Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Dipl.-Phys. Julian Stürmer

Born in Aschaffenburg

Oral examination: 29.01.2015
Technologies for improving the precision of radial velocity measurements with high-resolution, fibre-fed échelle spectrographs

Referees:

Prof. Dr. Andreas Quirrenbach

Apl. Prof. Dr. Reinhard Mundt
Abstract

Technologies for improving the precision of radial velocity measurements with high-resolution, fibre-fed échelle spectrographs

The objective of the dissertation is to investigate instrumental limitations of high-precision radial velocity measurements with fibre-fed échelle spectrographs and to develop technologies to overcome these limitations. It is composed of two parts:

The first part investigates the optimal use of optical fibres to reduce the sensitivity of spectrographs to unstable coupling conditions at a telescope. In the framework of CARMENES, a high-resolution spectrograph project, optical properties of appropriate fibres are investigated with the aim of determining the optimal solution for the particular requirements of this instrument. Scrambling and focal ratio degradation behaviour of non-circular fibres are measured and optical simulations of the stability performance of CARMENES are presented. The results provide a strong indication that a 1 m/s radial velocity measurement precision, which is the goal of CARMENES, can be reached by combining an octagonal with a circular fibre.

The second part describes a novel method of stabilizing a broadband Fabry-Pérot etalon for spectrograph calibration. By using Doppler-free saturated absorption spectroscopy, a single etalon transmission peak is locked to hyperfine transitions of rubidium. For reasonable astronomical integration times, a locking precision of 3 cm/s is demonstrated.
Zusammenfassung

Techniken zur Verbesserung der Präzision von Radialgeschwindigkeitsmessungen mit hochauflösenden, fasergekoppelten Échelle Spektrographen

Zielsetzung der Dissertation ist es, instrumentelle Limitierungen von hoch-präzisen Radialgeschwindigkeitsmessungen mit fasergekoppelten Échelle Spektrographen zu untersuchen und Techniken zu entwickeln, um diese zu überwinden. Die Arbeit besteht aus zwei Teilen:


Der zweite Teil beschreibt eine neuartige Methode zur Stabilisierung eines breitbandigen Fabry-Perot Etalons zum Zwecke einer Spektrographenkalibration. Mit Hilfe von Doppler-freier Absorptionsspektroskopie wird eine einzelne Etalon Transmissionslinie bezüglich der Hyperfeinübergänge von Rubidium stabilisiert. Für typische astronomische Integrationszeiten kann hiermit eine Präzision der Stabilisierung des Etalons von $3 \text{ cm/s}$ erreicht werden.
### Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADEV</td>
<td>Allan deviation</td>
</tr>
<tr>
<td>APC</td>
<td>angled physical contact</td>
</tr>
<tr>
<td>CFP</td>
<td>confocal Fabry-Pérot</td>
</tr>
<tr>
<td>CTE</td>
<td>coefficient of thermal expansion</td>
</tr>
<tr>
<td>CU</td>
<td>calibration unit</td>
</tr>
<tr>
<td>ECDL</td>
<td>external cavity diode laser</td>
</tr>
<tr>
<td>EE</td>
<td>encircled energy</td>
</tr>
<tr>
<td>FE</td>
<td>front-end</td>
</tr>
<tr>
<td>FF</td>
<td>far field</td>
</tr>
<tr>
<td>FFP</td>
<td>fibre Fabry-Pérot</td>
</tr>
<tr>
<td>FoV</td>
<td>field of view</td>
</tr>
<tr>
<td>FPE</td>
<td>Fabry-Pérot etalon</td>
</tr>
<tr>
<td>FRD</td>
<td>focal ratio degradation</td>
</tr>
<tr>
<td>FSR</td>
<td>free spectral range</td>
</tr>
<tr>
<td>FWHM</td>
<td>full width half maximum</td>
</tr>
<tr>
<td>GIF</td>
<td>gradient index fibre</td>
</tr>
<tr>
<td>HCL</td>
<td>hollow-cathode lamp</td>
</tr>
<tr>
<td>LFC</td>
<td>laser frequency comb</td>
</tr>
<tr>
<td>MGF</td>
<td>mode-hop free</td>
</tr>
<tr>
<td>MM</td>
<td>multi-mode</td>
</tr>
<tr>
<td>NA</td>
<td>numerical aperture</td>
</tr>
<tr>
<td>NF</td>
<td>near field</td>
</tr>
<tr>
<td>PD</td>
<td>photo diode</td>
</tr>
<tr>
<td>PID</td>
<td>proportional-integral-derivative</td>
</tr>
<tr>
<td>PM</td>
<td>polarisation maintaining</td>
</tr>
<tr>
<td>PSF</td>
<td>point spread function</td>
</tr>
<tr>
<td>RV</td>
<td>radial velocity</td>
</tr>
<tr>
<td>SG</td>
<td>scrambling gain</td>
</tr>
<tr>
<td>SMF</td>
<td>single-mode fibre</td>
</tr>
<tr>
<td>SNR</td>
<td>signal to noise ratio</td>
</tr>
<tr>
<td>ThAr</td>
<td>thorium argon</td>
</tr>
</tbody>
</table>
Contents

1 Introduction 3
  1.1 Exoplanet Detection Methods 4
  1.2 Radial Velocities 7
  1.3 Échelle Spectrographs 9
  1.4 CARMENES 12
  1.5 Objective and Outline 15

2 Optical Fibres in High-Resolution Spectroscopy 19
  2.1 Theoretical Background 19
  2.2 Optical Properties of Fibres 26
    2.2.1 Scrambling 26
    2.2.2 Focal Ratio Degradation 29
  2.3 Requirements for CARMENES 30
  2.4 Experiments 32
    2.4.1 Optical Test Setup 32
    2.4.2 Nearfield Measurement Procedure 34
    2.4.3 Farfield Measurement Procedure 36
    2.4.4 Results 39
  2.5 Numerical Simulations 48
    2.5.1 Fibre Simulations 48
    2.5.2 Impact of Fibre Illumination Effects on RV Precision 54
    2.5.3 Pupil vs. Image Coupling 64
  2.6 Modal Noise 66
    2.6.1 Background 66
    2.6.2 Experiments 68
  2.7 Dynamical Billiard 72
    2.7.1 Billiard Systems 72
    2.7.2 Circular, Elliptical and Oval Boundaries 76
    2.7.3 Polygonal Boundaries 77
    2.7.4 Boundaries with Chaotic Dynamics 78
  2.8 Fibres for CARMENES 81
    2.8.1 Mechanical Considerations 83
    2.8.2 Fibre Preparation 84
  2.9 Conclusion and Future Work 87
## 3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

### 3.1 Introduction to Wavelength Calibration
- 3.1.1 Emission Lamps
- 3.1.2 Gas Absorption Cell
- 3.1.3 Laser Frequency Combs
- 3.1.4 Etalons

### 3.2 Fabry-Pérot Etalon
- 3.2.1 Theory
- 3.2.2 Types of Etalons
- 3.2.3 Etalon Mirror Coatings
- 3.2.4 Etalon Tuning

### 3.3 Laser Locked Etalon Concept

### 3.4 Experiments
- 3.4.1 Laser
- 3.4.2 Rubidium Doppler Free Spectroscopy
- 3.4.3 Etalons
- 3.4.4 Data Acquisition and Software
- 3.4.5 Locking Stability

### 3.5 Reliability and Systematics

### 3.6 Échelle Measurements

### 3.7 Conclusion and Future Work

## Bibliography
1 Introduction

Two possibilities exist: either we are alone in the universe or we are not. Both are equally terrifying.

Arthur C. Clarke

Are we alone in the universe? Is our solar system unique? Is our Earth and life unique? For the first time in the history of mankind, these fundamental questions can be scientifically investigated. One of the crowning achievements of modern astronomy is the discovery of the first planet around a Sun-like star by Mayor and Queloz (1995). Together with the earlier discovery of a planetary system around a pulsar by Wolszczan and Frail (1992), these studies initiated an entirely new research field in astronomy. In particular, the exploration of planetary companions around other stars, called extra solar planets or exoplanets, has become a rapidly expanding research area of modern astronomy.

As of today, more than 1500 confirmed exoplanets have been detected and the number of exoplanets has been steadily increasing over the last decades, thus allowing scientists to improve their statistics on the existence of planets.\footnote{Source: \url{http://exoplanets.org/}} Recent publications suggest that there is at least one planet per star in our galaxy (Cassan et al., 2012). In addressing the question of the uniqueness of our Earth, planets around stars within their habitable zone (HZ) are of particular interest. Usually, the HZ is defined as the region around a star in which the surface temperature of a potential planet is such that water would be in liquid form. Current estimates indicate that about one in five Sun-like stars hosts an Earth-sized planet within its HZ (Petigura et al., 2013). Occurrence rates of planets around low mass stars, the most abundant type of stars in our universe, are possibly even higher (Tuomi et al., 2014). Considering the 100 to 300 billion stars in our galaxy, these estimates put into new perspective the unanswered question of whether our Earth is unique.

However, an actual Earth twin, i.e. a planet with environmental conditions comparable to the ones on our planet, has not been detected so far. It still remains
technically extremely challenging, since such a discovery requires the exploration of
the composition of an exoplanet including its atmosphere. Even so, there exist some
candidates such as Kepler-22b, which is a planet with 2.4 times the radius of Earth
orbiting within the HZ of a solar type star (Borucki et al., 2012). Up to now, little
is known about its mass, its composition, and its potential atmosphere. However,
some ongoing studies already probe the atmospheres of individual exoplanets and
first claims of detection of water in atmospheres have been made, see e.g. Fraine
et al. (2014).

The enormous progress in exoplanetary research over the last decades has only
been possible due to advances in technology that allowed for suitable high-precision
measurements. This development is likely to even accelerate in the foreseeable fu-
ture, considering the numerous facilities that are either currently being built or
planned, such as ground based spectrograph projects (see Section 1.3 and Table 1.1
in particular) as well as space missions like PLATO and the James Webb Space
Telescope. These projects concentrate on further refining the statistics on exoplanet
occurrence rates, learning about planet formation and evolution, as well as actually
probing physical and chemical properties of individual planets.

1.1 Exoplanet Detection Methods

In the following, I summarize the most successful methods that are currently be-
ing used for exoplanet detection. A more detailed overview of instrumentation for
exoplanetary research can be found in Pepe et al. (2014a).

Direct Detection

Imaging, or direct detection, is the only technique that collects light from the exo-
planets themselves. This is hindered by the extreme contrast in apparent brightness
of host star and planet and their physical closeness. Only a few planets have been
directly observed and all of them are massive planets located at a great distance
from the host star. Such images are obtained in the infrared, because the contrast
ratio between the star and the planet is smaller at longer wavelength. One example
of a direct image of a planetary system is given in Fig. 1.1, which shows the four
massive planetary companions of HR8799, a young star with about 1.5 solar masses
(Marois et al., 2008, 2010). The stellar light is blocked by a coronagraph.

More recently, the Atacama Large Millimeter/Submillimeter Array (ALMA) has
1.1 Exoplanet Detection Methods

taken an image of a planet-forming disc around the star HL Tauri (Fig. 1.1). The image reveals structures in the accretion disc around the young star, indicating the ongoing formation process of planets within a system. However, only very few systems can be imaged directly and most detections of exoplanets have been made by indirect methods.

Figure 1.1 – Left: Direct image of HR8799 and its four companions (b-e). Source: Marois et al. (2010). Right: The protoplanetary disc around HL Tauri, a young star located approximately 450 light years from earth. Source: http://www.eso.org/public/news/eso1436/

Transiting

The transit technique monitors stars for periodic dips in their light curves that result from partial eclipses of the stellar disc while a planetary body is transiting. These dips can only be detected when the planetary system is observed edge-on from the Earth. Assuming a random orientation of planetary orbits around other stars, the probability of an edge-on orientation for a planet around a Sun-sized star is below 1%. The drop in brightness of a star due to a transiting planet is, depending on the size of both the planet and the star, on the order of a few percent or below. The first planet that has been detected by transiting was HD 209458b (Charbonneau et al., 2000), and follow-up observations allowed for a precise determination of its mass, its density, and even of its actual atmospheric composition. By far the most successful
1 Introduction

project using the transit method for detecting exoplanets was the Kepler mission, that looked for planets around 190000 stars in the constellations Cygnus, Lyra, and Draco. Kepler discovered over 3000 planet candidates and most of them are waiting for their confirmation by follow-up observations. It is expected that around 90% of these candidates will be confirmed as planets.

Dynamical Methods and Microlensing

All other indirect methods rely on the impact of the planets’ gravity.

The radial velocity (RV) method measures the line of sight velocity of a star to infer the existence of any planets. It will be described in more detail in the next section.

Astrometry can be used to map the orbit of a host star when circling the common barycentre with its planetary companion. The astrometric precision required lies beyond what ground-based telescopes can deliver. Benedict et al. (2002) published the confirmation of a planet, which had been discovered with the radial velocity method, by astrometric measurements using the Hubble Space Telescope. This was the first time, the astrometric signal of a planet could be detected. It is expected that the currently launched GAIA satellite will deliver astrometric data of unprecedented precision of about one billion stars and will discover thousands of planets via their astrometric signal.

Microlensing also detects the planets’ gravity field, but in a very different way. When two stars are almost perfectly aligned with respect to the line of sight from an observer on Earth, the foreground star acts as a lens that magnifies the light of the background star. This effect is called gravitational lensing. During a microlensing event, which may last days or weeks, the apparent brightness of the foreground star is changing in a very characteristic fashion. In case the foreground star has a planetary companion, the light curve is disturbed and by comparing models to the observed shape of the light curve, one can deduce physical parameters of the companion such as its mass and orbital distance. As an example, Fig. 1.2 shows the light curve of a microlensing event that led to the discovery of a 5.5 Earth-mass planet, which was the lowest mass of any exoplanet around a regular star that had been discovered at that time (Beaulieu et al., 2006). Compared to other methods, microlensing is quite sensitive to low-mass planets. However, one major drawback of the method is that it is not repeatable: The lensing event occurs just once, since Earth, the foreground and the background star are moving.
Of the detection methods discussed, the transit and RV methods have detected the bulk of currently known planetary systems. As a matter of fact, they may complement each other in the case where both methods can be used for the same target. The transit method gives insight into the orbital period, inclination (as a requirement), and ratio of radii between the host star and its companion. The RV method allows to determine period, eccentricity, and mass ratio modulo the inclination angle. Together, both methods yield extra information such as the average density of the planet, initiating further theoretical work on the formation mechanisms and inner composition of exoplanets, which are still mainly unexplored.

1.2 Radial Velocities

The radial velocity (RV) of an astronomical object is defined as its velocity projected onto the line of sight of an observer. Variations of this velocity can be measured by the use of Doppler spectroscopy: as an object moves towards the observer, its spectrum gets blue shifted and red shifted when the object recedes. When a planet moves around a host star, both objects circle the common barycentre. This interaction leads – depending on the inclination angle of the planet’s orbit – to a characteristic and periodic variation of the radial velocity (RV) of the host star. By observing the stellar spectrum and comparing it to some reference, the RV variations can be measured and parameters of the planetary orbit can be deduced.

However, the RV variations induced by a planet can be rather small. More precisely,
1 Introduction

Figure 1.3 – Left: Radial velocity (RV) signal of 51 Pegasi, the first exoplanet that has been discovered around a solar-type star in 1995 (Mayor and Queloz, 1995). The planet has a period of 4.2 d and its $M \sin(i)$ is 0.461 Jupiter masses. Right: RV signal of a planet around iota Draconis. This was the first detection of a planet around a giant star (Frink et al., 2002). Source: http://exoplanets.org/

The RV amplitude $K_p$ in m/s can be calculated by

$$K_p = 0.0895 \frac{M_p}{\sqrt{aM_*}} \sin(i) \text{ [m/s]}, \quad (1.1)$$

where $M_p$ is the mass of the planet given in Earth masses, $a$ is the semi-major orbital axis in astronomical units, $M_*$ is the mass of the host star in solar masses, and $i$ is the inclination angle of the orbit. As an example, Earth generates a RV signal of the Sun of only 9 cm/s with a period of one year, while Jupiter causes a 12.4 m/s modulation. To search for a planetary companion around a star, the star is regularly observed with a high-resolution spectrograph. The spectral shift of each observation with respect to a wavelength reference is determined with high precision, converted into a relative velocity, and compared to the first program observation. The velocities need to be corrected for the movement of the earth itself (called the barycentric correction). The resulting time series of relative velocities is analysed for periodic variations that stem from a Doppler signal of a sub-stellar companion. This yields the period and the mass ratio of host and companion times the sine of the inclination angle, which is however not known a priori.

Figure 1.3 shows RV data of the star 51 Pegasi, the first Sun-like star around which a exoplanet was detected. Most of the exoplanets that had at first been detected with the RV technique were relatively large planets with short orbital periods, as these are easiest to detect.
1.3 Échelle Spectrographs

The fractional shift in wavelength of a spectral line due to the Doppler shift is given by

\[ \Delta \lambda / \lambda_0 = \frac{RV}{c} , \]

(1.2)

where \( \lambda_0 \) is the wavelength if the star did not move relative to Earth, and \( c \) is the speed of light. In order to measure RV shifts in the meter per second regime or even below, fractional wavelength shifts on the order of \( 1 \times 10^{-9} \) have to be detected. This requires extraordinarily stable and precise spectrographs.

1.3 Échelle Spectrographs

Figure 1.4 – A true colour image of an échelle spectrum recorded using a commercial DSLR camera with the laboratory échelle spectrograph at the Landessternwarte Heidelberg. Each order covers only a few nanometres wavelength range, but due to the two-dimensional format of an échelle spectrum, a large bandwidth can be recorded at once.

The key component for any high-resolution spectroscopic measurements in astronomy is a so-called échelle spectrograph. The term échelle is derived from the French word for ladder, and is also used to name the diffraction grating, which is the characteristic element of an échelle spectrograph. Échelle gratings are characterized by a very large groove spacing and the groove profile is shaped for the use at large blaze angles and thus high diffraction orders. Because of the high order of diffraction, the
Theoretical resolving power \( R \) is also high:

\[
R = \frac{\lambda}{\Delta \lambda} = mN ,
\]

(1.3)

where \( m \) is the diffraction order and \( N \) is the total number of illuminated grooves. However, the above formula is only true in case of diffraction limited optics. When observing stars from the ground with large telescopes, the turbulence of the earth’s atmosphere leads to a blurring of the objects. Their angular size on the sky, \( \phi \), does not decrease for larger telescope mirrors. For this so-called seeing-limited case, the resolution of a spectrograph scales as

\[
R \propto \frac{W}{\phi D} ,
\]

(1.4)

where \( W \) is the width of the grating, and \( D \) is the diameter of the telescope (Schroeder, 1999). This relation is the reason why large telescopes require large spectrograph optics, including large échelle gratings for high-resolution spectroscopy. Throughout the thesis, I will use the terms resolving power and resolution interchangeably, following common nomenclature in the astronomical literature.

Échelles provide dispersion comparable to the highest line density holographic gratings, but as they operate in high orders, the free spectral range (FSR) for each order, i.e. the largest wavelength range that does not overlap with the same range of an adjacent order, is small:

\[
\text{FSR} = \frac{\lambda}{m} .
\]

(1.5)

For \( m \) of the order 100 and for wide spectral bandpass often spanning the whole visible wavelength range, this means that the spectrum is divided into several tens of orders. All orders of an échelle overlap, and one needs a secondary dispersive element, called the cross disperser, which is oriented perpendicular to the échelle’s dispersion direction. Essentially, the cross disperser slightly shifts the orders apart so that they become disentangled and can be recorded simultaneously on a rectangular detector array. This matching of the spectrum’s spatial distribution to commercially available detectors enables the wide simultaneous bandwidth desired in stellar spectroscopy, and is the prime advantage of the cross dispersed échelle configuration. A typical échelle spectrum is shown in Fig. 1.4. Échelle spectrographs used for Doppler studies operate with resolution of the order 100000. However, as stated
above a fractional precision of $1 \times 10^{-9}$ – four orders of magnitude below the native resolution of a spectrograph – is needed to detect RV signals of small planets. This corresponds to a shift of only several nanometres on the detector. To be able to measure shifts of the whole spectrum by this small fraction of a resolution element, and in addition over periods of many months or years, one needs an extraordinarily stable and well-calibrated instrument. Imperceptibly small movements of the optics or the spectrograph bench can move the spectrum well beyond the desired amount. Likewise, temperature changes in the mounts, optical elements or the detector may shift the spectrum. Finally, pressure changes in the refractive index of the medium surrounding the dispersive elements translate the spectrum on short time-scales. To counteract these effects, the leading instruments are mounted inside a vacuum vessel and their temperature is very tightly regulated. The opto-mechanical engineering aims to reduce the sensitivity to temperature changes as much as possible by using low CTE materials. Lastly, very precise wavelength calibration techniques are used to monitor instrumental drifts, often during the science exposure, and to calibrate out any remaining shifts of the spectrum that can mask a RV signal.

However, all these measures only concern the inherent stability of the spectrograph. The spectrum is also sensitive to the coupling conditions, namely the input illumination. This is because an échelle spectrograph images its input slit, which is – in case of a fibre-fed instrument – an image of the fibre itself. Illumination changes in the spatial and angular distribution of the light will lead to a misinterpretation of the spectral content of the light. Additionally to detection limitations, there are also intrinsic effects arising from the stars’ activity that hinder precise RV measurements. Sun-spots, stellar pulsation, flares and also rotation all effect the RV measured and it depends on the particular target star, which is what kind of noise is dominant.

Currently, the most precise spectrographs reach a RV precision of just below 1 m/s. The first instrument achieving this precision was the HARPS spectrograph (Mayor et al., 2003). For an actual image of the instrument, see Fig. 1.5. Inspired by the scientific success and the technological progress, a number of spectrographs for high-precision RV measurements is currently being built. A list of existing and future high-resolution spectrographs is presented in Table 1.1. One of the listed spectrographs is CARMENES. It will be described in more detail in the next section, since the first part of the thesis was closely related to this project.
1 Introduction

Figure 1.5 – The optical bench of the HARPS spectrograph is enclosed in a vacuum tank. At the centre, the large optical échelle grating can be seen. Source: http://www.eso.org

1.4 CARMENES

CARMENES is a high-resolution spectroscopy facility for the 3.5 m telescope at the Calar Alto Observatory in Granada (Spain), consisting of two $R \sim 82000$ échelle spectrographs, one for the visual range from 550 nm to 950 nm (VIS) and one for the near infrared range from 950 nm to 1700 nm (NIR) (Quirrenbach et al., 2012). The main scientific goal of CARMENES is to search for planets within the HZs of 300 low-mass stars. This is also suggested by the official logo of CARMENES (Fig. 1.6): the red circle represents the M-dwarf (also called red dwarf due to its typical apparent colour), the smaller black circle a potential planet.

Figure 1.6 – CARMENES [kár-men-es] stands for: Calar Alto high-resolution search for M dwarfs with exoearths with near-infrared and optical échelle spectrographs
There are several reasons why a study of exoplanets around M-dwarfs is a logical next step in the exoplanet research. As stated before, an Earth-like planet around a Sun-like star produces a RV signal of about 10 cm/s. However, such precision cannot be reached with today’s astronomical spectrographs and next generation instruments like ESPRESSO (Pepe et al., 2014b) have yet to prove that it is technically feasible. When looking for Earth-like planets around low-mass stars at the same orbital distance, the RV signature is considerably higher, because the amplitude of the signal increases with decreasing stellar mass. In addition, the HZ lies closer to M-dwarfs than to solar-type stars, which also implies an increased RV amplitude for planets within this zone according to Eq. (1.1). Stellar luminosity is a steep function of stellar mass that ultimately defines the actual location of the HZ. Taking all these effects into account, one can for example predict that a planet in the HZ around a 0.2 M$_\odot$ star generates a RV amplitude some ten times larger than for a corresponding planet around a solar-type star. Figure 1.7 illustrates how the HZ depends on stellar mass.

**Figure 1.7** – Habitable zone as a function of stellar mass and orbital distance (shaded area). The black dashed line refers to the distance where planets get tidally locked. The red lines show the orbital distance a 5 or respectively a 10 Earth-mass planet needs to have to generate a RV signal with an amplitude of 1 m/s. The left scale is fractional solar masses, the right scale indicates approximate stellar spectral types accordingly. For reference, some planets of our own solar system, namely Venus, Earth, and Mars are illustrated as black dots at 1 solar mass. The regions being accessible to CARMENES lie to the left of the red lines. Figure based on Quirrenbach et al. (2010).
1 Introduction

M-dwarfs are by far the most abundant stars in our milky way (\(\sim 75\%\)) and obtaining statistics on planet abundance around these stars is therefore pivotal to understanding planet formation and evolution. However, the number of planets that have been found around low-mass stars is low compared to those around solar-type stars. The main reason for this is that the stars are relatively faint in the visible wavelength range and that their intrinsic stellar jitter hinders the detection of planets with current high-resolution spectrographs.

**Figure 1.8** – Optical layout of the visual CARMENES spectrograph. Light from the telescope is entering the spectrograph at the fibre exit. The fibre is imaged onto an image slicer that effectively splits the fibre image up into two parts and puts them in top of each other, thereby reducing the fibre width in dispersion direction in order to increase the resolving power. The light is then collimated and falls onto the \(\text{échelle}\) grating where it gets dispersed. Falling back to the collimator, the light is focused onto the fold mirror, redirected, and collimated again. The cross disperser entangles the overlapping orders and the camera focuses the light onto the CCD.

