Topics in precision astrophysical spectroscopy

A dissertation presented

by

Aakash Ravi

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

October 2019

Photo: A. R.

Dissertation Advisor:
Ronald Walsworth

Author:
Aakash Ravi

Topics in precision astrophysical spectroscopy

Abstract

Applying advances in optical frequency metrology to astronomy has enabled extremely precise spectroscopic measurements of stars. Precision spectroscopy of stars can inform us not only about stellar physical processes but also about the existence of extrasolar planets. An indispensable tool for carrying out these measurements is the optical frequency comb. Using a turn-key Ti:sapphire optical frequency comb, we report the generation of a 16-GHz visible frequency comb suited to astrophysical spectroscopy (i.e., an "astro-comb"). The light from this source is used to calibrate the HARPS-N astrophysical spectrograph for precision radial velocity measurements. The comb-calibrated spectrograph achieves a stability of ~ 20 kHz within half an hour of averaging time. We also use the astro-comb as a reference for measurements of solar spectra obtained with a compact telescope and as a tool to study intrapixel sensitivity variations on the spectrograph detector. After the initial tests of the astro-comb, we describe the design and testing of a dispersion-engineered tapered photonic crystal fiber which is capable of producing a broad, flat, visible supercontinuum spectrum. Using 30 fs, 100 pJ pulses from our Ti:sapphire laser, we obtain an output spectrum that is flat to within 3 dB over the range 490-690 nm with a blue tail extending below 450 nm. This greatly extends the range of applicability of the astro-comb. Finally, we present a remote-control interface for the astro-comb and outline future automation prospects.

Beyond the detection of extrasolar planets, we show that the precision radial velocity technique is sensitive enough to observe the acceleration of stars in the galactic potential. Stellar acceleration

measurements would inform us about the dark matter density distribution in the Milky Way. This knowledge is crucial to both our understanding of the standard cosmological model and for grounding direct and indirect searches for the particles comprising dark matter. Current measurements of galactic dark matter content rely on model assumptions to infer the forces acting upon stars from the distribution of observed velocities. Using the precision radial velocity method instead, we can measure the change in the velocity of an ensemble of stars over time, thereby providing a direct probe of the local gravitational potential in the galaxy. Using numerical simulations, we develop a realistic strategy to observe the differential accelerations of stars in our galactic neighbourhood with next-generation telescopes, at the level of 10^{-8} cm/s^2. Our simulations show that detecting accelerations at this level with an ensemble of 10^3 stars requires the effect of stellar noise on radial velocity measurements to be reduced to < 10 cm/s. The measured stellar accelerations may then be used to extract the local dark matter density and morphological parameters of the density profile.

Many extrasolar planets have already been discovered, so it is natural to ask whether if any of them are habitable. In an effort to answer this question, we can search for the absorption signatures of oxygen, which is crucial to life on Earth, in the atmospheric spectra of exoplanets. This is an extremely challenging measurement to do with ground-based surveys, and one obvious systematic that threatens the quality of the data is the variability of the inevitable spectral background resulting from the Earth's oxygen absorption features. Using four hours of high-resolution atmospheric spectra obtained with a Fourier transform spectrometer, we examine correlations of line parameters with environmental parameters to try to assess whether this contribution may be modelled out or if in-situ monitoring of the atmosphere is required.

Contents

List of Figures

List of Tables

Citations to Previously Published Work

Parts of this dissertation cover results reported in the following articles:

1. A. Ravi, D. F. Phillips, M. Beck, L. L. Martin, M. Cecconi, A. Ghedina, E. Molinari, A. Bartels, D. Sasselov, A. Szentgyorgyi, and R. L. Walsworth, "Astro-comb calibrator and spectrograph characterization using a turn-key laser frequency comb," *Journal of Astronomical Telescopes, Instruments, and Systems* **3**, 045003 (2017).

2. A. Ravi, M. Beck, D. F. Phillips, A. Bartels, D. Sasselov, A. Szentgyorgyi, and R. L. Walsworth, "Visible-Spanning Flat Supercontinuum for Astronomical Applications," *Journal of Lightwave Technology* **36**, 5309 (2018).

3. A. Ravi, N. Langellier, D. F. Phillips, M. Buschmann, B. R. Safdi, and R. L. Walsworth "Probing Dark Matter Using Precision Measurements of Stellar Accelerations," *Physical Review Letters* **123**, 091101 (2019).

Acknowledgements

Graduate school has been quite the roller coaster. In June 2016, I was in dire straits, having just left a cold atoms group four years into my PhD. But I was not quite ready to quit. It was around this time that I first reached out to Ron. And the rest is history...

So first and foremost, I would like to thank my advisor Ron Walsworth for giving me the opportunity to work on a variety of very interesting physics problems. Ron's creativity as a physicist has been a continual source of inspiration for me. I am especially amazed by his ability to take precision-measurement tools from AMO physics and apply them in novel ways to solve important problems in other fields (e.g., from measuring magnetic fields produced by neurons all the way to enabling discovery of Earth-like planets outside our solar system). Ron has been a very kind and encouraging advisor, allowing me the academic freedom to pursue my own ideas as well as suggesting interesting projects for me to work on based on his ideas. It has been a true pleasure to work in his research group and I am certain I will miss it dearly in the days to come.

Without a doubt, the next person I need to thank is senior scientist David Phillips. David is an exceptional mentor with tremendous amounts of technical expertise and patience. It is not an understatement to say that nearly all of the technical results in this thesis can be traced back, in one way or another, to discussions I had with David. Hence, many thanks to David for always leaving his door ajar so that I could always drop in and pepper him with silly questions! I have learned so much in the last three years and I feel very fortunate to have been able to work with him.

Next, I would like to thank my colleagues Nick Langellier and Tim Milbourne (a.k.a. the "A-team"). I could truly not ask for better office mates. Nick and Tim, being data analysis wizards and knowing far more about astronomy than I probably ever will, have helped me many, many times in my research. Together, we've shared countless laughs, consumed much fine coffee, eaten many baked goods, and played nearly every day's New York Times Spelling Bee word puzzle. Not to

mention all the time spent tinkering with CoffeeBot and the "board meetings" held at Jose's. The jovial atmosphere of the office has been an extremely positive influence on me, especially in the final stretch as I was trying to wrap up my thesis.

Professionally, I have had the great fortune to collaborate with a number of very smart people on all my projects, and I am very thankful for this. I am grateful to Chih-Hao Li, Andrew Benedick, Matthias Beck, Albrecht Bartels, Leopoldo Martin, the TNG staff, Andrew Szentgyorgi, Dimitar Sasselov, Guoquing Chang, Franz Kärtner, Pawel Latewiec, Fiorenzo Omenetto, Gábor Fűrész, Tim Hellickson, Mike Harju, Boris Desiatov, Amirhassan Shams-Ansari and Willa Dworschack for their invaluable help on the comb-related projects. Of particular note, Matthias Beck (Product Manager at Laser Quantum GmbH at the time) was instrumental in the success of the dispersion-engineered photonic crystal fiber project. We collaborated on other ideas as well, culminating in a trip I took to the company headquarters in Konstanz, Germany, a visit I remember very fondly. I also want to highlight the wonderful work of an extremely talented undergraduate student, Willa Dworschack. Her LabVIEW code, once deployed, will enable the astro-comb to be remotely-operable, a very exciting moment for HARPS-N exoplanet science!

For their theory contributions to the stellar accelerations project, I would like to thank Benjamin Safdi and Malte Buschmann. And for the final topic of my thesis (regarding searches for oxygen on exoplanets), I am greatly indebted to Mercedes López-Morales for bringing me on board to work on such an ambitious and exciting research project. She has been an extremely warm and supportive mentor to me, and I have learned much about exoplanet astronomy as a result. I would also like to acknowledge the help of Iouli Gordon and thank Geoffrey Toon, Ralf Sussmann and Markus Rettinger of the TCCON collaboration for supplying the data for the oxygen project.

Outside the immediate sphere of my research, I also have many people to thank – Cora Dworkin, Marko Lončar and Susanne Yelin for serving on my committee, my former advisor John Doyle for giving me the opportunity and the environment to learn the skills required of an AMO physicist, Börge Hemmerling, Eunmi Chae, Garrett Drayna for their mentorship, Wolfgang Ketterle for his brilliant ideas and technical advice, Jim MacArthur for his help with electronics, Stan Cotreau and

Steve Sansone for their help with machining, and Lisa Cacciabaudo and Jacob Barandes for their administrative support. I also want to thank the entire Walsworth group for all their support and the #lunch crew for many a lunchtime well spent.

Outside the lab, there were many whose company kept me sane through all the years. I especially owe a great deal of thanks to my good friend Jacob Baron, who was there for me the entire way, through good times and bad. I also want to thank Tony Zhu for accompanying me on weekend climbing trips no matter how cold it was outside and no matter how many times I would get us lost in the woods. I also want to thank members of my extended family for all their support.

Finally, I would like to thank my loving parents. I owe more to them than I can ever write here. Their unwavering support and encouragement has been the one constant in my life over the last seven years. I want to thank them for taking a very active interest in my progress, for listening, for sharing in both the joys of my successes and the hardships of failures, and for all of their sound advice through the years. I could not have done it without them.

To my parents

Preface

It has been over 200 years since absorption lines were first detected in the Sun[1]. This discovery led to tremendous insights regarding the elemental composition of the Sun and was arguably the birth of astrophysical spectroscopy.

In the intervening two centuries, tremendous advances in atomic physics and optical engineering have completely revolutionized the way we study astrophysical objects. Notably, the invention of large, high-quality reflecting telescopes, echelle gratings and optical frequency combs simultaneously enable us to see extremely faint stars as well as measure the frequency of light with extraordinary precision (e.g. resolving sub-part-per-billion Doppler shifts).

Information about various physical phenomena can be encoded in the spectrum of starlight. This work covers three of these phenomena: the reflex motion of stars as a result of the gravitational interaction with orbiting planetary bodies, the acceleration of stars in the gravitational potential of the galaxy and the passage of starlight through the atmospheres of extrasolar planets. In more detail, **Chapter 1** discusses the development of highly-robust optical frequency combs as calibration tools for exoplanet science. **Chapter 2** discusses a proposal to measure the local dark matter density by tracking the velocity of stars in our galaxy over decade timescales. And finally, **Chapter 3** discusses a proposal to detect oxygen in the atmosphere of extrasolar planets using ground-based telescopes, and tries, using high-resolution spectroscopic data, to quantify an important systematic: namely, the variability of the Earth's atmospheric spectra during such observations.

I hope that these examples provide a convincing demonstration of the power of precision astrophysical spectroscopy.

[1] W. H. Wollaston (1802), J. Fraunhofer (1814)

Co-worker contributions

Chapter 1

David Phillips (CfA), Matthias Beck (Laser Quantum) and I jointly gathered the data for the initial astro-comb tests presented in §1.5. David analyzed all the data in that section. I prepared the manuscript which summarized our results from the initial demonstration[2].

I came up with the design for the tapered photonic crystal fiber described in §1.6 and performed all the numerical simulations. The fiber was fabricated by Mike Harju (Vytran LLC) and tested by Matthias at Laser Quantum in Germany. I interpreted the results and summarized the results in a short paper[3].

Willa Dworschack (Lawrence University) and I designed and implemented the remote-control interface for the comb described in §1.7.

Together, Matthias and I brainstormed the idea for an astro-comb based on a 10 GHz source comb. I designed the silicon nitride chip in §1.8 with help from Boris Desiatov of Prof. Marko Lončar's group (Harvard SEAS). Boris fabricated the chip. Matthias and I did the tests.

Chapter 2

Ron came up with the idea for this project. The theory was developed by our collaborators Malte Buschmann and Benjamin Safdi (University of Michigan). I generated the synthetic radial velocity time series data, did the naive version of the data analysis and wrote much of the manuscript[4]. The more sophisticated analysis with Gaussian process regression was done by my colleague Nicholas

[2] *Journal of Astronomical Telescopes, Instruments, and Systems* **3**, 045003 (2017)
[3] *Journal of Lightwave Technology* **36**, 5309 (2018)
[4] *Physical Review Letters* **123**, 091101 (2019)

Langellier (Harvard). David did much of the early feasibility studies, and provided extremely valuable technical advice throughout.

Chapter 3

Mercedes Lopez-Morales (CfA) came up with the idea for this project. The data was collected by Ralf Sussmann and Markus Rettinger of the Total Carbon Column Observing Network (TCCON) collaboration. I did all the data analysis and documentation of results. Mercedes and David co-supervised the project and provided me lots of useful technical guidance and feedback.

1

Optical frequency combs for exoplanet astronomy

1.1 Radial velocity method

The existence of inhabitable planets outside our solar system has long been a tantalizing prospect. Due to the vast distances separating Earth from such extrasolar planets, it is not generally possible to directly observe them with telescopes at present. However, one observable signature is a change in the spectrum of emitted light from the host star due to gravitational forces between the planet and the star. When two massive bodies interact via the gravitational force, they both execute orbits about their centre of mass. Though the planet is typically much lighter than the star, the motion of the orbiting planet causes the star to move a small amount. This effect may be observed as periodic Doppler shifting of the starlight at a period given by the planet's orbit, as shown in Figure 1.1. In this way, we obtain evidence of the planet's existence without actually directly resolving the disk of either the star or the planet with a telescope. Since the orbital information is encoded in the spectrum of the starlight, one can use a high-resolution echelle spectrograph to observe the Doppler shifts. This method is termed the radial velocity (RV) method, and it has been extremely successful in finding exoplanets over the last few decades [1, 2]. The RV method provides information about the planet's mass, distance from the star, and orbital period. When combined with observations from

a complementary transit method [3], giving information about the size of the planet, we can make estimates of the planet's density and thereby draw some basic conclusions about its composition.

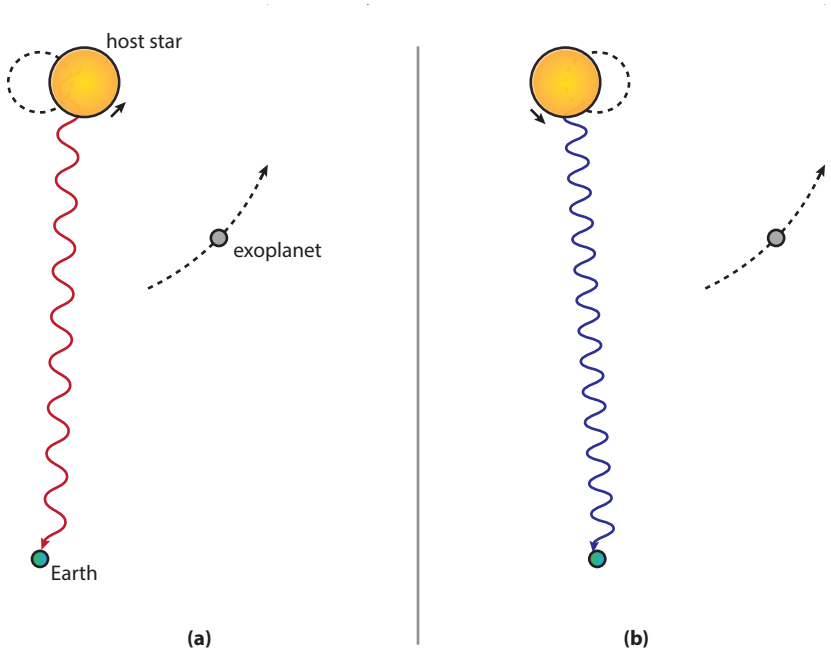

Figure 1.1: The gravitational interaction of a host star and an orbiting exoplanet causes the star to wobble. When the star is moving away from the Earth as in panel (a), the light is red-shifted. Half a cycle later, when the star is moving toward the Earth as in panel (b), the light is blue-shifted. This can be detected by collecting the starlight with a telescope and then using a dispersive system (spectrograph) to observe the spectrum.

To get an estimate of the size of Doppler shift associated with reflex motion of a Sun-like star being orbited by an Earth-like planet, we will work out the numbers for the Sun-Earth system. To start, let us make three simplifying assumptions: (1) mass of star $m_s \gg m_p$ mass of planet, (2) the orbit has no eccentricity and (3) the system is edge-on and not tilted with respect to the line of sight to Earth.[1]

Kepler's third law tells us that square of the period T_p of any planet is proportional to the cube of the semimajor axis of its orbit a_p, i.e.

$$4\pi^2 a_p^3/T_p^2 = G(m_s + m_p) \approx G m_s, \tag{1.1}$$

[1]For a much more rigorous treatment of the orbital mechanics involved in such calculations, see Ch. 2 of Ref. [3].

where G is the gravitational constant. Solving for the semimajor axis, we get

$$a_p = (G m_s T_p^2 / 4\pi^2)^{1/3}. \tag{1.2}$$

It follows from the solution of the gravitational two-body problem that the centre of mass is invariant with time and that star's reflex motion about the centre of mass is weighted by a factor of approximately m_p/m_s. Therefore, the size of star's semimajor axis is given by

$$a_s = (m_p/m_s) a_p. \tag{1.3}$$

The star executes this wobble over the same period T_p as the planet. We can then model its projection x on the semimajor axis as a sinusoidal oscillation,

$$x = a_s \sin(2\pi t/T_p). \tag{1.4}$$

The velocity along this coordinate is

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = 2\pi a_s/T_p \cos(2\pi t/T_p). \tag{1.5}$$

The RV amplitude is then the amplitude of this function, i.e.

$$\mathrm{RV} = 2\pi a_s/T_p = \left(\frac{G m_s T_p^2}{4\pi^2}\right)^{1/3} \times \frac{m_p}{m_s} \times \frac{2\pi}{T_p}, \tag{1.6}$$

putting all the formulas together. Plugging in numbers for Sun-Earth, we get 9 cm/s, and for Jupiter, we get 12.5 m/s. The Doppler shift $\delta\nu$ is then $\mathrm{RV}/c \times \nu$ ($\sim$ 200 kHz for starlight in the optical range $\nu \sim 600$ THz). Therefore, the task of finding a Earth-like exoplanet around a Sun-like stars amounts to making a sub-ppb Doppler shift measurement.[2]

[2] In reality, it is even more difficult. The magnetic activity of stars as well as instrumental systematics limit us long before we get to this Doppler shift precision [4–6].

1.2 Optical frequency combs

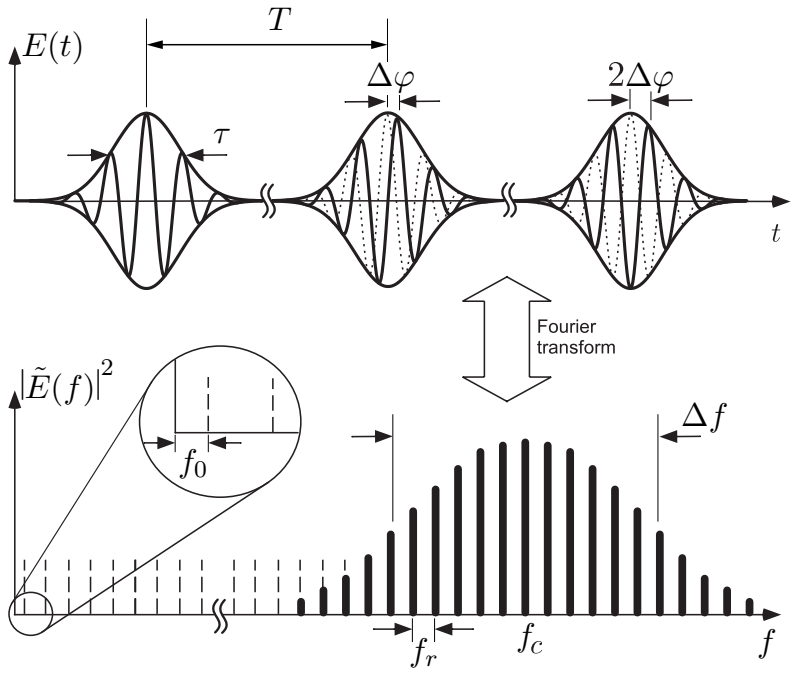

Figure 1.2: Time-frequency correspondence for a pulse train. In the time domain, the electric field is characterized by a regular pattern of pulses. The time interval between adjacent pulses is the period T (given by the cavity roundtrip time). The electric field oscillates at carrier frequency f_c and is amplitude-modulated by an envelope of characteristic width τ. As a result of cavity dispersion, the phase between the carrier and envelope slips by an amount $\Delta\varphi$ from pulse to pulse. When we examine the power spectrum of the electric field, we observe a set of regularly spaced modes (comb teeth), where neighbouring teeth are separated by the repetition rate $f_r = 1/T$, and are offset from zero by the carrier-envelope offset frequency $f_0 = f_r \cdot \Delta\varphi/(2\pi)$. The comb spectrum is centred at the carrier frequency f_c and has a bandwidth $\Delta f \sim 1/\tau$. Figure adapted from Ref. [7].

Finding and characterizing Earth-like planets orbiting Sun-like stars is one of the most challenging goals of modern radial velocity (RV) exoplanet science [8]. Such searches place very demanding requirements on the wavelength calibration of astrophysical spectrographs: specifically, detecting $\sim$ 10 cm/s RV shifts over the course of months to years, corresponding to sub-MHz changes in Doppler-broadened stellar absorption lines that are many GHz broad. Atomic emission lines from hollow cathode lamps and absorption lines from iodine vapour cells have been the workhorse calibration tools. However, these sources neither have uniform spectral coverage nor long-term stability. A

frequency comb that is referenced to an atomic clock provides an excellent solution to this problem, as it provides a very large set of equispaced frequency markers with very accurately known absolute frequencies [9–11].

An optical frequency comb is generated from a pulsed femtosecond laser. In a pulsed femtosecond laser, a delicate balance of dispersion and nonlinearity result in the laser cavity supporting a train of pulses that propagate without spreading, i.e. solitons. Seen in the wave picture, this result is obtained as a result of constructive interference when many cavity modes are phase-locked. To gain more intuition for the structure of the pulse train, it is instructive to examine the electric field (or intensity) associated with the light in both the time and frequency domain, as shown in Figure 1.2. The frequencies of the comb modes are described by only two degrees of freedom, the repetition rate f_r and the carrier-envelope offset frequency f_0. The frequency of the n^{th} mode f_n is then given by the following simple formula:

$$f_n = f_0 + nf_r. \tag{1.7}$$

This simple linear equation is extremely powerful as it describes a direct relationship between radiofrequencies f_r, f_0 and optical frequency f_n.

Stabilizing the pulse train amounts to locking both degrees of freedom f_r, f_0. Detecting f_r is relatively straightforward as it requires only a fast photodiode. Stabilizing it is then simply a matter of comparing the detected f_r to a local oscillator by mixing and low-pass filtering the result to obtain an error signal, which is easily corrected using a proportional-integral-derivative (PID) controller. The feedback is applied to one of the cavity mirrors using a piezoelectric actuator, as the period T is simply given by c/L, where L is the cavity roundtrip length. The carrier-envelope-offset, on the other hand, is quite nontrivial to detect. One popular method, known as *self-referencing*, requires the comb to be octave-spanning. This means that the comb spectrum contains both some frequency f as well as double this frequency $2f$. With the exception of certain few-cycle lasers, the output bandwidth of frequency combs is insufficient to meet this requirement, so accomplishing this is done with the aid of a microstructured fiber which performs spectral broadening. Once the comb is octave-spanning, one can divert some comb mode from the red tail $f_0 + nf_r$ and frequency

double it using a second harmonic generation crystal to obtain $2f_0 + 2nf_r$. When compared to the comb mode $f_0 + 2nf_r$ already existing in the blue tail of the original comb by interfering the two beams of light, a beat note at frequency f_0 is observed. Since this interferometry requires a nonlinear process, this setup is called a f:$2f$ interferometer. Once the beat is detected, it is mixed with a local oscillator (much like repetition rate) to generate the error signal for PID locking. The feedback is applied typically applied to the pump power, as the cavity dispersion is nonlinear. It is important to remember that the repetition rate and the carrier-envelope offset frequency are not independent degrees of freedom in terms of laser parameters, so in practice, simultaneous locking of the two frequencies can prove challenging.

1.3 Astro-combs

High-resolution astrophysical spectrographs used for Doppler exoplanet searches have extremely high dispersion and are able to achieve frequency resolutions of order ~ 1 GHz in the visible. However, this is still at least an order of magnitude larger than the ~ 100 kHz Doppler shifts associated with an exo-Earth. Since such measurements are inherently photon-flux-limited and require spectral sensitivity much better than the resolution of the spectrograph, they require combining information from thousands of just-resolved spectral lines across the passband of the instrument. Doing this reliably over orbital timescales requires an extremely stable calibration source with a large bandwidth and uniform spectral coverage. State-of-the-art spectrographs used in Doppler exoplanet searches such as HARPS [12] and HARPS-N [13] operate in the visible wavelength range (400-700 nm). Frequency combs are well suited to calibrating these instruments, and several designs have been successfully demonstrated to date (see Ref. [14] and references therein). Combs specifically optimized for astrophysical applications are termed *astro-combs*.

Existing astro-comb architectures are based on near-infrared source combs (e.g. Ti:sapphire, Yb/Er fiber), so providing visible calibration light for an astrophysical spectrograph typically requires a nonlinear optical element to coherently shift and broaden the source comb radiation. Early astro-combs derived from Ti:sapphire source combs relied on second harmonic generation [15, 16] but had

limited utility due to extremely low output bandwidth ($\sim$ 15 nm). This is a serious shortfall because the exoplanet detection sensitivity depends on the bandwidth of the observed stellar light that is calibrated. A calibration source with constant line spacing over a larger bandwidth can therefore enable more precise determination of stellar Doppler shifts.

A better alternative to frequency doubling is to pump a highly nonlinear photonic crystal fiber (PCF) with the source comb to take advantage of supercontinuum generation [17], an effect where a narrowband high-intensity pulse experiences extreme spectral broadening as a result of interactions with the medium through which it propagates. The dispersion and nonlinearity of PCFs may be engineered via suitable changes in geometry; for example, two commonly used parameters are the pitch and the diameter of the air holes. PCFs also typically exhibit very high nonlinearities compared to standard optical fibers due to their small effective mode field diameters. As a result of these attractive properties, broadband supercontinuum generation using PCFs has found many applications, from optical coherence tomography [18] to carrier-envelope phase stabilization of femtosecond lasers [19], as well as calibration of astrophysical spectrographs [20–24].

Astro-combs need to have a large mode spacing ($\gtrsim$ 10 GHz) to match the resolution of astrophysical spectrographs (optimal calibration requires comb teeth spaced apart by $\sim$ 2.5 resolution elements [25]). Ti:sapphire-based astro-combs facilitate this because these lasers are available with intrinsically larger mode spacing ($\sim$ 1 GHz) compared to fiber lasers ($\sim$ 250 MHz). But contrary to their fiber laser counterparts, these lasers require occasional realignment. Recently, however, this disadvantage has been overcome with the advent of turn-key Ti:sapphire frequency combs. It should be noted that even though 1 GHz is closer to the target spacing of $\gtrsim$ 10 GHz, there is still a gap to bridge to achieve this coarser comb structure. This is achieved with the use of mode filtering, i.e. one or more Fabry-Perot cavities with a free spectral range (FSR) that is an integer multiple m of the source comb's repetition rate f_r. In other words, the mode filter performs the repetition rate multiplication mf_r. The main challenge with designing mode filters is dispersion-engineering; in the absence of dispersion control, the cavities do not have a constant FSR over the entire passband, and the comb teeth quickly fall out of lockstep with the cavity modes. Zero group delay dispersion (GDD) mirror pairs [26] provide an excellent solution to this problem, at least for bandwidths of

∼100 nm. In a nutshell, the idea of zero GDD mirror pairs is to construct a sets of mirrors such that the sum of penetration depths for both mirrors is constant regardless of wavelength. The mirror coatings are not symmetric, so long-wavelength light may penetrate more deeply into the one mirror and less deeply into the other mirror (and vice versa for short-wavelength light).

In addition to large spectral bandwidth and large mode spacing, another property that is desired for astronomical calibration applications is spectral flatness, i.e., low intensity variation across the band of the calibration source. Spectral flatness is valued because calibration precision is both shot-noise- and CCD-saturation-limited. Therefore the largest photoelectron count below CCD saturation per exposure provides the optimal calibration. This condition is best fulfilled for a uniform intensity distribution over all the comb peaks [21].

Although astro-combs have been successfully demonstrated [9–11, 20, 22, 23, 27–30], they have not yet seen widespread adoption as primary calibrators in the astronomical community due to their complexity and cost. The operation of astro-combs has so far required significant laser expertise. It is therefore imperative to simplify the use of astro-combs [22, 31] to make them viable for the next generation of high-precision RV measurements.

In this remainder of the chapter, we describe in detail our astro-comb setup, a set of test measurements, and the design and testing of a optimized PCF for our application. Finally, we discuss progress on remote control of the comb and outline automation possibilities as well as ideas for future astro-comb designs.

1.4 Optical setup

The astro-comb is located at the Telescopio Nazionale Galileo (TNG), on the island of La Palma in the Canary Islands, where it is used to calibrate the HARPS-N spectrograph [13]. Figure 1.3 shows a schematic of the experimental setup. Here, we describe the second generation of the astro-comb (see Ref. [23] for more details on the first generation). The light source is the taccor comb

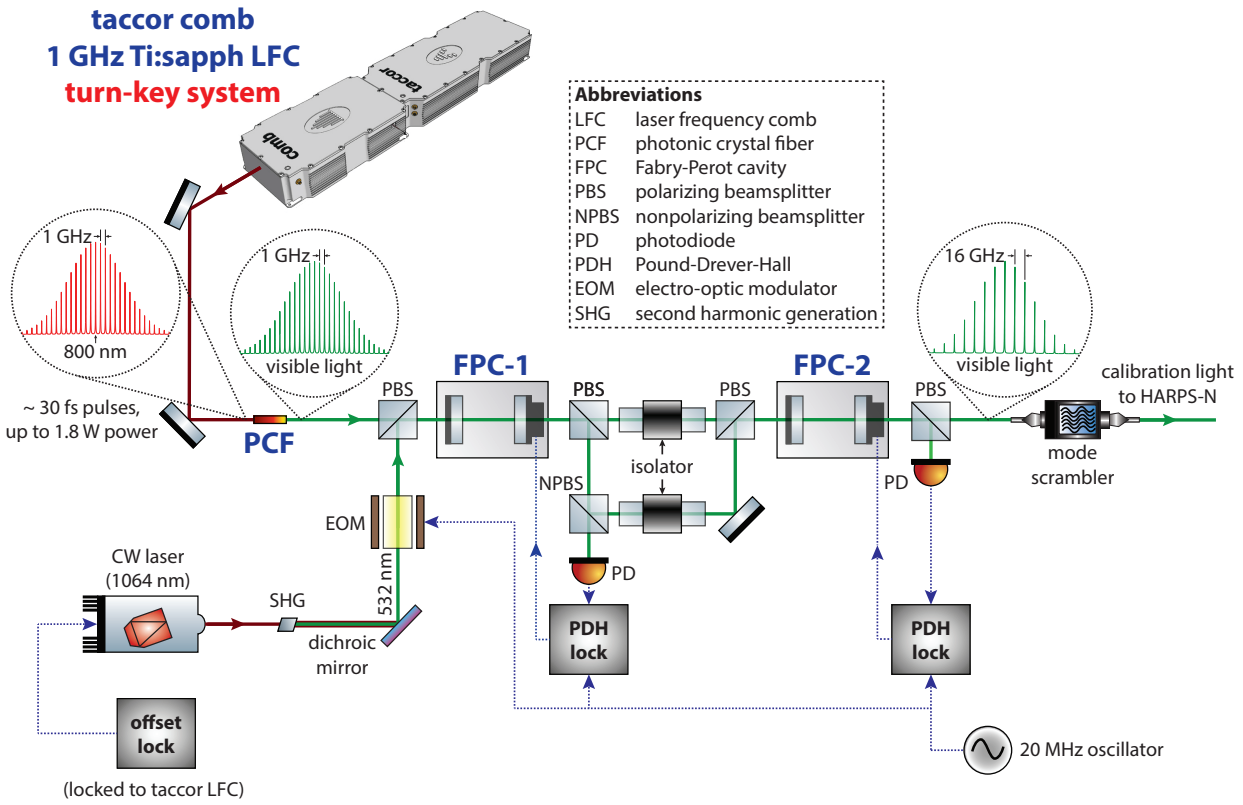

Figure 1.3: Astro-comb block diagram showing a turn-key frequency comb that generates 1 GHz-spaced comb teeth about a centre wavelength of 800 nm, a photonic crystal fiber that coherently shifts the light into the visible wavelength range, and custom-built Fabry-Perot cavities that suppress 15 out of every 16 lines so as to match the resolution of the astrophysical spectrograph to be calibrated. (Abbreviations defined in figure; see text for more details.)

(Laser Quantum), based on a turn-key 1 GHz Ti:sapphire mode-locked laser with a carrier-envelope offset (CEO) detection unit. The frequency comb operates at a centre wavelength of about 800 nm and outputs $\sim$ 30 fs pulses at a repetition rate f_r stabilized to a $\approx$ 1 GHz sourced from a RF synthesizer. The frequency comb carrier-envelope offset frequency f_0 is detected by sampling the optical output and sending it through a short length of nonlinear fiber to an f-2f detection unit, which directly locks the carrier-envelope offset frequency to 60 MHz. Both synthesizers are referenced to a GPS-disciplined 10 MHz Rb clock.

Approximately 100 mW of the source light is coupled into a dispersion-engineered tapered PCF. This light is then filtered by two 16 GHz FSR Fabry-Perot cavities in series, achieving $>$ 40 dB suppression of undesired comb teeth [23]. The broadband cavities, which are based on zero group delay dispersion mirror pairs [26], are optimized for operation between 500-650 nm. The cavity lengths are stabilized to a frequency-doubled 1064 nm CW single-frequency laser (JDSU NPRO 126N-1064-500) using a Pound-Drever-Hall scheme [32] in transmission. As the cavities have residual dispersion, the frequency of the CW laser is empirically tuned to maximize the bandwidth transmitted by the cavities. This laser, in turn, is offset-locked to the comb.

The spectrally broadened and filtered astro-comb light is coupled into a multimode fiber and sent through a mode scrambler [23] to eliminate dynamic modal noise. The light is then sent via multimode fiber to HARPS-N [13], which is a high resolution ($R = 115{,}000$) cross-dispersed echelle spectrograph with spectral coverage from 380 nm to 690 nm. HARPS-N achieves $\sim$ few m/s RV stability prior to calibration by careful design, operation in vacuum, and temperature stabilization on the millikelvin level. Crucial to achieving sub-m/s RV observations is wavelength calibration, as well as simultaneous monitoring of potential calibration drifts while science exposures are performed. To this end, two input channels to HARPS-N are present: star light is injected into the "science channel" and calibration light into the "reference channel."

