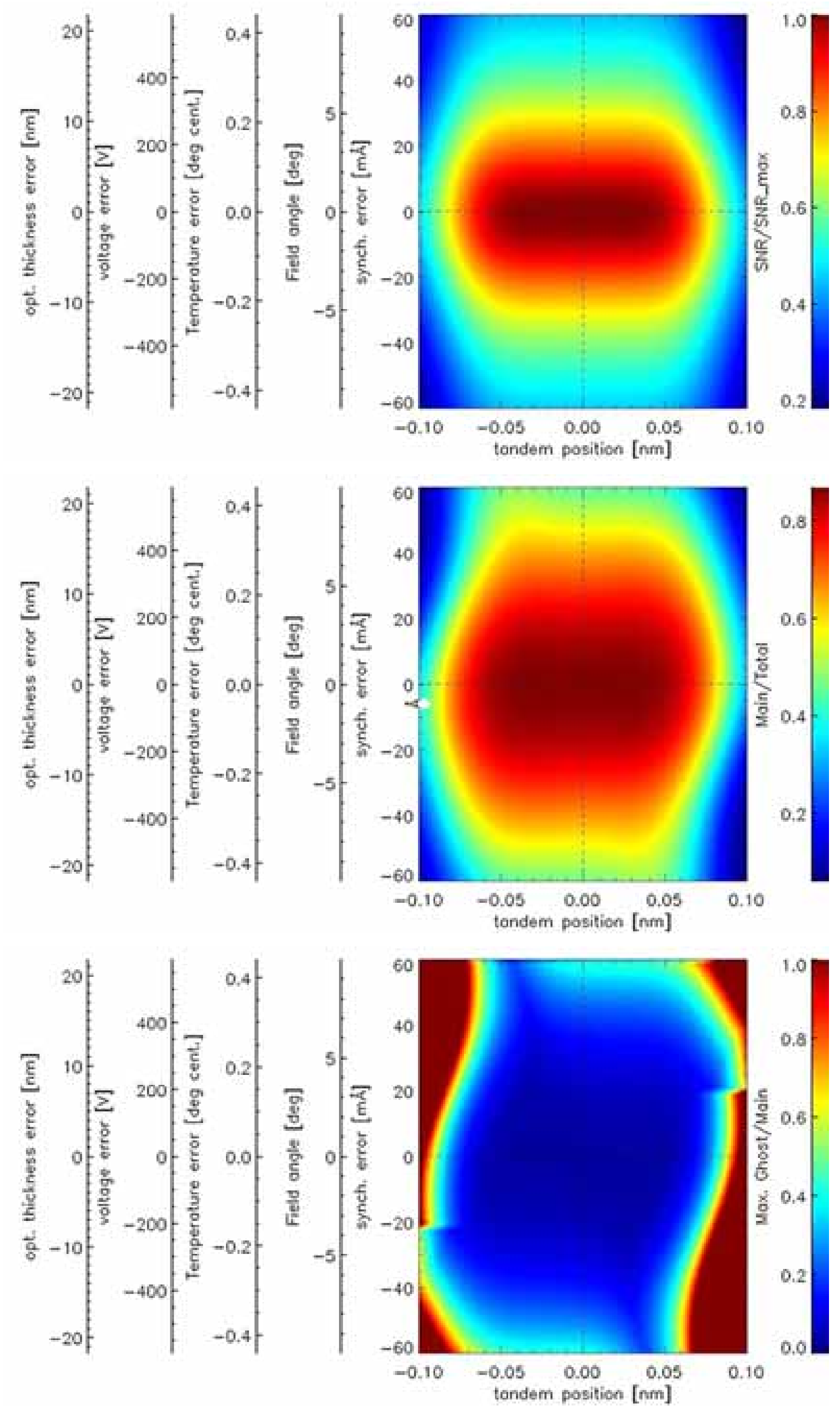


Fig. 5.8: Performance parameters $\sqrt{W_l}$ (top, normalized to the max.), W_l/W (middle), W_g/W_l (bottom) as a function of the synchronization error between the FP etalons and the distance of the tandem FP transmission peak from the center of the prefilter transmission profile. This example is computed for the 460.7 nm prefilter (cf. Tab. 2.3). The y-axes on the left translate the synchronization error to the corresponding approx. errors in optical thickness, FP voltage, temperature and field angle (in the collimated setup).

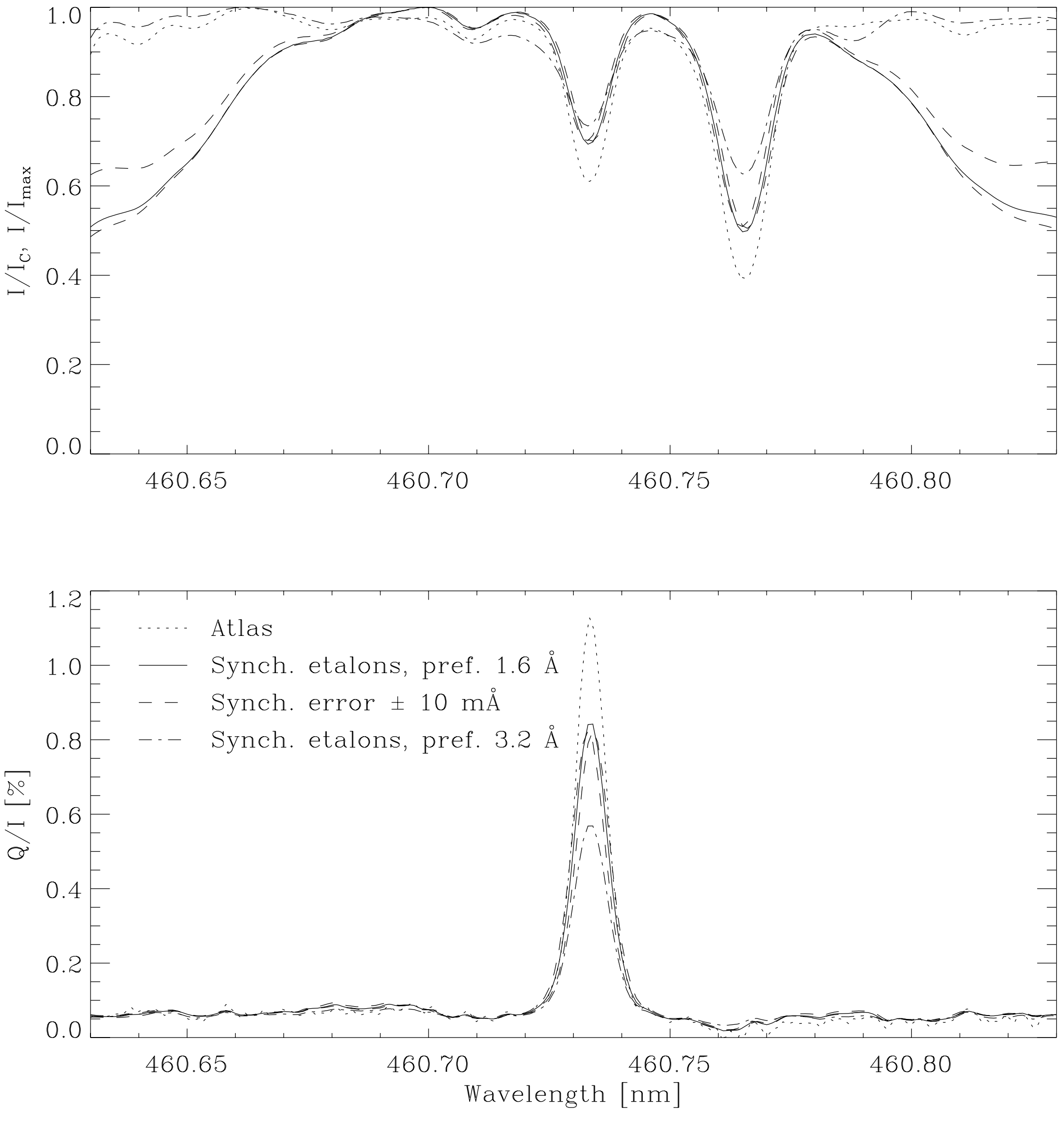


Fig. 5.9: Simulation of a TNF scan through the Sr I prefilter range. The I and Q/I spectra denoted with "Atlas" are taken from the atlas of the second solar spectrum (Gandorfer, 2002). The I spectrum from the atlas is normalized to the continuum intensity I_c, the scanned spectra are normalized to the maximum intensity $I_{\rm max}$.For the actual prefilter FWHM of 1.6 Å three scanned spectra are plotted: one spectrum produced by perfectly synchronized etalons and two spectra where one etalon is off by ± 10 mÅ respectively. To demonstrate the effect of the prefilter bandwidth, another spectrum is shown which corresponds to perfectly synchronized etalons but a prefilter which is twice as broad.

5.3 Imaging performance

In this section the effect of FP etalons on the imaging performance is evaluated in terms of a diffraction model based on Fourier optics (e.g. Goodman, 1988; Gaskill, 1978).

A similar evaluation has been done by Martinez Pillet & al. (2004) for IMaX (collimated setup), by von der Lühe & Kentischer (2000) for TESOS and by Scharmer (2006) for the future stationary FP system at the SST (telecentric setups).

Let us look at the monochromatic wave propagating from an object point. In the plane of the etalon surface (perpendicular to the optical axis) the wave has the (complex) electrical field distribution $E_0(x, y)$. After the etalon the field is

$$E(x, y) = \tau(x, y)\, E_0(x, y) \tag{5.23}$$

We recall that τ is the complex amplitude transmission given by (2.5). The spatial variations of $T = |\tau|^2$ are commonly referred to as *pupil apodization* and its effect on the imaging performance of FP etalons was first investigated by Beckers (1998). The spatial variations in the phase Φ of the complex number τ are commonly called pase error but we will use the expression *wave phase error* (WPE) to avoid a confusion with the phase ϕ i.e. the argument of the amplitude transmission function. Scharmer (2006) has shown that for small errors $\delta\phi$

$$\delta\Phi \approx f(\phi, R)\, \delta\phi \tag{5.24}$$

The WPE amplification function f is maximal at the center of a transmission peak and the maximum value is $R/(1-R)$ where R is the reflectivity of the etalon surfaces. This has an important consequence: the higher the reflectivity, the stronger the effects of CTE or oblique rays (see below) on the imaging performance of a FP etalon.

The *monochromatic* point-spread function (PSF) i.e. the intensity distribution of the wave in the image plane, at a distance d_f from the etalon is

$$p(x, y) \propto \left| \hat{E}\left(\frac{x}{\lambda d_f}, \frac{y}{\lambda d_f} \right) \right|^2 \tag{5.25}$$

where $\hat{E}$ is the Fourier transform of E. We adopt here the convention to normalize the PSF with the total intensity passing through the etalon. Related to the PSF is the *Strehl ratio* which is defined as the ratio between the maximum of the PSF of a real optical system divided by the maximum of the PSF corresponding to diffraction limited imaging (Airy pattern). The basic idea behind the Strehl ratio is the fact that the integral over the PSF is equal to the total intensity transmitted by the optical system and therefore independent of the imaging. Thus, the lower the Strehl ratio, the broader the intensity distribution in the PSF and with it the imaging performance.

Another convenient way to evaluate the imaging performance is by means of the modulation transfer function (MTF) which is proportional to the Fourier transform of the PSF, or equivalently, the autocorrelation of E. The MTF is a measure of the loss in contrast of intensity patterns with different frequencies.

In our collimated setup, each image point is associated to a plane wave covering a large fraction of the etalon aperture. The CTE δnd induce small-scale errors $\delta\phi_{\mathrm{c}}$ in the phase ϕ of the complex amplitude transmission. As we are only looking at the small wavelength interval around one transmission peak we can use the approximation (2.9) to describe the phase errors as

$$\delta\phi_c(x,y) \approx \frac{2\pi}{\Delta\lambda_{\mathrm{FSR}}}\left(\lambda\frac{\delta nd(x,y)}{nd}\right) \tag{5.26}$$

In our evaluation the CTE are taken from the thickness maps of the etalons.

In our telecentric setups, each image point corresponds to a converging wave which only spans a very small fraction of the etalon aperture because of the high f-ratios. In a first approximation the CTE are therefore negligible and the phase errors can be described as

$$\delta\phi_t(x,y) \approx \frac{\pi}{\Delta\lambda_{\mathrm{FSR}}}\lambda\theta^2 \tag{5.27}$$

The angle θ varies over the wavefront as

$$\theta \approx \frac{r}{2f/\#} \tag{5.28}$$

where r is the wavefront radius normalized to unity at the edge and $f/\#$ is the f-ratio.

Fig. 5.10 shows the imaging performance for the collimated setup and for a monochromatic light source sampling the etalon across one transmission peak. It can be seen that the performance strongly varies with wavelength. Close to the center of the transmission peak the PSFs are heavily distorted and broadened. Further away they converge toward the Airy pattern. As we generally observe spectral features which are broader than the etalon passband we have to evaluate the PSF at each wavelength within the passband

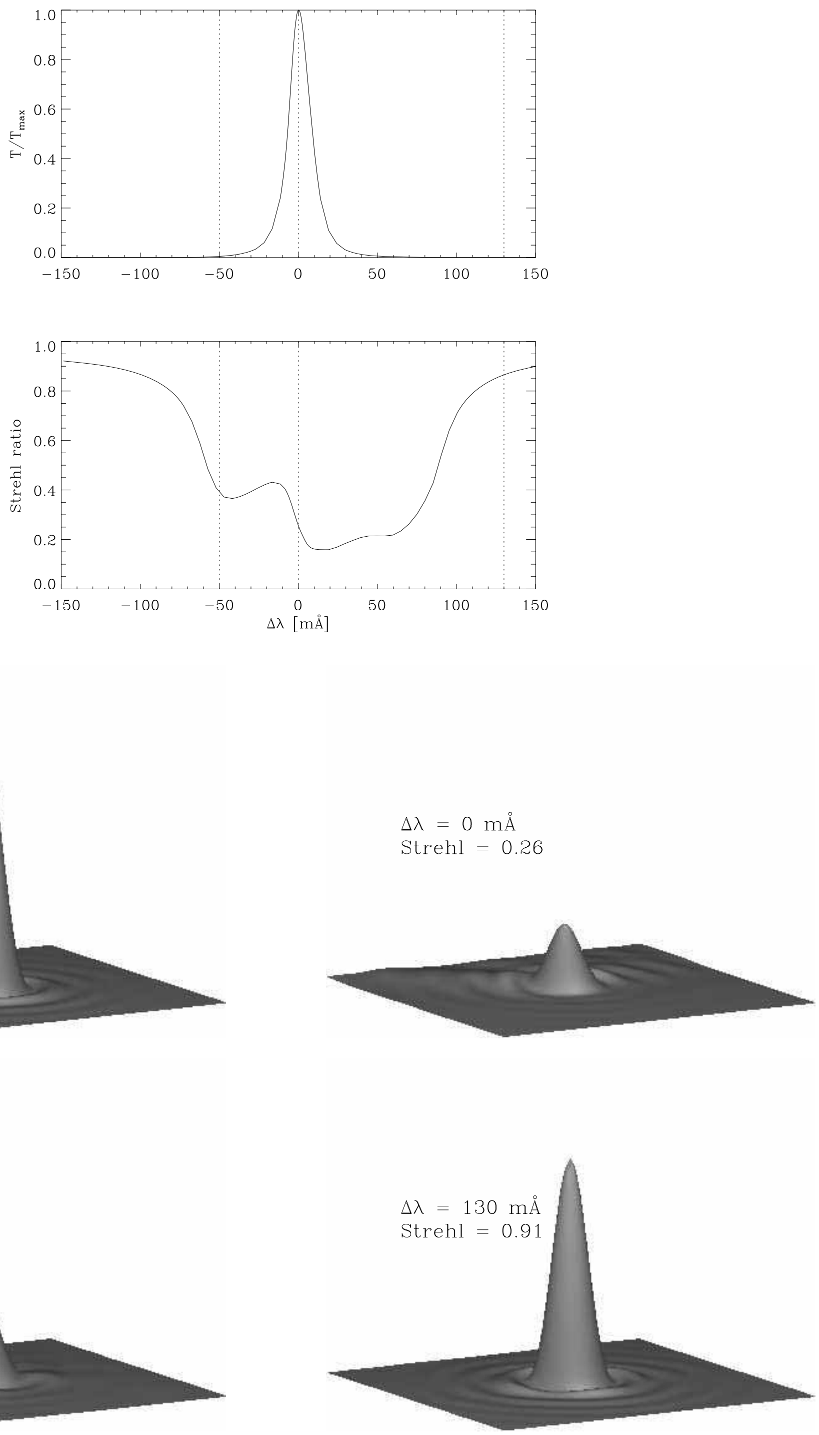


Fig. 5.10: TNF in the collimated setup; monochromatic imaging performance versus position in the transmission peak. **Top panels:** transmission and Strehl ratio. **Bottom panels:** Monochromatic point-spread functions for 3 different spectral positions (vertical lines in the upper plots). The Airy pattern is shown for reference. The x, y and z scales are equal for all images.

and integrate over the wavelengths.

$$P(x,y) = \int_{-\Delta\lambda_{\mathrm{FSR}}}^{\Delta\lambda_{\mathrm{FSR}}} T(\Delta\lambda) p(x,y,\lambda) \mathrm{d}\Delta\lambda \tag{5.29}$$

where T is the etalon transmission. We use the model for a single transmission peak given by (2.11). P is commonly called the *polychromatic* PSF (e.g. von der Lühe & Kentischer, 2000).

The MTF computed from P is shown in Fig. 5.12. When the full etalon aperture of 60 mm is used the image degradation is severe, especially in the low spatial frequencies. The imaging performance can be somewhat improved by limiting the beam to a smaller aperture. Here we show the MTF for an aperture of 40 mm, used in our collimated setup.

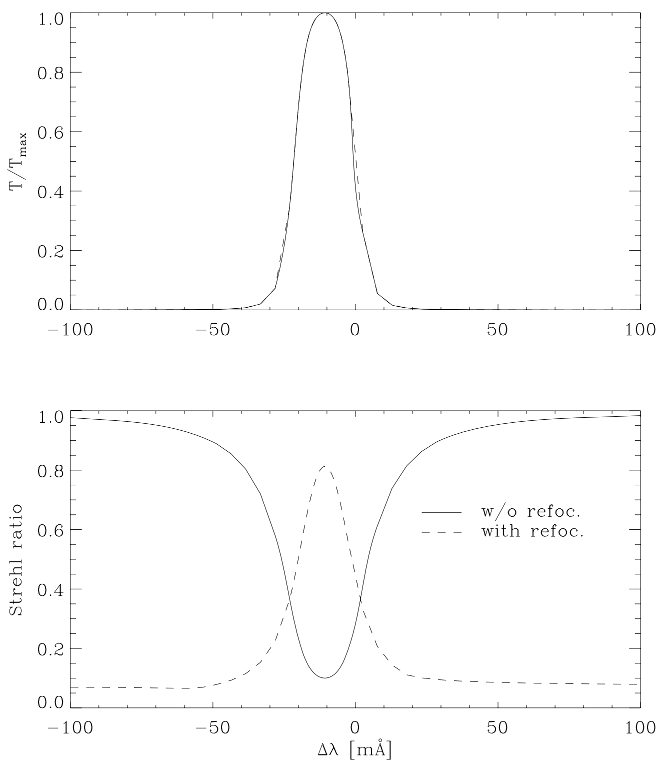


Fig. 5.11: TNF in the telecentric setup with $f/72$; transmission (top panel) and monochromatic Strehl ratio (bottom panel) versus position in the transmission peak. The WPE behaves closely like a focus shift that varies in amplitude across the peak. It can be partially compensated by a refocusing.

Fig. 5.11 shows the result of the monochromatic performance analysis for our telecentric setup, analogous to the collimated case. Here we only show

$f/\#$	k_4	Strehl w/o refoc.	Strehl with refoc
72	1.6	0.1	0.79
108	0.8	0.5	0.94
223	0.19	0.95	1.00

Tab. 5.5: Telecentric setups. Optimum refocusing and imaging improvement in terms of Strehl ratio.

the setup with the lowest magnification ($f/72$) as an example. Again the image degradation is largest close to the peak. In this case it is possible to substantially improve the imaging performance by refocusing (Scharmer, 2006). In table 5.5 we have compiled the computed optimum refocusing and the corresponding Strehl ratio for all three magnifications. The higher the f-ratio i.e. the magnification, the better the refocusing works.

Fig. 5.12 shows the MTFs for the telecentric setup and all three magnifications after an optimum refocusing has been applied. As in the collimated case the performance degradation is affecting above all the low spatial frequencies.

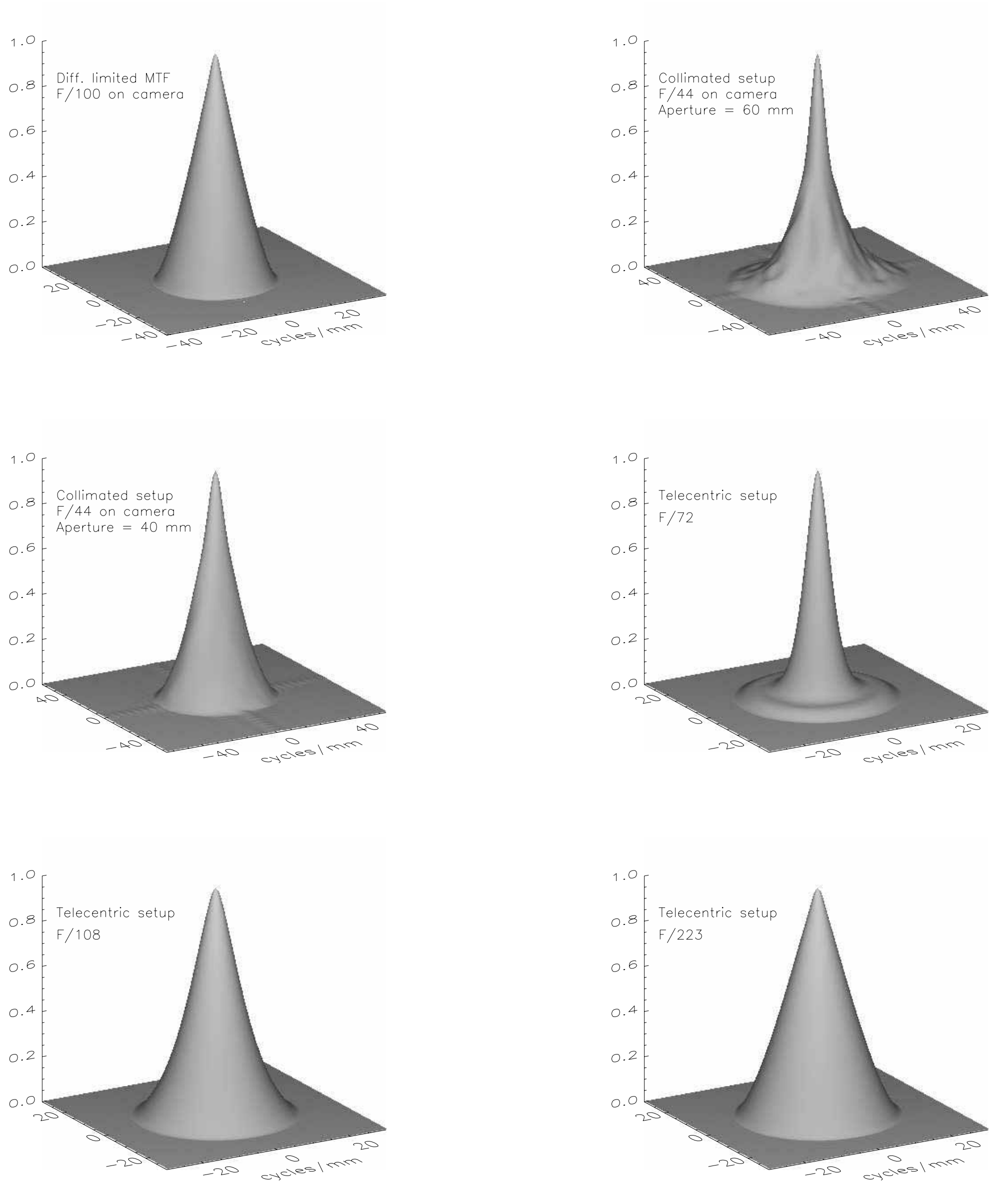


Fig. 5.12: Simulated modulation transfer functions of our TNF for different optical setups. The MTFs are computed for zero field angle. The diffraction limited MTF at F/100 (first panel) is shown for reference, its width scales with $1/f/\#$. The x and y coordinates are in units of cycles/mm in the detector plane.

Chapter 6

Polarimetric calibration

During a polarization measurement the true solar Stokes vector $\boldsymbol{S} = (I, Q, U, V)$ is linearly[1] mapped to the measured intensities $\bar{\boldsymbol{I}} = (\bar{I}_1, \ldots, \bar{I}_n)$:

$$\bar{\boldsymbol{I}} = \mathscr{O}\mathscr{T}\boldsymbol{S}. \tag{6.1}$$

The 4×4 Mueller matrix $\mathscr{T}$ includes in principle everything in between the Sun and the polarization modulator that has an effect on $\boldsymbol{S}$. As the influence of the Earth's atmosphere can be neglected, $\mathscr{T}$ is reduced in practice to the Mueller matrix of the telescope which we will simply refer to as *telescope matrix* in the following. The polarimeter, including the modulator and the detector, is described by the $n \times 4$ *modulation matrix* $\mathscr{O}$. By *polarimetric calibration* we understand the problem of defining $\mathscr{O}$ and $\mathscr{T}$ and then solving (6.1) for $\boldsymbol{S}$. This problem is split up into independent stages.

The *polarimeter calibration*, described in section 6.1 deals with finding the elements of $\mathscr{O}$, with the help of calibration optics inserted in front of the polarization modulator. This process is completely decoupled from the telescope calibration because the calibration optics are generating a 100% polarized state which means that no previous polarization effects of the telescope are leaking through. The calibration optics are not imperatively required to be in front of the modulator. Any components in between these optics and the analyzer of the modulator are automatically included in the calibration as integral part of the polarimeter.

After the polarimeter calibration, the Stokes vector in front of the modulator is known which then permits to find the elements of $\mathscr{M}_{\mathrm{T}}$ within the scope of a *telescope calibration.* Two different approaches are actually used to this end. The first approach, based on a-priori assumptions about certain characteristics of the true Stokes parameters, is working well on spectropolarimetric data. The second approach, which makes use of a model for the telescope matrix, is the preferred method for monochromatic imaging polarimetry, as the first approach mostly fails here due to the sparsely available spectral line information. The model approach is therefore extensively discussed in section 6.2 and its practical suitability demonstrated with a model of the SST, valid for the whole working range of our TNF.

---

[1] The influence of detector nonlinearities on polarization measurements has been studied by Keller (1996) and the author has defined the boundary conditions where the nonlinearities can be neglected. These conditions are almost always met by our current ZIMPOL system and we aware of the cases where one has to exercise caution.

6.1 Polarimeter

6.1.1 General procedure

As the telescope polarization is neutralized by the calibration optics in front of the modulator the problem we have to solve at this stage of the polarimetric calibration reduces to

$$\bar{\boldsymbol{I}} = \mathscr{O}\boldsymbol{S}. \tag{6.2}$$

This corresponds exactly to (6.1), but stripped from the telescope matrix, and where $\boldsymbol{S}$ is the Stokes vector generated by the calibration optics. The general procedure is to determine the elements $O_{i,j}$ of the modulation matrix by measuring a set of m well-known polarization states and then solving the linear system

$$\mathscr{C}\begin{pmatrix} O_{11} \\ O_{12} \\ \vdots \\ O_{n4} \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{I}}_1 \\ \vdots \\ \bar{\boldsymbol{I}}_m \end{pmatrix} \tag{6.3}$$

for the O_{ij}, where the $mn \times 4n$ *calibration matrix* $\mathscr{C}$ is defined by the Stokes vectors $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m$. The concrete expressions for the elements C_{ij} can be easily found by algebraically evaluating (6.2) for each calibration measurement $k = 1 \ldots m$, defined by the known vectors $\boldsymbol{S}_k$ and $\bar{\boldsymbol{I}}_k$ on the right-hand and left-hand side respectively. In principle the problem can be solved in the form (6.3) with a minimum of $m = 4$ independent[2] calibration measurements.

In practice however it is difficult to rely on absolute intensities. Normally the Sun is used as a light source and the absolute intensities can vary significantly between the different calibration measurements. One could think of overcoming this complication by using a second detector on a reference channel that is branching off between the calibration optics and the modulator. In imaging polarimetry however we want to calibrate for each image element separately which means that the reference image has to be mapped exactly on the polarimetric images. In short, as the investment for obtaining accurate absolute intensity images is considerable, it is well worth thinking of a way to avoid this.

---

[2] Independent in the sense that the $\boldsymbol{S}_k$ have to be chosen in a way to obtain a full-ranked matrix $\mathscr{C}$.