CARMENES has been optimised for the search for planets around low mass stars both by instrument design and observing strategy. As it consists of two spectrographs, spectra can be taken simultaneously in the visual and in the near infrared. This will help to distinguish wavelength-dependent stellar activity effects from real (achromatic) planetary signals. In order to detect Earth-size planets, CARMENES will have to be able to measure radial velocities with a precision better than 3\(\text{m s}^{-1}\), while the aim is to even reach 1\(\text{m s}^{-1}\). As already mentioned, attaining such a level of precision is very challenging from a technical point of view. The resolving power of both spectrographs is slightly above 80000. As long as there is enough light, a higher resolution is favourable, because potentially blended stellar lines can be
1.5 Objective and Outline

resolved, thus allowing for more precise RV determination (Bouchy et al., 2001). However, above \( \sim 80000 \) the gain in RV precision with increasing resolution is small in case of typical M-dwarf spectra. According to Eq. (1.4), a higher resolving power also needs either larger optics, or the field of view (and therefore the number of available photons) has to be reduced. There is another possibility to increase the resolving power: In the optical design of CARMENES (see Fig. 1.8), an image slicer is used to cut the fibre image in half in dispersion direction and reassemble it such that the two halves are on top of each other, roughly doubling the image size in cross dispersion direction but reducing it by the same factor in dispersion direction, effectively doubling the resolution (see also Section 2.5.2). Due to its non-linear nature, the usage of an image slicer is delicate and it comes at the price of higher cross dispersion needed to entangle the \( \text{échelle} \) orders and more detector pixels to cover the same bandwidth.

According to the current schedule of CARMENES, the VIS spectrograph will be delivered to Calar Alto in April 2015 and the NIR spectrograph in September 2015. First scientific results are expected to be obtained by December 2015.

1.5 Objective and Outline

The objective of this thesis is to investigate instrumental limitations of high precision RV measurements with fibre coupled \( \text{échelle} \) spectrographs. The thesis is split into two parts that can be read independently. The first part investigates the optimal use of optical fibres to reduce the sensitivity of spectrographs to unstable coupling conditions at a telescope. The second part of the thesis describes a novel method to generate an extremely stable and precise wavelength calibrator.

Chapter 2 deals with the properties of optical fibres and their optimal use for spectroscopy in the framework of the CARMENES project. After a general introduction to the optical properties of fibres in Section 2.1, I discuss the most important properties in the context of high-resolution spectroscopy in Section 2.2. The requirements regarding the CARMENES project are introduced in Section 2.3.

I present experimental data in Section 2.4 and numerical simulations on illumination effects in fibres and their corresponding impact on spectrum stability are discussed in Section 2.5. In Section 2.6, I investigate methods to mitigate the consequences of the limited number of modes in an optical fibre. Section 2.7 provides a theoretical foundation based on dynamical billiard systems to explain the differing scrambling performances of various types of fibres. In Section 2.8, I present an overview of
1 Introduction

the CARMENES fibre link with an emphasis on its actual manufacturing. Finally, Section 2.9 concludes the first part of the thesis.

Chapter 3 describes a novel precise wavelength calibration method. After an introduction on current wavelength calibration methods in Section 3.1, I discuss in detail the theory of Fabry-Pérot etalons and present own simulations on etalon mirror coatings in Section 3.2. Section 3.3 presents an active locking concept based on laser spectroscopy that enables the usage of an etalon as a stable and precise wavelength calibrator. Section 3.4 describes the conducted experiments and analyses the obtained locking stability. In Section 3.5, the reliability of the setup is discussed. I present basic échelle measurements of the transmission spectrum of etalons in Section 3.6. Finally, Section 3.7 concludes the second part of the thesis and provides an outlook on future work.
1.5 Objective and Outline

For the spectral band and the spectral resolution the maximum value is given. The total efficiency has been extrapolated to include slit losses, and telescope and atmospheric throughput. The radial-velocity precision was estimated from published orbits or standard star’s velocities. Historical instruments have not been listed. It is interesting to note that most of the planets discovered between 1995 and 2003 were detected using a small number of precision instruments: High Resolution Echelle Spectrometer (HIRES) at the 10-m Keck I telescope in Hawaii, CORALIE at the European Southern Observatory (ESO) 3.6-m telescope in La Silla. The Hamilton Spectrophotometer at the Shane 120-inch telescope at Lick, ELODIE at the 1.93-m telescope of the Haute-Provence Observatory, Advanced Fiber-Optic Echelle (AFOE) on the 1.5m telescope at the Whipple Observatory, University College London Echelle Spectrograph (UCLES) at the Anglo-Australian Telescope (AAT), HARPS at the 3.6-m telescope, the Sandiford Cassegrain Echelle spectrograph on the 2.1-m telescope and the High-Resolution Spectrograph (HRS) at the Hobby-Eberly Telescope (HET), all of them at the McDonald Observatory. After 2003 the HARPS spectrograph opened a new window on the domain of super-Earths and mini-Neptunes by improving the radial-velocity precision below the metre-per-second level. Since then, the metre-per-second precision has become a ‘standard’ and a goal for most of the Doppler-velocimeter projects presented in the tables.

<table>
<thead>
<tr>
<th>Instrument/technique</th>
<th>Telescope/observatory</th>
<th>Start of operations</th>
<th>Band (μm)</th>
<th>Spectral resolution</th>
<th>Efficiency (%)</th>
<th>Precision (m s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamilton[69]/self-calibration</td>
<td>Shane 3 m/Lick</td>
<td>1986</td>
<td>0.34-1.1</td>
<td>30,000-60,000</td>
<td>3-6</td>
<td>3</td>
</tr>
<tr>
<td>UCLES[69]/self-calibration</td>
<td>Keck I/AAT</td>
<td>1986</td>
<td>0.3-1.0</td>
<td>25,000-85,000</td>
<td>6</td>
<td>1-2</td>
</tr>
<tr>
<td>CORALIE[69]/sim. reference</td>
<td>ELODIE/ESO La Silla</td>
<td>1998</td>
<td>0.3-0.69</td>
<td>60,000</td>
<td>5</td>
<td>3-6</td>
</tr>
<tr>
<td>UVES[69]/self-calibration</td>
<td>UT2-VLT/ESO Paranal</td>
<td>1999</td>
<td>0.3-1.1</td>
<td>30,000-110,000</td>
<td>4-5</td>
<td>2-5</td>
</tr>
<tr>
<td>HRS[69]/self-calibration</td>
<td>HET/McDonald</td>
<td>2000</td>
<td>0.42-1.1</td>
<td>15,000-120,000</td>
<td>6-9</td>
<td>3-6</td>
</tr>
<tr>
<td>HD[69]/self-calibration</td>
<td>Subaru/Mauna Kea</td>
<td>2001</td>
<td>0.3-1.0</td>
<td>90,000-160,000</td>
<td>6-13</td>
<td>5-6</td>
</tr>
<tr>
<td>HARPS[69]/sim. reference</td>
<td>3.6-m/ESO La Silla</td>
<td>2003</td>
<td>0.38-0.69</td>
<td>115,000</td>
<td>6</td>
<td>&lt;0.8</td>
</tr>
<tr>
<td>FEROS[69]/sim. reference</td>
<td>2.2-m/ESO La Silla</td>
<td>2003</td>
<td>0.36-0.92</td>
<td>48,000</td>
<td>20</td>
<td>10-15</td>
</tr>
<tr>
<td>MIKE[69]/self-calibration</td>
<td>Magellan II/Las Campanas</td>
<td>2003</td>
<td>0.32-1.00</td>
<td>65,000-83,000 and 22,000-28,000</td>
<td>20-40</td>
<td>5-3</td>
</tr>
<tr>
<td>SOPHIE[69]/sim. reference</td>
<td>1.93-m/OHP</td>
<td>2006</td>
<td>0.38-0.69</td>
<td>39,000 and 75,000</td>
<td>4 and 8</td>
<td>2</td>
</tr>
<tr>
<td>OMEGA[69]/self-calibration</td>
<td>UT1-VLT/ESO Paranal</td>
<td>2007</td>
<td>0.95-5.2</td>
<td>&lt;100,000</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>PFS[69]/self-calibration</td>
<td>Magellan II/Las Campanas</td>
<td>2010</td>
<td>0.39-0.67</td>
<td>38,000-190,000</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>PARS[69]/sim. reference</td>
<td>4-m/MAO</td>
<td>2010</td>
<td>0.37-0.86</td>
<td>63,000</td>
<td>10</td>
<td>3-5</td>
</tr>
<tr>
<td>CAFE[69]/sim. reference</td>
<td>2.2-m/Calar Alto</td>
<td>2011</td>
<td>0.39-0.95</td>
<td>67,000</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>CHIRON[69]/self-calibration</td>
<td>1.5-m/CTIO</td>
<td>2011</td>
<td>0.41-87</td>
<td>80,000</td>
<td>15</td>
<td>&lt;1</td>
</tr>
<tr>
<td>HARPS[69]/P[69]/sim. reference</td>
<td>TNG/ORM</td>
<td>2012</td>
<td>0.38-0.69</td>
<td>115,000</td>
<td>10-15</td>
<td>&lt;1</td>
</tr>
<tr>
<td>LEVY[69]/self-calibration</td>
<td>APF/Lick</td>
<td>2013</td>
<td>0.37-0.97</td>
<td>114,000-150,000</td>
<td>10-15</td>
<td>&lt;1</td>
</tr>
<tr>
<td>EXPERT[69]/PNA</td>
<td>2-m/AST/Paraná</td>
<td>2013</td>
<td>0.39-0.9</td>
<td>100,000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>GIANO[69]/self-calibration</td>
<td>TNG / ORM</td>
<td>2014</td>
<td>0.95-2.5</td>
<td>50,000</td>
<td>20</td>
<td>NA</td>
</tr>
<tr>
<td>SALT-HRS[69]/self-calibration</td>
<td>SALT/SAO</td>
<td>2014</td>
<td>0.38-0.89</td>
<td>16,000-67,000</td>
<td>10-15</td>
<td>3-4</td>
</tr>
<tr>
<td>FIRST[69]/PNA</td>
<td>2-m/AST/Paraná</td>
<td>2014</td>
<td>0.8-1.8</td>
<td>60,000-72,000</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>IRD[69]/sim. reference</td>
<td>Subaru/Mauna Kea</td>
<td>2014</td>
<td>0.96-1.75</td>
<td>70,000</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>NR[69]</td>
<td>6-1-m/MAO</td>
<td>2015</td>
<td>0.39-0.86</td>
<td>53,000</td>
<td>3-5</td>
<td></td>
</tr>
<tr>
<td>MINERVA[69]/self-calibration</td>
<td>4-1-m/M. Hopkins</td>
<td>2015</td>
<td>0.39-0.86</td>
<td>NA</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CARMENES[69]/sim. reference</td>
<td>Zeem S5/Calar Alto</td>
<td>2015</td>
<td>0.55-1.77</td>
<td>80,000</td>
<td>10-13</td>
<td>1</td>
</tr>
<tr>
<td>PEPES[69]/sim. reference</td>
<td>LBT/M. Graham</td>
<td>2015</td>
<td>0.38-0.91</td>
<td>120,000-320,000</td>
<td>10</td>
<td>NA</td>
</tr>
<tr>
<td>HET[69]/sim. reference</td>
<td>HET/McDonald</td>
<td>2016</td>
<td>0.98-1.40</td>
<td>50,000</td>
<td>4</td>
<td>1-3</td>
</tr>
<tr>
<td>OMEGA[69]/self-calibration</td>
<td>VT/ESO Paranal</td>
<td>2017</td>
<td>0.95-5.2</td>
<td>&lt;100,000</td>
<td>15</td>
<td>&lt;5</td>
</tr>
<tr>
<td>ESPRESSO[69]/sim. reference</td>
<td>All UTs-VLT/ESO Paranal</td>
<td>2017</td>
<td>0.38-0.78</td>
<td>60,000-200,000</td>
<td>6-11</td>
<td>1</td>
</tr>
<tr>
<td>SPIRou[69]/sim. reference</td>
<td>CFHT/Mauna Kea</td>
<td>2017</td>
<td>0.96-2.35</td>
<td>70,000</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>G-CLEF[69]/sim. reference</td>
<td>GMT/Las Campanas</td>
<td>2019</td>
<td>0.35-0.95</td>
<td>120,000</td>
<td>20</td>
<td>0.1</td>
</tr>
</tbody>
</table>

For the spectral band and the spectral resolution the maximum value is given. The total efficiency has been extrapolated to include slit losses, and telescope and atmospheric throughput. The radial-velocity precision was estimated from published orbits or standard star’s velocities. Historical instruments have not been listed. It is interesting to note that most of the planets discovered between 1995 and 2003 were detected using a small number of precision instruments: High Resolution Echelle Spectrometer (HIRES) at the 10-m Keck I telescope in Hawaii, CORALIE at the European Southern Observatory (ESO) 3.6-m telescope in La Silla. The Hamilton Spectrophotometer at the Shane 120-inch telescope at Lick, ELODIE at the 1.93-m telescope of the Haute-Provence Observatory, Advanced Fiber-Optic Echelle (AFOE) on the 1.5m telescope at the Whipple Observatory, University College London Echelle Spectrograph (UCLES) at the Anglo-Australian Telescope (AAT), HARPS at the 3.6-m telescope, the Sandiford Cassegrain Echelle spectrograph on the 2.1m telescope and the High-Resolution Spectrograph (HRS) at the Hobby-Eberly Telescope (HET), all of them at the McDonald Observatory. After 2003 the HARPS spectrograph opened a new window on the domain of super-Earths and mini-Neptunes by improving the radial-velocity precision below the metre-per-second level. Since then, the metre-per-second precision has become a ‘standard’ and a goal for most of the Doppler-velocimeter projects presented in the tables.

Table 1.1 – Non-exhaustive table of present and future spectrographs for high-precision RV science. Source: Pepe et al. (2014a)
2 Optical Fibres in High-Resolution Spectroscopy

The results presented in this chapter have partly been published in Stürmer et al. (2014). Unless stated otherwise, all figures presented are based on own calculations or measurements.

2.1 Theoretical Background

The optical properties of fibres are usually described by the means of ray or wave optics. For fibres with a large core size, the ray optical treatment is a good approximation but becomes more and more inaccurate when the core size of the fibre decreases towards diameters that are only a few times larger than the wavelength. Following the more descriptive approach, we will introduce the basic properties of optical fibres by ray optics. Figure 2.1 illustrates the ray propagation inside a step-index fibre, which is the most commonly used type of optical fibre in astronomy. This fibre consists of a core with a refractive index $n_1$ and a cladding with a slightly lower index $n_2$. When light is coupled into the fibre from an ambient medium with index $n_a$, its angle with respect to the optical axis of the fibre is refracted according to Snell’s law to $\sin(\beta) = n_a/n_1 \sin(\alpha)$. The ray is propagating inside the core and eventually hits the cladding, where it gets reflected by total internal reflection as long as the incidence angle $\gamma$ is smaller than $\sin(\gamma_{\text{max}}) = n_2/n_1$. This can be translated into the maximum acceptance angle of the fibre

$$\alpha_{\text{max}} = \arcsin\left(\sqrt{n_1^2 - n_2^2}\right) .$$  \hspace{1cm} (2.1)

The argument of the arcsin is called numerical aperture (NA):

$$\text{NA} = \sqrt{n_1^2 - n_2^2} .$$  \hspace{1cm} (2.2)
A step-index multi-mode fibre consists of a glass core and a cladding that has a slightly lower refractive index than the core. A ray that hits the fibre under an angle $\alpha$ travels along the fibre due to total internal reflection, as long as $\sin(\gamma) < n_2/n_1$. The coating (e.g. Acrylate, Polyimide) protects the fragile glass fibre from mechanical damage and also provides the peculiar mechanical flexibility of a fibre.

It measures the difference between the refractive index of the core and the cladding and is an important design parameter of a fibre. Its values typically range between $\text{NA}=0.12$ and $\text{NA}=0.28$, but there exist also high-NA fibres with numerical apertures up to 0.5. The most common type of large core fibres has $\text{NA}=0.22$, which corresponds to an acceptance angle of $\alpha_{\text{max}} \sim 12.7^\circ$. In the visible and the near infrared, fibres are usually made of pure fused silica ($\text{SiO}_2$) and their cladding is additionally doped with fluorine to lower the refractive index and match the design numerical aperture of the fibre. For longer wavelengths, other materials such as fluoride glasses or chalcogenide glasses have to be employed, because silica has a fair transmission only up to about $2\,\mu\text{m}$ (Haynes et al., 2006).

Within the scope of this thesis, only fused silica fibres are investigated, since they show a good transmission throughout the full wavelength range of CARMENES, as illustrated in Fig. 2.2. Ray optics are particularly useful to describe the optical properties of fibres with large cores, which are the most common ones in astronomy. However, their characterization by means of ray optics is not completely accurate. In reality, wave phenomena such as interference or cladding penetration will occur so that a proper wave-optical treatment is still necessary.
Theoretical Background

Figure 2.2 – Typical transmission spectrum of a fused silica fibre. Different absorption and scattering processes contribute to the total intrinsic attenuation. Source: Mitschke (2010)

The derivation of wave equations follows the one of Mitschke (2010) in structure and notation. Using the assumptions that there are no free charges, no currents and that there is no magnetisation within the material, Maxwell’s equations in MKS units can be simplified to

\[ \nabla \cdot \mathbf{D} = 0, \]  \hspace{1cm} (2.3)

\[ \nabla \cdot \mathbf{B} = 0, \]  \hspace{1cm} (2.4)

\[ \nabla \times \mathbf{B} = \mu_0 \frac{\partial \mathbf{D}}{\partial t}, \]  \hspace{1cm} (2.5)

\[ \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \]  \hspace{1cm} (2.6)

where \( \mathbf{E} \) is the electric field strength, \( \mathbf{B} \) the magnetic induction, \( \mathbf{D} \) the dielectric displacement, and \( \mu_0 \) the vacuum permeability. The curl of Eq. (2.6) gives the wave equation for the electric field:

\[ -\nabla (\nabla \cdot \mathbf{E}) + \nabla^2 \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu_0 \frac{\partial^2 P}{\partial t^2}, \]  \hspace{1cm} (2.7)

using the relation \( \frac{1}{c^2} = \mu_0 \varepsilon_0 \) between the vacuum permeability, the permittivity and \( c \). The polarisation \( P \) is connected to the dielectric displacement via \( \mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} \).
Equation (2.7) can be derived for the magnetic field in an analogous way. An optical fibre made of ultra-pure fused silica can well be approximated by a homogeneous medium, which means one can assume that the polarisation is always parallel to the field strength. Therefore, the dielectrics are isotropic: \( \epsilon = \epsilon_0 \epsilon_r \) with \( \epsilon_r \) being a scalar. Additionally, light intensities are always low in astronomical spectroscopy, which implies a linear relationship between \( P \) and \( E \). The wave equation can therefore be further simplified to

\[
\nabla^2 E = \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} E,
\]

(2.8)

with \( n^2 = \epsilon \) being the refractive index and respectively for the magnetic component of the wave

\[
\nabla^2 H = \frac{n^2}{c^2} \frac{\partial^2}{\partial t^2} H.
\]

(2.9)

In case of a step index fibre in the linear regime (low light intensities), the ansatz

\[
E = E_0 N Z T
\]

(2.10)

allows to separate the wave function into its field distribution in the plane normal to the \( z \)-axis \( N = N(x, y) \), along the \( z \)-axis \( Z = Z(z) \), and a monochromatic wave with angular frequency \( \omega \): \( T(t) = e^{i\omega t} \).

Finally, the wave equation can be simplified to the Helmholtz equation:

\[
\nabla^2 N Z + n^2 k_0^2 N Z = 0.
\]

(2.11)

The solution to this equations can be derived analytically only in few cases when symmetries of the underlying geometry can be used to simplify the problem. For a fibre, \( Z \) takes the form \( e^{-i\beta z} \) with wave number \( \beta \).

In a circular step index fibre, the equation can be solved by using a cylindrical coordinate system and the solutions take the general form:

\[
N(r, \phi) = c_1 J_m(wr/a) \cos(m\phi + \phi_0) : \quad r \leq a
\]

(2.12)

\[
N(r, \phi) = c_2 K_n(wr/a) \cos(m\phi + \phi_0) : \quad r > a,
\]

(2.13)

where \( a \) is the core radius and \( c_1, c_2 \) are normalization constants, which account for the smoothness of the solution at the core cladding transition at \( r = a \). The integer
2.1 Theoretical Background

$m$ is derived from the azimuthal part of Eq. (2.11). $J_m$ is the Bessel function of first kind and $K_m$ is the modified Bessel function of second kind.

The dimensionless quantities $u$ and $w$ depend on the refractive indices of the fibre, its diameter, and the wavelength. The constant $\sqrt{u^2 + w^2}$ is called normalized frequency or V-number, and is defined as

$$V = \sqrt{u^2 + w^2} = k_0 a \text{NA} = \frac{2\pi}{\lambda_0} a \sqrt{(n_1^2 - n_2^2)} .$$

(2.14)

The V-number is of great importance, since it determines all relevant information about the fibre as well as the wavelength.

The conditions for smooth transition at $r = a$ and recursion relations between Bessel functions allow to derive the characteristic equations:

$$\frac{J_m(u)}{u J_{m+1}(u)} = \frac{K_m(w)}{w K_{m+1}(w)} \quad \text{and} \quad w = \sqrt{V^2 - u^2} .$$

(2.15)

Obviously, only certain combinations of arguments yield valid solutions of these equations. The discrete set of solutions is called \textit{(guided) modes} and is denoted by $LP_{mp}$, where the index $m$ specifies the number of pairs of azimuthal maxima and $p$ labels the number of maxima in radial direction. As an example, Fig. 2.3 shows the intensity patterns of all existing modes in a circular fibre where $V = 6$. As noted above, the V-number contains information about the particular experimental conditions (wavelength) and the fibre. $V$ also defines the number of guided modes inside a fibre. For $V \approx 2.405$, there exists only one mode, the $LP_{01}$, and the fibre is called a single-mode fibre (SMF). In practice, this means that - assuming a standard NA – the core radius is sufficiently small, i.e. a few microns for visible light.

As it is impossible to couple light from a telescope into a SMF efficiently (at least for large telescopes without the use of adaptive optics), fibres with core radii from 50 $\mu$m to 400 $\mu$m are commonly used in astronomy. In case of a step index fibre, the number of supported modes for large V-numbers can be approximated by

$$N_{\text{modes}} \approx \frac{4}{\pi^2} V^2 \approx V^2 / 2 .$$

(2.16)

For a 100 $\mu$m diameter fibre, the number of modes available in the visible is on the order of a few thousands and when coupling light into it, a superposition of these modes will get excited. The limited number of modes within a fibre will lead to an additional noise contribution in a fibre-fed spectrograph as will be described in more
Figure 2.3 – Intensity distributions of all existing modes for a fibre where $V=6$. In a weakly guiding fibre, i.e. a fibre where $(n_1 - n_2)/n_2 \ll 1$, the modes can be simplified to the linearly polarised modes, the $LP$ modes. They are denoted by $LP_{mp}$, where the index $m$ specifies the number of pairs of azimuthal maxima and $p$ labels the number of maxima in radial direction.

detail in Section 2.6.

Throughout the thesis, the terms near field and far field are used to describe the intensity pattern directly at the fibre face and in a plane a couple of centimetres behind the fibre exit, respectively. It should be emphasized that the near field (NF) and the far field (FF) are not uncorrelated. To the contrary, one can in principle fully reconstruct the NF from the FF and vice versa by Fourier transform. However, in practise this is not possible, because one only measures an intensity distribution and not a complex amplitude.

The nomenclature NF and FF is commonly used in the literature, however its definition is often flawed. A measurement of the NF would require to scan across the fibre tip with an appropriate small aperture and map out the power distribution (see Chapter 7.4 in Mitschke, 2010). In the astronomical literature, measurements of the NF of a fibre are always performed by imaging the fibre tip with additional optics. This is problematic in the sense that the optics used always suffer from aber-
2.1 Theoretical Background

![Figure 2.4](image)

**Figure 2.4** – Typical NF (left column) and FF (right column) intensity distributions, when a large core fibre (here a 100 µm circular fibre) is illuminated by polychromatic light (top row) or by monochromatic light (bottom row), respectively.

Aberrations. Typically, these aberrations depend on the exact illumination of the optics, i.e. the fibre FF in this case. Therefore, variations in the FF will contaminate the NF measurements. Nevertheless, I will follow the standard approach of imaging the fibre when investigating the fibre NF. A typical NF and FF of a large core fibre are shown in Fig. 2.4, where the fibre was illuminated both by monochromatic and by polychromatic light.
2.2 Optical Properties of Fibres

Fibres are a key optical element of a fibre-fed spectrograph. Not only do they allow to separate the spectrograph from the highly variable environmental conditions the telescope is exposed to. They also stabilize the input illumination of the spectrograph, which is an important factor for high-precision spectroscopy in terms of attainable RV precision. However, one major drawback is that they do not conserve the \textit{étendue}.\footnote{The \textit{étendue} is a characteristic property of an optical system and measures the extent of a light beam both in area and solid angle. It is a conserved quantity in geometric optics and also known as $A\Omega$ product.} The main optical properties of fibres that are of particular interest for astronomical spectrographs will be discussed in more detail in the next two subsections.