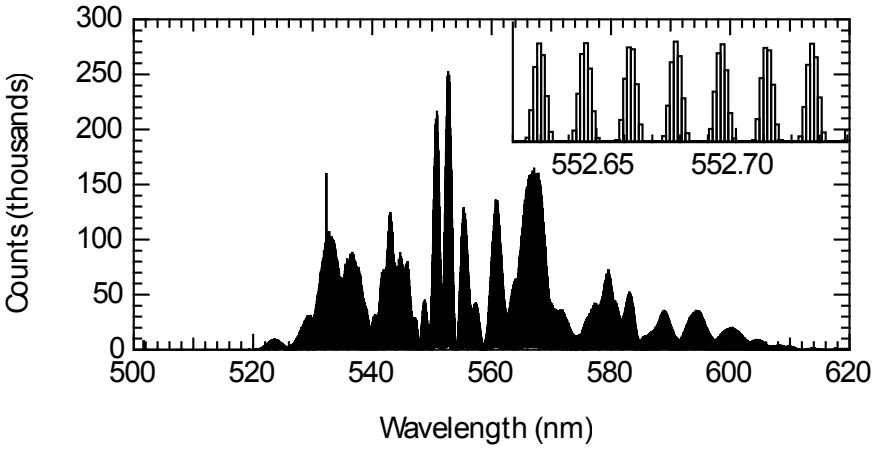

Figure 1.4: Example astro-comb spectrum measured with HARPS-N spectrograph. This particular spectrum was recorded during a test with a commercial supercontinuum device (not an optimized photonic crystal fiber). The strong peak at 532 nm is the CW reference laser used to stabilize the filter cavity lengths. (Inset) At higher spectral resolution, individual comb teeth (spectral lines) are observed.

1.5 Testing the astro-comb

1.5.1 Measurement of the astro-comb spectrum

Figure 1.4 shows an example spectrum of the astro-comb, as measured on the HARPS-N spectrograph. On a finer scale, individual comb spectral lines spaced by about 16 GHz are visible. The contrast of astro-comb peaks (peak height divided by background level) on HARPS-N is approximately 100. We believe this is due to long tails in the HARPS-N instrument profile as the contrast of comb lines from a similar source observed on a high resolution Fourier transform spectrometer is >10,000 [23]. The strong modulation in the envelope of the spectrum is a result of the nonlinear processes in the supercontinuum device used for spectral broadening. Improved spectral uniformity and even extending spectral coverage can be addressed by designing an optimized photonic crystal fiber for this frequency comb.

Ultimately, the bandwidth of our astro-comb is limited by the residual dispersion of filtering Fabry-Perot cavities [33]. Extending cavity mirror spectral coverage is currently under investigation.

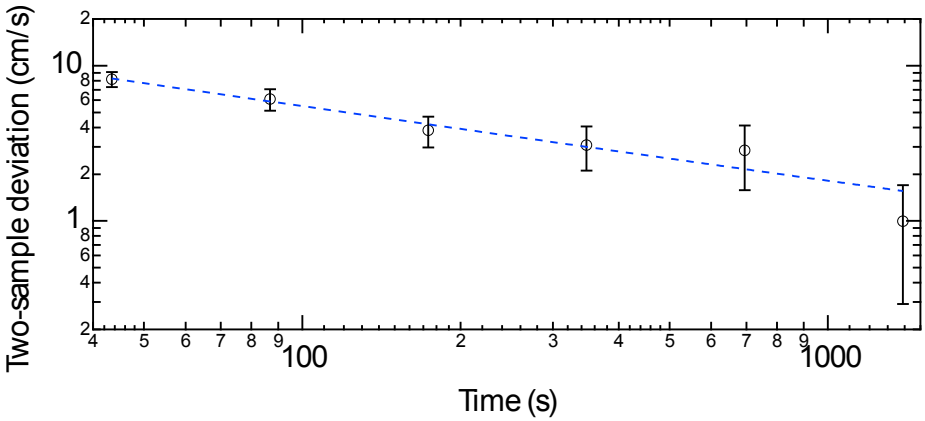

Figure 1.5: Two-sample deviation of the measured frequency stability (in radial velocity units) of the spectrograph vs. averaging time. The black circles are the measured stability and dashed line is a fit consistent with the photon (shot) noise floor. The overall duration of the measurements was approximately 1 hour. (Figure courtesy of D. F. Phillips)

An alternative approach that has been successfully demonstrated [20, 28, 29, 31] is to filter a narrow-band comb with a series of narrow-band high-finesse cavities and then perform the spectral broadening with a nonlinear fiber. This approach extends the bandwidth of the comb at the expense of system complexity, including added demands on Fabry-Perot cavity performance.

1.5.2 Spectrograph stability characterization

To collect information about frequency stability of the spectrograph, we injected astro-comb light into both channels of the HARPS-N spectrograph and monitored the deviation of their difference as a function of averaging time as shown in Fig. 1.5. We consistently achieved RV sensitivity of nearly 1 cm/s at one half hour with no signs of spectrograph drift. Moreover, the two-sample deviation is consistent with the photon noise limit up to this point. This wavelength calibration is more than sufficient for RV detection of an Earth-analog exoplanet ($\sim$ 9 cm/s RV modulation).

1.5.3 Comb-referenced solar spectra

As an example astronomical application, we performed solar spectral observations with HARPS-N referenced to the astro-comb. Using an automated compact solar telescope located at the TNG

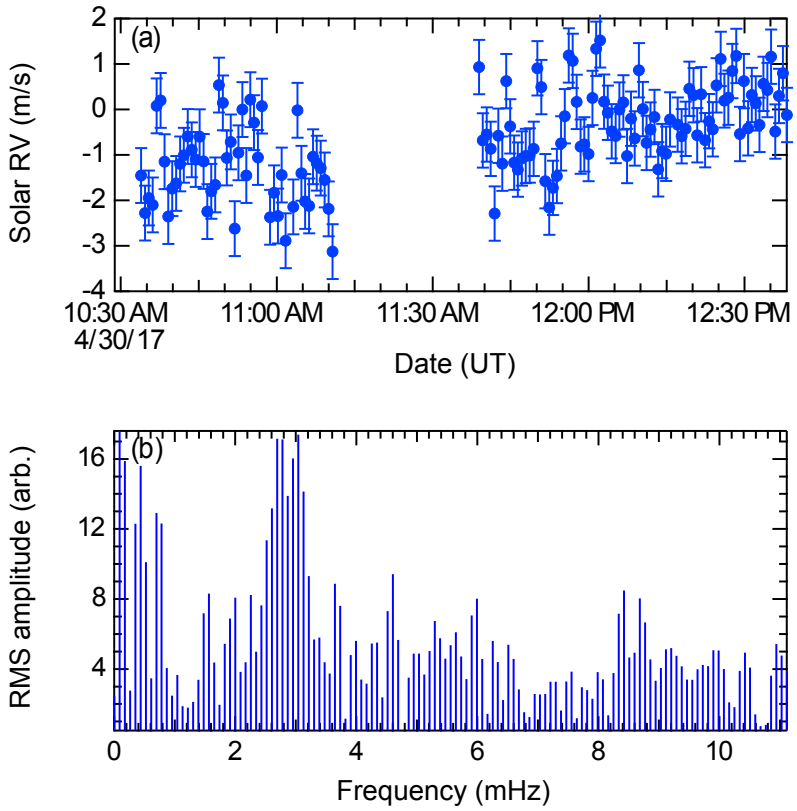

Figure 1.6: (a) Comb-referenced solar RV observations. (b) Power spectrum of the data. (Figure courtesy of D. F. Phillips)

facility [34], we collect light from the full disk of the Sun and feed an integrating sphere to discard spatial information. This light is then injected into the science channel of the HARPS-N spectrograph, with the astro-comb light simultaneously injected into the reference channel. We took 20 second exposures for several hours, with a break due to clouds. In these short-term measurements, the astro-comb primarily provides a simultaneous reference as the radial velocity is only measured relative to the initial reference exposures for a few hours. We also derive a wavelength solution from the astro-comb spectrum injected in the science channel. Figure 1.6a shows the difference between the observed RVs (calculated using a cross correlation technique with an empirical template and averaged across the orders of HARPS-N with significant astro-comb light) and the expected RVs from the JPL *Horizons* ephemeris. The power spectrum of these differences (Fig. 1.6b) prominently shows the 5-minute (3 mHz) p-mode solar acoustic oscillations. However, there is also significant low-frequency noise present that is likely associated with granulation effects in the Sun [35].

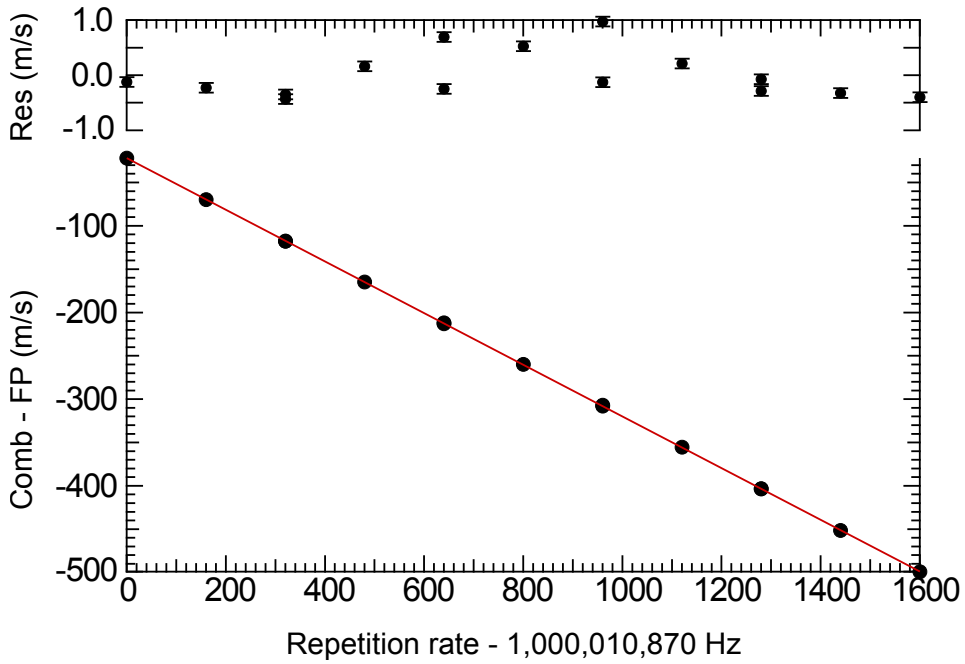

Figure 1.7: Measured mean difference in astro-comb spectral lines relative to a white light FP spectrum (black circles) as a function of comb repetition rate, with a linear fit (red line), and residuals (top trace). The error bars of the residuals plot are the statistical uncertainty of ≈ 10 cm/s (see text). (Figure courtesy of D. F. Phillips)

1.5.4 Studies of intrapixel sensitivity variations of the spectrograph CCD

A source of systematic error in RV measurements with an astrophysical spectrograph is non-uniformity in the detector. One example is intrapixel sensitivity variations [36–38] in the CCD. To study this effect, we took a sequence of exposures while shifting the astro-comb lines across half a pixel. Specifically, we shift the comb repetition rate f_r since even a small change in f_r gets magnified by the mode number n through the relation defining the frequency of the n^{th} line, $f_n = f_0 + n f_r$. For example, a step of 3.2 kHz in the repetition rate amounts to moving a comb tooth across an entire pixel (~ 1.6 GHz). Note that we continue to take $f_r = 1.000010870$ GHz to remain consistent with our discussion of the source comb; however, only every 16^{th} line appears on the spectrograph.

To track the shifting astro-comb teeth as the repetition rate is changed, we inject comb light into the science channel of the spectrograph and light from a from a passively stabilized Fabry-Perot (FP) cavity illuminated by broadband white light (a laser-driven plasma source) into the reference

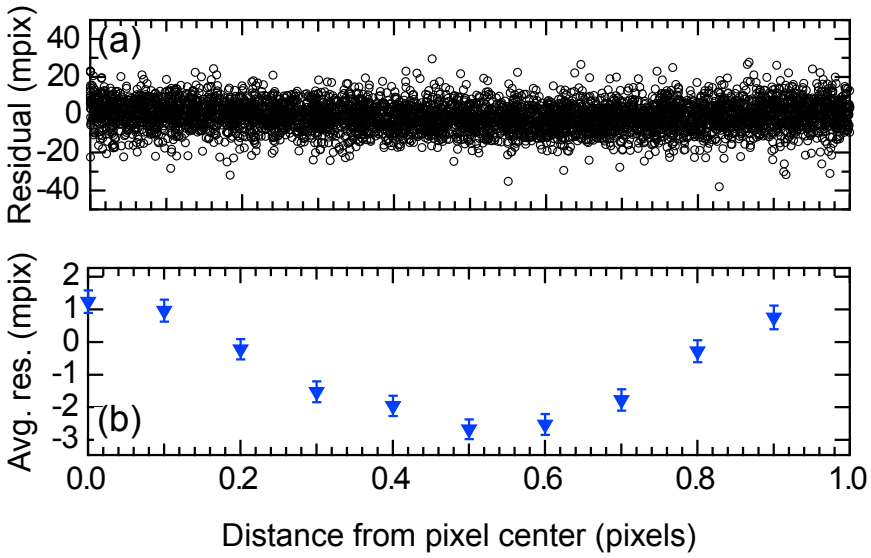

Figure 1.8: (a) Difference between expected and actual astro-comb peak position, as a function of peak position inside a pixel. The uncertainty for each point is consistent with the $\sim$ 10 millipixel scatter observed. (b) Data shown in panel (a) averaged into bins with 0.1 pixel width. (Figure courtesy of D. F. Phillips)

channel as a fixed reference. Using a standard cross-correlation analysis, we calculate the mean shift in all the lines relative to the FP spectrum, assuming that the drift of the FP cavity is negligible during the entire measurement. Figure 1.7 shows the results of this analysis: a linear shift due to a change in repetition rate of the astro-comb of 0.298(2) m/s/Hz, implying $n \approx 523{,}000$ which is a good estimate for the mean value of n in $f_n = f_0 + nf_r$. The upper panel of Fig. 1.7 shows the residuals from the fit with error bars representing the statistical uncertainty of $\approx$ 10 cm/s. We attribute additional scatter in the residuals to systematic errors induced by changes in astro-comb power vs. wavelength as the repetition rate and reference laser are retuned. These variations do not affect the results reported below and in Fig. 1.8 as the analysis is henceforth done for each astro-comb peak independently.

To search for intrapixel sensitivity variations, we fit all 365 astro-comb spectral peaks in a single echelle order at $\sim$ 570 nm (near the centre of the astro-comb band) on the HARPS-N spectrograph to Gaussian profiles. As seen in the inset of Fig. 1.4, typical peaks have a 3-4 pixel full width at half maximum. We model the mean intrapixel sensitivity variation of all the pixels in this order as follows: let P_i^j be the pixel value of the fitted peak position of the i^{th} peak in the j^{th} exposure,

where i runs from 0 to 364, while j runs from 0 to 15. $\text{frac}\left(P_i^j\right) = P_i^j - \left\lfloor P_i^j \right\rfloor$ gives the fractional part of the peak position. Note that a value of 0 or 1 corresponds to pixel centre while a value of 0.5 corresponds to a pixel boundary. We then compute a residual R_i^j of the shift of the observed peak i in exposure j relative to the initial exposure compared to the expectation from shifting the repetition rate f_r as

$$R_i^j = \left(P_i^j - P_i^0\right) - \Lambda\left(\Delta f_r, P_i^j\right) - \Delta_{\text{FP}}^{j,0}\left(P_i^j\right) \tag{1.8}$$

where $\Lambda\left(\Delta f_r, P_i^j\right)$ is the expected shift in pixels of the comb peak at P_i^j due to the change in repetition rate Δf_r. The additional correction $\Delta_{\text{FP}}^{j,0}\left(P_i^j\right)$ is the local shift (i.e. interpolated to P_i^j) of the Fabry-Perot simultaneous reference between the initial exposure and the j^{th} exposure. This parameter is extracted from a wavelength solution derived from the FP spectrum. Figure 1.8a shows R_i^j as a function of $\text{frac}\left(P_i^j\right)$, for all i, j. Averaging the data shown in Fig. 1.8a into bins of 0.1 pixels leaves a systematic residual at roughly the 5 millipixel level, as shown in Fig. 1.8b, which corresponds to $\sim$ 5 m/s in RV units. Averaging over all the observed lines in all the orders with different fractional CCD pixel values should reduce this systematic error below the 1 m/s level. Key challenges for future work are to determine such intrapixel sensitivity variations across the full HARPS-N spectrum, and to mitigate its effects on astronomical RV observations.

1.6 Visible frequency comb generation using dispersion-engineered waveguides

In this section, we show that careful dispersion-engineering of a PCF by tapering enables the production of a very broad, flat, and high-intensity optical supercontinuum spectrum from pump radiation emitted by a turn-key Ti:sapphire comb. Previous attempts to produce such a supercontinuum from a Ti:sapphire comb have reported relatively low spectral coverage (500-620 nm) and large intensity variations ($\sim$ 10 dB) across the band [23]. A flat 235 nm-wide visible supercontinuum has been demonstrated using a Yb:fiber comb [31], but at the expense of losses incurred by spectral flattening using a spatial light modulator [21].

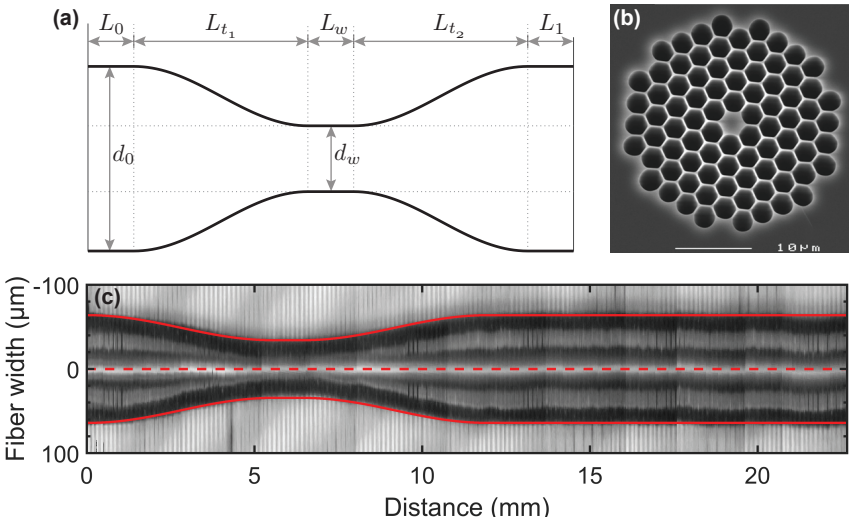

Figure 1.9: (a) General tapered photonic crystal fiber (PCF) geometry showing varying core size vs. length. The quantities d_0, d_w, $L_{0,1}$, $L_{t_{1,2}}$, and L_w parametrize the geometry; see text for details. (b) SEM micrograph (courtesy of NKT Photonics A/S) of end face of NL-2.8-850-02 PCF. (c) Composite microscope image of fabricated tapered PCF. Overlaid red curves represent the proposed geometry, scaled by the cladding/core size ratio.

1.6.1 Fiber geometry and parameters

In our application, we consider a PCF with a 2.8 µm core diameter and an 850 nm zero dispersion wavelength (NKT Photonics NL-2.8-850-02). The cross section of the fiber is shown in Figure 1.9b. We chose a large core diameter to facilitate coupling and have the zero dispersion wavelength (ZDW) near our 800 nm pump wavelength. Changing the PCF core diameter as a function of distance along the fiber via tapering modifies both the dispersion and nonlinearity of the PCF. Sometimes termed "dispersion micromanagement" in the literature, such techniques have been pursued before with PCFs to enable generation of light with increased bandwidth and flatness [39–42], but not targeted specifically toward uniform visible wavelength coverage using GHz repetition rate lasers.

To design a device capable of producing a flat visible-spanning spectrum when pumped with 800 nm femtosecond pulses, we model optical pulse propagation in the PCF by solving the generalized nonlinear Schrödinger equation (GNLSE) as outlined in Appendix A. In our design, we consider a specific taper geometry, as shown in Figure 1.9a. The core diameter d changes smoothly over

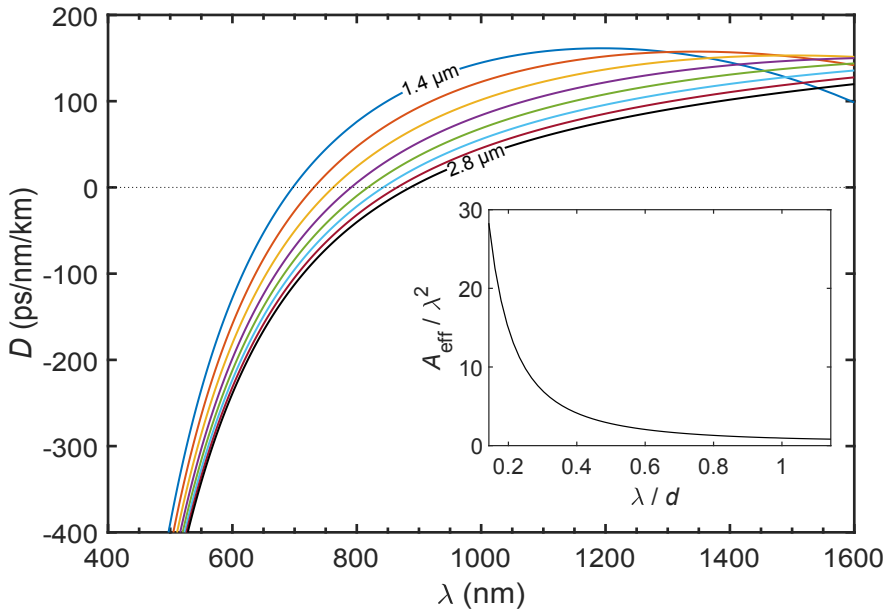

Figure 1.10: Simulated optical properties of tapered PCF. Main plot: chromatic dispersion curves for varying core diameter d (curves spaced apart by $\Delta d = 0.2$ μm). Inset: normalized effective area vs. normalized wavelength.

the tapers with a cosine function $d(\zeta) = d_i + \frac{1}{2}(1 - \cos\pi\zeta)(d_f - d_i)$ over $\zeta \in [0, 1]$, where d_i and d_f are, respectively, the initial and final core diameters of the down- or up-taper. More explicitly, $d_i = d_0$, $d_f = d_w$ for the down-taper and $d_i = d_w$, $d_f = d_0$ for the up-taper. The nondimensionalized variable ζ parametrizes the distance along the taper, i.e., $\zeta = (z - L_0)/L_{t_1}$ for the down-taper and $\zeta = (z - (L_0 + L_{t_1} + L_w))/L_{t_2}$ for the up-taper.

Over the desired range of core diameters (1.4 μm $\leq d \leq$ 2.8 μm), we compute the PCF dispersion and nonlinearity as inputs for the pulse propagation calculations. We use a commercial finite-difference mode solver (Lumerical MODE Solutions) to calculate the PCF properties (see Appendix B for more details). Figure 1.10 shows the chromatic dispersion $D = -\frac{\omega^2}{2\pi c}\frac{\partial^2\beta}{\partial\omega^2}$ versus wavelength λ as a function of core diameter d (β is the propagation constant of the fundamental mode and ω is the angular frequency of light). Additionally, the inset of Figure 1.10 shows the modal area A_{eff} calculation for the PCF. This calculation is required only for a single core diameter because A_{eff} respects the scale invariance of Maxwell's equations, as pointed out in Ref. [43], whereas dispersion does not. In subsequent results where we solve the GNLSE in a tapered PCF, the dispersion was interpolated to the local fiber diameter from a pre-computed set of dispersion curves spaced apart

by $\Delta d = 0.05$ µm. In addition, the effective area was evaluated for each PCF diameter according to the computed curve in terms of the normalized variables shown in the inset of Figure 1.10.

1.6.2 Tapered fiber design

We design a taper geometry for our PCF with the following parameters: $d_0 = 2.8$ µm, $d_w = 1.5$ µm, $L_0 = 50$ µm, $L_{t_1} = L_{t_2} = 5.5$ mm, $L_w = 900$ µm and $L_1 = 9.55$ mm, as shown in Figure 1.11a. These geometric parameters were manually chosen to produce a supercontinuum with low intensity variation in the visible wavelength range while trying to push the blue edge of the supercontinuum to wavelengths as short as possible. A design derived using an optimization method with a merit function capturing these desired qualities may lead to enhanced performance, but is outside the scope of this work.

The pump radiation for the PCF is sourced from a taccor comb (Laser Quantum). This system is a turn-key 1 GHz Ti:sapphire mode-locked laser with repetition rate and carrier-envelope offset stabilization. The Ti:sapphire mode-locked laser is based on the technology described in Ref. [44], and is referenced to a GPS-disciplined Rb frequency standard, yielding a fractional stability of better than 10^{-12} [9]. Therefore, the source comb will not limit the frequency stability of the astro-comb [24]. The comb centre wavelength is ~ 800 nm and the laser output bandwidth is approximately 35 nm, so the transform-limited intensity full width at half maximum (FWHM) is 27 fs for a Gaussian pulse shape.

By integrating the GNLSE (see Appendix A) with a fourth-order Runge-Kutta method, we investigate idealized pulse evolution in the tapered PCF assuming an initial Gaussian pulse shape of the form $\sqrt{P_0}\exp\left(-2\ln 2\,(T/T_0)^2\right)$. Here, T_0 is the pulse FWHM, $P_0 \approx 0.94E_p/T_0$ is the pulse peak power, and E_p is the pulse energy. Figure 1.11b and d show the pulse evolution in the spectral and time domain, respectively for a 215 pJ pulse injected into the tapered PCF, whose geometry is represented in Figure 1.11a. Panels c and e show the spectrum and time trace of the pulse at the fiber output. Panel c also shows the magnitude of the first-order coherence $g_{12}^{(1)}$ of the output (see

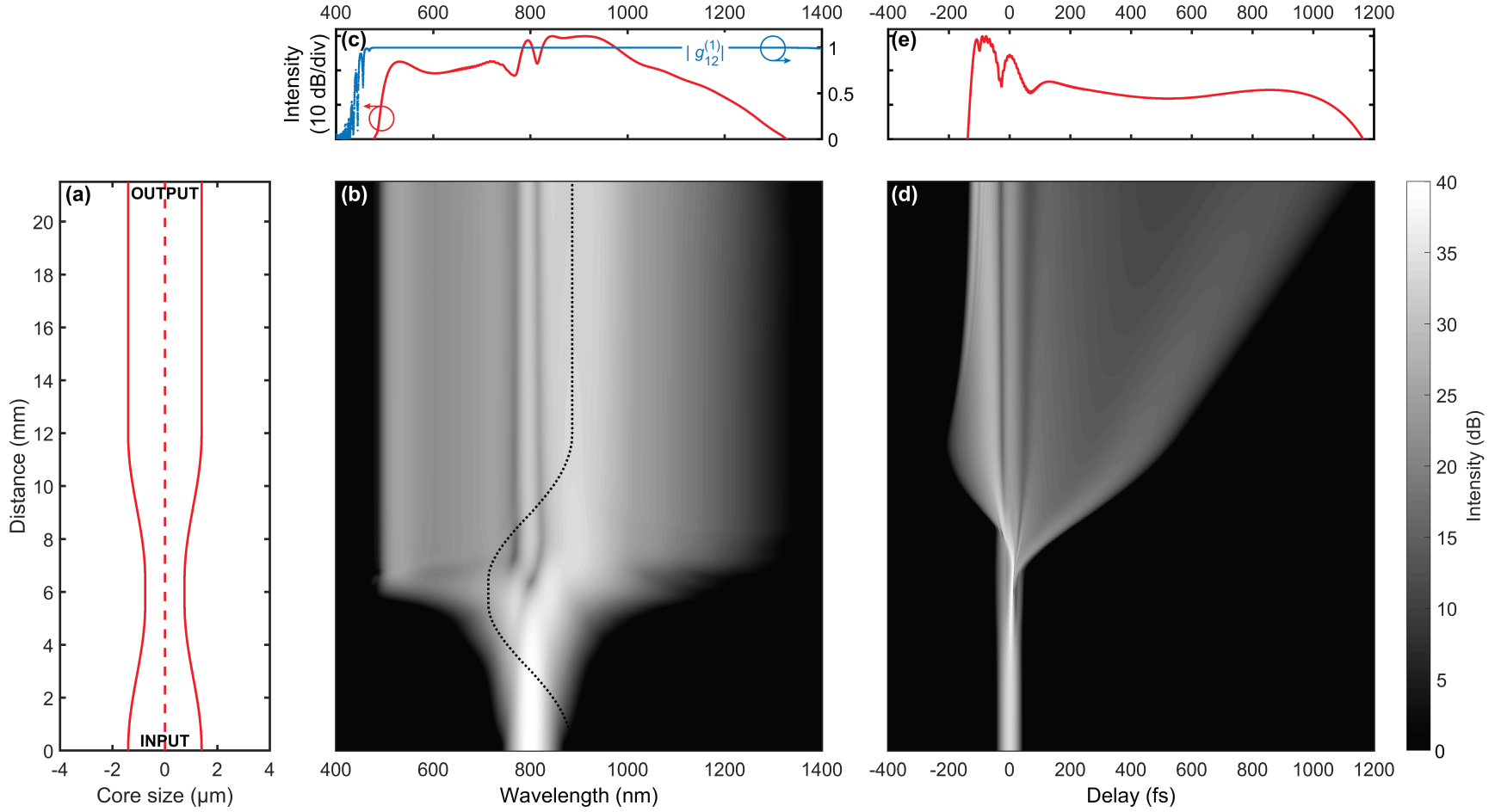

Figure 1.11: Simulated spectral and temporal evolution in a tapered PCF of a 215 pJ, 27 fs Gaussian pulse at 800 nm. (a) PCF core size vs. longitudinal distance (0 is the fiber input face). (b) Pulse evolution vs. distance in the spectral (wavelength) domain. The PCF's computed zero dispersion wavelength (ZDW) is superimposed as a dotted black curve. (c) Pulse spectrum and magnitude of the first-order coherence $g_{12}^{(1)}$ at the fiber output face (d) Pulse evolution vs. distance in the time domain. (e) Pulse time trace at fiber output face.

Appendix A for more details). Focusing on the panels b and d, we see that the initial dynamics correspond to symmetric spectral broadening and an unchanged temporal envelope associated with self-phase modulation [45]. This is because the pump pulse is initially propagating in the normal dispersion regime (i.e., chromatic dispersion $D < 0$). As the PCF narrows, the pulse crosses over into the anomalous ($D > 0$) dispersion regime. Here, perturbations such as higher-order dispersion and Raman scattering induce pulse break-up in a process called soliton fission [45]. Subsequent to fission, the solitons transfer some of their energy to dispersive waves (DW) propagating in the normal dispersion regime according to a phase matching condition [46]. DW generation (also known as fiber-optic Cherenkov radiation), the phenomenon responsible for generating the short-wavelength radiation in this case, has been thoroughly studied in the past [47, 48]. The supercontinuum bandwidth is dependent on the pump detuning from the ZDW, so typically one can achieve larger blue shifts in constant-diameter PCFs by using smaller core sizes, but at the expense of a spectral gap opening up between the pump and DW radiation [45]. This issue can be addressed by using a tapered geometry, where the sliding phase matching condition along the narrowing PCF can generate successively shorter wavelength DW components [39–42]. Beyond the taper waist, the spectrum stabilizes as a result of the relaxation of the light confinement and corresponding reduction in nonlinearity. This enables us to obtain a structure-free, flat band of light spanning 500-700 nm (Fig. 1.11c). The light is also extremely coherent ($\left|g_{12}^{(1)}\right| \approx 1$) across the whole spectral region containing significant optical power.

1.6.3 Experimental results

Following the above design, the tapered PCF was fabricated from NL-2.8-850-02 fiber by M. Harju at Vytran LLC on a GPX-3000 series optical fiber glass processor using a heat-and-pull technique. Measurements from a micrograph of the fabricated device (Fig. 1.9c) agree well with the design geometry. We tested the tapered PCF by clamping it in a V-groove mount and pumping it with light from a taccor comb (Laser Quantum), recording the output spectrum on an optical spectrum analyzer (OSA). An 8 mm effective focal length aspheric lens (Thorlabs C240TME-B) was used for in-coupling, and a microscope objective (20× Olympus Plan Achromat Objective, 0.4 NA, 1.2 mm

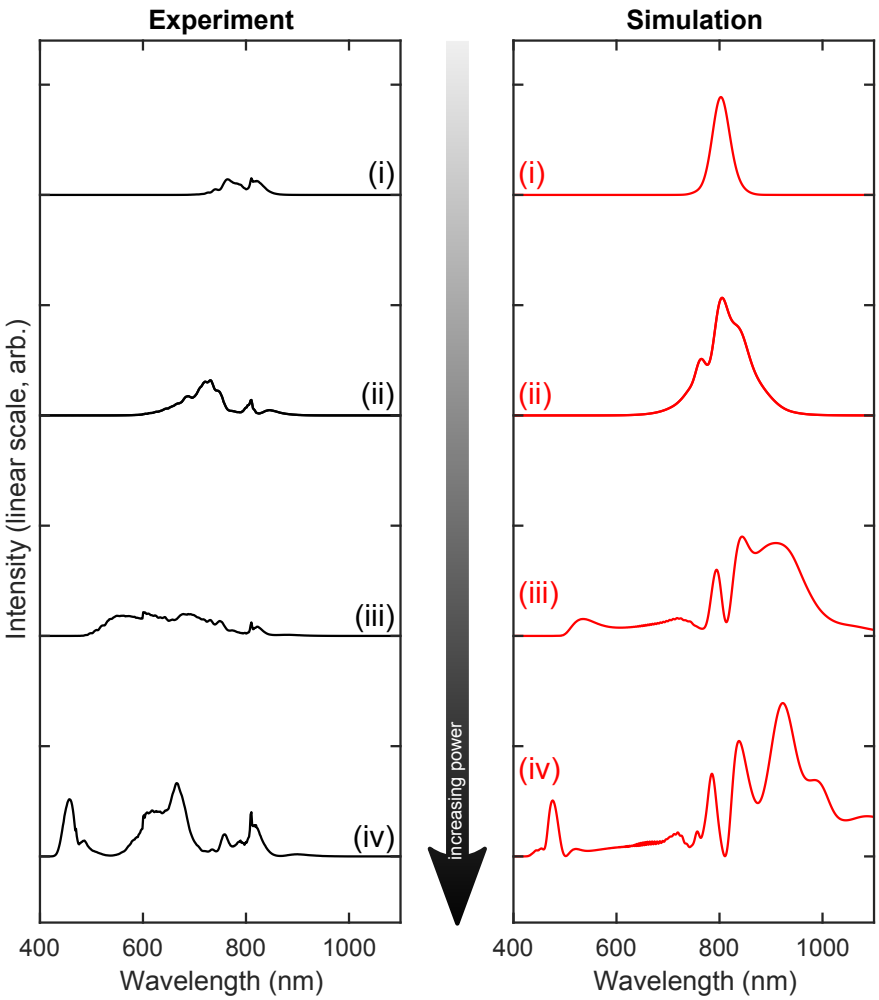

Figure 1.12: Comparison of experimental (left panel) and simulated (right panel) output spectra from tapered PCF pumped by taccor source comb as a function of coupled pulse energy. The spectra are offset for clarity. Except where noted, all traces within each panel have the same scale and are normalized such that the area under each curve is proportional to the pulse energy. Experimental and simulated pulse energies for each row are, respectively, (i) 24 pJ, 50 pJ, (ii) 63 pJ, 130 pJ, (iii) 100 pJ, 215 pJ, (iv) 190 pJ, 290 pJ. The discrepancies in the pulse energies and spectral features are discussed in the text.