As the polarization state is fixed by the calibration optics the intensity variations manifest themselves as a time-dependent scaling of the Stokes vector

$$\boldsymbol{S} = I(t)\,\boldsymbol{s} \tag{6.4}$$

where $\boldsymbol{s}$ denotes the normalized Stokes vector $(1, q, u, v)$ with $q = Q/I$, $u = U/I$ and $v = V/I$. When we do not want to rely on absolute intensity measurements we have to introduce m additional unknowns $I_1, \dots I_m$ into (6.3), ending up with a nonlinear system needing at least $4n/(n-1)$ independent calibration measurements. However, in general we are only interested in the normalized Stokes parameters q, u and v, which permits us to reduce and linearize the problem, as will be shown in the following.

When giving up absolute intensity information, the problem becomes scalable in a way that we can always divide (6.2) by the first measured intensity $\bar{I}_1$, without loss of generality, to obtain the following set of equations:

$$\bar{i}_j = \frac{O_{j1} + O_{j2}\,q + O_{j3}\,u + O_{j4}\,v}{O_{11} + O_{12}\,q + O_{13}\,u + O_{14}\,v}, \quad j = 1, \dots, n \tag{6.5}$$

with $\bar{i}_j = \bar{I}_j/\bar{I}_1$. Again without loss of generality, we can normalize the matrix elements $o_{jl} = O_{jl}/O_{11}$ to obtain a new linear system in the $(4n-1)$ unknowns $o_{12}, \dots, o_{n4}$. Let us now define the vector $\bar{\boldsymbol{i}} := (i_2, \dots, i_n)$ by just throwing away the first intensity $\bar{i}_1$, which is equal to 1 by definition. As before, we repeat this procedure for each calibration measurement $k = 1 \dots m$ with known $\boldsymbol{s_k}$ and $\boldsymbol{i}_k$ to finally obtain the full linear system

$$c \begin{pmatrix} o_{12} \\ o_{13} \\ \vdots \\ o_{n4} \end{pmatrix} = - \begin{pmatrix} \bar{\boldsymbol{i}}_1 \\ \vdots \\ \bar{\boldsymbol{i}}_m \end{pmatrix} \tag{6.6}$$

with the $m(n-1) \times (4n-1)$ *reduced calibration matrix* c whose elements can be determined by algebraically evaluating (6.5) for each calibration measurement. This reduced calibration problem can be solved with a minimum of $m = (4n-1)/(n-1)$ independent calibration measurements.

As c can be ill-conditioned in some cases, even with carefully chosen $\boldsymbol{s}_k$, the numerical solution of (6.6) may become unstable. We therefore apply a singular value decomposition (SVD) on c:

$$c = \mathscr{U}\mathscr{W}\mathscr{V}^T \tag{6.7}$$

with $\mathscr{U}$, $\mathscr{V}$ orthogonal $m(n-1) \times (4n-1)$ and $(4n-1) \times (4n-1)$ matrices respectively and $\mathscr{W}$ the diagonal $(4n-1) \times (4n-1)$ matrix, containing the positive or zero singular values [3].

When c is determined, the final step of the polarimeter calibration is to find $\boldsymbol{s}$ by solving the linear system

$$\begin{pmatrix} 1 \\ \bar{\boldsymbol{i}} \end{pmatrix} = o\,\mu\boldsymbol{s} \tag{6.8}$$

o is the $n \times 4$ *reduced modulation matrix* with $o_{11} = 1$ and the other elements defined by (6.6). The previously applied scaling of the calibration problem appears here in the form of a factor $\mu = o_{11}/\bar{I}_1$, which will disappear again by solving for $\mu\boldsymbol{s}$ and then normalizing to $\boldsymbol{s}$. To numerically solve (6.8) we also apply a SVD $o = \mathscr{U}\mathscr{W}\mathscr{V}^T$. Whereas numerical stability is not an issue here (unless the modulator is badly designed), the SVD technique has the advantage that it allows to implement a common solution algorithm for (6.8) for all types of modulators.

The solution, expressed in terms of the SVD components, is

$$\boldsymbol{s} = \mathscr{V}\mathscr{W}^{-1}\mathscr{U}^T \begin{pmatrix} 1 \\ \bar{\boldsymbol{i}} \end{pmatrix} =: o^{-1} \begin{pmatrix} 1 \\ \bar{\boldsymbol{i}} \end{pmatrix} \tag{6.9}$$

where $\mathscr{W}^{-1}$ is the diagonal matrix defined by $W_{jj}^{-1} = 1/w_{jj}$, $j = 1 \ldots 4$ We branch the discussion of (6.9) out to the different modulator types, as there are important differences in the interpretation of the respective solution $\boldsymbol{s}$.

6.1.2 Different modulator types

True full-Stokes modulator When o is full-ranked, we speak of a *true full-Stokes modulator.* In other words, a true full-Stokes modulator permits to measure all Stokes parameters simultaneously[4]. In the case of a true full-Stokes modulator, the system (6.8) is either determined or overdetermined, which means that (6.9) either delivers the unique solution or the closest solution in the least squares sense, in analogy to the above discussion of the solution of (6.6).

---

[3] It is out of the scope of this thesis to present the SVD in detail, an extensive treatise of this well-established technique can be found in textbooks of numerical mathematics.

[4] Simultaneously in the sense that intensity variations of $\boldsymbol{S}$ during the $\bar{\boldsymbol{I}}$ measurement can be neglected, which is the case for all ZIMPOL modulators due to their operation at frequencies well above the seeing frequency spectrum.

Partial-Stokes modulator When o has a rank inferior to 4, we will call the modulator a *partial-Stokes modulator.* In this case (6.8) is underdetermined and one or more singular values (diagonal elements of $\mathscr{W}$) are equal to zero. The states of polarization that the modulator is sensitive and not sensitive to can directly be read off the SVD. The Stokes vectors that can be measured form the rank of o and are spanned by the columns j of $\mathscr{U}$ whose corresponding singular values W_{jj} are nonzero. The Stokes vectors that cannot be measured form the nullspace of o and are spanned by the columns l of $\mathscr{V}$ whose corresponding singular values W_{ll} are zero.

It is important to note that a modulator can very well be a partial modulator only, even though it is sensitive to some q, u and v signals. For instance, a mixture of q and u can easily happen, when the modulator is slightly rotated relative to the Stokes reference system defined by the calibration optics.

PEM as pseudo full-Stokes modulator Let us consider a Stokes modulator where the assumption that all Stokes parameters are recorded simultaneously within the same modulation process does not hold. This is for instance the case for our PEM modulator, which is often used in the following mode:

1. IQV measurement
2. Rotate the modulator by 45 degrees
3. IUV measurement

where the time passing between steps 1 and 3 is a few seconds.

Let us denote the intensities measured in IQV and IUV mode by $\bar{\boldsymbol{I}}_{1\ldots3}$ and $\bar{\boldsymbol{I}}_{4\ldots6}$. As the length of the Stokes vector $\boldsymbol{S}$, generated by the calibration optics, can vary between the IQV and IUV measurements due to absolute intensity variations, (6.2) with $\bar{\boldsymbol{I}} = \bar{I}_1 \ldots \bar{I}_6$ is not a valid description of the calibration problem. This special case can however be handled by splitting up the problem into two separate calibration problems, well described by (6.2), namely one for the IQV part with $\bar{\boldsymbol{I}} = \bar{\boldsymbol{I}}_{1\ldots3}$ and one for the IUV part with $\bar{\boldsymbol{I}} = \bar{\boldsymbol{I}}_{4\ldots6}$ and then combining the independently obtained modulation matrices o_1 and o_2 as will be shown in the following.

When reliable absolute intensities are available from a reference channel, one can combine the modulation matrices o_1 and o_2 after a correct scaling of o_2:

$$o = \begin{pmatrix} o_1 \\ \frac{I_2}{I_1}\, o_2 \end{pmatrix} \tag{6.10}$$

If not, one can at least seek out a best-guess solution by determining the scaling factor for o_2 which minimizes the residuals of all available calibration measurements in a least squares sense. So let us take all the m calibration measurements, defined each by the pair $(\bar{\boldsymbol{i}}_k, \boldsymbol{s}_k)$, $k = 1, \ldots, m$, and the two modulation matrices o_1 and o_2, found by splitting the calibration problem as described above. The unknown $2m$ factors $\mu_{1,1\ldots3}$, $\mu_{1,4\ldots6}$, $\mu_{2,1\ldots3}$, $\ldots$, $\mu_{k,4\ldots6}$ of (6.8) are all assumed to be randomly distributed around a common value μ. Let us now arrange the calibration data in the form

$$\bar{\boldsymbol{i}}_{k,1\ldots3} = o_1\, \mu\, \boldsymbol{s}_k \tag{6.11}$$

$$\bar{\boldsymbol{i}}_{k,4\ldots6} = o_2\, \mu\, \frac{O_{2,11}}{O_{1,11}}\, \boldsymbol{s}_k \tag{6.12}$$

The best-guess of the scaling factor $O_{2,11}/O_{1,11}$ between the two modulation matrices o_1 and o_2 is found by solving the systems (6.11) and (6.12) in the least squares sense for μ and $\mu\,(O_{2,11}/O_{1,11})$ and then dividing the second solution by the first one:

$$\frac{O_{2,11}}{O_{1,11}} = \frac{(\bar{\boldsymbol{i}}_{1,4\ldots6}, \ldots, \bar{\boldsymbol{i}}_{m,4\ldots6}) \cdot (o_2\boldsymbol{s}_1, \ldots, o_2\boldsymbol{s}_m)}{(\bar{\boldsymbol{i}}_{1,1\ldots3}, \ldots, \bar{\boldsymbol{i}}_{m,1\ldots3}) \cdot (o_1\boldsymbol{s}_1, \ldots, o_1\boldsymbol{s}_m)} \cdot \frac{(o_1\boldsymbol{s}_1, \ldots, o_1\boldsymbol{s}_m)^2}{(o_2\boldsymbol{s}_1, \ldots, o_2\boldsymbol{s}_m)^2} \tag{6.13}$$

After this sticking point specific to pseudo full-Stokes modulation has been overcome, we can combine the two modulation matrices and the corresponding measured intensities to

$$\begin{pmatrix} 1 \\ \bar{\boldsymbol{i}} \end{pmatrix} = \begin{pmatrix} o_1 \\ \frac{O_{2,11}}{O_{1,11}}\, o_2 \end{pmatrix} \mu \boldsymbol{s} \tag{6.14}$$

which is equivalent to (6.8) and go on from there with the SVD inversion as described before.

The random distribution of the absolute intensity variations during the calibration measurements is the critical assumption this solution is based upon. It could possibly happen that our "best-guess" is not a very good one, for example when the intensity is changing systematically during the calibration measurements (e.g. due to clouds, ...). But, in default of reliable absolute intensity information, it seems to be the most rational thing to do and it has well stood the test in practice, as will be demonstrated with a concrete example of a PEM calibration in the next section.

	$\boldsymbol{s}_1$	$\boldsymbol{s}_2$	$\boldsymbol{s}_3$	$\boldsymbol{s}_4$	$\boldsymbol{s}_5$	$\boldsymbol{s}_6$
q	1	-1	0	0	0	0
u	0	0	1	-1	0	0
v	0	0	0	0	1	-1

Tab. 6.1: The 6 Stokes vectors generated by the calibration optics for the PEM calibration.

6.1.3 PEM calibration

To illustrate the ideas of the previous sections and to demonstrate their practicability, we are now going through a concrete example of a PEM calibration. We have a look at the PEM used as pseudo full-Stokes modulator at the SST and at 460.7 nm. In this case no reference channel was available, so we resort to the technique described in the previous section.

We perform 2×6 calibration measurements, generating 6 different Stokes vectors $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_6$ for each PEM mode, IQV and IUV. The $\boldsymbol{s}_k$ are outlined in Tab. 6.1. With $m = 6$ calibration measurements and $n = 3$ intensity measurements per PEM mode we solve the linear system (6.6) with $(n-1)\,m = 12$ independent equations for the $4n-1 = 11$ unknown elements of o_1 and o_2. We explicitly construct here the reduced calibration matrix c_1, to clarify the process. So, by successively evaluating (6.5) for each calibration measurement:

$$c = \begin{pmatrix} \bar{i}_{1,2} & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\bar{i}_{2,2} & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \bar{i}_{3,2} & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\bar{i}_{4,2} & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{i}_{5,2} & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\bar{i}_{6,2} & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \bar{i}_{1,3} & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ -\bar{i}_{2,3} & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & \bar{i}_{3,3} & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & -\bar{i}_{4,3} & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & \bar{i}_{5,3} & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & -\bar{i}_{6,3} & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix} \tag{6.15}$$

The matrix c_2 is constructed in a completely analogous way and has the exact same form despite replacing the $\bar{i}_{k,2}$ and $\bar{i}_{k,3}$ by the $\bar{i}_{k,5}$ and $\bar{i}_{k,6}$.

Tab. 6.2 shows the resulting reduced modulation matrix o, after the combination of the individual solutions o_1 and o_2 with the technique described

$$\begin{pmatrix} 1.000 & 0.043 & -0.038 & -0.022 \\ 0.018 & 0.488 & -0.060 & 0.083 \\ -0.010 & -0.029 & 0.264 & 0.308 \\ 1.025 & -0.054 & 0.001 & -0.008 \\ -0.003 & 0.102 & 0.404 & -0.247 \\ -0.005 & 0.016 & 0.219 & 0.360 \end{pmatrix} \begin{pmatrix} 0.000 & 0.012 & 0.033 & 0.012 \\ 0.003 & 0.009 & 0.012 & 0.010 \\ 0.003 & 0.017 & 0.069 & 0.056 \\ 0.035 & 0.061 & 0.012 & 0.009 \\ 0.005 & 0.007 & 0.058 & 0.092 \\ 0.003 & 0.014 & 0.080 & 0.054 \end{pmatrix}$$

Tab. 6.2: Mean (left matrix) and rms (right matrix) values of the reduced PEM modulation matrix for 460.7 nm. To facilitate the interpretation of the matrix, the raw measured intensities are combined in a way that rows 2 and 3 are closely representing the Stokes parameters q, v and rows 5 and 6 the parameters u, v. The elements (3,3), (5,4) and (6,3) show that this setup is suffering from a significant crosstalk between Stokes U and V, which is caused by oblique mirrors in the beam path between the calibration optics and the modulator.

in the previous section. The calibration is done for each pixel separately, so o appears in the form of a 6×4 matrix of images, each image representing one element o_{kl}. Here we only show the mean and rms of each element o_{kl}.

residuals for	s_1	s_2	s_3	s_4	s_5	s_6
q	-0.059	-0.012	-0.021	-0.045	0.032	0.022
u	0.028	0.015	-0.024	0.008	-0.012	-0.002
v	0.003	-0.004	0.002	0.004	0.005	0.010

Tab. 6.3: Mean Residuals of the PEM calibration with respect to the six calibration measurements $s_1, \ldots, s_6$.

The residuals with respect to the calibration measurements are displayed in Tab. 6.3. They are defined as the difference between the reproduced Stokes vectors, obtained by applying the demodulation matrix on the measured intensities of the calibration measurements, and the true Stokes vectors s_k (cf. Tab. 6.1). As can be seen from this table all the residuals are in the order of percent.

Another test of the calibration technique is shown in Fig. 6.1. Series of measurements with rotating calibration optics are compared with the theoretical curves. The residuals are found to be in the order of percent, which is consistent with the residuals of Tab. 6.3.

The question whether the calibration is good enough or not, depends of course on what we want to measure. Let us illustrate this with a concrete scenario: we are interested in detecting structures of scattering polarization whose amplitude in polarization degree is in the order of 10^{-3}. Let us assume that we are observing an active region at the limb showing strong Zeeman signals. After the polarimeter calibration we have to be suspicious about any structures at the 10^{-4} level, which is the level the Zeeman signatures are reduced to by the residual calibration errors. In the case of the SST observations the Zeeman signatures always are an order of magnitude fainter, which means that we can trust the structures in the calibrated images down to amplitudes of 10^{-5}, which is far below the noise level.

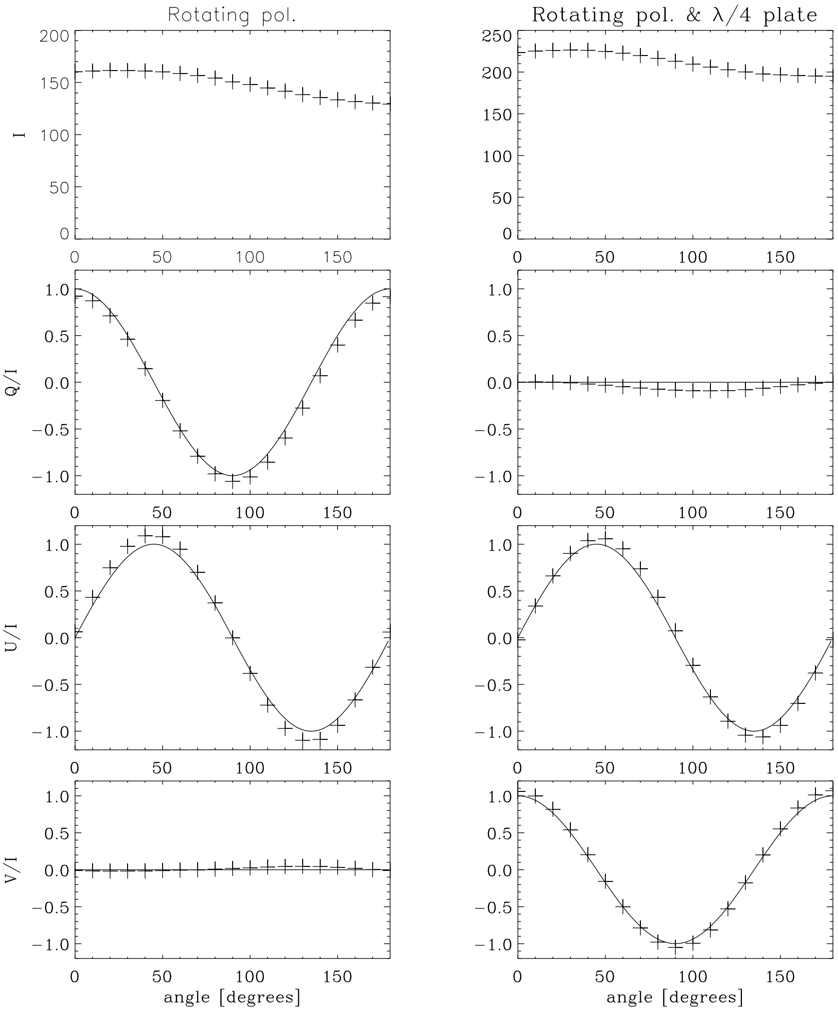


Fig. 6.1: As another test of the polarimetric calibration two series of special measurements are recorded with the calibration optics. In the first series, represented by the panels in the left column, the polarizer is rotated between 0 and 180 degrees in steps of 10 degrees and a Stokes image is recorded for each orientation. In the second series (right column) the same polarizer is rotated again but after the polarizer the light is going through the fixed quarterwave plate of the calibration optics. All these measurements are calibrated with the procedure described in this section. The plots in this figure show the results (crosses) together with the theoretical curves. The residuals are in the order of percent.

6.2 Telescope

In general the optical components of a telescope modify the true incoming solar Stokes vector in a significant way. On most telescopes the polarization effects are changing during the day, as the telescope is constantly changing its orientation to follow the Sun. To get rid of this *telescope polarization* several strategies have been developed.

The best way is of course to avoid any telescope polarization by design. Examples of such polarization-free telescopes are the Richard B. Dunn telescope (Sacramento Peak, USA) or the French THEMIS (Tenerife, Spain). The Gregory-Coudé telescope at IRSOL is not completely free from polarization effects but they only depend on the Sun's declination and are therefore practically constant during one observing day.

Another way consists in introducing optical elements in front of the modulator to compensate the telescope polarization (e.g. Gisler, 1999; Thalmann, 2008).

The most commonly used method is to develop a model of the telescope polarization. To do this, two different approaches are used. The first approach, (Holzreuter, 2003; Stenflo et al., 2002; Holzreuter & Stenflo, 2007) is based on a priori assumptions about the Stokes parameters: one knows that the true spectral Zeeman profiles must have a symmetric shape in linear polarization and an antisymmetric shape in circular polarization, or that coherent scattering produces linear polarization only, with an overall orientation parallel to the solar limb. With the additional assumption that the telescope matrix conserves the total degree of polarization one can then determine a unique rotation of the vector (q, u, v) on the Poincaré sphere so that the a-priori conditions in the Stokes parameters are met or closely met in a least-squares sense.

In narrow-band imaging polarimetry Stokes images are often recorded at a few spectral positions only to keep the total integration time at a minimum. When the spectral information is not sufficient for the a-priori method to work satisfactorily we resort to a telescope model in form of a Mueller matrix (called *telescope matrix* hereafter) which is obtained by multiplying the Mueller matrices of the individual optical components of the telescope. The unknown physical parameters of some components are considered as free parameters and the model is then fitted to calibration data where the true Stokes vector in front of the telescope is known. This approach has been applied to some solar telescopes hosting ZIMPOL observing campaigns like e.g. the McMath-Pierce telescope (Bernasconi, 1997), the Gregory-Coudé telescope at IRSOL (Sanchez Almeida et al., 1991) and, very recently to the Swedish Solar Telescope (Selbing, 2005), hosting the first observing campaign with our TNF. This approach is however not free of problems. For instance the parameters characterizing a mirror surface can change significantly in time due to contamination by a thin oil film or to progressive aluminum oxidation. Therefore, at each observing campaign the parameters should be determined again.

In the following we present a polarization model of the Swedish Solar Telescope, adapted to our ZIMPOL polarimeter setup, and the model calibration procedure for 422 nm, the position of the Ca I line observed during our 2006 campaign.

6.2.1 Model of the Swedish Solar Telescope

The SST model used in this work is described in detail by Selbing (2005). It includes the main lens, the elevation and azimuth mirrors, field mirror, Schupmann corrector and the field lens (cf. Fig. 6.2). The AO system is not included as it is either behind our analyzer or in between the polarimeter calibration optics and our modulator and therefore part of the polarimeter calibration.

The Stokes entrance reference frame in front of the telescope is fixed to the main lens. Positive Q is vertical when the turret is at 0° elevation. To switch to a reference frame fixed to the Sun with positive Q along the solar meridian, one just has to rotate the lens reference frame by an angle $s - p$. Here p is the angle between the declination and vertical circles through the Sun center, measured ccw from the vertical circle and s is the angle between the declination and the solar meridian, measured ccw from the solar meridian. The angle $s - p$ (mainly p) changes during the day and is conveniently logged by the turret control software. The exit reference frame after the field lens is defined by the positive Q direction along terrestrial North-South.

The models of the individual optical components are discussed by Selbing (2005). In this place we just highlight the most important features. The main lens does clearly not behave as a single retarder with one global retardation and axis, as the Stokes images of the pupil in Fig. 6.4 demonstrate. The stress in the glass, mainly produced by the vacuum load, with possible contributions from gravitational forces or from the manufacturing process, is inducing a complex non-symmetrical birefringence distribution. Therefore, the lens is described with a 5-parameter Mueller matrix $\mathscr{L}$, which is deduced from an

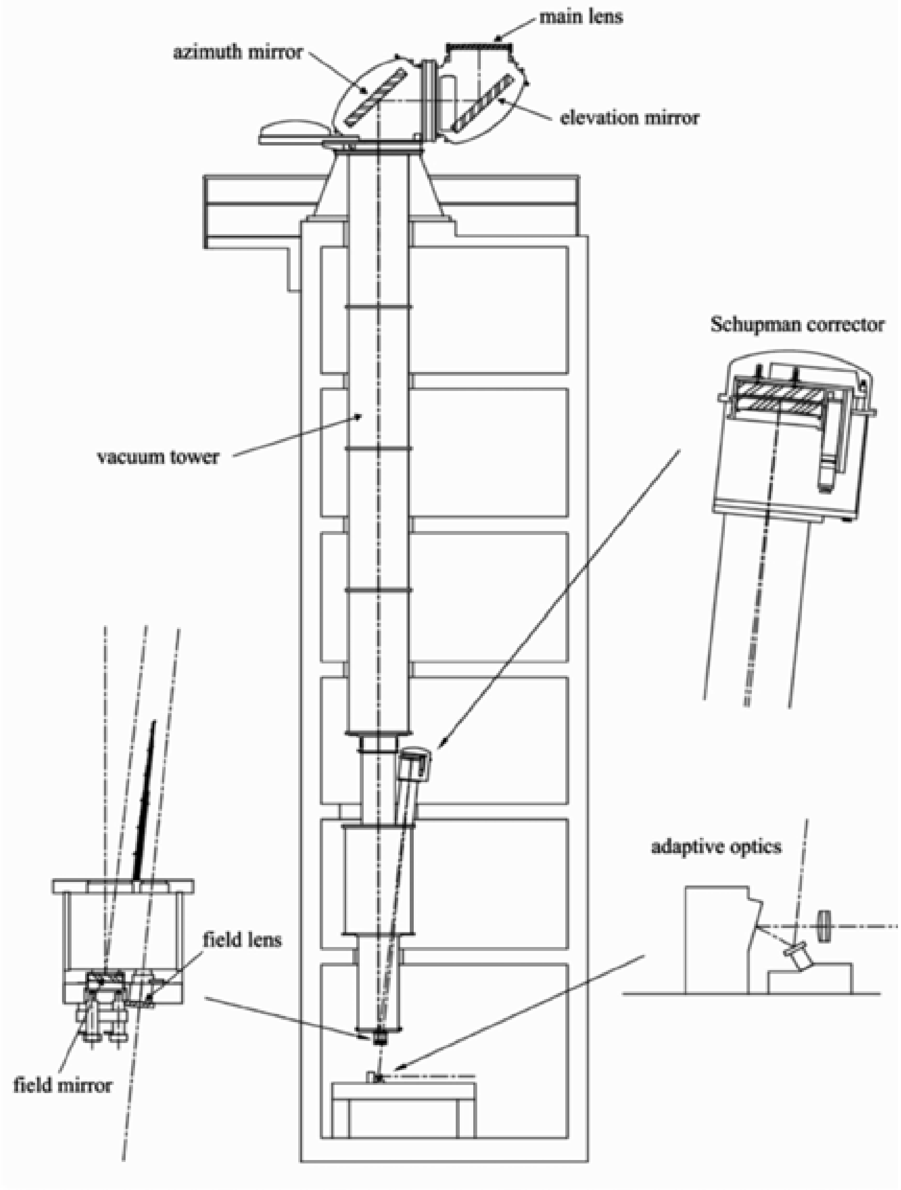

main lens
azimuth mirror
elevation mirror
Schupman corrector
vacuum tower
adaptive optics
field lens
field mirror

Fig. 6.2: SST optical setup. The sunlight enters the telescope through the main lens which acts as a vacuum window at the same time. The beam is then deflected by the elevation and azimuth mirrors down to the field mirror, at the focus of the main lens. The field mirror projects the entrance pupil onto the Schupmann system (consisting of a lens and a mirror) for achromatic correction. The Schupmann corrector generates a secondary focus on the field lens which is also the vacuum exit window. All these optical components are included in the telescope matrix for polarization correction.

aperture average of local retarders with arbitrary retardance and axis:

$$\mathscr{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & A & B & -C \\ 0 & B & D & E \\ 0 & C & -E & A+D-1 \end{pmatrix} \tag{6.16}$$

A better approach is to model each image element (or a reasonable subaperture) of the pupil image as an individual retarder and then averaging the corresponding Mueller matrices over the lens aperture. The big advantage of this approach is that the resulting model can easily be converted to other wavelengths. It is however not used in this work because - unfortunately - we did not have this insight during the 2006 campaign and have performed most calibration measurements with the camera in the telecentric beam of the science setup, i.e. without imaging the pupil.

For the mirrors the so called *free mirror model* is used. Its Mueller matrix is

$$\mathscr{M} = \frac{1}{2} \begin{pmatrix} 1+R & 1-R & 0 & 0 \\ 1-R & 1+R & 0 & 0 \\ 0 & 0 & -2\sqrt{R}\cos\delta & 2\sqrt{R}\sin\delta \\ 0 & 0 & -2\sqrt{R}\sin\delta & -2\sqrt{R}\cos\delta \end{pmatrix} \tag{6.17}$$

with the two free parameters R, related to linear polarization effects and δ, describing a retardation. This model better describes a mirror which deviates from a strict physical mirror, e.g. due to surface contaminations.

The Schupmann corrector, consisting of a lens and a mirror is modeled as a single zero-degree mirror $\mathscr{M}_0$, i.e. a free mirror with $R = 1$ and $\delta = 0$. This is reasonable because the mirror is illuminated virtually at normal incidence. One the other hand the Schupmann lens, which is entirely embedded in the vacuum tower and therefore lacking any major mechanical stress, is assumed to have a negligible effect on polarization. The field lens is not considered at all. Although it is exposed to pressure stress, like the main lens, its effects on polarization are assumed to be negligible, as the ratio between thickness and diameter is much higher than for the main lens.