2.2.1 Scrambling

During the late 1970s and early 1980s, astronomers realized that optical fibres could significantly reduce systematic wavelength errors in spectroscopic measurements compared to slit spectrographs, due to their incomplete transfer of spatial image content (Serkowski et al., 1979; Heacox, 1988). In case of a slit spectrograph, errors of the guiding of the telescope, different focus positions and seeing variations all lead to a direct change of the illumination of the spectrograph. These illumination effects cause systematic wavelength shifts that limit the RV precision attainable for Doppler spectroscopic measurements. However, in case of a fibre-coupled spectrograph, guiding errors and other environmental parameters such as temperature entail a change of the particular coupling conditions of the fibre. Luckily, while the light is travelling along the fibre, this spatial information is partly lost. This beneficial property of a fibre is called \textit{scrambling}. It is one of the greatest benefits of a fibre-fed spectrograph compared to a slit spectrograph. Yet, the output of a fibre is not fully independent of its input. As a consequence, changes in the fibre illumination may cause a varying fibre output. In case of an échelle spectrograph, this can easily be misinterpreted as a change of the spectral content of the light. This phenomenon is illustrated in Fig. 2.5. Usually, the scrambling gain (SG) is defined as the ratio of the relative displacement of the barycentre at the entrance of the fibre $d_i/D_i$ and the relative displacement of the spectral line at the detector:

$$\text{SG} = \frac{d_i/D_i}{f/F},$$ \hspace{1cm} (2.17)
2.2 Optical Properties of Fibres

Figure 2.5 – This schematic illustrates how a variation at the fibre input translates into a shift of the spectral line. When a star moves across the fibre entrance, e.g. due to imperfect guiding, parts of the spatial information is transported to the spectrograph input. Here, the illumination variation causes an apparent shift of the spectral line, although the spectral content of the light did not change. The scrambling gain (SG) of the fibre link is defined as the ratio of the relative movements of the barycentres at the fibre input and the spectral line.

where \( f \) is the displacement of the line and \( F \) is its full width half maximum (FWHM), see also Fig. 2.5.

Heacox (1987) had already noted early on that in case of standard circular step index fibres, the azimuthal scrambling is nearly perfect, while some spatial information is propagated radially along the fibre. Hunter and Ramsey (1992) were among the first to extensively investigate scrambling properties of circular optical fibres in the context of astronomical spectrographs. They also commented on the importance of the far-field illumination for the stability of a spectrograph. The authors introduced what they called a double scrambler, which further helps increasing the homogeneity of the fibre output beam. They split the fibre and used lenses to interchange the angular and positional rays, i.e. the pupil of the first fibre is imaged onto the image
plane of the second fibre and vice versa. By doing so, they increased the poor radial scrambling of the circular fibre and at the same time stabilized the far-field of the entire fibre link. Their device had a throughput of only 20%, but the principle of exchanging near field and far field became the standard solution for many high stability spectrographs. Examples are the HERMES spectrograph (Raskin et al., 2011), the HR+ mode of the SOPHIE+ spectrograph (Perruchot et al., 2008) or HARPS (Pepe et al., 2000). By optimisation of the design, the throughput could further be increased and lies between 70% to 85% for more recent devices (Barnes and MacQueen, 2010).

Avila et al. (2006) and Avila and Singh (2008) investigated alternative scrambling solutions, such as solid light pipes, mechanical squeezing and fibre agitating, single lens scramblers, and beam homogenizers. A promising candidate with good scrambling efficiencies is the mechanical squeezing of fibres. Hereby, the fibres are slightly bent at small radii, which increases the mode-to-mode coupling and results in a more homogeneous near field and far field pattern. However, stress on fibres is known to increase focal ratio degradation (FRD) (see next section) and the experiments mentioned above showed indeed a reduced FRD performance. Avila et al. (2010) and Chazelas et al. (2010) presented first results for the scrambling performance of fibres with a non-circular core geometry. Square, rectangular, and octagonal geometries were shown to have very promising scrambling properties. These results have been confirmed by subsequent work, see Avila (2012), Chazelas et al. (2012), and Spronck et al. (2012). In particular, octagonal fibres are nowadays used on-sky in several high-resolution spectrographs with excellent results, e.g. at CHIRON, HAPRS-North or SOPHIE+ (Tokovinin et al., 2013; Consento et al., 2012; Perruchot et al., 2011), where in case of the latter they are used in combination with an optical double scrambler.
2.2 Optical Properties of Fibres

2.2.2 Focal Ratio Degradation

![Figure 2.6](image)

Figure 2.6 – The schematic illustrates focal ratio degradation (FRD) in an optical fibre: The light cone emerging from the fibre is larger than the input cone. FRD is an unavoidable property that occurs in any optical multi-mode fibre. Different scattering processes as well as environmental conditions and fibre handling contribute to the total amount of FRD.

A second fundamental property of optical fibres is that they do not conserve the étendue, i.e. fibres tend to scatter light into a cone that is larger than the one of the incoming light. This effect is often called focal ratio degradation (FRD) and is illustrated in Fig. 2.6. Parry (1998) emphasized that FRD represents a specific type of loss and is therefore undesirable for an astronomical spectrograph. In case the optics subsequent to the fibre do not account for FRD, light that emerges at an angle larger than the input angle will simply be lost. In case the collimator of an échelle spectrograph is adapted to collect all the light and the F-ratio of the camera is kept constant, FRD entails that more detector pixels are needed in order to record a single spectral resolution element. Also, larger optics, in particular the échelle grating, become more expensive. FRD has been encountered ever since multi-mode fibres have been used in astronomy (Angel et al., 1977) and numerous studies have investigated this effect in the context of astronomical instrumentation (Avila, 1988; Ramsey, 1988; Barden, 1998; Crause et al., 2008). However, the results of the different studies are far less consistent and conclusive than in the case of scrambling properties. This is because FRD is influenced by various parameters and measurement methods are not standardized among research groups. Haynes et al. (2011) described in detail what kind of physical processes contribute to FRD:
scattering, diffraction, and modal diffusion. Material impurities, micro- and macro-bending, mechanical stress, and fibre end-face quality all affect FRD. As these processes have different wavelength dependencies, there is yet no consensus in the literature on the overall wavelength dependency of FRD. Rather, it is determined by the dominant FRD source of the particular fibre under test. There is also no agreement on whether FRD has a dependency on fibre length. If so, it is a small one, at least for the relatively short fibre lengths that typically astronomical fibres have (compared to e.g. telecommunication). Furthermore, precise FRD measurements are challenging, as for example already slight misalignments of the optical axis of the fibre and the rest of the optical system lead to an additional artificial FRD (see Avila, 1998). One aspect that all research groups agree on is that FRD losses are generally more prominent when a fibre is fed by a beam that is slow compared to the NA of the fibre. As a consequence, the light from the telescope is usually coupled into fibres at F-ratios faster than \( \approx F/5 \) to limit FRD losses.

### 2.3 Requirements for CARMENES

Scrambling and FRD are important quantities for high-resolution spectrographs that need to be understood and carefully tested when designing a fibre link. This part of the thesis deals with the following aspects of designing a suitable fibre link for CARMENES: understanding the particularities of optical fibres, determining the relevant parameters for an elaborate fibre link and eventually providing such fibres. In doing so, financial constraints had to be taken into account, since the costs of a custom fibre draw can easily reach a few ten thousand Euro, which is beyond the project’s budget.

The design parameters for the CARMENES fibre link were determined by the specific conditions at the Calar Alto observatory and scientific top level requirements. According to internal CARMENES documents, the 80 percentile seeing at Calar Alto is measured to be \( \sim 1.05'' \) in the V-band and scaled to \( \sim 0.9'' \) at 1\( \mu \)m. The 3.5m telescope has a focal ratio of F/10, which corresponds to a plate scale of 169.78\( \mu \)m/". In order to collect \( \sim 90\% \) of the light of a star for these given seeing conditions, the field of view (FoV) of the fibre was set to \( \sim 1.47'' \). Figure 2.7 shows the relation of coupling loss and fibre FoV/seeing. The actual fibre size then solely depends on the F-ratio that is used for feeding. In order to keep FRD losses low, while also taking into account the availability of appropriate fibre core sizes, the feeding optics convert the F/10 beam of the telescope to a F/3.9 beam. The 1.47"
are then covered by a 100µm fibre. As the spectrograph optics accept a F/3.5 beam emerging from the fibre, the design accounts for quite some FRD of the fibre link. The project’s minimum requirement on total throughput of both the VIS and NIR spectrograph including atmosphere and telescope is 5%, while the project’s goal is 7%. This translates into a throughput of the fibre of 70%, including all types of losses, such as coupling, FRD, intrinsic absorption. The same factor of 70% is used for a potential scrambler, since this efficiency is typical for a double scrambler as noted earlier. Therefore, the fibre link that includes the scrambler is required to have a throughput of at least 49%.

The requirement for the scrambling gain is derived from the scientific goal of keeping the total error budget within 1 m/s RMS. The guiding accuracy of the 3.5m telescope is reported to be 0.12'' RMS, which translates to a shift of about 8.2µm at the fibre entrance. A resolution element in CARMENES has a width of 3.75 km s⁻¹, and therefore the scrambling gain has to be \( SG = \frac{0.12''/1.47''}{0.5 \text{ m s}^{-1}/3.75 \text{ km s}^{-1}} \approx 600 \) to keep the systematic shifts below 0.5 m/s. In order not to let the guiding errors be the dominant noise source of the spectrograph, the goal for the scrambling factor was set to 1000.

Regarding FRD, there is no specific requirement, as FRD losses are already included by their contribution to the total throughput losses.

**Figure 2.7** - **Left:** Coupling loss versus fibre field of view/seeing. **Right:** Gaussian disk with a typical FWHM according to seeing measurements on Calar Alto and the CARMENES FoV of the fibre.
2.4 Experiments

This section describes the experiments that I performed to investigate possible solutions for the CARMENES fibre link. The optimal solution needs to maximize the throughput while fulfilling the project’s requirements stated in the previous section.

2.4.1 Optical Test Setup

First, I describe my optical setup of the fibre test-bench. I do so in detail, as reliable and reproducible results strongly dependent on a correct setup. The whole setup is depicted schematically in Fig. 2.9, while a photo is presented in Fig. 2.10. The basic purpose of the setup is to reimage, under well known conditions, a polychromatic source onto a fibre and to investigate its output both in NF and FF. Figure 2.8 shows the illumination system of the pinhole. If not stated otherwise, a white LED, which has its spectral peak around 450 nm is used as a light source. I placed two diffuser plates in front of the LED to eliminate structures of the light source and homogenize the beam. For the scrambling tests, I used a 10 µm pinhole, while a 50 µm pinhole is inserted for FRD measurements.

The pinhole is imaged one to one onto a fibre, which is mounted on a XYZ stage and a gimbal mount for tip/tilt adjustments. The Y axis is motorized, to enable automated measurement series. A beamsplitter in the collimated beam allows to look at the fibre input directly with a camera. By doing so, one can precisely determine where the spot hits the fibre.

![Schematic diagram of the optical test setup](image)

**Figure 2.8** – Optical setup for illuminating the pinhole that is imaged onto the fibre under test. Two diffusers eliminate spatial structures in the LED that could contaminate precise measurements. The microscope lens focuses the light onto a pinhole with a fast beam.
2.4 Experiments

**Figure 2.9** – Optical setup for reimaging the pinhole onto the fibre under test. An aperture stop after the first collimator allows to set the F-ratio of the light feeding the fibre. A camera monitors the fibre entrance via a beam splitter and allows therefore to determine the exact position where the light enters the fibre. A shear plate interferometer is used to ensure that the light is properly collimated.

**Figure 2.10** – Photo of the optical setup for illuminating the fibres under test. The optics are assembled within a rigid and stable cage system to ensure good mechanical stability.
2.4.2 Nearfield Measurement Procedure

In order to measure the scrambling gain, I took a series of images of the NF for varying positions of the spot on the entrance of the fibre. The spot is displaced across the fibre entrance by the motorised axis of the XYZ-stage. According to the definition of the scrambling gain (SG), see Eq. (2.17), one would need to measure the SG directly with the spectrograph, which is often not possible. Instead, the SG is derived by measuring

\[ SG = \frac{d_i/D_i}{d_o/D_o}, \]  

where \( d_o/D_o \) is the relative displacement of the barycentre directly at the fibre output (i.e. the shift of the barycentre of the NF).

However, this definition is somewhat problematic due to several reasons. First, the SG generally varies for different spatial displacements \( d_i \). Thus, it is always more conclusive to look at a series of displacements rather than presenting just a single measured SG. Second, the definition only accounts for movements of the barycentre of the fibre output, which is not the only possible source of line distortion or line shift in a spectrograph. In fact, simulations show that even for circular fibres (which are known to have relatively poor scrambling gain), the contribution of imperfect scrambling on the barycentre shift is negligible (Allington-Smith et al., 2012). In reality, barycentric shifts are still measured, due to inhomogeneities of the fibre surface, due to illumination dependent aberrations of the optics used for investigating the NF, and due to an inhomogeneous sensitivity of the detector. All these effects are also present in a spectrograph. As a consequence, if one only looks at the barycentric shift of the fibre output, the effect of imperfect scrambling can be underestimated. On the other hand, the barycentric shift can easily be converted to a radial velocity error, whereas other measurements of the inhomogeneity of the NF, such as RMS values or higher image moments, do not have a direct relation to the RV error. They are therefore only useful for relative comparisons of different fibres or scrambling methods.

For precise scrambling gain measurements, the setup needs to be mechanically very stable. Any movement of the fibre that is imaged by the microscope objective contributes to the overall measured barycentric shift and contaminates precise SG measurements. Avila et al. (2006) suggested to mount a reference fibre next to the fibre under test and to measure only differential shifts. In case of connectorized
fibres, this was not possible, because the field of view of the camera was too small to image two fibres at the same time. Instead, I developed an alternative procedure that identifies fibre movements and allows to reject contaminated measurements, or in some cases, to correct for the movements: After dark subtraction for each NF image, a mask is created by blurring and thresholding the image. The mask covers the fibre NF plus a small margin around it. All remaining pixel values are ignored for further processing. By analysing the mask position of the image series, one has a good indicator of how stable the setup was during the measurement. Ideally, the mask would not move at all, but only the illumination within it would change. In reality, the image of the NF will not stay at the exact same position, due to mechanical instabilities of the setup. In principle, one could correct for these movements of the fibre by subtracting the mask position from the measured barycentres, but non-linearities and inhomogeneities might lead to ambiguous results. Therefore, I only used the mask position as an indicator for the setup stability. This method has another advantage over a simple weighted mean of the whole detector: It efficiently suppresses noise from outside the region of interest that could otherwise massively distort the results, especially when calculating higher image moments. Figure 2.11 shows the NF setup and a measurement of the bench stability in more detail. To measure the intrinsic stability, I kept the illumination of the fibre constant. Any remaining barycentre movements indicate instabilities, mainly of mechanical nature, over the plotted timescale. The maximum deviations of the barycentre are below 10 nm over a timescale of 5 min, which validates the good stability of the setup.

Figure 2.11 – **Left:** Near field (NF) setup for measuring the scrambling gain. **Right:** The stability of the setup is shown for a series of NF images of a 100 µm fibre that were taken over a timespan of 300 s. In order to measure high scrambling gain, the setup must be mechanically very stable and the optics should have a sufficient magnification.
2 Optical Fibres in High-Resolution Spectroscopy

2.4.3 Farfield Measurement Procedure

I measured the FF of the fibre to check both FRD and the influence of a moving spot on the FF structure. For measuring the FRD, the angle of the cone of light exiting the fibre has to be determined. Since the exact distance between the CCD (FLI ProLine 16803) and the fibre is unknown at the beginning of a measurement, I used a differential method. I calculated the angle of the outgoing beam $\theta_0$ using a linear fit to the sizes of the FF at different distances between the detector and the fibre (Fig. 2.12). The F-ratio of the beam is then defined by

$$F/# = \frac{1}{2 \tan(\theta_0)}$$

(2.19)

The CCD is mounted with a motorized focuser (FLI Atlas) that allows to control very precisely the relative distance between the fibre and the CCD, as the resolution of a single step is only 8 nm. After dark subtraction, the size of the light cone is determined by calculating its radius of 95% encircled energy (EE95). As mentioned above, FRD measurements can easily be contaminated by a misalignment of the optical axis of the fibre with respect to the illumination system. In order to avoid this effect, I adjusted the tip/tilt angle at the beginning of the procedure until the diameter of the cone recorded on the CCD was minimised at a fixed distance of the fibre to the CCD, as suggested by Murphy et al. (2008). Other error contributions to FRD measurements are non-perfect collimation, errors in the aperture stop diameter, and deviations from a top-hat profile of the pupil illumination. To check for correct collimation of the beam, I used a shearing interferometer. As aperture stops, I used a set of fixed apertures to ensure repeatability in setting the input F-ratio. It should further be noted that the calculated FRD value also depends on the encircled energy level used for the determination of the spot size. Choosing EE99 or EE90 will lead to slightly different results (see Wang et al., 2013). This makes the comparison between the absolute values I measured and the ones proposed in the literature very difficult.

Additionally to the FRD images, I took another set of images in similar fashion as for the scrambling gain measurements in the NF: while moving the spot across the fibre entrance, I recorded a series of FF images. As mentioned before, there exists no simple relation between variations in the FF and the RV shifts they entail. I link RV shifts to FF variations via numerical simulations that will be presented in Section 2.5.2.
The whole procedure of fibre scrambling and FRD characterisation is summarized in Fig. 2.13. I designed a custom software that controls the camera, the linear stage, and the focuser, thus facilitating a convenient and quick measurement procedure.

\[ \theta_o = \arctan \left( \frac{\Delta y}{\Delta x} \right) \]

**Figure 2.12** — **Top left:** Schematic of the differential FRD measurement procedure: the camera is moved by well known steps, and for each position the size of the light cone exiting the fibre is determined. **Bottom left:** A typical result from this procedure indicates the size of the cone at five different positions of the camera. A linear fit allows to determine the F-ratio and the distance between the fibre and the CCD. **Right:** Image of the FF setup in the laboratory. The assembly is light-tight to avoid any contamination of the measurements by background lighting.
Figure 2.13 – Flow chart of focal ratio degradation (FRD) and scrambling measurements. The calibrated far field (FF) images and the near field (NF) images are stored for further analysis.
2.4 Experiments

2.4.4 Results

Using the setup described in the previous section, I tested several fibres with core sizes being appropriate for CARMENES, which I all polished in-house. Technical specifications of the fibres are listed in Table 2.1. The scrambling results are depicted in Fig. 2.16, while FRD measurements are illustrated in Fig. 2.19. The plotting style for the SG measurements is adapted from Feger et al. (2012).

All results presented in this section were derived from a series of measurements that had been conducted within a short time frame (∼2-3 days) in order to guarantee consistent test conditions and minimize systematic errors, e.g. by a change in setup.

<table>
<thead>
<tr>
<th>Type</th>
<th>Core shape</th>
<th>Core [µm]</th>
<th>Cladding [µm]</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>WF 100/125A</td>
<td>circular</td>
<td>100</td>
<td>125</td>
<td>0.22</td>
</tr>
<tr>
<td>WF 100/187P</td>
<td>octagonal</td>
<td>100</td>
<td>187</td>
<td>0.22</td>
</tr>
<tr>
<td>WF 67/125P</td>
<td>octagonal</td>
<td>67</td>
<td>125</td>
<td>0.22</td>
</tr>
<tr>
<td>WF 100x100/123x123A</td>
<td>square</td>
<td>100x100</td>
<td>123x123</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 2.1 – Technical information on the tested fibres from CeramOptec.

Scrambling Gain

In case of 100µm circular fibres, the scrambling gain ranges from 100 to 200. This is consistent with results for circular fibres obtained by Avila et al. (2010) and Feger et al. (2012). The output illumination in the NF depends strongly on the coupling conditions, which is shown in Fig. 2.14. In these measurements, SG does not noticeably depend on fibre length.

Non-circular fibres clearly improve the homogeneity of the fibre output: by simple visual inspection, there is no noticeable difference in the NF patterns if the spot is moved across the fibre entrance. This leads to an improvement of the SG of a factor of ∼10 in case of the square fibre, compared to a circular one. For the octagonal fibre, the measurements yielded an even higher SG: depending on the actual spot displacement, it exceeded 1000 (except for the very smallest spot displacements) and for larger initial displacements values even 2000. The same performance was obtained by the circular-octagonal fibre link, with a circular fibre coupled to an octagonal one with a FC-FC junction, see Fig. 2.15. The shift of the barycentre of the output is reduced by a factor of almost 20 when using the circular-octagonal link. The RMS shift of the barycentre is about $0.16 \cdot d/1000$, which corresponds to $16\,µm$ or $\sim 1.2\,m/s^{-1}$ in case of the CARMENES spectrograph. Due to the definition of the scrambling gain, systematic errors mostly affect the SG values obtained for small
2 Optical Fibres in High-Resolution Spectroscopy

Figure 2.14 – Top row: Image of the fibre entrance. A 10 µm spot is moved across the 100 µm circular fibre. Middle row: The corresponding near field of the fibre shows strong variations. Bottom row: The far field shows also prominent structures that vary when coupling conditions change.
2.4 Experiments

initial displacements. This manifests itself in the small vertical distance between the regions indicating different SGs in Fig. 2.16. Considering the uncertainty of the measurement, the lower SG for small displacements (in the centre of the bottom plot of Fig. 2.16) is not considered to be problematic for our case. We note that the similar SG obtained with the octagonal fibre alone indicates that this fibre is mostly responsible for the good scrambling performance of the combined circular-octagonal link.

![Figure 2.15 - Top row: Image of the fibre entrance. A 10µm spot is moved across the 100µm circular fibre of the circular-octagonal fibre link. Middle row: By visual inspection, the corresponding near field of the octagonal fibre does not show any noticeable illumination variations any more. Bottom row: The far field is also free of structures and it stays relatively constant for the different coupling conditions. However, it shows an increased focal ratio degradation.](image-url)
Figure 2.16 – Barycentric movement of the fibre output for different spot displacements at the fibre entrance. **Top left:** 100 µm circular fibre, **Top right:** 100 µm square fibre (note different scaling), **Middle left:** 100 µm octagonal fibre, **Middle right:** 100 µm circular fibre + octagonal fibre. **Bottom:** 1D version of the circular + octagonal scrambling, also showing regions that indicate different levels of scrambling gain as defined in Eq. (2.18).
2.4 Experiments

**FF homogeneity and stability**

In general, FF structures are much more prominent in short fibres than in longer ones, see Fig. 2.17. This might be an indication that FRD, which in general smoothes out FF structures, is indeed dependent on fibre length. Still, the fibre length seems to have only a minor contribution to FRD: measurements of short and long fibres show a very similar performance, as illustrated in Fig. 2.19.

![Cross sections of the FFs of four different fibres.](image)

**Figure 2.17** – Cross sections of the FFs of four different fibres. The shorter circular fibre (blue line) shows more prominent features than the longer one (purple line). The octagonal fibre (black line) shows a good homogeneity, however, some coupling dependent structures near the centre can still be noticed. For the circular-octagonal link (red line), no irregular structures can be seen any more. This comes at the cost of a higher FRD, which manifests itself in faint wings on the outer part of the FF.

The FF of the octagonal fibre is relatively homogeneous, see Fig. 2.17. Some structures are visible close to the centre of the illumination pattern, however, they change rapidly with varying coupling conditions. Variations in the FF cannot directly be translated to spectral line shifts, but qualitatively the same arguments as in case of the NF hold: the more stable the FF is, the more stable is the spectral line on the detector. Therefore, a homogeneous and stable intensity distribution in the
2 Optical Fibres in High-Resolution Spectroscopy

Figure 2.18 – FF pattern of the 100 µm square fibre, when tilting the fibre with respect to the input light beam. Radial structures arise when the fibre is illuminated at different angles of incidence (AOI). The diameter of the light cone is also increasing when tilting, due to geometrical FRD.

FF is desired to ensure a stable spectrum. The circular-octagonal link shows good performance in this regard, and meets the requirements for the CARMENES spectrographs. The resulting FFs are free from any prominent structures and changes of the FF pattern for different spot displacements are minimal. Section 2.5.2 summarizes the results in a quantitative fashion.

FRD

The good performance of the circular-octagonal link, i.e. its smooth FF patterns as well as its high SG, comes at the price of slightly increased FRD losses (Fig. 2.19). My measurements show that the FRD performance of non-circular fibres is comparable to that of circular fibres, which is in agreement with more recent publications (Feger et al., 2012; Avila, 2012). For instance, at F/3.9 the octagonal fibre shows an output F-ratio of \(\sim F/3.8\), while for the circular-octagonal link the output F-ratio ranges from F/3.25 to F/3.55, depending on the particular link. The increase in FRD is therefore on the order of 5% to 14%. These values are comparable to the increase in FRD that Avila (2012) reported for his method of squeezing fibres in order to homogenize the FF. However, the benefit of a dual fibre link is the interface at the vacuum tank. This allows for easy replacement of the long fibre between the telescope and the spectrograph without breaking the vacuum. These fibres are somewhat vulnerable as they are moved constantly. In case of CARMENES, the interesting quantity regarding FRD is the encircled energy (EE) within a F/3.5 beam when the fibre is illuminated by F/3.9, because this matches the coupling conditions at the telescope and the accepted F-ratio of the spectrograph optics. Figure 2.20 compares the measured EE of the prototypes of the circular-octagonal link for the best and the worst cases. The values range from 91.0% to 95.4%.
One particular result that requires further comment is the seemingly inferior performance of the square fibre. In the course of the FRD measurement, prominent patterns in the FF of the square fibre appeared. This prompted further investigation of the fibre after the end of the test. During a visual inspection of the fibre with a microscope to check for defects, the fibre broke, most probably due to a fibre crack located near the connector. The high FRD values are not in agreement with older measurements that clearly showed that the FRD is comparable to that of the octagonal or the circular fibre. Also, Chazelas et al. (2010) reported that in their tests the square fibres, obtained from the same vendor as our samples, suffered from strange patterns in the FF, whose origin was not clear. In previous measurements, no irregularities in the FF of our square fibre under test were detected. However, prominent radial features were visible in the FF when I tilted the fibre with respect to the incoming beam (Fig. 2.18). The patterns I recorded for the damaged square fibre looked very similar to the ones for a high angle of incidence.