WD) was used for out-coupling. A dispersion compensation stage based on chirped mirrors was used prior to coupling to counteract the chirp induced by downstream optical elements and bring the pulse FWHM at the fiber input close to its transform limit. A half-wave plate was also inserted into the beam path before the in-coupling lens for input polarization control. Coupled power was measured with a thermal power meter just after the out-coupling microscope objective. Light was coupled into the OSA via multimode fiber (Thorlabs FG050LGA) using a fixed focus collimator (Thorlabs). We recorded spectra for a variety of femtosecond laser pump powers, optimizing the coupling before each measurement. For this series of measurements, the wave plate angle $\phi_{\lambda/2}$ was kept constant. All measurements were performed with the same equipment, the same dispersion compensation, and on the same day.

We compared our measured and simulated output spectra for the tapered PCF (Fig. 1.12) as a function of increasing pump power. The pulse energies shown in the right panel (simulations) were chosen to produce representative spectra. In comparing simulations and measured spectra (Fig. 1.12), we find two principal discrepancies: (1) the simulated pulse energies are a factor of ~ 2 larger than the corresponding experimental pulse energies, and (2) there is very little radiation observed at wavelengths longer than the pump wavelength in the experimental spectra. The first discrepancy may be due to the combined effect of (geometric) out-coupling loss and reduced infrared transmission through the out-coupling objective, which is optimized for visible light; this mechanism lowers the measured out-coupled power compared to simulated spectra since the power is measured after the objective. These effects are thought to also contribute to the second discrepancy to some extent, but it is suspected that the dominant contributions come from chromatic effects in coupling into the multimode fiber used with the OSA as well as the reduced long-wavelength sensitivity of the OSA. Finally, there are uncertainties in fiber properties (both geometric and optical) and laser parameters (e.g., coupled bandwidth) which lead to uncertainties in the expected spectral profiles.

Despite the differences described above, the simulations qualitatively reproduce the trend in the visible region quite well: initial symmetric broadening is followed by a wide, flat spectral plateau forming to the blue side of the pump; also, at high pump powers, the blue-shifted radiation separates into a distinct feature with a spectral gap between the pump and the dispersive waves. Experimen-

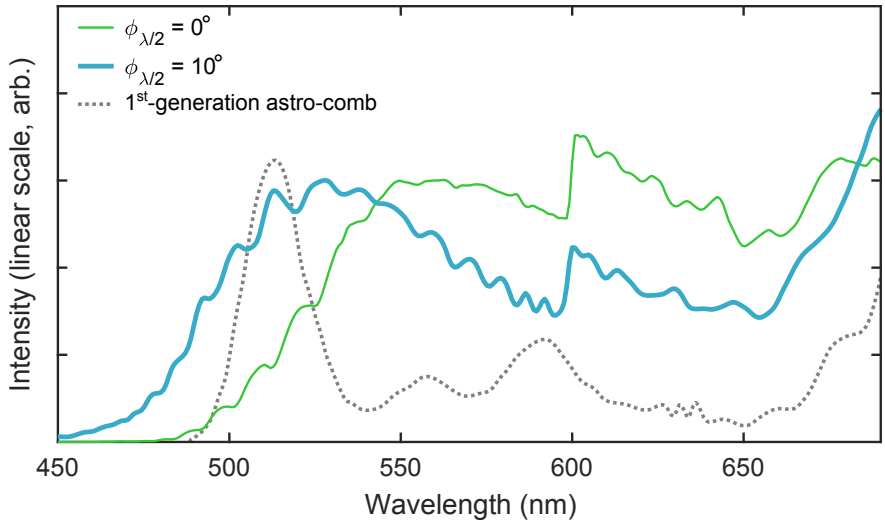

Figure 1.13: Measured output spectra for tapered PCF pumped by taccor source comb at different input polarizations (thin green and thick blue traces): a trade-off between spectral bandwidth and flatness is observed by rotating the input polarization using a half-wave plate at angle $\phi_{\lambda/2}$ (see text for definition). The spectra were recorded with 100 pJ coupled into the tapered PCF. The bandwidth and flatness of the green and blue traces can be compared to the PCF output spectrum from the first-generation astro-comb system [23] (dotted grey trace).

tally, $E_p \approx 100$ pJ is an ideal operating point for an astro-comb, because it produces the flattest spectrum.

In a second test, we varied the input polarization using the half-wave plate at constant coupled pump power. The results are shown in Figure 1.13 (the kink in the thin green and thick blue traces near 600 nm is thought to be an artifact from the OSA). The condition denoted $\phi_{\lambda/2} = 0°$ indicates the wave plate angle that gives the flattest spectrum at $E_p = 100$ pJ (this is the angle at which all traces in the left panel of Figure 1.12 were recorded). At this nominal angle, the intensity variation over the 530-690 nm range is only 1.2 dB. Rotating the wave plate by 10 degrees results in a spectrum with increased bandwidth at the expense of flatness: 3.2 dB variation over 490-690 nm, with a blue tail extending below 450 nm. The polarization dependence likely comes from the fact that single-mode fibers support two modes with orthogonal polarizations. Natural birefringence [17] then leads to different output supercontinua for the two modes. This behaviour cannot be quantitatively understood using our simulations as we use a model formulated using a single mode. A complete multi-mode formulation of the GNLSE [49] would allow us to obtain more insight into the nature of these dynamics, but this is beyond the scope of the current work.

Comparing the usable (visible) comb light from our new design to the previous-generation astro-comb, which was based on a different frequency comb and tapered PCF [23], our new design provides improved bandwidth and spectral flatness. We also estimate the power per comb mode by measuring the transmission through a 10 nm-wide bandpass filter around 532 nm, obtaining values of $\sim 10^2$ nW/mode, which is comparable to the result in Ref. [23]. It thus satisfies the requirements for the astro-comb application. Moreover, the new tapered PCF enables the use of a turn-key Ti:sapphire laser, which greatly simplifies astro-comb design and operation [24].

Our new astro-comb (employing the turn-key laser and tapered PCF described in this work) is expected to reach a radial velocity precision of < 10 cm/s (i.e., < 200 kHz in units of optical frequency) in a single exposure, required for detection of terrestrial exoplanets in the habitable zone around Sun-like stars. Once the new astro-comb is fully deployed, we will verify the stability by injecting the astro-comb light into both channels of the astrophysical spectrograph simultaneously and compute the two-sample deviation of the spectral shift between channels as a function of averaging time [23, 24]. The next major step is to improve the residual dispersion of the Fabry-Perot mode filters [26] used for repetition rate multiplication, so as to preserve all of the bandwidth generated by the PCF. Another viable option would be to split the comb light into several bands and filter each band separately using narrowband cavities [11].

In addition to the calibration of astrophysical spectrographs used for Doppler velocimetry of stars, our tapered PCF design may find applications in optical coherence tomography (OCT) [18]. In OCT systems, the axial (spatial) resolution scales as $\sim \lambda_0^2 / (\Delta\lambda)$, where λ_0 is the centre wavelength and $\Delta\lambda$ is the bandwidth of the source. Hence, the resolution benefits from reducing the centre wavelength and increasing the bandwidth, which is a similar design problem to the one addressed here. Ultrahigh-resolution visible-wavelength OCT has enabled optical sectioning at the subcellular level [50] as well high-speed inspection of printed circuit boards [51]. In such applications, spectral gaps in the output band of supercontinuum sources used in OCT studies degrade the axial resolution below that possible with the full band. Thus, the spectral uniformity possible from the present tapered PCF design may improve resolution further; it may also obviate some of the challenges associated with dual-band OCT, where sophisticated signal-processing techniques are required to

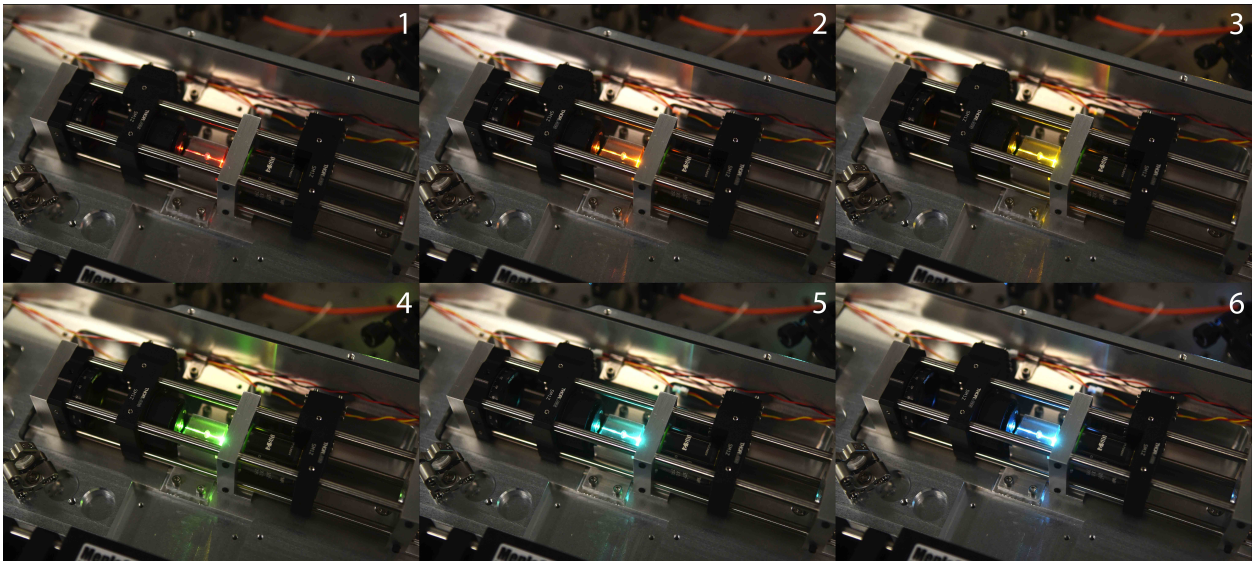

Figure 1.14: Visual appearance of the PCF vs. coupled power (increasing 1 through 6)

combine information from spectrally separated bands [52].

In summary, we demonstrated a tapered PCF that produces spectrally flat light almost spanning the entire visible range when pumped by a turn-key GHz Ti:sapphire laser. Our result represents a marked improvement in the amount of optical bandwidth available to calibrate a visible-wavelength spectrograph. This work also enables a simple visible frequency comb system without the need for spectral shaping.

1.7 Remote-control interface and automation prospects

In an effort to make the astro-comb available for routine calibration and accessible to operators without laser expertise, it is imperative to automate the operation of the system. As a first step toward this, we constructed a remote-control interface so that the comb may be operated without the user being nearby. Figure 1.15 shows a schematic of this interface, including all relevant electronic connections.

The laser is switched on (or off) by using a microcontroller (Arduino Uno) to control a power relay

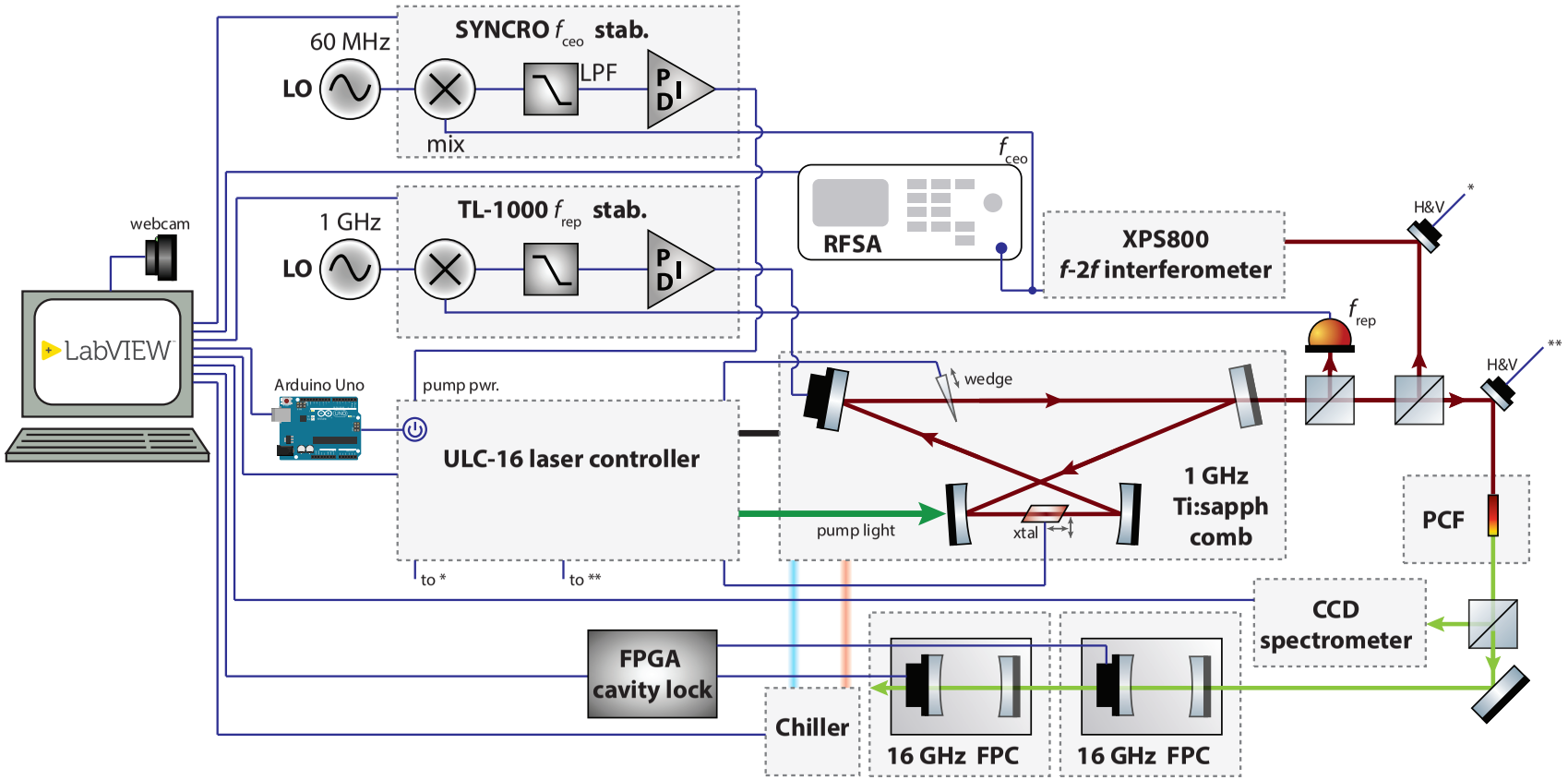

Figure 1.15: Schematic of astro-comb remote control interface (see text for details). Light paths are shown with directional arrows, and electronic connections are denoted by dark blue lines. Reference laser omitted for simplicity.

(Digital Loggers IoT Relay). Upon startup, the laser controller (Laser Quantum ULC-16) goes through an automatic initialization routine which involves temperature stabilization and establishing modelocked operation. The laser controller is capable of adjusting the pump laser power, moving the Ti:sapphire crystal, translating the intracavity wedge, as well as performing horizontal (H) and vertical (V) steering of beams into the PCF and f-$2f$ interferometer. The pulse train from the laser is directly detected on a fast photodiode and compared to a 1 GHz frequency source (LO) for locking by a repetition rate stabilization unit (Laser Quantum TL-1000); feedback is applied by changing the cavity length using a piezoelectric actuator. After this, some light is diverted to a carrier-envelope phase detection unit (Menlo Systems XPS800). The beat note from this module is monitored on a radio frequency spectrum analyzer (RFSA) as well as sent to a stabilization unit (Menlo Systems SYNCRO) which compares it to a 60 MHz frequency source (LO) for locking; feedback is applied by modulating pump power. The remain output light is then sent into our tapered PCF to perform spectral broadening. The resulting light is monitored on a CCD spectrometer (Thorlabs CCS100) before travelling into the final part of the astro-comb, i.e. the 16 GHz Fabry-Perot mode filters. The cavity lengths are locked using a Pound-Drever-Hall scheme with a frequency-doubled Nd:YAG laser, and the control loop is implemented using a field-programmable gate array (FPGA) system. Coarse tuning of the repetition rate can be achieved by changing the laser baseplate temperature with the chiller (Solid State Solutions TCube edge) temperature setpoint.

All the relevant instruments are connected (typically via USB or RS-232 serial interface) to a PC running LabVIEW 2017 (National Instruments). We designed a virtual instrument (VI) to let the user monitor laser status information, tweak various laser parameters, lock f_r, f_0, and optimize the output spectrum from the PCF. The user interface is shown in Figure 1.16. The implementation uses a queued state machine with a producer-consumer architecture, where user events are added to a first-in-first-out queue and passed to consumer loops, which correspond directly to various instruments.

Separate VIs control the Arduino and filter cavities, and the user interfaces for these are shown in Figure 1.17 and 1.18, respectively.

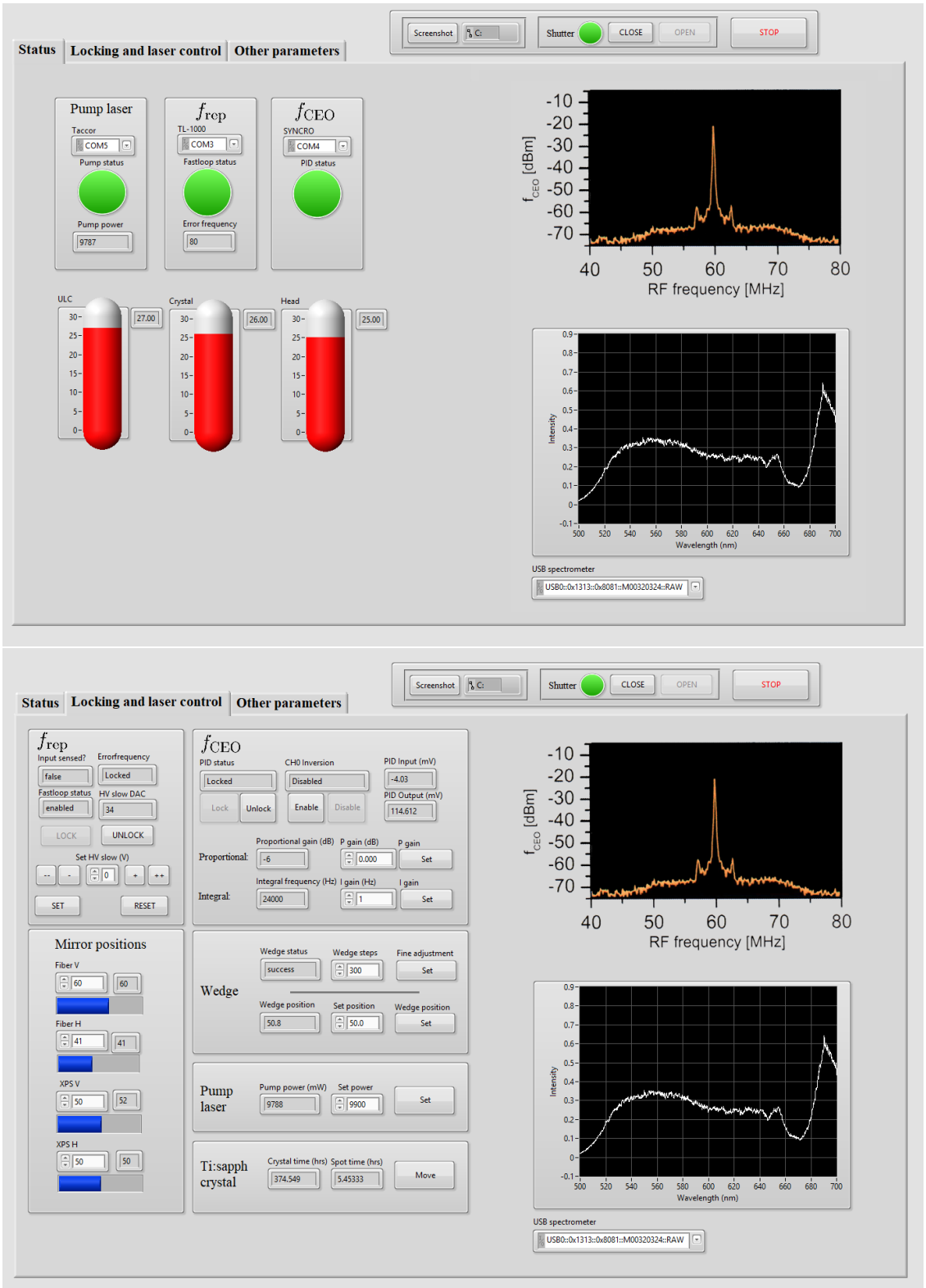

Figure 1.16: LabVIEW user interface for a remote-controlled comb. (Top) The first tab shows laser status, f_0 beat note and PCF spectrum. (Bottom) The second tab contains all the laser controls: lock loop settings for f_r, f_0, wedge picomotor control, pump laser power, crystal spot translation, and piezo mirror alignment for PCF and f-$2f$ interferometer. (Not shown) A third tab has auxiliary information. (Figure courtesy of W. Dworschack)

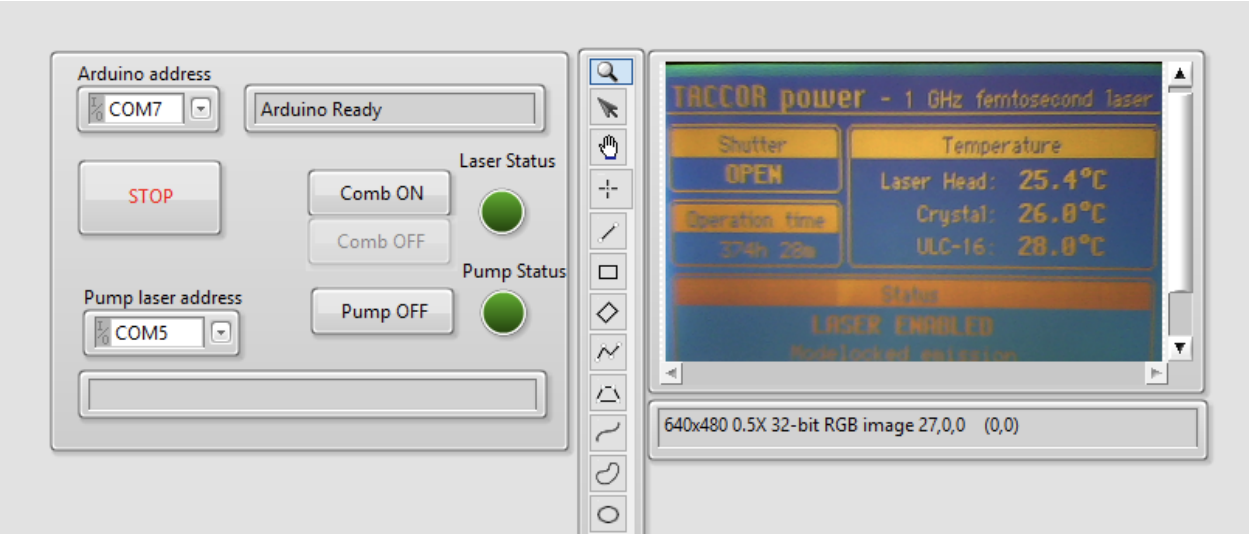

Figure 1.17: LabVIEW user interface to remotely turn on and off the laser. This VI has built-in safety features, e.g., it does not turn off the laser without first turning off the pump laser. It also features pop-up dialogs to confirm or cancel user actions. On the right, a webcam is shown to monitor the physical state of the front panel of the laser controller. (Figure courtesy of W. Dworschack)

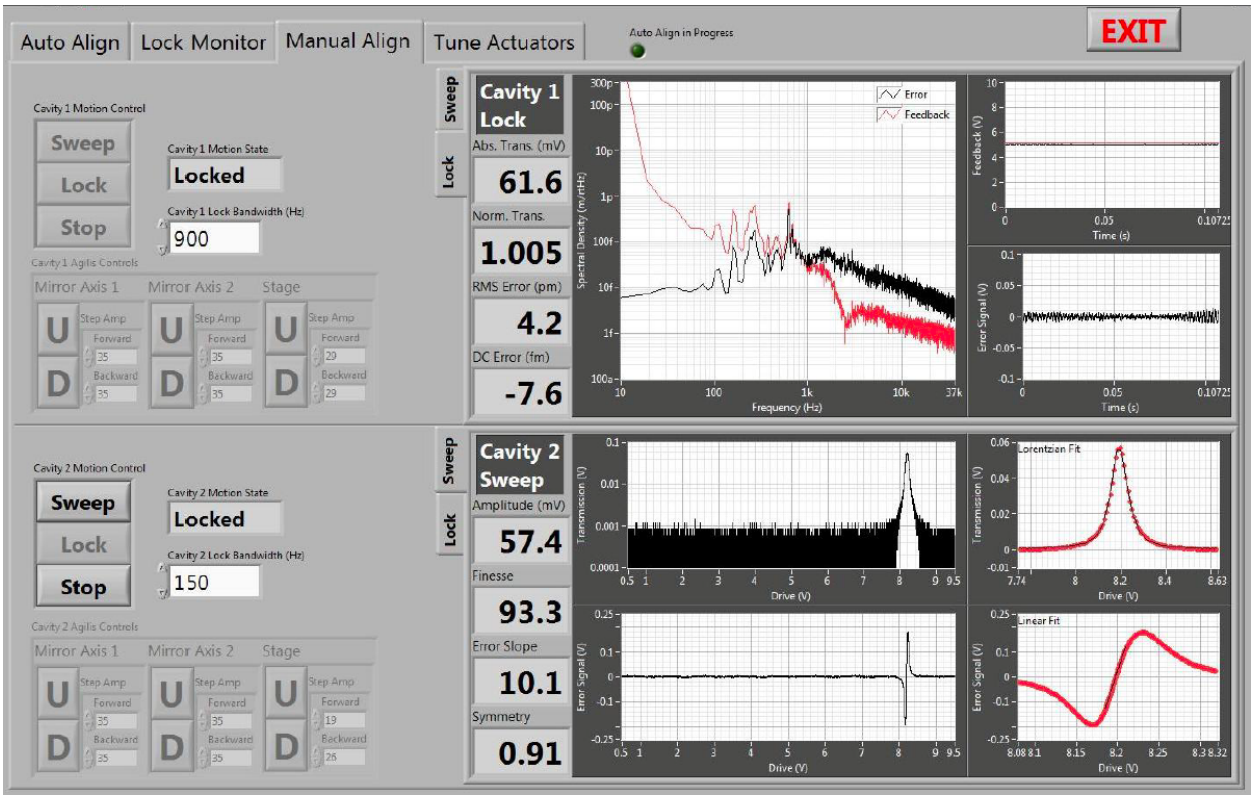

Figure 1.18: LabVIEW user interface to control filter cavities. (Figure from Ref. [53])

The next steps include automated locking of f_r, f_0 and optimization of PCF output spectrum. The automated locking of f_r should be relatively straightforward as it amounts to a simple 1D search in the slow piezo high voltage bias with the target set such that error frequency falls below the ~ 200 Hz capture range of the PID (at which point we can engage the TL-1000's lock). Next, we tackle f_0, which is somewhat more challenging. Assuming the f_0 beat SNR is sufficient ($\gtrsim 40$ dB), we would perform a 1D search with the wedge with the target set such that error frequency falls below the capture range of the SYNCRO PID. This of course assumes that the beat is visible at all within the ~ 20 MHz bandwidth of the loop. If the f_0 beat SNR is insufficient, we would perform two independent 1D searches with the XPS mirror in the H and V directions to optimize the SNR. PCF spectrum optimization is also done similarly, but constructing a cost function is more complicated than simply using beat SNR. Essentially, one would need to devise a metric which evaluates the similarity between the current spectrum and the desired spectrum (e.g., a broad, flat template). This is a highly simplified picture, and in reality, things will be much more challenging. Among other things, various degrees of freedom (e.g. f_0, f_r) are coupled, so the order of operations will certainly be important.

1.8 Conclusions and outlook

In summary, we have demonstrated a robust astro-comb design based on a turn-key Ti:sapphire femtosecond laser, a dispersion-engineered tapered photonic crystal fiber and Fabry-Perot cavities constructed from zero GDD mirror pairs. It is our hope that (1) remote control and (2) eventual automation of the system leads to the widespread adoption of astro-combs at facilities operating precision RV spectrographs.

One possible avenue to the further simplification of the design of astro-combs is to move to higher repetition rate source combs. To motivate this, consider a simple Lorentzian model for transmission T of n sequential filter cavities, where

$$T(\delta; \mathcal{F}, \mathrm{FSR}, n) = \left(\frac{4\mathcal{F}^2}{\mathrm{FSR}^2} \delta^2 + 1 \right)^{-n}. \tag{1.9}$$

$\mathcal{F}$ is the finesse and δ is the detuning from resonance. The FSR (which stands for free spectral range) is the cavity mode spacing. Our current scheme provides $T\left(\pm 1 \text{ GHz}; 100, 16 \text{ GHz}, 2\right) \approx -44$ dB of suppression of the nearest side mode. Note that if the teeth of the source comb are spaced apart by 10 GHz, instead of 1 GHz, and we use 20 GHz FSR filter cavities, we automatically get the roughly same suppression with a single cavity, i.e., $T\left(\pm 10 \text{ GHz}; 100, 20 \text{ GHz}, 1\right) \approx -40$ dB. Moreover, it may even be possible to completely eliminate the filter cavities altogether! Future spectrographs such as EXPRES [54] are moving to higher resolution ($R = 150,000$). With all other parameters kept the same as HARPS-N (spectral sampling, spectral coverage), moving to $R = 150,000$ gives an optimal calibration spacing of $2.5 \times \nu/R = 10$ GHz (for optical frequencies of ~ 600 THz). In this case, there is no need for filtering − this is highly advantageous as it automatically allows one to use the full band produced by supercontinuum generation.

The major challenge with supercontinuum generation with high repetition rate sources (e.g., at 10 GHz instead of 1 GHz) is the low peak pulse power. The peak pulse power $P_{\text{peak}} \sim E_p/T_0 = P_{\text{avg}}/\left(f_r T_0\right)$, where E_p is the pulse energy, T_0 is the pulse FWHM, P_{avg} is the average power of the laser, and f_r is the repetition rate. Given a fixed amount of average power, the peak pulse power of a 10 GHz laser is 10× lower than that of a 1 GHz laser. Since the frequency conversion stage of the astro-comb is inherently nonlinear, this means accomplishing supercontinuum generation at effectively 10× lower power.

One solution to this problem is to take advantage of advances in new photonic materials for supercontinuum generation. Silicon nitride is an excellent platform for this application because the nonlinear refractive index $n_2 = 24 \times 10^{-20}$ m^2/W *is* 10× larger than that of silica [55]. Since the index change is given by $n_2 I$, where I is the intensity, this solves our problem of having 10× lower intensity. Silicon nitride also has negligible Raman gain and $\chi^{(2)}$ effects (which makes it easier to simulate). Some other materials that are also good candidates are lithium niobate (see [56] and references therein) and diamond [57, 58].

Using the same principles as the dispersion-engineered PCF (i.e., width modulation), we can design a tapered planar waveguide to produce the desired visible-spanning supercontinuum. It is important

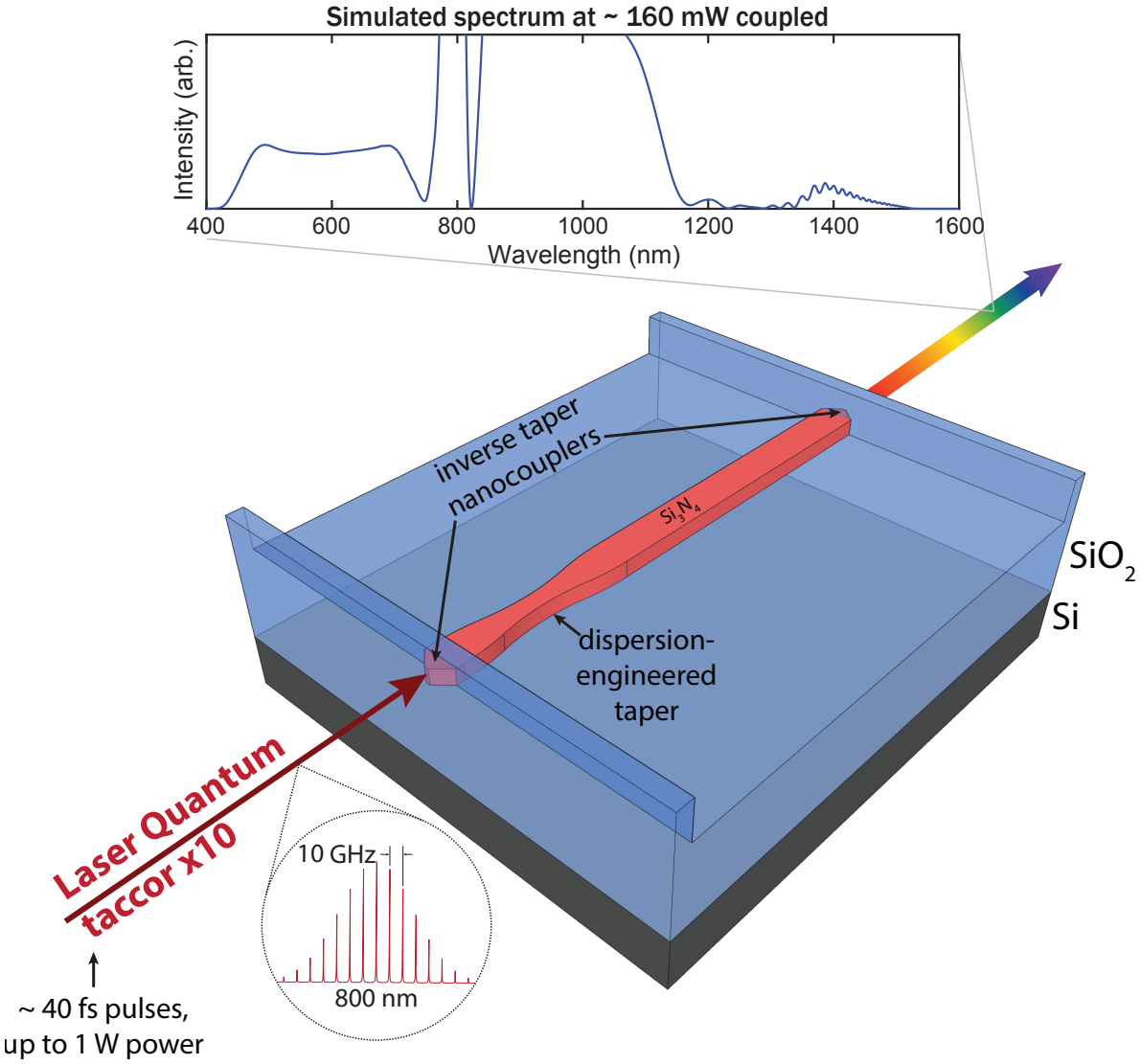

Figure 1.19: (Not to scale) Design of a planar waveguide for supercontinuum generation from a 10 GHz source comb. The substrate is a 0.5 mm thick silicon wafer. On top of the substrate, a 2 μm thermal oxide (SiO_2) layer is deposited, and this is followed by a 400 nm-thick silicon nitride layer which is deposited using low-pressure chemical vapour deposition (SVM Microelectronics). The shape of the waveguide is lithographically patterned, with the following parameters: $d_0 = 1.30$ μm, $L_0 = 0$ μm, $d_w = 0.60$ μm, $L_{t_1} = 1.24$ mm, $L_{t_2} = 1.00$ mm, $d_1 = 1.35$ μm (on the output side the taper) and $L_1 = 2.76$ mm, where we use the same conventions as in Figure 1.9(a). Inverse taper (of typical length ~ 100 μm and final width ~ 100 nm) spot size converters [59, 60] are added to both ends to facilitate coupling. After defining the waveguides, a top cladding of a few μm thickness is deposited and patterned so that it covers only the nanotaper coupling regions. The final step is mechanical polishing or etching [61] of the facets to minimize scattering from surface roughness. (Upper plot) After mode solving using WGMODES [62], we run our GNLSE simulations with a 16 pJ, 40 fs Gaussian pulse at 800 nm from a 10 GHz Ti:sapphire laser coupled into the waveguide. This results in a nearly visible-spanning supercontinuum.

to make use of waveguide geometry (width×height) and surroundings (silica vs. air cladding) to produce anomalous dispersion at the 800 nm pump wavelength [63]. This is shown in Figure 1.19. The experimental realization and preliminary results are shown in Figure 1.20.