The entire telescope matrix is finally modeled as

$$\begin{aligned} \mathscr{T} \;=\; & \mathscr{R}(\theta_{\mathrm{mod}})\, \mathscr{M}_0\, \mathscr{R}(-\theta_{\mathrm{fld}}) \mathscr{M}(R_{\mathrm{fld}}, \delta_{\mathrm{fld}})\, \mathscr{R}(\theta_{\mathrm{fld}}) \\ & \mathscr{R}(\theta_{\mathrm{az}})\, \mathscr{M}(R_{\mathrm{az}}, \delta_{\mathrm{az}})\, \mathscr{R}(\pi/2 - \theta_{\mathrm{el}}) \\ & \mathscr{M}(R_{\mathrm{el}}, \delta_{\mathrm{el}})\, \mathscr{L}(A, B, C, D, E) \end{aligned} \tag{6.18}$$

where $\mathscr{R}(\theta)$ is the Mueller matrix that rotates the Stokes reference frame ccw by an angle θ and $\mathscr{M}_0, \mathscr{M}, \mathscr{L}$ are the Mueller matrices of mirrors and the

Parameters	Description
θ_{fld}	Angle of the plane of incidence of the field mirror measured ccw from terrestrial N-S
$\mathbf{R_{fld}}, \boldsymbol{\delta_{fld}}$	Field mirror parameters, cf. (6.17)
θ_{az}	Turret azimuth angle (East is at $+90°$)
$\mathbf{R_{az}}, \boldsymbol{\delta_{az}}$	Parameters of the turret azimuth mirror
θ_{el}	Turret elevation angle
$R_{\mathrm{el}} = R_{\mathrm{az}}, \delta_{\mathrm{el}} = \delta_{\mathrm{az}}$	Parameters of the turret elevation mirror. The polarization effects of the azimuth and elevation mirrors are assumed to be identical.
$\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}$	Main lens parameters, cf. (6.16)
$\boldsymbol{\theta_{mod}}$	Angle of the modulator +Q direction, measured ccw from terrestrial N-S

Tab. 6.4: Parameters of the SST telescope matrix. The parameters written in boldface are used as free parameters in the model fit.

main lens as defined before. The model parameters, written out explicitly in the above equation, are recapitulated in Tab. 6.4.

6.2.2 Measurements and model fitting

The SST model fit is applied on two types of data sets. The first data set is formed of measurements with a polarizer mounted in front of the main lens. To cover varying turret orientations we have recorded two series separated by a few hours. Each series is composed of Stokes images taken at different angles of the polarizer. The first series was has been recorded with the original telecentric setup, using the prefilter for 422.7 nm. For the second series the optical setup was changed to image the pupil on the camera. This has two advantages: it permits to check the orientation of the polarizer and it provides the spatial distribution of the main lens parameters[5] (cf. Fig. 6.4). At wavelengths below 500 nm the polarizer cannot be considered ideal anymore and the partial cancellation has to be taken into account (cf. Fig. 6.5).

The second data set is given by the science Stokes images themselves. The mean values of these images practically correspond to unpolarized light in

---

[5] In fact, the polarized image of the main lens is mixed with the image of the Schupmann corrector which is close to a pupil too.

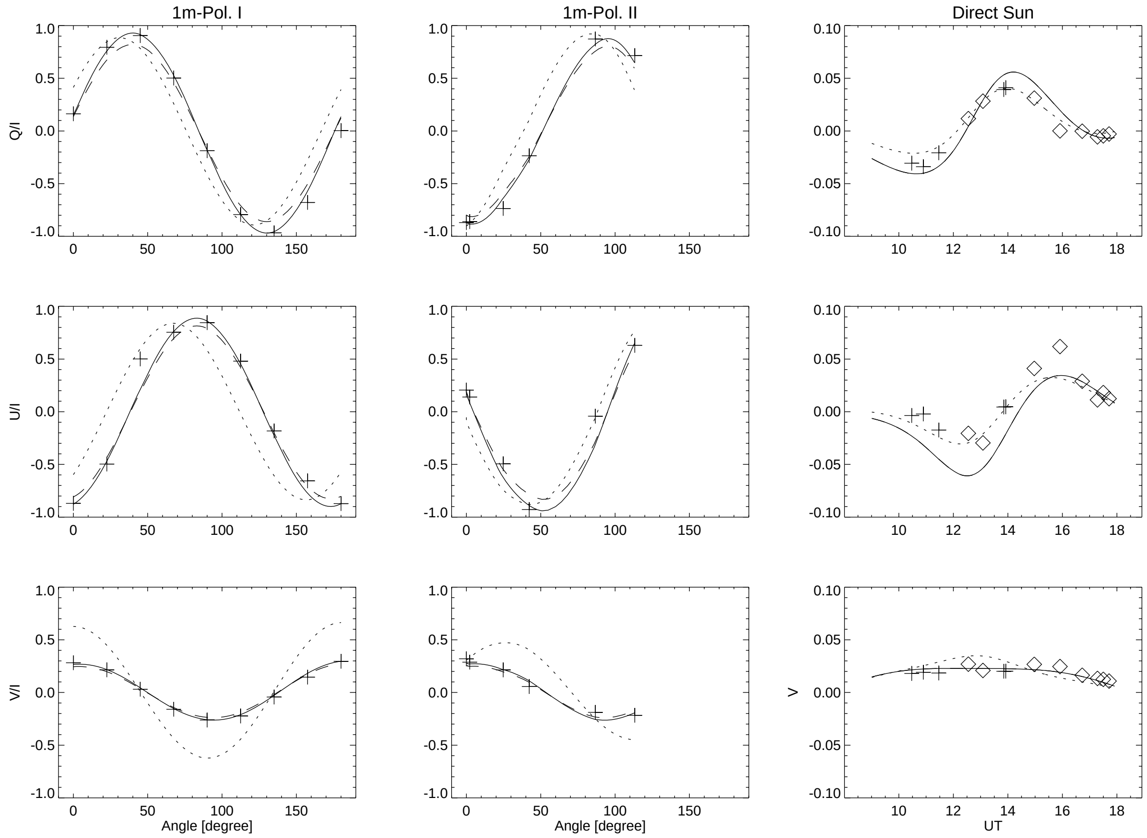


Fig. 6.3: SST polarization. Calibration measurements and model fit for 422.7 nm. The first two columns represent the two measurement series with the polarizer in front of the main lens. The third column shows the data set obtained from the science images, spread over two subsequent days (crosses and diamonds). The solid curves represent the model fit of all 10 free parameters to all data combined. The dashed line shows the effect of setting the main lens parameters A and D to 1. The mirror parameter fit applied solely on the second data set is represented by the dotted curves.

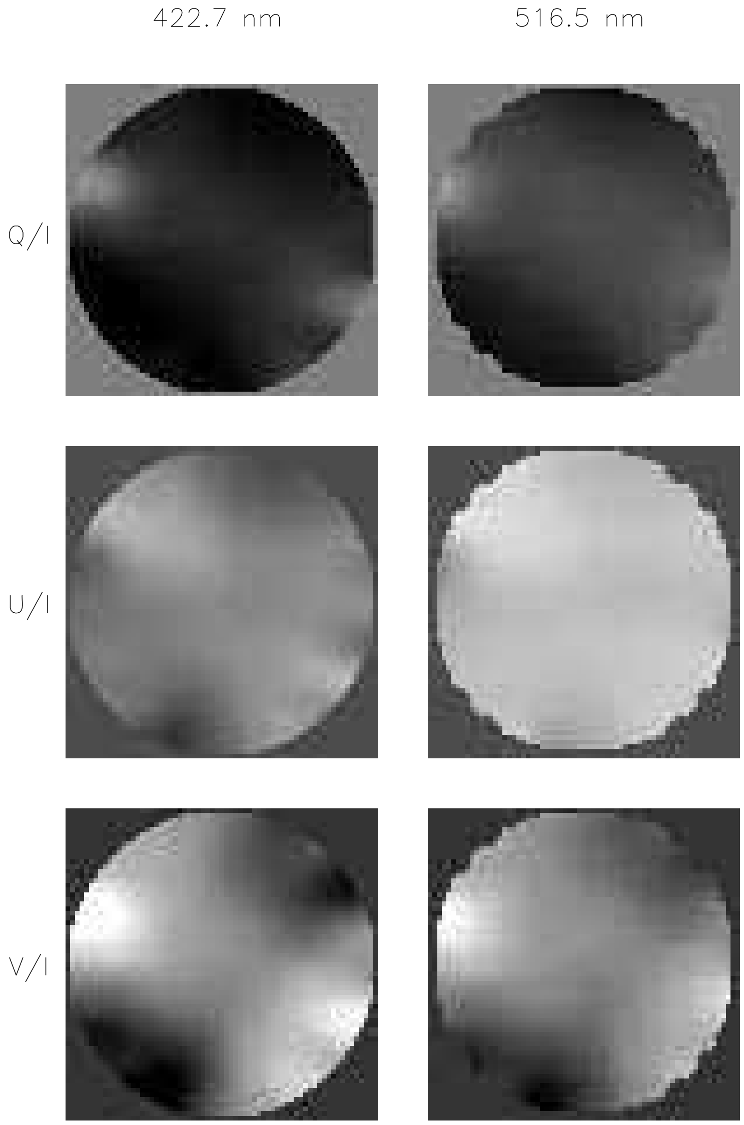


Fig. 6.4: Stokes images of the SST pupil at two different wavelengths, with the polarizer mounted in front of the main lens. The grayscale range is ±0.5. Note that the amplitude of the polarization effects is wavelength dependent, but not the spatial distribution. This is consistent with the picture that the lens locally behaves like a retarder, whose retardation scales with the wavelength.

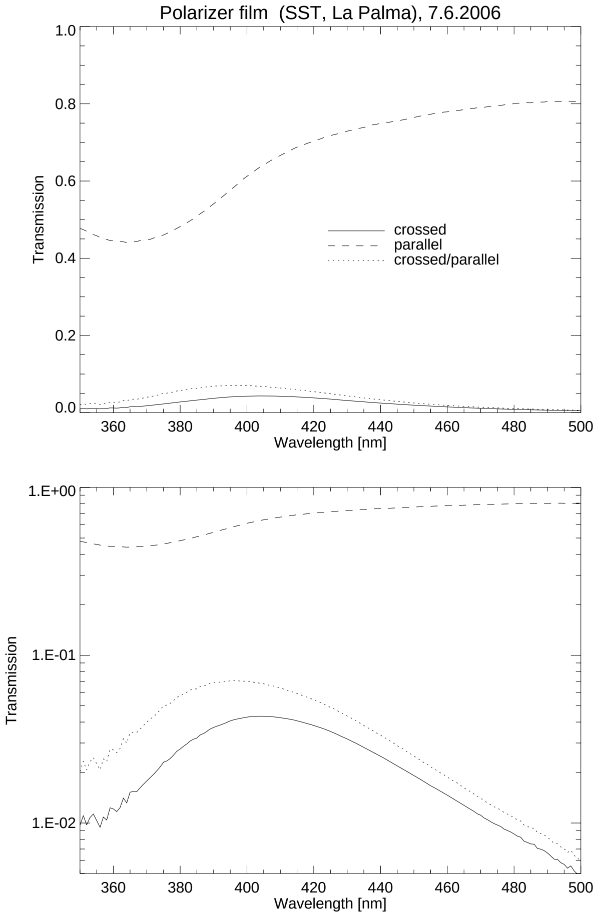


Fig. 6.5: Properties of the 1m polarizer used in front of the SST main lens for telescope calibration (measurements: D. Gisler).

front of the telescope[6] which means that they contain information about the first column of the telescope matrix. This data set is however only sensitive to the mirror properties.

We think that the parameters of the main lens are stable over many years[7]. Once a good set of polarized pupil images is available at a couple of wavelengths, it should be possible to translate the main lens parameters to any wavelength by scaling the local retardation. The mirror parameters are assumed to be prone to a more rapid change and should be recalibrated at each observing run. This calibration is coming quasi for free along with the science data, as described before.

The 10 free parameters (cf. Tab. 6.4) are determined by a nonlinear fit to the two data sets described above, using the Levenberg-Marquardt algorithm (Marquardt, 1963). The points of the second data set are weighted with a factor of 10 in χ^2 to yield an equilibrated impact of both data sets on the fit. The Levenberg-Marquardt algorithm only finds a local χ^2 minimum in the neighborhood of the start parameters. For this reason the data were fitted 500 times, each time with another set of starting parameters, randomly selected within the region of the parameter space that is constrained by the physical limits. Finally we run two Monte-Carlo simulations, as described in Chapter 15.6 of Press et al. (2002), to estimate the confidence limits of the individual parameters.

6.2.3 Results

The statistical results of the fits with different start parameters are shown in Tab. 6.5. The solutions are tightly accumulating around a well-defined global χ^2 minimum, except for the lens parameters C and E, which are not well determined. The lens parameters A and D are settling at unphysical values. Fig. 6.3, shows that resetting the values of these two parameters to the physical limit of 1, only affects the q and u amplitudes. It seems that the partial cancellation of the polarizer, measured separately in the lab (cf. Fig. 6.5) is not perfectly consistent with these data obtained on the telescope.

The results of the two Monte-Carlo simulations are shown in Tab. 6.6. One simulation adds normally distributed errors to the polarizer angle, corresponding to a standard deviation of 2.5° for the first series and 0.25° for the second series. The angle errors are smaller for the second series because the orientation of the polarizer can be verified on the pupil images. The mean deviation of the true polarizer angles read from the pupil images and the set angle values is used as error estimate for the first series. As can be seen from the confidence intervals, the lens crosstalk elements B, C and E are very sensitive to errors in the polarizer angle. The second simulation is adding Poisson-distributed errors on the polarization values q, u and v, corresponding to a standard deviation of 10^{-3}.

---

[6] The mean values of the Stokes images are computed over structureless regions. They correspond to truly unpolarized light in the sense that the residual solar polarization is much smaller than the telescope effects.

[7] The stability has yet to be confirmed with another series of polarization measurements.

The lens parameters are not sensitive to the second data set, as the light passing the main lens is practically unpolarized. When fitting the remaining 5 mirror parameters on the second data set exclusively, the fit is improved for this data set but worsened for the first data set (cf. Fig. 6.3). The q and u plots suggest a systematic offset error in the polarizer angles of about 10°. Moreover, the crosstalk between linear and circular polarization is overestimated.

The overall fit of both data sets and all free parameters, represented by the solid curves in Fig. 6.3 is used for the calibration of the science data obtained during our 2006 campaign. For this fit the average residual in polarization degree is of order 3%. This means that relative variations of polarization signals in the images are reliable down to a level of 10^{-4} even when strong Zeeman structures are present. The SST model on its own is not accurate enough to provide reliable absolute polarization values on the 10^{-3} level and below where the scattering polarization signals, we are mainly interested in, are settled. For some data sets it is however possible to further enhance the accuracy by combining the SST model calibration with a-priori assumptions about the Stokes signals, as described in the introduction of this chapter. The possible higher order corrections that this technique can provide are however limited by the fact that residual errors that generate or destroy polarization are eluding. On the other hand the SST model itself can be enhanced above all by improving the precision in the absolute orientation of the polarizer in front of the main lens. We therefore suggest to repeat the measurements with the polarizer during the next campaign. This will also allow to retroactively correct the data of the 2006 campaign as the lens parameters are assumed to be stable over a long period of time.

	Average	Std. dev.	Min.	Max.
Main lens				
A	1.09e+00	3.61e-03	1.07e+00	1.09e+00
B	-1.73e-02	1.04e-03	-1.89e-02	-1.46e-02
C	-1.10e-02	3.97e-02	-6.63e-02	1.04e-01
D	1.13e+00	3.15e-03	1.12e+00	1.13e+00
E	3.56e-02	7.02e-03	1.21e-02	4.51e-02
Turret				
R_{tur}	9.11e-01	1.10e-03	9.10e-01	9.15e-01
δ_{tur} [°]	-1.49e+02	4.00e+00	-1.60e+02	-1.43e+02
Field mirror				
R_{fld}	9.99e-01	4.50e-03	9.91e-01	1.01e+00
δ_{fld} [°]	1.05e+00	4.97e-02	9.08e-01	1.25e+00
Modulator				
θ_{mod} [°]	-5.16e+01	5.67e-01	-5.24e+01	-5.02e+01
χ^2	7.94e-04	2.13e-06	7.85e-04	7.97e-04

Tab. 6.5: SST free model parameters. Statistical results of 500 fits with different sets of start parameters, randomly distributed the physically allowed region of the parameter space.

	Errors in polar. angle 95% conf.		Errors in pol. meas. 95% conf.	
Main lens				
A	1.09e+00	1.15e+00	1.08e+00	1.09e+00
B	-4.51e-02	1.08e-02	-1.75e-02	-1.71e-02
C	-5.57e-02	4.02e-02	-2.65e-02	5.80e-03
D	1.12e+00	1.20e+00	1.13e+00	1.13e+00
E	1.84e-02	5.07e-02	3.30e-02	3.78e-02
Turret				
R_{tur}	9.03e-01	9.24e-01	9.09e-01	9.13e-01
δ_{tur} [°]	-1.53e+02	-1.44e+02	-1.50e+02	-1.47e+02
Field mirror				
R_{fld}	9.80e-01	1.02e+00	9.94e-01	1.00e+00
δ_{fld} [°]	6.49e-02	1.91e+00	9.98e-01	1.10e+00
Modulator				
θ_{mod} [°]	-5.30e+01	-5.01e+01	-5.18e+01	-5.13e+01

Tab. 6.6: SST free model parameters. Results of two Monte-Carlo simulations. Left part: effect of random errors in the orientation of the polarizer in front of the main lens. Right part: effect of Poisson noise in the polarization signals.

Chapter 7

First observations

In this chapter we show first observations with the IP. Figs. 7.2 and 7.3 have been obtained with the collimated setup at IRSOL in the spectral lines of Hα (6563 Å) and Ca I (4227 Å) respectively. Fig. 7.4 and the corresponding flatfield scan shown in Fig. 7.1 have been recorded with the telecentric setup at the SST during our campaign in 2006. The images are not of great scientific interest and rather have the character of test images but they clearly show signatures of Zeeman effect and scattering polarization (Ca I) on the level of 10^{-3} in polarization degree.

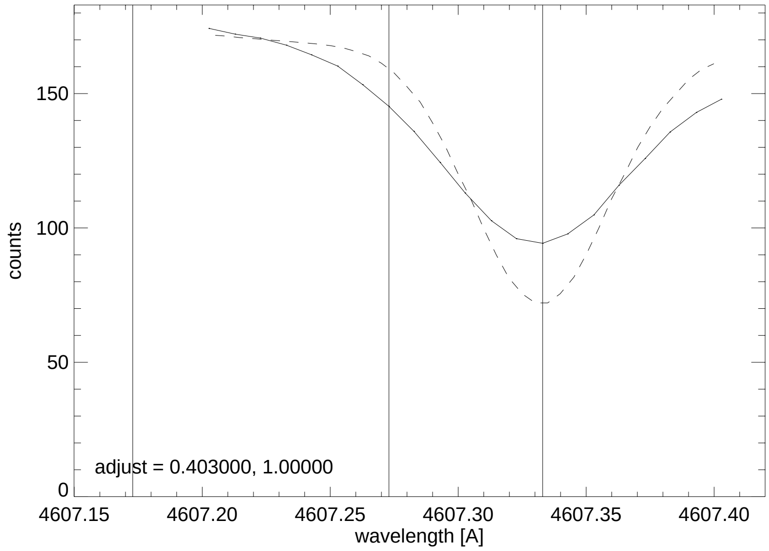


Fig. 7.1: Averaged flatfield scan (solid line) through the Sr I line and comparison with the FTS atlas (dashed line). The vertical lines denote the spectral positions of the observations shown in Fig. 7.4.

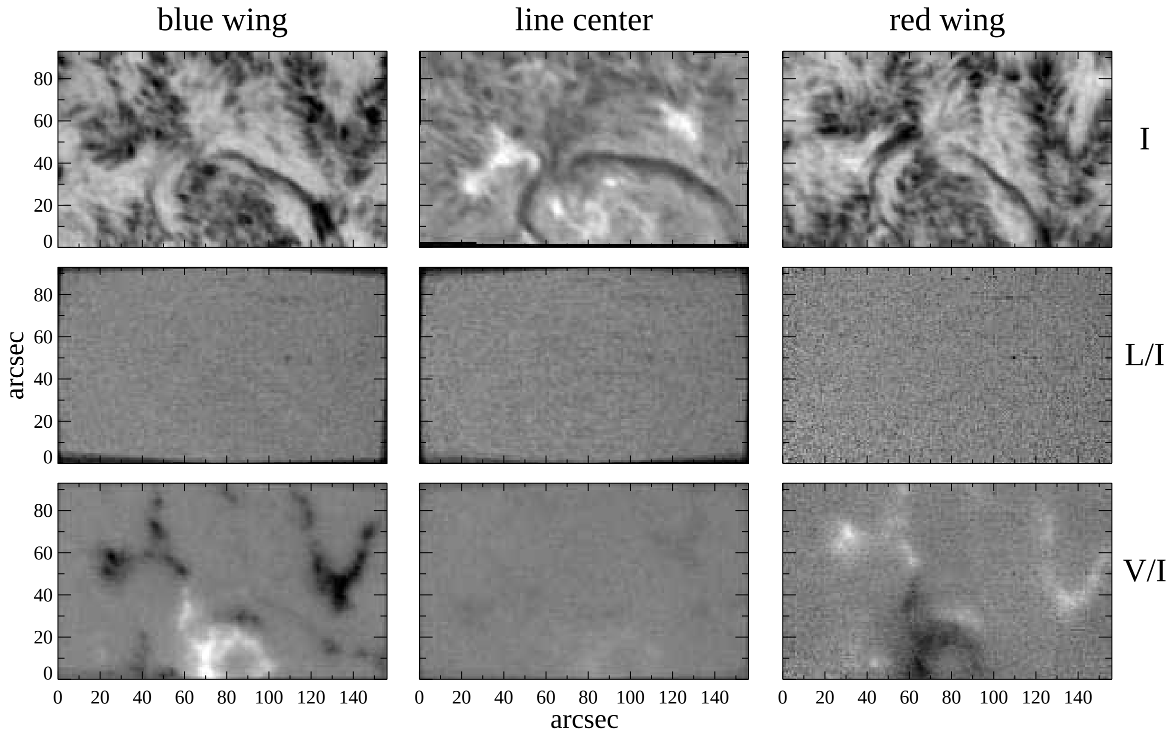


Fig. 7.2: "First light" observation with the TNF in Hα. Within the passband of the Daystar prefilter with 0.5 Å FWHM, the TNF was tuned to the blue wing, red wing, and line center, respectively. For each spectral position a series of Stokes image sets have been recorded. The total integration time per series is about 50s. Since the linear polarization is weak, Q/I and U/I have been combined to an amplitude of linear polarization $L/I = [(Q/I)^2 + (U/I)^2]^{1/2}$. The amplitude of the grayscale corresponds to 40% of the mean value for I, and to 1.5% for L/I and V/I. Prior to averaging, the individual images have been numerically corrected for rotation and shifts, which produces artefacts at some of the image edges. Note the sign change of the Stokes V Zeeman features between blue and red wing.

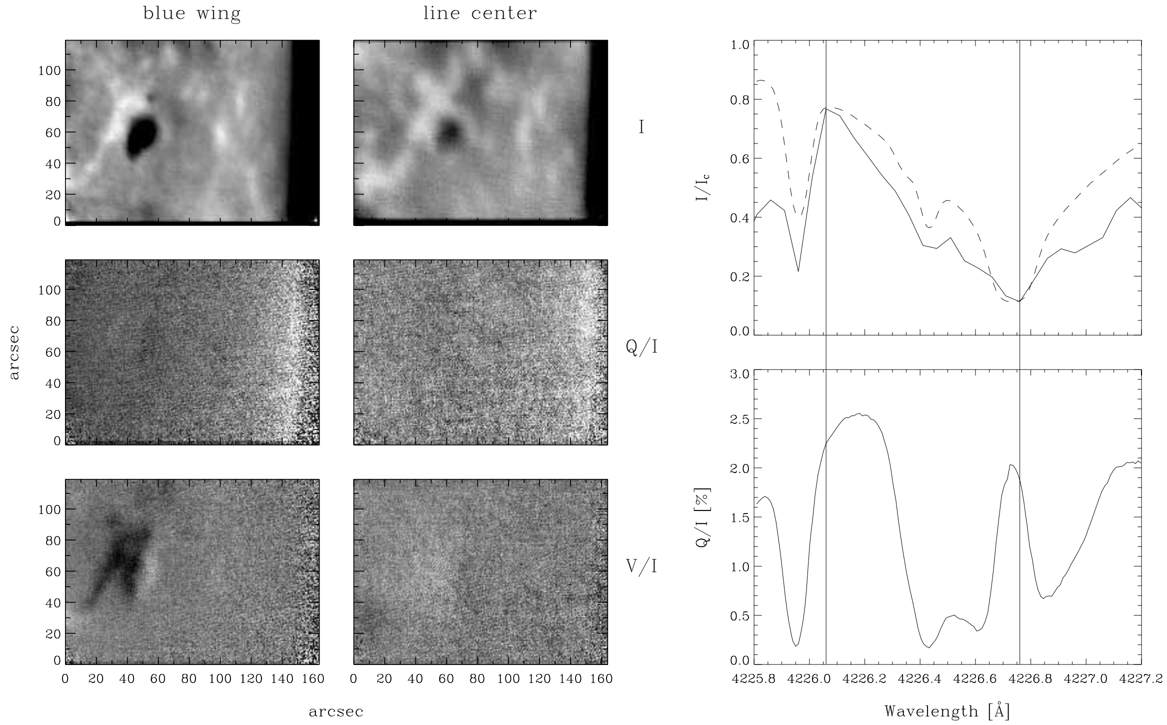


Fig. 7.3: Scattering polarization and Zeeman effect in Ca I observed with the collimated setup at IRSOL next to the solar limb. **Upper right:** flatfield scan with the TNF, averaged across the image (solid line). As comparison, the corresponding spectrum from Gandorfer (2002) is overplotted (dashed line). **Lower right:** Q/I from Gandorfer (2002) for a limb distance of $\mu = 0.1$. **Left:** Series of Stokes images recorded with ZIMPOL at the two spectral positions indicated by the vertical lines in the scan. The amplitude of the grayscale for Stokes Q/I and V/I is 2%. To enhance the contrast, the limb darkening has been removed in the Stokes I images.

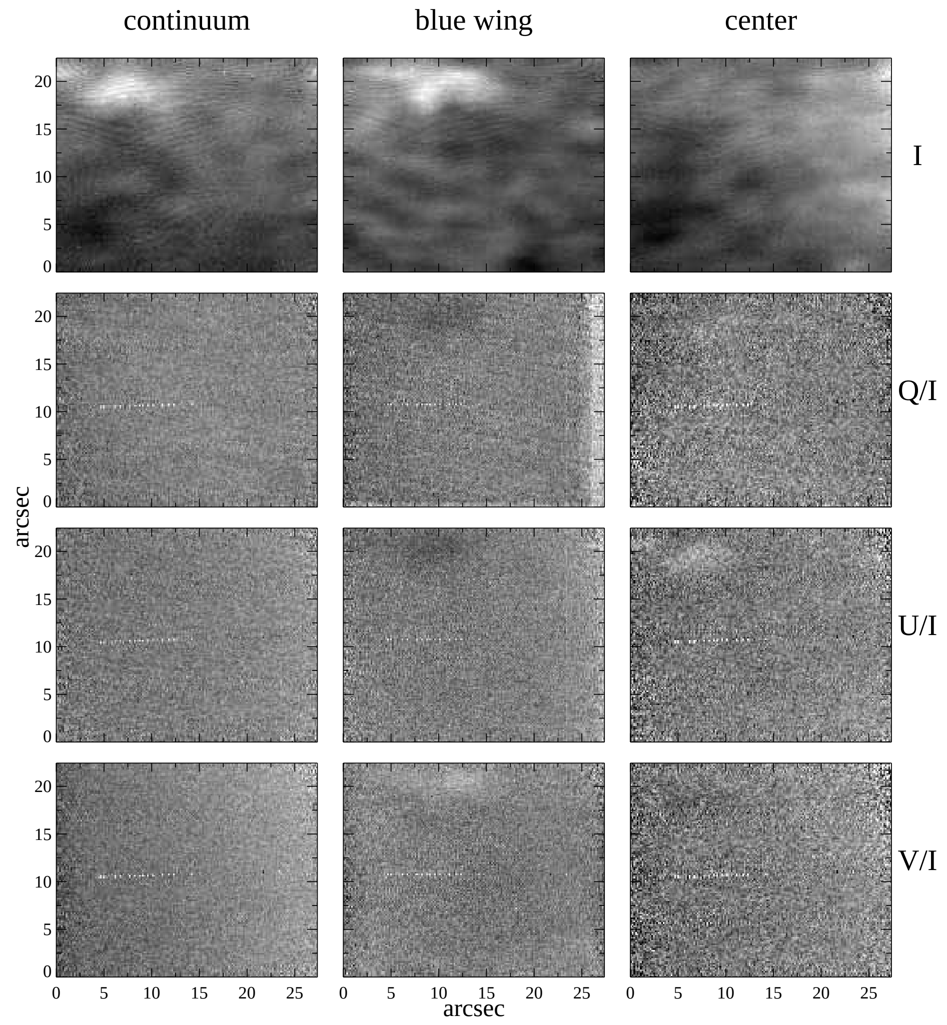


Fig. 7.4: Region with plage. Observation with the Swedish Solar Telescope during our campaign in 2006. Unfortunately the seeing conditions have been very bad most of the time. This image has been obtained without the full adaptive optics support, only the tip-tilt mirror (correction of image shifts) was working. Nevertheless, the faint Zeeman signature of a small plage can be discerned in the Stokes images. The signal is comparable to the expected amplitude of spatial variations in scattering polarization next to the solar limb. The three columns represent different spectral positions around Sr I (c.f. Fig. 7.1). For each position some 15 1s-exposures, spread over 10 minutes, are averaged. The amplitude of the grayscale in Q/I, U/I and V/I is $\pm 5 \cdot 10^{-3}$

.