In conclusion, my measurements show that non-circular fibres have a similar FRD performance compared to circular ones, while their scrambling is at least ten times better. In order to eliminate FF structures, which are most prominent in short fibres, one has to induce some FRD. This can be done by squeezing in the spirit of Avila (2012) or by linking fibres of different cross sections. A circular-octagonal link is a good solution for two reasons: First, circular geometries match the geometry of optical components, which are also mostly circular symmetric. In case of CARMENES, the field of view of the fibre is circular, because in the focal plane the light has to pass a small hole in a mirror that is used for telescope guiding. In case an octagonal fibre would be used directly at the telescope, one would either have to underfill the fibre or drill up the hole in the mirror to fully cover the octagon. The latter would reduce the light available for guiding of the telescope, while not all of the light, which now passes the circular hole, is coupled into the fibre. Second, coupling two octagonal fibres requires either more advanced fibre couplers that allow to rotate the fibres with respect to each other, or one has to accept higher coupling losses due to a larger geometrical mismatch (see also Section 2.8.1).
Figure 2.19 – Focal ratio degradation measurements of different fibres and circular-octagonal fibre junctions. In general, FRD is higher for slow input F-ratios and gets close to the theoretical maximum (blue solid line) for fast input beams. The behaviour of circular and octagonal fibres is very similar. However, the circular-octagonal fibre junctions show an increased FRD. The square fibre shows a seemingly inferior performance. However, the fibre turned out to be damaged (see also main text), and former measurements showed a similar performance compared to octagonal fibres. The measurement was still included in this figure, to illustrate that FRD measurements can readily reveal partial damages of fibres.
2.4 Experiments

Figure 2.20 – Measured encircled energy versus F-ratio of the output beam for a F/3.9 input beam of two prototypes of the circular-octagonal fibre link. For CARMENES, the encircled energy (EE) within F/3.5 is the most interesting quantity, as it determines how much light exiting the fibre is accepted by the spectrograph optics. This is indicated by the grey solid vertical line. Depending on the particular link, the EE within F/3.5 lies between 91.0% to 95.4% as indicated by the horizontal solid lines. For reference, also the EE within the F/3.9 output cone is indicated by the dashed lines.


2.5 Numerical Simulations

In this section, I present a couple of numerical simulations that I ran in order to better understand the optical properties of fibres and to translate the measured illumination variations in NF and FF into RV errors.

2.5.1 Fibre Simulations

Precise measurements of scrambling gain and FRD are rather involved and time-consuming. Moreover, non-circular fibres are not available with arbitrary core shapes and diameters and customized fibre manufacturing is very costly. Simulations are therefore a convenient way of predicting and analysing the potential performance of new fibre types. In the context of circular optical fibres, an early analytical approach of image transfer was presented by Heacock (1987). He theoretically derived the poor radial scrambling of circular fibres, which was confirmed experimentally at about the same time. More recently, Allington-Smith et al. (2012) performed raytracing simulations for fibres with various core shapes with the aim of providing an insight into their scrambling and FRD behaviour. The authors wrote their own raytrace method to account for FRD. They limited the region where FRD happens to a small section of the fibre, because the authors argued in former studies that FRD does not depend on fibre length (Poppett and Allington-Smith, 2010).

For the simulations presented here, I used the commercial software ZEMAX to investigate some basic scrambling behaviour of various fibres. My simulations mimic the experiments for scrambling measurements as presented in Section 2.4.4. In the non-sequential mode of ZEMAX, a light spot is imaged onto a fibre and moved across the fibre surface. At the fibre output, both NF and FF are recorded. The fibre is modelled by a solid silica tube surrounded by another silica tube, while the difference in their refractive indices was manually set to match a standard NA 0.22 fibre. The fibre is assumed to be straight, i.e. without bends and non-uniformities of the fibre cross-section.

In a first test to validate this simulation approach, a spot with 5μm diameter and wavelength $\lambda = 550\,\text{nm}$ is moved across a 100μm circular fibre and the NF is compared to actual measurements. Indeed, the simulated and the measured NF show

---

2 [https://www.zemax.com](https://www.zemax.com)

3 In the sequential mode of ZEMAX, rays travel from surface to surface in a predefined order. In the non-sequential mode, ZEMAX allows for tracing rays when there are multiple optical paths. This is useful for the implementation of e.g. scattering or extended light sources.
a qualitative match (see second and third row of Fig. 2.21). However, a closer look at the cross section in the middle of the fibre reveals that the simulations generally show more pronounced features in the NF compared to the measurements (Fig. 2.22). This is not surprising, as by default no scattering or other FRD effects are simulated. The resulting NF pattern is thus purely dependent on the geometry of the boundary. In principle, ZEMAX can handle both bulk and surface scattering and it even allows for user defined scattering functions. Any scattering will effect the NF as well as the FF, and it is a valid assumption that scattering processes will happen within a fibre. As an example, I turned on Mie bulk scattering. As depicted in Fig. 2.22, a persuasive match between the simulation and experiment could be reached by appropriately adapting the scattering parameters such as mean free path and the size of the particles. This points towards the possibility of providing full quantitative models in ZEMAX if the scattering properties of the fibres are investigated empirically, and accounted for accordingly.

It is clear that when turning on bulk scattering, effects such as FRD become dependent on fibre length. As stated in Section 2.2.2, the literature does not agree as to whether FRD shows a length dependence or not. It seems to be a reasonable assumption that FRD is mainly determined by fibre end effects and stress from the adhesive that is used to hold the fibre. This could easily be implemented in ZEMAX by restricting the scattering processes to small sections of the fibre in the spirit of Allington-Smith et al. (2012). However, I did not attempt to trace the sources of FRD to their actual physical processes, such as different scattering processes within the fibre. Figure 2.22 should therefore only be seen as a strong indication that realistic NF patterns can be simulated when taking scattering into account. Nevertheless, even without considering scattering effects, I was able to verify with my simulation that ray tracing is a feasible way to describe multi mode fibres, and in particular that circular fibres are poor radial scramblers.

The same is true for fibres with an elliptical core, see Fig. 2.23. In general, the illumination is far from homogeneous and strongly dependent on the input parameters. The results indicate, that in case a homogeneous illumination is desired, it is not sufficient to remove circular symmetry from the core shape. A theoretical explanation will be given in Section 2.7.
Figure 2.21 - Experimental and simulated NF images of a 100μm circular fibre for three different input conditions. A 5μm spot is moved across the fibre. **Top row:** recorded images of the spot on the entrance of the circular fibre. **Middle row:** the corresponding output NF. **Bottom row:** ZEMAX simulations of the NF of a 50mm long, 100μm diameter circular fibre for the same input conditions, indicated by the grey circle. For the first two rows, the units of the x and y axis of the images are in pixels, while in the bottom row, the units are in micron. A clear qualitative match between the simulation and the experiments validates the approach to simulate large core multi-mode fibres by raytracing methods.
2.5 Numerical Simulations

Figure 2.22 – Comparison of simulated and experimental data of a NF cross section for a 100 µm circular fibre when illuminated with a 5 µm spot with a 37 µm lateral offset. By including scattering effects, a good quantitative match between the simulations and the experiments can be achieved.

Figure 2.23 – Simulated NF images of a fibre with an elliptical core shape when illuminated with a small spot at the geometrical centre (left), when illuminated with a random offset (middle), and when illuminated near the focal point of the ellipse (right). In general, the illumination strongly depends on the lateral displacement of the spot. Removing circular symmetry is therefore not sufficient for a homogeneous output illumination.
I expanded the simulations to non-circular core shapes, in particular to polygonal shapes, because these fibres experimentally showed a good scrambling performance. Figure 2.24 shows the simulated NF of fibres where the core is a regular polygon with $n = 4$ to $n = 12$ vertices. Here, the spot was located exactly at the centre of the fibre. For all other positions that have been investigated, the NF of all polygonal fibres were equally homogeneously illuminated. Only when the spot was placed at the centre of the polygons, inhomogeneities arise. A few interesting conclusions can be drawn from these results: In general, the scrambling gets better with increasing length of the fibre, but only a relatively short length (longest simulated length in the figure is 100 mm) is needed to homogenize the output. Polygons with a low number of sides show a better scrambling performance than geometries with a greater number of sides. Also, polygons with an even number of sides scramble the input better than odd numbered ones. The first statement can be explained by the fact that the more sides a polygon has, the better its approximation by a circle. The second conclusion, namely that odd numbered polygons have an inhomogeneous NF when the spot is placed in their geometrical centre, should be taken with a pinch of salt as numerical issues due to the singular nature of the sharp corners cannot fully be excluded. Up to now, there is no experimental evidence that odd numbered polygons show this behaviour, simply because of a lack of availability of such fibres.

Allington-Smith et al. (2012) noted that the lack of scrambling (even for a circular fibre) has no effect on the barycentre, because the inhomogeneities show axial symmetry. Although this statement is correct for an ideal fibre and an ideal detector, it will not hold for real fibres. Slight scratches on the surface, dust or other irregularities paired with inhomogeneities of the detector or the optics that image the fibre onto it will break the axial symmetry and lead to a barycentric shift as soon as the output illumination changes.

One result that requires further comment is the effect of macroscopic bends on the fibre output. Experimental measurements have shown that bending a fibre homogenizes the NF, but induces FRD. When simulating bends solely by geometrical means, the homogeneity of the NF indeed increases. However, instead of an increasing light cone exiting the fibre, rays with a large angle with respect to the optical axis get lost as they exceed the critical angle. This is also found by Allington-Smith et al. (2012). This is in clear contrast to what is found experimentally. Most probably, the discrepancy comes from the fact that a macroscopic bend also changes the refractive index (Sokkar et al., 2014), which is neglected in the current simulations.
Figure 2.24 – Simulated NF images of fibres with polygonal core shapes when illuminated with a small spot at the geometrical centre of the fibre. The simulated length of the fibre is increasing from top to bottom row (16 mm, 36 mm, 64 mm, 100 mm). The number of sides is increasing from left to right from a square to a dodecagon. For each geometry, the bottom row shows the projection of five rays onto the plane that is perpendicular to the optical axis of the fibre. Radial structures and deviations from a homogeneously illuminated core only occurred when the spot was placed exactly in the geometrical centre. For all other positions, the output illumination was equally homogeneous.
2.5.2 Impact of Fibre Illumination Effects on RV Precision

The stability of illumination of the spectrograph optics is an essential factor for precise RV measurements. In the literature, it is often called point spread function (PSF) stability of the spectrograph. Yet, as pointed out by Spronck et al. (2013), the term PSF should rather not be used in the context of illumination stability of spectrographs, as this is not the appropriate term from an optical point of view. This is because the PSF is the response function of optics for a point source. However, when a multi mode fibre is used to feed a spectrograph, the fibre is not a point source but rather an extended object. Variations in both the NF and the FF of the fibre will affect the attainable RV precision.

I investigate these effects by raytracing methods using the ZEMAX models of the CARMENES spectrographs to translate the variations of the measured NF and FF into RV errors. Figure 2.25 displays the traced image of a 100µm fibre on the CARMENES detector. It is generated by emerging a few thousand rays from one of the two input fibres and raytracing them through the CARMENES optics in the sequential mode of ZEMAX, including the image slicer that was implemented by a user defined surface. For each field point, i.e. point on the fibre, only the chief ray is traced. It is a reasonable assumption that for a particular wavelength the optical aberrations over the field of view, which is in this case the fibre, are constant. Obviously, optical aberrations will vary with wavelength as will be discussed later, but when looking only at one wavelength, all rays will be affected equally. It is therefore a valid simplification to neglect rays other than the chief ray when looking at the NF effects. Figure 2.25 shows that the two slices of the circular fibre are not perfect half circles, but distorted ones. In fact, the two slices correspond to two halves of an ellipse. The optical distortions of the échelle grating cause the circular shape of the fibre to be elliptically deformed. However, it is a valid approximation to translate the barycentric shifts of the fibre NF one to one onto the detector. According to internal CARMENES documents, the effect of different flux ratios of the two half ellipses is more than a factor of 10 smaller than actual barycentric shifts of the fibre NF. This justifies calculating the scrambling gain and therefore the RV shifts directly from the barycentric movements in Section 2.4.4.

In order to simulate the impact of variations of the FF on RV precision, I used the ZEMAX models of the CARMENES spectrographs to generate spot diagrams for a number of different wavelengths. For a particular wavelength, the optical aberrations are effectively constant over the field of view, which allows to separate
2.5 Numerical Simulations

Figure 2.25 – Raytrace of a 100µm circular fibre through the VIS CARMENES optics onto the detector at 867.7278nm in échelle order 70. The grid indicates the 15µm pixels of the CCD. The fibre was sampled with ~ 31000 rays. On the left, all rays are plotted with the same colour. For the other two images, the colour indicates the relative illumination. The intensities are interpolated from two recorded NF images of a 100µm fibre.

The NF from the FF effects. For each ray of each spot diagram, the intersection coordinates with the pupil plane of the respective spectrograph are stored. Each ray can therefore be uniquely mapped from the detector plane to a position on the pupil plane. This information allows one to determine the barycentre of a spot diagram by calculating a weighted mean of all ray positions for different pupil illuminations. Different FF illuminations therefore correspond to different weightings of the rays. I made sure that the procedure is numerically stable, by comparing the results I obtained with an increasing number of sampling rays in the pupil plane.

Extreme cases like D-shaped apertures of different orientation can easily cause shifts in the order of several tens of meters per second. This is also confirmed both by simulations (Boisse et al., 2010) and by experimental investigation at other spectrographs: Perruchot et al. (2011) measured a RV shift in the order of several meters per second when vignetting the telescope aperture by the dome. Although these cases are not realistic for normal telescope operation, they illustrate the importance of the FF illumination for the stability of a spectrograph.
2 Optical Fibres in High-Resolution Spectroscopy

Figure 2.26 – Simulation of the impact of FF variations: A FF image, here an artificial radial gradient (left), is interpolated onto the pupil plane of the spectrograph (middle). Raytrace simulations allow to map each ray of a spot diagram to a certain position in the pupil (right). Different FF illuminations induce a different weighting of the rays, therefore shifting the barycentre of the spot.

Boisse et al. (2010) only simulated partial vignetting of the telescope pupil by an aperture stop. Instead, the method I present here allows for a more detailed analysis and yields estimates of the RV errors for more realistic FF variations. In the laboratory, I took a series of images of the FF for different coupling conditions: Slight defocus as well as various spot displacements at the fibre entrance have been used to generate a series of FF images. I extracted the FF within the F/3.5 cone that is accepted by the CARMENES optics and used the intensity distribution as the weighting function for the spot diagrams. The impact on RV for both the VIS and NIR channel of CARMENES for different coupling conditions is illustrated in Figs. 2.27 to 2.32. The two FFs that have been used to generate the RV error maps in Fig. 2.28 and Fig. 2.31 have been generated by offsetting a 50 µm spot by 22.3 µm, which corresponds to a guiding error of 0.13'' - the RMS guiding error at Calar Alto. During these measurements, a prototype of the F/N system that converts the F/10 beam to a F/3.9 beam was attached to the fibre tip.

Generally, the error maps reveal that distortion effects of a varying FF are larger where optical aberrations are more prominent, i.e. in the corners of the detectors. In other words, the larger the spots are, the more sensitive they are to illumination effects, which is intuitively plausible. Having a good control over optical aberrations is therefore crucial for instrument stability. It can also be seen that the effects are mostly symmetric across the detector. Averaging over many lines along one échelle order will therefore help to cancel out illumination effects. This is a consequence of the symmetry of the spot diagrams, but it is not generally true for any spectrograph design. For CARMENES, the total shift for a single line caused by typical changes of the FF is below 1 m s$^{-1}$. However, it can rise up to several m/s for more extreme
lateral displacements as illustrated in Fig. 2.29 for the VIS channel, and in Fig. 2.32 for the NIR, respectively.

However, it is difficult to predict how the FF varies on a real telescope over a longer period of time. Guiding, focussing, and seeing have an impact on the exact shape of the FF. It is still very instructive to perform these simulations to get a better understanding of the impact of illumination variations on spectrograph stability and to possibly incorporate these results for new spectrograph designs. In this spirit, Fürész et al. (2014) explicitly included a term in the merit function during the G-CLEF (GMT-CfA, Carnegie, Catolica, Chicago Large Earth Finder) spectrograph design that minimizes shifts caused by pupil illumination variations.

One limitation of the presented raytrace approach is that it neglects diffraction effects. Also, it is assumed that the FF variations, which are measured with the fibre being illuminated with white light, are on average the same for different wavelengths. Still, the method provides insight into the complex relation of optical aberrations, FF illumination, and radial velocity stability.
Figure 2.27 – Spot diagrams of different spots across the CCD of the VIS channel of CARMENES. The black circle indicates the airy disk at the particular wavelength. The colours indicate the relative weight of each ray. The weights are extracted from FF images recorded in the laboratory. The positions of the spots on the detector, indicated by the number at the top right of each spot diagram, are shown in Fig. 2.28.
Figure 2.28 – Two FF images are recorded in the lab with different coupling conditions that are expected to occur at the telescope. The image intensities are used as relative weights of each ray depending on its pupil coordinate. Calculating the weighted mean for both illuminations gives an estimate on how barycentres of a spectral line will shift. In most cases, the errors show a symmetry with respect to the centre of the échelle order. The numbers on the red dots refer to the different spot diagrams shown in Fig. 2.27.
Figure 2.29 – RV shifts on the VIS detector due to FF illumination changes for a series of FF images simulating a guiding offset of $-0.52''$ (top left), $-0.26''$ (top right), $0.26''$ (bottom left), and $0.52''$ (bottom right), respectively.
2.5 Numerical Simulations

Figure 2.30 – Spot diagrams of different spots across the detector of the NIR channel of CARMENES. The black circle indicates the airy disk at the particular wavelength. The colours indicate the relative weight of each ray. The weights are extracted from FF images recorded in the laboratory. The positions of the spots on the detector, indicated by the number at the top right of each spot diagram, are shown in Fig. 2.31.
Figure 2.31 – Shifts of the spectral lines due to typical FF illumination effects in the NIR channel of CARMENES. The numbers on the red dots refer to the different spot diagrams shown in Fig. 2.30.
2.5 Numerical Simulations

Figure 2.32 - RV shifts on the NIR detector due to FF illumination changes for a series of FF images simulating a guiding offset of 

$0.52\arcsec$ (top left), $-0.26\arcsec$ (top right), $0.26\arcsec$ (bottom left), and $0.52\arcsec$ (bottom right), respectively.
2.5.3 Pupil vs. Image Coupling

The results described in the previous section may have an impact on the fibre coupling technique. There seems to be no agreement in the literature as to whether feeding the pupil of the telescope or feeding the image of the telescope is superior in terms of spectrograph stability. The two feeding methods are illustrated in Fig. 2.33.

Figure 2.33 - Left: Pupil coupling with a microlens. Rays for two different field positions are shown. The translation of spatial variations into angle variations is obvious. Right: For image coupling, additional optical elements are needed. Because of the (nearly) telecentric mapping, rays with different field positions end up under the same angle on the fibre. The rod lens is used to minimize glass-air transitions.

In the case when no optical double scrambler is used, but solely non-circular fibres instead (or a combination of fibres with different shapes), a telecentric feeding optic is clearly the best choice. While the fibres will efficiently stabilise the NF, the FF will not get much further stabilised. This means error sources that usually dominate at the telescope, namely guiding errors and seeing variations, can efficiently be suppressed by the fibre link. In contrast, the telescope pupil, which is known to be much more stable than the image, will only gain a further stabilization due to smoothing caused by FRD.

In total, my simulations and the tests I conducted in the laboratory led to a change of the coupling strategy for CARMENES: Initially, it had been planned to use microlenses and couple the telescope pupil on the NF of the fibre, as it is done for instance at the FEROS spectrograph. However, any movement of the star across the fibre will induce a dramatic change of the FF. This is because any lateral changes translate into angular variations that cannot be efficiently scrambled without a double scrambler.
2.5 Numerical Simulations

**Figure 2.34** – FF images of a 100µm circular fibres for three different lateral displacements of a spot when coupled via a microlens. Due to the pupil coupling, lateral displacements translate into angle variations that are not scrambled by the fibre. Such FF variations would cause a shift of the spectral line of several ten meters per second.

This is illustrated in Fig. 2.34, where a series of FF images of a circular fibre is shown for different lateral displacements of the spot and micro-lens pupil coupling. Although these variations diminish considerably when the circular fibre is coupled to an octagonal, the FF remains still very sensitive to lateral displacement and simulated RV shifts are up to 50 m/s.

The coupling of the CARMENES fibres at the telescope will now be done via a telecentric F/N system that feeds the fibres with F/3.9, thus projecting the telescope’s focus directly onto the entrance surface of the fibre.
2 Optical Fibres in High-Resolution Spectroscopy

2.6 Modal Noise

2.6.1 Background

Generally speaking, an optical fibre supports only a limited number of modes. Therefore, additional noise is introduced in fibre-coupled spectrographs, which was first observed by Baudrand et al. (1998). Both radial velocity stability as well as the signal-to-noise-ratio (SNR) that is achievable are influenced by the illumination variations due to interference effects on the output of the fibre. The number of excited modes of a fibre that is illuminated with unpolarised light of a fixed F-ratio, \( F/\# \), is given by

\[
M = 2\pi^2 \left( \frac{d}{2\lambda} \right)^2 \sin^2 \left( \arctan \left( \frac{1}{2F/\#} \right) \right) .
\]

(2.20)

The modal noise limits the signal to noise ratio (SNR) to

\[
\text{SNR} = \frac{\rho}{\nu} \cdot \sqrt{\frac{M+1}{1-\rho^2}} ,
\]

(2.21)

where \( \nu \) is the Michelson contrast of the speckle pattern and \( \rho^2 \) is the fraction of light reaching the detector (Goodman and Rawson, 1981; Kanada and Aoyama, 1983; Lemke et al., 2011). The number of modes \( M \) is determined by fibre size, wavelength as well as coupling conditions and is therefore a fixed number for a given spectrograph system. So is \( \rho^2 \), but exact numbers are hard to predict a priori. In principle, \( \rho^2 \) should be kept as high as possible by avoiding overfilling of any spectrograph optics including artificial slits after the fibre. But even in the case of CARMENES, where no vignetting should occur, inhomogeneities of the grating, scattering, and other sources of light loss lead to a \( \rho^2 \) smaller than unity. For instance, the amount of scattering at the slicing edge strongly depends on the spatial distribution of the modes and is therefore wavelength dependent. The wavelength dependency of the light loss modulates the spectrum and is the origin of modal noise (Chen et al., 2006).

For a given spectrograph, the speckle contrast \( \nu \) is the only variable in Eq. (2.21) that can be manipulated to increase the SNR limit. In Fig. 2.35, the number of available modes and the corresponding SNR that can be achieved are illustrated for different spectrographs. Astronomical integration times are usually of the order of a few seconds to several minutes. The speckle contrast can therefore be reduced
by inducing changes of the exact speckle pattern within this timeframe. Agitating the fibre is the simplest way to do so. Using piezo elements allows for well defined stretching or twisting of the fibre, but requires access to the bare fibre (see Chen et al., 2006). Mahadevan et al. (2014) recently developed a very effective method to minimise speckle contrast for the calibration light by using an integrating sphere combined with a dynamical diffuser. Because of the light loss of this solution, it is not applicable for fibres transporting stellar light, but shows an impressive performance in case of highly coherent calibration light sources.

![Figure 2.35 - Top: Number of excited modes for various spectrographs according to Eq. (2.20). In the visual, CARMENES has as many modes available as e.g. HARPS does. In the near-infrared, the number of modes is substantially lower, leading to a lower attainable SNR of a spectrum. Bottom: The maximum attainable SNR of a spectrum according to Eq. (2.21) for two different values of the speckle contrast $\nu$. Agitating the optical fibres leads to a decrease of $\nu$ and allows to obtain spectra with higher SNR.](image)

2 Optical Fibres in High-Resolution Spectroscopy

2.6.2 Experiments

In order to find an appropriate way to reduce the speckle contrast, I tested different ways of agitating the fibres. I compared the methods by calculating the Michelson contrast of the speckle pattern in the NF of the fibre:

$$\nu = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}.$$  \hspace{1cm} (2.22)

A laser is fed into a SMF that is re-imaged one to one on the fibres under investigation, namely either a 100µm circular fibre, a 100µm octagonal fibre or a 67µm octagonal fibre. Using this kind of illumination, only a small number of modes is excited. Hereby, I simulate the low number of modes available in the 100µm fibre in the H-band of CARMENES, which could not be tested directly due to the lack of an appropriate detector. In my analysis, I use only an inner region of the recorded NF for calculating $\nu$, since it is very difficult to determine the exact border of the fibre. Including areas outside the NF will heavily contaminate the contrast measurements, which has to be avoided by all means. The region of analysis needs to be large enough to provide good statistics, but as long as this criterion is fulfilled, the contrast does not significantly depend on the exact size of the region any more. Before extracting the contrast value, I perform dark subtraction and flatfielding. The flatfield is created by illuminating the fibre with white light and agitating it. In case of the circular fibre, flatfielding was necessary to cancel out the artificial increase of the speckle contrast due to inhomogeneous illumination, while for the octagonal fibre, it did not make a substantial difference. This is yet another indicator of the superior scrambling properties of these fibres.

Fig. 2.36 shows speckle patterns and corresponding contrast values for different methods of agitation. As expected, the contrast for a non-agitated fibre is near unity (0.97). A gentle and regular movement of the fibre already reduces the contrast by a factor of $\sim$2.5. By using two subsequent mechanical agitators (wiper motors), a contrast reduction by a factor of $\sim$4.5 was achieved, roughly equalling the reduction that can be reached by hand agitation. Although this solution does not seem very elegant, the method successfully imitates agitation by hand, which is known for being an effective method of modal noise reduction (McCoy et al., 2012). Still, the mechanical solution is very easy to implement and maintain, in addition it is also both cheap and robust.
2.6 Modal Noise

Figure 2.36 – Speckle patterns of a 100µm circular fibre for different kinds of fibre agitation. The red boxes indicate the areas where the contrast values are derived from. **Left:** no agitation. **Middle:** gentle regular mechanical movement. **Right:** two subsequent mechanical agitators with moderate agitation. The exposure time of all images was 5s.