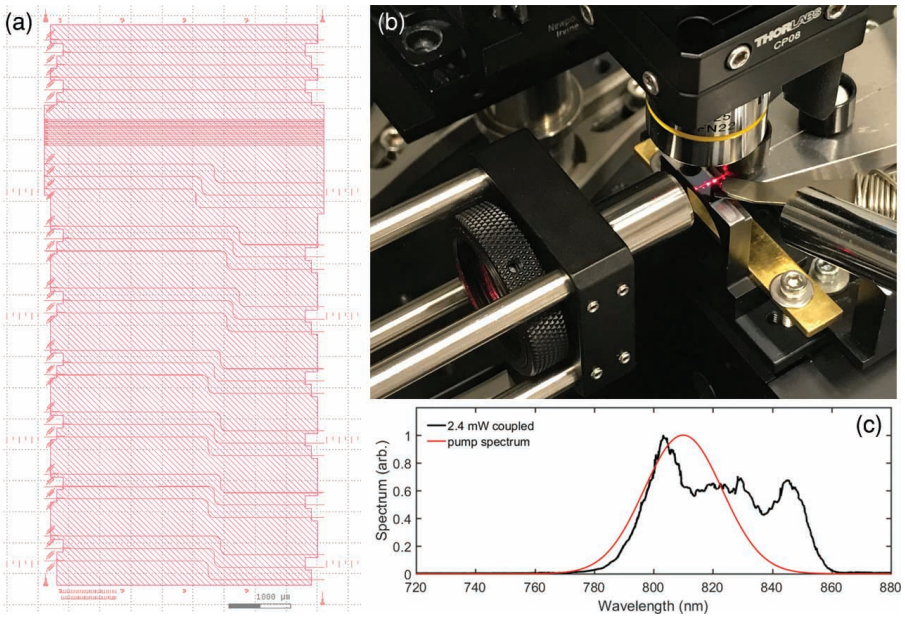

Figure 1.20: (a) Layout artwork for silicon nitride chip (hatched fill indicates oxide top cladding). Multiple devices variants and additional test structures are pictured. The bends in waveguides are to prevent spurious reflections and enable easy verification of coupling. (b) Test setup for silicon nitride chip. Light from a 10-GHz laser (taccor x10) is coupled in using an 8 mm aspheric lens (Thorlabs C240TME-B) and the output is collected using a lensed fiber (OZ Optics BC#62002 TSMJ-X-680-3.5/125-0.25-10-2-9-1-S630-HP). (c) Preliminary evidence of spectral broadening with a 1-GHz laser coupled into the chip.

More work is needed to improve the coupling efficiency, but initial results appear promising. Currently, there are issues with back-reflections from the chip which cause laser instabilities (CW breakthrough or lasing in the wrong direction), but this can be suppressed with some modifications to the setup (e.g., by angling the chip).

2

Measuring the local dark matter density using stellar accelerations

2.1 Introduction

Having developed astro-combs for RV exoplanet science, it is natural to ask if the superb precision they enable could be useful for other applications. One possibility is to measure the acceleration of stars in the galactic gravitational potential to better understand the distribution of dark matter in our galaxy. Given that stars in our vicinity take approximately 250 million years to execute an orbit about the galactic centre, one might expect that this effect is not appreciable on human timescales. But a simple back-of-the-envelope calculation shows

$$\frac{1 \text{ cm/s}}{3 \text{ yr}} \sim 10^{-8} \text{ cm/s}^2,$$

where 1 cm/s is the precision comb-calibrated spectrographs can have, and 3 years is a typical planetary orbital period. Given the rotational velocity of the Sun about the galactic centre $v_{\text{circ}}(r_0) \equiv v_0 \approx 220$ km/s [64], and our galactocentric distance $r_0 \approx 8$ kpc [64], the local centripetal acceleration is $a_r(r_0) \equiv a_0 \approx 2 \times 10^{-8}$ cm/s^2. This is of the same order of magnitude as our estimated acceleration sensitivity.

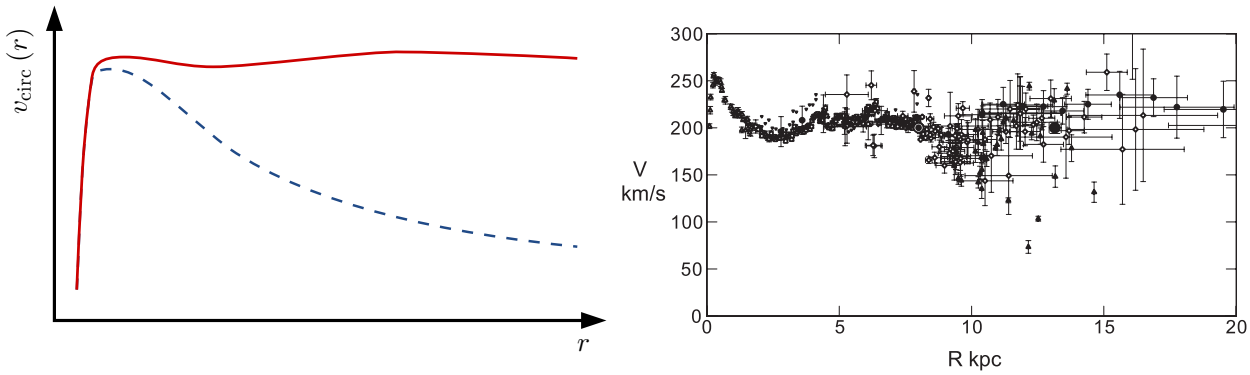

Figure 2.1: (Left, adapted from a version by P. Hibbs, CC-BY-SA-3.0) An example of a galactic rotation curve (red). The Newtonian prediction is shown in dotted blue. (Right) Measured rotation curve of the Milky Way (from Ref. [68]).

Early evidence for dark matter comes from galactic rotation curves of disk galaxies [65–67]. These measurements measure the rotational velocity of stars, v_{circ}, about the galactic centre versus their distance from the galactic centre, r. If a galaxy behaved as a solid disk, we would expect the rotational velocity to increase linearly with distance, i.e. $\sim r$. This holds well for stars near the centre of the galaxy. If on the other hand, the most of the mass was concentrated at the centre, the velocity should fall off as $\sim 1/\sqrt{r}$ according to Newtonian mechanics. For most disk galaxies, this is not observed, and the velocity levels off and remains constant beyond some distance from the centre, as shown in Figure 2.1. One possible explanation for the discrepancy is the existence of additional matter that has gravitational interactions but is not visible, i.e. *dark matter* (DM).

Understanding the nature of DM [69] is one of the most pressing issues in modern physics. Many particle DM models, such as those employing weakly-interacting massive particles and axions, predict unique laboratory and astrophysical signatures, which are being searched for at a variety of experiments and observatories [70]. However, knowledge of the local DM density is crucial for interpreting the results of these efforts. Unfortunately, current methods for determining the local properties of DM (i.e., within our region of the galaxy) are indirect and subject to large systematic uncertainties [71]. In addition to aiding searches for particle DM, better certainty of the local DM distribution may provide key insights into the history of the Milky Way (MW). In this work, we propose a new approach – direct measurements of stellar accelerations – to determine the local DM density and morphological parameters in the MW. Our technique circumvents many of the

systematic issues faced by existing methods.

Currently, the DM density in the MW is inferred from either the galactic rotation curve, measured via Doppler shifts, or the dispersion of local stellar velocities in the vertical direction about the galactic mid-plane [71–75], measured using astrometry. However, implicit in both these analyses is the assumption of equilibrium, *i.e.*, that dynamics have reached steady-state. In particular, the velocity distribution does not directly probe the gravitational potential and thus the DM distribution: only the equilibrium velocity distribution is determined by the potential. Given the presence of density waves in the MW, for example those causing the local North-South asymmetry recently studied in [76] using *Gaia*, as well as other out-of-equilibrium processes, such as the continuing interactions of the galaxy with massive satellites, the equilibrium assumption used in traditional determinations of the local DM density is open to question. Stellar accelerations, on the other hand, are directly determined by the forces acting on a star, and thus the gravitational potential, with no modelling assumptions. Having such a direct probe would allow, in principle, for an unbiased mapping of the gravitational potential of the MW. This approach opens up, for example, the possibility of searching for low-mass DM subhalos, which are predicted in the standard cosmological framework but are absent in certain DM models such as warm DM and fuzzy DM [77].

Stellar accelerations may also be used to probe both the spatial morphology of the density profile of the bulk DM halo and the galactic disk, in addition to possible DM subhalos (see [78–83] for related but more indirect proposals). The radial density profile in particular plays a key role in interpreting searches for DM annihilation [84], while the halo shape is influenced by baryonic feedback and potentially DM self-interactions [85].

In this chapter, we show how precision radial velocity (RV) measurements of the accelerations of individual stars in our galactic neighbourhood can be used to measure the local DM density and constrain the spatial morphology of the DM density distribution. To our knowledge, a similar idea has only been suggested once before, in the context of testing modified Newtonian gravity with MW globular clusters [86, 87]. We highlight the work of Silverwood and Easther [88], contemporaneous to our own, in which they also propose using stellar accelerations to map the galactic gravitational

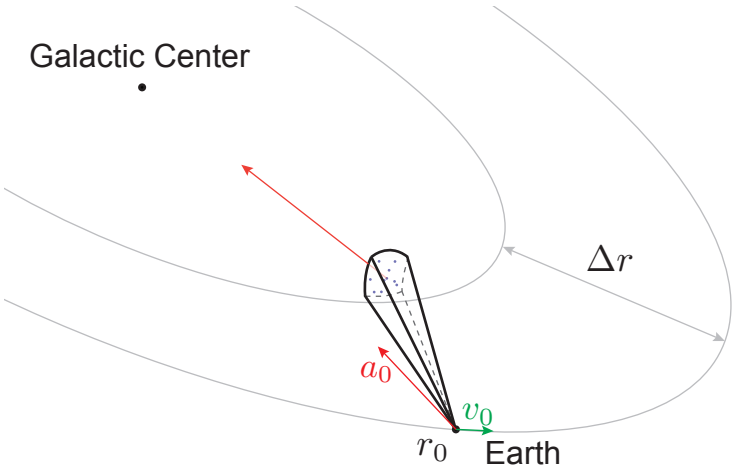

Figure 2.2: Geometry for observing stellar accelerations in the Milky Way. The solar system is at a distance r_0 from the galactic centre (origin), has a rotational velocity v_0 and feels an acceleration a_0 due to the Milky Way gravitational potential. Stars further inward feel a stronger acceleration. From Earth, we can observe the radial velocity of stars Δr away. By measuring small changes in these velocities over time, we directly determine stellar accelerations and hence the Milky Way gravitational potential. The diagram above is not to scale and angles are exaggerated for effect.

field, and provide a complementary analysis.

We propose to use the precision RV method, honed in the search for extrasolar planets [8], to measure directly the change over time of the velocities of an ensemble of individual stars – and thus the forces acting upon those stars. Since the Sun is also accelerating in the gravitational potential of the MW, measurements must be performed on stars sufficiently distant from the Sun – either closer or further from the galactic centre (GC) – for a difference in acceleration to be observed. This situation is depicted in Figure 2.2.

2.2 Theoretical framework

To obtain the galactic potential from stellar accelerations, we start with the gravitational Poisson equation

$$\nabla \cdot \mathbf{a} = -4\pi G\rho, \tag{2.1}$$

where $\mathbf{a}$ is the acceleration field, G is the gravitational constant. As depicted in Figure 2.2, our proposed observing region will be intentionally slightly misaligned from the GC (vertically and horizontally) to avoid extinction in the galactic midplane and overcrowding of stars. To a good approximation, pointing in this manner primarily gives us a measurement of Δa_r, where a_r is the centripetal acceleration. Dividing by the separation between the target volume and Earth, Δr, gives $\partial a_r/\partial r$. We assume azimuthal symmetry and discuss errors from contamination by vertical gradients later in this section.

The energy density is composed of two parts, a disk (baryonic) component and a (likely spherical) dark matter component, i.e.,

$$\rho = \rho_{\text{disk}} + \rho_{\text{DM}}. \tag{2.2}$$

Substituting Eq. 2.2 back into 2.1 gives us

$$\nabla \cdot \mathbf{a} = -4\pi G \left(\rho_{\text{disk}} + \rho_{\text{DM}}\right). \tag{2.3}$$

Writing the gradient out in spherical coordinates, we get

$$\nabla \cdot \mathbf{a} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 a_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta a_\theta\right) + \frac{1}{r\sin\theta}\frac{\partial a_\varphi}{\partial \varphi}, \tag{2.4}$$

where θ is the angle from the z-axis (the galactic disk is in the xy-plane), and φ is the azimuthal angle. Since we are assuming azimuthal symmetry, we can set $a_\varphi = 0$. Confining ourselves to the vicinity of the disk (i.e. $z \approx 0$), we make a change of coordinates to cylindrical coordinates. Recall that $z = r\cos\theta$, so $\partial z = -r\sin\theta\partial\theta$. Additionally, $\sin\theta \approx 1$ and $a_\theta \approx -a_z$ near the disk. Putting all of this together yields

$$\nabla \cdot \mathbf{a} = \underbrace{\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 a_r\right)}_{\text{primarily } \rho_{\text{DM}}} + \underbrace{\frac{\partial a_z}{\partial z}}_{\text{primarily } \rho_{\text{disk}}} \tag{2.5}$$

Vertical accelerations near the disk are dominated by the disk (which is compact in the z-direction) and radial accelerations are dominated by the dark matter. Under this assumption, we can focus

on the first term,

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 a_r\right) = \frac{2a_r}{r} + \frac{\partial a_r}{\partial r} \tag{2.6}$$

$$= -2\frac{v_{\rm circ}^2}{r^2} + \frac{\partial a_r}{\partial r}, \tag{2.7}$$

using the definition of the centripetal acceleration $a_r = -v_{\rm circ}^2/r$. Rearranging, we get

$$\rho_{\rm DM} \approx \frac{1}{4\pi G}\left(2\frac{v_{\rm circ}^2}{r^2} - \frac{\partial a_r}{\partial r}\right). \tag{2.8}$$

Evaluating at $r = r_0$ gives us the local dark matter density

$$\rho_{\rm DM}\left(r_0\right) \approx \frac{1}{4\pi G}\left(2(A-B)^2 - \left.\frac{\partial a_r}{\partial r}\right|_{r=r_0}\right), \tag{2.9}$$

where $A - B = v_0/r_0$, and $A = 15.3 \pm 0.4$ km s^{-1} kpc^{-1} , $B = -11.9 \pm 0.4$ km s^{-1} kpc^{-1} are the Oort constants [89]. Even though the disk energy density, $\rho_{\rm disk}$, can be $\sim 10\times$ larger than $\rho_{\rm DM}$ [71], we find that completely neglecting the disk is still a good approximation, as discussed below.

In the absence of other matter, for a spherical DM density profile $\rho_{\rm DM}\left(r\right)$, we can write

$$a_r(r) = -GM_{\rm DM}(r)/r^2, \tag{2.10}$$

where $M_{\rm DM}(r) = 4\pi\int_0^r R^2\rho_{\rm DM}\left(R\right)dR$ is the DM mass enclosed within the radius r. Substituting this into right-hand side of Eq. (2.6) and simplifying, we obtain

$$\rho_{\rm DM} = \frac{M'_{\rm DM}\left(r\right)}{4\pi r^2}, \tag{2.11}$$

where the $'$ denotes a radial derivative. Given the definition of $M_{\rm DM}$, this is self-consistent.

If instead, we did the same calculation for a disk density profile $\rho_{\rm disk}$, we can still write

$$a_r\left(r\right) = -GM_{\rm disk}(r)/r^2, \tag{2.12}$$

where $M_{\rm disk}\left(r\right) = 2\pi\int_0^r R\Sigma_{\rm disk}dR$ is the disk mass enclosed within the radius r. $\Sigma_{\rm disk}$ is the disk surface density (the disk is assumed to be thin and uniform). Repeating the calculation, we see that

$$\rho_{\rm disk} = \frac{M'_{\rm disk}\left(r\right)}{4\pi r^2} = \frac{\Sigma_{\rm disk}}{2r} \tag{2.13}$$

This leads to a fictitious contribution to $\rho_{\rm DM}\left(r_0\right)$ in Eq. (2.9), which we refer to as $\rho_{\rm DM}^{\rm fict.}(r_0)$, given by $\rho_{\rm DM}^{\rm fict.}\left(r_0\right) = \Sigma_{\rm disk}/(2r_0)$. According to measurements, $\Sigma_{\rm disk} \approx 50\ M_\odot/{\rm pc}^2$ [71, 75], where $M_\odot$ is the solar mass. Plugging this and $r_0 = 8$ kpc into Eq. (2.13) results in $\rho_{\rm DM}^{\rm fict.}\left(r_0\right) \approx 0.003\ M_\odot/{\rm pc}^3$. $\rho_{\rm DM}\left(r_0\right) \approx 0.01\ M_\odot/{\rm pc}^3$ [71] according to recent estimates, so we find that $\rho_{\rm DM}^{\rm fict.}\left(r_0\right)/\rho_{\rm DM}\left(r_0\right) \approx 30\%$. It should be noted however, that modelling the disk contribution can enable subtracting it out at high precision.

Propagating fractional uncertainty in the radial acceleration gradient $\frac{\delta a'_r(r_0)}{a'_r(r_0)}$ gives the fractional uncertainty in the local DM density

$$\frac{\delta\rho_{\rm DM}\left(r_0\right)}{\rho_{\rm DM}\left(r_0\right)} \approx \frac{\delta a'_r\left(r_0\right)}{a'_r\left(r_0\right)}\frac{\left(A-B\right)^2}{2\pi G\rho_{\rm DM}\left(r_0\right)} \approx 2.7\frac{\delta a'_r\left(r_0\right)}{a'_r\left(r_0\right)}\,. \tag{2.14}$$

2.3 Observational considerations

If we select stars 3 kpc away (beyond the star field observed by the *Kepler* spacecraft [90] but potentially observable with the Giant Magellan Telescope [91]), the fractional change in acceleration compared to the local acceleration should be approximately 0.75, i.e., $\Delta a_r = 1.5\times10^{-8}$ cm/s^2. In more convenient units, this is 0.5 cm/s/year. Over 10 years, one would then expect a typical stellar velocity change due to the MW gravitational potential of approximately 5 cm/s, which is similar to the RV amplitude associated with an Earth-like exoplanet in the habitable zone around a Sun-like

star [3]. We note that next-generation instruments designed for RV studies of exoplanets, including ESPRESSO [92] and G-CLEF [91], are expecting to achieve 10 cm/s or better RV precision and long-term stability in order to pursue such exoplanet astronomy.

At a target signal-to-noise ratio of 100, the faintness of stars sets the exposure times to approximately 12 minutes per star with a 30-m telescope at $\Delta r = 1$ kpc or with a future 100-m telescope at $\Delta r = 3$ kpc. Including the length of nights as well as star visibility leads to $\sim 10^4/N$ observations per star per year with a single telescope for a N-star survey where each night is time-shared between various targets (see §2.4.1).

Necessary for measuring small stellar accelerations is extremely stable calibration of the spectrograph used to determine the RVs over several years. The ideal tool for this task is a laser frequency comb optimized for calibrating spectrographs. These specialized instruments, known as "astro-combs" [9–11, 22, 23, 31], may be referenced to GPS-disciplined atomic clocks. Thus, spectrograph wavelength solutions are easily trustworthy over a decade; and even measurements from multiple comb-calibrated observatories can be combined into a single data set if the same reference clock is used for the astro-combs at all observatories.

An additional effect which needs to be considered is the contribution to radial velocities from the motion of stars in the plane of the sky, the so-called perspective acceleration. This effect can be removed using high-precision astrometric surveys such as *Gaia* [93]. As pointed out by Silverwood and Easther [88], this effect can be subtracted out at the $\sim 1\%$ level by choosing stars below a transverse velocity threshold of ~ 55 km/s, and such stars are quite abundant in the *Gaia* catalogue.

When designing the stellar acceleration survey, one consideration is the volume of space needs that to be surveyed to observe about 10^3 target Sun-like stars. The local stellar density in the solar neighbourhood[1] is 0.1 pc^{-3}, and we conservatively assume the same density at $\sim$ few kpc inward toward the galactic centre (GC) relative to the Sun. Further, if we assume only approximately 5% of the stars are of spectral type G, we would need an observation box of side length ~ 60 pc $= 200$ ly.

[1] RECONS Census Of Objects Nearer Than 10 Parsecs

Another important consideration is star visibility. Using Sagittarius A* as an indicator of the GC and a sky model[2] for the Paranal Observatory in Chile (24°37´38″S 70°24´15″W), we find that the GC is visible about 8 out of 12 months a year.

Finally, we also consider exposure times for targets. First we compute the apparent magnitude m of a Sun-like star at distance Δr away from its absolute magnitude $M = 4.83$. The apparent magnitude at various distances are: 14.8 (1 kpc away) and 17.2 (3 kpc away). We feed these magnitudes into exposure time calculators for current precision RV spectrographs at 3.6 m telescopes[3] and 10 m telescopes[4], and appropriately scale by telescope diameters, i.e., $t_{\rm exp}/\left(D_{\text{next-gen}}/D_{\text{current}}\right)^2$. With a target signal-to-noise ratio of 100, we get $t_{\rm exp} = 12$ min for a target $\Delta r = 1$ kpc away with a 30-meter telescope, or alternatively for a target $\Delta r = 3$ kpc away with a 100-meter telescope (e.g., Overwhelmingly Large Telescope [94]). Given an average night of 8 hours, and the visibility issue discussed above (duty cycle 2/3), we calculate that this exposure time would limit the number of observations to $10^4/N$ per star per year for a N-star survey if a single telescope was the only resource for the project. It should be noted that the stellar acceleration change Δa_r increases linearly with distance from the Sun Δr (as long as the potential remains effectively linearized), but the total telescope exposure time grows quadratically with distance. So one wins by doing very precise measurements at shorter range until more advanced telescope technology becomes available. One can also go to intermediate distances and measure for longer than a decade.

2.4 Simulation scheme

To determine whether sufficiently sensitive stellar acceleration measurements are possible, given other sources of Doppler shift "systematics" (e.g., stellar companions, planets, stellar noise), we simulate a measurement campaign with synthetic RV time series from a population of stars including all the above effects and a realistic observing schedule. We then analyze the time series to try to recover an injected acceleration signal of order few cm/s over a decade.

[2] SkyCalc Sky Model Calculator
[3] HARPS Exposure Time Calculator
[4] HIRES Exposure Time Calculator

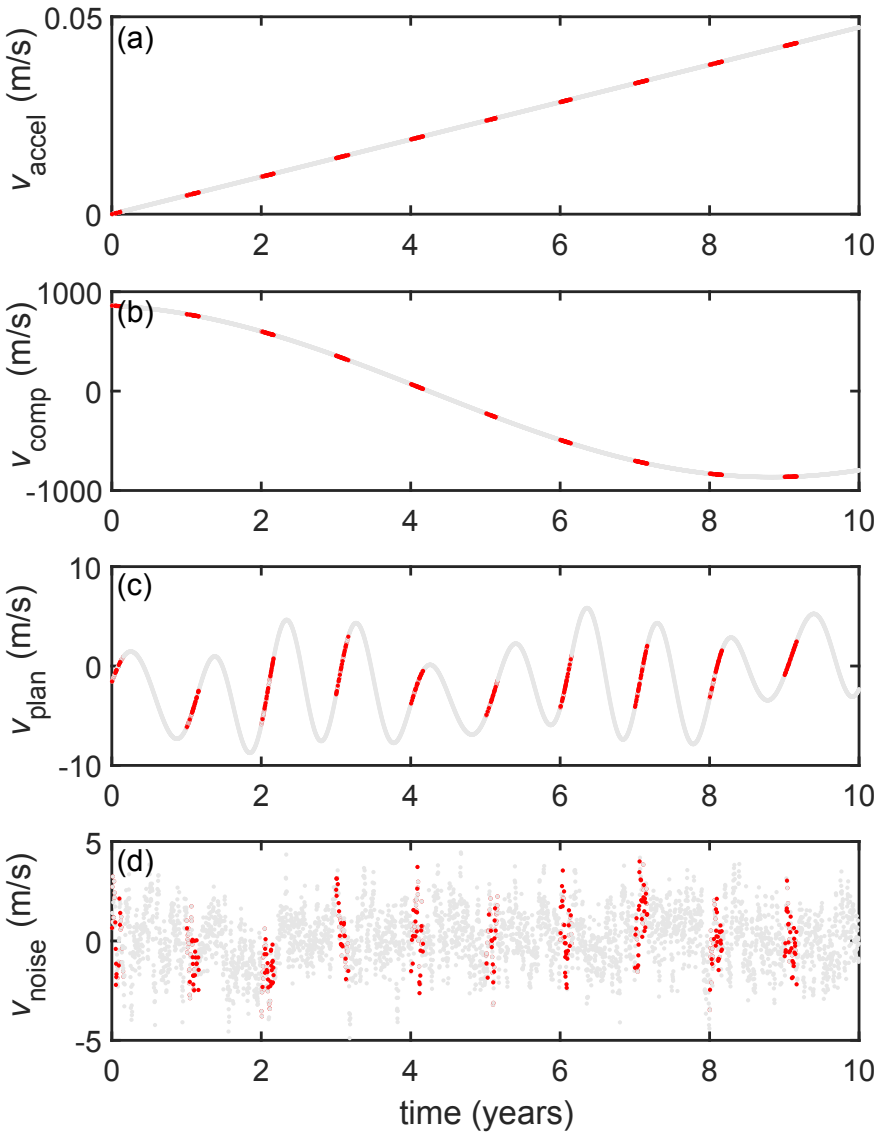

Figure 2.3: Example of a synthesized RV time series for a single primary star, showing the four contributing mechanisms considered in our analysis. The effects of (a) the acceleration due to the Milky Way gravitational potential, (b) stellar companions, (c) planets (multiple planets in this case), and (d) "noise" including stellar magnetic activity and instrumental effects are depicted. Red dots are the observed nights and grey dots represent all nights.

An observing schedule is generated for a total measurement campaign of 10 years (see §2.4.1). The observing schedule is applied to the generation of N time series (10^3 or larger in this work), representing N candidate stars in an initial sample. Though the telescope time associated with the observing schedule used in this work exceeds the limits posed by the exposure times calculated in the previous section for a single telescope, we use it to aid understanding confounding effects in the search for the acceleration signal. It is important to note that fewer than N stars will actually be followed in a real campaign because many stars will be poor targets for detecting an acceleration signal (due to reasons discussed below). The full campaign would then consist of a target selection phase (where one prunes the list of candidate targets) lasting a few years followed by a ∼decade observations devoted exclusively to measuring the acceleration signal.

Each star in the simulation is assigned a number of stellar companions and planets. The multiplicity and orbital parameters of these orbiting bodies are determined using known statistical distributions (see §2.4.1). The total RV for a given primary star is:

$$v_{\rm total}\left(t\right) = v_{\rm accel}\left(t\right) + v_{\rm comp}\left(t\right) + v_{\rm plan}\left(t\right) + v_{\rm noise}\left(t\right), \tag{2.15}$$

where $v_{\rm accel}\left(t\right) = \Delta a_r t$ according to the acceleration change Δa_r at target distance Δr away, due to the MW gravitational potential. The three terms following $v_{\rm accel}$, arise from stellar companions, planets, and noise (see §2.4.1). Since the survey volume is relatively small and $\Delta r = 3$ kpc, we take $\Delta a_r = 1.5 \times 10^{-8}$ cm/s^2 (~ 5 cm/s/decade) for all stars. Figure 2.3 shows a typical example of each of the components plotted separately as time series; note the scale of the various effects. In this particular example, $v_{\rm total}$ would be dominated by $v_{\rm comp}$. The ideal case (occurring about 10% of the time) is a star with no stellar companions or planets. This leaves us with only the MW acceleration plus noise.

2.4.1 Generation of synthetic RV time series

The observing schedule $\{t_i\}$ is generated for a total measurement campaign of 10 years, with observing blocks lasting 60 days (to average over stellar rotations), occurring at 1-year intervals. Within

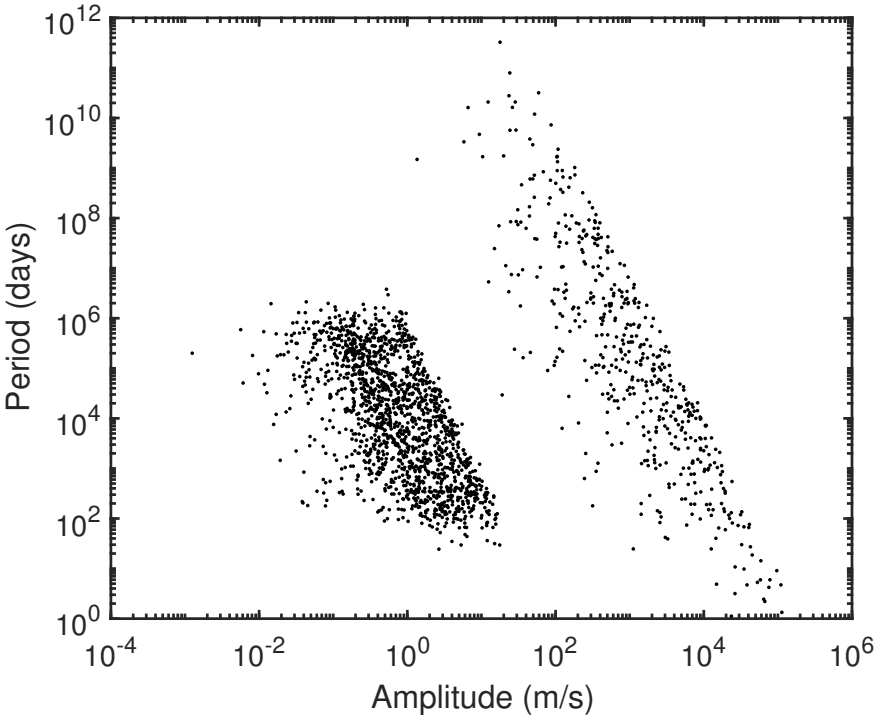

Figure 2.4: Distribution of periods and semi-amplitudes (weighted by inclination factor $\sin i$) in our synthesized dataset of 10^3 primary stars. The group on the left is the set of planets while the group on the right is the set of stellar companions.

the observing blocks, we assume that, for a given target, single exposures are taken every night, and that 1/3 of the nights (generated randomly) are discarded due to the presence of clouds.

We restrict ourselves to solar-type FGK stars (mass range $0.7 - 1.3 M_\odot$) so that we may apply well-known models developed for such stars. Taking the simplest possible assumption for the mass distribution, we assume the primary star masses M_p to be uniformly distributed within this range. FGK stars are second in abundance only to M stars and these four groups almost comprise all stars in the galaxy [95]. Unfortunately M stars are typically too faint to be observed by spectrographs operating in the visible region of the electromagnetic spectrum. Planet occurrences are roughly similar for F, G and K spectral types according to the *Kepler* mission statistics [96], but radial velocity noise generally decreases from F through K [97]. However, the luminosity also decreases from F through K, so there may be a sweet spot for the best stars to choose (likely G stars).

Each stellar system is given a random inclination angle i uniformly distributed between 0 and π, as

we are not aware of any anisotropy in this distribution. All bodies orbiting the star are assumed to lie in a single plane inclined at this angle. The inclinations are used to project the RVs along the line of sight. Orbital eccentricities are neglected. Ignoring eccentricities leads to simpler fitting of the time series as it suffices to use sinusoidal functions instead of transcendental equations [3]. Furthermore, frequency domain analyses of the time series data are simplified by the absence of harmonics. These simplifications may lead to a slightly more favourable scenario for detecting stellar accelerations, but a full treatment of planetary dynamics is beyond the scope of our preliminary assessment, where we simply hope to capture the salient physics of orbiting bodies.

First, we will consider the largest Doppler shift contribution: stellar companions. According to stellar multiplicity statistics for solar-type stars [98], 56% of the systems are single, 33% are binary, 8% are triple, and 3% are quadruple and higher. In practice, we truncate the distribution at quadruples for simplicity, i.e., 3% of the stars are given three companions. If a star has a companion, we generate a mass M_s and period T_s for this secondary object. The mass may be generated from a mass ratio $q = M_s/M_p, 0 \leq q \leq 1$. The mass ratio follows a power-law distribution [99], $\sim q^{0.3}$. Similarly, the period of stellar companions follows a lognormal distribution [98] with mean $\mu_{\log_{10}(T/\text{days})} = 5.03$ and standard deviation $\sigma_{\log_{10}(T/\text{days})} = 2.28$. Using Kepler's third law, we can compute the semimajor axis of the secondary as $a_s = \sqrt[3]{GM_pT_s^2/(4\pi^2)}$, where M_p is the mass of the primary star. Once we have these orbital parameters, we can sum the radial velocity (RV) contributions of stellar companions for each primary star as

$$v_{\text{comp}}(t) = \sum_k \frac{2\pi}{T_k} q_k a_k \sin(2\pi t/T_k + \phi_k) \sin i, \tag{2.16}$$

where T_k is the period of the k^{th} stellar companion in the stellar system. Similarly, q_k, a_k and ϕ_k are, respectively, the mass ratio, the semimajor axis and orbital phase offset (uniformly distributed between 0 and 2π) for that companion.

Next, we consider the effect of planets in a very similar manner. Statistics from microlensing observations indicate that each Milky Way star hosts, on average, 1.6 planets [100]. We then assign stars n planets, where n is a Poisson random variable with mean 1.6, i.e., $n \sim \text{Pois}(\bar{n} = 1.6)$. If a

given primary star has any planets, then for each planet, we generate a semimajor axis $a_{\rm pl}$ and mass $M_{\rm pl}$. We assume $\ln a_{\rm pl}$ is uniformly distributed between 0.4 AU (inner cutoff) and 120 AU (the size of our solar system) as per Ref. [101]. We smooth the edges of this log-uniform distribution appropriately to prevent hard cutoffs. From this information, we can calculate the period using Kepler's third law $T_{\rm pl} = \sqrt{4\pi^2 a_{\rm pl}^3 / (GM_p)}$. We then compute the planet mass using the statistics for planet radii [102] and use the well-known planetary mass-radius relations [103] to convert radii to masses. We assume the planet radius $R_{\rm pl}$ is power-law distributed as $\sim R_{\rm pl}^{-2}$ and draw numbers between $4R_\oplus$ and $30R_\oplus$ (cutoffs), where $R_\oplus$ is the radius of the Earth. Once we have both the semimajor axis and mass of the planet, we can sum the RV contributions of the planets for each primary star as for stellar companions in Eq. (2.16) to obtain $v_{\rm plan}$. The only difference is that q in the case of planets is defined as $M_{\rm pl}/M_p$. Together, we refer to $v_{\rm comp}$ and $v_{\rm plan}$ as Keplerian terms as they are associated with orbiting bodies. Figure 2.4 shows the distribution of period versus semi-amplitude for the orbiting bodies in our synthesized dataset.