Chapter 8

Conclusions and outlook

In this part of my thesis I have presented the development and implementation of a tunable narrow-band filter system (TNF) to be used in combination with the Zurich Imaging Polarimeter (ZIMPOL) for quasi-monochromatic imaging vector polarimetry on the Sun. The complete system, TNF plus ZIMPOL, is referred to as Imaging Polarimeter (IP). In the current first version two lithium niobate Fabry-Pérot (FP) etalons are used in series in a single-pass tandem configuration.

We have designed and implemented two complementary optical setups for the TNF. The telecentric setup aims for work at high magnifications and is mainly used at the Swedish Solar Telescope, one of the world's leading telescopes in terms of spatial resolution. In contrast, the collimated setup has a large field of view and higher throughput. It exploits the specific potential of the telescope at the Istituto Ricerche Solari Locarno (IRSOL) where it is permanently installed and operational since 2006. First Stokes images obtained with both optical setups during the first year of operation, prove that the IP works.

The implementation of the IP has been preceded by an in-depth study of the FP etalons in terms of photometric sensitivity as well as spectral, polarimetric and imaging performance. The corresponding models and simulations have allowed to define initial design specifications and requirements for the optics, the prefilters and the etalon control. During the implementation they have been continuously cross-checked with the real filter behavior and iteratively refined. I will recapitulate here the main results of these studies.

The photometric sensitivity of the IP is directly linked to the polarimetric performance since ZIMPOL is essentially limited by photon noise down to a level of 10^{-5} in the degree of polarization. Pure intensity measurements or Zeeman diagnostics with a polarimetric sensitivity of a few percent can get by with a modest S/N ratio. Other imaging polarimeters can thus be operated with integration times below one second which dramatically facilitates the job of getting spatially well resolved images. In our case a S/N of order 10^3 or better is required to detect the faint polarization signatures of the Second Solar Spectrum. It turns out that integration times of 5 minutes or longer (depending on the particular setup, spectral position and location on the Sun) are needed to meet this requirement, a time span comparable to the lifetime of targeted solar structures like the granulation for instance. The photometric sensitivity is therefore identified as a clear limitation of the IP. Our current ZIMPOL cameras harbor the biggest potential for improvement as some 70% of the photons get lost here at wavelengths below about 450 nm. An improvement by a factor of 3 in photon flux with a more sensitive camera is a realistic estimate. Moreover, plans for new solar telescopes of the 4m class have emerged on the horizon of the not too far future. The North-American Advanced Technology Solar Telescope (ATST) is supposed to become operational in the next decade and European countries are seriously thinking of a joint effort in the form of a European Solar Telescope (EST). In view of such powerful light-gathering tools high-precision imaging polarimeters like our current system have a very promising future.

To analyze the spectral and polarimetric performance of the IP we have defined a set of performance parameters and developed a computer program to evaluate these parameters as a function of spectral position, synchronization error (between the two etalons) and prefilter bandwidth. The average result is an optimum prefilter bandwidth between 1.5 and 2 Å and a maximum synchronization error of 10 mÅ, although these figures can vary substantially, depending on the particular scientific requirements. Nevertheless, the program permits to study each application individually, and to make appropriate trade-offs before buying a new prefilter or working on the TNF tuning accuracy.

The imaging performance has been analyzed separately for the actual optics that come with the etalons and for the etalons themselves. The optics for both the collimated and the telecentric setup are designed to work close to the diffraction limit. The influence of the FP etalons on the imaging performance has been studied by means of a diffraction model based on Fourier optics. It turns out that the MTF of the collimated setup is significantly degraded by the small-scale cavity thickness errors. As a first countermeasure the effect can be attenuated by limiting the aperture. A careful combination of selected sub-apertures can further improve the imaging. The telecentric setup is suffering from aberrations induced by the angle-dependent etalon transmission. These aberrations can be partially compensated by refocusing. Whereas the

MTF degradation is still substantial at our lowest magnification (after refocusing), it becomes negligible at the highest one. We also note that the effect is less pronounced in lithium niobate than in air-spaced etalons because of the higher refractive index of the etalon cavity.

Based on the early performance studies and later test observations specific calibration and tuning techniques for the IP have been developed. They are necessary for the reliable operation of this complex instrument system. To meet the tight requirements on the TNF synchronization we have developed a calibration technique for the etalon tuning parameters that permits to keep the aperture-averaged synchronization error below 5 mÅ. The basic calibration is performed once for each prefilter range, with the help of the high-resolution spectrograph at IRSOL. The fine tuning can be conveniently done in the measurement configuration and with the solar spectrum, on the verge of the actual scientific recordings. The fine tuning procedure is highly automated and takes only a few minutes which allows for occasional verifications and readjustments of the TNF settings between the observations. In addition to the fine tuning the so called flatfield scanning procedure, which is also automated and fast, permits to determine small local shifts of the TNF passband and to estimate the spectral purity of the transmission. This information is very valuable for the later data analysis.

On the level of polarimetry we have elaborated a new general algorithm to determine the complete modulation matrix of basically any type of polarimeter with a minimum of calibration measurements. It comes to grips with polarimeters that do not measure the complete Stokes vector as well as with full Stokes polarimeters or polarimeters that do not measure all Stokes components simultaneously. An example of the latter type is the frequently used piezoelastic modulator (PEM) which alternates between Stokes IQV and IUV modes. It has been used during the 2006 observing campaign at the SST where the new algorithm has been successfully applied for the first time.

Another problem associated with polarimetry is the instrumental polarization of the telescope, especially when this polarization varies during the day as in the SST case. Based on the work of Selbing (2005) we have adapted an improved SST model to our ZIMPOL polarimeter and we have been able to successfully calibrate it for the wavelength 422 nm, one of the wavelengths used during our 2006 campaign. The basic calibration is done once with the help of a 1m polarizer sheet in front of the main lens. The ordinary science data can be used thereafter to continuously refine and improve the model. During the next campaign in 2007 the calibration with the 1m polarizer will be repeated at another wavelength which then hopefully permits to interpolate with sufficient accuracy across the whole working range of the TNF (390-660 nm).

Lithium niobate etalons have recently drawn the attention of the astrophysical world as an attractive alternative to the classical air-spaced etalons. They are already used or foreseen to be used in current solar balloon and space based missions like, for instance, IMaX on board SUNRISE (Martinez Pillet & al., 2004), or the proposed VIM (Visible Imager and Magnetograph) for the Solar Orbiter mission. In this context it is of great interest to the scientific community to further explore the potential of this type of filter system. At IRSOL we will experiment with two new configurations. (1) A hybrid system consisting of the spectrograph and the FP etalons. The basic idea is to install the two etalons in series in the spectrograph beam. The resulting spectrum seen by the camera consists of the discrete and narrow transmission peaks separated by the free spectral range of the filter system (e.g. some 5 Å at 4600 Å). The spectrograph entrance slit is replaced by a rectangular aperture whose size exactly matches the free spectral range and the original filter spectrum then appears as a series of adjacent 2D images. This would allow to obtain monochromatic images of any spectral position without the use of a prefilter. However, at this stage the feasibility of this concept has not been proven yet. (2) A special double-pass configuration described theoretically by Netterfield et al. (1997) where the light is fed twice through one or even both etalons. The optical property that distinguishes our etalons from all other etalons currently in use in solar physics is their birefringence. Between the two passes the plane of polarization of the light can therefore be rotated by 90° to switch between the optical axes of the crystal. The result is equivalent to a series of two etalons with different cavity thicknesses.

The initial single-pass tandem version of the IP has been completed. The behavior of this complex instrument system is sufficiently well understood and all the calibration and tuning procedures, which are necessary for a reliable operation, have been developed. With this version first science goals are readily approached. The focus is on a domain in solar polarimetry where the excellent polarimetric accuracy of ZIMPOL can be exploited, namely the exploration of the spatial structuring of the polarimetric signatures in the Second Solar Spectrum. During observing campaigns at the Swedish Solar Telescope we will try to detect fluctuations in scattering polarization both spatially at sub-arcsec resolution and temporally on a time scale of a few minutes. We have prefilters for several extraordinary sources of scattering polarization in the Second Solar Spectrum that form at different heights in the solar atmosphere. These are the atomic lines of Ca II K (3933 Å), Ba II (4554 Å), Sr I (4607 Å), and the molecular band of C_2 (5141 Å). The goal is to investigate the height dependence of photospheric and chromospheric inhomogeneities on a granule spatial scale in temperature, weak magnetic fields, and velocity fields. At IRSOL we will also investigate spatial variations of scattering polarization but at larger spatial scales of a few arcsec.

In view of the science goals mentioned above it is important to conserve the effective spatial resolution of images with long integration times as far as possible. The first step in this direction will be the application of a so-called destretching technique (e.g. November & Simon, 1988) to the individual short-exposure frames before averaging. The basic idea of this technique is to divide the image into a certain number of sub-apertures and to track the seeing-induced relative movements of these sub-apertures between subsequent frames by means of a cross-correlation. The resulting vector fields of shifts are then applied to correct the frames. Further steps will consist in an experimentation with special image reconstruction techniques where the IP images are corrected with the optical transfer function obtained from simultaneously recorded broadband images (Keller, 1992). This technique has been seriously elaborated and refined during the last years as can be seen from the recent work of e.g. Puschmann et al. (2006); Sánchez-Andrade Nuño et al. (2005); van Noort et al. (2006) and Mikurda et al. (2006).

Bibliography

Atherton, P. D., Reay, N. K., Ring, J., 1981, *Tunable Fabry-Pérot filters*, Opt. Eng. **20**, no. 6, 806

Beckers, J. M., 1998, *On the effect of narrow-band filters on the diffraction limited resolution of astronomical telescopes*, A&A suppl. ser. **129**, 191

Bernasconi, P., 1997, *Stokes Vector Polarimetry: Observation and Analysis of Solar Magnetic Fields*, Ph.D. thesis, ETH Zürich

Bianda, M., Ramelli, R., Feller, A., Stenflo, J. O., Küveler, G., 2007, *Instrumental developments at the Gregory Coudé Telescope (GCT) at IRSOL*, in F. Kneer, K. G. Puschmann, A. D. Wittmann (eds.), *Modern Solar Facilities - Advanced Solar Science*, 59–62, Universitätsverlag Göttingen

Bonaccini, D., Raymond, N. S., 1988, *Lithium niobate double channel Fabry-Pérot interferometer for solar corona uses*, Appl. Opt. **27**, no. 24, 5095

Cavallini, F., 1998, *The Italian Panoramic Monochromator*, A&A suppl. ser. **128**, 589

Cavallini, F., 2002, *IBIS (Interferometric BIdimensional Spectrometer)*, in H. Sawaya-Lacoste (ed.), *ESA SP-477: Solspa 2001, Proceedings of the Second Solar Cycle and Space Weather Euroconference*, 585–587

Cavallini, F., Baffa, C., Reardon, K., Berrilli, F., Cantarano, S., Egidi, A., 2003, *Recent results from IBIS*, Memorie della Societa Astronomica Italiana **74**, 796

Cavallini, F., Berrilli, F., Cantarano, S., Egidi, A., 2000, *IBIS: A Purely Interferometric Instrument for Solar Bidimensional Spectroscopy*, in A. Wilson (ed.), *ESA SP-463: The Solar Cycle and Terrestrial Climate, Solar and Space weather*, 607

Darvann, T., Owner-Petersen, M., 1994, *Prestudy of a FPI Filter for LEST*, Tech. Rep. 57, LEST Foundation

Gandorfer, A., 2002, *The Second Solar Spectrum: A high spectral resolution polarimetric survey of scattering polarization at the solar limb in graphical representation. Volume II: 3910 Å to 4630 Å*, VdF, Zurich

Gandorfer, A. M., Steiner, H. P., Povel, P., Aebersold, F., Egger, U., Feller, A., Gisler, D., Hagenbuch, S., Stenflo, J. O., 2004, *Solar polarimetry in the near UV* *with the Zurich Imaging Polarimeter ZIMPOL II*, A&A **422**, 703

Gaskill, J. D., 1978, *Linear Systems, Fourier Transforms and Optics*, Wiley, New York

Gisler, D., 1999, *Polarisationskompensation an Sonnenteleskopen*, Master's thesis, ETH Zürich

Gisler, D., 2005, *Instrumentierung für hochpräzise Vektorpolarimetrie in der Astronomie*, Ph.D. thesis, ETH Zurich

Goodman, J. W., 1988, *Introduction to Fourier Optics*, Mac Graw-Hill Publ. Co., New York

Holzreuter, R., 2003, *Telescope-independent Removal of Instrumental Polarization in High-precision Vector Polarimetry*, in J. Trujillo-Bueno, J. Sanchez Almeida (eds.), *Astronomical Society of the Pacific Conference Series*, vol. 307 of *Astronomical Society of the Pacific Conference Series*, 71–+

Holzreuter, R., Stenflo, J. O., 2007, *Scattering polarization in strong chromospheric lines. II. Influence of the temperature curve on the Ca II K line*, A&A **467**, 695

Keller, C. U., 1992, *High Resolution Observations of Solar Magnetic Fields*, Ph.D. thesis, ETH Zürich

Keller, C. U., 1996, *Recent Progress in Imaging Polarimetry*, Solar Phys. **164**, 243

Kentischer, T. J., Schmidt, W., Sigwarth, M., Uexkuell, M. V., 1998, *TESOS, a double Fabry-Pérot instrument for solar spectroscopy*, A&A **340**, 569

King, D. L., 1985, *Atmospheric extinction at the Roque de los Muchachos Observatory, La Palma*, La Palma technical note 31, RGO

Kleint, L., 2007, *Spectropolarimetry of the chromosphere with IBIS*, Master's thesis, ETH Zurich

Mack, J. E., McNutt, D. P., Roesler, F. L., Chabbal, R., 1963, *The PEPSIOS purely interferometric high-resolution scanning spectrometer. I: The pilot model*, Appl. Opt. **2**, 873

Marquardt, D. W., 1963, *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, Journal of the Society for Industrial and Applied Mathematics **11**, 431

Martinez Pillet, V., al., 2004, *The Imaging Magnetograph eXperiment for the SUNRISE balloon Antartica project*, Proc. of SPIE **5487**, 1152

Mikurda, K., Tritschler, A., Schmidt, W., 2006, *The influence of image reconstruction on two-dimensional spectrograms of the solar photosphere*, A&A **454**, 359

Neckel, H., 1996, *On the Wavelength Dependency of Solar Limb Darkening (\lambda 303 to 1099 nm)*, Solar Phys. **167**, 9

Netterfield, R. P., Freund, C., Seckold, J. A., Walsh, C. J., 1997, *Design of a* *lithium-niobate Fabry-Pérot etalon-based spectrometer*, Appl. Opt. **36**, no. 19, 4556

November, L. J., Simon, G. W., 1988, *Precise proper-motion measurement of solar granulation*, ApJ **333**, 427

Pérot, A., Fabry, C., 1897, *Sur les franges des lames minces argentées et leur application à la mesure de petites épaisseurs d'air*, Ann. Chim. Phys. **12**, 459

Pérot, A., Fabry, C., 1899, *On the Application of Interference Phenomena to the Solution of Various Problems of Spectroscopy and Metrology*, ApJ **9**, 87

Povel, H., 1995, *Imaging Stokes polarimetry with piezoelastic modulators and charge-coupled-device image sensors*, Opt. Eng. **34**, 1870

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 2002, *Numerical Recipes in* `C++`, Cambridge University Press

Puschmann, K. G., Kneer, F., Seelemann, T., Wittmann, A. D., 2006, *The new Göttingen Fabry-Pérot spectrometer for two-dimensional observations of the Sun*, A&A **451**, 1151

Sanchez Almeida, J., Martinez Pillet, V., Wittmann, A. D., 1991, *The instrumental polarization of a Gregory-coude telescope*, Solar Phys. **134**, 1

Sánchez-Andrade Nuño, B., Puschmann, K. G., Sánchez Cuberes, M., Blanco Rodríguez, J., Kneer, F., 2005, *Chromospheric Dynamics of a Solar Active Region*, in *ESA Special Publication*, vol. 600 of *ESA Special Publication*

Scharmer, G. B., 2006, *Comments on the optimization of high resolution Fabry-Pérot filtergraphs*, A&A **447**, 1111

Schlarb, U., Betzler, K., 1993, *Refractive indices of lithium niobate as a function of temperature, wavelength, and composition: A generalized fit*, Phys. Rev. B **48**, 15613

Selbing, J., 2005, *SST polarization model and polarimeter calibration*, Master's thesis, Stockholm University

Stenflo, J. O., 2004, *The new world of scattering physics seen by high-precision imaging polarimetry*, Rev. Modern Astron. **17**, 269

Stenflo, J. O., Gandorfer, A., Holzreuter, R., Gisler, D., Keller, C. U., Bianda, M., 2002, *Spatial mapping of the Hanle and Zeeman effects on the Sun*, A&A **389**, 314

Thalmann, C., 2008, Ph.D. thesis, ETH Zürich, in preparation

Tritschler, A., Schmidt, W., Langhans, K., Kentischer, T., 2002, *High-resolution solar spectroscopy with TESOS*, Solar Phys. **211**, 17

van Noort, M., Rouppe van der Voort, L., Löfdahl, M., 2006, *Solar Image Restoration by use of Multi-Object Multi-Frame Blind Deconvolution*, in J. Leibacher, R. F. Stein, H. Uitenbroek (eds.), *Solar MHD Theory and Observations: A High*

Spatial Resolution Perspective, vol. 354 of *Astronomical Society of the Pacific Conference Series*, 55–+

Volkmer, R., Kneer, F., Bendlin, C., 1995, *Short-period waves in small-scale magnetic flux tubes on the Sun.*, A&A **304**, L1+

von der Lühe, O., Kentischer, T. J., 2000, *High spatial resolution performance of a triple Fabry-Pérot filtergraph*, A&A **146**, 499

Part II

Polarized scattering at sodium vapor

Chapter 9

Introduction

With the advent of high-precision imaging polarimetry through the introduction of the ZIMPOL (**Z**urich **Im**aging **Pol**arimeter) technology (Povel, 1995, 2001; Gandorfer et al., 2004), the rich world of coherence phenomena on the Sun that had largely been hidden from view, due to lack of instrument sensitivity, became fully accessible to observations (Stenflo & Keller, 1996, 1997). The linearly polarized solar spectrum that is formed by coherent scattering processes is as richly structured as the ordinary intensity spectrum but looks very different and unfamiliar. Ivanov (1991) introduced the apt term "second solar spectrum" for this linearly polarized spectrum, to express that we are dealing with a new and very different spectrum that is governed by different physical processes and contains different information. An atlas of the second solar spectrum over the entire visible range and the ultraviolet down to the atmospheric cut-off has been produced, published as books in three volumes (Gandorfer, 2000, 2002, 2004).

The systematic exploration of the second solar spectrum has brought to light a large number of surprising and apparently anomalous polarized spectral structures (cf. reviews of Stenflo, 2003b, 2004). It has been possible to find explanations for most of them in terms of effects that were previously rather unfamiliar to the world of astrophysics, like quantum interference between states of different total angular momentum, optical pumping of atomic ground states, or coherent scattering by molecules. Much theoretical work in quantum physics has been invested to resolve these various "mysteries", examples of which are the contributions of Landi degl'Innocenti (1996), Bommier (1997a,b), Stenflo (1997) and Trujillo Bueno (2001, 2003). An overview of the subject can be found in the proceedings of the third Workshop on "Solar Polarization" (Trujillo-Bueno & Sanchez Almeida, 2003).

However, one of these "mysteries" has stubbornly resisted all explanation attempts and remained highly enigmatic, since it was first presented many years ago (Stenflo & Keller, 1996). It concerns the observed polarization peak in the core of the well-known line Na I D_1 5896 Å, which according to standard quantum mechanics should be intrinsically unpolarizable as a consequence of its quantum-number structure. This enigma was deepened by the discovery of a polarization peak in the Ba II 4934 Å line, which has the identical quantum-number structure as the Na I D_1 line and therefore also would be expected to be intrinsically unpolarizable (Stenflo et al., 1998, 2000b). These peaks are most prominent in solar regions that appear devoid of magnetic fields, indicating that it is an effect of non-magnetic scattering.

An initially promising-looking explanation was advanced by Degl'Innocenti (1998, 1999) in terms of optical pumping of the hyperfine-structure components of the ground state of the sodium D_1 transition. The idea is that the atomic polarization that is induced in the excited state by the anisotropic illumination gets partly transferred to the ground state in the spontaneous emission process. After many such scattering processes a statistical equilibrium is established, in which the ground state is partly polarized. Scattering from a polarized initial atomic state would result in polarization of the emitted radiation. Later quantitative analysis of this process however showed that it predicted polarization signatures that were 1–2 orders of magnitude too small and had the wrong symmetry as compared with the observed polarization (Trujillo Bueno et al., 2002; Kerkeni & Bommier, 2002; Klement & Stenflo, 2003). The D_1 enigma therefore remains unexplained.

The fundamental question concerning the D_1 mystery remained: Is it a problem of solar physics or of atomic physics/quantum mechanics ? The solar atmosphere is a complex environment in many aspects: It is an optically thick medium where radiative-transfer physics with effects like partial frequency distribution plays an important role. It is a magnetized medium permeated by intermittent and tangled magnetic fields on practically all scales with an apparently fractal structure (cf. Stenflo, 2004). It is a medium where atomic collisions play a significant role. Although none of these effects should be able to give rise to a core polarization of the D_1 line if it were intrinsically unpolarizable, the solar evidence is not convincing enough to be considered as proof that the D_1 problem is one of atomic physics/quantum mechanics. For such a proof, the Sun appears to be too messy an object.

To resolve this fundamental question we have therefore carried out a laboratory experiment to study 90° scattering at sodium atoms under optically thin conditions and without optical pumping effects (unpolarized initial state).

Chapter 10

Solar evidence for a fundamental problem

The exploration of the second solar spectrum has brought many unexpected surprises, which generally have found an explanation before too long. Examples are quantum interferences between atomic states of different total angular momentum, reported and explained in Stenflo (1980) for the Ca II K and H ultraviolet lines, partial frequency redistribution (Rees & Saliba, 1982; Saliba, 1985) as a radiative-transfer effect to explain the triplet peak structure of the polarized Ca I 4227 Å profile observed by Stenflo (1974), hyperfine structure and isotope effects (Stenflo, 1997) to explain the triplet peak of the Ba II 4554 Å line observed by Stenflo & Keller (1996, 1997), and optical pumping of atomic ground states (Trujillo Bueno, 2001; Manso Sainz & Trujillo Bueno, 2001, 2003) to explain the anomalous polarization ratios in multiplets like the Mg I b lines (5167, 5173, and 5184 Å) and the Ca II infrared triplet (8498, 8542, and 8662 Å) observed and reported in Stenflo et al. (2000b). In contrast, the Na I D_1 5896 and Ba II D_1 4934 Å lines have stubbornly defied all explanation attempts since the problem was first pointed out by Stenflo & Keller (1996).

10.1 The Na I D_1 evidence

Figure 10.1 shows the shape of the remarkable polarization signature of the sodium D_2–D_1 system in non-magnetic regions on the Sun. The recording was made near the heliographic south pole (to avoid the much stronger magnetic fields at lower latitudes) with the slit parallel to and 5 arcsec inside the solar limb. The top panel gives the ordinary intensity spectrum (Stokes I, normalized to the intensity of the adjacent continuum I_c), the bottom panel the fractional linear polarization (Stokes Q/I), defined such that positive Q/I

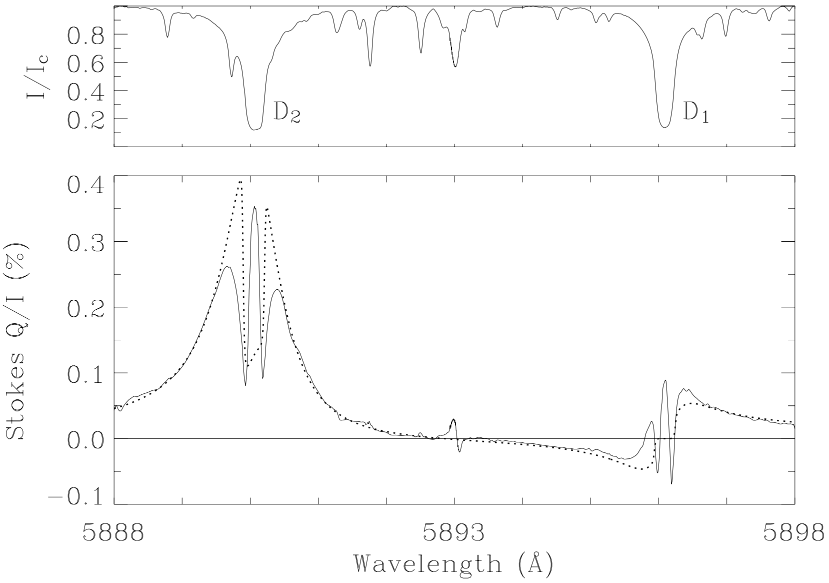


Fig. 10.1: Intensity (upper panel) and fractional linear polarization (lower panel, solid line) across the Na I D_2 5889.97 and D_1 5895.94 Å lines recorded with ZIMPOL at the National Solar Observatory (Kitt Peak) at 5 arcsec inside the solar limb near the heliographic south pole (Stenflo et al., 2000a). The dotted curve in the lower panel represents a fit based on a simplified model as described in section 11.4.

represents polarization with the electric vector parallel to the limb, negative Q/I polarization with the perpendicular orientation. The solid curve represents the observations, while the dashed curve refers to a simplified model fit that will be described in section 11.4 below.

The polarizability of a scattering transition depends on the total angular momentum quantum numbers that are involved. Ignoring the nuclear spin of sodium for a moment, the ground state has electronic angular momentum $J = 1/2$, common for both the D_2 and D_1 transitions, while the excited state has $J = 3/2$ for D_2, $1/2$ for D_1. As we will see in section 11.1 below, a $J = 1/2 \rightarrow 1/2 \rightarrow 1/2$ scattering transition should be intrinsically unpolarizable, while a $J = 1/2 \rightarrow 3/2 \rightarrow 1/2$ transition is partially polarizable. We thus expect the D_2 line to show polarization, while the D_1 line should have zero polarization. At intermediate frequencies away from the resonances of either line we have a coherent combination of both scattering transitions, which leads to quantum interference between the $J = 3/2$ and $1/2$ excited states. In other words, the intermediate state is a mixed quantum state, a “Schrödinger cat”. The quantum interference term causes the scattering polarizability to change sign between the D_2 and D_1 resonant wavelengths (Stenflo, 1980). Without this interference, the polarization would remain positive (electric vector parallel to the limb) throughout the whole wavelength range.

The scattering polarizability is however modified by the nuclear spin, which is $3/2$ for sodium and combines with the electronic angular momentum J to give rise to new angular momentum states F. Thus the $J = 1/2$ state is split into the two hyperfine states with $F = 1$ and 2, while the $J = 3/2$ state is split into the four states with $F = 0$, 1, 2, and 3. Scattering transitions between these various hyperfine states have different intrinsic polarizabilities, which when combined lead to an effective polarizability that is reduced in the D_2 line core with respect to the polarizability without nuclear spin (cf. section 11.1 and Fig. 11.3). The polarizability in the D_1 core however remains zero, due to perfect cancellation between the resonant and fluorescent contributions, as will be explained in more detail in section 11.1 below. The quantum mechanical prediction is therefore zero polarization in the D_1 line core, in contradiction with the observations, which show a very significant polarization peak symmetrically positioned at the line center wavelength.

The dashed model curve in Fig. 10.1, which will be explained in more detail in section 11.4 below, has an abrupt dip in the Doppler core of the D_2 line, while the observed polarization profile instead shows a narrow peak there, surrounded by deep minima. This particular shape, with the conversion of the trough-like intrinsic polarizability shape into a triplet-type polarization profile has recently been explained as a result of partial frequency redistribution (PRD) in the polarized radiative-transfer problem in the optically thick solar atmosphere (Fluri et al., 2003). In contrast, the simplified model curve in Fig. 10.1 ignores radiative transfer and assumes frequency coherence, which is not valid in the line core and its immediate surroundings. It is therefore possible to say that the triplet structure of the D_2 polarization profile is sufficiently understood. However, the same radiative transfer with PRD completely fails to produce any polarization at the center of the D_1 line, no matter how we choose the free or uncertain parameters in the radiative-transfer problem (Fluri et al., 2003). The reason is simply that because the intrinsic polarizability of D_1 is zero, there is nothing there to redistribute through radiative transfer.