In case of the octagonal fibres, I measured even lower values for the speckle contrast when the fibres were agitated by two subsequent mechanical agitators (see Fig. 2.37). Most probably this is because even with flatfielding the images of the circular fibre might be contaminated by the inhomogeneous output illumination that contributes to the measured speckle contrast. The values for the smaller octagonal fibre are not worse than those of the 100µm one. These measurements do not suggest that the speckle contrast for an agitated fibre depends on the exact number of modes available, at least as long as a few thousand modes are still available.

Also, I did not find evidence for a dependency on the specific location where the fibre is agitated as stated by (Baudrand and Walker, 2001). In order to test this, I agitated a 50m long circular fibre at ~1m distance from the beginning and end, respectively. In both cases, I measured the same contrast value.

In the laboratory, I also tested other methods, including low amplitude modulation of the bare fibre with frequencies of 20Hz to 120Hz and ultrasonic transducers. Although these methods work equally well, the handling of bare fibres is delicate. Also, merely by mounting the bare fibres, additional FRD may easily be introduced due to the external stress. A three dimensional model of the fibre agitator for CARMENES is shown in Fig. 2.38, which was designed by Gerardo Veredas based on my experiments in the laboratory. The device does not perform a chaotic movement as suggested by Grupp (2003). However, since the resulting speckle contrast is as low as in the case of non-regular hand agitation, a chaotic movement seems not to be essential for speckle suppression as long as the amplitude of the agitation is high enough.
Figure 2.37 - Speckle patterns of a 100μm octagonal fibre at two different wavelengths. **Left column:** no agitation, $\lambda = 633\text{nm}$. **Middle and right column:** two subsequent mechanical agitators with moderate agitation (532nm and 633nm, respectively). The exposure time of all images was 5s.

My tests also showed that if a double scrambler is used, it is the fibre behind the bulk optics that should be agitated. This might be problematic if the second fibre lies completely within the tank (as it does for HARPS), in which case any moving parts are unwanted. Agitating only the first fibre has a significantly lower efficiency in speckle contrast reduction.
Figure 2.38 – Design of the fibre shaker for CARMENES by Gerardo Veredas. Two wiper motors that are arranged 90° to each other agitate the fibre similarly to irregular hand agitation. The fibres are moved with a frequency of $\sim 1$ Hz, which gives a sufficient decrease of the speckle contrast during an astronomical observation.
2 Optical Fibres in High-Resolution Spectroscopy

2.7 Dynamical Billiard

This section links the light propagation in optical fibres to the theory of dynamical systems. I discuss how ray propagation along a fibre relates to the time evolution of a dynamical billiard. By using this mathematical framework, many of the formerly described properties of fibres can be explained. Furthermore, it can be used to theoretically investigate properties of fibres with arbitrary shapes. Finally, I discuss fibre cross-sections that induce chaotic dynamical behaviour. These fibres have the potential to offer superior scrambling and modal noise properties.

2.7.1 Billiard Systems

A billiard system (or simply billiard) is a standard problem in dynamical system theory (see Chapter 3, Korsch et al., 2008). It describes the frictionless motion of a particle within a region $\Omega$ of a (piecewise) smooth boundary curve $r(\varphi) \equiv \partial \Omega$. The particle is reflected elastically at the boundary. $\Omega$ is called billiard table, or simply table, while the trajectory of the particle is also called orbit.

Despite the simplicity of the problem, the system can show all forms of dynamical behaviour from integrability to chaos, depending on the shape of the boundary curve and the initial parameters of the particle.

**Figure 2.39** - Angles and geometry of a billiard: The trajectories can fully be described by a set of angles $\varphi_i$ denoting the polar angle of the intersection point with the boundary, and $\alpha_i$ denoting the angle between the trajectories and the tangent of the boundary at the intersection point.
The geometry and relevant angles of a billiard system are shown in Fig. 2.39. A billiard can be described by a Hamiltonian system:

$$H(p_x, p_y, x, y) = \frac{1}{2m}(p_x^2 + p_y^2) + V(x, y) \quad ,$$

with

$$V(x, y) = \begin{cases} 0 & (x, y) \in \Omega \\ \infty & (x, y) \notin \Omega \end{cases} \quad .$$

In classical mechanics, the systems treated are almost always integrable. This means that for a system with $N$ degrees of freedom, there exist $N$ constants of motion which are in involution. By definition, the energy is always conserved in a billiard system. Therefore, the time evolution of the particle in the four-dimensional phase space $(x, y, p_x, p_y)$ can be restricted to a three-dimensional sub-manifold, often called the energy surface. In case a second constant of motion exists (e.g. constant angular momentum for rotationally symmetric boundary curves), the trajectory of the particle usually lies on a torus. Exceptions from this rule will be discussed later. If we follow the trajectories for a typical billiard in a Cartesian coordinate system, the resulting orbits show a complicated form. It is therefore very useful to switch to a different coordinate system, the Poincaré-Birkhoff coordinate system. These coordinates describe a Poincaré section that naturally emerges from the geometry of a billiard system. The first coordinate corresponds to the length of the perimeter at the point of impact at angle $\varphi$

$$S(\varphi) = \frac{1}{L} \int_{0}^{\varphi} \sqrt{r'^2(\varphi') + (dr/d\varphi')^2} d\varphi' \quad ,$$

where $L$ is the total length of the boundary curve. The second coordinate is $p = \cos(\alpha)$, where $\alpha$ denotes the angle between the trajectory after reflection and the boundary (see Fig. 2.39). It can therefore be interpreted as the parallel momentum of the point-mass. As the total momentum is constant and is usually set to unity, the four-dimensional phase space can be reduced to two dimensions. The underlying geometry allows to construct a mapping $T$ that relates each point in the Poincaré section to the succeeding point of impact:

$$(S_n, p_n) \xrightarrow{T} (S_{n+1}, p_{n+1}) \quad .$$
In these new coordinates, the dynamics of the system can now be investigated much more easily. An example for trajectories in Cartesian and Poincaré-Birkhoff coordinates for a circular billiard table is shown in Fig. 2.40.

**Figure 2.40** – Three different orbits are generated for a circular billiard, showing periodic (a) and non-periodic orbits (b). The orbits are plotted in the Poincaré map (c). Periodic orbits appear as a finite set of points, while non-periodic orbits lie on an invariant curve in the phase space. In the Cartesian coordinate system, non-periodic orbits lead to caustics, i.e. curves for which trajectories remain tangent after consecutive reflection.
2.7 Dynamical Billiard

One can derive the equivalence of a dynamical billiard to the ray propagation in an optical multi-mode fibre in two ways. The intuitive way is to further simplify ray-tracing methods by assuming the fibre to be ideal along the z-direction. In fact, this is an appropriate assumption, as modern fibres do have very tight tolerances on the diameter deviations in the order of 1%. It should also be noted that in principle, billiard systems could account for such variations by adding a time dependence to the potential \( V = V(x, y, t) \). For simplicity, we will however not include a time dependent potential for the following considerations. Total internal reflection at the core-cladding transition turns into specular reflection at the billiard boundary.

A second way to show the equivalence of ray propagation in a weakly guiding multi-mode fibre and a billiard system is given in Doya et al. (2002). The authors showed that the Helmholtz equation (Eq. (2.11)) corresponds to the pseudo time-dependent Schrödinger equation, where instead of a dependency on time \( t \) the equation depends on the coordinate \( z \). By deriving the eikonal equation in paraxial approximation, i.e. for small angles between the optical axis of the fibre and the incoming light, they showed that the wave propagation of light within the fibre can be reduced to the dynamics of a billiard.

In order to investigate the dynamical properties of a billiard system, it is useful to follow the time evolution of a set of initial points. Usually, a small region in the phase space is selected as an initial set. The behaviour of this region under the mapping \( T \) is used to reveal the underlying dynamics and to distinguish between regular and chaotic systems. Here, we pick our initial values in a slightly different way, in order to make the simulations similar to the experiments in Section 2.5.1. We pick a location \( r_0 \) within the region \( \Omega \) and create initial values by choosing random vectors in a small disk around \( r_0 \) with random directions. Both location and direction are chosen from uniform distributions over the following intervals:

\[
\phi_{\text{initial}} \in [0, 2\pi] \quad , \quad r_{\text{initial}} \in D_{d/2}(r_0) \quad ,
\]

(2.27)

where \( D \) is the disk with radius \( d/2 \) around \( r_0 \). For a large number of initial values, this is equivalent to imaging a small spot with diameter \( d \) onto the fibre with a homogeneous far field, as it was done in the raytracing simulations and experiments in Section 2.5.1.

---

\(^4\)A multi-mode fibre is called weakly guiding if \( (n_1 - n_2)/n_2 \ll 1 \), where \( n_1 \) and \( n_2 \) are the refractive indices of the core and the cladding, respectively. All standard silica fibres are weakly guiding fibres.
2 Optical Fibres in High-Resolution Spectroscopy

2.7.2 Circular, Elliptical and Oval Boundaries

The simplest boundary curve is a circle with

\[ r(\varphi) = R \quad . \]  \hspace{1cm} (2.28)

For a circular billiard, Eq. (2.23) is integrable, as energy and angular momenta are conserved. The motion is regular for any initial condition. On the Poincaré map, any orbit lies on a horizontal line as the angle \( \alpha_n = \alpha_0 \) stays constant. Skew rays, i.e. rays that do not pass the axis of the fibre, show a clearly visible caustic or are periodic (see Fig. 2.40b). Each skew ray creates such a pattern and when many (random) start-values are used, the superposition of many caustics generates a ring whose radius depends on the initial location \( r_0 \), as shown in Fig. 2.41. The Poincaré map in Fig. 2.41 immediately reveals the two empty horizontal regions above \( p = 1/\sqrt{2} \approx 0.7 \) and below \( p = -1/\sqrt{2} \) that are a direct consequence of the starting value \( r_0 \).

![Figure 2.41](image)

**Figure 2.41** – A circular billiard with \( r_0 = (0.5, 0.5) \), 200 random initial values, and 100 iterations. **Left:** The superposition of caustics gives rise to a ring-like structure. The disc of initial values is indicated by the black circle. **Right:** The Poincaré map shows regions that are not covered, depending on the specific starting values.

The integrability of the motion and the existence of caustics imply a strong coupling of initial conditions and the resulting orbit pattern. This behaviour can therefore be seen as an explanation for the bad scrambling properties of a circular fibre.

Elliptic billiards show a similar behaviour and are also integrable. Although their caustics look different, their billiard motion remains regular due to their integrability. Therefore, elliptic fibres are not a suitable candidate for a fibre with good
scrambling properties. This observation puts into perspective the suggestion of using elliptical fibres in astronomy by Allington-Smith et al. (2012). However, the elliptical form is beneficial in other applications: fibres with a small elliptical core are commonly used for polarisation maintaining fibres.

Elliptical and circular boundaries are the only twice differentiable boundaries known to be integrable.\(^5\) In general, oval forms show a mixture of chaotic and regular motion that depends on the initial condition (Lopac et al., 2002).

In summary, neither elliptical nor oval billiards seem to be a good choice if a homogeneous and stable output illumination is desired.

### 2.7.3 Polygonal Boundaries

Billiards with polygonal boundaries are somewhat special, because their boundaries are not a smooth function that is twice differentiable. We will restrict our view to rational polygons, where each angle is a rational fraction of \(\pi\). In such billiards, each trajectory has only a finite number of directions. Therefore, the phase space is at most two-dimensional (Richens and Berry, 1981). Zemlyakov and Katok (1975) showed that unlike the case of integrable systems, rational polygons possess an invariant surface whose topology differs from a torus. A few exceptions exist, such as rectangles and special triangles, which are integrable systems. For all other regular polygons, no unique action-angle variables can be defined.\(^6\) This is the reason why regular polygonal systems are called pseudointegrable (Richens and Berry, 1981).

The dynamics in a regular polygon are non chaotic in the sense that the Lyapunov exponent is zero (Chernov and Troubetzkoy, 1998), i.e. neighbouring trajectories do not exponentially diverge.\(^7\) By looking at Fig. 2.42, one sees that \(\Omega\) is homogeneously filled and that there are no visible caustics as in the case of a circular boundary. In fact, one property of rational polygons is that they are dynamically optimal. This means that any trajectory is either periodic or it uniformly covers the table (DeMarco, 2011).

---

\(^5\)This statement is referred to as the Birkhoff-Poritsky conjecture. As of today, it has not formally been proven yet.

\(^6\)Action-angle variables are a set of coordinates that provide a convenient way to describe integrable systems in classical mechanics.

\(^7\)The Lyapunov exponent measures how fast two orbits starting with slightly different initial conditions diverge. In a chaotic system, the distance of orbits increases exponentially.
2 Optical Fibres in High-Resolution Spectroscopy

Figure 2.42 – Square billiard with \( r_0 = (0.25, 0.25) \), 100 random initial values, and 100 iterations. **Left:** The region \( \Omega \) is homogeneously filled by the orbits. The disc of initial values is indicated by the black circle. A single, non-periodic trajectory (red lines) illustrates how the whole billiard table gets successively filled. However, in any rational polygon the number of directions, i.e. different angles \( \alpha \), for each trajectory is limited. **Right:** In the Poincaré map, the trajectory does not lie on an invariant curve, but it rather ‘jumps’ in four planes, which is a consequence of the limited number of directions.

2.7.4 Boundaries with Chaotic Dynamics

As a matter of fact, boundaries that are integrable and imply regular motion constitute an exception. In general, the motion in an arbitrary boundary shows a mixture of regular and irregular behaviour (see Chapter 3 in Korsch et al., 2008). These systems are called chaotic. Examples for boundaries that are well studied in the literature and exhibit chaotic dynamics are the stadium, i.e. two half circles connected by straight lines, the oval, and the D-shaped boundary. We will focus on the latter one, whose dynamics are extensively studied both from the classical and the quantum perspective in Ree and Reichl (1999). The dynamics of the D-shaped billiard depends on the parameter \( w \), which equals the width of the cut divided by the radius, see Fig. 2.43 and Fig. 2.44. Only when \( w \in \{1, 2\} \), i.e. the boundary is a full or a half circle, the system is integrable (Ree and Reichl, 1999). For \( 1 < w < 2 \), the dynamics are chaotic and ergodic (Bunimovich, 1979), which means that the iterations of almost any initial value will asymptotically come arbitrarily close to any point in
2.7 Dynamical Billiard

phase space. Put in other words, the iterates of (almost) any initial condition will fill the whole phase space. Also, chaotic dynamics are always characterized by an extreme sensitivity of the orbit to initial conditions due to the positive Lyapunov exponent. At first glance, a high sensitivity to initial conditions seems to be an undesirable property in the context of scrambling. However, a positive Lyapunov exponent can also be related to an average loss of information during the process of iteration, i.e. the system loses the memory of its initial conditions (Korsch et al., 2008). One could argue that a perfect scrambler would have no information of its initial conditions any more. From this perspective, a fibre which induces chaotic ray motion seems highly suitable.

So far, we have restricted our view to the classical dynamics of billiards. However, fibres clearly show quantisation effects, i.e. modes, and a proper treatment should account for this. Doya et al. (2002) investigated D-shaped fibres theoretically and experimentally and they indeed found that the mode statistics in these fibres differ significantly from the ones in a circular fibre. They showed that the intensity of generic modes in a D-shaped fibre is uniformly distributed over the fibre cross section. However, there are also few modes that show a higher intensity along the classical trajectories of periodic orbits. These modes are called scars and are of special interest for applications dealing with fibre amplification (Michel et al., 2009).

Figure 2.44 – A billiard with a D-shaped boundary with $w = 1.5$, $r_0 = (0.25, 0.25)$, 100 random initial values, and 100 iterations. Left: The billiard table is homogeneously filled by the orbits. The disc of initial values is indicated by the black circle. Right: Due to the chaotic dynamics of the billiard, a single trajectory (red dots) does not lie on an invariant curve, but will eventually fill the whole phase space.
Concerning spatial scrambling, it remains an open question, whether polygonal or chaotic fibres are better suited. However, from a theoretical point of view, the latter are suitable candidates for superior scramblers. These fibres also potentially differ in their modal behaviour, as in the classical limit, all orbits, particularly all periodic orbits, are unstable. Perturbations such as seeing variations or fibre agitation could have larger impact on the resulting speckle patterns, which would be beneficial in order to mitigate modal noise. Finally, further experiments that directly address the questions of scrambling and modal noise behaviour of chaotic multi mode fibres are needed.
2.8 Fibres for CARMENES

In total, CARMENES uses 14 fibres for different purposes as described in Fig. 2.45: For each channel, large core fibres (D=400µm) guide calibration light from the two calibration units (CUs) towards the front-end (FE) at the telescope. CARMENES also uses large core fibres (D=600µm) to guide light collected from the 0th order of the échelle gratings towards a photon counter to provide information on the exposure for both arms. The light from the target star is directly split up in the FE into its visual and near-infrared components. In each wavelength range, the star light as well as the calibration light is fed into two 100µm circular fibres with a length of 42m (VIS) / 50m (NIR) each. Before entering the vacuum tank of each spectrograph, the circular fibres are linked to fibres with an octagonal core with a size of 100µm by a FC-FC junction, as these fibres offer the desirable scrambling properties discussed in Section 2.4.4. The fibre type used is CeramOptec Optran WF 100/187P. In the following, the linked circular octagonal fibres will be called
science fibres, as they carry the stellar and calibration light. For both the VIS and the NIR channel, the same type of fibre with a low concentration of hydroxyl (OH) groups (CeramOptec WF) is used, thus providing high internal throughput within the whole spectral range of both instruments. Because of advances in technology that were primarily driven by the telecommunication industry, the light losses in fibres decreased during the past decades to $\sim 0.2 \text{dB/km}$ or $\sim 2.3\%$/km at some wavelength regimes. However, these values are only reached for SMFs. For multi-mode fibres, the losses are about a factor ten higher. Different physical processes contribute to light loss in optical fibres. In the UV, the glass itself begins to absorb light, whereas in the visual wavelength range, the loss process is dominated by Rayleigh scattering up to about $1.6\mu$m. The two bumps that are observed in Fig. 2.46 at $\sim 1.24\mu$m and $\sim 1.39\mu$m are absorption features from OH molecules. For the latter regime, even for such relatively short fibres at CARMENES compared to the ones used in telecommunications, the attenuation increases significantly. However, this occurs in the regime between the J- and H-band, where the atmospheric transmission is also very poor, because of the same physical process of absorption by water molecules. There exist also ultra-low OH fibres with an even more reduced OH concentration that show a better transmission in these wavelength regimes. However, such fibres were not available with the diameters and geometries that are appropriate for CARMENES. Figure 2.46 shows the wavelength dependency of the internal transmission of the VIS and the NIR fibres that is calculated from data provided by CeramOptec. The fibre transmission lies around 93\% for most part of the wavelength range of CARMENES.

![Figure 2.46](image)

**Figure 2.46** - **Left:** Attenuation of a typical CeramOptec Optran WF fibre, based on data provided by CeramOptec. The bumps around $1.4\mu$m are due to light absorption by OH molecules. **Right:** Internal transmission of the CARMENES VIS and NIR fibres calculated from this data.
2.8 Fibres for CARMENES

2.8.1 Mechanical Considerations

The operating temperature of the science fibres for CARMENES ranges from about −20°C to 40°C at the FE and from room temperature to −135°C at the NIR channel. In case of the latter regime, not all protective coatings for fibres are applicable any more. Different coatings do not only have various effects on the fibres’ mechanical properties, they may also effect their optical properties (Padilla Michel et al., 2012). For the circular part of the science fibres, CeramOptec Optran WF 100/125A fibres are used. This is a fused silica fibre with a 100µm core, a 125µm fluorine-doped fused silica cladding, and a 250µm acrylate coating that can withstand the prevalent temperatures at the FE. The octagonal fibre has a polyimide coating, which is also applicable to the NIR temperature range.

Almost all interfaces in CARMENES use FC-connectors with ceramic ferrules, because they have tighter tolerances on their inner diameters as well as better concentricities compared to metal ferrules or standard SMA connectors. The most precise ceramic ferrules have tolerances of only ∼1µm on bore diameter and concentricity. FC connectors also have a key that prevents the fibre from rotating when plugged into a receptacle. Their ferrules are spring loaded and their respective receptacles have a limit stop. Therefore, the fibre always ends up at the same angle and in the same plane when being reconnected. This allows for very good mechanical repeatability and should minimize realignment of optics in case of any fibre replacement, e.g. in the FE. It is also very important to impose tight tolerances on inner diameter and concentricity at the junction of the circular and the octagonal fibres. Figure 2.47 shows the throughput of various fibre-fibre junctions as a function of their lateral mismatch. A prototype of the circular octagonal FC-FC junction had a throughput of ∼94%. However, one should note that this prototype used metal connectors that were drilled out to the cladding diameter of the fibre. The tolerances are expected to be worse compared to high precision ferrules, which were not available at that time for the particular diameter of the octagonal fibre. Using high-precision custom Zirconia ferrules, it is expected that the light loss of the final junction will be well below 5%.
2 Optical Fibres in High-Resolution Spectroscopy

Figure 2.47 - Geometrical fibre mismatch and corresponding loss in throughput for a circular-circular, a circular-octagonal, and an octagonal-octagonal fibre junction. The diameter of the circular fibres is set to 100µm, for the octagonal fibre the inscribed circle is also set to 100µm. This corresponds to the sizes of the Optran WF 100/125A and octagonal Optran WF 100/187P, respectively. The ‘best’ and the ‘worst’ case corresponds to the different orientation of the two fibres with respect to each other and denote the orientations with the lowest and the largest light loss, respectively. For circular fibres, there is only one case due to symmetries.

2.8.2 Fibre Preparation

For CARMENES, the circular and octagonal science as well as the reference fibres have all been assembled and prepared in-house. At the time this thesis was written, they had either already been built or were still under construction. As stated before, both the assembly and preparation of the fibres influences their optical properties such as FRD. By doing the relevant work in-house, I could extensively investigate important design aspects of the fibre feed. One important aspect was that the physical properties of the glue used to permanently adhere the fibre into the ferrules, but also the exact method how it is applied, influence the final FRD of the fibre to some degree. This is largely due to the stress induced by the physical strain occurring when the thin layer of glue around the fibre shrinks upon hardening. From this perspective, it is beneficial to keep the glued-in length as short as possible and the layer of glue as thin as possible. Further, glue that stays softer and has lower shrinkage rates is preferred. Conversely, the glue needs to be hard enough
after curing to enable polishing the fibre without 'smearing' its surface. In practice, the lower limit for glue film thickness is about 1µm to 2µm for the gap between fibre and ferrule. Below this range, it becomes hard to insert the fibre into the ferrule, as the tolerances on the diameter of the fibre as well as the bore of the ferrule are in the same regime.

Two methods for fibre end-face preparation exist: Polishing and cleaving. The first technique applies a set of successively finer grades of lapping sheets to polish the fibre, which is glued into a ferrule or a connector, to the surface roughness being requested. This method is rather time consuming, because after each polishing step, an inspection of the surface is highly recommended in order to guarantee good optical quality. The polishing process causes sub-surface damages that can propagate from the surface down the fibre. It is believed that these damages propagate to a depth that is roughly three times the depth of the biggest defect on the fibre surface. This makes it inevitable to remove sufficient material after each polishing step. Figure 2.48 illustrates the polishing steps of one of the CARMENES circular science fibres that I prepared. Care was taken to produce surfaces with low surface roughness. However, in case of the circular science fibres of CARMENES, this is not particularly critical. Both ends of the fibres are immersed to the other components, i.e. to a rod lens in the FE and to the octagonal fibre at the other end, by applying an index matching gel that reduces irregularities on the fibre surface. The second method that is used for fibre end face preparation is cleaving. Hereby, the fibre is put under axial tension and is slightly scratched from the side. Depending on the exact tension and fibre parameters, the glass cracks in a plane that is perpendicular to the optical axis of the fibre. In principle, cleaving is easier and faster than polishing, and when utilising sophisticated cleavers, the repeatability of high-quality end faces is also very good. For CARMENES, cleaving can only be used for the end faces of the octagonal fibres at the spectrograph side, because all other fibres are in physical contact with other optical components. However, cleaving of non-standard fibres such as the octagonal fibre used for CARMENES is not straightforward, because the cleaving parameters are not known a priori. Using wrong parameters may lead to a bad cleave with chips or lips on the edges of the fibre. In addition, the end face may show zones with high surface roughness, called mist, or with even larger irregularities, called hackle. Even when the right cleaving parameters are applied, these problems might still occur, for instance when there is intrinsic stress from the manufacturing process frozen within the fibre.

Both are equally terrifying.
Polishing process of a 100µm circular fibre for CARMENES. End-face images at \( \sim 400 \times \) magnification after polishing with a diamond lapping paper with a grid size of 6µm, 3µm, 1µm, and 0.5µm (from left to right).

Images of the end face of the CeramOptec 100µm octagonal fibre after cleaving. The fibre was cleaved at two different axial tensions: 280g (left) and 415g (right). The nick at the cleave initiation point is larger for the lower tension, while for the higher tension, the contamination with mist/hackle is more pronounced.