Finally, we include the effects of stellar activity (henceforth referred to as stellar noise). For our purposes, stellar activity consists of contributions to radial velocity measurements from stellar rotation, convection on the stellar surface and surface magnetic features such as spots and plages [4]. For a given primary star, the stellar noise contribution is taken to be distributed as a multivariate normal, i.e., $v_{\rm noise}(t) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K})$. The noise is taken to have zero mean (i.e., $\boldsymbol{\mu} = \mathbf{0}$) and covariance given by the kernel $\mathbf{K}$. The matrix $\mathbf{K}$ in turn may be expressed as the sum of two parts $\mathbf{K}_{\rm WN} + \mathbf{K}_{\rm QP}$, a white noise component and a quasiperiodic component. The white noise component, which models the instrumental noise (same for all stars), is given as $K_{\rm WN}(t_i, t_j) = \sigma_{\rm WN}^2 \delta(t_i, t_j)$, where $\sigma_{\rm WN}^2$ is the white noise variance and δ is a Kronecker delta. The elements of the quasiperiodic component are given as [104, 105]

$$K_{\rm QP}(t_i, t_j) = h^2 \exp\left(-\frac{\sin^2\left[\pi\left(t_i - t_j\right)/T_r\right]}{2w^2} - \frac{\left(t_i - t_j\right)^2}{\lambda^2}\right) \tag{2.17}$$

where h is scale of the magnetic activity (in RV units), T_r is the rotation period of the star, w is

a dimensionless "roughness" parameter and λ is the (stellar) spot lifetime. While this quasiperiodic model can capture both the periodic (i.e., stellar rotation) and transient (e.g., convection, spots, plages) aspects of activity phenomena, it is strictly phenomenological. Nevertheless, it has been found to be a reasonably successful proxy for a physics-based model [105]. In our simulations, we take $\sigma_{\rm WN} = 0.1$ or 0.6 m/s (depending on the scenario), $h \sim U\left(0.5, 2.5\right)$ m/s, $T_r \sim U\left(14, 35\right)$ days, $w \sim U\left(0.4, 0.6\right)$, and $\lambda \sim U\left(20, 30\right)$ days. $U\left(a, b\right)$ denotes a uniform distribution between a and b. The mean hyperparameter choices are estimates for the Sun, and we take reasonable ranges about these means. In the cases distributions are given, the parameters are drawn from these distributions for each star. The presence of stellar magnetic cycles further causes modulation of $\boldsymbol{\mu}, \mathbf{K}$ over the cycle period [106] and is straightforward to include in the simulation but is ignored here for simplicity.

It is worth noting that the mean planet occurrence rate of 1.6 planets per star is only valid up to an upper bound of 10 AU on $a_{\rm pl}$ [100], and this could lead to an underestimation of the number of planets when the range is extended up to 120 AU. Therefore, we also run some tests by inflating the mean number of planets per star.

2.5 Results

In a series of numerical experiments, we attempt to recover an injected ~ 5 cm/s/decade acceleration signal from a large ensemble of synthesized $v_{\rm total}$ time series for N simulated primary stars.

First, to assess what is possible with present-day state-of-the-art instruments and analysis techniques, we set the Gaussian white noise standard deviation $\sigma_{\rm WN}$ to 60 cm/s and add correlated noise with amplitudes ranging from 50 cm/s to 250 cm/s on a given star (the range of noise levels observed in the Sun [106]). In order to fit the more complicated noise model, we employ a Gaussian process (GP) regression (see Appendix C for details) with a quasiperiodic kernel function [104, 107, 108]. The GP regression is also able to simultaneously fit the Keplerian components of the time series (i.e., those associated with planets and stellar companions). With the GP fit

(N = 75,044 stars) we obtain a mean (standard error of the mean σ given in parentheses) of $1.42(0.24)\times10^{-8}$ cm/s^2. In contrast, a simple linear fit (N = 72,425 stars) to the time series yields a mean acceleration of $2.00(0.58)\times10^{-8}$ cm/s^2. Thus the GP fit reduces the uncertainty by more than a factor of 2 compared to the linear fit, recovers the injected stellar acceleration signal of 1.5×10^{-8} cm/s^2 within 1σ and is $\approx 6\sigma$ away from a null result. The probability density functions for the fitted accelerations are shown in Figure 2.5(a).

We can vary the observed number of stars N to determine what is required for a statistically significant detection. For our purposes, a detection at the level $n\sigma$ is defined as the mean lying $n\sigma$ away from zero. Figure 2.5(b) shows σ as a function of N. Unfortunately, nearly 2×10^4 stars are required for a 3σ detection ($\sigma \approx 0.5 \times 10^{-8}$ cm/s^2). This is a prohibitively large sample size given realistic quantities of observation time. However, we reiterate that N represents the number of stars in an initial sample before target selection.

Next, we study the case where $v_{\rm noise}$ contains only Gaussian white noise with $\sigma_{\rm WN}$ = 10 cm/s. This is a futuristic scenario we envision where data processing techniques developed in exoplanet astronomy have matured to the point of being able to effectively filter out the effects of correlated stellar noise. At present, this is an unsolved problem, but it is being worked on very actively [4, 5, 8]. Once stellar noise is removed, we reach the instrumental noise limit.

For this dataset of $N = 10^3$ stars, we use a linear fit for each time series (i.e., fit only the acceleration component), and construct a histogram of the fitted slopes. To avoid broadening the histogram with Keplerian signals, we employ Lomb-Scargle periodograms [3] to identify Keplerian signals and reject time series containing planets and stellar companions (without a priori knowledge of their existence).

Setting a threshold on the maximum allowed periodogram power close to the maximum observed from a pure noise signal rejects significant Keplerian signals. Our threshold allows 277 stars to be examined, including all 132 lone stars (i.e., free of orbiting bodies) in the sample. Figure 2.6 shows the result of such an analysis. Using this subset, we obtain a mean acceleration of $1.46(0.21)\times10^{-8}$ cm/s^2. This result is consistent with the injected stellar acceleration signal of 1.5×10^{-8} cm/s^2

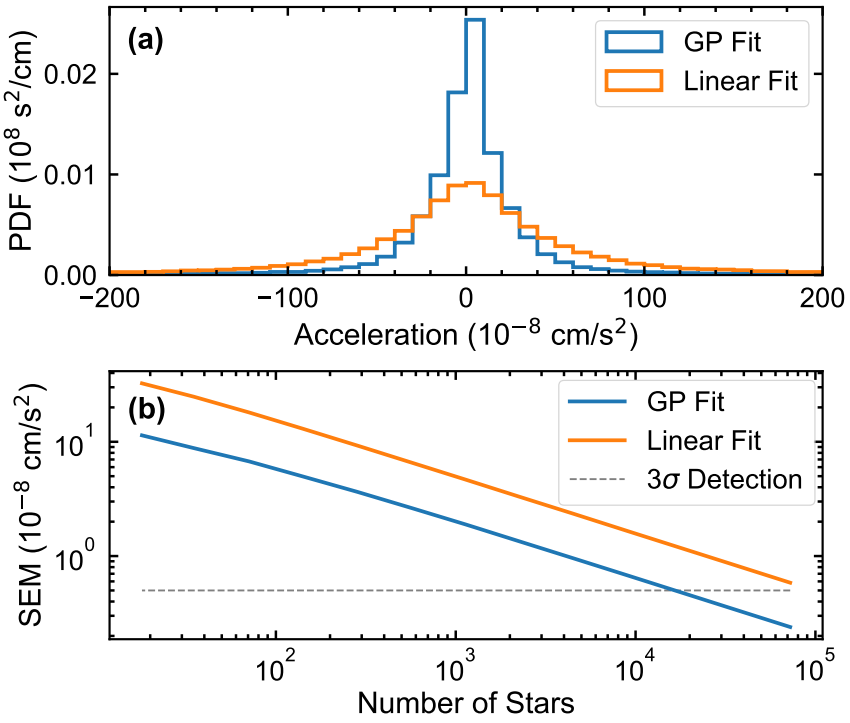

Figure 2.5: (a) Probability density function (PDF) of fitted stellar accelerations for $\sim 7.5 \times 10^4$ stars using Gaussian processes (GP) (blue) and a simple linear fit (orange). 60 cm/s of white noise and quasiperiodic correlated noise ranging from 50 cm/s to 250 cm/s is added to each time series. (b) Standard error of the mean (SEM) vs. number of stars observed. A 3σ detection (dotted grey line) is obtained after about 2×10^4 stars for the GP fit. (Figure courtesy of N. Langellier)

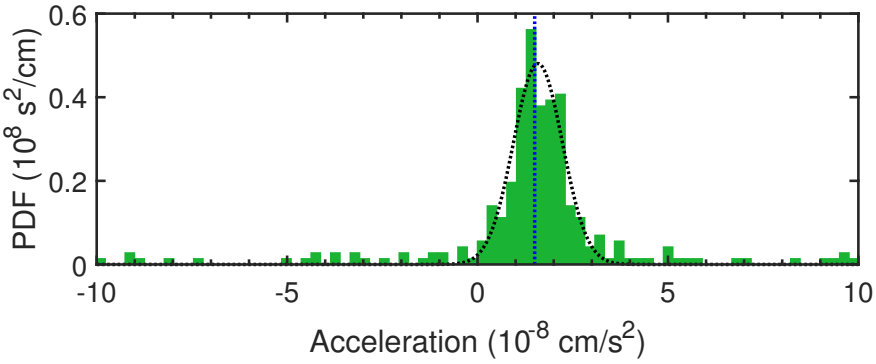

Figure 2.6: Probability density function (PDF) of fitted stellar accelerations for 10^3 stars, using a simple linear fit and filtering using periodogram power. The only source of noise included is 10 cm/s white noise. The dotted black curve is a Gaussian fit to the histogram (a guide to the eye). The injected acceleration signal of 1.5×10^{-8} cm/s^2 is shown as a blue dotted line.

and $\approx 7\sigma$ away from a null result. Note that an acceleration measurement with 14% uncertainty implies, using Eq. (2.14), a local DM density measurement with 38% uncertainty.

We also test the dependence of the acceleration sensitivity with planet occurrence by doubling the mean number of planets per star, keeping all other parameters fixed. This also doubles, on average, the number of long-period (>10 year) planets, which are difficult to remove without sufficient coverage of the orbital period. With the added planets it takes roughly 4 times as many stars to reach the same detection significance.

Instead of rejecting companions via periodograms, they can be used to help fit Keplerian signals. This could enable reaching the same precision in the result with far less telescope time. In the future, we plan to simulate the target selection program in more detail, keeping track of the time overhead associated with following poor targets and also do a study of the stellar acceleration precision versus number of observations, as telescope time is an expensive resource.

2.6 Conclusions and outlook

In this chapter, we put forth the idea of using precision RV measurements to quantify the local DM density. More specifically, we propose to track the velocity of stars over time to extract their accelerations, thereby directly probing the local gravitational potential and foregoing the equilibrium assumption used with static measurements of stellar velocities. The exquisite RV precision and stability achievable with astro-comb wavelength calibrators and exoplanet spectrographs, combined with next-generation large telescopes, should make it feasible to measure stellar accelerations directly at the necessary level of 10^{-8} cm/s^2. Importantly, detecting accelerations at this level with an ensemble of 10^3 stars requires the reduction of the effect of stellar noise on RV measurements to <10 cm/s. In the future, measurements over a wide range of pointing directions could allow construction of a map of the gravitational potential of the galaxy.

We conclude by emphasizing two key points. First, though the technical challenges are daunting

for a large observing campaign to map stellar accelerations and constrain dark matter models, we believe such a program will become feasible in the next decade. Second, the dataset that would result from this effort, providing precision RVs from $\sim 10^3$ stars, would be rich and valuable for other areas of astrophysics: e.g., many long-period exoplanets and stellar companions would likely be detected and characterized in the process. It is our hope that the exciting prospects for exoplanet and stellar astronomy as well as dark matter physics will encourage others to consider the stellar acceleration idea and pursue it further.

While we do not explore this possibility in detail here, the local DM density can also be measured by using vertical acceleration measurements of stars in the local neighbourhood but at sufficiently high vertical displacement from the disk z, such that the dominant vertical acceleration is from DM and not the disk. However, the vertical acceleration from DM is suppressed compared to the radial acceleration by the factor z/r_0. In addition, parameters fully characterizing the DM density profile, such as the local slope, may be determined from higher precision acceleration measurements.

Another idea that is similar in spirit to this proposal is using pulsars for measuring Doppler shifts (alluded to in Ref. [88]) in the time domain. Pulsars, which have been extensively studied by astronomers [109], are rotating neutron stars that emit beams of electromagnetic radiation. The rotation period of pulsars is very well-defined. Some fast-rotating pulsars ("millisecond pulsars") reach a fractional frequency stability of $\sim 10^{-14}$ at 1 year, which is only one order of magnitude worse than microwave clocks on Earth[110]. Thus, pulsars make relatively good natural clocks. When the beam of radiation sweeps by the Earth repeatedly, it registers as a sequence of pulses on terrestrial detectors (hence the name "pulsar").

Pulsars timing has been extensively studied by radio astronomers, especially in the context of detecting gravitational waves [111]. In the case of a perfectly defined pulse frequency ν, the galactic acceleration a causes a chirping $\dot{\nu}_{\rm accel} = a\nu/c$ as a result of the Doppler effect. For a typical millisecond pulsar ($\nu = 300$ Hz), and a typical galactic acceleration ($a = 10^{-10}$ m/s^2), we get $\dot{\nu}_{\rm accel} = 10^{-16}$ Hz/s $\approx 3 \times 10^{-9}$ Hz/yr, which is easily within reach of current instrumentation. An effect which is impossible to separate from this signal is the natural braking of the pulsar as a

consequence of radiative loss. The best pulsars[5] (which happen to be fast rotators in binary systems) have $\dot{\nu}_{\rm braking} \sim 10^{-16}$ Hz/s, so this becomes a challenging (SNR ≈ 1) but doable measurement if we can appropriately combine data from multiple pulsars. One final complication is that the timing measurement of pulsar braking requires correction for proper motion [112, 113]. More work is needed to fully evaluate the feasibility of this idea, but initial simulations look very promising.

[5] ATNF Pulsar Catalogue

3

High-resolution spectroscopic studies to enable oxygen detection on exoplanets

3.1 Introduction: Detecting Earth twins

Transit photometry is an extremely successful method of detecting exoplanets [3, 114]. This technique measures the amount of light coming from a star versus time. In the event of a planet passing in between the star and the observer, the amount of light received at the photometer is reduced. This dip is the telltale planetary signature. Transit photometry is capable of measuring the size of a planet as well as its period. In some cases, even the inclination of the system may be determined. When combined with radial velocity information (which gives the period and a mass weighted by the sine of the inclination), it can yield estimates of the density of exoplanets.

Combining virtues of the two techniques, one can perform spectroscopic measurements during a transit (so-called transit spectroscopy). In the transit configuration, some the host star's light passes through the exoplanet's atmosphere and is absorbed before it reaches the observer, as shown in Figure 3.1. The absorption signature is suppressed by a factor $\epsilon \sim 10^4 - 10^5$, where ϵ is the area ratio of the stellar disk and the planet's atmospheric ring[1]. For ground-based exoplanet searches, there is

[1]This range of ϵ is for planets of radius similar to Earth and stars of radius similar to the Sun.

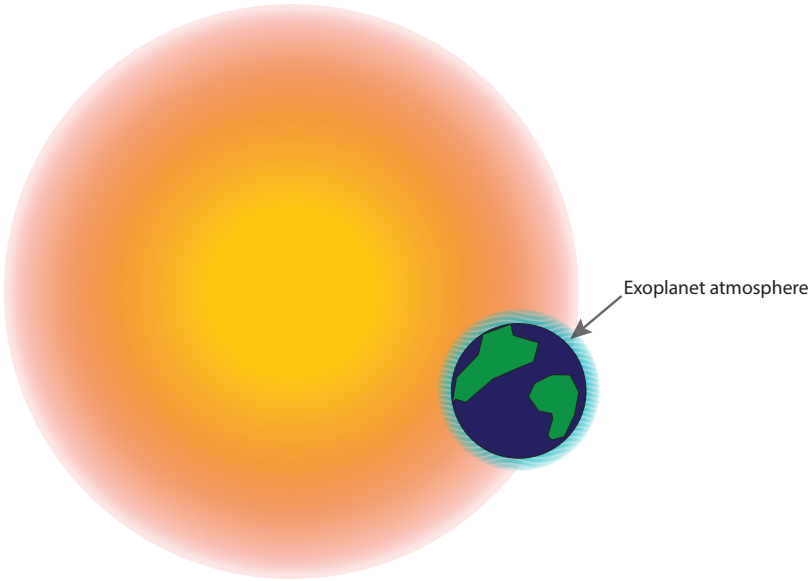

Figure 3.1: (As seen from Earth, but not to scale) When the exoplanet passes in front of the host star, some of the light from the star passes through the exoplanet's atmosphere before reaching the Earth. Transit spectroscopy aims to detect the properties of the exoplanet's atmosphere by performing spectroscopic measurements during the transit. ϵ (described in the text) is ratio of the area of the stellar disk seen in the background to the light blue ring around the exoplanet.

the confounding effect of the Earth's atmospheric lines, also known as telluric lines. However, it has been shown that atmospheric signal of the exoplanet may be recovered with very high-resolution spectrographs provided that the systemic Doppler shift of the extrasolar system is large enough to suitably separate the exoplanet signal from the telluric signal [115–119].

The variability of the telluric signal can be an important systematic for the detection of the exoplanet lines. If the lines are extremely stable, one can easily remove them using a model, leaving behind the exoplanet's signal. If on the other hand, the lines are perturbed strongly by environmental effects (e.g., pressure, temperature, humidity, wind speed and direction), then their contribution is not easy to subtract without simultaneous *in-situ* monitoring. The goal of this work is to assess the variability of the tellurics and make a recommendation as to whether simultaneous *in-situ* monitoring is required or not.

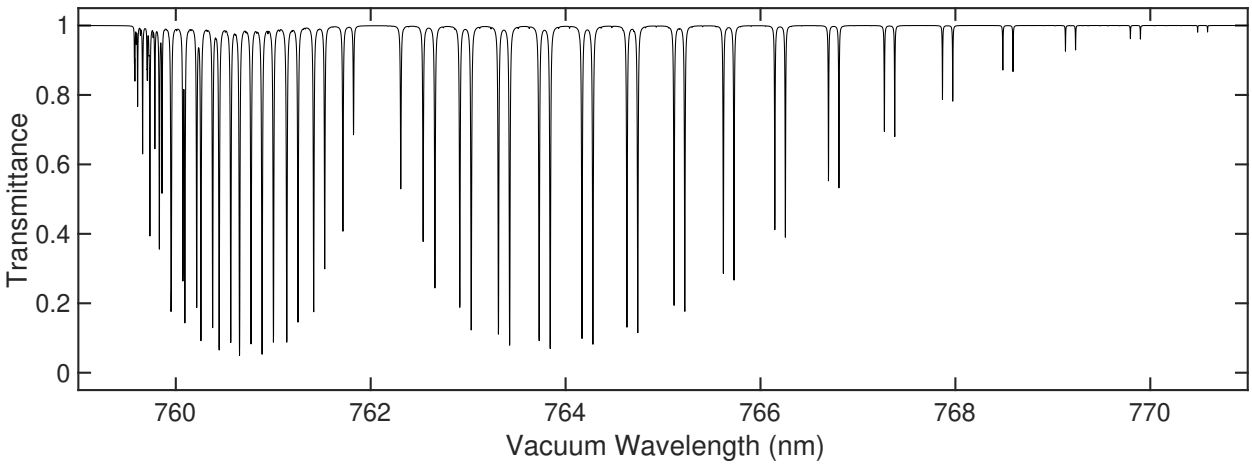

Figure 3.2: Simulated absorption spectrum of oxygen near 0.76 µm using HITRAN parameters. In the simulated spectrum, we include $^{16}O^{16}O$, $^{16}O^{17}O$ and $^{16}O^{18}O$ isotopologues of oxygen. Line profiles are taken to be purely Lorentzian, and widths are assumed to be air-broadened. The optical depth is chosen to be modest so that even the strongest lines are not saturated (purely for illustrative purposes). The group of dense lines from $\sim 759 - 762$ nm is the R-branch, and the sparser set spanning $\sim 762 - 771$ nm is the P-branch.

3.2 Spectroscopy of A-band of molecular oxygen

Oxygen is a very important biosignature on Earth [120], so it is instructive to search for spectroscopic evidence of oxygen when performing transit measurements. The A-band of molecular oxygen near 760 nm is particularly attractive due to the strength of the absorption, the transparency of the atmosphere down to the surface[2], and it being free from contamination by spectral lines from other atmospheric gases (e.g. water vapour, etc.). The A-band refers to the rovibronic transition $X^3\Sigma_g^-\,(v=0) \rightarrow b^1\Sigma_g^+\,(v=0)$. It has a rich rotational structure, further divided into P-branch ($\Delta N = -1$) and R-branch ($\Delta N = +1$), as shown in Figure 3.2. There are approximately 55 strong lines in this band and it has been extensively studied by spectroscopists in the laboratory [121–123] and in the atmosphere [123–127].

If we examine the A-band lines closely, we can see that they always seem to occur in doublets. To understand the source of this structure, we have to delve a bit into molecular structure [128]. A molecular term symbol for an electronic state reads ${}^{2S+1}\Lambda_{g,u}^{\pm}$, where S is the spin, Λ is the projection

[2]In the transit configuration, refraction limits lowest observable atmospheric depth however [116].

of the orbital angular momentum onto the internuclear axis, $\pm$ is the reflection symmetry of the electronic orbital along an arbitrary plane containing the internuclear axis. Finally, g, u is the inversion symmetry of the electronic orbital across the centre of mass — ungerade vs. gerade (only for homonuclear diatomics). The Laporte rule forbids $g \to g$ and $u \to u$ as electric dipole transitions involve a change in parity. Consequently, the A-band transitions ($g \to g$) are magnetic dipole transitions (i.e., 9 orders of magnitude weaker than the Rb D_2 line).

To understand the doublets, let us now consider the rotational levels. We restrict ourselves to the isotopologue $^{16}O_2$ in this discussion. Since ^{16}O is a boson, the overall spatial wavefunction needs to be symmetric with respect to interchange of the atoms. From the Born-Oppenheimer approximation, we can write the overall wavefunction as $\psi = \psi_{\text{el}}\psi_{\text{vib}}\psi_{\text{rot}}\psi_{\text{nuc}}$ where the overall parity is then determined by a product of parities. The nuclear wavefunction can only be symmetric in this case. Vibration need not be considered because interchanging particles leaves the internuclear separation the same. The electronic wavefunction parity comes from the superscript $\pm$ because it has been demonstrated [129, 130] that the inversion in the lab frame (with origin at the centre of mass) is equivalent to reflection in the molecule frame (z-axis pinned to internuclear axis).

Knowing that the ground state $X^3\Sigma_g^-$ has antisymmetric electronic wavefunction, we then require an antisymmetric rotational wavefunction to get an symmetric overall wavefunction. The rotational wavefunction changes parity as $(-1)^N$, where N is the rotational quantum number. Thus, only odd N is allowed. Given that this is a triplet state, i.e. $2S + 1 = 3$ or $S = 1$, we know that the values for the total angular momentum $J = \left|\vec{N} + \vec{S}\right|$ are $N - 1, N, N + 1$.

For the excited state $b^1\Sigma_g^+$ ($v = 0$), we have a symmetric electronic wavefunction, so we need a symmetric rotational wavefunction to retain symmetric overall wavefunction. Therefore, only even values of N are allowed in the upper state. Furthermore $J = N$ because we have a singlet state, i.e. $2S + 1 = 1$ or $S = 0$.

From the selection rule $\Delta J = 0, \pm 1$ for magnetic dipole transitions, one gets the following spin-rotation transitions [121]

- $J'' = N'' - 1, N'' \rightarrow J' = N' = N'' - 1$ PQ and PP
- $J'' = N'', N'' + 1 \rightarrow J' = N' = N'' + 1$ RR and RQ

Double primes denote the lower state and single primes denote the upper state. The transitions are divided into P-branch ($\Delta N = -1$) and R-branch ($\Delta N = +1$). In the figure above, the R-branch is on the left and the P-branch is on the right. The transitions are then subdivided by ΔJ using the same conventions. Therefore, each rotational transition consists of a doublet PP and PQ or RQ and RR. The somewhat regular spacing of the rotational doublets comes from the form of the rigid rotor Hamiltonian $H_{\text{rot}} = BJ(J+1)$, and the doublets arise from spin-rotation interaction. Deviations from the regularity of the spacing arise from so-called centrifugal distortion: as the rotor spins faster, the centrifugal force stretches bond and increases moment of inertia (leading to a second-order correction $H_{\text{rot}} \rightarrow H_{\text{rot}} - DJ^2(J+1)^2$, where D is the centrifugal distortion). The overall envelope of the P- and R-branches come from the thermal occupation of the rotational states in the ground manifold.

The A-band spectral lines are subject to both homogeneous and inhomogeneous broadening mechanisms in the Earth's atmosphere [128]. The natural Lorentzian lineshape has a width given by the inverse of the excited state liftetime. Due to the fact that the A-band transitions are dipole-forbidden, the natural linewidth is on the order of 10 mHz. However, the true linewidths in the atmosphere are affected by both the finite pressure and temperature of molecules. The effect of pressure is essentially thought of as reducing the excited state lifetime via collisions. Therefore the pressure-broadened line remains Lorentzian, but the width grows with pressure. The Lorentzian lineshape is given as

$$f_L(\nu; \nu_0, \alpha_L) = \frac{1}{\pi} \frac{\alpha_L}{\alpha_L^2 + (\nu - \nu_0)^2}, \tag{3.1}$$

where ν_0 is the centre frequency of the line, α_L is the half width at half maximum (HWHM). It should be noted that pressure can also impart line shifts; intermolecular interactions cause energy level shifts in the molecules, thereby changing the transition frequencies. The next effect to consider is the temperature of the gas. At finite temperature, a range of velocity classes are populated according to a Gaussian distribution, where the width $\sim \sqrt{T}$. This leads to a range of Doppler

shifts, effectively convolving the natural lineshape with the velocity distribution. Given that the Doppler width greatly exceeds the natural linewidth, the Doppler-broadened lineshape essentially reduces to a Gaussian

$$f_D\left(\nu;\nu_0,\alpha_D\right)=\frac{\sqrt{\ln 2/\pi}}{\alpha_D}\exp\left(-\ln 2\left(\frac{\nu-\nu_0}{\alpha_D}\right)^2\right), \tag{3.2}$$

where the HWHM $\alpha_D=\nu_0\sqrt{2k_BT\ln 2/\left(mc^2\right)}$. k_B is Boltzmann's constant, T is the temperature, m is the mass of the atom (or molecule) and c is the speed of light in vacuum. To incorporate the effects of both pressure and Doppler broadening, it is common to use a Voigt lineshape,

$$f_V=f_L*f_D, \tag{3.3}$$

where $*$ denotes a convolution integral. Fitting with this lineshape is more computationally intensive, but algorithms to compute it have improved recently [131]. At ambient conditions (1 bar, 300 K), the pressure broadening ($\sim$ 3 GHz) [121, 122] dominates the width for the A-band lines (Doppler width $\sim$ 1.2 GHz).

Summing over many transitions, we get the absorption spectrum of the atmosphere according to the Beer-Lambert law

$$T\left(\nu\right)=\exp\left[-nl\sum_i S_i f_i(\nu;\nu_i,\Gamma_i)\right], \tag{3.4}$$

where T is the transmittance (dimensionless, ranging from 0 to 1), n is the number density of molecules (i.e., molecules/cm^3) and l is the optical path length (cm). S_i is the intensity of the i^{th} transition (units of cm/molecule), and f_i is the lineshape function of the i^{th} transition (units of cm). The product S_if_i can be thought of as the i^{th} cross-section σ_i (cm^2/molecule). This formulation assumes a constant density atmosphere.

If one draws the photon path from the source (in this case, the Sun) to the detector on the surface of the Earth, l is equal to the length of the ray confined to within the atmosphere (often taken to be $\sim$ 9 km thick). For simplicity, we neglect any atmospheric inhomogeneities. The length of this

ray naturally changes over the course of the day due to the motion of the Sun in the sky. We can easily relate it to the the relative airmass z commonly used in astronomy

$$z = \sqrt{\left(\tilde{r} + \tilde{h}\right)^2 \cos^2\theta + 2\tilde{r}(1-\tilde{h}) - \tilde{h}^2 + 1} - \left(\tilde{r} + \tilde{h}\right)\cos\theta, \tag{3.5}$$

where $z = l/h_{\text{atm}}$, and h_{atm} is the thickness of the Earth's atmosphere. Nondimensionalized variables $\tilde{r} = R_\oplus/h_{\text{atm}}$ and $\tilde{h} = h_{\text{obs}}/h_{\text{atm}}$. $R_\oplus$ is the radius of the Earth and h_{obs} is the elevation of the observer. The solar zenith angle θ (i.e. angle from the vertical) is obtained from a solar position calculator [132] given the observer latitude, longitude and time.

3.3 Assessments of telluric variability

3.3.1 Gathering data

The data for this project was generously supplied by Ralf Sussmann and Markus Rettinger of the Total Carbon Column Observing Network (TCCON) collaboration. "TCCON is a network of ground-based Fourier Transform Spectrometers recording direct solar spectra in the near-infrared spectral region."[3] TCCON instruments track the Sun, and take spectra at $\sim$ 1-minute cadence. TCCON data is primarily used for environmental science [133].

To replicate the conditions of the Giant Magellan Telescope ($h_{\text{obs}} \sim 2.5$ km) [91], where the search for Earth twins will eventually take place, we requested data from a TCCON site with similar altitude. We were able to get data over two dry winter days (January 16 and 17, 2019) from the Zugspitze site (47.421078° N, 10.983358° E, $h_{\text{obs}} = 2.961$ km). This site has a Bruker 125HR Fourier transform with the following detectors: InGaAs, Si, InSb, HgCdTe. The Si detector allows us to observe the near-infrared signal of the oxygen A-band at 760 nm. The instrument has a maximum optical path difference of $L = 45$ cm, so the theoretical resolution is $\frac{1}{2L} \approx 0.01$ cm^{-1} = 300 MHz [134]. A Fourier transform spectrometer (essentially a Michelson interferometer) records a signal vs.

[3] tccon.caltech.edu

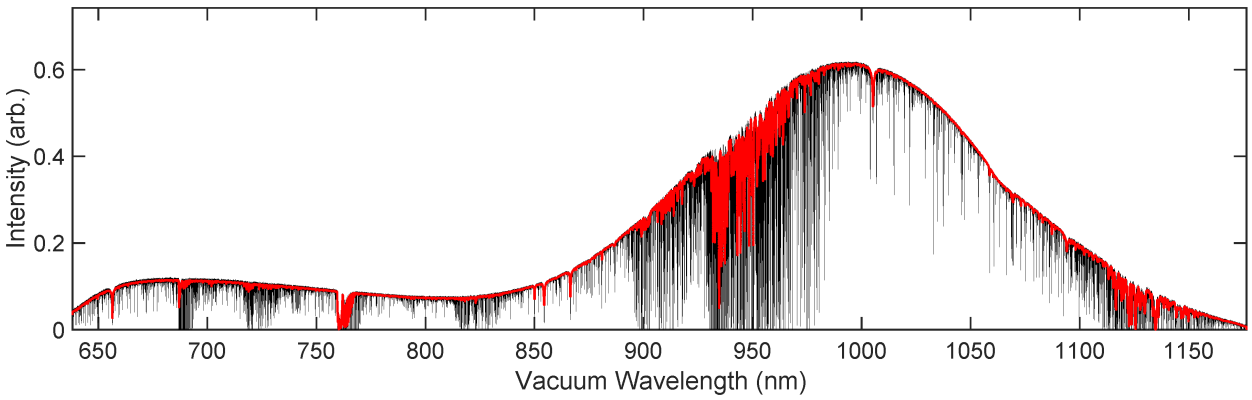

Figure 3.3: Example of raw FTS spectrum (black) with median-filtered data shown in red (used for continuum normalization). The A-band is seen as a pronounced dip near 760 nm. It is also important to fit the zero-level offsets as discussed in the text.

optical path difference (a so-called interferogram) and this signal is Fourier transformed to retrieve the spectrum. We received nearly 1800 spectra, recorded over the course of 2 days, with a spectrum taken approximately every 20 s.

Along with the spectra, we also received logs of the temperature, pressure, relative humidity, wind speed and wind direction at the instrument (at hourly intervals). These telemetric data are extremely valuable in searching for correlations in the atmospheric spectra. Finally, the peak interferogram amplitude was also recorded and archival webcam snapshots are available for cloud monitoring.