The mechanism of optical pumping to produce atomic polarization in the $F = 1$ and 2 ground states (Degl'Innocenti, 1998, 1999) seemed promising at first, but self-consistent treatments within a radiative-transfer context (Trujillo Bueno et al., 2002; Kerkeni & Bommier, 2002; Klement & Stenflo, 2003; Fluri et al., 2003) show that it fails both quantitatively (by giving polarizations 1–2 orders of magnitude too small) and qualitatively (by predicting anti-symmetric core profiles, while the observed peak is symmetric). The D_1 polarization thus remains enigmatic.

The D_2–D_1 wavelength range has been observed repeatedly over a number of years and in many different types of solar regions, both magnetic and non-magnetic. In magnetic regions, where the core peak in the D_2 line varies strongly along the spectrograph slit due to the Hanle depolarization effect from spatially structured solar magnetic fields, the D_1 peak is suppressed or invisible (Stenflo et al., 2001). It is only consistently seen in "non-magnetic" regions. With non-magnetic we mean solar regions where we see no trace of Zeeman-effect signatures (circular or linear) in any of the spectral lines (in particular in the narrow surrounding lines, which produce more Zeeman-effect polarization than the broader sodium lines), and where we also see no significant spatial variations of the D_2 core peak due to the Hanle effect. Such regions are found at high heliographic latitudes. A careful empirical determination of the center-to-limb variation of the D_2–D_1 polarization profile in non-magnetic regions has been presented in (Stenflo et al., 2000a). The profile in Fig. 10.1 is a typical example. While the polarization amplitudes decrease rapidly as we move away from the limb toward disk center, the complex shape of the polarization profile remains almost unchanged. In particular, the ratio between the D_1 and D_2 line core polarization amplitudes stays almost invariant with limb distance. The general center-to-limb variation of the polarization amplitude is common to all non-magnetic scattering polarization on the Sun: at disk center the non-magnetic polarization vanishes for symmetry reasons (zero net scattering angle), while the net scattering angle (determined by the net anisotropy of the radiation field that is due to the limb darkening of the solar disk) approaches 90° as we approach the limb.

10.2 Comparison with related multiplets

The Na I D_1 anomaly would be less convincing if it were the only such case in the whole solar spectrum. One might for instance worry that there could be some unidentified polarizing blend line that by chance happens to lie at exactly the same wavelength as the D_1 line. Is the problem unique to sodium, with no relevance to any other element? Although we are convinced that all spurious instrumental effects, like the influence of stray light in the core of deep absorption lines, have been fully accounted for and cannot produce false anomalies like this, there may always remain a small suspicion, at least

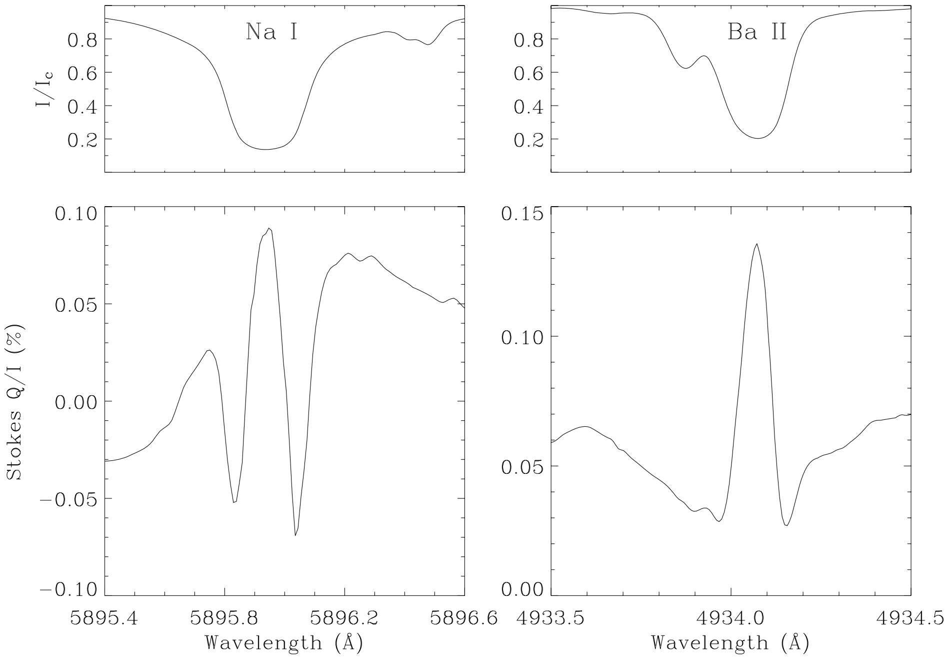


Fig. 10.2: Comparison of the polarization profiles of the Na I D$_1$ and Ba II D$_1$ lines. The left panel represents a blown-up portion of Fig. 10.1. The observations on which the right panel is based have been reported in Stenflo et al. (2000b).

for those not familiar with the measurement procedures, that we might have overlooked something. It is therefore particularly reassuring that there exists another clean case exhibiting the same kind of anomaly, namely the Ba II 4934 Å line.

The Na I D$_2$ and D$_1$ lines make up multiplet no. 1 of sodium. Similarly, multiplet no. 1 of singly ionized barium contains two lines with the identical quantum number structure as the two sodium lines, including nuclear spin and hyperfine structure. Due to this quantum equivalence it is convenient to use the labels D$_2$ and D$_1$ for them. The D$_2$ line is the strongly polarizing Ba II 4554 Å line, whose scattering polarization and profile asymmetry was studied already decades ago (Stenflo et al., 1980). The D$_1$ line is Ba II 4934 Å, which, like Na I D$_1$, exhibits a very significant narrow polarization peak centered on the line.

Figure 10.2 shows a comparison of the polarization profiles of the Na I and Ba II lines, recorded at the same limb distance (5 arcsec inside the limb). Since the ratio of their wavelengths is 1.2, we have chosen a spectral window that is 1.2 times wider for the sodium plot as compared with the barium plot. The barium curve represents an average of five recordings made on different days and years, but, as shown in Stenflo et al. (2000b), all these five recordings show the same profile shape, which demonstrates the extremely high reproducibility of the polarization profile, not only from day to day but also from year to year. In comparing the Na and Ba polarization profiles, it is the remarkable agreement in the shape, width, and centering of the polarization peak that is relevant here. We should not be distracted by the non-relevant differences: The surroundings of the central peak are highly asymmetric in the Na case, due to the quantum interference with the D_2 line. In the Ba case that quantum interference does not play a role since the Ba D_1 and D_2 lines are separated by 380 Å and are thus much further away from each other than the corresponding Na lines (separation of 6 Å). Instead, there is a slight asymmetry around the Ba line which seems to be related to the rather strong blend line (of Fe I) in the far blue wing of the Ba line. Further, since the continuum polarization becomes more pronounced with decreasing wavelength, the region around the Ba line is elevated in polarization, while this is much less the case for the Na line. Accounting for these expected differences, we see how the sodium D_1 anomaly is also appearing in the same way as a barium D_1 anomaly.

Several other chemical elements have multiplets with the same J quantum numbers as the Na D_2 and D_1 lines, both with and without nuclear spin. The other cases with nuclear spin $3/2$ and observed polarization profiles are the Li I D_2 and D_1 lines at 6707.74 and 6707.89 Å (reported in Stenflo et al., 2000b), and the Be II D_2 and D_1 lines at 3130.414 and 3131.058 Å (reported in Stenflo, 2003a). These two cases however do not give "clean" answers for the following reasons: In the case of the lithium lines, their fine structure separation is only 0.15 Å, of the same order as the width of the D_2 polarization profile. This means that quantum interference between the D_2 and D_1 transitions is of major importance, and the two lines blend together in the radiative-transfer process. In the case of the beryllium lines deep down in the ultraviolet, the effects are instead masked by the high level of the continuum polarization at these wavelengths, in combination with the spectrum being very crowded with numerous lines that depolarize the continuum. What can be seen is that the Be II D_2 line clearly polarizes more than the surrounding continuum, while the Be II D_1 line polarizes significantly less than the continuum. To determine if the Be II D_1 line polarizes at all requires very careful untangling of its interaction with the continuum in the line-formation process, again is a complex and model-dependent radiative-transfer problem.

For these reasons the lithium and beryllium observations are inconclusive with respect to the D_1 anomaly. The Na and Ba D_1 lines are the only "clean" cases that we know that are not subject to these complications. For them the D_1 anomaly is evident from direct visual inspection, without the need for any modeling or radiative transfer.

Let us now compare with a multiplet that has the same fine structure as the D_2–D_1 system but without nuclear spin or hyperfine structure. The by far most prominent such case is that of the Ca II K 3933 and H 3968 Å lines. The K line has the J quantum numbers of a D_2 transition, H of a D_1 transition. As shown in Stenflo (1980) quantum interference between the K and H scattering transitions has a profound effect on the polarization profile of this line system, which dominates the observed polarization over more than 200 Å of the spectrum, because the line opacity remains high in the dispersion wings very far from the resonances due to the high calcium abundance on the Sun. Due to the quantum interference, the polarization changes sign and becomes negative between the K and H lines. This extended polarization curve however has a zero crossing exactly at the center of the H line, with no indication of any intrinsic line polarization from the H line transition (cf. the observations reproduced in Stenflo, 2003b). The observed polarization from the Ca II K and H line system is in perfect agreement with the expectations from standard quantum mechanics, with no sign of any anomaly.

Chapter 11

Why the D_1 line is expected to be unpolarized

In this chapter we will give an abbreviated overview of how scattering polarization is calculated in quantum mechanics, without giving the full expressions needed to carry out the calculations explicitly. Various versions of this quantum theory have been developed in the past, but they all give the same predictions. The two main theoretical frameworks are on the one hand the density matrix formalism in terms of irreducible tensors (cf. Landi degl'Innocenti, 1996; Bommier, 1997a,b; Trujillo Bueno, 2001, 2003), on the other hand the formalism based on the Kramers-Heisenberg scattering amplitudes. Here we will use the latter approach, since we find it conceptually more transparent, and it directly lends itself to the definition of intrinsic scattering polarizabilities. For the detailed exposition of the theory we refer to the monograph of Stenflo (1994) and to later elaborations of the theory (Stenflo, 1997, 1998).

The basic component of the theory is the complex probability amplitude for scattering from initial magnetic substate a to final substate f via all possible intermediate substates b. It is given by the Kramers-Heisenberg formula

$$w_{\alpha\beta} \sim \sum_b \frac{\langle\, f\,|\,\hat{\boldsymbol{r}}\cdot\boldsymbol{e}_\alpha\,|\,b\,\rangle\,\langle\, b\,|\,\hat{\boldsymbol{r}}\cdot\boldsymbol{e}_\beta\,|\,a\,\rangle}{\omega_{bf}-\omega-i\gamma/2}\,. \tag{11.1}$$

Here ω is the angular frequency of the scattered radiation, $\hbar\omega_{bf}$ the energy difference between the excited and final states, and γ the damping constant that accounts for the broadening of the excited state. Because of energy conservation $\omega_{bf}-\omega$ could also be replaced by $\omega_{ba}-\omega'$, where ω' is the frequency of the incident radiation. $\hat{\boldsymbol{r}}$ is the position operator (which is proportional to the dipole moment operator) and $\boldsymbol{e}_{\alpha,\beta}$ the linear unit polarization vectors for the radiation. The dependence of the scattering amplitude on the quantum numbers of the scattering transition is contained in the matrix elements, which are expanded via the Wigner-Eckart theorem, resulting in an expression that contains the oscillator strengths and 3-j symbols with the quantum-numbers (cf. Stenflo, 1998).

To get from probability amplitudes to probabilities (scattering cross sections), we have to form bilinear products between the scattering amplitudes. The scattering probability for the components I, Q, U, and V of the Stokes vector that represents the full polarization state is given by the Mueller scattering matrix $\mathscr{M}_s$. It describes how the incident Stokes vector is transformed to the scattered Stokes vector and is a function of both frequency, scattering angles, and quantum numbers. If we ignore an unimportant proportionality factor, it can be written as

$$\mathscr{M}_s = \mathscr{T} \left(\sum_{m_a} \rho_{m_a m_a} \sum_{m_f} \boldsymbol{w} \otimes \boldsymbol{w}^* \right) \mathscr{T}^{-1} , \tag{11.2}$$

(Stenflo, 1994, pp. 41, 122 and 174). The symbols $\otimes$ and $*$ stand for tensor product and complex conjugation, respectively. Note in (11.1) that the scattering amplitude $w_{\alpha\beta}$ is obtained by a *coherent* summation over all possible intermediate states b while the summations over the initial and final magnetic substates m_a and m_f in (11.2) are done in an *incoherent* way. In other words, the intermediate or excited state is a mixed quantum state ("Schrödinger cat"), while the initial and final states are definite states. $\rho_{m_a m_a}$ represent the relative populations of the initial substates and are only needed if the initial atomic state is polarized. If it is unpolarized, as in our laboratory scattering experiment (cf. section 12.1), it can be omitted from (11.2), since then all ρ are the same and only contribute to a proportionality constant that we can ignore. $\mathscr{T}$ in (11.2) is a purely mathematical transformation matrix without physical contents.

In the limit of weak magnetic fields, when the Zeeman splitting is much smaller than the Doppler width, all components of the Mueller matrix have the same frequency profile Φ, which therefore can be factorized out as a scalar:

$$\mathscr{M}_s = \Phi(\omega)\, \mathscr{P} . \tag{11.3}$$

The matrix part $\mathscr{P}$ that is left is called the *phase matrix*. It is frequency independent except for effects of quantum interference between different states of total angular momentum (see below). If we disregard the circular polarization, which is not relevant here, since it is entirely decoupled from the linear polarization in the weak-field case, the phase matrix can be written as

$$\mathscr{P} = W_2\, \mathscr{P}_R + (1 - W_2)\, \mathscr{E}_{11}\,. \tag{11.4}$$

Here $\mathscr{P}_R$ is the Rayleigh phase matrix that describes classical dipole scattering. With the Stokes $+Q$ direction perpendicular to the scattering plane

$$\mathscr{P}_R = \frac{3}{4} \begin{pmatrix} 1 + \cos^2\theta & \sin^2\theta & 0 & 0 \\ \sin^2\theta & 1 + \cos^2\theta & 0 & 0 \\ 0 & 0 & 2\cos\theta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{11.5}$$

where θ is the angle between incident and scattered radiation. $\mathscr{E}_{11}$ is the matrix that has its first component equal to unity while all the other components are zero. W_2 therefore represents the fraction of scattering processes that behave like polarizing, classical dipole scattering, while the remaining fraction, $1 - W_2$, represents unpolarized, isotropic scattering.

11.1 Dependence of polarizability on angular momentum

The parameter W_2 in (11.4) is the parameter that represents the intrinsic polarizability of a scattering transition. If we have a transition with total angular momentum quantum numbers $J = J_a \rightarrow J_b \rightarrow J_f$, and we ignore quantum interference between excited states with different J_b, then W_2 can be expressed explicitly in a compact way as follows (Landi Degl'Innocenti, 1984):

$$W_2 = c_{af}^{(2)} / c_{af}^{(0)}\,, \tag{11.6}$$

where the coefficients are given by 6-j symbols:

$$c_{af}^{(K)} = \begin{Bmatrix} 1 & 1 & K \\ J_b & J_b & J_a \end{Bmatrix} \begin{Bmatrix} 1 & 1 & K \\ J_b & J_b & J_f \end{Bmatrix}\,. \tag{11.7}$$

In this special case W_2 is a number that is a function of the three quantum numbers J_a, J_b, and J_f involved but does not depend on wavelength. When there is quantum interference between two or more excited states of different total angular momentums J_b, then W_2 will vary with wavelength. The interferences appear as additional terms in the nominator of (11.6) but do not affect the denominator, which only contains the squared scattering amplitudes weighted according to the transition strengths. A complete set of algebraic expressions for calculating the interference terms, the relative weights, and the functional form of W_2 for any combination of quantum numbers have been given in Stenflo (1997). All these formulas are also valid in the hyperfine structure case simply by using quantum number F instead of J for the total angular momentum. When we want to add together the contributions from a whole multiplet with different initial and final states J_a and J_f, then we need to weigh the contributions to the nominator and denominator in (11.6) with the multiplicity products $(2J_a + 1)(2J_f + 1)$.

W_2 as calculated with (11.6) and (11.7) is always positive for Rayleigh or resonant transitions, for which the initial and final states a and f are the same. In contrast, W_2 is often negative for Raman or fluorescent transitions, for which the initial and final states are different. This sign difference leads in the case of the D_1 hyperfine system to exact cancellations between the resonant and fluorescent contributions, which is the reason why the net intrinsic polarizability is predicted to be exactly zero according to the described quantum theory, as we will see in section 11.3 below.

The maximum value W_2 can have is unity, which corresponds to classical dipole scattering. According to (11.4) and (11.5) for 90° scattering, it results in 100 % linear polarization with the electric vector perpendicular to the scattering plane.

11.2 Hyperfine structure of D_2 and D_1

Sodium's nuclear spin of $I = 3/2$ couples to the electronic angular momentum J to split the $J = 1/2$ state (common ground state for both the D_2 and D_1 transitions, and also representing the excited state of D_1) into the two hyperfine states with total angular momentum $F = 1$ and 2, while the $J = 3/2$ state (excited state of D_2) is split into the four hyperfine states $F = 0$, 1, 2, and 3. The possible transitions between these various F states result in 6 spectral components for D_2 and 4 for D_1. Since the size of the hyperfine splitting is about an order of magnitude larger in the ground state as compared with the excited states, the components appear clustered in two groups: one blueshifted group related to $F = 1$ in the ground state, and a redshifted group related to $F = 2$ in the ground state.

Figure 11.1 shows the relative wavelength displacements and absorption strengths of the various hyperfine structure components. The transition strengths are given by $g_a f_{ab}$, where f_{ab} is the absorption oscillator strength,

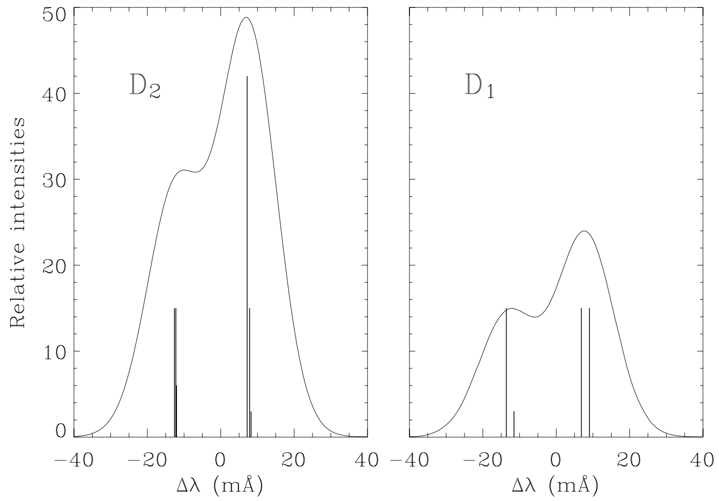


Fig. 11.1: Hyperfine structure of sodium. The vertical lines indicate the relative wavelengths and absorption transition strengths of the various hyperfine structure components, while the solid curves represent the optically thin absorption probability profiles, assuming a thermal Doppler width for a temperature of 150° C, which is typical for the conditions in the sodium vapor cell used for our scattering experiment.

and $g_a = 2F_a + 1$ is the statistical weight of the lower state. The various hyperfine structure components have in principle different intrinsic polarizabilities W_2, but the D_1 components turn out to have zero polarizabilities due to exact cancellation between the resonant and fluorescent contributions, as we will see in the next section. The shape of the absorption profiles is given by Voigt profiles defined by the two parameters damping and Doppler width. The damping width is determined by the inverse life time of the excited state, which is $6.15 \times 10^7\,\mathrm{s}^{-1}$ for D_2, $6.14 \times 10^7\,\mathrm{s}^{-1}$ for D_1 (Volz et al., 1996). The Doppler width is in general determined by both thermal and turbulent velocities. For our laboratory experiment significant turbulent velocities are not present, and the typical cell temperatures used are 100°– 200° C. In Fig. 11.1 we have plotted the composite shapes of the D_2 and D_1 absorption profiles assuming a temperature of 150° C for the thermal Doppler width, which in velocity units becomes $0.553\,\mathrm{km\,s}^{-1}$ and in wavelength units 10.8 mÅ. We see that the Doppler smearing is comparable to the hyperfine structure splitting, so that the two groups of hyperfine components are only partially resolved.

On the Sun the Doppler smearing is much larger. Typical thermal velocities are $1.5\,\mathrm{km\ s}^{-1}$, corresponding to a temperature of 5533 K, while typical turbulent velocities are $2\,\mathrm{km\ s}^{-1}$. Combining these values quadratically we obtain a net Doppler width of 2.5 km s^{-1} in velocity units, 49.1 mÅ in wavelength units. This is approximately five times broader than for the laboratory experiment, implying that the hyperfine splitting will be completely masked by the velocity smearing on the Sun.

11.3 Exact cancellations in the D_1 system

In the preceding sections we have introduced the formalism which allows us in principle to calculate the polarizabilities for any atomic transition. The explicit calculations go beyond the scope of this work, so we limit ourselves to a phenomenological treatment instead, in terms of transition strengths and branching ratios, and by introducing a fiducial magnetic field that we let go to zero while invoking the principle of spectroscopic stability. This phenomenological treatment gives the same results as the Kramers-Heisenberg scattering theory while allowing us to get a more intuitive physical understanding of the problem. We will use this approach explicitly for the D_1 system to clarify why the intrinsic polarizabilities remain zero in the D_1 line core according to standard quantum mechanics in spite of the hyperfine structure splitting.

D_1 scattering transitions between each separate combination of F quantum numbers have intrinsic polarizabilities which exactly cancel each other. An overview how this cancellation occurs for all the possible scattering transitions in the D_1 system is given in Fig. 11.2. There are 8 ways in which the F_a, F_b, and F_f states can be combined via a scattering transition. We have in Fig. 11.2 represented them by 4 diagrams, each containing one resonant and one fluorescent transition, both having a common absorption transition. The number given beside each absorption transition is the transition strength $g_a f_{ab}$ in relative units (relative to the absorption strength of the first diagram). The emission transitions are labeled by two numbers: first the transition strength, and below it, larger and in slanted bold face, the scattering polarizability W_2. The emission transition strengths, which are proportional to the oscillator strengths f_{bf}, have been normalized so that they sum up to unity for each of the four diagrams. They therefore represent branching ratios.

Let us consider the first diagram. Once the $F = 1$ level has been radiatively excited from ground state $F = 1$, the spontaneous decay can occur via two channels: $1/6$ of the cases go via the resonant channel (back to level $F = 1$), $5/6$ of the cases via the fluorescent channel. The respective polarizabilities W_2 are 0.25 and -0.05. As the relative polarization contribution is proportional to the product between the branching ratio and the polarizability, the net polarization from the first diagram is proportional to $\frac{1}{6} \times 0.25 + \frac{5}{6} \times (-0.05) = 0$. Similarly, the products of the branching ratios and the polarizabilities cancel each other exactly in each of the diagrams.

Since the two emission profiles of each diagram have a small relative wavelength shift, the exact cancellation that we speak of refers to the wavelength-integrated case. Due to the wavelength shift a residual polarization is possible if only single wavelengths are considered. However, due to the Doppler broadening, which in the solar case is much larger than the hyperfine structure splitting, the resulting spectral smearing constitutes effectively a wavelength integration, suppressing the residual polarization effects to insignificant values. In the case of our laboratory experiment the Doppler smearing is about five times less than for the Sun, but as we have to do with an optically thin medium and integrate the scattered radiation over a spectral window that is much broader than the hyperfine structure splitting, the polarization cancellation is exact for all practical purposes.

The representation in Fig. 11.2 does not include the quantum interferences between the excited $F = 1$ and 2 states, resulting from the coherent sum in (11.1). It can however be shown that the interference terms for the fluorescent transitions are of equal magnitude but opposite sign with respect to the interference terms for the resonant transitions. The contributions to W_2 of the interference terms therefore also cancel exactly.

We have seen that the cancellations depend on the product of transition strengths and polarizabilities. Transition strengths, or oscillator strengths, are abundantly used and firmly established in atomic physics. The polarizability for scattering via a given set of total angular momentum states F_a, F_b, and F_f can be obtained by first considering scattering between each combination of magnetic substates, and then sum over the contributions from all combinations of such substates, as in (11.2).

11.4 Relation between W$_2$ and the polarized spectrum

From the quantum theory, accounting for nuclear spin and quantum interferences, we can obtain $W_2(\nu)$ as a function of frequency ν. The resulting functional form of W$_2$ is given in Fig. 11.3 for the neighborhood of the D$_2$ line and for two values of the Doppler width corresponding to the solar and to our laboratory conditions respectively. Due to the cancellation effects in

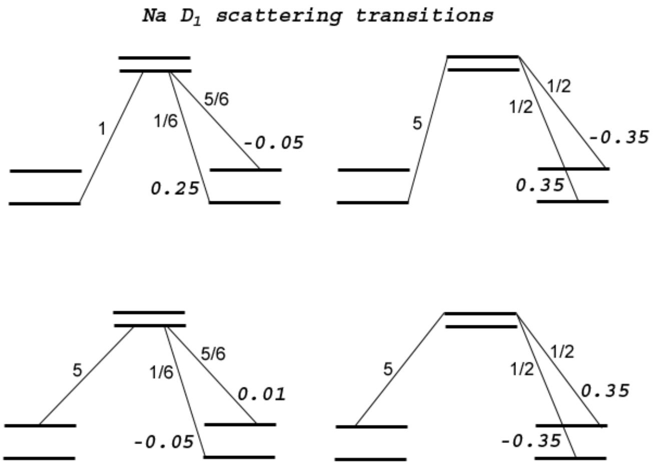


Fig. 11.2: Overview of the perfect cancellations between the resonant and fluorescent polarizabilities for the scattering transitions within the D$_1$ hyperfine structure multiplet. There are 8 possible ways in which the initial, intermediate, and final states can be combined (plus the interferences between the excited states, not shown in this figure). Each transition is labeled by its strength, in relative units. The strengths of the emission transitions have been normalized to represent branching ratios. We also give in slanted bold face the polarizability W_2 of the corresponding scattering transition. The net polarizability is obtained by averaging the individual polarizabilities, weighted by the respective transition strengths, which results in zero because of cancellations.

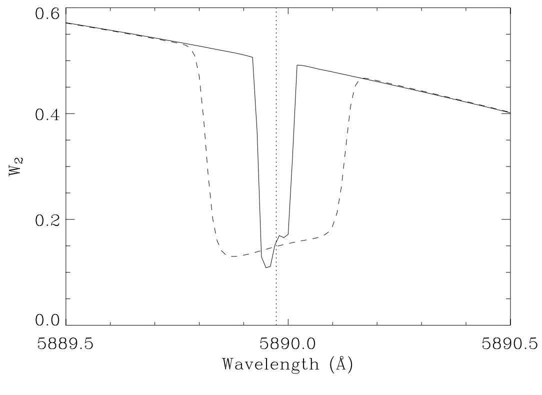


Fig. 11.3: Wavelength dependence of the scattering polarizability W_2 around the core of the D$_2$ line, for different values of the Doppler width. The solid curve represents the case of a thermal Doppler width for a temperature of 150° C, typical for our laboratory experiment, while the dashed curve corresponds to a Doppler width of 2.5 km s^{-1}, which is typical for the solar case. Without nuclear spin there would be no depression of W_2 in the Doppler core of the line, but W_2 would be 0.5 at line center. The slant of the W_2 function outside the line core is due to quantum interference between the D$_2$ and D$_1$ scattering transitions.

the D$_1$ system, W$_2$ remains zero throughout the line core region, defined by the Doppler-smearing interval around the hyperfine components.