In order to find out if cleaving is an option for the octagonal fibre, I removed the polyimide coating and cleaved the fibre under different axial tensions. I ran these experiments at the fibre preparation facilities at InnoFSPEC in Potsdam, who kindly let me use their professional Vytran LDC-400 fibre cleaver. The resulting end faces of the octagonal fibre are displayed in Fig. 2.49 for two different tensions. A lower tension requires a deeper nick to initiate the propagation of the crack across the surface, as illustrated by the left image. The nick at the cleave initiation point is an undesirable feature of this technique. However, it does most probably not affect the optical properties a lot, as it only appears in the secondary cladding of the fibre. For higher fibre tensions, the crack propagation is initiated with a smaller nick,
causing yet other inconveniences. The greater the tension, the higher the energy of the shock wave that propagates across the fibre surface. For all tensions applied, the cleaved surface showed some mist/hackle regions at the opposite side of the cleave initiation point (see Fig. 2.49). This is a consequence of how the shock wave propagates across the fibre and of how it gets reflected at the fibre boundary. In the right image, a higher tension was applied, yielding a more pronounced mist zone that spreads to a greater extend into the core. In addition, the mist zone is redirected when penetrating the secondary cladding. This might be an indication that the fibre suffers from locked-in stress.\footnote{Comparison with images from http://www.ardenphotonics.com/products/cleaves.htm} Further test have to clarify which cleaving parameters deliver better performance from an optical point of view.

2.9 Conclusion and Future Work

In this part of the thesis, I discussed the experiments and simulations that I conducted in order to find the optimal solution for the fibre link of CARMENES. I devised an optical test setup and implemented a strategy to reliably measure optical properties of fibres. The measurement procedure has been automated as far as possible to minimize systematic errors caused either by time dependent processes, such as intensity fluctuations of the light source, or by mechanical misalignment of the setup due to temperature variations.

I measured and compared the scrambling and FRD behaviour of various types of fibres. My measurements confirmed that normal circular fibres do not provide a sufficient image stabilization for CARMENES. I investigated the features of fibres with non-circular cross section and verified their superior NF scrambling. In contrast to some previous studies in the literature, I found that the square and the octagonal fibre have a relatively homogeneous FF under normal fibre illumination conditions. Also, their FRD behaviour does not show an inferior performance compared to that of circular fibres. These results are in agreement with more recent publications on non-circular fibres. Overall, I found that combining fibres with different cross sections, in particular circular and octagonal fibres, provides both a good scrambling in NF and FF, while having a tolerable FRD loss. A careful mechanical design keeps the coupling losses at the fibre junction at a minimum. The final manufacturing of the fibres that will be used for CARMENES is still ongoing. However, the prototypes already fulfilled the project’s requirements.
In addition, I simulated the scrambling behaviour of fibres using the commercial raytracing software ZEMAX in order to theoretically investigate fibres with cross sections that have not been available for testing. My simulations of circular fibres show a clear qualitative agreement with my measurements, which justifies the approach to simulate multi mode fibres by raytracing. The results for polygonal shaped fibres indicate that the lower the number of their sides, the better their scrambling. Finally, I analysed the impact of variations in the FF of a fibre on RV precision by raytracing simulations. For both CARMENES spectrographs, I estimated the sensitivity of the position of a spectral line due to illumination changes by calculating the barycentric shifts of the spot diagrams when the pupil illumination is changing. The simulations showed that for FF variations that are expected to be typical at a telescope, the shifts for single lines are smaller than 1 m/s to 2 m/s. The barycentric shifts strongly depend on the optical aberrations and therefore also on wavelength. In case of the CARMENES spectrographs, the shifts along an échelle order typically show a reflection symmetry with respect to the centre of the order. Therefore, averaging over many spectral lines along an order will further reduce the systematic shifts that are caused by these FF variations.

Although a fair number of publications in the last few years investigated the scrambling properties of fibres with various cross sections, there is, up to my best knowledge, no such study that gives a satisfying theoretical explanation for the different scrambling performances of these fibres. I suggested that the theory of dynamical billiards provides a suitable theoretical framework to explain and further investigate the scrambling and modal noise properties of multi mode fibres. I concluded that fibres whose underlying geometry induce chaotic ray motion will potentially help to further reduce the impact of insufficient scrambling and modal noise on radial velocity precision.
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

This chapter presents the work on an actively stabilized Fabry-Pérot etalon for spectrograph calibration. Many of the results are published either in Schwab et al. (2014) or Gurevich et al. (2014). In collaboration with my co-authors, I developed a novel etalon locking concept. I was the main responsible for the experimental setup in the optical laboratory at the Landessternwarte Heidelberg. I implemented all of the relevant software for locking and performed the data analysis.

3.1 Introduction to Wavelength Calibration

3.1.1 Emission Lamps

Wavelength reference spectra are typically generated by using a hollow-cathode lamp (HCL). This lamp consists of an anode and a cathode made of a specific material that is enclosed by a glass tube filled with a buffer gas. By applying a high voltage between the anode and the cathode, the buffer gas is ionized, resulting in gas atoms being accelerated onto the cathode sputtering off atoms. Both gas and cathode atoms get excited by collision with each other. When decaying into lower states, they emit light in a rich spectrum of emission lines, whose absolute wavelengths can be measured with high precision in the laboratory.

In astronomy, two methods for calibrating spectrographs with HCLs are in use. One is to take an emission lamp spectrum before and after an observation and compare it to the intermediate science spectrum. In most cases, the short term stability of the spectrograph ensures sufficiently small drifts between the calibration and the science image. However, a major drawback of this method is the overhead generated by the consecutive recording of calibration and science spectra that reduces the number of possible targets per time unit. To overcome this problem, Baranne et al. (1996) introduced a second fibre that allows for simultaneous measurement of science and calibration light. The fibre is placed such that each order of the calibration spectrum
lies in between two consecutive orders of the science spectrum. As a consequence, this solution demands a higher cross dispersion and a larger detector area. In the visible part of the spectrum, thorium argon (ThAr) HCLs, which are commercially available at relatively low cost, are commonly used. For a typical ThAr spectrum see Fig. 3.1.

![Figure 3.1 - Part of a ThAr spectrum measured with the laboratory échelle spectrograph that is currently being built at the Landessternwarte Heidelberg (see also Section 3.6). Spectral lines have very different brightness and their spacing is also irregular.](image)

The thorium lines provide an absolute wavelength solution for the échelle spectrum. However, the spectral properties of ThAr lamps are far from being ideal for the purpose of calibration: the lines are unevenly distributed in wavelength, they have very different brightness, and most lines are blended at typical spectrograph resolution (Lovis and Pepe, 2007). Figure 3.2 shows the distribution of the closest neighbour from the ThAr-catalogue provided by Lovis and Pepe (2007) with blended lines removed. The catalogue covers the region from 378 nm to 691 nm and contains about 8400 lines. The figure also shows the distribution of the relative intensity of the lines illustrating the large brightness differences among the lines. In the red part of the spectrum, ThAr lamps become less suitable due to contamination from very bright argon lines. This is illustrated in Fig. 3.3, which shows the larger closest neighbour separations as well as the unequal brightness distribution of thorium and argon lines. In addition to the sparse wavelength coverage in the near infrared, the total lifetime of HCLs is only a few hundred hours (Murphy et al. (2007); SJ Juniper ThAr lamp manual) and their spectra change with age (Mayor et al., 2009). Different lamps, even when obtained from the same supplier, may also show significant deviations among their respective spectra. In case of a long term survey such as CARMENES, these effects necessitate the use of a set of master and slave lamps.
3.1 Introduction to Wavelength Calibration

Figure 3.2 – Distribution of the closest neighbour separation between lines (left) and relative intensity of ThAr lines from 378 nm to 691 nm (right). Data taken from Lovis and Pepe (2007).

Figure 3.3 – Distribution of the closest neighbour separation between lines (left) and relative intensity of ThAr lines from 691 nm to 5804 nm (right). Data taken from Nave et. al, http://www.nist.gov/pml/data/th-arlamp/

that are calibrated to each other on a regular time base (Quirrenbach et al., 2010). The slave lamps are considered for a daily usage, while the master lamps are only sparingly used, thereby preventing the lamps from fast ageing. This method complicates the usage of ThAr lamps and is also time consuming, yet it allows for reliable long term calibration. Additionally, due to regulations concerning the radioactive and toxic thorium, the further commercial availability of high quality ThAr lamps is put in jeopardy.
Nevertheless, ThAr lamps are the primary calibration source for the HARPS spectrograph and Lovis and Pepe (2007) report a calibration precision with the simultaneous calibration method of 20 cm/s. They are not a good option for the near infrared though, where more recently other emission lamps have been used for calibration. Here, uranium-neon lamps are good candidates (Ramsey et al., 2010), since they provide a larger number of lines from 900 nm to 1300 nm. However, these lamps suffer from similar problems regarding line spacing and brightness, limited lifetime, and changing spectral properties over time.

Boesch et al. (2014) are investigating electrodeless microwave discharge lamps, where CN molecules produce a feature rich spectrum with many lines between 1000 nm and 2000 nm. Yet, the practicability of these lamps for astronomical spectrograph calibration and their durability still needs to be further studied.

3.1.2 Gas Absorption Cell

The second well established method of calibrating a spectrograph is to pass the starlight through a gas cell that acts as a transmission filter. The cell imprints known, stable, and sharp absorption lines onto the stellar spectrum. Therefore, any illumination-dependent effects are eliminated, since the starlight and the calibration light always use the same optical path. For slit-fed spectrographs, a radial velocity (RV) precision in the meter per second regime is possible, as this method relaxes the stability requirements for the spectrograph and its guiding system. As no interleaved orders for the calibration light are used, gas cells also save detector area. However, by introducing a gas cell in the light path of the telescope, the throughput is reduced by around 15%.

Campbell and Walker (1979) used heated cells filled with hydrogen fluoride (HF). They have deep, but few lines for calibrating. Due to its toxic and corrosive nature, HF has to be handled very carefully. As an alternative, iodine cells can be used, as they provide a very large number of sharp lines in the range 500 nm to 600 nm (Butler et al., 1996; Marcy and Butler, 1992). The narrow bandwidth of the iodine cell limits its applicability to bright targets that are rich in spectral features in the visible part of their spectrum. In order to reach the same RV precision, the signal to noise ratio (SNR) of the obtained spectra must be at least twice as large as the one needed for simultaneous ThAr calibration (Bouchy et al., 2001). Still, RV precisions

---

1. This value is taken from the CHIRON efficiency test document http://www.ctio.noao.edu/~atokovin/echelle/efficiency_jun2011.pdf
better than 2 m/s has been reported (Tokovinin et al., 2013).
Mahadevan and Ge (2009) suggested to use gas cells in a similar way as emission lamps, i.e. by illuminating them with broadband white light. This would eliminate the SNR problem and increase the throughput of the system. However, now one loses the benefit that the calibration light follows the same optical path as the stellar light. New gas mixtures for absorption cells that provide lines in a wide range in the NIR are also extensively being investigated (Mahadevan and Ge, 2009; Redman et al., 2012; Plavchan et al., 2013; Seemann et al., 2014). In some cases, it is not even necessary to put another gas mixture in the line of sight between the observer and the star, but one can rather use what is already out there: the Earth’s atmosphere. Rotational-vibrational transitions from molecules such as CO₂ provide a set of known fiducial marks that can be used for calibration. The width of the lines, their exact position, and their depth depend on weather conditions. Pressure, temperature, and most importantly wind limit the precision of this method to a few meters per second (Figueira et al., 2010).

3.1.3 Laser Frequency Combs

A laser frequency comb (LFC) is a coherent light source that generates a comb like spectrum of very narrow lines, whose frequencies are stabilized to an atomic clock. This extreme accuracy allows for atomic and molecular transition measurements with unprecedented precision. For their contribution to this field of research, T. W. Hänsch and J. L. Hall were awarded the 2005 Nobel Prize in Physics. During the last few years, LFCs have been adapted to the needs of RV measurements (Steinmetz et al., 2008; Li et al., 2008; Stark et al., 2011). Usually, LFCs have a narrow line spacing around 1 GHz that remains unresolved in astronomical spectrographs. In order to block out most of these narrow lines, Fabry-Pérot etalons have to be used as filter cavities. The design of these cavities is challenging, because they suffer from dispersion effects and need a high finesse value over a large bandwidth. The comb itself also has a limited bandwidth of about one octave and its lines vary substantially in brightness over the wavelength range, which demands additional flattening techniques (Ycas et al., 2012; Probst et al., 2013). Wilken et al. (2010) reported a 2.5 cm/s calibration precision at HARPS, which demonstrates the outstanding performance of a LFC for calibration. Still, the intrinsic precision of a LFC cannot fully be translated into RV precision. This is due to several sources of noise that occur in an astronomical spectrograph, most noticeable modal noise when multi-mode fibres
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

are used (Wilken et al., 2012). LFCs offer the most precise calibration source, their accuracy and long-term stability far surpassing the requirements of RV observations. However, the costs and the high complexity of the systems, which require a significant operational maintenance, limit their applicability to the largest observational facilities.

3.1.4 Etalons

A Fabry-Pérot etalon is an optical resonator that acts as a wavelength filter with a comb-like transmission spectrum. The theory of Fabry-Pérot etalons (FPEs) is discussed in detail in the next section. Wildi et al. (2010, 2011, 2012) developed a FPE for the wavelength calibration of the HARPS spectrograph. They use plane, dielectric coated mirrors with a compensated zero drur spacer in between and control the temperature and pressure of the assembly within tight limits. During one night, they achieved a system stability of $10 \text{cm/s}$. However, the repeatability over a longer timespan remains challenging, mainly because of illumination effects. Halverson et al. (2012, 2014) are developing a fibre Fabry-Pérot (FFP) that is insensitive to illumination effects, at the cost of a much larger temperature sensitivity. Still, a short term stability of about $80 \text{cm/s}$ has been demonstrated at the APOGEE spectrograph and their experiments indicated that the internal precision of the etalon is even better.
3.2 Fabry-Pérot Etalon

3.2.1 Theory

The working principle of a Fabry-Pérot etalon is based on multiple beam interferometry. It is named after Charles Fabry and Alfred Pérot, who developed an interferometer consisting of two partially reflective surfaces (e.g. a glass plate with parallel and semi-transparent surfaces). When the distance \( L \) between the reflecting surfaces remains constant, the device is called a Fabry-Pérot etalon or a cavity. In case of an adjustable distance, the instrument is called a Fabry-Pérot interferometer, which can be used for various spectroscopic measurements (Born and Wolf, 1999). Fig. 3.4 shows a plane parallel plate with refractive index \( n \), surrounded by a medium of refractive index \( n_1 \). Consider an incoming plane wave with wavelength \( \lambda_0 \) that hits the first surface under an angle \( \theta' \). As the plate is semi-transparent, one part of the beam is reflected, while the other enters the plate under the angle \( \theta \) according to Snell’s law. At the backside of the plate, the beam is divided in the exact same manner, as indicated in Fig. 3.4.

![Diagram of a Fabry-Pérot etalon](image-url)
The optical path difference between two consecutive reflections is $2\lambda_0 \cos(\theta)$ and hence the phase difference can be calculated as

$$\delta = \frac{4\pi}{\lambda_0} \ln \cos(\theta).$$  \hspace{1cm} (3.1)

Following the notation in Born and Wolf (1999), the complex amplitude of the electric field of the incident wave is denoted by $A^{(i)}$. Let $r$ be the reflection coefficient and $t$ the transmission coefficient at the first surface and let $r'$ and $t'$ be the corresponding coefficients at the backside of the plate. Then, the complex amplitudes of the reflected waves are obtained as

$$rA^{(i)}, \quad tt' rA^{(i)} e^{i\delta}, \quad tt' r^3 A^{(i)} e^{i(2\delta)}, \quad \ldots, \quad tt' r^{2p-3} A^{(i)} e^{i(p-1)\delta},$$  \hspace{1cm} (3.2)

and in the same way, the transmitted waves are given by

$$tt' A^{(i)}, \quad tt' r^2 A^{(i)} e^{i\delta}, \quad tt' r^4 A^{(i)} e^{i(2\delta)}, \quad \ldots, \quad tt' r^{2(p-1)} A^{(i)} e^{i(p-1)\delta}.$$  \hspace{1cm} (3.3)

The reflection and transmission coefficients can be calculated by the Fresnel formulae, but here only the following relations are needed:

$$tt' = T, \quad r^2 = r'^2 = R, \quad T + R = 1,$$  \hspace{1cm} (3.4)

where $T$ and $R$ denote the transmissivity and reflectivity of the plate surface.

The resulting electric field of all $p$ superposed waves of Eq. (3.2) and Eq. (3.3) is given by the sum of all complex amplitudes. This sum takes the form of a geometrical series and can for $p \to \infty$ be reduced to:

$$A^{(r)} = \frac{(1 - e^{i\delta}) \sqrt{R}}{1 - Re^{i\delta}} A^{(i)}, \quad A^{(t)} = \frac{T}{1 - Re^{i\delta}} A^{(i)}.$$  \hspace{1cm} (3.5)

The intensity of the reflected light is now given by

$$I_r = A^{(r)} A^{(r)*} = I_0 \frac{4R \sin^2(\delta/2)}{(1 - R)^2 + 4R \sin^2(\delta/2)}. $$  \hspace{1cm} (3.6)

Similarly, the transmitted light is determined by

$$I_t = A^{(t)} A^{(t)*} = I_0 \frac{1}{1 + \frac{4R}{(1 - R)^2} \sin^2(\delta/2)} = I_0 \frac{1}{1 + F \sin^2(\delta/2)},$$  \hspace{1cm} (3.7)
where the parameter $F$ is called the coefficient of finesse. It solely depends on the reflectivity $R$ and is given by

$$F = \frac{4R}{(1 - R^2)}.$$  \hfill (3.8)

The transmission function of an etalon is of special interest. It has its maxima when the phase shift $\delta$ is an integer multiple $m$ of $2\pi$:

$$\delta = 2\pi m = \frac{4\pi}{\lambda_0} nL \cos(\theta).$$  \hfill (3.9)

An important characteristic of a FPE is its free spectral range (FSR). It is defined as the distance in wavelength between two adjacent maxima of the transmission function. The free spectral range (FSR) can be derived from Eq. (3.9) as

$$\text{FSR}_\lambda = \frac{\lambda_0^2}{2nL \cos(\theta)},$$  \hfill (3.10)

or when expressed in terms of frequency:

$$\text{FSR}_\nu = \frac{c}{2nL \cos(\theta)}.$$  \hfill (3.11)

The FSR is inversely proportional to the distance of the two reflective surfaces. Aside from the FSR, the so-called finesse $\mathcal{F}$ is the second characteristic quantity describing an etalon. It is defined as the ratio of the FSR to the full width half maximum (FWHM) of a peak:

$$\mathcal{F} = \frac{\text{FSR}_\lambda}{\text{FWHM}_\lambda} = \frac{\text{FSR}_\nu}{\text{FWHM}_\nu}.$$  \hfill (3.12)

The transmission functions for etalons with different finesses and FSRs are plotted in Fig. 3.5. For a sufficiently high reflectivity of the mirrors, the finesse can be calculated as

$$\mathcal{F} = \frac{\pi \sqrt{F}}{2} = \frac{\pi \sqrt{R}}{(1 - R)}.$$  \hfill (3.13)

This close relation between $F$ and $\mathcal{F}$ is the reason why the distinction between both quantities is often flawed in the literature. The wavelength dependency of the FSR, see Eq. (3.10), has two interesting consequences for the use of etalons in échelle
spectrographs. First, since the resolving power of an échelle is constant, the distance of the peaks in the échelle spectrum measured in pixels, for example, is linearly increasing with wavelength. In order to be able to resolve the peaks throughout the whole wavelength range of the échelle spectrograph, they need to be resolvable at the shortest wavelength. At longer wavelengths, there are fewer lines available for calibration. The second consequence is that for an etalon with constant finesse, the width of an etalon peak in an échelle is also a linearly increasing function of $\lambda$. In case it is desired that a single peak is spectroscopically unresolved, the finesse has to be chosen such that the peak width is smaller than a resolution element of the échelle over the whole wavelength range. All formulas shown above are derived for nearly monochromatic light in a perfect FPE. However, these derivations are based on the slightly inaccurate assumption that the coating forming the mirror is infinitely thin and has no absorption. As noted by e.g. Monzón et al. (1991), it is more appropriate to consider the FPE to be composed of three media separated by two thin films.
By using transfer matrix methods, the authors showed that the above formulae hold when \( \delta \) is replaced by

\[
\delta = \frac{4\pi}{\lambda_0} \ln \cos(\theta) + 2\Phi ,
\]

(3.14)

where \( \Phi \) is the phase change on internal reflection. Additionally, \( R \) should be interpreted as the reflectivity for internal reflection, yielding

\[
R + T + A = 1 ,
\]

(3.15)

where \( A \) is the amount of light absorbed by the thin film. This is of particular importance in case the mirror is formed by a semi-transparent metal layer, where absorption can be significant compared to the transmission. Finally,

\[
I_t = I_0 \left( \frac{T}{1 - R} \right)^2 \frac{1}{1 + F \sin^2(\delta/2)} ,
\]

(3.16)

\[
= I_0 \left( \frac{1 - R - A}{1 - R} \right)^2 \frac{1}{1 + F \sin^2(\delta/2)} .
\]

(3.17)

The effect of absorptive layers is thus reflected by a reduction of the peak transmission by a factor of

\[
T_{\text{peak}} = \left( \frac{1 - R - A}{1 - R} \right)^2 .
\]

(3.18)

It is worth noting that this simple result does not hold for the intensity profile of the reflected part of the light introduced in Eq. (3.6). Here, the finite thickness and the absorption lead to an asymmetry in the resulting function. Due to scattering effects, the transmission function is found to be of the form as in Eq. (3.16) even in case of absorption-free dielectric coatings (Born and Wolf, 1999). Section 3.2.3 presents some simulations of metal coatings for FPEs that are of interest for astronomy.

A second simplification was made by assuming nearly monochromatic light, which neglects all dispersion effects. Yet, for real etalons, \( n_1, n \) as well as the additional phase shift \( \Phi \) in Eq. (3.14) depend on wavelength. Dispersion effects lead to a varying line spacing, which prevents the easy use of FPEs as absolute wavelength references. A priori, the exact position of a transmission line is not known. It could still be measured though, by comparing it with known lines from an emission lamp or a laser frequency comb, or by using a Fourier transform spectrometer.
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

3.2.2 Types of Etalons

There are various types of FPEs with different optical properties. This subsection briefly describes the differences between the two most commonly used etalon types: the confocal and the plane-parallel FP. As many of the results presented later are obtained with fibre Fabry-Pérots, these devices will be described in more detail.

Confocal Etalons

A confocal Fabry-Pérots (CFPs) consists of two spherical mirrors with equal radii of curvature $r$ that are separated by $L = r$. The transmission function is also of the form in Eq. (3.16), but the phase difference $\delta$ for light of normal incidence is

$$\delta = \frac{2\pi}{\lambda_0} \left( 4L + \frac{\rho^4}{r^3} \right),$$

(3.19)

where $\rho$ is the off-axis distance of the incoming light, see Hercher (1968). The FSR, i.e. the distance between the interference maxima, is given by

$$\text{FSR}_{\nu}^{\text{CFP}} = \frac{c}{4L + \frac{\rho^4}{r^3}},$$

(3.20)

which differs from that of a plane etalon, see Eq. (3.11). CFPs are commonly used in laser physics, as they have several advantages over plane etalons. In general, they are much less sensitive to a tilt of the mirrors, as the optical path difference does not change for spherical mirrors in first approximation. Usually, higher Finesse values can be achieved with CFPs compared to plane-parallel ones, because spherical mirrors can be polished to higher precision than plane mirrors (Demtröder, 2008). Astronomical etalons on the other hand, demand a FSR of several GHz in order to yield resolvable lines in an échelle spectrograph. According to Eq. (3.20), this requires small radii of curvature for the mirrors. The optical path difference for a ray increases with distance $\rho$ from the optical axis of the CFP by $\rho^4/(4r^3)$. For large beams, this causes an asymmetric line broadening, which is clearly undesirable. If one requires that the optical path difference is smaller than $\lambda/100$ in order to guarantee sharp, symmetric lines, the beam size for an astronomical etalon becomes very small (on the order of a few hundred micron). Also, to achieve the confocal condition, the mirror separation must match their radii within tight tolerances (a few tenth of microns). All in all, the assembly and alignment of a CFP with a large FSR is challenging.
3.2 Fabry-Pérot Etalon

**Plane-parallel etalons**

The theory of a plane-parallel etalon has already been described in Section 3.2.1. A common macroscopic realisation of such an etalon, often called air-gap etalon, is obtained by optically contacting two mirrors to a spacer with very low thermal expansion. For example, the etalons for HARPS (Wildi et al., 2010) and CARMENES are built this way. Alternatively, the mirror coatings can be put on both sides of a highly parallel substrate. This resulting solid etalon is strictly monolithic and its tuning and stability behaviour differs from an air-gap etalon. While the sensitivity to small tilts of the incoming light is reduced by a factor of \( n \) (the refractive index), the sensitivity to temperature is much higher. This is mostly due to the change of the refractive index with temperature. For more details on the stability and tuning properties of etalons, see Section 3.2.4.

**Fibre Fabry-Pérot etalons**

The fibre Fabry-Pérot (FFP) is a miniaturised version of a solid plane parallel etalon. Here, the resonator is formed by a piece of single-mode fibre (SMF) that is coated on both sides. The etalon is placed in between two SMFs, either by splicing it onto the fibres or simply by physical contact. For easy handling, the fibre can be glued into a ferrule and used in standard FC-FC sleeves (see Fig. 3.6). Due to the use of SMFs,

![Figure 3.6](image)

**Figure 3.6** - A fibre Fabry-Pérot (FFP) consisting of a ceramic ferrule and a single-mode fibre glued into it. Usually, the fibre does not sit perfectly concentric within the hole of the ferrule, leading to anisotropic adhesive shrinkage that can cause birefringence (left). One of the FFPs that has been built in-house (right). For comparison, a standard integrated circuit is shown next to it.
any alignment sensitivity is mitigated from the system, which allows building very compact and robust etalons. However, FFPs suffer from birefringent effects within the fibre as noted by Halverson et al. (2014). For a birefringent etalon, the optical path length depends on the polarisation state of the incoming light: $n_x L \neq n_y L$, where $n_x$ and $n_y$ denote the refractive index of the fast and the slow axis of the fibre etalon respectively. As a consequence, each transmission peak of the etalon is split into two peaks, with their relative intensities being dependent on the polarisation state of the incoming light. A priori, it is not clear whether a broad band source used for illuminating the etalon has a long term stable polarisation behaviour, as for example the fibre used for feeding the etalon can introduce polarisation that might depend on environmental conditions. For a long term stable FFP, it is therefore crucial to control the polarisation of the light by placing polarising filters either in front or behind the FFP or both.