3.3.2 Data reduction algorithms

The first step in processing the data is averaging several raw spectra to increase our signal-to-noise ratio. Typically, we average 5 min worth of data ($\sim$ 15 spectra) to get sufficiently high SNR (see Figure 3.3 for an example). After this step, we have to normalize by the continuum to get a transmittance spectrum. The continuum level defined by the broadband emission of the Sun as well as the instrument response. To normalize this level, we first median filter the raw spectra (red trace in Figure 3.3), which effectively removes all the absorption features as they constitute a "salt-and-pepper" noise due to their relative sparsity and sharp, impulse-like characteristics. After median

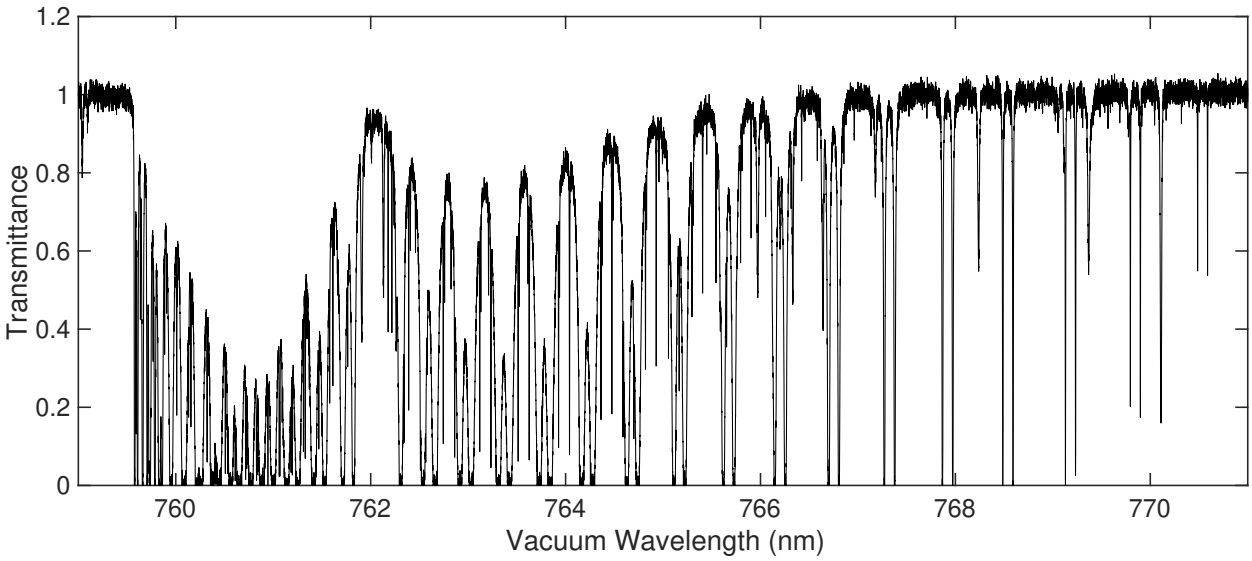

Figure 3.4: Full A-band transmittance spectrum taken at Zugspitze on June 16, 2019 at 12:30 PM local time.

filtering, the we fit a third-order polynomial to the continuum in the vicinity of the A-band and divide the data by the fitted polynomial to obtain a transmittance. The final step is to correct for zero-level offsets in the data. A zero-level offset is an artifact where the mean signal level at 100% absorption is nonzero (sometimes slightly negative). Within the A-band, we pick a few spectral regions of where the atmosphere is known to be completely opaque and take the mean to find the zero-level offset. We then vertically shift the spectrum accordingly to correct for this offset h, and divide the spectrum by $1-h$ to rescale such that its values lie between 0 and 1. One can interchange the order of the the zero-level offset correction and continuum normalization.

Once we have completed these pre-processing steps, we have taken the raw FTS data and converted it into a transmittance spectrum, as shown in Figure 3.4. Doing this for every 5-minute interval over four hours leads to a time series of transmittance spectra, as shown in Figure 3.5. Working with the minimum airmass data (as these best replicate the nighttime observing conditions) around midday on January 16, we select a range of four hours to simulate the duration of a transit observation[4]. We restrict our spectral region to a portion of the P-branch ($762.42 - 765.42$ nm) to avoid deeply

[4]This estimate is based on the transit of an Earth-sized planet in the habitable zone of an M4V star ($\backsimeq 0.2 M_{\odot}$). M4V stars are chosen as they are optimal for observations [116]. The transits are expected to last $\backsimeq$ 2 hours (and reoccur every $\backsimeq$16 days). In addition to the duration of the actual transit, we need some amount of time to set an out-of-transit baseline to separate the stellar signal from the signal containing the star as well as the planet. This is why 4 hours of observation are required.

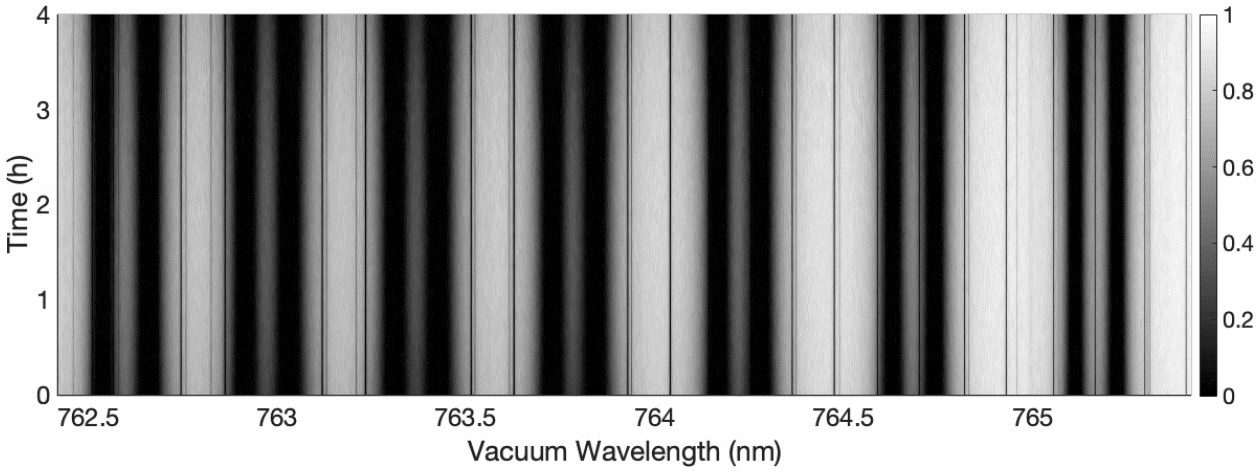

Figure 3.5: Time series of A-band transmittance spectra (3 nm-wide slice in the P-branch) taken at Zugspitze on June 16, 2019 from 10:30 AM ($t = 0$ h) to 2:30 PM ($t = 4$ h) local time. Colour bar indicates the transmittance.

saturated lines, strong solar features[5] and have a reasonable number of lines (~ 50) to fit. We assume that this set of lines is representative of the behaviour of all the lines in the A-band.

At this point, there are still major variations in the spectra due to the airmass variation over the day. If one takes the mean spectrum over the four hours as a template $\bar{T}$, and computes a cross correlation of T_i (spectrum at time t_i), we get a time series of cross-correlation functions $\mathrm{CCF}_i(\nu) = T_i(\nu) \star \bar{T}(\nu)$, where $\star$ denotes a cross-correlation. Plotting the CCFs at three different times ($t = 0$ designates the start of the four-hour time period), we see that the primary difference between the CCFs is the peak value (at zero delay) — this is a proxy for the offset. Plotting this peak value vs. the relative airmass reveals a nice linear correlation as shown in Figure 3.6.

To search for residual effects beyond the airmass variation, we follow Ref. [135] and take the logarithm of T_i and divide it by z_i (relative airmass at time t_i). Before doing so, we threshold the data such that we discard data within 2% of $T = 0$ or $T = 1$, such that we avoid introducing excess noise from taking the logarithm. The next step is pruning the lines to avoid strong degeneracies in the fitting of blended lines. First we search for and flag very strong lines of the isotopologue $^{16}\mathrm{O}^{16}\mathrm{O}$

[5] Solar absorption lines may be fitted with a phenomenological solar lineshape given by Toon: $f_S(\nu) = A \exp\left(-\frac{(\nu-\nu_9)^2}{\sqrt{d^4+(\nu-\nu_9)^2\Gamma^2}}\right)$. A is the line amplitude, ν_0 is the centre frequency, Γ is related to the width, and d is a shape factor. The spectrum is then adjusted by directly subtracting solar lines, i.e. $T(\nu) \to T(\nu) - \sum_j f_{S_j}(\nu)$, where f_{S_j} is the j^{th} solar line.

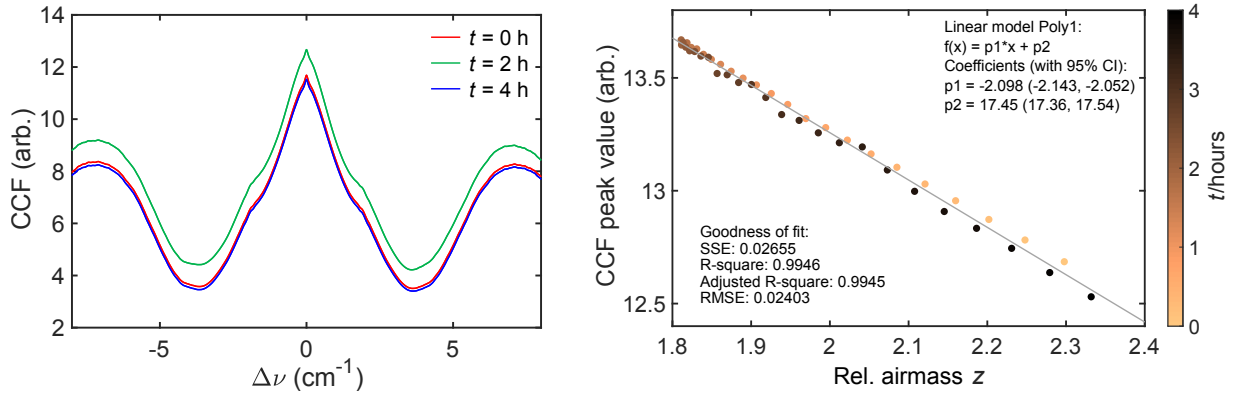

Figure 3.6: (Left) Cross-correlation of A-band transmittance spectra at various times ($t = 0, 2, 4$ h) with the template spectrum obtained by averaging all four hours worth of data. The x-axis is the frequency shift. (Right) The primary variation in the CCFs is the explained by the peak value (at $\Delta\nu = 0$), and this fits very well to a model with a linear dependence on relative airmass z (correlation coefficient $R^2 > 0.99$). The scatter plot is also colour-coded by time.

($S > 1$ cm/molecule). In the vicinity of these strong lines, we check to see if there are other lines within 0.5 cm^{-1} ($\sim 3\times$ FWHM) of line centre of the strong line. If there are other lines, they are dropped from the list. Finally, we also drop extremely weak lines ($S < 10^{-4}$ cm/molecule) for all isotopologues.

Pressure shifts for the A-band are ~ 100 ppb, comparable to the absolute wavenumber accuracy of the FTS[6], so the wavenumber axis also needs to be calibrated. We do this by fitting two strong solar lines on both sides of the A-band as fiducial markers and linearly interpolate the frequencies in between. Our lines of choice[7] are the K I line at 12,985.185724 cm^{-1} and the Fe I line at 13,174.547 cm^{-1}. To obtain the marker frequencies in the Sun, the final step is to apply two corrections: (1) a gravitational redshift of 633 m/s [136] and (2) a barycentric correction[8].

3.3.3 Multipeak fitting

After preprocessing, the data is ready for fitting. To fit the data, we employ a sum of Lorentzian functions (ignoring Doppler broadening for simplicity), each having a line amplitude, a centre fre-

[6] IFS 125HR User Manual, Bruker Optik GmbH
[7] Frequencies listed are from the NIST Atomic Spectra Database.
[8] Barycentric corrections are performed using the JPL HORIZONS System ephemeris.

quency and width. Guesses for these parameters are generated from the high-resolution transmission molecular absorption database (HITRAN - see Ref. [137]). We include all three isotopologues of molecular oxygen: $^{16}O^{16}O$, $^{16}O^{17}O$ and $^{16}O^{18}O$. Contamination from other gases (e.g., water vapour, carbon dioxide) is negligible in this spectral region. The guesses for the parameters defining the k^{th} line are based on the following formulae:

$$A_k^* = S_k/\Gamma_k^*, \tag{3.6}$$

$$\nu_k^* = \nu_k + \delta_k\left(p_{\text{ref}}\right)p, \tag{3.7}$$

$$\Gamma_k^* = \left(\frac{T_{\text{ref}}}{T}\right)^{n_{\text{air},k}}\left[\Gamma_{\text{air},k}\left(p - p_{\text{self}.k}\right) + \Gamma_{\text{self},k}p_{\text{self}.k}\right]. \tag{3.8}$$

A^*, ν^* and Γ^* are, respectively, the guesses for the amplitude, centre frequency and width of the line. p is the local pressure, $p_{\text{ref}} = 1$ bar and δ is the pressure shift ($\text{cm}^{-1}/\text{bar}$). T is the local temperature, $T_{\text{ref}} = 296$ K, and n_{air} is the coefficient of the temperature dependence of the air-broadened half width Γ_{air}. p_{self} is the partial pressure of the species corresponding to the k^{th} line and Γ_{self} is the self-broadened half width. S is the line intensity, as before. The quantities $\nu_k, S_k, n_{\text{air},k}, \Gamma_{\text{air},k}, \Gamma_{\text{self},k}$ are sourced from HITRAN. Using these guesses, we can construct the full fit function

$$f\left(\nu; \mathcal{C}, A_1, \ldots, A_N, \nu_1, \ldots, \nu_N, \Gamma_1, \ldots, \Gamma_N\right) = \mathcal{C}\sum_{k=1}^{N} A_k f_L(\nu; \nu_k, \Gamma_k), \tag{3.9}$$

where $\mathcal{C}$ is a overall prefactor. N is the number of lines. As we can see, f has $3N+1$ free parameters. We initialize all the free parameters to the guess values denoted by $*$ superscripts.

Our aim is to fit the data F to f. We use a constrained optimization method (trust-region-reflective) to perform nonlinear least squares regression as we have physically-motivated constraints which can reduce the size of the search space dramatically. The next step is to properly set bounds for the parameters. Considering the centre frequencies, we fix the frequencies where there is no data due

to the thresholding, i.e., $\nu_k = \nu_k^*$ if $F(\nu_k^*) = \text{NaN}$. Otherwise,

$$\nu_k^* - R \leq \nu_k \leq \min\left(\nu_k^* + R, \nu_{k+1}^*\right) \text{ if } k = 1, \tag{3.10}$$

$$\max\left(\nu_k^* - R, \nu_{k-1}^*\right) \leq \nu_k \leq \min\left(\nu_k^* + R, \nu_{k+1}^*\right) \text{ if } 2 \leq k \leq N-1, \tag{3.11}$$

$$\max\left(\nu_k^* - R, \nu_{k-1}^*\right) \leq \nu_k \leq \nu_k^* + R \text{ if } k = N. \tag{3.12}$$

In practice, this means the centre frequencies are allowed to roam a distance $R = 0.5$ cm^{-1} ($\sim 3 \times$ FWHM), unless they run into next line (above or below in frequency). The line amplitudes are fixed, the widths are allowed to vary by approximately 50%, and the overall prefactor $\mathcal{C}$ is allowed to vary between 10% and 200% of its initial guess (i.e., very widely). This configuration works extremely well and leads to convergence of the multipeak fit over the four-hour duration of the dataset with none of the parameters saturating their bounds. An example is shown in Figure 3.7.

3.3.4 Telluric variability from multipeak fit parameters

We now have $3N + 1$ fitted parameters for 48 instances in time. To distill the information into a few manageable metrics, we can construct three intensity-weighted averages

$$\hat{S} = \sum_{k=1}^{N} s_k \cdot S_k, \tag{3.13}$$

$$\hat{\delta\nu} = \sum_{k=1}^{N} s_k \left(\delta\nu\right)_k, \tag{3.14}$$

$$\hat{\Gamma} = \sum_{k=1}^{N} s_k \cdot \Gamma_k, \tag{3.15}$$

where $s_k = S_k / \sum_k S_k$, and $S_k = A_k \Gamma_k$. The centre frequency shift $\delta\nu_k$ is defined as $\nu_k - \nu_k^*$ (ν_k is the fitted centre frequency and ν_k^* is the initial guess). Once we have these weighted averages, we can see how they vary with environmental parameters. One can also do much more detailed line-by-line correlation analyses, but this is beyond the scope of the current work. First, we can plot the telemetric data vs. time to see if there are any obvious trends. This is shown in Figure 3.8.

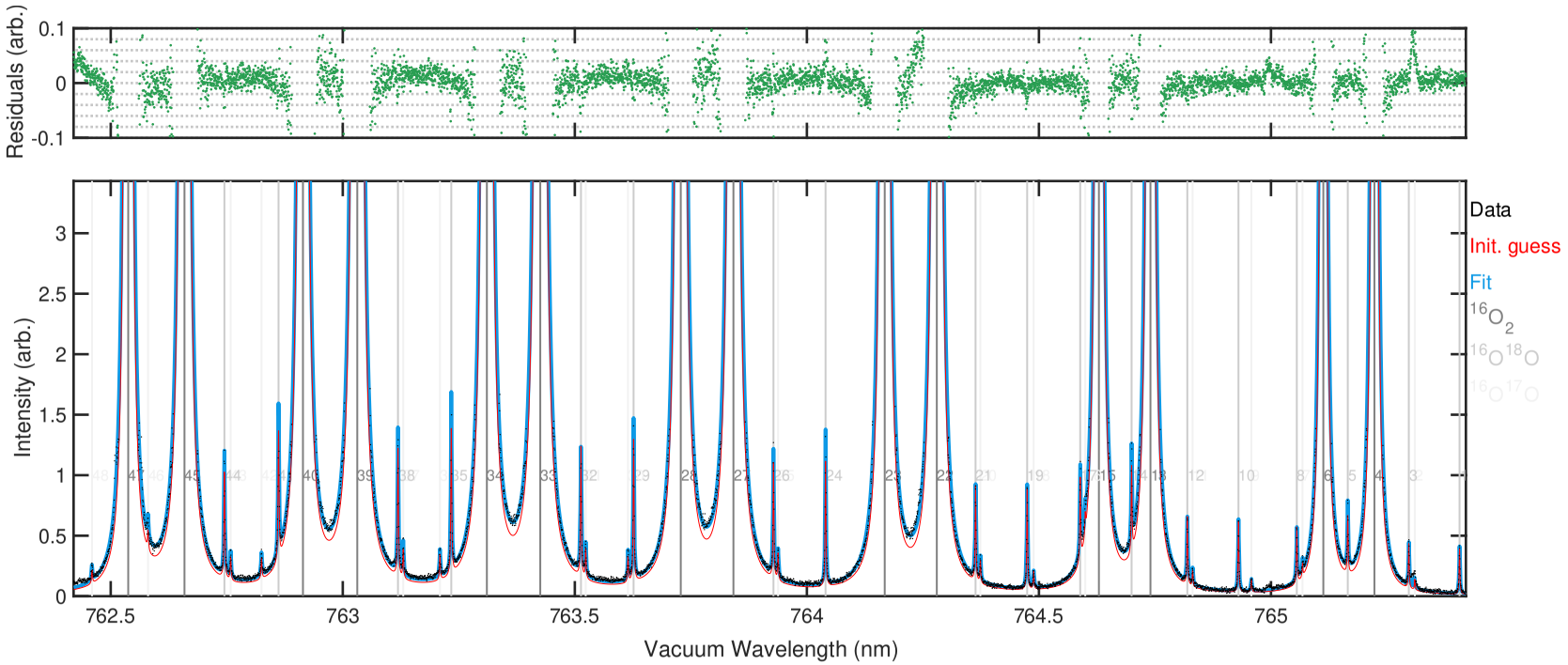

Figure 3.7: (Main panel) Example fit to A-band spectrum from $762.42 - 765.42$ nm. The quantity on the y-axis is derived by taking $-\frac{1}{z}\log T(\nu)$, where T is the transmittance and z is the relative airmass. The data (black dots) is fitted to a sum of Lorentzians using nonlinear least-squares regression. The gaps in the data arise from the thresholding procedure discussed in the text. The solver is started at the initial guess indicated by the thin red curve and converges to the light blue curve. The fit function contains spectral lines (see vertical line markers) associated with $^{16}O^{16}O$ (dark grey), $^{16}O^{18}O$ (medium grey), $^{16}O^{17}O$ (light grey). (Upper panel) Fit residuals (green dots) vs. vacuum wavelength.

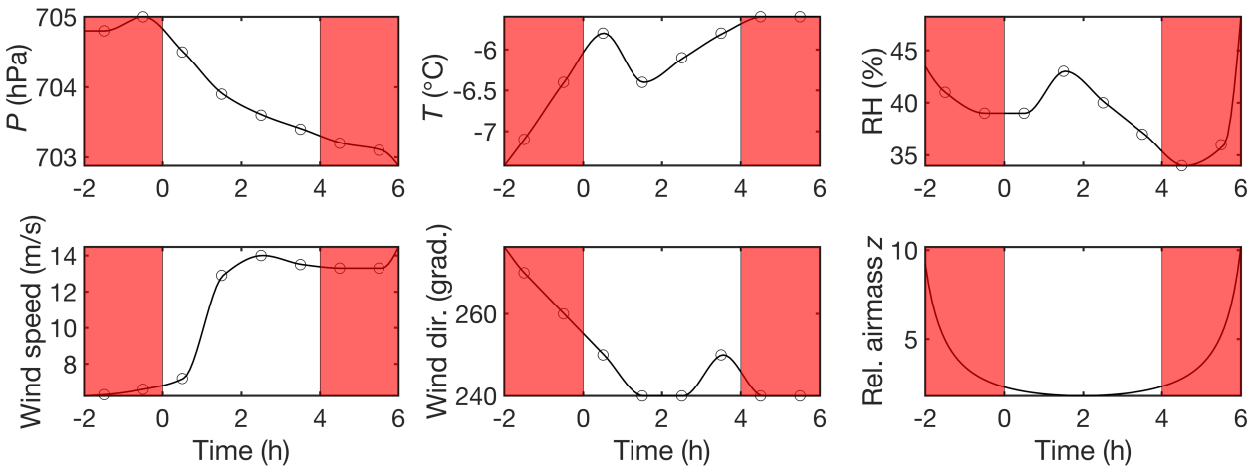

Figure 3.8: (Top row, from left to right) Pressure, temperature and relative humidity at Zugspitze summit on June 16, 2019 from 8:30 AM ($t = -2$ h) to 4:30 PM ($t = 6$ h) local time. (Bottom row, again from L to R) Wind speed, wind direction and relative airmass (computed) vs. time. Actual data points are indicated by black circles. Given the poor time resolution of the telemetric data, we assume a smooth variation of environmental parameters and interpolate using a shape-preserving cubic spline (black curves). Our analysis are performed only with the low-airmass data ($0 \text{ h} \leq t \leq 4 \text{ h}$ range shown between shaded red regions) to best replicate eventual nighttime observing conditions, but the full dataset is shown for completeness.

Next, we can plot the weighted averages vs. the telemetric data to search for correlations, as shown in Figure 3.9. There is little correlation in either the top or bottom row (line intensity and width). The strongest correlation visible is the increase of the centre frequency shift with time. The reason for the shapes of the correlations in the second row is because the environmental parameters are themselves correlated with time (i.e., pressure drops monotonically, wind speed increases monotonically and saturates, etc.). Assuming our calibration is perfect, it is difficult to say what actually causes the shifts. One possibility is that the effective pressure of the atmosphere is poorly represented by the pressure measured on the ground.

It is reassuring to note that there is very little correlation of any of the intensities or widths with relative airmass z. In fact, we can reconstruct the CCF from the fitted multipeak functions to see how well the airmass correction works. The result is shown in Figure 3.10. As we can see in the right panel, there is very little variation of the CCF with time, and the residual variations can be attributed to environmental fluctuations. To reiterate, the airmass variation was removed by taking the (negative) logarithm of i^{th} transmittance spectrum T_i and dividing it by z_i (relative

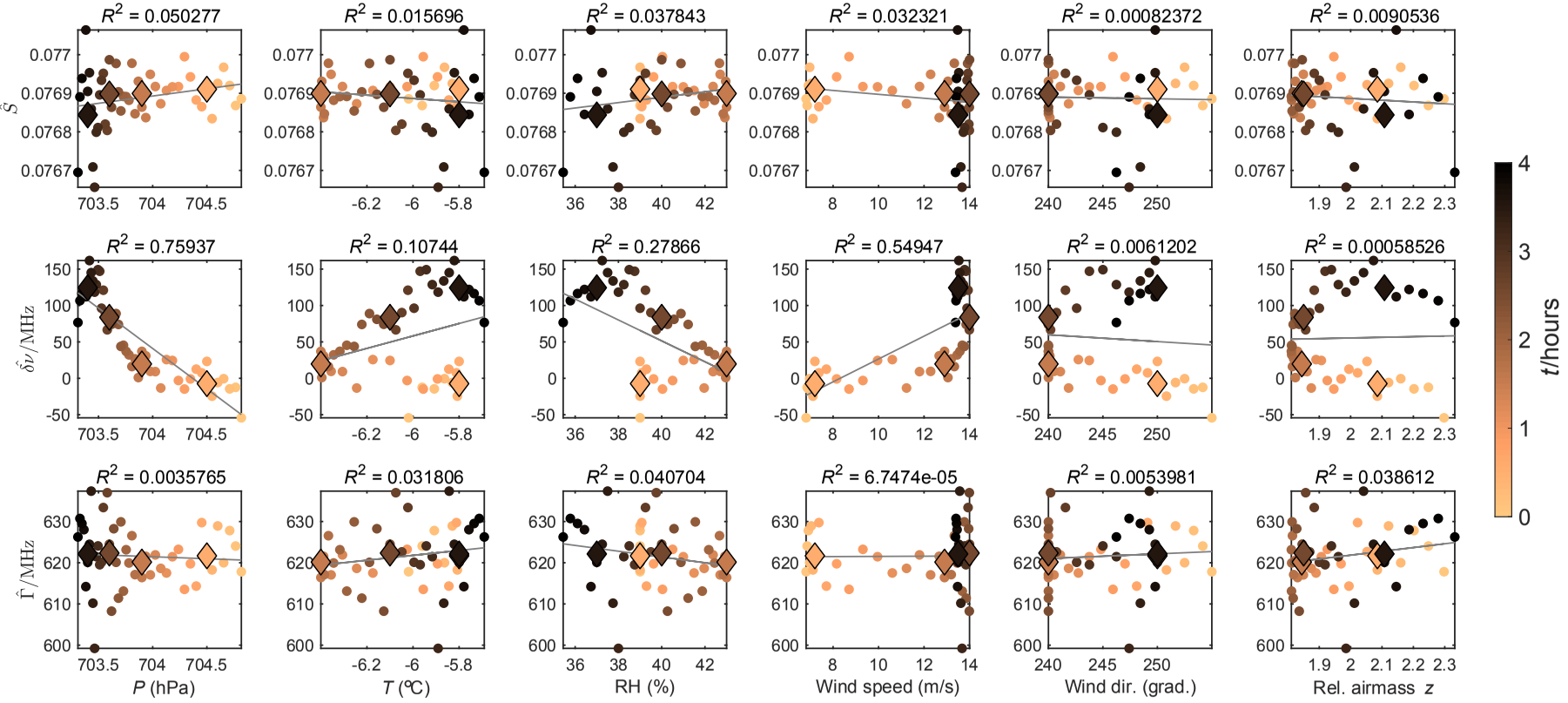

Figure 3.9: Line parameters vs. environmental parameters. Rows (top to bottom): Averaged intensity $\hat{S}$ (arb.), averaged frequency shift $\hat{\delta\nu}$, and averaged half-width $\hat{\Gamma}$. Columns (left to right): Local pressure, temperature, relative humidity, wind speed, wind direction and relative airmass at Zugspitze summit on June 16, 2019 from 12:30 PM ($t = 0$ h) to 2:30 PM ($t = 4$ h) local time. Data is colour-coded by time (as per the colour bar on the right). In each plot, the dots use the cubic spline interpolation of environmental parameters given in Figure 3.8, while the larger diamonds use the raw environmental data and hourly averages of the line parameters. Grey lines are fits of the dots to a linear model, with the correlation coefficient R^2 given above each plot.

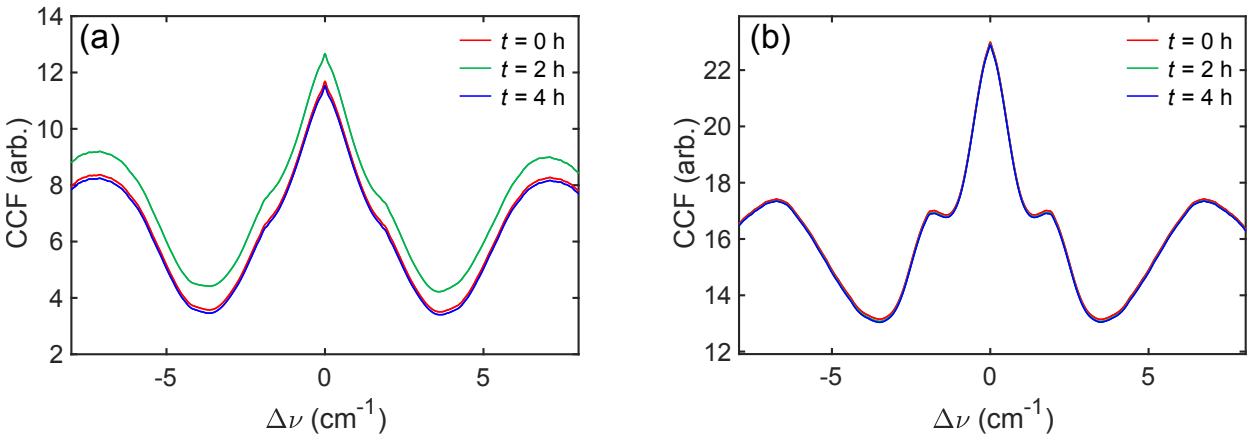

Figure 3.10: (a) Cross-correlation of A-band transmittance spectra with airmass variation included (same as Figure 3.6, but repeated for comparison) (b) Reconstructed cross-correlation of A-band transmittance spectra after removing effect of airmass variation.

airmass at time t_i). The transformed data is then fitted to a model f, and the transmittance is then reconstructed by taking the $\exp(-f)$. CCFs are computed as before. We use the f instead of the data since the data has discontinuities which make computing the CCFs nontrivial.

At this point, the data quality is too low (due to the infrequent telemetric data) to draw serious conclusions about telluric variability. Having telemetric data at the same cadence as the FTS spectra would certainly improve our confidence in correlations we might observe. However, it is still unclear how effectively local environmental parameters represent the bulk properties of the atmosphere (one might need balloon measurements at various altitudes). We leave such analyses for future work.

3.4 Conclusions and outlook

This work is still much in its infancy. So far, we have developed a suite of software tools to analyze the telluric variability in atmospheric spectra. Some future steps include:

- Do full P-branch fits including all oxygen lines and solar lines (R-branch is likely too saturated to be useful).

- Use more sophisticated lineshapes, e.g., Voigt, or speed-dependent Voigt [127].
- Develop more sophisticated data reduction and fit procedures: avoid continuum normalization and taking logarithms of the raw data. For example, the GFIT [124] algorithm directly fits the raw data by incorporating zero level offset, continuum and instrument lineshape as fit functions. Furthermore, the transmittance function uses a layered atmospheric model instead of a homogenous atmosphere.
- Understand instrument lineshapes [134] and sources of correlated noise in spectra.
- Understand line-by-line parameters variations to find indicators of atmospheric instability.
- Repeat the measurements at nighttime with a bright star to better replicate observing conditions.
- Inject a synthetic exoplanet signal with a predetermined area ratio ϵ into the time series of spectra and try to recover it (e.g., Ref. [119]).

Appendices

A

Simulation of nonlinear pulse propagation in optical fibers

A.1 Generalized nonlinear Schödinger equation

The generalized nonlinear Schödinger equation (GNLSE) is used to model pulse propagation in fibers. In the time domain, it can be written as

$$\frac{\partial}{\partial z}A\left(z,t\right)=\mathrm{i}\sum_{k\geq 2}\frac{\mathrm{i}^{k}\beta_{k}}{k!}\frac{\partial^{k}}{\partial t^{k}}A\left(z,t\right)+\mathrm{i}\gamma\left(1+\frac{\mathrm{i}}{\omega_{0}}\frac{\partial}{\partial t}\right)\left[A\left(z,t\right)\int_{-\infty}^{\infty}R\left(\tau\right)\left|A\left(z,t-\tau\right)\right|^{2}\,\mathrm{d}\tau\right], \quad \text{(A.1)}$$

where $A\left(z,t\right)$ is the pulse (electric field) envelope, $\beta\left(\omega\right)$ is the propagation constant, and $\beta_{k}=\left.\frac{\partial^{k}\beta(\omega)}{\partial\omega^{k}}\right|_{\omega=\omega_{0}}$ are the dispersion coefficients at the pulse carrier frequency ω_{0}. $\gamma=\frac{n_{2}\omega_{0}}{cA_{\mathrm{eff}}}$ is the nonlinear coefficient, with n_{2} being the nonlinear refractive index and A_{eff} is the effective area of the propagating mode. The time t given here is actually measured in a comoving frame, i.e. $t=t_{r}-\beta_{1}z$ where t_{r} is the time in the rest frame. Since $\beta_{1}=1/v_{g}$, the frame moves with the pulse at the group velocity, which is why dispersion is of importance at only 2^{nd}-order or higher. $R\left(t\right)$ is the Raman response function given as

$$R\left(t\right)=\left(1-f_{r}\right)\delta\left(t\right)+f_{r}h_{r}\left(t\right)\Theta\left(t\right), \quad \text{(A.2)}$$

where f_r is the fractional contribution of delayed Raman response. $h_r(t) = \frac{\tau_1^2+\tau_2^2}{\tau_1\tau_2^2}\exp(-t/\tau_2)\sin(t/\tau_1)$. $\Theta(t)$ is the Heaviside step function. Substituting equation A.2 into equation A.1 gives:

$$\frac{\partial}{\partial z}A(z,t) = \mathrm{i}\sum_{k\geq 2}\frac{\mathrm{i}^k\beta_k}{k!}\frac{\partial^k}{\partial t^k}A(z,t) + \\ \mathrm{i}\gamma\left(1+\frac{\mathrm{i}}{\omega_0}\frac{\partial}{\partial t}\right)\left[(1-f_r)A|A|^2 + f_r A\int_{-\infty}^{\infty}h_r(\tau)\Theta(\tau)|A(z,t-\tau)|^2\,\mathrm{d}\tau\right].$$

It is convenient to now define two operators

$$\begin{aligned}\mathcal{D} &= \mathrm{i}\sum_{k\geq 2}\frac{\mathrm{i}^k\beta_k}{k!}\frac{\partial^k}{\partial t^k},\\ \mathcal{N} &= \frac{\mathrm{i}\gamma}{A}\left(1+\frac{\mathrm{i}}{\omega_0}\frac{\partial}{\partial t}\right)\left[(1-f_r)A|A|^2 + f_r A\left(H_r * |A|^2\right)\right],\end{aligned}$$

where $\mathcal{D}$ denotes dispersion, and $\mathcal{N}$ denotes nonlinearity. $*$ indicates a convolution and $H_r(t) = h_r(t)\Theta(t)$. With these definitions, we can rewrite the GNLSE as

$$\frac{\partial}{\partial z}A = (\mathcal{D}+\mathcal{N})A. \tag{A.3}$$

A.2 Fourier representation of operators

Let $\mathcal{F}$ be an integral transform operator that has the following action:

$$\mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)\,\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t \equiv \tilde{f}(\omega),$$

where $\mathcal{F}$ is normally the inverse Fourier transform (`ifft` in MATLAB) as defined in most literature. However, we align with this nonstandard convention used by Dudley and coauthors in Ch. 3 of Ref. [45] so as to make interpretation of their `source code` transparent. Applying our transform to a time derivative gives

$$\mathcal{F}\{f'(t)\} = \int_{-\infty}^{\infty}\frac{\mathrm{d}f}{\mathrm{d}t}\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t = -\mathrm{i}\omega\tilde{f}(\omega),$$

where we have used integration by parts and the fact that $f(t) \to 0$ as $t \to \pm\infty$. Generalizing this gives

$$\mathcal{F}\left\{f^{(n)}(t)\right\} = (-\mathrm{i}\omega)^n \tilde{f}(\omega).$$

Applying the transform to our dispersion operator $\mathcal{D}$ then gives

$$\tilde{\mathcal{D}} = \mathcal{F}\left\{\mathrm{i}\sum_{k\geq 2}\frac{\mathrm{i}^k\beta_k}{k!}\frac{\partial^k}{\partial t^k}\right\} = \mathrm{i}\sum_{k\geq 2}\frac{\mathrm{i}^k\beta_k}{k!}(-\mathrm{i}\omega)^k = \mathrm{i}\sum_{k\geq 2}\beta_k\omega^k/k! \tag{A.4}$$

Note that ω is not the true frequency, but a frequency associated with the rate of change of the slowly varying envelope. In other words, the transform of the time-domain pulse envelope occupies a window of some characteristic width $[-\Delta\omega, \Delta\omega]$ in frequency space. However, the true pulse actually occupies a window $[\omega_0 - \Delta\omega, \omega_0 + \Delta\omega]$ in frequency space as it the envelope is modulated by the carrier frequency.