Next let us convert the polarizability parameter W_2 to the polarization p for 90° scattering. Since for this particular scattering case Stokes U and V are zero, we only need to consider the part of the phase matrix $\mathscr{P}$ in (11.4) that relates to I and Q. The Rayleigh scattering part of $\mathscr{P}$ is then, in the special case of 90° scattering,

$$\mathscr{P}_{R,\,90^\circ} = \tfrac{3}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} . \tag{11.8}$$

For incoming radiation that has 100 % Q/I linear polarization, the Stokes vector is

$$\boldsymbol{I}'_{\perp,\,\parallel} = \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} . \tag{11.9}$$

The $+$ and $-$ signs are for polarization perpendicular and parallel to the scattering plane, respectively. Applying the phase matrix $\mathscr{P}$ of (11.4) on $\boldsymbol{I}'$, using $\mathscr{P}_R$ of (11.8), we obtain I and Q, and thereby $p = Q/I$ of the scattered radiation. The result is

$$p = \frac{\frac{3}{2}W_2}{1 + \frac{1}{2}W_2} \tag{11.10}$$

for the $\boldsymbol{I}_\perp$ case, while p is always zero in the $\boldsymbol{I}_\parallel$ case (incoming radiation polarized parallel to the scattering plane). Inverting (11.10), we obtain

$$W_2 = \frac{2p}{3 - p} . \tag{11.11}$$

To model the observed polarization on the Sun, we need to take into account that (a) the line opacity is competing with the background continuum opacity, (b) that the background continuum is also polarized, and (c) that the anisotropy of the radiation field that induces the scattering polarization is small in comparison with the 90° scattering case. To account for these three aspects in a very simplified model (as introduced in Stenflo, 1980) we start from the generalized version of (11.6), where

$$W_2(\nu) = C^{(2)}(\nu)/C^{(0)}(\nu) . \tag{11.12}$$

The quantum interferences only occur in the numerator $C^{(2)}(\nu)$, whereas the denominator $C^{(0)}(\nu)$ represents the combined line opacity from D_2 and D$_1$. Doppler smearing has been applied separately to $C^{(2)}(\nu)$ and $C^{(0)}(\nu)$. Let us now introduce an "effective" W_2 that accounts for the continuum opacity and polarization, defined as

$$W_{2,\,\mathrm{eff}}(\nu) = W_2(\nu)\,\frac{C^{(0)}(\nu)}{C^{(0)}(\nu)+a} + b\,\frac{a}{C^{(0)}(\nu)+a}\,. \tag{11.13}$$

The frequency-independent parameters a and b represent the continuum opacity and polarization, respectively. The first term in (11.13) shows how W_2 is scaled by the fraction of line opacity relative to the total (line plus continuum) opacity, while the continuum polarizability b is scaled (in the second term) by the ratio of continuum opacity to total opacity. To compare with the observed solar polarization we rescale $W_{2,\,\mathrm{eff}}(\nu)$ with a frequency-independent scaling factor c, to account for the geometric dilution factor (small anisotropy of the incident radiation field). With $W_2(\nu)$ given by quantum physics and using a Doppler broadening that is representative of the solar case (dashed line in Fig. 11.3), we can now choose the three free parameters a, b, and c to obtain a good fit to the observed polarization in the line wings of the D$_2$–D$_1$ system. The result is shown as the dotted line in Fig. 10.1.

While the agreement between the simplified model and the observations is remarkably good outside the two line cores, it fails in the core regions. If the modeling is refined by using full numerical radiative transfer for a realistic model atmosphere, the observed polarization profile in the core of the D$_2$ line can be successfully modeled when including the effects of partial frequency redistribution PRD (Fluri et al., 2003). In contrast, PRD effects are unable to produce any polarization peak in the D$_1$ core, since there is no intrinsic polarization there to redistribute. No playing around with the free parameters of the models is able to change this. Even if we are given the freedom to choose the most unrealistic model atmospheres just for the purpose of finding something that could fit the observed polarization, it does not help the D$_1$ problem. If there is no intrinsic polarizability there because of atomic physics/quantum mechanics, then no choice of solar model will help.

11.5 Influence of collisions and magnetic fields

Collisional effects can appear in various forms. The elastic collision rate γ_c is responsible for line broadening and frequency redistribution. The depolarizing collision rates are denoted $\gamma_c^{(K)}$, where K refers to the $2K$-multipole that is involved. Let us here note that it is the same K that appears in the intrinsic polarizabilities W_K. Here we will only consider $K = 2$, since it is this multipole that relates to the atomic alignment and is most relevant for the scattered linear polarization. For details and a more complete treatment we refer to chapter 10 in the monograph by Stenflo (1994).

The depolarizing collisions both suppress the amount of scattering polarization. The scattered linear polarization in the line core is changed by the collisional depolarization factor

$$k_c^{(K)} = \frac{\gamma_N}{\gamma_N + \gamma_c^{(2)}}\,, \tag{11.14}$$

where γ_N is the natural line width (inverse life time of the state). A simplified classical collision theory gives $\gamma_c^{(K)} = 0.5\gamma_c$ for both $K = 1$ and 2 (Stenflo, 1994), while detailed quantum mechanical calculations give

$$\gamma_c^{(2)} = 0.38\gamma_c \tag{11.15}$$

(Berman & Lamb, 1969).

In the solar chromosphere, where the cores of the sodium D_2 and D_1 lines are formed, $k_c^{(K)} \approx 1$ due to the low density in these higher atmospheric layers. In our laboratory experiment, however, $k_c^{(K)} < 1$ due to collisions between the sodium atoms and the buffer gas of argon (cf. section 12.1).

Let us now discuss the role of magnetic fields in our experiment. First, we have verified with a magnetometer that the strength of magnetic field in the sodium cell is less than 1 G. Besides the Earth's magnetic field of about 0.3 G, there are no major magnetic disturbances from the outside world.

The transverse Zeeman effect, which generates linear polarization that would interfere with our measurements, can be ignored in this context. For weak, optically thin spectral lines, which we have in the case of our laboratory experiment, we have (e.g. Stenflo, 1994)

$$Q/I \approx \left(\frac{v_H \sin\gamma}{2}\right)^2 \frac{1}{I}\frac{\partial^2 I}{\partial v^2} \tag{11.16}$$

for an emission line I, if the positive Q direction is defined to be along the projection of the magnetic field vector, and γ is the angle between the magnetic field and the line of sight. $v = \Delta\lambda/\Delta\lambda_D$ is the wavelength scale expressed in Doppler units, $v_H = \Delta\lambda_H/\Delta\lambda_D$ is the Zeeman splitting in Doppler units.

The optically thin intensity profile I is given by a Voigt function, but since a Voigt function can be closely approximated by a Gaussian in the line core, we can for estimation purposes easily compute the core polarization analytically using $I = \exp(-v^2)$. With the effective Landé factor 4/3, valid for the D_1 line for field strengths larger than the hyperfine level-crossing field strengths, and a Doppler width of 10.8 mÅ, which represents the thermal width for a temperature of 150° C, typical for our laboratory experiment, we find $Q/I = 2.0 \times 10^{-6} B^2$, if B is given in units of G. For $B = 1$ G the polarization level is two orders of magnitude lower than the sensitivity of our experiment. Below the level-crossing fields (below about 10 G) the effective Landé factor is on average 1.75 for the hyperfine transitions. This gives $Q/I = 2.6 \times 10^{-6} B^2$, which does not change the argument. This demonstrates that the transverse Zeeman effect does not play any role for the laboratory experiment.

On the Sun the transverse Zeeman effect is about 20 times weaker than in the laboratory case, due to the about 4.5 times larger Doppler width (produced by the much higher temperatures plus the turbulent broadening). Values of $Q/I = 10^{-3}$, comparable to those from the non-magnetic scattering polarization near the solar limb, are reached for fields of 100 G.

Chapter 12

Laboratory experiment

We have seen in chapter 11 why D_1 type scattering transitions must be unpolarized according to standard quantum mechanics, regardless of what quantum formalism we adopt. This theoretical prediction is at odds with the solar observations reviewed in chapter 10, which for quiet solar regions always show a symmetric polarization peak in the core of both the Na I D_1 5896 and Ba II 4934 Å lines. Since this contradiction has been known for nearly a decade now but no explanation has yet been found, the question has been raised whether this is a problem for solar physics or for quantum mechanics. The Sun is a "messy" object with an optically thick and structured atmosphere and a complex and intermittent magnetic field that may carry much of its magnetic flux in a tangled form on scales far smaller than can be resolved by our instruments. It is therefore highly desirable to perform a clear-cut experiment under controlled and highly simplified conditions. It is for this purpose that we have set up our laboratory experiment.

In comparison with the solar case, the laboratory environment has a number of major advantages: (1) A 90° scattering angle can be chosen, rather than having a radiation field with a small anisotropy and integration over a large range of scattering angles. (2) The incident radiation can be chosen to be 100 % linearly polarized, both perpendicular and parallel to the scattering plane, rather than having nearly unpolarized incident radiation on the Sun. (3) Single-scattering events in an optically thin medium can be observed, rather than being limited to the case of multiple scattering in an optically very thick medium.

12.1 Sodium cell

A sodium cell tailor-made for the purpose of this experiment has been built by A. Cacciani in Rome. Sodium cells based on the Cacciani design have

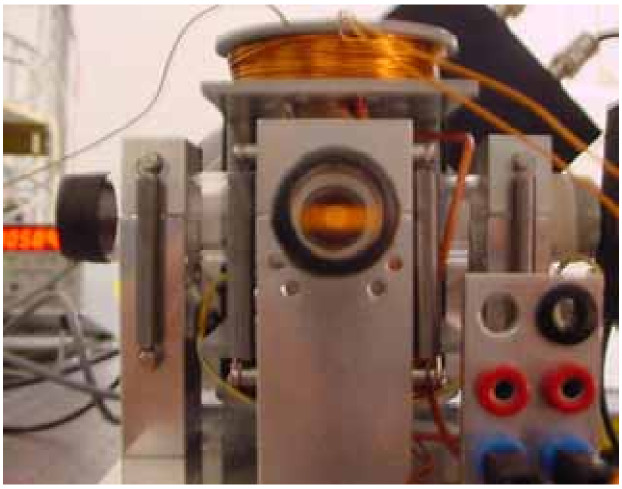


Fig. 12.1: Photograph of the sodium vapor cell. The picture is taken through the optical arm facing the exit channel, the light enters the cell from the right. When the cell temperature is below about 160°C the sodium gas becomes optically thin. Through the yellowish glowing band of scattered radiation in the cell center one can discern the circular field aperture of the detector optics.

been used for decades in solar physics, in particular for the imaging of solar magnetic and velocity fields (Cimino et al., 1968; Cacciani & Moretti, 1994; Cacciani et al., 1997) and for helioseismology (Tomczyk et al., 1995). These cells are generally equipped with strong magnets that produce kG magnetic fields so that a narrow-band filter with two pass bands, one in each line wing, can be obtained, and modulation between the pass bands can be done by polarization switching. In contrast our sodium cell does not have any fixed magnets, since we are interested in the non-magnetic scattering case. Instead of having a straight cell with one entrance and one exit window, our cell is in the form of a cross with four arms, having four entrance/exit windows. This allows us to have the viewing direction at right angles to the incident, illuminating radiation, to record the photons that are scattered by 90°.

Perpendicular to the plane with the four optical arms are two other arms that contain the pits with the solid sodium, surrounded by heating mats to produce sodium in gaseous form. A temperature sensor allows us to regulate the heating such that a given cell temperature can be chosen. The number density of sodium atoms increases with temperature, but when the temperature is reduced below about 160° C the sodium gas in the cell becomes optically thin (Fig. 12.1)

The cell is filled with argon buffer gas at a pressure of about 27 mbar.

The buffer gas is needed to prevent the sodium to condensate on the four windows, which would make them opaque, but it has the side effect of introducing collisional depolarization. To obtain an estimate of the expected depolarization we can use the Lindholm approximation of collisions with van der Waals interactions (Mihalas, 1978) between the neutral atoms of sodium and argon in the cell. We find that the undisturbed atomic life time of sodium between collisions (with argon atoms) is of order 10^{-8} s, a value that consistently explains the discrepancy between the observed and theoretically expected D_2 polarization (cf. section 12.5). The expected temperature dependence of the collisional depolarization is slight. According to the van der Waals theory, the collision rate scales with $v^{3/5}$. Since the typical thermal velocities $v \sim \sqrt{T}$, the collision rate is expected to scale with $T^{0.3}$. This corresponds to an increase of only 7 % when the temperature increases from 100° to 200° C, which is a relatively minor change.

The estimate of the collision rate can also be used to answer the question if the sodium atoms are optically pumped or not in our experiment. Optical pumping is a mechanism that was proposed by Degl'Innocenti (1998) and at first it seemed to be a convincing and promising way to resolve the D_1 mystery. But later more self-consistent treatments for realistic solar conditions (Trujillo Bueno et al., 2002; Kerkeni & Bommier, 2002; Klement & Stenflo, 2003; Fluri et al., 2003) have shown that the radiation anisotropy in the solar atmosphere is too small to produce the desired pumping effect. The theoretically predicted polarization is 1-2 orders of magnitude too weak. Moreover, the predicted polarization profiles turned out to have the wrong symmetry. The optical pumping process can only work if the atom is radiatively excited many times before a collision happens. With the light source used in our experiment the time between two successive radiative excitations is of order 10^{-4} s which is several orders of magnitude longer than the time between collisions. We can therefore decidedly rule out optical pumping in our experiment.

12.2 Experimental setup

12.2.1 Light source

As light source we use a 15 W low-pressure sodium vapor lamp. It has the advantage over a broad-band lamp that all its radiative energy is concentrated around the D_1 and D_2 lines, where it is needed, with no significant contributions from unwanted wavelengths. This greatly helps in the elimination of stray light. To make absolutely sure that any spurious light from other wavelengths than D_1 and D_2 is reduced to a negligible level, we in addition use, in front of the detector, a 10 Å wide interference filter whose passband is centered at 5893 Å (average wavelength of D_1 and D_2).

12.2.2 Imaging optics

Illumination optics The lens L_1 ($f_1 = 25$ mm) focuses the entrance pupil in front of the lamp onto lens L_2 ($f_2 = 80$ mm) which acts as field lens. L_2 re-images the entrance pupil into the center of the cell. The f-ratio of the beam entering the cell is f/10. This so-called Köhler configuration ensures a homogeneous illumination of the region where the scattering happens. The diameter D_c of this region is about 7 mm. The ratio

$$H = \frac{D_c}{f/\#} \propto \sqrt{n_{\mathrm{phot}}} \tag{12.1}$$

between D_c and the f-ratio $f/\#$ of the entering light cone, the so-called (paraxial) Helmholtz-Lagrange invariant H (cf. section 3.2) defines the number of photons n_{phot} that are available for scattering. The optical design is chosen so that H is maximized. The constraints are the entrance arm of the cell where reflections off the glass walls must be avoided and the size of the region around the cell center where the gaseous sodium is confined. The total distance between the lamp and the cell center is 383 mm. This scaling leaves enough space in between the lenses to install the polarization optics.

Detector optics The light scattered at 90° gets transferred to the detector, a ZIMPOL camera, by 3 further lenses. Lens L_3 ($f_3 = 60$ mm) re-images the cell center onto field lens L_4 ($f_4 = 150$ mm). L_4 acts as collimator for the Fabry-Pérot (FP) etalon (see below) by projecting the entrance pupil of the detector optics, defined by L_3, to infinity. Lens L_5 ($f_5 = 250$ mm) finally re-images the pupil onto the camera. The imaging of the entrance pupil ensures a homogeneous intensity distribution within the sodium lines in the recorded FP spectrum.

12.2.3 Fabry-Pérot etalon

One of our FP etalons (which are extensively discussed in part I of this book) acts as spectrograph. We make use of the angle dependent blueshift of the FP transmission spectrum in a collimated setup (cf. sections 2.3

and 3) to separate the sodium lines. Fig. 12.3 shows an example of such a sodium spectrum. The D_1 and D_2 lines appear as concentric rings. We have scaled the detector optics such that at least one spectral order of each line is projected as a complete ring onto the camera.

12.2.4 Polarization optics and detector

The polarization optics comprises two polarization modulators and the corresponding calibration optics. The time modulated intensity signal generated by these modulators is in turn demodulated by a ZIMPOL camera to retrieve the polarization information. In this experiment we use a special configuration of two so called piezoelastic modulators (PEMs) acting as variable retarders (for a detailed description of this modulator type see e.g. Gandorfer, 1996; Bernasconi, 1997). The modulators are arranged in a symmetric way around the sodium cell and are alternately enabled and disabled during measurement. In doing so we can trace potential instrumental errors that could arise from an asymmetric polarization behavior of the cell. In case no such errors occur, it allows to confirm the symmetry of the scattering phase matrix expected from theory (cf. 11.5 and section 12.5.3). Modulator no. 1 is installed in front of the cell entrance arm. First, the light passes the linear polarizer and then goes through the PEM. The transmission axis of the polarizer is oriented perpendicularly to the scattering plane during normal operation but it can be turned by 90° to perform a reference measurement with zero scattering polarization (cf. section 12.4.2). Modulator no. 2 is situated after the cell exit arm and the light passes the PEM first and then the polarizer, always oriented perpendicularly to the scattering plane. The PEMs modulate three Stokes components simultaneously, namely Stokes I (DC component), Stokes V (at the fundamental frequency of 42 kHz) and Stokes Q (at the first harmonic). Stokes I and Q are the components we are actually interested in. Stokes V serves as null reference and additional tracer of systematic errors.

12.3 Calibration of the light source

For verification purposes and for a proper quantitative interpretation of the scattering measurements it is essential to test and calibrate the optical components used. One crucial component that we will discuss here is the light source, a low-pressure sodium vapor lamp. The interpretation of the mea-

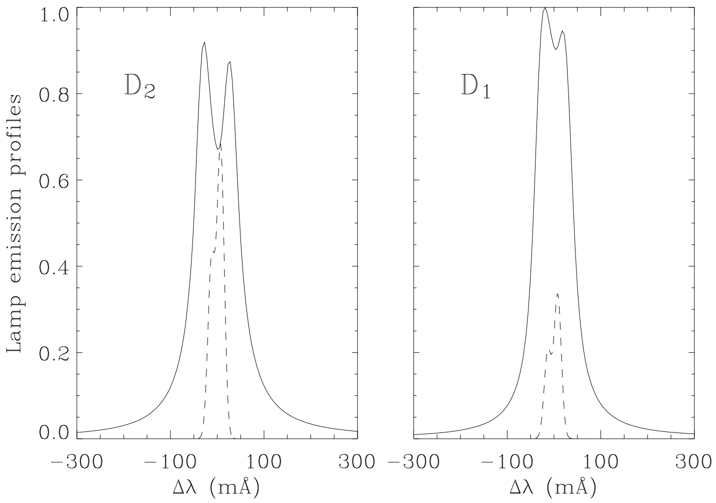


Fig. 12.2: The fully resolved spectral profiles of the D_2 and D_1 lines (solid curves), in relative units. For comparison, the dashed curves show (also in relative units) the same absorption probability profiles that were given in Fig. 11.1. The relative number of photons scattered in the two lines is determined by the integrated product of the solid and dashed profiles.

surement results in terms of true polarizabilities requires the knowledge of the spectral profiles and the relative contributions of the D_2 and D_1 lines.

The lamp profiles were recorded with the high-resolution spectrograph at the Istituto Ricerche Solari Locarno (IRSOL). First, the profiles were recorded with lower spectral resolution such that both the D_2 and D_1 emission profiles could fit on the same CCD frame. After accounting for possible flat-field effects we used these recordings to determine the ratio between the wavelength-integrated emission profiles of the D_2 and D_1 lines. It was verified that this ratio did not vary spatially along the slit over the emitting region of the lamp. Then each of these profiles was recorded separately with the highest spectral resolution, to obtain the fully resolved profiles. A global scale factor was then applied to each high-resolution profile to ensure that the ratio between their integrated intensities agreed with the ratio determined from the lower resolution measurements. The so determined profiles are illustrated in Fig. 12.2 as the solid curves. The superposed dashed curves are the optically thin absorption probability curves that were given in Fig. 11.1.

When comparing the four curves in Fig. 12.2 we notice the following properties: (1) The lamp profiles are much broader than the optically thin profiles, which shows that they are strongly saturated. (2) While the D_1 lamp profile has a larger amplitude than the corresponding D_2 profile, the integrated area of the D_2 profile is actually 10 % larger than that of D_1. The larger width of the D_2 profile is further evidence that it is more saturated than D_1. (3) This larger saturation of D_2 is expected, since as shown by the optically thin profiles, the D_2 line is twice as strong (in its optically thin area) as compared with the D_1 line. (4) Further evidence for the larger saturation of the D_2 profile is seen in the considerably larger self-absorption at line center.

The deviation from a spectrally flat illumination over the small wavelength range covered by the optically thin absorption profiles is largely determined by this self-absorption. The relative number of photons absorbed by each line in the scattering process is proportional to the wavelength-integrated products of the solid and dashed curves in Fig. 12.2. These products closely reflect the amplitude ratio between the line center intensities of the lamp profiles, not the ratio of their areas, since the narrow, optically thin probability functions only sample the core regions but ignore the wings. Since the line-center D_1/D_2 amplitude ratio is 1.35, the product value is half of this (since the areas of the probability functions are in proportion 1:2), so that the D_1/D_2 product ratio is about 0.67. This is the value that we need to use when interpreting the observed scattered intensities (Stokes I) as functions of the field angle (cf. section 12.2). However, the exact value used has little or no influence on the interpretation of the polarization values (Q/I or U/I) of the scattered radiation.

12.4 Absence of spurious signals

To be absolutely sure that no spurious signals can infiltrate and generate fictitious non-zero scattering polarization in the D_1 line great precautions must be taken to verify that any such spurious effects can be ruled out. We have taken great care in baffling the optics to make it very unlikely for any photons to reach the detector other than via scattering at the sodium gas in the cell.

The dark current of the ZIMPOL camera can be measured either by directly covering the CCD with a lid, or by simply putting a screen in front of the lamp to block it from directly illuminating the cell. In both cases, when the cell is cold or hot there is no discernible difference between these two versions of the dark current measurement. This verifies that stray light diffusing in the room cannot reach the detector, and that thermal emission from collisionally excited sodium atoms is insignificant.

Still there are other potentially serious effects that could invalidate the measurements, like reflections off the glass walls of the cell, and possible polarization effects that could have to do with stress in the cell windows or gradient effects in the cell of unidentified origin. Next we will indicate how all these effects can be ruled out.

12.4.1 Reflections off the glass walls

If light reflected off the glass walls of the cell could reach the detector, and as such reflections may be highly polarizing, fictitious polarization results could be created. Wall reflections that do not involve scattering in the sodium gas can however be measured separately by recording the apparent scattering signal when the cell is cold (while being illuminated by the lamp). These cold cell recordings do not show any discernible intensity signal above the dark current level. This verifies that such reflections can be ruled out.

Although cold cell reflections can be ruled out, another secondary reflection effect, although very unlikely, may be considered: if a photon scattered by a sodium atom in the hot cell is reflected on a glass wall before reaching the detector, then it could in principle acquire polarization that does not have its origin in the scattering process. This possibility can be ruled out because we observe null polarization in the scattered radiation when the entrance polarizer is oriented to give 100 % horizontal input polarization, as discussed in the next subsection.

12.4.2 Null polarization with horizontal entrance polarizer

We have seen in section 11.3 that for incident radiation that is linearly polarized parallel to the scattering plane, the 90° scattered radiation (in the non-magnetic case) is for geometrical reasons always unpolarized, regardless of what the intrinsic polarizability W_2 of the scattering transition is. This null prediction must be reproduced by the observations, since it should not depend on the intricacies of quantum mechanics (like the D_1 mystery) but instead relate to simple geometry, which must be obeyed by any modifications of the theory. If our observations would fail to reproduce this null prediction, then this would be a serious indication that the measurements are contaminated by spurious effects.

Fortunately, as we will see in section 12.5 below, our observations indeed give zero scattering polarization in both the D_2 and D_1 lines when the entrance polarizer is oriented parallel to the scattering plane. This is contrasted with the very strong polarization signals obtained in D_2 when the entrance polarizer is oriented to polarize perpendicular to the scattering plane. If there were hidden, subtle instrumental effects that we have not properly taken into account but which were a source of spurious polarization signals, then these effects should also produce spurious polarization signals when a horizontal entrance polarizer is used. The circumstance that our polarization experiment passes the null polarization test gives us confidence that all the polarization signals that we see when a vertical entrance polarizer is used are indeed real and produced by the scattering events.

12.4.3 Temporal instabilities

As we have seen in section 12.2.4, a polarization measurement is eventually the measurement of an intensity varying in time. The demodulation basically consists in integrating the intensity during certain phases of the modulation period and in subtracting the respective values. This is the critical point where a systematic error, caused by temporal instabilities in the experiment, can sneak in. If the intensity would vary between two integrations because of a disturbance in the experiment that has nothing to do with the polarization modulation, a false polarization signal would be generated. However, with 42 kHz, the fundamental modulation frequency is well above the frequency spectrum of any disturbances like e.g. typical 50 Hz variations in the light source or small cell temperature drifts on the time scale of seconds.

12.5 Results

12.5.1 Interpretation model

The intensity modulation sampled with the ZIMPOL camera is the product of components of three different types. (1) A PEM with Mueller matrix $\mathscr{M}_{\mathrm{PEM}}(t)$. A detailed description of the PEM can be found for instance in Gandorfer (1996) or Bernasconi (1997), we only recapitulate the properties that are relevant here. The PEM can be modeled as a retarder with a particular time dependent and periodic retardation $\phi(t)$. The optical axis of the PEM is at 45° to the $+Q$ direction.

$$\mathscr{M}_{\mathrm{PEM}}(t) = \mathscr{R}(-45^\circ)\,\mathscr{M}_r(\phi(t))\,\mathscr{R}(45^\circ) \tag{12.2}$$

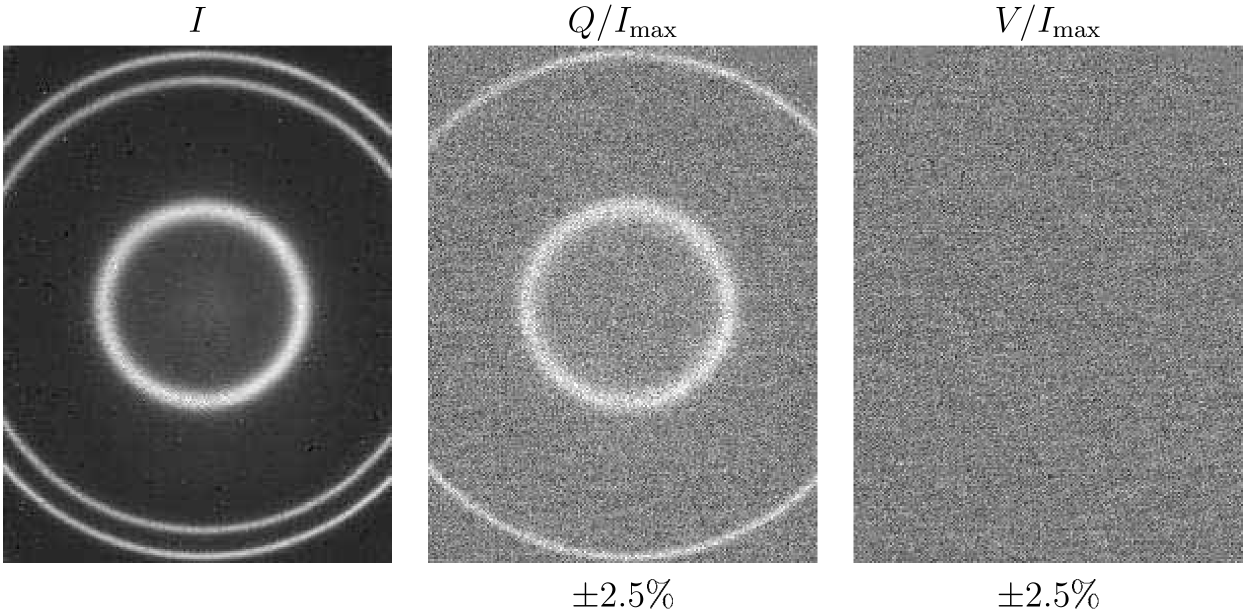


Fig. 12.3: Example from the series of 2D Stokes images the spectra in Fig. 12.4 are obtained from. Stokes V is recorded simultaneously with Q and serves as reference. The first and third rings (counted from the center) correspond to two successive orders of the D_2 line, the second ring corresponds to D_1. The amplitudes of the grayscales of Q/I_{max} and V/I_{max} are noted below the images.

where $\mathscr{M}_r$ is the Mueller matrix of a retarder with optical axis along $+Q$ and $\mathscr{R}$ the Mueller matrix describing a rotation of the Stokes reference frame. The exact form of the time dependence $\phi(t)$ is not relevant here. (2) A linear polarizer $\mathscr{M}_p$ with transmission axis along $+Q$ and (3) the phase matrix $\mathscr{P}$ describing the scattering in the cell. For the 90° scattering of the experiment the explicit form of the phase matrix is

$$\mathscr{P} = \frac{1}{4}\begin{pmatrix} 1-W_2 & 3W_2\,P_1 & 0 & 0 \\ 3W_2\,P_2 & 3W_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{12.3}$$

In the standard configuration a PEM is combined with a polarizer and receives an incoming Stokes vector $\boldsymbol{S} = (I, Q, U, V)^T$. The ZIMPOL camera samples the time-modulated intensity i.e. the first component $\bar{I}(t)$ of the outgoing Stokes vector

$$\bar{\boldsymbol{S}}(t) = \mathscr{M}_p\,\mathscr{M}_{\mathrm{PEM}}(t)\,\boldsymbol{S} \tag{12.4}$$

which is of the form

$$\bar{I}(t) = \frac{1}{2}\,[I + Q\cos\phi(t) + V\sin\phi(t)] \tag{12.5}$$

I is obtained from the time average of $\bar{I}$, and Q, V by demodulating the periodic patterns $\cos\phi(t)$ and $\sin\phi(t)$ respectively.