Birefringence within the etalon can be caused either by residual internal stress of the fibre that results from the manufacturing process or by external forces. By gluing the fibre into a ferrule, asymmetric stress can occur, leading to a polarisation dependent change of the refractive index. Tateda et al. (2007b) derived a model to explain and quantify the birefringence caused by the shrinkage of the adhesive. The authors found a proportional relation between the amount of birefringence $B = \Delta n/n$ and the product of the difference in the coefficient of linear expansion of the fibre and the ferrule material, temperature change, Young’s modulus of the adhesive, and the anisotropy of the adhesive layer width. The authors also measured $B$ for three different adhesives with different viscosities and shrinkage properties, finding that $B$ ranges from 3 to $18 \times 10^{-6}$, and a temperature dependence $\partial B/\partial T$ of about $5 \times 10^{-8}$ K$^{-1}$. For these values, the separation of the two polarisation peaks lies in the order of a few GHz and might be unresolved by an astronomical spectrograph. However, the line barycentre as well as its shape will change if the polarisation state of the incoming light varies over time.

Beside the stress induced by adhesives, other forms of external forces can cause stress. Let

$$\sigma = F/A$$ (3.21)

denote extrinsic stress, where $F$ is the force and $A$ the cross section of the etalon. This stress is for instance induced by mounting the fibre in between two SMFs using physical contact. In order to determine its impact on the refractive indices and
the cavity length, one has to distinguish between axial and radial stress. Following Tateda et al. (2007a), we obtain the relative change of the peak transmission frequency by differentiating Eq. (3.11)

\[ \frac{d\nu}{\nu} = -(dn/n + dL/L) \]  \hspace{1cm} (3.22)

In case the force is purely axial, we get

\[ \frac{dL}{L} = \epsilon_z = \frac{\sigma_z}{E} \]  \hspace{1cm} (3.23)

where \( \epsilon_z \) is the strain in z-direction, \( \sigma_z \) is the stress and \( E \) is Young’s modulus. For the change of the refractive index, Tateda et al. (2007a) derived

\[ \frac{dn_x}{n_x} = \frac{dn_y}{n_y} = \left( \frac{n^2 \sigma_z}{2E} \right) (-p_{11} \mu + p_{12} - p_{12} \mu) \]  \hspace{1cm} (3.24)

Here, \( p_{11} \) and \( p_{12} \) denote the photo-elastic (or opto-elastic) constants of the fibre material, \( \mu \) is its Poisson’s ratio and \( E \) is Young’s modulus of the ferrule. This leads to a total relative shift of

\[ \frac{d\nu}{\nu} = \left( \frac{\sigma_z}{E} \right) \left[ \left( \frac{n^2}{2} \right) (-p_{11} \mu + p_{12} - p_{12} \mu) - 1 \right] \]  \hspace{1cm} (3.25)

Dixon (1967) measured \( p_{11} = 0.12 \) and \( p_{12} = 0.27 \) for fused silica, while \( \mu = 0.16 \). Usually, standard ferrules are made of zirconia, where \( E = 1.9 \times 10^{11} \text{ N m}^{-2} \). These ferrules typically have a 2.0 mm tip diameter, resulting in a peak shift caused by axial pressure (at 780 nm) of

\[ d\nu = 494 \text{ MHz/N} \]  \hspace{1cm} (3.26)

or, when expressed in terms of RV:

\[ \Delta \text{RV} \approx 385 \text{ m s}^{-1} \text{ N}^{-1} \]  \hspace{1cm} (3.27)

To keep the lines of the etalon stable close to 3 cm/s, the force onto the FFP may change no more than \( \sim 0.08 \text{ mN} \), which is comparable to the weight of a large snowflake (on Earth). It is important to note that \( d\nu/\nu \) - hence also \( d\lambda/\lambda \) - remains constant in first order, which means that all etalon peaks move by the same amount on the échelle spectrum.
Pure axial stress causes only a shift of the transmission peak frequency, whereas radial stress actually influences the birefringence properties of the etalon. Tateda et al. (2007a) found

$$B = \sigma_y \left( \frac{n^2}{2E} \right) (p_{11} - p_{12}) (1 + \mu)$$

for unidirectional side pressure $\sigma_y$. In case of fused silica fibres and zirconium ferrules, the birefringence is

$$B = -1 \times 10^{-12} \sigma_y .$$

I built a 6mm FFP in-house (see Section 3.4.3) and measured its birefringence, which is shown in Fig. 3.7. The distance of the two etalon transmission peaks is about 2.09 GHz, which corresponds to $B = 5 \times 10^{-6}$. This is a minimal value if the order of interference of both peaks is the same. The value is in agreement with the $B = 3 - 18 \times 10^{-6}$ measured by Tateda et al. (2007a), which is an indication that both peaks are indeed of the same order of interference. In order to measure the dependency of the peak structure on the polarisation state of the light, I placed the FFP in between a polarisation maintaining (PM) fibre and a standard SMF. I fed laser light through a polariser and a subsequent half-wave plate into the PM fibre, thus allowing for a manipulation of the polarisation direction of the light. Figure 3.7 shows how the intensities of the two transmission peaks vary with changing incoming polarisation. However, I could not fully suppress the double peak structure, which is most probably due to some polarisation cross-talk in the PM fibre. In order to completely get rid of one peak, a high contrast polariser should be inserted directly in front (or after) the FFP (see Halverson et al., 2014).

The applicability of a single mode FFPs is limited to the bandwidth where the fibre operates in the single mode regime. In a SMF, light with shorter wavelengths excites several modes, each of which having a different propagation constant, leading to similar problems like birefringence, while for longer wavelengths the throughput quickly drops. In the visual regime for example, a typical single mode fibre has a bandwidth ranging from 400nm to 680nm (Thorlabs S405-XP). In the near-infrared, a larger absolute bandwidth can be achieved as the relative wavelength change determines the type of guidance, e.g. 970nm to 1650nm (Thorlabs SM980-5.8-125).
Baranov et al. (1989) reported the fabrication of multi-mode (MM)-gradient index fibre (GIF)-FFPs with a finesse up to approximately 50. Using multi-mode fibres in FFPs, could mitigate the bandwidth limitation. However, all of the three GIF-FFPs that I built in-house (with Thorlabs GIF50C and Thorlabs GIF625 fibres) showed an irregular and illumination-dependent transmission spectrum. It therefore remains unclear, whether the optical path-length differences of the different guided modes in other GIFs are sufficiently small to form a stable transmission spectrum.

![Figure 3.7](image_url) - Measured birefringence of a 6mm fibre Fabry-Perot (FP). From the peak distance, the calculated birefringence is about $5 \times 10^{-6}$. The relative peak intensities are manipulated by controlling the polarisation in front of the polarisation maintaining (PM) fibre that feeds the FFP. In an astronomical spectrograph, the peaks cannot be spectrally resolved. However, a changing polarisation would shift the barycentre of the (unresolved) line.
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

3.2.3 Etalon Mirror Coatings

Optical coatings for FPEs are either metallic or consist of a set of different dielectric layers. In laser physics, dielectric coatings are commonly used, as high finesse values can be achieved, while absorption remains negligible. This is of special interest for high power laser applications. Also, dielectric coatings are very durable. Astronomical spectrographs usually demand a large operational wavelength range, while the illumination source operates at low power compared to many laser applications. Both, the HARPS etalon (Wildi et al., 2010) as well as the CARMENES etalon (Schäfer and Reiners, 2012) use dielectric coatings. However, in some applications metallic coatings are of particular interest, as they work naturally over a broad wavelength regime and their phase shift upon reflection varies only slowly with wavelength. In the following, I present some simulations suggesting that in some cases metallic coatings could provide a good alternative for astronomical etalons.

For the simulations, I used the Python library TMM, which calculates light propagation in planar multilayer films as well as in thick films by using transfer matrix methods. I took all refractive index data from the refractiveindex.info database, where data from multiple scientific publications is collected and made publicly available. Figure 3.8 shows the wavelength dependence of the reflectivity for a 50 nm metallic layer and Fig. 3.9 illustrates the corresponding phase shifts upon reflection. Aluminium has the highest reflectivity in the UV, but shows poor performance around 800 nm. Moreover, even thin layers show a high absorption. In the visible range, silver works best and a reflectivity above 90% can be reached. In case of aluminium and silver, one needs to put on a protective layer in order to avoid corrosion. As in the case of dielectric coatings, this additional layer usually introduces dispersion. Gold, which shows a good performance from 600 nm to beyond 10 µm, does not need a corrosion protection and is therefore a good candidate for an astronomical etalon working in the NIR.

One major difference between dielectric and metallic coatings is the relatively high absorption of metals, which limits the layer thickness and therefore the achievable finesse. Fig. 3.10 shows the finesse and the peak transmission as function of layer thickness and wavelength. Here, the peak transmission includes absorption losses and is calculated via Eq. (3.18). The peak transmission as function of finesse shows very different behaviours for silver and for gold.

https://github.com/sbyrnes321/tmm
http://refractiveindex.info/
3.2 Fabry-Perot Etalon

Figure 3.8 – Calculated reflectivity curves of aluminium (Al), silver (Ag), and gold (Au) for a 50nm thick metallic layer for light under normal incidence. The reflectivity is calculated by using the TMM library (see main text) and refractive index data from http://refractiveindex.info/. The bottom figure is a zoom-in of the top one.

Figure 3.9 – Calculated phase shift upon reflection for a 50nm thick aluminium (Al), silver (Ag), and gold (Au) metallic layer. The surrounding medium is air. The phase shift is calculated by using refractive index data from http://refractiveindex.info/.
While for a particular finesse the peak transmission for silver is strongly wavelength dependent, gold shows a rather constant transmission for all wavelengths (Fig. 3.11). This means that for silver, high finesse values in the blue are very hard to achieve, since their peak transmission quickly drops to an impractically low amount. However, finesse is a function of wavelength for metallic etalons, as Fig. 3.11 clearly indicates. By a quirk of nature, the finesse of a silver etalon is approximately linearly increasing with wavelength over the VIS range. This means that the line width measured in resolution elements (i.e. pixels on the detector) remains approximately constant in an échelle spectrograph. Also, silver is naturally well suited for the experiments presented later, where a narrow etalon peak at 780 nm is desirable. The higher finesse values at longer wavelengths result in sharp lines, while the lower finesse at short wavelength still allows for a decent throughput.

In general, one can say that metallic mirror coatings are an adequate choice for astronomical etalons, at least for prototyping. Clearly, light absorption is the major drawback of metal coatings and needs to be taken into account when designing such an etalon. In case of silver, corrosion effects may limit their long-term applicability.
3.2 Fabry-Perot Etalon

Figure 3.10 – Calculated finesse (*top row*) and peak transmission (*bottom row*) of silver (Ag) and gold (Au) etalons as function of layer thickness and wavelength.

Figure 3.11 – Calculated peak transmission as function of finesse for silver (Ag) and gold (Au) coated etalons (*top row*). For different constant layer thickness, the finesse increases with wavelength for both coatings (*bottom row*).
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

3.2.4 Etalon Tuning

Rewriting Eq. (3.9) gives the wavelength of the etalon peaks:

$$\lambda_{\text{peak}} = \frac{2\pi L \cos(\theta)}{\pi m - \Phi}$$  \hspace{1cm} (3.30)

For illumination angles around $\theta \approx 0^\circ$, the variation of the peak position that is due to external parameters can be expressed in terms of a radial velocity error as follows:

$$\Delta RV = c \left( \frac{d\lambda_{\text{peak}}}{\lambda_{\text{peak}}} \right) = c \left( \frac{dL}{L} + \frac{dn}{n} - \frac{d\theta^2}{2} \right)$$ \hspace{1cm} (3.31)

All terms on the right hand side can either be seen as possible control parameters for an actively controlled etalon or as error sources which have to be kept constant to the level of $\sim 10^{-10}$ for a sub m/s stability. It depends strongly on the etalon design, which constitutes the dominant term in Eq. (3.31).

For a FFP with a SMF, the illumination angle dependence is zero, leaving a temperature and a pressure dependence of the length and the refractive index as follows:

$$\left( \frac{d\lambda}{\lambda} \right)_{\text{FFP}} = \left( \frac{1}{L} \frac{\partial L}{\partial T} + \frac{1}{n} \frac{\partial n}{\partial T} \right) dT + \left( \frac{1}{L} \frac{\partial L}{\partial \sigma_z} \frac{1}{n} \frac{\partial n}{\partial \sigma_z} \right) d\sigma_z$$

$$= \left( \alpha + \frac{1}{n} \frac{\partial n}{\partial T} \right) dT + \left( \frac{\sigma_z}{E} + \frac{1}{n} \frac{\partial n}{\partial \sigma_z} \right) d\sigma_z$$

$$\approx \left( \alpha + \frac{1}{n} \frac{\partial n}{\partial T} \right) dT$$ \hspace{1cm} (3.32)

Here, $\alpha$ is the coefficient of thermal expansion (CTE) of the fibre and $\frac{\partial n}{\partial T}$ is often referred to as the thermo-optic coefficient. In the last step of Eq. (3.32), it is assumed that the pressure onto the FFP is constant. Variation in pressure will lead to birefringence and significant peak shifts as described in Section 3.2.2. The setup of the FFP has to be mechanically stable for these effects to be negligible. For a fused silica fibre one has $\alpha_{\text{SiO}_2} = 0.55 \times 10^{-6} \text{K}^{-1}$ and $\frac{\partial n}{\partial T} = 8.72 \times 10^{-6} \text{K}^{-1}$ (Leviton and Frey, 2006), which means that for a solid etalon the major contribution to the peak shift is due to the change of the refractive index. In total, the peak shift in RV is given by

$$\frac{d}{dT} \text{RV} = \frac{d}{dT} \left( \frac{d\lambda}{\lambda} \right)_{\text{FFP}} c = 1.96 \text{km s}^{-1} \text{K}^{-1} = 1.96 \text{m s}^{-1} \text{mK}^{-1}$$ \hspace{1cm} (3.33)
Figure 3.12 – Measurement of the temperature dependency of a 6mm FFP by investigating the step response of the FFP to a 300MHz set-point change. The left axis indicates the relative etalon position with respect to the rubidium hyperfine lines (see next section). The right axis indicates the temperature, which was measured with a PT100 sensor mounted within the copper block that encased the FFP. The derived tuning rate of $2.09\,\text{MHz}\,\text{mK}^{-1}$ is close to the theoretically expected value.

Figure 3.12 illustrates the temperature dependency of one of the self-built FFPs. The tuning rate is derived as $2.09\,\text{MHz}\,\text{mK}^{-1}$, which corresponds to $1.56\,\text{m}\,\text{s}^{-1}\,\text{mK}^{-1}$ and is thus close to the theoretically expected value (cf. Eq. (3.33)). For a stability of $10\,\text{cm/s}$, the FFP has to be stabilized to $\sim 51\,\mu\text{K}$. At this level of precision, measuring and controlling the temperature is challenging. However, in the case of a FFP, the volume which has to be stabilized is quite small and it can well be isolated from environmental changes. Most importantly, the measurement of the peak position of the etalon itself can be seen as a high-precision temperature measurement (see next section), which allows for a much better control compared to an electric temperature sensor.

For an air-gap etalon with plane parallel mirrors and a zero dur spacer (such as the HARPS or the CARMENES etalons), the CTE is $\alpha = 0.02 \times 10^{-6}\,\text{K}^{-1}$. This translates to a necessary temperature stability of some mK for sub m/s precision, which is a much less rigorous requirement compared to that for a solid etalon. Here, the most dominant error source is the illumination dependency of the peak positions (Schäfer and Reiners, 2012).
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

3.3 Laser Locked Etalon Concept

The impressive results of Wildi et al. (2010, 2011, 2012) demonstrated the feasibility of using FPEs for calibration of astronomical spectrographs. However, as noted before, the long term stability of these devices remains problematic, since the authors monitored only secondary parameters, such as temperature and pressure, to ensure the transmission spectrum being stable. It is known that illumination effects can easily change the line positions on the order of several meters per second (Schäfer and Reiners, 2012). Additionally, the spacer material can age over time (as for example zero duration does), and so can coatings. These effects cannot be tracked by a passive stabilisation of a FPE and therefore it is still necessary to recalibrate the etalon spectrum against a known absolute spectrum (e.g. ThAr) on a regular timebase for stable long-term usage.

To overcome the intrinsic limitations of a passive stabilization, we developed a more direct approach in order to provide a long-term stable etalon spectrum. The basic idea is to directly measure parts of the etalon spectrum with very high precision in real time and to compare them to a known absolute wavelength reference. This can be accomplished by using a tunable laser that simultaneously scans both the etalon and a reference. As an absolute reference, we use the $D_2$ hyperfine transition of rubidium (Rb) at 780 nm. By continuously comparing the position of one single etalon peak to the Rb spectrum, the offset between the peaks can be determined with very high precision. This information can then be used to lock the etalon to any frequency offset to the Rb transitions. Ideally, the offset would be around zero to avoid problems arising from non-linearities of the scan. The control mechanism of the etalon peak position depends on the actual etalon design as described in the previous section. Even in the absence of an active control parameter for the etalon, this new concept is useful. It allows for a precise determination of the offset of the etalon during an astronomical observation, which makes recalibration against an emission lamp redundant. A schematic and an actual photo of the experimental setup are presented in Fig. 3.13.

Compared to other locking techniques that are typically used in laser physics, such as the lock-in technique (Wallard, 1972) or the Pound-Drever-Hall method (Drever et al., 1983), the concept presented here requires less sophisticated hardware and is less complex. It has been optimised with the aim of serving the needs of RV observations, while keeping the setup as simple as possible, thus providing a device that is well suited for an observatory environment. More advanced locking setups
allow for even higher stabilities of $< 10^{-12}$. They do however not counterbalance the additional costs, the more complex setup, and the laborious long-term operation as in any case such precision levels can not be reached in astronomical observations.
3.4 Experiments

3.4.1 Laser

The narrow-bandwidth tunable light source of the experiment is an external cavity diode laser (ECDL) that we built at the Landessternwarte Heidelberg (Fig. 3.14). The design is similar to that of Ricci et al. (1995). The ECDL consists of a laser diode (QDLaser QLF073D) located behind a collimating lens. An optical grating (Thorlabs G13-18V) mounted in Littrow configuration forms the external cavity, which provides a frequency dependent feedback to the laser diode. By slightly tilting the grating, the amplified wavelength of the laser can be tuned. Fine adjustment by a piezo as well as a coarser manual tilting using a fine-pitched screw allow to tune the laser frequency to the desired value around the rubidium transitions. The lasing wavelength is not only determined by the grating position, but also by the temperature of the laser head assembly and the current the laser diode is operated with. In fact, an ECDL does not only exhibit an external cavity, but the end faces of the diode itself form another one. Therefore, the total gain at a certain wavelength corresponds to the product of the gain profiles of all cavities. Temperature variations lead to a physical change of cavity length, while the current changes the optical length by varying the refractive index inside the laser diode. Figure 3.15 shows the lasing wavelength at a fixed grating position as function of the laser current at three different temperatures. An ECDL allows for single-frequency operation, which
Figure 3.15 – Wavelength versus laser current at different temperatures (T) of the laser assembly. The plateaus arise from different lasing modes of the ECDL.

means that the laser operates on a single resonator mode providing a very narrow linewidth. However, ECDLs can support different modes and the relative intensities among them strongly depend on the feedback, temperature, and current. When one mode dominates, the laser is said to be in single-mode operation. Each step seen in Fig. 3.15 is caused by a change of the dominant mode either of the external cavity (small steps) or of the diode cavity (larger steps). An ECDL can achieve very narrow linewidths. The attainable values depend on the stability of the lasing wavelengths over short timescales, that is, on the duration over which the linewidth is measured. Noise in the driving current leads to high frequency jitter that effectively broadens the line. The current source of the laser was built at TU Darmstadt (Führer and Walther, 2012) and provides extremely low noise, which allows for a linewidth below 10kHz. The linewidth of the laser that is used here has not been measured, but it remained unresolved with a high-finesse scanning FPE, indicating that the width is below the 2.5 MHz FWHM linewidth of the FPE. The mode-hop free (MHF) tuning range of the laser was measured to be \( \sim 3 \) GHz, which is typical for laser diodes without anti-reflection coatings like the QLF073D used. The MHF range can be increased by simultaneously modifying the laser current while sweeping the optical
grating (Dutta et al., 2012). By implementing this so called feed-forward method, we could increase the MHF scan range to about 15 GHz. This range is sufficiently large to scan all components of the Rb D₂ line that span over $\sim 8$ GHz. When locking the etalon, the scan range can be reduced such that one peak of the etalon is fully covered. Yet, it is very useful to have a large tuning range. As mentioned before, it is in principle possible to lock the etalon to an arbitrary offset to the rubidium lines within the scanning range. This allows to dither the etalon in known wavelength steps through its FSR, which can be used to derive a very detailed map of the spectrograph’s detector. Wilken et al. (2010) reported discontinuities in the pixel structure of the CCD of HARPS that arise from the manufacturing process of the detector. When comparing a LFC to a ThAr spectrum, the authors found discontinuities in the pixel structure of up to 12 nm, which in their case corresponds to 60 m/s. While such large scale structures can already be seen in any etalon spectrum due to the large number of lines, dithering will in principle also allow to determine a precise sub-pixel structure of the detector.

### 3.4.2 Rubidium Doppler Free Spectroscopy

The absolute reference marks used to measure the exact etalon peak position are the hyperfine transitions of rubidium. Under normal conditions, absorption lines of an atomic or molecular gas are broadened due to the velocity distribution of the particles. Thermal Doppler broadening causes an increase of the FWHM of

$$\delta \nu = \frac{v_0}{c} \sqrt{\frac{8kT \ln 2}{m}}.$$  \hspace{1cm} (3.34)

The atomic mass of rubidium is $m = 84.911\,794$ u for $^{85}$Rb and $m = 86.909\,187$ u for $^{87}$Rb. At room temperatures, the broadening amounts to $\sim 514$ MHz and $\sim 508$ MHz respectively, which is almost a factor of 100 larger compared to the natural linewidth of 6 MHz for the Rb D₂ line (Steck, 2010). A common method in laser physics to remove the Doppler broadening from the signal is the use of saturated absorption spectroscopy. Hereby, two laser beams derived from the same laser overlap in a counter propagating way inside the rubidium gas. One beam, called the pump beam, is propagating in $-z$ direction, while the other one, called the probe beam, is propagating in $+z$ direction. The pump beam has a significantly larger intensity than the probe beam. Although both beams have the same frequency as they originate from the same laser, they address different ensembles of atoms: An atom
3.4 Experiments

With velocity \( v_z \) in \(+z\) direction will see the pump beam frequency being shifted to \( \nu'_{\text{pump}} = \nu_L(1 + v_z/c) \) due to the Doppler shift, while it will see the probe beam at frequency \( \nu'_{\text{probe}} = \nu_L(1 - v_z/c) \). In general, when \( \nu_L \neq \nu_0 \), the pump and the probe beam are resonant for atoms with different velocities \( \nu_{\text{pump}} = c(\nu_0/\nu_L) - 1 \) and \( \nu_{\text{probe}} = c(1 - \nu_0/\nu_L) \), respectively. If \( \nu_L = \nu_0 \), both beams interact with the ensemble of atoms with zero velocity in \( z \)-direction. As the pump beam has a stronger intensity, a large portion of the atoms is excited, thereby reducing the absorption of the probe beam near the atomic resonance. Within the Doppler-broadened, Gaussian-shaped absorption line small dips (often called Lamb-dips) appear when the transition frequency of the atoms in the vapour is probed by the laser. Figure 3.16 shows the full rubidium spectrum of the Rb D\(_2\) line measured with the setup. In total, one can see four nearly Gaussian-shaped absorption features, which are all showing some finer substructure. The natural abundance of rubidium is 72.2\% for \( ^{85}\text{Rb} \) and 27.8\% for \( ^{87}\text{Rb} \). Thus, a common gas cell always shows absorption for both isotopes. Figure 3.17 is a zoom-in of the \( ^{87}\text{Rb} \) F2 transition, whose lines will later be used as an absolute reference and whose frequencies are listed in Table 3.1.

<table>
<thead>
<tr>
<th>Transition</th>
<th>Frequency ( \nu_0 = 384.2291414844837 \text{ THz} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2F1</td>
<td>( \nu_0 - 229.8518 \text{ MHz} )</td>
</tr>
<tr>
<td>F2F2</td>
<td>( \nu_0 - 72.9112 \text{ MHz} )</td>
</tr>
<tr>
<td>F2F3</td>
<td>( \nu_0 + 193.7407 \text{ MHz} )</td>
</tr>
</tbody>
</table>

**Table 3.1** — \(^{87}\text{Rb}\) F2 transition frequencies for the three different hyperfine transitions. Data taken from Steck (2010).