Similarly, applying the transform to the nonlinearity $\mathcal{N}$ gives

$$\begin{aligned}
\tilde{\mathcal{N}} &= \mathcal{F}\left\{\frac{\mathrm{i}\gamma}{A}\left(1+\frac{\mathrm{i}}{\omega_0}\frac{\partial}{\partial t}\right)\left[(1-f_r)A|A|^2 + f_r A\left(H_r * |A|^2\right)\right]\right\} \\
&= \frac{\mathrm{i}\gamma}{\tilde{A}}\left(1+\frac{\mathrm{i}}{\omega_0}(-\mathrm{i}\omega)\right)\left[(1-f_r)\mathcal{F}\left\{A|A|^2\right\} + f_r\mathcal{F}\left\{A\left(H_r * |A|^2\right)\right\}\right] \\
&= \frac{\mathrm{i}\gamma\Omega}{\tilde{A}\omega_0}\left[(1-f_r)\mathcal{F}\left\{A|A|^2\right\} + f_r\mathcal{F}\left\{A\mathcal{F}^{-1}\left\{\tilde{H}_r\cdot\mathcal{F}\left\{|A|^2\right\}\right\}\right\}\right] \\
&= \frac{\mathrm{i}\gamma\Omega}{\tilde{A}\omega_0}\mathcal{F}\left\{(1-f_r)A|A|^2 + f_r A\mathcal{F}^{-1}\left\{\tilde{H}_r\cdot\mathcal{F}\left\{|A|^2\right\}\right\}\right\},
\end{aligned} \tag{A.5}$$

where $\Omega = \omega + \omega_0$. We have used the convolution theorem to simplify the expression $\mathcal{F}\left(H_r * |A|^2\right)$.

A.3 Interaction picture

The dispersive component of our PDE is now very simple algebraically as it is just a polynomial of ω in the frequency domain. In order to separate the effect of dispersion from the nonlinearity, we

apply a transformation to our envelope and operators, i.e.

$$A_I = \mathrm{e}^{-(z-z')\mathcal{D}} A, \tag{A.6}$$

$$\mathcal{N}_I = \mathrm{e}^{-(z-z')\mathcal{D}} \mathcal{N} \mathrm{e}^{(z-z')\mathcal{D}}, \tag{A.7}$$

to move into the so-called *interaction picture* [138], indicated by subscript I. z' is the distance between the interaction and normal pictures. Differentiating equation A.6 and using equations A.3, A.7 gives

$$\begin{aligned}
\frac{\partial}{\partial z} A_I &= \frac{\partial}{\partial z} \left[\mathrm{e}^{-(z-z')\mathcal{D}} A \right] \\
&= -\mathcal{D} \mathrm{e}^{-(z-z')\mathcal{D}} A + \mathrm{e}^{-(z-z')\mathcal{D}} \frac{\partial}{\partial z} A \\
&= -\mathrm{e}^{-(z-z')\mathcal{D}} \mathcal{D} A + \mathrm{e}^{-(z-z')\mathcal{D}} \left(\mathcal{D} + \mathcal{N} \right) A \\
&= \mathrm{e}^{-(z-z')\mathcal{D}} \mathcal{N} A \qquad \text{(A.8)} \\
&= \mathrm{e}^{-(z-z')\mathcal{D}} \mathcal{N} \mathrm{e}^{(z-z')\mathcal{D}} A_I \\
&= \mathcal{N}_I A_I, \qquad \text{(A.9)}
\end{aligned}$$

where we have used the fact that $\mathcal{D}$ commutes with itself. Equation A.9 is a very simple differential equation, but the one that is actually more useful in practice is equation A.8. Applying the transform $\mathcal{F}$ to both sides of equation A.8 gives

$$\frac{\partial}{\partial z} \tilde{A}_I = \exp\left(-\tilde{\mathcal{D}} \Delta \right) \tilde{\mathcal{N}} \tilde{A}, \tag{A.10}$$

where $\Delta = z - z'$. Letting $z' = 0$, and substituting equation A.5 into equation A.10 gives

$$\frac{\partial}{\partial z} \tilde{A}_I = \mathrm{i}\gamma \frac{\Omega}{\omega_0} \exp\left(-\tilde{\mathcal{D}} z \right) \mathcal{F} \left\{ (1 - f_r) A |A|^2 + f_r A \mathcal{F}^{-1} \left\{ \tilde{H}_r \cdot \mathcal{F} \left\{ |A|^2 \right\} \right\} \right\}. \tag{A.11}$$

Notice that the $\tilde{A}$'s in the two equations cancel. Equation A.11 can be integrated with the 4^{th}-order (adaptive) Runge-Kutta method (e.g. `ode45` in MATLAB). The initial condition for the solver is

$\tilde{A}_I(0,\omega) = \tilde{A}(0,\omega) = \mathcal{F}\{A(0,t)\}$. At every point z_i, the integrator can take the value of $\tilde{A}_I(z_i,\omega)$ and convert it to $A(z_i,t)$ by computing $\mathcal{F}^{-1}\left\{\exp\left(\tilde{\mathcal{D}}z_i\right)\tilde{A}_I(z_i,\omega)\right\}$.

The output of the simulation is processed via the exact same transformation, i.e. $\tilde{A}(z,\omega) = \exp\left(\tilde{\mathcal{D}}z\right)\tilde{A}_I(z,\omega)$ and $A(z,t) = \mathcal{F}^{-1}\tilde{A}(z,\omega)$. Conversion to wavelength space is done by plotting $\tilde{A}$ versus $\lambda = 2\pi c/\Omega$, where $\Omega = \omega + \omega_0$. This permits tracking the behaviour of the pulse along the fiber in both the time and spectral domain.

A.4 Test cases

To check the supercontinuum simulations, we run a few test cases to make sure the code reproduces the expected spectra. The first case is that of Fig. 3 in Ref. [17]. The initial pulse is described by a hyperbolic secant packet, i.e.

$$A(0,t) = \sqrt{P_0}\mathrm{sech}(t/\tau),$$

where P_0 is the peak power of the pulse, and τ characterizes the width of the pulse. The intensity full width at half maximum $T_{\mathrm{FWHM}} = 2\ln\left(1+\sqrt{2}\right)\tau \approx 1.763\tau$. Often the pulse energy E is given instead of the peak power, so it is helpful to also remember the relation $P_0 = \ln\left(1+\sqrt{2}\right)E/T_{\mathrm{FWHM}} \approx 0.881E/T_{\mathrm{FWHM}}$.

The parameters for Fig. 3 in Ref. [17] are: $\lambda_0 = 835$ nm, $P_0 = 10$ kW, $\tau = 28.4$ fs, $\beta_2 = -11.830 \times 10^{-3}$ ps^2/m, $\beta_3 = 8.1038 \times 10^{-5}$ ps^3/m, $\beta_4 = -9.5205 \times 10^{-8}$ ps^4/m, $\beta_5 = 2.0737 \times 10^{-10}$ ps^5/m, $\beta_6 = -5.3943 \times 10^{-13}$ ps^6/m, $\beta_7 = 1.3486 \times 10^{-15}$ ps^7/m, $\beta_8 = -2.5495 \times 10^{-18}$ ps^8/m, $\beta_9 = 3.0524 \times 10^{-21}$ ps^9/m, $\beta_{10} = -1.7140 \times 10^{-24}$ ps^{10}/m, $\gamma = 0.11$ W^{-1}m^{-1}, $f_r = 0.18$, $\tau_1 = 12.2$ fs, $\tau_2 = 32.0$ fs. λ_0 is the pulse centre wavelength, β_i, $i = 2,\ldots,10$ are dispersion coefficients, γ is the nonlinear parameter, f_r is the fractional Raman contribution, and $\tau_{1,2}$ are characteristic timescales associated with the Raman processes. P_0 and τ are as defined above.

The comparison of spectra obtained from our simulations and those calculated by Dudley *et al.* in

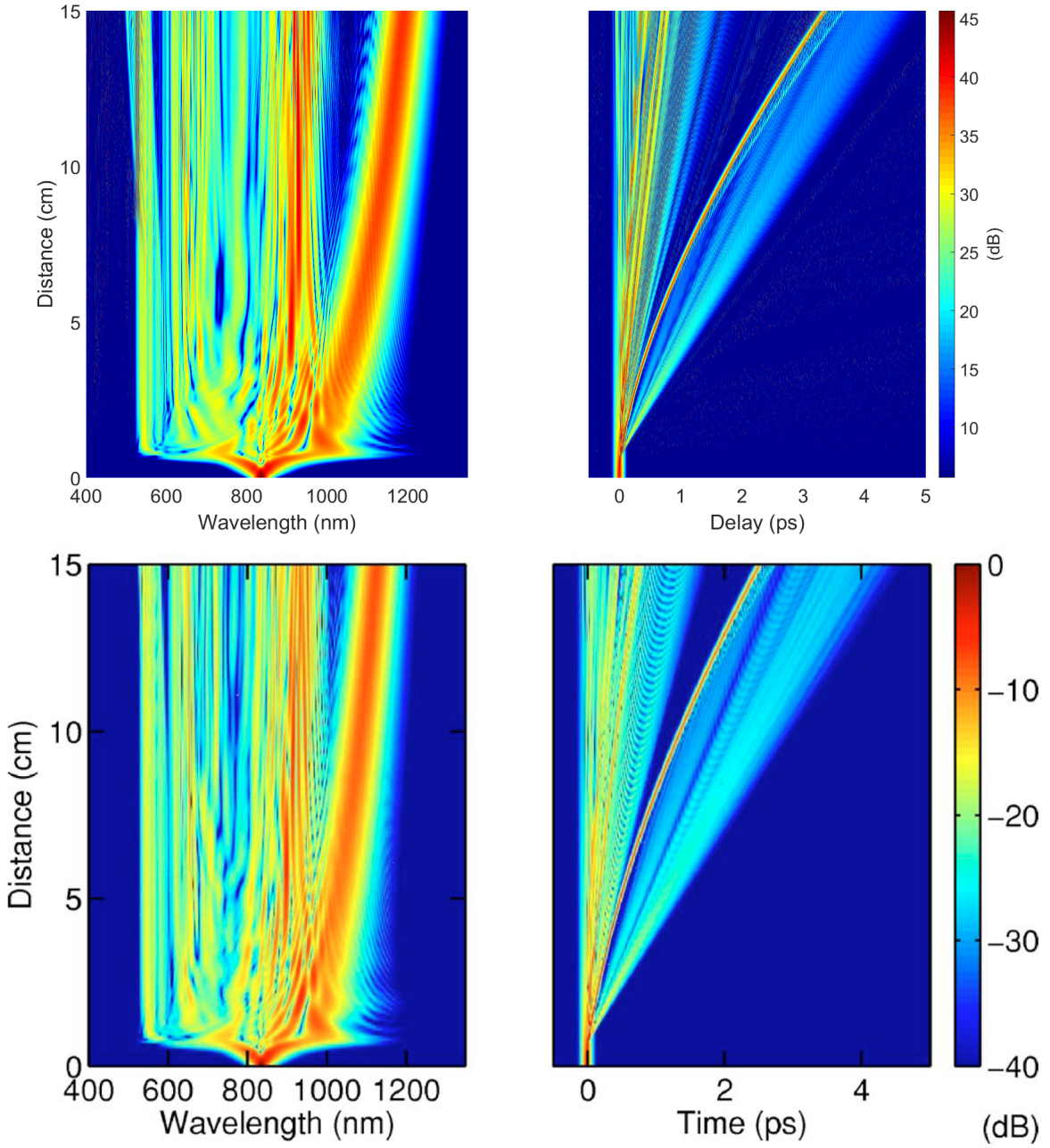

Figure A.1: Comparison of GNLSE simulation (top row) vs. Figure 3 of Ref. [17] (bottom row). The left column shows the spectral evolution vs. length and the right column shows temporal evolution vs. length. 0 cm on the length axis indicates the input of the fiber, so the light is launched at the bottom of the figure and propagates upwards.

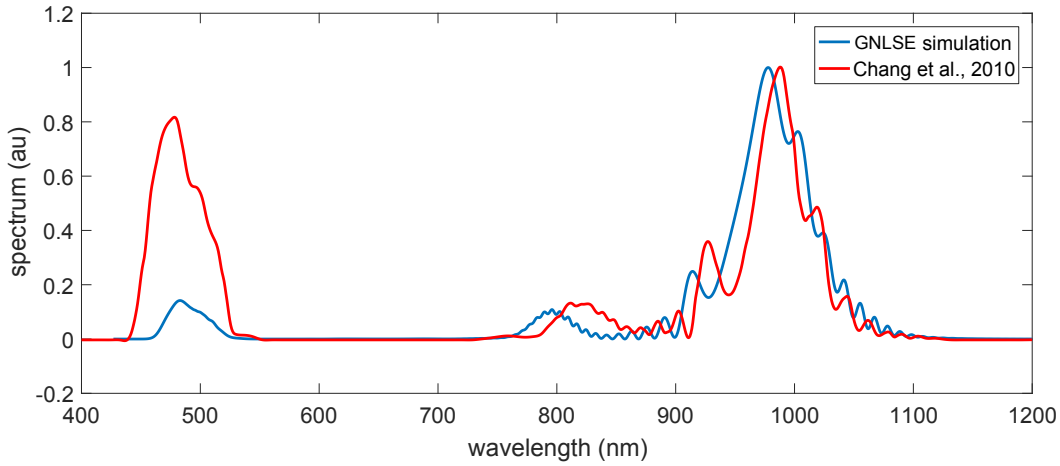

Figure A.2: Comparison of GNLSE simulation vs. Figure 1a of Chang *et al.*, *Opt. Lett.* 35, 2361 (2010)

Ref. [17] is shown in Figure A.1. There are slight differences in the results, probably attributable to the inclusion of a frequency-dependence in the nonlinear parameter as well as a slight difference in the definition of the Raman response in Ref. [17] .

The second test case is Fig. 1(a) of Ref. [47]. In this paper, Chang and coworkers demonstrate efficient Cherenkov radiation generation from short lengths of photonic crystal fiber pumped by few-cycle lasers. This type of work ultimately led to the design of the PCF used in the first-generation astro-comb. In this work, the parameters are: $\lambda_0 = 800$ nm, $E = 300$ pJ, $T_{\mathrm{FWHM}} = 10$ fs, $\beta_2 = -23.851 \times 10^{-3}$ ps^2/m, $\beta_3 = 9.0377 \times 10^{-5}$ ps^3/m, $\beta_4 = -9.4016 \times 10^{-8}$ ps^4/m, $\beta_5 = 1.4476 \times 10^{-10}$ ps^5/m, $\beta_6 = -2.0436 \times 10^{-13}$ ps^6/m, $\beta_7 = 1.6707 \times 10^{-16}$ ps^7/m, $\beta_8 = -6.082 \times 10^{-20}$ ps^8/m, $\beta_9 = 1.8973 \times 10^{-22}$ ps^9/m, $\gamma = 0.139$ W^{-1}m^{-1}, $L = 2$ cm. L is the length of the fiber. The Raman parameters are taken to be same as before, and dispersion coefficients come from fitting a polynomial to the $D(\lambda)$ data given in Fig. 1(a). The nonlinear parameter was obtained from a datasheet for NL-1.8-710 fiber (the fiber used in the paper). The results of the calculation are shown in Fig. A.2. As is visible, the red-shifted features (Raman solitons) are reproduced relatively nicely, but the transfer of power into the blue-shifted Cherenkov radiation band is much lower in our computation compared to the quoted 40% in Ref. [47].

A.5 Updated model

In the final simulations, we used a more sophisticated GNLSE model which incorporates the frequency-dependence of the effective area as well [45], i.e.,

$$\begin{aligned}\frac{\partial}{\partial z}\tilde{A}_I &= \mathrm{i}\bar{\gamma}(\omega)\exp\left(-\mathcal{D}(\omega)z\right) \\ &\quad\times\mathcal{F}\left\{\bar{A}(z,T)\times\left(R(T)*\left|\bar{A}(z,T)\right|^2\right)\right\}. \end{aligned} \tag{A.12}$$

All symbols used above in are defined in in Table A.1.

In terms of the GNLSE, the PCF is entirely described by $\mathcal{D}(\omega)$, $\bar{\gamma}(\omega)$ and $R(t)$. The fiber material is fused silica, so we take $f_r = 0.18$, $\tau_1 = 12.2$ fs, and $\tau_2 = 32$ fs as given in Ref. [139], and $n_2 \approx 2.5\times10^{-20}$ m^2/W (Crystal Fibre A/S, NL-2.8-850-02 datasheet). for all our calculations. We neglect any losses, i.e., $\alpha(\omega) = 0$ dB/m.

Note that in tapered geometries, both the dispersion operator and the frequency-dependent nonlinear parameter become functions of z as well, i.e., $\mathcal{D}(\omega)\to\mathcal{D}(\omega,z)$ and $\bar{\gamma}(\omega)\to\bar{\gamma}(\omega,z)$ [140, 141]. This approach has been pointed out to not be strictly correct as it does not conserve the photon number [43, 142], but we have adopted it here for simplicity. Our solver codes and data are available online.

Table A.1: Definitions for GNLSE model

Symbol	Definition
$*$	convolution
$\mathcal{F}\{f(T)\}$	$= \int_{-\infty}^{\infty} f(T)\, \mathrm{e}^{\mathrm{i}(\omega-\omega_0)T} \mathrm{d}T$, Fourier operator
$\mathcal{F}^{-1}\{g(\omega-\omega_0)\}$	$= (2\pi)^{-1} \int_{-\infty}^{\infty} g(\omega-\omega_0)\, \mathrm{e}^{-\mathrm{i}(\omega-\omega_0)T} \mathrm{d}\omega$, inverse Fourier operator
z	longitudinal coordinate
ω	angular frequency
t	time
ω_0	reference frequency (set to center of computational window)
ω_c	pulse carrier frequency
$\beta(\omega)$	$= n_{\mathrm{eff}}(\omega) \times \omega/c$, propagation constant
$n_{\mathrm{eff}}(\omega)$	effective index
c	speed of light in vacuum
$\beta_n(\omega)$	$= \partial^n \beta / \partial \omega^n$, n^{th}-order dispersion
T	$= t - \beta_1(\omega_c)\, z$, time in comoving frame
$\tilde{A}(z, \omega-\omega_0)$	spectral envelope of pulse
$A(z,T)$	$= \mathcal{F}^{-1}\left\{\tilde{A}(z,\omega-\omega_0)\right\}$, time-domain envelope of pulse
$\mathcal{D}(\omega)$	$= \mathrm{i}[\beta(\omega) - \beta(\omega_0) - \beta_1(\omega_c)(\omega-\omega_0)] - \frac{1}{2}\alpha(\omega)$, dispersion operator
$\alpha(\omega)$	frequency-dependent loss
$\tilde{A}_I(z, \omega-\omega_0)$	$= \mathrm{e}^{-\mathcal{D}(\omega)z} \tilde{A}(z, \omega-\omega_0)$, interaction picture spectral envelope of pulse
$\bar{\gamma}(\omega)$	$= \frac{n_2 n_{\mathrm{eff}}(\omega_0)\omega}{c n_{\mathrm{eff}}(\omega) A_{\mathrm{eff}}^{1/4}(\omega)}$, frequency-dependent nonlinear parameter
n_2	nonlinear refractive index
$A_{\mathrm{eff}}(\omega)$	$= \frac{\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lvert F(x,y,\omega)\rvert^2 dxdy\right)^2}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lvert F(x,y,\omega)\rvert^4 dxdy}$, frequency-dependent mode effective area
$F(x,y,\omega)$	transverse modal distribution
$\bar{A}(z,T)$	$= \mathcal{F}^{-1}\left\{\tilde{A}(z,\omega-\omega_0)/A_{\mathrm{eff}}^{1/4}(\omega)\right\}$
$R(t)$	$= (1-f_r)\,\delta(t) + f_r h_r(t)\,\Theta(t)$, Raman response function
f_r	fractional contribution of delayed Raman response
$h_r(t)$	$= \frac{\tau_1^2+\tau_2^2}{\tau_1\tau_2^2} \exp(-t/\tau_2) \sin(t/\tau_1)$, with Raman fit parameters τ_1, τ_2
$\Theta(t)$	Heaviside function

A.5.1 Calculation of supercontinuum coherence

We use a first-order measure $g_{12}^{(1)}$ to evaluate the coherence [17] of the output optical field $\tilde{A}(\omega)$,

$$\left|g_{12}^{(1)}(\omega)\right| = \frac{\left|\left\langle \tilde{A}_i^*(\omega)\,\tilde{A}_j(\omega)\right\rangle_{i\neq j}\right|}{\sqrt{\left\langle\left|\tilde{A}_i(\omega)\right|^2\right\rangle\left\langle\left|\tilde{A}_j(\omega)\right|^2\right\rangle}} \tag{A.13}$$

where $\langle\cdots\rangle$ is an ensemble average over N propagations of the simulation.

Each run of the simulation differs by some noise injected into the input field. We include only a shot noise seed and no spontaneous Raman noise, as shot noise has been shown to be the dominant noise process [143]. To perturb the input pulse (in the time domain), we follow Ref. [144, 145]: to each temporal bin of both the real and imaginary components of $A(T)$, we add a random number drawn from a normal distribution with zero mean and variance $\hbar\omega_c/(4\Delta T)$, where ω_c is the pulse carrier frequency and ΔT is the temporal bin width. In Figure 1.11c, we evaluate $\left|g_{12}^{(1)}\right|$ over $N = 100$ runs of the simulation.

A.6 Useful literature resources

For more in-depth explanations regarding the GNLSE, please consult the following texts:

- Christian Skovmøller Agger, *Infrared Supercontinuum Generation in Soft-glass Fibers*, PhD thesis, Technical University of Denmark, 2013
- Simon Toft Sørensen, *Deep-blue supercontinuum light sources based on tapered photonic crystal fibres*, PhD thesis, Technical University of Denmark, 2013

B

Calculation of optical properties of photonic crystal fibers

B.1 Glass rod model of a photonic crystal fiber

Now that we have developed the tools for simulating pulse propagation in Appendix A, we focus next on the calculation of fiber parameters such as dispersion and nonlinearity, which are necessary inputs for the propagation calculation. The geometry of the fiber (i.e. core size, air fill fraction, etc.) directly governs the dispersion and the nonlinear parameter as it defines the mode of propagation radiation. One degree of freedom, as far as choosing dispersion characteristics and nonlinearity, is the fiber type itself. Nonlinear fibers are often characterized by their core size and zero dispersion wavelength (ZDW, when $\partial^2\beta(\omega)/\partial\omega^2 = 0$). However, another (often overlooked) degree of freedom is the core size as a function of distance. Changing the core size over some length by heating and pulling is termed *tapering*. Tapering allows for micromanaging the dispersion and nonlinear characteristics of the fiber as well as ease of coupling. One can start with a large core fiber on the input face and draw it down to a small diameter to exploit confinement of the light to generate a supercontinuum. Here, the dispersion coefficients (and the zero dispersion wavelength) as well as the nonlinear parameter are changing along the length of the fiber, so we need to understand how to calculate these properties as a function of core radius.

Let's begin with the simplest model of a high air-fill fraction PCF. This structure with a glass core and a honeycomb of air surrounding it may be approximated as as a glass rod in air (also known as silica strand model in literature) because the the width of struts holding the core is negligible.

B.1.1 Dispersion vs. core radius

Simulation of tapered PCFs involves taking into account the changes in both dispersion and non-linearity as a function of core radius. This section will discuss the former. Let n_{cl} be the cladding refractive index (= 1 for air) and n_{co} be the core refractive index ($n_{\mathrm{co}}(\lambda)$ is given by the Sellmeier equation for fused silica).

With cylindrical symmetry of the problem, the vector wave equations can be approximated as an eigenvalue equation [146] describing the modes in the fiber:

$$\left(\frac{J_m'(\kappa a)}{\kappa J_m(\kappa a)}+\frac{K_m'(\gamma a)}{\gamma K_m(\gamma a)}\right)\left(\frac{J_m'(\kappa a)}{\kappa J_m(\kappa a)}+\frac{K_m'(\gamma a)}{\gamma K_m(\gamma a)}\frac{n_{\mathrm{cl}}^2}{n_{\mathrm{co}}^2}\right)=\left(\frac{m\beta k_0\left(n_{\mathrm{co}}^2-n_{\mathrm{cl}}^2\right)}{an_2\kappa^2\gamma^2}\right)^2,$$

where $k_0 = 2\pi/\lambda$ is the vacuum wavenumber, a is the core radius, $\kappa = \sqrt{n_{\mathrm{co}}^2 k_0^2 - \beta^2}$, and $\gamma = \sqrt{\beta^2 - n_{\mathrm{cl}}^2 k_0^2}$. m is the mode index (azimuthal). J_m and K_m are the m^{th}-order Bessel function of the first kind and modified Bessel function of the second kind, respectively. Primes $'$ denote derivatives. Both waveguide and material dispersion are taken into account in this equation because n_{co} is a function of λ via the Sellmeier equation.

Note the following two recurrence relations for Bessel functions:

$$J_m' = \frac{1}{2}\left(J_{m-1} - J_{m+1}\right),$$
$$K_m' = -\frac{1}{2}\left(K_{m-1} + K_{m+1}\right).$$

With this, we can then determine the eigenvalues β_{mp} numerically. Note that p labels the radial mode index. The lowest order mode is the hybrid mode HE_{11}, so we are interested in finding β_{11}. We know that $n_{\mathrm{cl}}k_0 \le \beta \le n_{\mathrm{co}}k_0$.

Another way of saying this is that the effective index n_{eff} of the mode is in between n_{cl} and n_{co} since $\beta = 2\pi/k_0 \times n_{\text{eff}}$. Since the index step between glass and air is very large, the fiber is strongly guiding. We then expect the effective index to be very close to that of the core, n_{co}.

Thus, we recast our problem as a numerical root-finding problem (i.e. LHS $-$ RHS $= 0$ for the eigenvalue equation) and search starting from $n_{\text{co}}k_0$ downward, and the first zero-crossing is then β for our mode HE_{11}. If this is done at regular frequency spacing $\{\omega_i\}$ (instead of wavelength spacing), we can numerically evaluate the second derivative as follows: $\beta_2(\omega_i) \approx \frac{\omega_{i-1}+2\omega_i+\omega_{i+1}}{(\Delta\omega)^2}$, where we have used a centred finite-difference approximation.

Note that using $\beta_2(\omega)$ is a simpler and more accurate way of modelling dispersion than a series expansion at the carrier frequency ω_0 with coefficients β_k.

At any rate, we are now interested in verifying if this procedure gives the correct results. We prepare a regular wavelength grid $\{\lambda_i\}$ and find β for each value of λ, as described above. we then convert β to n_{eff} using the relation $\beta = 2\pi/k_0 \times n_{\text{eff}}$. Then, the dispersion, D is evaluated by computing

$$D = -\frac{\lambda}{c}\frac{d^2 n_{\text{eff}}}{d\lambda^2}.$$

An example dispersion calculation is shown in Figure B.1, directly comparable to Fig. 2 of Ref. [147].

B.1.2 Nonlinearity vs. core radius

Having developed a simple dispersion model, we move on to a discussion of nonlinearity as a function of core radius. For this analysis, we will assume fixed indices for core $n_{\text{co}} = 1$ and cladding $n_{\text{cl}} = 1.45$. We can take the wavelength to be $\lambda = 800$ nm, but it is technically irrelevant as the final plot we will make has nondimensionalized axes. Let a be the core radius.

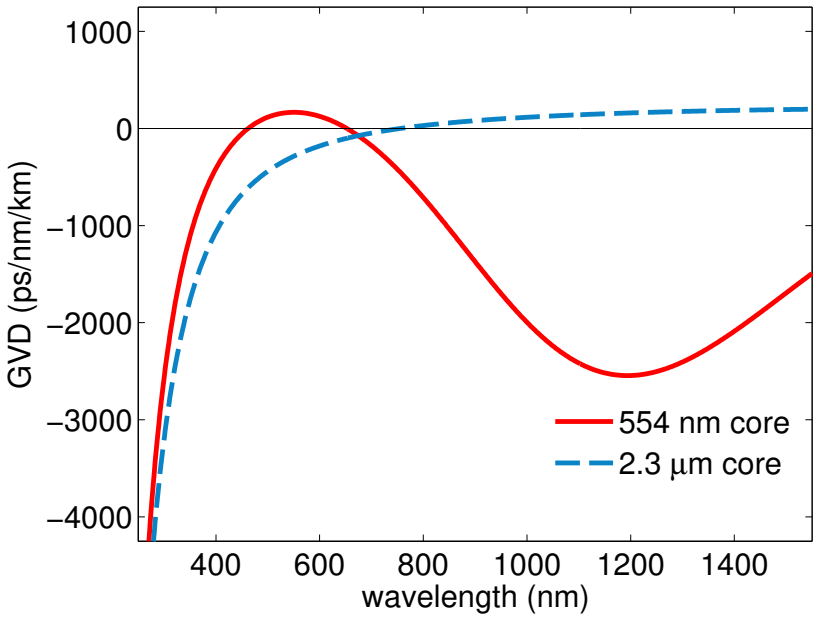

Figure B.1: Calculated dispersion for 554 nm core and 2.3 µm core silica rods

The nonlinear parameter of γ of a waveguide [148] is defined as

$$\gamma = \frac{2\pi}{\lambda} n_2 \frac{\int_{\text{core}} S_z^2 dA}{\left(\int S_z dA\right)^2},$$

where S_z is the longitundinal component of the Poynting vector and n_2 is the nonlinear refractive index of the core material. $dA = rdrd\theta$ in cylindrical coordinates. Note that the integral in the denominator is computed over the whole plane. The integral in the numerator is computed only over the core since $n_2^{\text{co}} \gg n_2^{\text{cl}}$. To compute the quantity γ as a function of core radius a, we need expressions for S_z in the mode HE_{11}. Note the following conventions:

$$
\begin{aligned}
&&&a_1 = \frac{1}{2}\left(F_2 - 1\right),\\
\Delta = \frac{1}{2}\left(1 - \frac{n_{\text{co}}^2}{n_{\text{cl}}^2}\right), &&b_1 = \frac{J_0\left(U\right) - J_2\left(U\right)}{2UJ_1\left(U\right)}, &&a_2 = \frac{1}{2}\left(F_2 + 1\right),\\
\kappa = \sqrt{n_{\text{co}}^2 k_0^2 - \beta^2},\, U = \kappa a, &&b_2 = -\frac{K_0\left(W\right) + K_2\left(W\right)}{2WK_1\left(W\right)}, &&a_3 = \frac{1}{2}\left(F_1 - 1\right),\\
\gamma = \sqrt{\beta^2 - n_{\text{cl}}^2 k_0^2},\, W = \gamma a, &&F_1 = \left(UW/V\right)^2\left[b_1 + \left(1 - 2\Delta\right)b_2\right], &&a_4 = \frac{1}{2}\left(F_1 + 1\right),\\
V = k_0 a\sqrt{n_{\text{co}}^2 - n_{\text{cl}}^2}, &&F_2 = \left(V/UW\right)^2/\left(b_1 + b_2\right), &&a_5 = \frac{1}{2}\left(F_1 - 1 + 2\Delta\right),\\
&&&a_6 = \frac{1}{2}\left(F_1 + 1 - 2\Delta\right).
\end{aligned}
$$

With these, we can then write down the longitundinal component of the Poynting vector in the core and cladding [149].

$$
\begin{aligned}
S_z\left(0 < r < a, \phi\right) &\propto \frac{1}{J_1^2\left(U\right)}\left[a_1 a_3 J_0^2\left(UR\right) + a_2 a_4 J_2^2\left(UR\right) + \frac{1 - F_1F_2}{2} J_0\left(UR\right) J_2\left(UR\right)\cos 2\phi\right],\\
S_z\left(a \le r < \infty, \phi\right) &\propto \frac{U^2}{W^2 K_1^2\left(W\right)}[a_1 a_5 K_0^2\left(WR\right) + a_2 a_6 K_2^2\left(WR\right)\\
&\quad - \frac{1 - 2\Delta - F_1F_2}{2} K_0\left(WR\right) K_2\left(WR\right)\cos 2\phi]
\end{aligned}
$$

where $R = r/a$.

Now, it's just a matter of doing the integrals for each value of $\beta(a)$ as we vary the core radius a. β comes from solving the eigenvalue equation noted in the previous section. Note that $\cos 2\phi$ terms integrate to 0 over $\phi \in [0, 2\pi]$, so these terms are irrelevant for the computation of $\int S_z dA$. However, we must keep the angular term in the core when computing $\int S_z^2 dA$ since $\int_0^{2\pi} \cos^2 2\phi d\phi = \pi$. The integrals are numerically computed using the trapezoid rule, and the integrals are truncated at $10a$, where the evanescent field is negligible. A total of 10^5 evenly spaced points are used to span this interval in the radial dimension. The result is shown in Figure B.2.

Note that $\frac{\int_{\text{core}} S_z^2 dA}{\left(\int S_z dA\right)^2}$ can be thought of as an inverse effective area $1/A_{\text{eff}}$. One can ask whether A_{eff}

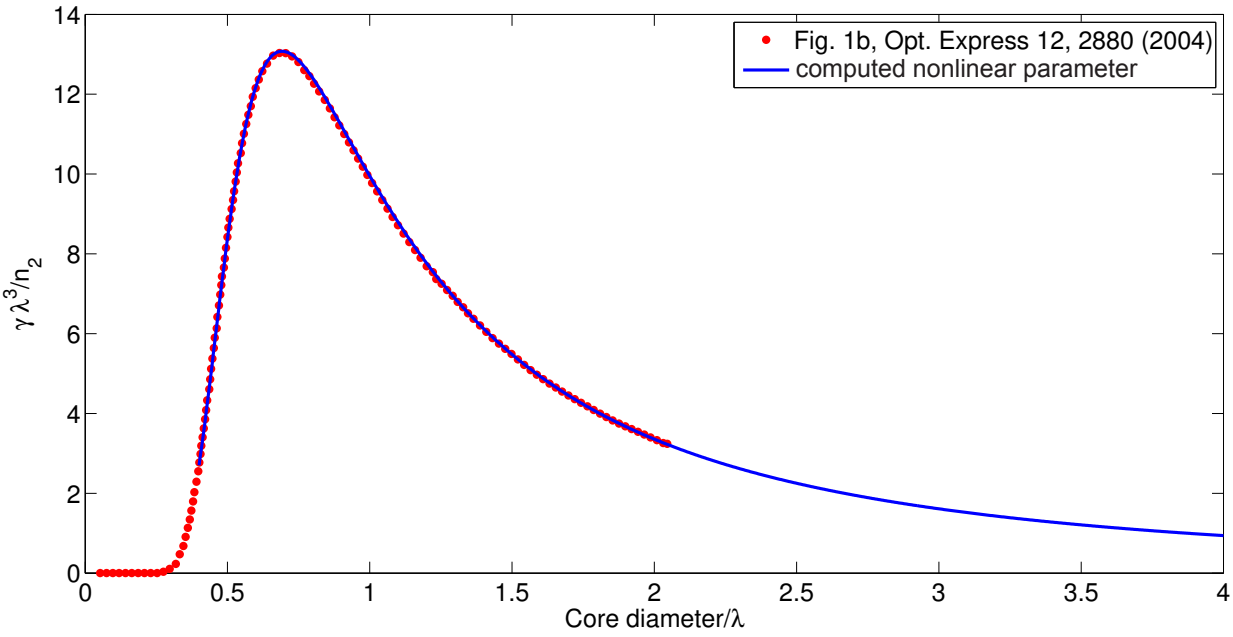

Figure B.2: Calculated nondimensionalized nonlinear parameter vs nondimensionalized core diameter for glass rod model, compared with result in Ref. [147]

depends on the frequency ω since the mode in the fiber changes with ω. There is indeed a dependence on ω, and this correction (i.e. a frequency-dependent effective area) is discussed in detail in Ref. [150]. However, the theory shows that "*the characteristics of SC generated using near-infrared femtosecond pump pulses around 800 nm show negligible dependence on the frequency-dependent effective area over a wide parameter range typical of that encountered in previous experiments*" [150].