In the PEM1 configuration of our experiment PEM no. 2 is switched off and neutral to polarization. First, the light passes a linear polarizer with transmission along $+Q$. It then goes through the active PEM no. 1, enters the cell, gets scattered, passes through the passive PEM no. 2 and again through a polarizer before it hits the ZIMPOL camera. The time-modulated intensity seen by the camera is

$$\begin{aligned}\bar{I}_1(t) &= \left[\bar{\boldsymbol{S}}_1(t)\right]_1 = [\mathscr{M}_p\,\mathscr{P}\,\mathscr{M}_{\mathrm{PEM}}(t)\,\mathscr{M}_p\,\boldsymbol{S}]_1 \\ &= \frac{1}{2}\left[\left(1 - \frac{W_2}{4} + \frac{3W_2\,P_2}{4}\right) + \left(\frac{3W_2}{4} + \frac{3W_2\,P_1}{4}\right)\cos\phi(t)\right]\end{aligned} \tag{12.6}$$

Comparing this expression with (12.5) one can see that the degree of scattering polarization $q_1 := Q_1/I_1$ obtained from the PEM1 measurement is

$$q_1 = \frac{3W_2(1 + P_1)}{4 + W_2\,(3P_2 - 1)} \tag{12.7}$$

In the PEM2 configuration PEM no. 1 is switched off. First, the light passes the linear polarizer in front of PEM no. 1, then the passive PEM 1, the cell, the active PEM no. 2, and finally the second linear polarizer. The outgoing Stokes vector is thus

$$\bar{\boldsymbol{S}}_2(t) = \mathscr{M}_p\,\mathscr{M}_{\mathrm{PEM}}(t)\,\mathscr{P}\,\mathscr{M}_p\,\boldsymbol{S} \tag{12.8}$$

and, following a reasoning analogous to the PEM1 case, the degree of scattering polarization obtained from the PEM2 measurement is

$$q_2 = \frac{3W_2(1 + P_2)}{4 + W_2\,(3P_1 - 1)} \tag{12.9}$$

12.5.2 Upper limit for the D_1 polarization

In the case of D_1 the measured polarization (more precisely the upper limit for the polarization, as we will see below) is so faint that a possible asymmetry of the phase matrix will not be detectable. We can therefore assume that $P_1 = P_2$ and therewith $q_1 = q_2$ as can be read off (12.7) and (12.9).

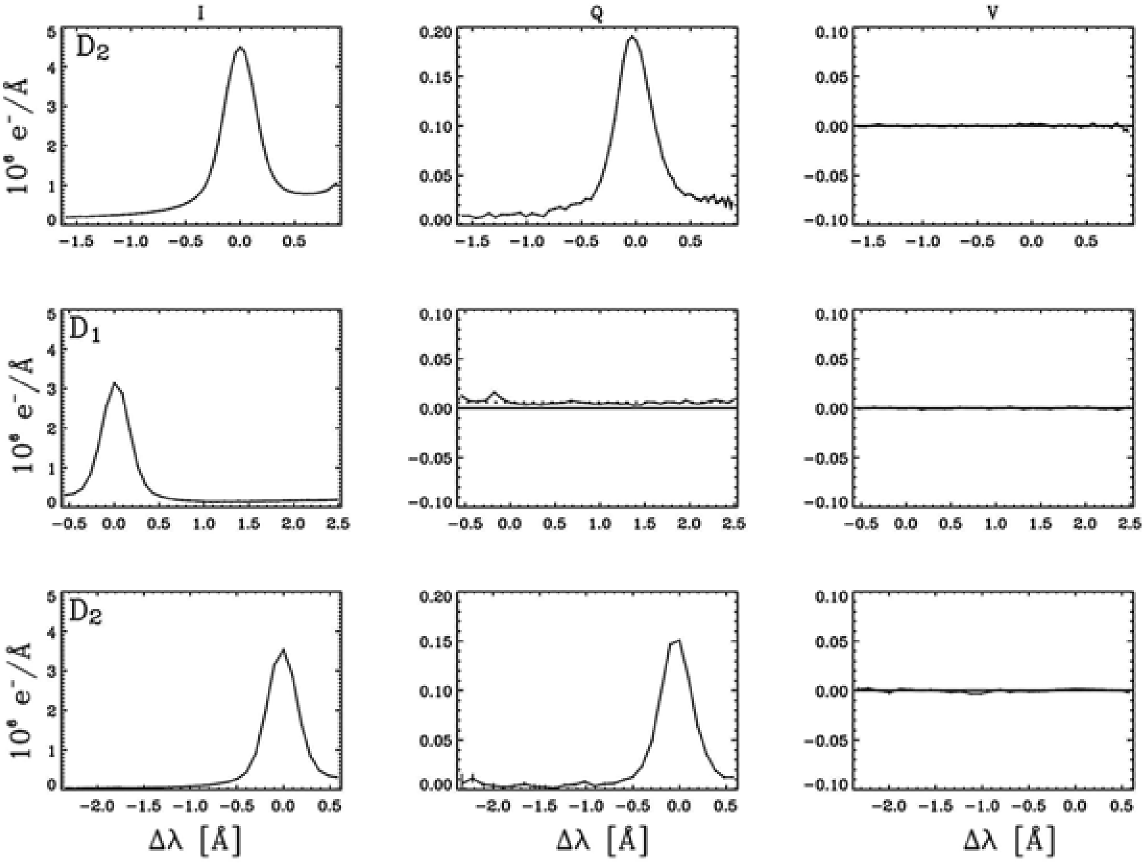


Fig. 12.4: Stokes spectra obtained from the 2D Stokes images (cf. Fig. 12.3). Each data point, at a given wavelength, is the average along the circle with the radius corresponding to this wavelength. The error bars indicate the rms value of the pixels contributing to the point.

Fig. 12.4 shows an overview of the measured Stokes spectra corresponding to the D_2 and D_1 rings of Fig. 12.3. The angular distribution is translated to a wavelength scale in Å relative to the respective line center. The spectral resolution drops from the inner rings (top panels) to outer rings (bottom panels) because of the quadratic behavior of the FP blueshift (cf. section 2.3).

Note that Stokes Q and V are not transformed to the usual polarization degrees Q/I and V/I but are plotted here as Q and V normalized to the maximum intensity of the respective spectral line they belong to. The main reason for this display format is the low intensity. By doing so we avoid any systematic errors from small dark current drifts. The dark current cancels out for Q and V which are basically intensity differences between different modulation phases.

Outside the spectral lines the Q signal does not drop to zero, but it stabilizes at a flat background value of about 0.2%. This polarized background comes from the D_2 line and is caused by the out-of-band transmission of the FP etalon.

Fig. 12.5 shows a detail view of the Q spectrum around the D_1 line. The six different measurement series averaged in Fig. 12.4 are plotted individually in this figure, together with the average spectrum. Except for a 3-sigma outlier, the signal varies randomly around the background value. On the extreme left border it begins to raise systematically to the D_2 value. The origin of the outlier is not conclusively identified. However, a physical connection with the D_1 line core is very unlikely because of its position 200 mÅ- more than a linewidth - off the D_1 line center.

The rms variation σ_Q of the Q signal permits to set an upper limit of 0.08% for Q in the D_1 line core which translates to a ratio $q_{\mathrm{D1}}/q_{\mathrm{D2}}$ between the D_1 and D_2 polarization degrees as follows:

$$\frac{q_{\mathrm{D1}}}{q_{\mathrm{D2}}} \leq \sigma_Q \frac{I_{\mathrm{D2}}}{Q_{\mathrm{D2}}\, I_{\mathrm{D1}}} \approx 2\% \tag{12.10}$$

I_{D1}, I_{D2} and Q_{D2} are the respective I and Q signals in the line centers of D_1 and D_2. In contrast the ratio measured on the Sun is of order 10% (cf. Fig. 11.4).

12.5.3 Upper limit for the asymmetry of the D_2 phase matrix

We have performed alternating measurements with two equal PEMs, one in front of the cell entrance arm and one behind the cell exit arm (cf. section 12.2.4). This particular arrangement permits to check the scattering phase matrix for a possible asymmetry in the form

$$\frac{\Delta P}{P} := \frac{P_{R,12} - P_{R,21}}{P_{R,12}} \neq 0 \tag{12.11}$$

where $P_{R,12}$ and $P_{R,21}$ are the respective elements of the Rayleigh phase matrix, defined in (11.5). To simplify the notation we will call them P_1 and P_2 respectively from now on. We also note that for the particular case of 90° scattering $P := P_1 \approx P_2 \approx 1$, one can say that $\Delta P/P = \Delta P$. According to the scattering theory presented in chapter 11 the phase matrix should be perfectly symmetric i.e. $P_1 = P_2$ should hold for all scattering angles, in particular for the 90° angle of our experiment. The measurement of ΔP

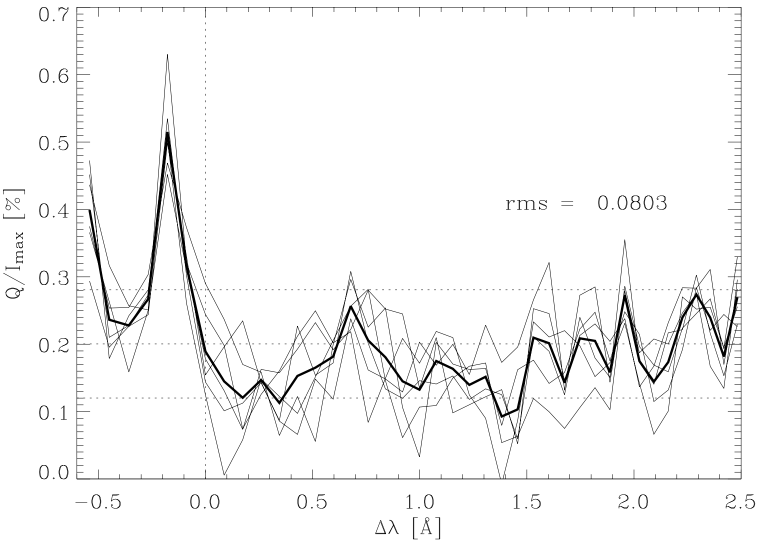


Fig. 12.5: Detail view of the D_1 Q-spectrum from Fig. 12.4. The six different measurement series that are averaged in Fig. 12.4 are plotted individually in this figure, together with the average (bold line). The three horizontal dotted lines indicate the spectral average ± 1 rms.

in the D_2 line, together with the measurement of the polarizability W_2 in the center of D_1 line are two important tests of the current theory. In the following paragraph we will explain the relation between ΔP and the two PEM measurements.

By combining the PEM1 and PEM2 measurements we can find the value of the asymmetry ΔP as a function of $\Delta q = q_2 - q_1$.

$$\begin{aligned} \Delta P &= \frac{(4 + P_1 - W_2)(4 + P_2 - W_2)}{(4 + P_1 + P_2 + 2W_2)} \Delta q \\ &\approx 2\frac{\Delta q}{q} \end{aligned} \tag{12.12}$$

To estimate the proportionality factor, which is important to define the uncertainty in Δp from the measurement accuracy, we have used the fact that $q := q_1 \approx q_2 \ll 1$ and $W_2 \approx 2q/3$, according to (11.11).

From the measurements we find

$$\Delta P = (-2 \pm 3)\%$$

which is consistent with the current theory.

Chapter 13

Conclusions and outlook

Solar observations of the unexpected polarization peaks in the cores of the D_1-type lines Na I 5896 Å and Ba II 4934 Å have brought our attention to a possible deficiency in the quantum mechanical theory of these transitions. With our laboratory experiment for 90° scattering at sodium atoms we set out to answer the question whether the observed solar anomaly is a problem of atomic or of solar physics.

The laboratory experiment allows us to eliminate many complicating factors that affect the solar observations. In the laboratory we have optically thin scattering with a specified geometry, the surrounding magnetic field is known and under control, and the polarization state of the exciting radiation can be chosen. Moreover, optical pumping can be ruled out.

In chapter 11 we have given an overview of the quantum mechanical model for D_1 and we have shown why this line is expected to be unpolarized. The laboratory result turns out to be consistent with the theoretical null prediction. The upper limit for the polarization ratio between the D_1 and D_2 lines, defined by the measurement uncertainty, is found to be about 2%, which is significantly smaller than the ratio observed on the Sun. Although this result does not completely answer the initial question it permits to conclude that the D_1 anomaly is at least to a certain degree a problem of solar physics.

The definitive answer probably involves magnetic fields. Recent (and yet unpublished) measurements with ZIMPOL show spatial variations of the D_1 line core polarization. In particular the latest observations of the K I D_1 line (7699 Å), another transition with the same quantum numbers than sodium D_1, show substantial spatial variations with clear signatures of the transverse Zeeman effect. This remarkably illustrates the ubiquity of solar magnetic fields and their possible effects in the context of the D_1 anomaly. Stenflo (2006) notes that the polarization in the K I D_1 line might be directly related to the transverse Zeeman effect. However, at the same time the author states that the Zeeman effect on its own is probably not sufficient to explain the signature seen in the broader sodium D_1 line since the required vertical field strengths in the order of 100 Gauss would not be reached in the lower chromosphere where sodium D_1 is formed. It becomes thus apparent that the D_1 anomaly probably reflects a certain effect of the Sun's magnetic field that is not yet identified.

Concerning the neighboring sodium D_2 line, we have obtained an interesting side result which also supports the validity of the current theory of atomic transitions. We can experimentally prove that (as theoretically expected) the scattering phase matrix is symmetric, at least as far as the measurement uncertainty of 3%.

In the meantime the sodium experiment has been replaced by a second generation experiment with K I (Thalmann et al., 2006) where, in contrast to sodium, a tunable solid-state laser is available as a light source. The laser provides a much higher S/N ratio than the sodium vapor lamp and it permits to resolve the spectral line profile at the excitation level so that no spectrograph is needed. The laser power range permits to work either in a regime with or without optical pumping. Moreover, the scattering can be studied under the influence of a magnetic field up to 30 G, provided by a pair of Helmholtz coils. This decisive improvements compared to the first generation experiment allow to further explore the D_1 anomaly from the viewpoint of atomic physics.

Bibliography

Berman, P. R., Lamb, W. E., 1969, *Influence of Resonant and Foreign Gas Collisions on Line Shapes*, Physical Review **187**, 221

Bernasconi, P., 1997, *Stokes Vector Polarimetry: Observation and Analysis of Solar Magnetic Fields*, Ph.D. thesis, ETH Zürich

Bommier, V., 1997a, *Master equation theory applied to the redistribution of polarized radiation, in the weak radiation field limit. I. Zero magnetic field case*, A&A **328**, 706

Bommier, V., 1997b, *Master equation theory applied to the redistribution of polarized radiation, in the weak radiation field limit. II. Arbitrary magnetic field case*, A&A **328**, 726

Cacciani, A., Moretti, P.-F., 1994, *Magneto-optical filter: concept and applications in astronomy*, in D. L. Crawford, E. R. Craine (eds.), *Proc. SPIE Vol. 2198, p. 219-228, Instrumentation in Astronomy VIII, David L. Crawford; Eric R. Craine; Eds.*, vol. 2198 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, 219–228

Cacciani, A., Moretti, P. F., Rodgers, W. E., 1997, *Measuring Doppler and Magnetic Fields Simultaneously*, Solar Phys. **174**, 115

Cimino, M., Cacciani, A., Sopranzi, N., 1968, *An Instrument to measure Solar Magnetic Fields by an Atomic-Beam Method*, Solar Phys. **3**, 618

Degl'Innocenti, E. L., 1998, *Evidence against turbulent and canopy-like magnetic fields in the solar chromosphere*, Nature **392**, 256

Degl'Innocenti, E. L., 1999, *Evidence for ground-level atomic polarization in the solar atmosphere*, in K. N. Nagendra, J. O. Stenflo (eds.), *ASSL Vol. 243: Polarization*, 61–71

Fluri, D. M., Holzreuter, R., Klement, J., Stenflo, J. O., 2003, *Radiative Transfer in Na I D_2 and D_1*, in J. Trujillo-Bueno, J. Sanchez Almeida (eds.), *Astronomical Society of the Pacific Conference Series*, 263–+

Gandorfer, A., 1996, *ZIMPOL - An Intermediate Step between ZIMPOL I and ZIMPOL II*, Master's thesis, University of Würzburg

Gandorfer, A., 2000, *The Second Solar Spectrum: A high spectral resolution polari-*

metric survey of scattering polarization at the solar limb in graphical representation. Volume II: 4625 Å to 6995 Å, VdF, Zurich

Gandorfer, A., 2002, *The Second Solar Spectrum: A high spectral resolution polarimetric survey of scattering polarization at the solar limb in graphical representation. Volume II: 3910 Å to 4630 Å*, VdF, Zurich

Gandorfer, A., 2004, *The Second Solar Spectrum, Vol. III: 3160 Å to 3915 Å*, VdF, Zurich

Gandorfer, A. M., Steiner, H. P., Povel, P., Aebersold, F., Egger, U., Feller, A., Gisler, D., Hagenbuch, S., Stenflo, J. O., 2004, *Solar polarimetry in the near UV with the Zurich Imaging Polarimeter ZIMPOL II*, A&A **422**, 703

Ivanov, V. V., 1991, *Analytical Methods of Line Formation Theory - are they Still Alive*, in L. Crivellari, I. Hubeny, D. G. Hummer (eds.), *NATO ASIC Proc. 341: Stellar Atmospheres - Beyond Classical Models*, 81–+

Kerkeni, B., Bommier, V., 2002, *Theoretical study of the collisional depolarization and of the Hanle effect in the Na I D_2 line observed on the solar limb*, A&A **394**, 707

Klement, J., Stenflo, J. O., 2003, *Influence of Collisions and Magnetic Fields on the Polarization of the Na I D_1 and D_2 Lines*, in J. Trujillo-Bueno, J. Sanchez Almeida (eds.), *Astronomical Society of the Pacific Conference Series*, 278–+

Landi Degl'Innocenti, E., 1984, *Polarization in spectral lines. III - Resonance polarization in the non-magnetic, collisionless regime*, Solar Phys. **91**, 1

Landi degl'Innocenti, E., 1996, *The Density Matrix Approach to Polarized Radiative Transfer*, Solar Phys. **164**, 21

Manso Sainz, R., Trujillo Bueno, J., 2001, *Modeling the Scattering Line Polarization of the Ca II Infrared Triplet*, in M. Sigwarth (ed.), *ASP Conf. Ser. 236: Advanced Solar Polarimetry – Theory, Observation, and Instrumentation*, 213–+

Manso Sainz, R., Trujillo Bueno, J., 2003, *Zero-Field Dichroism in the Solar Chromosphere*, Physical Review Letters **91**, no. 11, 111102

Mihalas, D., 1978, *Stellar Atmospheres*, Freeman and Company, San Francisco

Povel, H., 1995, *Imaging Stokes polarimetry with piezoelastic modulators and charge-coupled-device image sensors*, Opt. Eng. **34**, 1870

Povel, H., 2001, *Ground-based instrumentation for solar magnetic field studies, with special emphasis on the Zurich Imaging Polarimeters ZIMPOL-I and II*, in M. Mathys, S. K. Solanki, D. T. Wickramasinghe (eds.), *Magnetic Fields Across the Hertzsprung-Russel Diagram*, vol. 248 of *ASP Conf. Series*, 543, ASP, San Francisco

Rees, D. E., Saliba, G. J., 1982, *Non-LTE resonance line polarization with partial* *redistribution effects*, A&A **115**, 1

Saliba, G. J., 1985, *Non-LTE resonance line polarization with partial redistribution - The solar CA II K line*, Solar Phys. **98**, 1

Stenflo, J. O., 1974, *Observations of Resonance Polarization in Ca I 4227 Å*, Solar Phys. **37**, 31

Stenflo, J. O., 1980, *Resonance-line polarization. V - Quantum-mechanical interference between states of different total angular momentum*, A&A **84**, 68

Stenflo, J. O., 1994, *Solar magnetic fields: polarized radiation diagnostics*, Astrophysics and Space Science Library, Dordrecht; Boston: Kluwer Academic Publishers

Stenflo, J. O., 1997, *Quantum interferences, hyperfine structure, and Raman scattering on the Sun.*, A&A **324**, 344

Stenflo, J. O., 1998, *Hanle-Zeeman scattering matrix*, A&A **338**, 301

Stenflo, J. O., 2003a, *Imaging polarimetry: opportunities and limitations*, in S. Fineschi (ed.), *Polarimetry in Astronomy. Edited by Silvano Fineschi . Proceedings of the SPIE, Volume 4843, pp. 76-88 (2003).*, 76–88

Stenflo, J. O., 2003b, *Scattering Polarization in Magnetic Fields: Anomalies, Surprises and Enigmas*, in J. Trujillo-Bueno, J. Sanchez Almeida (eds.), *Astronomical Society of the Pacific Conference Series*, 385–+

Stenflo, J. O., 2004, *The new world of scattering physics seen by high-precision imaging polarimetry*, Rev. Modern Astron. **17**, 269

Stenflo, J. O., 2006, *Second Solar Spectrum: A Brief Overview*, in R. Casini, B. W. Lites (eds.), *Astronomical Society of the Pacific Conference Series*, vol. 358 of *Astronomical Society of the Pacific Conference Series*, 215–+

Stenflo, J. O., Baur, T. G., Elmore, D. F., 1980, *Resonance-line polarization. IV - Observations of non-magnetic line polarization and its center-to-limb variations*, A&A **84**, 60

Stenflo, J. O., Gandorfer, A., Keller, C. U., 2000a, *Center-to-limb variation of the enigmatic Na I D_1 and D_2 polarization profiles*, A&A **355**, 781

Stenflo, J. O., Gandorfer, A., Wenzler, T., Keller, C. U., 2001, *Influence of magnetic fields on the coherence effects in the Na I D_1 and D_2 lines*, A&A **367**, 1033

Stenflo, J. O., Keller, C. U., 1996, *New window for spectroscopy*, Nature **382**, 588

Stenflo, J. O., Keller, C. U., 1997, *The second solar spectrum. A new window for diagnostics of the Sun.*, A&A **321**, 927

Stenflo, J. O., Keller, C. U., Gandorfer, A., 1998, *Differential Hanle effect and the spatial variation of turbulent magnetic fields on the Sun*, A&A **329**, 319

Stenflo, J. O., Keller, C. U., Gandorfer, A., 2000b, *Anomalous polarization effects due to coherent scattering on the Sun*, A&A **355**, 789

Thalmann, C., Stenflo, J. O., Feller, A., Cacciani, A., 2006, *Laboratory Experiment for Polarized Scattering at Potassium Vapor*, in R. Casini, B. W. Lites (eds.), *Astronomical Society of the Pacific Conference Series*, vol. 358 of *Astronomical Society of the Pacific Conference Series*, 323–+

Tomczyk, S., Streander, K., Card, G., Elmore, D., Hull, H., Cacciani, A., 1995, *An Instrument to Observe Low-Degree Solar Oscillations*, Solar Phys. **159**, 1

Trujillo Bueno, J., 2001, *Atomic Polarization and the Hanle Effect*, in M. Sigwarth (ed.), *ASP Conf. Ser. 236: Advanced Solar Polarimetry – Theory, Observation, and Instrumentation*, 161–+

Trujillo Bueno, J., 2003, *New Diagnostic Windows on the Weak Magnetism of the Solar Atmosphere*, in J. Trujillo-Bueno, J. Sanchez Almeida (eds.), *Astronomical Society of the Pacific Conference Series*, vol. 307, 407–+

Trujillo Bueno, J., Casini, R., Landolfi, M., Landi Degl'Innocenti, E., 2002, *The Physical Origin of the Scattering Polarization of the Na I D Lines in the Presence of Weak Magnetic Fields*, ApJ lett. **566**, L53

Trujillo-Bueno, J., Sanchez Almeida, J. (eds.), 2003, *Solar Polarization*, vol. 307

Volz, U., Majerus, M., Liebel, H., Schmitt, A., Schmoranzer, H., 1996, *Precision Lifetime Measurements on NaI $3p^2P_{1/2}$ and $3p^2P_{3/2}$ by Beam-Gas-Laser Spectroscopy*, Physical Review Letters **76**, 2862

Part III

Polarimetry of the solar flash spectrum

Chapter 14

Introduction

The chromospheric emission spectrum, called flash spectrum, has already been recorded successfully during multiple eclipse expeditions (Cillié & Menzel, 1935; Mitchell, 1947; Dunn et al., 1968; Shen et al., 1981). The mentioned authors have compiled catalogues itemizing the calibrated absolute fluxes of more than a thousand emission lines in the range 300 to 910 nm. The spectra of Dunn et al. (1968) cover 4500 km on the Sun with a height resolution of 100 km in the direction of lunar movement. Although these data lack polarization information, they have been very helpful for the design of our instrument, allowing us to narrow down the required sensitivity range.

In terms of polarimetry we know of two past eclipse observations. The measurements in the Ca II H and K lines of Hanson et al. (1976) allowed a 5% upper limit to be set on the K line degree of polarization. Moreover, from his visual tracking of the intensity variations produced by a rotating polaroid screen, Stokley (1948) estimated that the overall degree of polarization of the flash spectrum is probably not above 10%. This result is however questionable because of the rapid exponential drop of the line intensities during the flash phase.

The flash spectrum is extremely difficult to observe outside an eclipse. The steep intensity decrease at the extreme limb of the Sun makes a coronagraph very vulnerable to stray light caused by the earth's atmosphere and by the optics of the instrument itself. The main challenge here is the occulting disc. It must not exceed the solar disc by more than a fraction of an arcsecond, to properly isolate the thin chromospheric layer from the photosphere while still leaving it uncovered. Another serious drawback of out-of-eclipse observations is the spatial resolution of current solar telescopes, which is limited to about 100 km. In contrast, as the eclipsing Moon is moving at a relative speed of about 350 km/s, a height resolution of order 10 km on the Sun can already be achieved with a moderate frame rate of 50 frames/s.

When using the Moon as occulting disc one has to take into account that the lunar limb is somewhat serrated. For the 29 March 2006 eclipse, the lunar limb profile at the positions of second and third contact (Espenak & Anderson, 2004) shows typical height variations corresponding to about 700 km on the Sun. On the other hand, the lunar mountains and valleys can be easily resolved with a modest spatial resolution of about 5 arcsec in the direction parallel to the Moon's limb, which allows for a later correction of the chromospheric height scale across the field of view.

The scientific rationale of our observations is described in detail by Stenflo (2007). A theoretical reference has been developed by Chandrasekhar (1950) in terms of an idealized model for a plane-parallel purely scattering atmosphere. The conditions at the extreme limb of the Sun are approaching this Chandrasekhar limit, but the real polarization can nevertheless be significantly influenced by different physical processes: the relative importance of different opacities, collisions, deviation from plane-parallel stratification (spherical geometry at the limb, small scale inhomogeneities), atomic physics (quantum interference, optical pumping), radiative-transfer physics and magnetic fields (Hanle effect). These processes affect the individual spectral lines in different ways. By recording a large part of the spectrum, we may be able to untangle and quantify them with the help of differential diagnostics.

Chapter 15

Instrument design

15.1 Photon budget

When designing a polarimeter one has to make a trade-off between spatial, spectral, and time resolution and polarimetric sensitivity (Tab. 15.1). The chromospheric sickle is not resolved in practice in the direction of lunar movement, because of the steep intensity drop with height (see Fig. 15.1). The spectral line profile is not resolved either. In this context it is convenient to define, for a given line i, an integrated flux F_i given by

$$F_i(h) = \int_h^\infty \int_{-1/2}^{1/2} \int_{\lambda_i-\Delta\lambda_i/2}^{\lambda_i+\Delta\lambda_i/2} f(\lambda)\, d\lambda\, dy\, dx \tag{15.1}$$

Thus $F_i(h)$ is the total number of photons per second and sterad, emitted from a one unit thick semi-infinite slab of the chromosphere at height h (see Fig. 15.2) and within a spectral window $\Delta\lambda_i$ around the line center λ_i.

Given the flux F_i, we can then determine the total number of photoelectrons generated in a camera pixel.

$$n_i(h) = F_i(h)\, \Delta\Omega\, \Delta y\ \Delta t\, \epsilon_{\rm opt}\, q_{\rm eff}(\lambda_i) \tag{15.2}$$

$\Delta\Omega$ is the solid angle covered by the instrument aperture, as seen from the Sun, Δy is the extent along the chromospheric sickle imaged by the pixel,

spectral range	340 - 870 nm
spectral sampling	0.21 nm/pixel
time resolution	13 - 40 ms
spatial sampling	5 arcsec/pixel
field of view	150 arcsec

Tab. 15.1: Main instrument characteristics.