Although there are only three hyperfine levels, in total six peaks appear in the spectrum. Three of them are so called crossover peaks. Their frequency lies exactly in between two transition frequencies \( \nu_1 \) and \( \nu_2 \). They occur because moving atoms see the counter-propagating pump and probe beam being resonant with the two separate transitions due to the Doppler shift. The depopulation of the common ground state by the pump beam causes the probe beam to see a lower absorption at \( (\nu_1 + \nu_2)/2 \). It is common in Rb saturated absorption spectroscopy experiments that the cross-over peaks are more pronounced than the actual transitions, which can also be seen in the measurements in Fig. 3.17.
Figure 3.16 – Measured saturated absorption spectrum of the Rb D\textsubscript{2} line. All four Gaussian-shaped absorption features of the two Rb isotopes can clearly be seen. Each of them shows finer substructure due to the underlying hyperfine structure.

Figure 3.17 – Zoom-in to the \textsuperscript{87}Rb F\textsubscript{2} absorption feature. The hyperfine lines and the crossover peaks (see main text) are used as the absolute reference, the etalon position is referred to. The absolute frequencies of these lines are known to very high precision and allow for a long-term calibration of the etalon lines. The pump beam only measures the Doppler-broadened absorption profile while the pump beam reveals the underlying hyperfine structure.
3.4 Experiments

3.4.3 Etalons

Several FFPs were built in-house by gluing fibres into 2.5 mm diameter zirconium ferrules and polishing them down from 10.5 mm to a length of ~6 mm. We used SMFs as well as GIFs, both with a pure fused silica core. In order to reduce stress on the fibre as far as possible, the adhesive applied was EPOTek 301-2, which is a glue with low shrinkage. As noted in Section 3.2.2, birefringence still occurred, which might also be a consequence of internal stress or asymmetries within the fibre. A thin layer of silver, overcoated with a SiO2 protection layer, was applied by e-beam deposition at Yale University.

Fig. 3.18 shows the measured transmission function for a single peak of a 6 mm long SM-FFP. The FSR of a 6 mm long fused silica (n ≈ 1.45) etalon is $\nu_{\text{FSR}} \approx 17.2$ GHz according to Eq. (3.11). The FWHM measured was 258 MHz at 780 nm, which therefore gives a finesse of

$$F_{780\text{nm}}^{\text{FFP}} = \frac{\text{FSR}_{\nu}}{\text{FWHM}_{\nu}} = 66.8.$$  \hspace{1cm} (3.35)
From the simulations in Section 3.2.3, it follows that the layer thickness of silver was about 50 nm, which is consistent with what was expected from the manufacturing process. For sufficiently sharp peaks, \( \sin\left(\frac{\delta - \delta_0}{2}\right) \approx \frac{\delta - \delta_0}{2} \) holds. This allows to approximate the transmission function of the etalon, see Eq. (3.7), by a Lorentzian.

In Fig. 3.18, the residuals from a Lorentzian fit show a low frequency sinusoidal signal that was suppressed to the 1% level by electrical shielding of the photo diode (PD) recording the etalon signal. Care had also to be taken when feeding the FFP with an optical fibre, because the fibre itself can act as an etalon. Figure 3.19 shows the transmission function when using a standard fibre with about 3 m length and flat end faces. The back reflection on the fibre entrance of about 4% forms a low finesse etalon, which modulates the overall transmission function. By using an angled physical contact (APC) fibre, the back reflection was reduced and the additional etalon effect was suppressed to the level of electronic noise.

In addition to the FFP, we also used a self-built 150 mm long CFP (FSR = 500 MHz), mainly to monitor the laser status. The design is similar to the cavity
Figure 3.20 – **Left:** The fibre Fabry-Pérot (FFP) inside the copper block, whose temperature is controlled by a cartridge heater and monitored by a thermistor. **Right:** the 150mm confocal Fabry-Pérot (CFP) made with a spacer of fused silica and brass in a thermally compensated manner.

presented in Barry (2013). Unwanted mode hops or multi-mode behaviour of the laser can easily be seen in the regular transmission spectrum of the CFP because of the large number of narrow peaks. Both the CFP as well as the FFP and its copper housing are shown in Fig. 3.20.

### 3.4.4 Data Acquisition and Software

For data acquisition, we used a National Instruments DAQ USB-4431, allowing for simultaneous sampling of four input channels and one output channel. The control software was entirely written in Python, using the Qt framework for the GUI. Via software, a smoothed sawtooth ramp is generated and applied as an output signal at the DAQ device. It is amplified by a factor of $\sim 4$ in the feed forward before being sent to the piezo. By applying the voltage ramp to the piezo, the laser frequency is swept as described in Section 3.4.1. The form of the signal (smoothed sawtooth) was chosen in order to prevent sharp voltage peaks that could damage the piezo.

While continuously sweeping the laser, the rubidium pump and probe beam as well as the signals from both etalons are acquired. The software removes the Doppler background of the Rb spectrum by subtracting the pump signal from the probe signal. By adjusting the polariser in front of the PD that records the pump beam,
the pump signal is attenuated to the same level as the probe signal. The five largest of the six Rb peaks (three transition peaks and three crossover peaks) are used to calibrate the ramp voltage scale in terms of frequency (Fig. 3.21).

To simultaneously determine the position of the etalon peak, a Lorentzian is fitted to its transmission spectrum. Numerically, the fit to a Lorentzian was found to be much faster and more robust compared to an Airy function, while still delivering a comparable fitting precision.

Fig. 3.23 shows a screen-shot of the control software and Fig. 3.24 illustrates the procedure in a flow chart. The PDs were usually sampled with frequencies of 40 kHz to 96 kHz and 4000 sampling points per ramp. The linearly increasing part of the ramp has a size of 3500 sampling points, the steeper declining part is neglected for further data processing. Depending on the sample frequency, the update rate of a whole spectrum lies between 10 Hz to 24 Hz. As fitting of the etalon signal and peak determination of the rubidium signal is required to be in real-time, the software signal processing needs sufficiently high performance. However, as consecutive spectra do not differ by much, using the previous fit results as a starting point for the current spectrum helps to keep the computing time to a minimum.

Figure 3.22 shows typical raw data of the four simultaneously recorded channels. The difference in the peak position of the etalon and the average of the highest five rubidium peaks is used as an error signal for a software proportional-integral-derivative (PID) control, whose parameters have to be optimised depending on the etalon design.
3.4 Experiments

Figure 3.22 – A typical raw spectrum recorded within one laser sweep (~88 ms). Signals from both etalons (grey and red lines, right scale) and the rubidium (blue and green lines, left scale) are retrieved simultaneously, thus allowing for a precise determination of the current etalon position with respect to the rubidium reference.

The output of the PID control was sent to a second National Instruments DAQ card (myDAQ), as the NI USB-4431 had only one output channel that was already used for the laser ramp. For the piezo driven CFP, the output of the myDAQ was attached directly to the piezo, while for the cartridge heater, the myDAQ could not provide sufficient power. Therefore, the signal was sent to the modulation input of a laser diode driver (Analog Technologies ATLS100mA103-D), which was used as a stable, low noise current amplifier to operate the heater.
Figure 3.23 – Screenshot of the control software of the etalon lock. The top panel shows the etalon transmission signal as well as the rubidium spectrum in real time. The lower panel shows the deviation between the etalon and the rubidium peak of the last 500 measurements.
3.4 Experiments

Figure 3.24 – Flow chart of the control software of the etalon lock.
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

3.4.5 Locking Stability

The locking performance was tested both with the piezo driven CFP and with the FFP. The former is not suitable for an astronomical spectrograph, as its lines are unresolved by any typical échelle spectrograph. However, it could be demonstrated that the locking technique works with various types of etalons, independently of its tuning method.

Confocal Fabry-Pérot Etalon

The CFP has an almost instantaneous response function due to the fast piezo reaction. This eases the determination of good PID parameters. Figure 3.25 shows the statistics of the best 2.5h long dataset of a running lock for the CFP, which ran with 10 Hz update rate. For a single measurement, the standard deviation of the error signal was $\sim 0.21$ MHz or $\sim 16$ cm/s. The relative frequency of the errors matches very well a Gaussian shape (see Fig. 3.25b). This is confirmed by an Anderson-Darling test for normally distributed errors. The test statistic had a value of 0.753, which implies that the null hypothesis of normally distributed errors could not be rejected at a 5% significance level. The behaviour of the error within a certain integration time is of particular interest, as the typical exposure times of an astronomical observation are in the range of a few seconds to half an hour. Figure 3.25c shows how the error evolves with increasing binning time. After $\sim 2.5$ s, the standard deviation of the binned error drops below 3 cm/s, while after 30 s the observed maximum error also drops below 3 cm/s. It can also be seen that the standard deviation of the CFP errors nicely follows $\sigma/\sqrt{N}$, where $N$ is the number of measured spectra. The optical table on which the setup was located had no vibrational damping and both laser and etalon are very sensitive to mechanical vibrations. For this reason, the air conditioning of the laboratory had to be switched off during the measurements. The setup was neither thermally nor pressure stabilized. As mentioned before, the piezo was driven directly by the NI myDAQ card, which had an output voltage range of only $\pm 10$ V. Over a few hours, air pressure variations required a correction of the piezo of more than 10 V, which is why the longest locking period was only about 8 h. However, there is no reason why the locking should be limited to short time scales. Putting the etalon inside a pressure stabilized or even evacuated chamber will greatly enhance the overall stability. Alternatively, using a piezo amplifier will increase the tuning range of the etalon to compensate for larger pressure variations.
3.4 Experiments

(a) Errors while lock is running

(b) Distribution of the errors

(c) Locking performance as function of binning time

Figure 3.25 – Measured performance of the locking stability of the piezo driven confocal Fabry-Pérot. During the 2.5h time span, the deviations of the etalon position with respect to the Rb were normally distributed with a standard deviation of 0.2 MHz. For binning times in the regime of seconds, the etalon position is stable better than 3 cm/s. For binning times of several minutes, the deviations drop even to the order of mm/s.

Fibre Fabry-Pérot Etalon

Figure 3.26 shows the performance of the locked FFP for the best 2.5h interval (15 Hz update rate). The standard deviation for a single measurement was 0.18 MHz and the distribution of the errors also looks Gaussian like, similar to the CFP. However, an Anderson-Darling test implied that the errors are not normally distributed, i.e. the null hypothesis of normally distributed errors was rejected at the 1% significance level. It takes about 150s of integration until the standard deviation of the error drops below 3 cm/s. This results from the fact that the FFP has a significantly longer response time to a change of the PID output than the CFP. While the CFP reacts almost instantaneously, the thermal mass of the copper block of the FFP and the bad thermal conductivity of the fibre ferrule cause a time lag between the change of the control output and a measurable peak shift. It should also be noted that Fig. 3.26 only shows the locking precision, while the position of the etalon peak for the corresponding binning time is known to a much higher precision.
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

Figure 3.26 – Measured performance of the locking stability of the thermally controlled fibre Fabry-Pérot. The relatively long response time of the FFP hinders a fast controlling of the etalon position. Compared to the CFP, the FFP shows a larger locking error for the same binning times. Still, for binning times above $\sim 150$ s the deviations drop below $3 \text{ cm/s}$.

ECDLs are extremely sensitive to vibrations. This is illustrated in Fig. 3.27, which shows the locking performance of a measurement where the air conditioning as well as a small vacuum pump, which was located next door ($\sim 7 \text{ m off}$), were running in the background. These vibrations caused a degradation in locking performance and they also manifested themselves in pronounced peaks in the periodogram of the errors.
3.4 Experiments

Figure 3.27 – Locking performance for the fibre Fabry-Pérot in case mechanical vibrations hindered an optimal performance (top left). The corresponding histogram of the errors (top right). A periodogram of the errors reveals several peaks that indicate a time correlation of the errors (bottom left). These vanished when switching off the air conditioning and the vacuum pumps next to the laboratory (bottom right).
Allan deviation

A good way to measure the residual frequency fluctuations of a locked etalon is the Allan deviation (ADEV) of the errors, which is often used to characterize frequency stability of clocks, oscillators or locked lasers, see Allan (1987) and Barnes et al. (1971). The Allan deviation (ADEV) is defined as

\[ \sigma_y(\tau) = \sqrt{\frac{1}{2} \left\langle (\bar{y}_{n+1} - \bar{y}_n) \right\rangle}, \quad (3.36) \]

where \( \bar{y}_n \) is the fractional frequency deviation, averaged over the period length \( \tau \), with \( \tau = n\tau_0 \), where \( \tau_0 \) is the sampling interval. \( \bar{y}_n \) is defined as

\[ \bar{y}_n = \left\langle \frac{\delta \nu}{\nu} \right\rangle_n, \quad (3.37) \]

where \( \delta \nu \) is the measured frequency deviation and \( \nu \) the reference frequency. The ADEV is a function of the time period used between samples. The interpretation of the ADEV is the following: the frequency measurements are averaged over \( \tau \) and the differences of successive intervals are used to calculate the ADEV. The value quantifies the instability between two observations that are \( \tau \) apart and averaged over \( \tau \), i.e. their relative root mean square is then \( \sigma_y(\tau) \). Figure 3.28 illustrates the ADEV for the locked CFP and FFP. The red dashed lines show how the curve would look like for normally distributed data with a standard deviation corresponding to the one measured. For the CFP, the measured data follows roughly the slope \( (1/\sqrt{\tau}) \) of a normally distributed dataset. Deviations from the line indicate other sources of noise rather than white noise from the measurement process. Deviations on the short time-scales for the CFP most likely arise from the sensitivity to mechanical vibrations. These effects are absent in the case of the FFP, where the ADEV follows that of white noise for the first \( \sim 3 \) s. After that, the ADEV is rising, which is most probably due to the slow response time of the FFP. It could also be the case that the PID parameter are not chosen optimally, causing a residual oscillation around the set point. In any case, the rising ADEV at \( \sim 3 \) s is the reason why in Fig. 3.26c, the slope of the standard deviation becomes rather flat after this time-scale.
3.4 Experiments

Figure 3.28 – Allan deviation (ADEV) of the locked CFP (top) and the FFP (bottom) shown as fractional frequency (left axis) and corresponding RV (right axis). The ADEV reveals on what timescales deviations of the etalon position with respect to the rubidium occur. For a white noise process, the ADEV would follow the red dashed lines. In case of the FFP, they clearly deviate after $\sim 10\text{s}$, indicating non-optimal locking parameters or inherent limits of the lock, e.g. due to the slow response of the FFP.
3 Stabilized Fabry-Pérot Etalon for High-Resolution Spectrographs

3.5 Reliability and Systematics

For a long term RV survey, the reliability and any potential systematic shifts of the calibration source are of great importance. In case of an actively locked etalon, the calibration spectrum, i.e. the etalon spectrum, is always compared to an absolute frequency reference. Therefore, all systematic shifts of the locked etalon result from systematics of the reference itself. In the case of rubidium, Affolderbach and Miletì (2005) investigated the sensitivity of the line positions to external parameters. They found a dependence on the absolute line position on the used pump laser power of

\[ \frac{0.2 \text{kHz}}{\mu\text{W cm}^{-2}} \]

The beam diameter of our laser was \( \sim 1.4 \text{mm} \), which gives a power dependent shift of \( \sim 13 \text{kHz} \mu\text{W}^{-1} \). For typical laser powers used, the absolute offset is a few hundred kHz or several tens of cm/s. Fluctuations of the laser power play the decisive role rather than the absolute laser power. We measured a relative power fluctuation of 0.175% in our setup over 9.6h, which translates into an uncertainty of only 0.59kHz over this timescale.

Temperature variations of the gas cell and external magnetic fields can also shift the rubidium lines and contribute to a systematic shift of lines. According to Affolderbach and Miletì (2005), the peak shift due to a magnetic field is \( 1.2 \text{kHz} \mu\text{T}^{-1} \) and the temperature dependence is about \( 4 \text{kHz K}^{-1} \). The authors also estimate that an angular misalignment of the pump and the probe beam causes a shift of \( 80 \text{kHz mrad}^{-1} \). By using a small aperture for beam alignment, the angle should be determined in the setup with an accuracy of more than 1 mrad, which is equivalent to 6 cm/s. In Fig. 3.16, it can be seen that the doppler broadened wing of the \(^{85}\text{Rb}\) isotope adds a slope to the used \(^{87}\text{Rb}\) lines. However, the contamination is very small. We fitted a Gaussian to the \(^{85}\text{Rb}\) spectrum and subtracted it from the \(^{87}\text{Rb}\). The resulting shift of the hyperfine lines was below 1 kHz.

The laser power is a critical parameter as it does not only shift the lines, but also influences the width of the lines and the achievable SNR of the spectrum. The increase of the linewidth with laser power is called power broadening, which can be seen in Fig. 3.29 and Fig. 3.30. Steck (2013) reported a saturation intensity for a circularly polarised laser of 1.67 mW cm\(^{-2}\), which corresponds to 26 \( \mu\text{W} \) for the 1.4 mm beam size used in our experiments. Beyond this threshold, power broadening becomes an important factor.
Figure 3.29 – Rubidium spectrum obtained with different laser power ranging from 10µW to 126µW. Increasing pump power leads to a larger signal, but also broadens the lines. In general, there is a trade off between signal strength and line width.
Figure 3.30 – Peak amplitude and FWHM as function of laser power (top left). Measurement of RMS noise versus laser power (points) and a fit to $\sqrt{\text{probe power} + \text{offset}}$ (top right). The dashed line at 42.82 $\mu$W is the fitted offset. SNR, calculated by the fitted peak amplitude divided by the RMS noise (bottom left). For determination of the line position, the line width has also to be taken into account. Plotting FWHM/SNR against laser power (bottom right) flattens for larger pump powers, which means that a further increase in power does not allow for a better position determination.
3.6 Échelle Measurements

I also investigated the FFP using a laboratory échelle spectrograph (R ∼ 80 000) that is currently being built at the Landessternwarte Heidelberg for exoplanetary research. Up to now, there is no data reduction pipeline available for this spectrograph. Therefore, I extracted the spectra rather naively, by summing up several pixels in cross dispersion direction along each order. The data was only used to check that the FFP is working as expected over a wider bandwidth. Figure 3.31 shows part of a recorded spectrum of a 3mm silver-coated FFP. It shows the behaviour one would expect over a large bandwidth. It visualizes the desirable high line density over a large bandwidth. Figure 3.32 illustrates an extracted etalon spectrum of the 6mm FFP that was actually used for the locking experiments described in Section 3.4. The etalon lines are slightly undersampled, however, particularities of an échelle spectrum like the anamorphic magnification, i.e. an optical distortion effect that leads to varying magnification along each order, can still be seen. A true colour image of a FFP transmission spectrum is shown in Fig. 3.33.

![Image of échelle spectrum](image)

**Figure 3.31** – Part of an échelle spectrum of a 3mm FFP illuminated with white light. The Wavelength increases from left to right. This can be seen by the larger separation of the orders on the left side due to the higher dispersion of the prism at the blue end of the spectrum, and also by the smaller separation of the etalon peaks, which scales with $\lambda^2$. The low brightness of the peaks in the range of column 600 – 1400 are due to the low efficiency of the LED that was used to illuminate the FFP in this wavelength regime.
Figure 3.32 – **Left:** Part of one order of an extracted échelle spectrum of a 6mm FFP illuminated with white light (left scale). For reference, a ThAr spectrum was also recorded and extracted (right scale). **Right:** A Gaussian is fitted to each etalon peak and the spacing of two consecutive peaks is plotted versus position on the CCD. The distance increases linearly in first order, which is a consequence of the anamorphic magnification of the échelle.

Figure 3.33 – A true colour image of an échelle spectrum recorded with a commercial DSLR camera with the laboratory échelle spectrograph at the Landessternwarte Heidelberg. It shows the transmission spectrum of a 3mm fibre Fabry-Perot.
3.7 Conclusion and Future Work

The presented locking stabilities of both the confocal Fabry-Pérot and the fibre Fabry-Pérot demonstrate the potential of our new active locking scheme. For both etalons, one transmission peak could be stabilized well below 10 cm/s for any integration time that is reasonable for astronomical observations. This is an improvement compared to the common ThAr technique of about one order of magnitude and is suitable for next-generation high-resolution spectrographs. Furthermore, the technique can be adapted to various etalon designs in the VIS as well as in the NIR spectral range. The setup is quite simple and the total cost, including operational cost, is comparatively low. This makes this technique particularly interesting for smaller observatories that might not be able to afford a laser frequency comb. The monitoring capabilities of the technique can also be used as an upgrade for passively stabilized etalons, where no active tuning was planned during the design phase.

The intent of this project was to deliver a proof of principle, while more work is needed to package the experiment into an out-of-the-box working system. In the following, some future work is suggested, addressing important open questions and improvements. The most crucial open question is, whether the stability of one single etalon peak can fully be translated into the stability of the whole transmission spectrum. While for short and medium time-scales (up to a few days) this might be a fair assumption, it remains unclear for longer time frames. Dispersion effects could be time-dependent, e.g. due to ageing of materials. In the case of a FFP, another scenario is conceivable: the pressure on the FFP changes over time leading to a line shift as described in Section 3.2.2. When these shifts are corrected by a temperature tuning, the errors caused by these corrections on peaks other than the locking peak depend on the differential dispersion $\frac{\partial \lambda}{\partial \sigma} d\sigma - \frac{\partial \lambda}{\partial T} dT$. However, this effect can be measured in a very similar way as the presented scheme. By using other gas cells (e.g. iodine, HF or caesium) and appropriate tunable lasers, the locking or monitoring of the lines with respect to a (different) absolute reference can be extended to other wavelength regimes. While locking at one transition, the second measurement would reveal dispersion related effects.

Another important question is, which etalon design is the most suited one for a long-term stable RV instrument. The design parameters include the type of the etalon (plane-parallel, confocal, solid, etc.), the coating (metallic, dielectric) as well as the tuning method (temperature, air pressure, mechanical). As an example, both the HARPS and the CARMENES etalon have a low finesse value around 10, which
is roughly constant with wavelength. A detailed study that takes into account the peculiarities of an échelle such as its constant resolution and incorporates them to current extraction schemes is needed.

This goes hand in hand with the question whether an etalon that is locked to an absolute reference could also be used as an absolute wavelength meter. This should in principle be possible, as the Fabry-Pérot can once be characterized against a LFC or a ThAr spectrum. Especially in the case of metallic mirrors, when the phase shifts are a low-order function of wavelength, it should be possible to determine the wavelength of each single transmission peak with good accuracy. Using two or three saturated absorption spectroscopic measurements at different wavelengths would deliver good constraints on the slope of the dispersion function.

More concretely, a couple of improvements to the presented setup would further enhance the precision and stability: As mentioned, the ECDL is very sensitive to vibrations and fluctuations of environmental conditions. An active stabilization scheme based on polarisation spectroscopy would reduce the sensitivity of the ECDL at the cost of adding only slightly more hardware complexity (Führer et al., 2011). One major contribution to the noise of a single measurement is the determination of the rubidium peak position, which is currently only implemented as a simple peak detection of the five largest peak in the rubidium spectrum. Himsworth and Freegarde (2010) compared measured Rb spectra to theoretical predictions and showed that their model based fit, not using any free parameters, has a residual difference of less than 1% over a large range of pump powers. Fitting an analytic function to the rubidium, would decrease the uncertainty of the peak position and would also allow for an easy monitoring of the status of the rubidium setup. For example, if spectral lines get dimmer or broader over time, it could indicate a gradual misalignment of the probe and pump beam or ageing of the cell. An analytic fitting would require more computational power, as it has to be done in real time. With a fully multi-threaded approach, such requirements should easily be fulfilled for a standard computer. So far, no optimisation of the sampling rate, the sampling size and the locking rate was performed. Investigating and optimising these parameters will most likely further improve the performance of the instrument.

Even in its current state, the laser locked etalon has significant benefits over other calibrators and could become an important asset for exoplanet spectrographs in the near future.
Bibliography


BIBLIOGRAPHY


BIBLIOGRAPHY


Doppler wavelength reference: a fiber Fabry-Perot interferometer. *Ground-based and Airborne Instrumentation for Astronomy IV.*


BIBLIOGRAPHY


BIBLIOGRAPHY


Acknowledgement

I would like to take the opportunity to thank all the people, who supported me during my PhD studies.

First, I would like to thank my supervisor Prof. Andreas Quirrenbach for giving me the opportunity to work on these interesting topics. I appreciated his helpful comments, his guidance, but also the freedom he granted me to focus on my own research interests. Thanks go to my second supervisor Otmar Stahl, who was always willing to listen to my problems and helping me. I also thank Prof. Reinhard Mundt for being the second referee for this thesis.

I acknowledge the support of the Studienstiftung des Deutsc hen Volkes, both financial and otherwise.

This work would not have been possible without the support of Christian Schwab. For the many fruitful (and sometimes exhausting) discussions, your great scientific guidance and your willingness to listen to my problems (and for keeping me busy in the rare moments I would have had some free time) around the clock from around the world: Thank you!

A big thank you goes to my co-authors, especially Yuliia Gurevich and Thorsten Führer, for sharing their knowledge with me.

I would also like to thank Christoph Koke, who patiently helped me whenever I ran into programming issues. He, Stephan Meister and Jens-Malte Gottfried, were also great company during 'the good old days' in Heidelberg. A special thank you goes to Tim-Oliver Müller. Not only for your useful comments on the manuscript, but for all your support in the last years. It is good to know to have such friends.

I want to thank my colleagues at LSW, Lutz Geuer and Ludwig Schäffner for their competent help whenever I needed something 'special' from the mechanical workshop and also my office colleagues during the last years, Andre, Kirsten, David, Markus and especially Christoph for keeping up with the chaos that tend to spread from my desk.

I am grateful for Uli Grötzingers help and his patience with trying to improve my knowledge of electronics.

To the people in Potsdam, Dionne and Roger Haynes and Andreas Kelz: thank you for letting me use your fancy fibre facilities and sharing your knowledge with me. This was extraordinary helpful.

None of this would have happened without the unconditional support of my family, especially my parents. Thank you for your trust in my abilities and your advice. All I have achieved is your credit as well.

To Karin: thank you for scrutinising the manuscript, for your constant encouragement and for your understanding for the long evenings this work took up. I am so grateful that I met you.