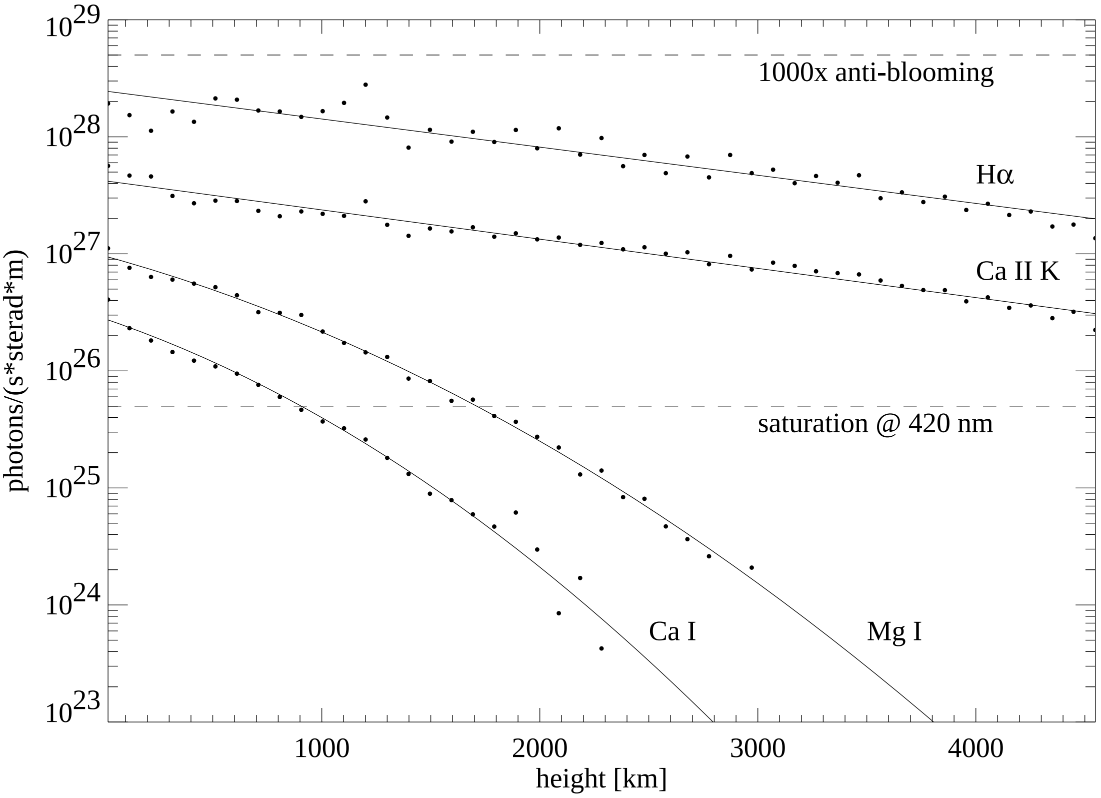


Fig. 15.1: Flux variation with height for some selected emission lines of Tab. 15.2. The data are from Dunn et al. (1968). For our purpose, the logarithm of the flux can be sufficiently well fitted with a linear function for the strong lines and a second order polynomial for the weak lines. The estimated saturation level around Ca I and the corresponding anti-blooming limit are indicated by the dashed lines.

Line	Wavelength nm	flux at h=0† photons $\cdot$ s^{-1} $\cdot$ sterrad^{-1} $\cdot$ m^{-1}
Hα	656.3	$2.0 \cdot 10^{28}$
Ca II	854.2	$9.4 \cdot 10^{27}$
Ca II	866.2	$8.7 \cdot 10^{27}$
Ca II K	393.3	$5.5 \cdot 10^{27}$
Hβ	486.1	$5.4 \cdot 10^{27}$
Ca II H	396.8	$4.8 \cdot 10^{27}$
Ca II	849.8	$4.3 \cdot 10^{27}$
He D$_3$	587.6	$3.5 \cdot 10^{27}$
Mg I	518.4	$1.3 \cdot 10^{27}$
Na I D$_2$	589.0	$8.9 \cdot 10^{26}$
Na I D$_1$	589.6	$8.9 \cdot 10^{26}$
Mg I	516.7	$7.8 \cdot 10^{26}$
Mg I	517.2	$7.8 \cdot 10^{26}$
Ca I	422.7	$4.3 \cdot 10^{26}$

† from Dunn et al. (1968)

Tab. 15.2: Lines of interest sorted by flux

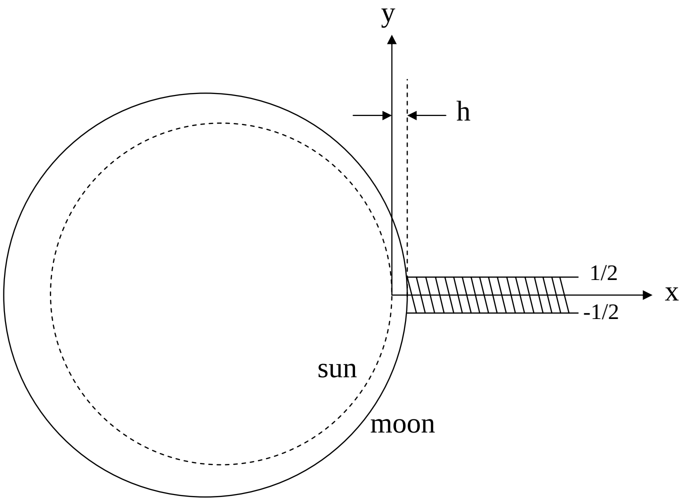


Fig. 15.2: Definition of the integrated flux from a chromospheric emission line

CCD size	4096 pixels × 96 pixels
pixel size	13 μm × 13 μm
dark current at 30 °C	$1.23 \cdot 10^3$ e$^-$/s
readout noise	70 e$^-$
sampling	50 e$^-$/count, 10 bit
readout time	6 ms/frame
frame rate†	1.53 - 75 frames/s
anti-blooming factor	1000

† limited by the mechanical shutter

Tab. 15.3: Camera characteristics

and Δt is the exposure time. ϵ_{opt} refers to the total transmission of the atmosphere and the optics and q_{eff} to the quantum efficiency of the camera. As the product of atmospheric extinction and quantum efficiency is changing by a factor of 5 across the spectral working range, its wavelength dependence must be taken into account. The optics are roughly estimated to contribute to the total efficiency with a factor of 0.1.

The huge dynamic range between Hα and the weak lines (see Fig. 15.1 and Tab. 15.2) cannot be handled by the camera at once. At maximum frame rate the instrument has to be sensitive enough to just saturate in the Ca I line at the beginning of the flash phase. The stronger lines will then drop into the sensitivity range later on. Unfortunately the Hα flux indicated by Dunn et al. (1968) for $h = 0$ is not consistent with the flux at the center of the Hα absorption line at the photospheric limb, which is an order of magnitude lower. This uncertainty is however not critical for the optical design as an instrument that is too fast can easily be dimmed. In addition the flux in the weak lines is decreasing steeply enough to stop saturating the camera after the first few seconds into the flash phase. On the other hand it is important to have a CCD with an anti-blooming feature of at least 1000 × saturation, to avoid any crosstalk from the strong lines during the saturated phase. As we were able to obtain a very fast F/2.8 spectrograph, we can adequately expose the camera with a modest and easily transportable 8-inch telescope (see Fig. 15.3).

The noise in the measured degree of polarization Q/I is, in case of a beam splitting polarimeter,

$$\sigma \approx \sqrt{\frac{1}{n} + \frac{n_{\mathrm{dc}}}{n^2} + \frac{\sigma^2_{\mathrm{readout}}}{n^2}} \quad , \tag{15.3}$$

where n is the number of photoelectrons, n_{dc} the number of electrons generated by the dark current and $\sigma_{\mathrm{readout}}$ the readout noise. All noise sources are assumed to be Poisson distributed. Close to saturation, σ is dominated by the shot noise $n^{-1/2} \approx 0.004$. This meets the requirements on polarimetric precision and is adequately sampled with a 10 bit A/D conversion (*cf.* Tab. 15.3). At less than 0.1 × saturation the readout noise begins to dominate. n_{dc} stays below 0.001 × saturation and it's contribution to the polarimetric noise is negligible.

15.2 Telescope and spectropolarimeter unit

The telescope is an 8-inch Dall-Kirkham Cassegrain reflector. The imaging is nearly diffraction limited within our small field of view and the Cassegrain design is virtually free of instrumental polarization.

A scaled drawing and a close-up photograph of the spectropolarimeter unit are shown in Fig. 15.3. To keep the optical efficiency high and chromatic aberrations at a minimum, the use of glass in the beam path is avoided, except for the half wave plate and the beam splitter. At the telescope focus one can choose between a 50 μm slit of length 1.6 mm and an aperture of 1.6 mm × 1.6 mm. With the slit we record calibration spectra of the solar disk or a spectral lamp. The aperture is used for the actual observation of the flash spectrum. This slitless mode has some important advantages: it allows the imaging of extended solar structures and relaxes the tolerances on guiding and pointing of the telescope. Then again we observed disadvantages like aberrations caused by an off-axis position of the sickle or the smearing of nearby spectral lines.

In terms of polarimetry we opt for a beam splitter, as opposed to the more accurate modulation/demodulation principle normally used in our Zurich Imaging Polarimeter (Povel, 1995; Gandorfer et al., 2004). The arguments in favor of a beam splitter are the intricacies involved in developing an adequate achromatic modulator, and the low frame rate of our ZIMPOL cameras ($\sim$ 1 frame/s). Therefore a Savart beam splitter is used. It is installed in the slower F/11.5 beam to keep its aberrations below the size of a pixel. The disadvantage of this position is the need of rather thick calcite elements. We have 2 × 22.8 mm, giving us a beam separation between 3.40 and 3.76 mm, depending on wavelength. To eliminate the effect of the polarization dependent grating efficiency, the beam splitter has to be turned so that the polarization direction of both beams is oriented ±45° to the grooves of the

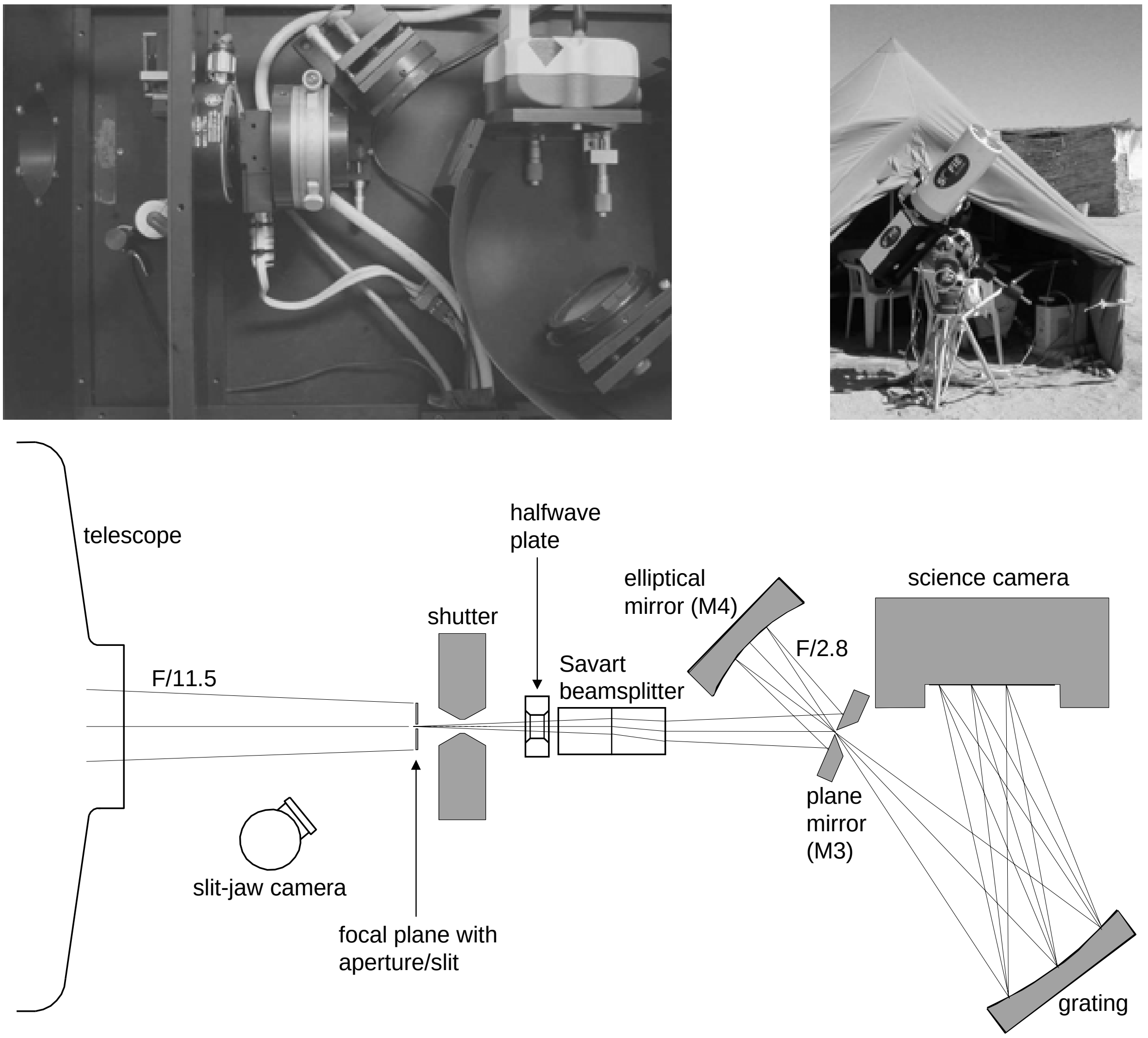

telescope
halfwave plate
elliptical mirror (M4)
science camera
shutter
F/2.8
Savart beamsplitter
F/11.5
plane mirror (M3)
slit-jaw camera
focal plane with aperture/slit
grating

Fig. 15.3: Bottom: optical setup. **Top left:** top view of the spectropolarimeter unit. **Top right:** overview of the setup at the eclipse site near Waw an Namos, Libya. The spectropolarimeter unit is mounted behind the telescope. In the background one can see the tent which is used as work place and protection for the electronic equipment. The electrical power of the whole setup (150-200 W) is provided by fuel generators and secured with a UPS.

grating. A half wave plate is needed in this context to realign the Moon's limb with the direction of Stokes $+Q$. We renounce the use of true beam exchange again because of the intricacies of getting an adequate fast-switchable and achromatic half wave plate. Instead the retarder is turned mechanically by 45° one single time between the two flash phases of second and third contact.

After the Savart plate the two beams are deflected by a plane mirror (M3) with a circular hole of 2.5 mm diameter, reduced by an elliptical mirror (M4) to F/2.8 and then focused through the hole in M3 into the spectrograph. This configuration allows to use M4 on-axis, and the hole only causes a small loss of about 5% of the flux. The focal points of M4 slightly deviate from their paraxial positions (telescope focus and intermediary focus at M3) to minimize the differential aberrations between cospatial image points in the two beams.

The spectrograph consists of an aberration corrected holographic concave grating with 405 grooves/mm, serving as both the dispersive and focussing element. Operated in first order it provides a practically linear dispersion of 0.21 nm/pixel over the whole wavelength range. The efficiency varies between about 45% (393 nm) and 23% (866 nm). We find however a significant residual astigmatism, clearly dominating the total aberrations of the instrument. As a trade-off we adjust the grating and camera position to align the detector plane as good as possible with the spectral focal plane, but accepting therefore a strong wavelength dependent spatial defocussing.

Both the telescope and the spectropolarimeter unit are assembled on a parallactic mounting with computer-controlled stepping motors on both axes (*cf.* section 15.3). The guiding and pointing accuracies in the hour axis are limited by the worm gear which has a period of about 12 minutes at guiding speed and a mean amplitude of some 40 arcsec. Due to their periodicity, the hour axis errors can at least be partly compensated with the guiding software. The guiding and pointing accuracies in the declination axis are of order 1 arcsec/minute and 10 arcsec respectively and can be neglected for our purpose. The knowledge of the pointing accuracies is crucial for the eclipse observations as the telescope has to be moved blindly to the opposite limb during totality (*cf.* section 15.3). It defines the minimum size of the aperture in the spectropolarimeter unit.

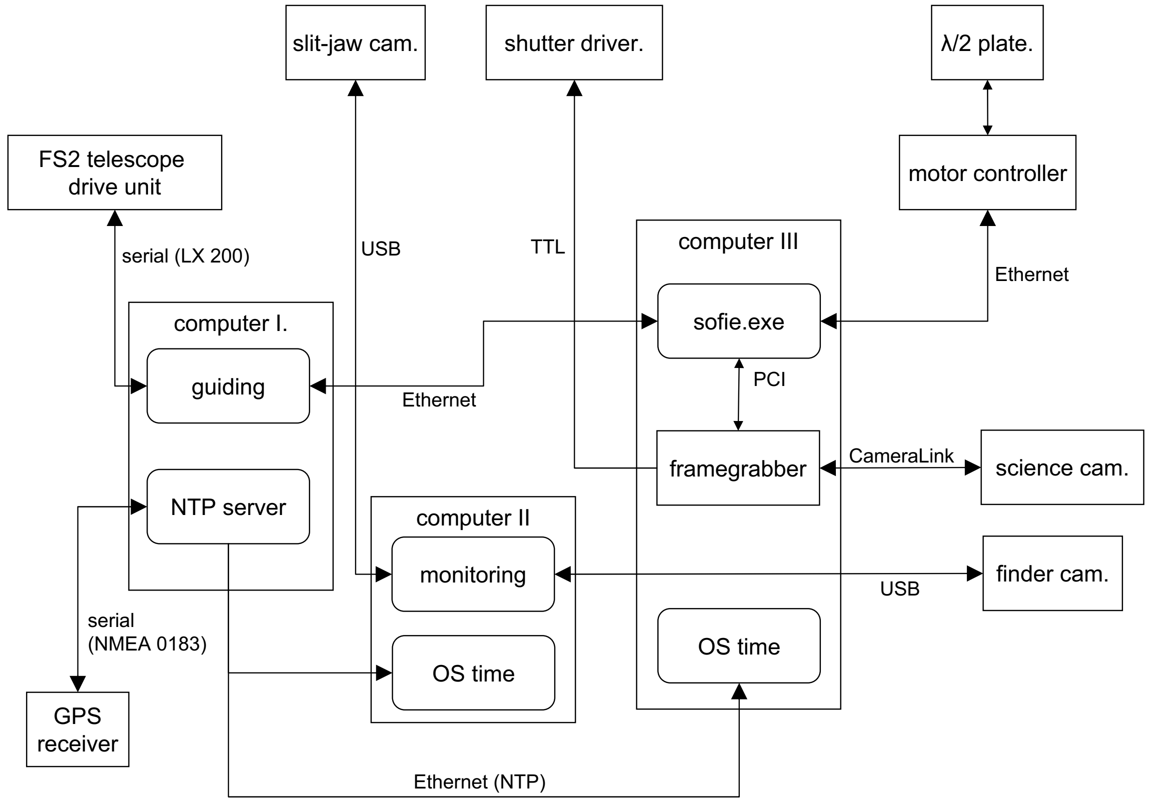

slit-jaw cam.
shutter driver.
λ/2 plate.
FS2 telescope drive unit
motor controller
serial (LX 200)
USB
TTL
computer III
Ethernet
computer I.
sofie.exe
guiding
PCI
Ethernet
framegrabber
CameraLink
science cam.
NTP server
computer II
monitoring
finder cam.
USB
serial (NMEA 0183)
OS time
OS time
GPS receiver
Ethernet (NTP)

Fig. 15.4: Control units and programs for the eclipse instrument. The boxes with the soft edges denote computer programs, the boxes with sharp edges hardware devices. The lines of communication are annotated with the connection type and/or the protocol.

15.3 Instrument control and observing procedure

Figure 15.4 shows a block diagram of the different control units and programs. The control tasks are allocated to three different computers: Computer I is responsible for the guiding and acts as time server, computer II is used for monitoring the solar disc, and computer III is handling the spectropolarimeter unit.

The telescope mounting is aligned by star tracking according to the Bigourdan procedure. For accurate pointing on the Sun a special guiding software has been developed in-house which uses the solar limb as reference frame. The pointing position is determined by scanning the solar disc in the

Latitude	24:28:03 N
Longitude	17:57:52 E
Altitude	450 m

Tab. 15.4: Observing location

two axis directions with the spectropolarimeter in slit-mode, and determining the limb position by the corresponding inflection point in the measured intensity.

The time server is providing UTC with an accuracy of about 10 ms, which corresponds to the time resolution of the instrument.

The science camera and shutter as well as the rotation of the half wave plate are controlled with software developed in-house and optimized for the particular requirements of the eclipse observation. A critical item is the reliable handling of the data stream, which is reaching 56 MB/s at maximum frame rate. Other important features are the live display of the science camera images and the possibility to quickly adjust the frame rate during operation.

Tab. 15.5 summarizes all the relevant observing and calibration procedures together with the respective timings. Our observing spot is located near Waw an Namos in the Sahara of southern Libya (*cf.* Tab. 15.4), on the center line of the eclipse path and close to the point of maximum duration of totality. The center-line position is necessary to avoid sickle movements during the flash phases, while the long duration of totality is of advantage for extending the duration of the flash phases.

UTC	Events
	Limb scan to set the guiding reference frame Position the slit at the predicted position of first contact (angle 135.2°)
08:55:54.67	First contact Verify the positioning accuracy and the orientation of the spectropolarimeter unit
	Move to the predicted position of second contact (angle -44.7°) Limb scan including the lunar limb to correct the guiding reference frame Change from slit to aperture
10:13:30	Open the telescope Start recording with a fixed rate of 25 frames/s
10:13:45.20	Second contact
10:14:34	Slow down to 1.5 frames/s for the coronal lines
	Move to predicted position of third contact (angle 136.9°) Rotate half wave plate by 45° for beam exchange
10:17:14	Start with variable frame rate program: the frame rates are incremented from 25 to 75 frames/s in steps of 25 frames/s, based upon a visual rating of the flash intensity
10:17:55.74	Third contact
10:18:16	Close telescope Stop recording
	Change back to slit Polarimetric calibration at disc center Dark current/bias and flat field measurements Spectral calibration with Rb I lamp

Tab. 15.5: Observing procedure

Chapter 16

Conclusions and outlook

The instrument described here has been developed and built from scratch within one year for the 29 March 2006 eclipse. The observing location in the desert, hundreds of km away from any infrastructure, called for robust, compact, reliable and completely autonomous equipment as well as elaborate logistics.

Despite all these challenges the equipment was running faultlessly. The eclipse was passing under ideal weather conditions and we were able to measure both flash phases of second and third contacts, on opposite solar limb positions (*cf.* Tab. 15.5).

Our requirements were however not entirely satisfied. The main drawback of the instrument is the residual aberrations caused by the grating, in particular the differential effects between the two beams (*cf.* Fig. 16.1). They have been discovered only during the tests of the final instrument, and we were not able to fully resolve them in due time. On the other hand the slitless mode has proven to be an advantage for isolated spectral lines like Hα, allowing to image extended structures like a prominence, but a disadvantage for adjacent lines like He D_3, Na D_1 and D_2.

The data reduction is still ongoing, and the results as well as the reduction techniques will be discussed in detail in a later publication. The main challenge is to study and model the aberrations in order to determine the cospatial regions in both images and to improve the spatial resolution in some parts of the spectrum.

We are already thinking of incorporating our experiences into a second generation eclipse experiment. The key enhancements we will be working on are the replacement of the slitless mode with an array of optical fibers, the use of an Echelle grating, and a true achromatic beam exchange.

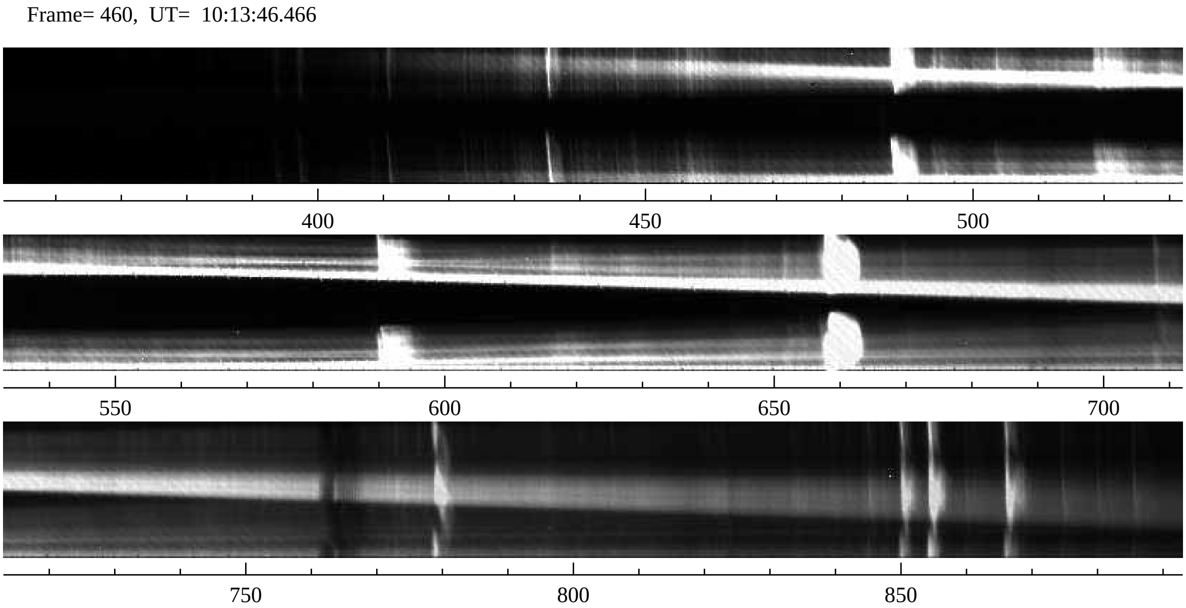


Fig. 16.1: Sample frame from the flash phase at second contact, representing the raw data (only corrected for bias and flat field). The scale gives the wavelength in nm. The two orthogonally polarized spectra are imaged above each other. Since the separation between them is strongly wavelength dependent, they are well separated in the UV (so well that vignetting is severe), while they significantly overlap in the infrared. While some strong lines, like Hβ, the Helium D3 lines, and Hα, are still saturated in this particular frame, the Ca II H and K lines in the UV are already quite faint. The almost horizontal bright band represents the remaining part of the photospheric sickle. In this frame one can discern the main instrumental problems: vignetting, overlap, and aberrations, in particular the differential effects between the two spectra.

Bibliography

Chandrasekhar, S., 1950, *Radiative Transfer*, Clarendon Press, Oxford

Cillié, G. G., Menzel, D. H., 1935, *The Physical State of the Solar Chromosphere.*, Harvard College Observatory Circular **410**, 1

Dunn, R. B., Evans, J. W., Jefferies, J. T., Orrall, F. Q., White, O. R., Zirker, J. B., 1968, *The Chromospheric Spectrum at the 1962 Eclipse*, ApJ suppl. ser. **15**, 275

Espenak, F., Anderson, J., 2004, *Total solar eclipse of 2006 March 29*, Total solar eclipse of 2006 March 29, by F. Espenak and J. Anderson. Washington, DC: NASA, 1993 - 2004

Gandorfer, A. M., Steiner, H. P., Povel, P., Aebersold, F., Egger, U., Feller, A., Gisler, D., Hagenbuch, S., Stenflo, J. O., 2004, *Solar polarimetry in the near UV with the Zurich Imaging Polarimeter ZIMPOL II*, A&A **422**, 703

Hanson, T., Stenholm, B., Söderhjelm, S., 1976, *An attempt to measure resonance polarization in the June 30, 1973 flash spectrum.*, Reports of the Lund Observatory **8**

Mitchell, S. A., 1947, *Chromospheric Spectrum from Ten Eclipse Expeditions.*, ApJ **105**, 1

Povel, H., 1995, *Imaging Stokes polarimetry with piezoelastic modulators and charge-coupled-device image sensors*, Opt. Eng. **34**, 1870

Shen, L.-X., Li, Q.-D., Yu, J.-Q., Sun, R.-Q., 1981, *Observation and preliminary analysis of the chromospheric flash spectrum obtained during the total solar eclipse of 1980 February 16*, Chinese Astronomy and Astrophysics **5**, 388

Stenflo, J. O., 2007, *Polarization at the extreme limb of the Sun and the role of eclipse observations*, in R. Ramelli, O. Shalabiea, I. M. Saleh, J. O. Stenflo (eds.), *Proc. Intern. Symp. on Solar Physics and Solar Eclipses*, 1–14, Sebha University publications, Sebha, Libya

Stokley, J., 1948, *Polarization of the flash spectrum.*, Astron. J. **53**, 117

Acknowledgments

This work has been financially supported by ETH Zürich and the Swiss National Science Foundation, which is gratefully acknowledged. I am also grateful to the Fonds National de la Recherche Luxembourg for financially supporting the publication of this thesis.

Curriculum Vitae

of Alex Jean Feller, born September 18, 1977 in Luxembourg

since 2003	Research assistant at the Institute of Astronomy, ETH Zürich
2002	Diploma thesis at the Institute of Astronomy, ETH Zürich
1997 – 2002	Studies in Physics at ETH Zürich
June 1997	Diplôme de fin d'études secondaires
1991 – 1997	High school in Luxembourg
1985 – 1991	Primary school in Luxembourg