B.2 Finite-difference model of a photonic crystal fiber

To fully take into account the microstructured nature of the photonic crystal fibers in our dispersion and nonlinearity models, we need to solve Maxwell's equations for the cross-section of the fiber. We do this using a commercial finite-difference mode solver (Lumerical MODE Solutions). To calculate the quantities of interest, we first solve for the modes of the geometry. The dispersion is evaluated directly in MODE by selecting the fundamental mode and and tracking it through a frequency sweep.

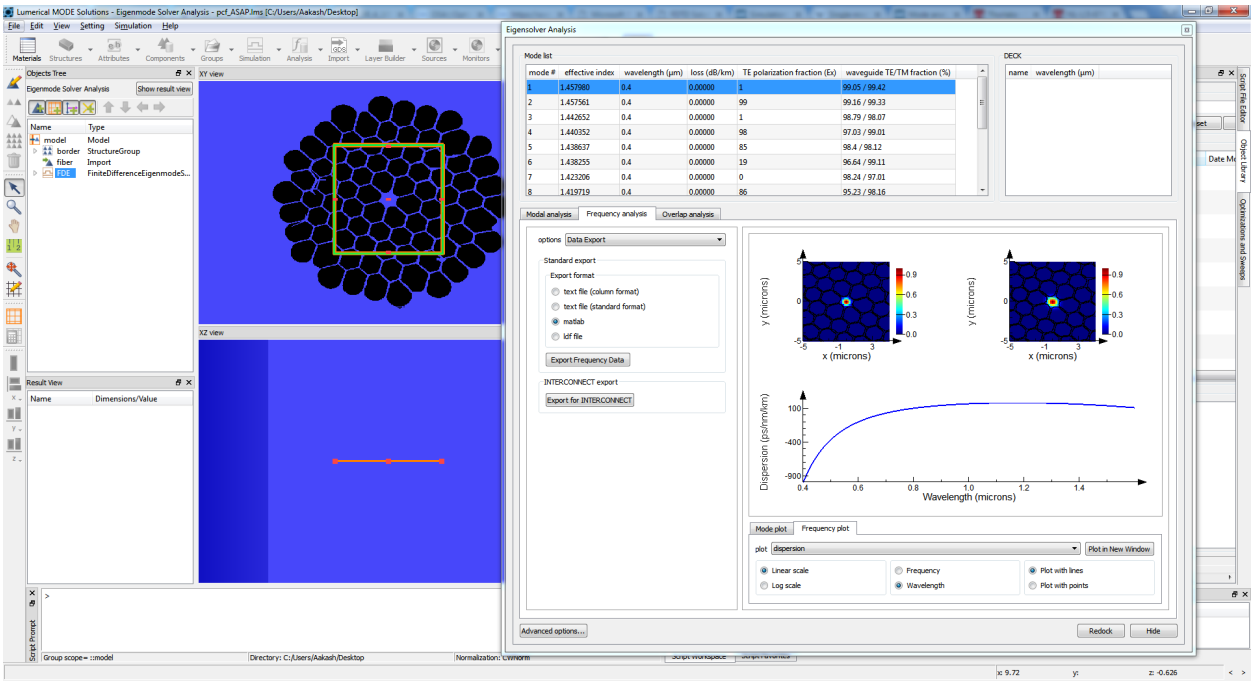

Figure B.3: MODE Solutions Eigensolver Analysis for NL-1.5-670 fiber

MODE has the capability to import geometry from a SEM micrograph of the fiber end face. This is a very convenient feature. The snapshot in Figure B.3 shows the package in action.

Calculating the dispersion for a variety of fibers and comparing with our simple glass rod model described in Section B.1 as well as the datasheet yields the result shown in Figure B.4.

The measured dispersion is always sandwiched between the two models, with the Lumerical simulation being biased toward lower dispersion and the simple glass-rod-in-air model being being biased toward higher dispersion. Lumerical generally has better agreement with the datasheet values, especially for small cores. However, the discrepancy is not very large.

Next, we calculate the nonlinear parameter using the exported electric field from the MODE software. To compare with the datasheet values of the nonlinear parameter, we numerically compute

$$A_{\text{eff}} = \frac{\left(\int\int |E(x,y)|^2 dxdy\right)^2}{\int\int |E(x.y)|^4 dxdy}.$$

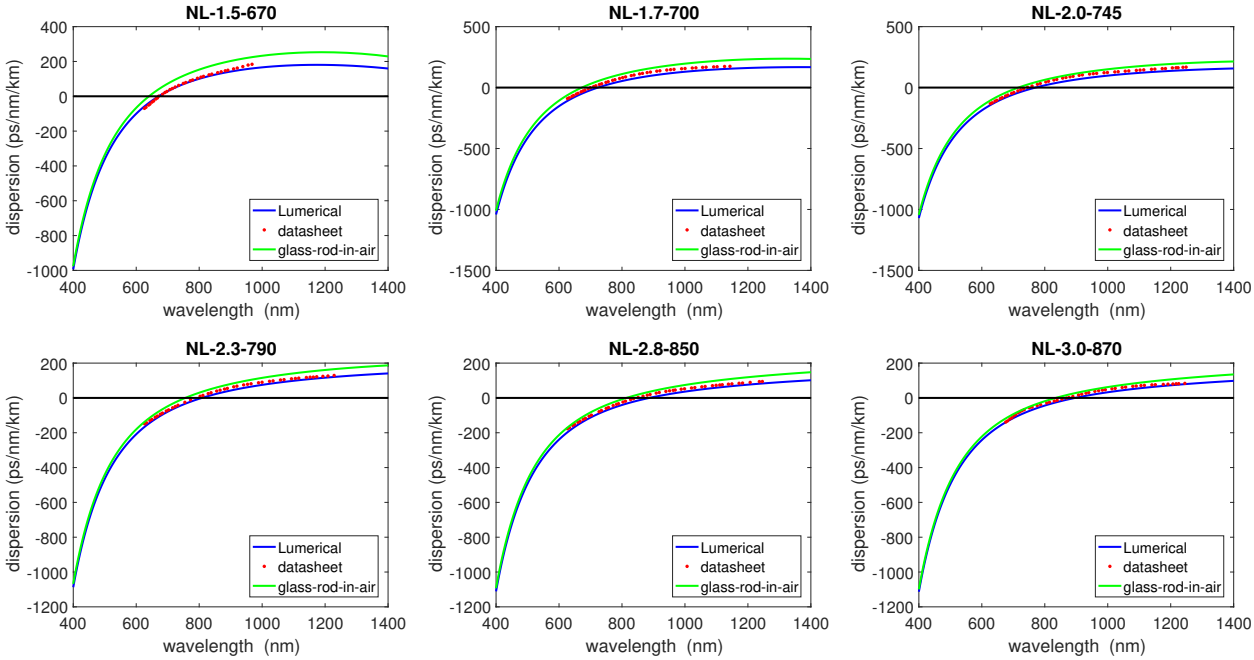

Figure B.4: Calculated dispersion of various nonlinear fibers from NKT photonics. NL-A-B indicates a nonlinear fiber with a core diameter A (in µm) and a zero dispersion wavelength B (in nm). Finite-difference calculations using MODE Solutions shown as blue curves, compared with glass-rod-in-air approximation in green as well as the measured dispersion quoted on the datasheet as red points.

We can then obtain the nonlinear parameter via $\gamma = 2\pi/\lambda_{\text{ZDW}} \times n_2/A_{\text{eff}}$ since the datasheet nonlinear parameters are quoted at the zero-dispersion wavelengths (ZDW). A comparison is shown in Table B.1.

Lumerical typically underestimates the nonlinearity (except at very small core sizes) and the simple glass-rod-in-air model typically overestimates it. In addition, Lumerical typically gets closer, but it's not a very dramatic competitive advantage.

Fiber	**Datasheet**	**Lumerical**	**glass-rod-in-air**
NL-1.5-670	190	216	229
NL-1.7-700	148	139	173
NL-2.0-745	104	94	120
NL-2.3-790	75	68	86
NL-2.8-850	47	38	55
NL-3.0-870	42	34	47

Table B.1: Nonlinear parameter of various nonlinear fibers from NKT photonics. γ is in units of $\text{W}^{-1}\text{km}^{-1}$

C

Time series analysis using Gaussian processes

The following text, written by N. Langellier for the Supplemental Material of Ref. [151], is reproduced here for completeness.

In order to test the feasibility of a stellar acceleration survey campaign with present-day technology and analysis techniques, RV time series are generated with correlated noise as described above. In order to model the correlated noise, a Gaussian process (GP) regression is employed [104, 107, 108]. A GP assumes the data are normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K}$. This allows for a simultaneous fit of any deterministic physics with a model of the correlated noise.

For this analysis, the mean function models both the stellar acceleration and M Keplerian signals (stellar companions and planets):

$$\mu(t_i) = \Delta a_r t_i + v_{\text{off}} + \sum_{j=1}^{M} A_j \sin\left(\frac{2\pi t_i}{T_j} + \phi_j\right), \tag{C.1}$$

where the relative stellar acceleration is Δa_r and the offset velocity is v_{off}. Each orbiting body

is assumed to have zero eccentricity so A_j, T_j, and ϕ_j are the RV amplitude, orbital period, and orbital phase offset for the j^{th} orbiting body, respectively. The number of Keplerian signals, M, is chosen by assuming we would have knowledge of some (but not necessarily all) orbiting bodies for a given star. For our fits, we look at each orbiting body injected into the simulation and include it in Eq. (C.1) if the RV amplitude is greater than 1 m/s and the orbital period is less than 10^4 days. This corresponds to an optimistic cutoff where an orbiting body at the limit of state of the art techniques ($\sim$ 1 m/s) has an orbital period almost three times the observing campaign length. Such a body would have a change in RV over the observing campaign comparable to the injected instrumental white noise ($\sim$ 60 cm/s). Keeping all orbiting bodies that meet these criteria provide a best case scenario wherein the feasibility of this analysis is easily tested.

The covariance matrix $\mathbf{K}$ is populated using the quasiperiodic kernel function given in Eq. (2.17) with an additive white noise term on the main diagonal:

$$K_{i,j}(t_i, t_j) = K_{\text{QP}}(t_i, t_j) + \sigma^2_{\text{WN}}\delta(t_i, t_j). \tag{C.2}$$

The GP model thus contains a total of $7 + 3M$ fit parameters,

$$\boldsymbol{\theta}_{\text{GP}} = \langle \Delta a_r, v_{\text{off}}, \{A_j\}, \{T_j\}, \{\phi_j\}, h, T_r, w, \lambda, \sigma_{\text{WN}} \rangle \tag{C.3}$$

with a likelihood function given by

$$\mathcal{L}\left(\boldsymbol{\theta}_{\text{GP}} | \mathbf{v}, \boldsymbol{t}\right) = \frac{1}{\sqrt{|2\pi\mathbf{K}|}} \mathrm{e}^{-\frac{1}{2}(\mathbf{v}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{v}-\boldsymbol{\mu})}, \tag{C.4}$$

where $\mathbf{v}$ is a vector of the measured RVs at times $\boldsymbol{t}$, and $\boldsymbol{\mu}$ and $\mathbf{K}$ are functions of $\boldsymbol{t}$ and $\boldsymbol{\theta}_{\text{GP}}$ as described in Equations (C.1) and (C.2). The relative stellar acceleration, Δa_r, is given a Gaussian prior with a mean equal to the injected stellar acceleration signal and a standard

deviation of 30 times the mean, i.e.,

$$P(\Delta a_r) \sim \mathcal{N}\left(1.5 \times 10^{-8}\ \mathrm{cm/s^2},\ 45 \times 10^{-8}\ \mathrm{cm/s^2}\right). \tag{C.5}$$

This prior is chosen to include the expected value of the relative stellar acceleration and reasonable deviations about this value while rejecting accelerations from long-period orbiting bodies that are degenerate with the stellar acceleration signal on the timescale of a decade. The remaining model parameters are allowed to vary freely.

The maximum a posteriori is taken as the parameter estimate, $\hat{\boldsymbol{\theta}}_{\mathrm{GP}}$. Minimization of the negative log posterior yields the desired result:

$$\hat{\boldsymbol{\theta}}_{\mathrm{GP}} = \arg\min_{\theta_{\mathrm{GP}}} \left[-\ln \mathcal{L}\left(\boldsymbol{\theta}_{\mathrm{GP}} | \mathbf{v}, \boldsymbol{t}\right) - \ln P\left(\Delta a_r\right)\right]. \tag{C.6}$$

Minimization is performed using the quasi-Newton method of Broyden, Fletcher, Goldfarb, and Shanno [152]. Initial guesses for the model parameters, $\boldsymbol{\theta}_{\mathrm{GP},0}$, were chosen assuming we would have knowledge of their true values, $\boldsymbol{\theta}_{\mathrm{GP,\ true}}$. The values were normally distributed about their true value with a standard deviation of about 3% of the mean, i.e.,

$$P(\boldsymbol{\theta}_{\mathrm{GP},0}) \sim \mathcal{N}\left(\boldsymbol{\theta}_{\mathrm{GP,\ true}},\ \frac{1}{30}\boldsymbol{\theta}_{\mathrm{GP,\ true}}\right). \tag{C.7}$$

The Python [153] packages NumPy [154], SciPy [155], and George [156] were used for this analysis.

Bibliography

[1] J. T. Wright, O. Fakhouri, G. W. Marcy, E. Han, Y. Feng, John Asher Johnson, A. W. Howard, D. A. Fischer, J. A. Valenti, J. Anderson, and N. Piskunov, "The Exoplanet Orbit Database," Publications of the Astronomical Society of the Pacific **123**, 412–422 (2011).

[2] Eunkyu Han, Sharon X. Wang, Jason T. Wright, Y. Katherina Feng, Ming Zhao, Jacob I. Brown, and Colin Hancock, "Exoplanet Orbit Database. II. Updates to Exoplanets.org," Publications of the Astronomical Society of the Pacific **126**, 827–837 (2014).

[3] Michael A C Perryman, *The Exoplanet Handbook* (Cambridge University Press, 2011).

[4] X. Dumusque, "Radial velocity fitting challenge – I. Simulating the data set including realistic stellar radial-velocity signals," Astron. Astrophys. **593**, A5 (2016).

[5] X. Dumusque, F. Borsa, M. Damasso, R. F. Díaz, P. C. Gregory, N. C. Hara, A. Hatzes, V. Rajpaul, M. Tuomi, S. Aigrain, G. Anglada-Escudé, A. S. Bonomo, G. Boué, F. Dauvergne, G. Frustagli, P. Giacobbe, R. D. Haywood, H. R. A. Jones, J. Laskar, M. Pinamonti, E. Poretti, M. Rainer, D. Ségransan, A. Sozzetti, and S. Udry, "Radial-velocity fitting challenge – II. First results of the analysis of the data set," Astronomy & Astrophysics **598**, A133 (2017).

[6] T. W. Milbourne *et al.*, "HARPS-N Solar RVs Are Dominated by Large, Bright Magnetic Regions," The Astrophysical Journal **874**, 107 (2019).

[7] Jun Ye and Steven T. Cundiff, eds., *Femtosecond Optical Frequency Comb: Principle, Operation, and Applications* (Kluwer Academic Publishers, 2004).

[8] Debra A. Fischer, Guillem Anglada-Escude, Pamela Arriagada, Roman V. Baluev, Jacob L. Bean, Francois Bouchy, Lars A. Buchhave, Thorsten Carroll, Abhijit Chakraborty, Justin R. Crepp, *et al.*, "State of the Field: Extreme Precision Radial Velocities," Publ. Astron. Soc. Pac. **128**, 066001 (2016).

[9] Chih-Hao Li, Andrew J. Benedick, Peter Fendel, Alexander G. Glenday, Franz X. Kärtner, David F. Phillips, Dimitar Sasselov, Andrew Szentgyorgyi, and Ronald L. Walsworth, "A laser frequency comb that enables radial velocity measurements with a precision of 1 cm s-1," Nature **452**, 610–612 (2008).

[10] Tilo Steinmetz, Tobias Wilken, Constanza Araujo-Hauck, Ronald Holzwarth, Theodor W. Hänsch, Luca Pasquini, Antonio Manescau, Sandro D'Odorico, Michael T. Murphy, Thomas Kentischer, *et al.*, "Laser Frequency Combs for Astronomical Observations," Science **321**, 1335–1337 (2008).

[11] D A Braje, M S Kirchner, S Osterman, T Fortier, and S A Diddams, "Astronomical spectrograph calibration with broad-spectrum frequency combs," Eur. Phys. J. D **48**, 57–66 (2008), doi: 10.1140/epjd/e2008-00099-9.

[12] M. Mayor *et al.*, "Setting New Standards with HARPS," The Messenger **114**, 20–24 (2003).

[13] Rosario Cosentino *et al.*, "HARPS-N: the new planet hunter at TNG," Proc. SPIE **8446**, 84461V (2012), doi: 10.1117/12.925738.

[14] Richard A. McCracken, Jake M. Charsley, and Derryck T. Reid, "A decade of astrocombs: recent advances in frequency combs for astronomy [Invited]," Opt. Express **25**, 15058–15078 (2017), doi: 10.1364/OE.25.015058.

[15] Andrew J Benedick, Guoqing Chang, Jonathan R Birge, Li-Jin Chen, Alexander G Glenday, Chih-Hao Li, David F Phillips, Andrew Szentgyorgyi, Sylvain Korzennik, Gabor Furesz, Ronald L Walsworth, and Franz X Kärtner, "Visible wavelength astro-comb." Optics Express **18**, 19175–84 (2010).

[16] David F. Phillips, Alexander G. Glenday, Chih-Hao Li, Claire Cramer, Gabor Furesz, Guoqing Chang, Andrew J. Benedick, Li-Jin Chen, Franz X. Kärtner, Sylvain Korzennik, Dimitar Sasselov, Andrew Szentgyorgyi, and Ronald L. Walsworth, "Calibration of an astrophysical spectrograph below 1 m/s using a laser frequency comb," Optics Express **20**, 13711 (2012).

[17] John M. Dudley, Goery Genty, and Stephane Coen, "Supercontinuum generation in photonic crystal fiber," Reviews of Modern Physics **78**, 1135–1184 (2006).

[18] I. Hartl, X. D. Li, C. Chudoba, R. K. Ghanta, T. H. Ko, J. G. Fujimoto, J. K. Ranka, and R. S. Windeler, "Ultrahigh-resolution optical coherence tomography using continuum generation in an air-silica microstructure optical fiber," Optics Letters **26**, 608 (2001).

[19] David J. Jones, Scott A. Diddams, Jinendra K. Ranka, Andrew Stentz, Robert S. Windeler, John L. Hall, and Steven T. Cundiff, "Carrier-Envelope Phase Control of Femtosecond Mode-Locked Lasers and Direct Optical Frequency Synthesis," Science **288**, 635–639 (2000).

[20] Tobias Wilken, Gaspare Lo Curto, Rafael A Probst, Tilo Steinmetz, Antonio Manescau, Luca Pasquini, Jonay I González Hernández, Rafael Rebolo, Theodor W Hänsch, Thomas Udem, and Ronald Holzwarth, "A spectrograph for exoplanet observations calibrated at the centimetre-per-second level," Nature **485**, 611 (2012).

[21] R. A. Probst *et al.*, "Spectral flattening of supercontinua with a spatial light modulator," Proc. SPIE **8864**, 88641Z (2013), doi: 10.1117/12.2036601.

[22] Richard A. McCracken *et al.*, "Wavelength calibration of a high resolution spectrograph with a partially stabilized 15-GHz astrocomb from 550 to 890 nm," Opt. Express **25**, 6450–6460 (2017), doi: 10.1364/OE.25.006450.

[23] Alexander G. Glenday, Chih-Hao Li, Nicholas Langellier, Guoqing Chang, Li-Jin Chen, Gabor Furesz, Alexander A. Zibrov, Franz Kärtner, David F. Phillips, Dimitar Sasselov, Andrew Szentgyorgyi, and Ronald L. Walsworth, "Operation of a broadband visible-wavelength astro-comb with a high-resolution astrophysical spectrograph," Optica **2**, 250–254 (2015), [doi:10.1364/OPTICA.2.000250].

[24] Aakash Ravi *et al.*, "Astro-comb calibrator and spectrograph characterization using a turn-key laser frequency comb," J. Astron. Telesc. Instrum. Syst. **3**, 045003 (2017), doi: 10.1117/1.JATIS.3.4.045003.

[25] M. T. Murphy, T. Udem, R. Holzwarth, A. Sizmann, L. Pasquini, C. Araujo-Hauck, H. Dekker, S. D'Odorico, M. Fischer, T. W. Hansch, and A. Manescau, "High-precision wavelength calibration of astronomical spectrographs with laser frequency combs," Monthly Notices of the Royal Astronomical Society **380**, 839–847 (2007).

[26] Li-Jin Chen *et al.*, "Broadband dispersion-free optical cavities based on zero group delay dispersion mirror sets." Opt. Express **18**, 23204–23211 (2010), doi: 10.1364/OE.18.023204.

[27] T. Wilken *et al.*, "High-precision calibration of spectrographs," Mon. Not. R. Astron. Soc. **405**, L16–L20 (2010), doi: 10.1111/j.1745-3933.2010.00850.x.

[28] F. Quinlan, G. Ycas, S. Osterman, and S. A. Diddams, "A 12.5 GHz-spaced optical frequency comb spanning $>$ 400 nm for near-infrared astronomical spectrograph calibration," Rev. Sci. Instrum. **81**, 063105 (2010), doi: 10.1063/1.3436638.

[29] Gabriel G. Ycas *et al.*, "Demonstration of on-sky calibration of astronomical spectra using a 25 GHz near-IR laser frequency comb," Opt. Express **20**, 6631–6643 (2012), doi: 10.1364/OE.20.006631.

[30] X. Yi, K. Vahala, J. Li, *et al.*, "Demonstration of a near-IR line-referenced electro-optical laser frequency comb for precision radial velocity measurements in astronomy," Nature Comm. **7**, 10436 (2016), doi: 10.1038/ncomms10436.

[31] Rafael A. Probst, Gaspare Lo Curto, Gerardo Avila, Bruno L. Canto Martins, José R. de Medeiros, Massimiliano Esposito, Jonay I. González Hernández, Theodor W. Hänsch, Ronald Holzwarth, Florian Kerber, *et al.*, "A laser frequency comb featuring sub-cm/s precision for routine operation on HARPS," Proc. SPIE **9147**, 91471C (2014).

[32] R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, A. J. Munley, and H. Ward, "Laser phase and frequency stabilization using an optical resonator," Appl. Phys. B **31**, 97–105 (1983), [doi:10.1007/BF00702605].

[33] Guoqing Chang *et al.*, "Optimization of filtering schemes for broadband astro-combs," Opt. Express **20**, 24987–25013 (2012), doi: 10.1364/OE.20.024987.

[34] Xavier Dumusque, Alex Glenday, David F Phillips, *et al.*, "HARPS-N observes the Sun as a star," Astophys. J. Lett. **814**, L21 (2015), [doi:10.1088/2041-8205/814/2/L21].

[35] Xavier Dumusque, Stephane Udry, Christophe Lovis, N. C. Santos, and M. J. P. F. G. Monteiro, "Planetary detection limits taking into account stellar noise. I. Observational strategies

to reduce stellar oscillation and granulation effects," Astron. Astrophys. **525**, A140 (2011), [doi:10.1051/0004-6361/201014097].

[36] Albert Piterman and Zoran Ninkov, "Subpixel sensitivity maps for a back-illuminated charge-coupled device and the effects of nonuniform response on measurement accuracy," Opt. Eng. **41**, 1192–1202 (2002), [doi:10.1117/1.1476691].

[37] Hiroyuki Toyozumi and Michael C B Ashley, "Intra-Pixel Sensitivity Variation and Charge Transfer Inefficiency - Results of CCD Scans," Publ. Astron. Soc. Aust. **22**, 257–266 (2005), [doi:10.1071/AS05013].

[38] Michael T. Murphy, Clayton R. Locke, Philip S. Light, Andre N. Luiten, and Jon S. Lawrence, "Laser frequency comb techniques for precise astronomical spectroscopy," Mon. Not. R. Astron. Soc. **422**, 761–771 (2012), [doi:10.1111/j.1365-2966.2012.20656.x].

[39] Fei Lu, Yujun Deng, and Wayne H. Knox, "Generation of broadband femtosecond visible pulses in dispersion-micromanaged holey fibers," Opt. Lett. **30**, 1566–1568 (2005), doi: 10.1364/OL.30.001566.

[40] Fei Lu and Wayne H. Knox, "Generation, characterization, and application of broadband coherent femtosecond visible pulses in dispersion micromanaged holey fibers," J. Opt. Soc. Am. B **23**, 1221–1227 (2006), doi: 10.1364/JOSAB.23.001221.

[41] A. Kudlinski *et al.*, "Zero-dispersion wavelength decreasing photonic crystal fibers for ultraviolet-extended supercontinuum generation," Opt. Express **14**, 5715–5722 (2006), doi: 10.1364/OE.14.005715.

[42] Parama Pal, Wayne H. Knox, Ingmar Hartl, and Martin E. Fermann, "Self referenced Yb-fiber-laser frequency comb using a dispersion micromanaged tapered holey fiber," Opt. Express **15**, 12161–12166 (2007), doi: 10.1364/OE.15.012161.

[43] Jesper Lægsgaard, "Modeling of nonlinear propagation in fiber tapers," J. Opt. Soc. Am. B **29**, 3183–3191 (2012), doi: 10.1364/JOSAB.29.003183.

[44] Long-Sheng Ma *et al.*, "Optical frequency synthesis and comparison with uncertainty at the 10^{-19} level." Science (New York, N.Y.) **303**, 1843–5 (2004), doi: 10.1126/science.1095092.

[45] J. M. Dudley and J. R. Taylor, eds., *Supercontinuum Generation in Optical Fibers* (Cambridge University Press, 2010).

[46] Nail Akhmediev and Magnus Karlsson, "Cherenkov radiation emitted by solitons in optical fibers," Physical Review A **51**, 2602–2607 (1995).

[47] Guoqing Chang, Li-Jin Chen, and Franz X Kärtner, "Highly efficient Cherenkov radiation in photonic crystal fibers for broadband visible wavelength generation," Optics Letters **35**, 2361–2363 (2010).

[48] Guoqing Chang, Li-Jin Chen, and Franz X Kärtner, "Fiber-optic Cherenkov radiation in the few-cycle regime," Optics Express **19**, 6635–6647 (2011).

[49] Francesco Poletti and Peter Horak, "Description of ultrashort pulse propagation in multimode optical fibers," J. Opt. Soc. Am. B **25**, 1645–1654 (2008), doi: 10.1364/JOSAB.25.001645.

[50] Paul J. Marchand *et al.*, "Visible spectrum extended-focus optical coherence microscopy for label-free sub-cellular tomography," Biomed. Opt. Express **8**, 3343–3359 (2017), doi: 10.1364/BOE.8.003343.

[51] J. Czajkowski *et al.*, "Sub-micron resolution high-speed spectral domain optical coherence tomography in quality inspection for printed electronics," Proc. SPIE **8430**, 84300K (2012), doi: 10.1117/12.922443.

[52] Peter Cimalla, Maria Gaertner, Julia Walther, and Edmund Koch, "Resolution improvement in dual-band OCT by filling the spectral gap," Proc. SPIE **8213**, 82132M (2012), doi: 10.1117/12.908331.

[53] Nicholas Langellier, Chih-Hao Li, Alexander G. Glenday, Guoqing Chang, Hung-Wen Chen, Jinkang Lim, Gabor Furesz, Franz Kärtner, David F. Phillips, Dimitar Sasselov, Andrew Szentgyorgyi, and Ronald Walsworth, "Green astro-comb for HARPS-N," *SPIE Proceedings*, Proceedings of SPIE **9147**, 91478N (2014).

[54] C Jurgenson, D Fischer, T Mccracken, D Sawyer, A Szymkowiak, A B Davis, G Muller, and F Santoro, "EXPRES: A Next Generation RV Spectrograph in the Search for Earth-like Worlds," Proceedings of SPIE **9908**, 99086T (2016).

[55] Kazuhiro Ikeda, Robert E. Saperstein, Nikola Alic, and Yeshaiahu Fainman, "Thermal and Kerr nonlinear properties of plasma-deposited silicon nitride/ silicon dioxide waveguides," Optics Express **16**, 12987 (2008).

[56] Marko Lončar, "Integrated Lithium Niobate Photonic and Applications," Conference on Lasers and Electro-Optics (2019), 10.23919/CLEO.2019.8750031.

[57] B. J. M. Hausmann, I. Bulu, V. Venkataraman, P. Deotare, and M. Lončar, "Diamond nonlinear photonics," Nature Photonics **8**, 369–374 (2014).

[58] Benjamin Feigel, David Castelló-Lurbe, Hugo Thienpont, and Nathalie Vermeulen, "Opportunities for visible supercontinuum light generation in integrated diamond waveguides," Optics Letters **42**, 3804 (2017).

[59] T. Tsuchizawa, K. Yamada, H. Fukuda, T. Watanabe, Jun-ichi Takahashi, M. Takahashi, T. Shoji, E. Tamechika, S. Itabashi, and H. Morita, "Microphotonics devices based on silicon microfabrication technology," IEEE Journal of Selected Topics in Quantum Electronics **11**, 232–240 (2005).

[60] Lukas Chrostowski and Michael Hochberg, *Silicon Photonics Design – From Devices to Systems* (Cambridge University Press, 2015).

[61] Jaime Cardenas, Carl B. Poitras, Kevin Luke, Lian-Wee Luo, Paul Adrian Morton, and Michal Lipson, "High Coupling Efficiency Etched Facet Tapers in Silicon Waveguides," IEEE Photonics Technology Letters **26**, 2380–2382 (2014).

[62] Arman B. Fallahkhair, Kai S. Li, and Thomas E. Murphy, "Vector Finite Difference Modesolver for Anisotropic Dielectric Waveguides," Journal of Lightwave Technology **26**, 1423–1431 (2008).

[63] Haolan Zhao, Bart Kuyken, Stéphane Clemmen, François Leo, Ananth Subramanian, Ashim Dhakal, Philippe Helin, Simone Severi, Edouard Brainis, Gunther Roelkens, and Roel Baets, "Visible-to-near-infrared octave spanning supercontinuum generation in a silicon nitride waveguide," Optics Letters **40**, 2177 (2015).

[64] Jo Bovy, Carlos Allende Prieto, Timothy C. Beers, Dmitry Bizyaev, Luiz N. da Costa, Katia Cunha, Garrett L. Ebelke, Daniel J. Eisenstein, Peter M. Frinchaboy, García Pérez, *et al.*, "The Milky Way's Circular-velocity Curve between 4 and 14 kpc from APOGEE data," Astrophys. J. **759**, 131 (2012).

[65] Vera C. Rubin and Jr. Ford, W. Kent, "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions," The Astrophysical Journal **159**, 379 (1970).

[66] V. C. Rubin, N. Thonnard, and Jr. Ford, W. K., "Extended rotation curves of high-luminosity spiral galaxies. IV - Systematic dynamical properties, SA through SC," The Astrophysical Journal **225**, L107 (1978).

[67] V. C. Rubin, N. Thonnard, and Jr. Ford, W. K., "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4kpc) to UGC 2885 (R = 122 kpc)," The Astrophysical Journal **238**, 471 (1980).

[68] Yoshiaki Sofue, Mareki Honma, and Toshihiro Omodaka, "Unified Rotation Curve of the Galaxy – Decomposition into de Vaucouleurs Bulge, Disk, Dark Halo, and the 9-kpc Rotation Dip –," Publications of the Astronomical Society of Japan **61**, 227–236 (2009).

[69] G. Bertone and D. Hooper, "A History of Dark Matter," Rev. Mod. Phys. **90**, 045002 (2018).

[70] C. Patrignani *et al.* (Particle Data Group), "Review of Particle Physics," Chin. Phys. C **40**, 100001 (2016).

[71] Justin I. Read, "The local dark matter density," J. Phys. G: Nucl. Part. Phys. **41**, 063101 (2014).

[72] Prajwal R. Kafle, Sanjib Sharma, Geraint F. Lewis, and Joss Bland-Hawthorn, "Kinematics of of the stellar halo and the mass distribution of the Milky Way using blue horizontal branch stars," Astrophys. J. **761**, 98 (2012).

[73] Jo Bovy and Scott Tremaine, "On the local dark matter density," Astrophys. J. **756**, 89 (2012).

[74] Miguel Pato, Fabio Iocco, and Gianfranco Bertone, "Dynamical constraints on the dark matter distribution in the Milky Way," J. Cosmol. Astropart. Phys. **1512**, 001 (2015).

[75] Katelin Schutz, Tongyan Lin, Benjamin R. Safdi, and Chih-Liang Wu, "Constraining a Thin Dark Matter Disk with *Gaia*," Phys. Rev. Lett. **121**, 081101 (2018).

[76] Morgan Bennett and Jo Bovy, "Vertical waves in the solar neighbourhood in *Gaia* DR2,"

Mon. Not. Roy. Astron. Soc. **482**, 1417–1425 (2019).

[77] Matthew R. Buckley and Annika H. G. Peter, "Gravitational probes of dark matter physics," Phys. Rep. **761**, 1–60 (2018).

[78] Ethan R. Siegel, M. P. Hertzberg, and J. N. Fry, "Probing Dark Matter Substructure with Pulsar Timing," Mon. Not. Roy. Astron. Soc. **382**, 879 (2007).

[79] Yashar Hezaveh, Neal Dalal, Gilbert Holder, Michael Kuhlen, Daniel Marrone, Norman Murray, and Joaquin Vieira, "Dark Matter Substructure Detection Using Spatially Resolved Spectroscopy of Lensed Dusty Galaxies," Astrophys. J. **767**, 9 (2013).

[80] Yashar Hezaveh, Neal Dalal, Gilbert Holder, Theodore Kisner, Michael Kuhlen, and Laurence P. Levasseur, "Measuring the power spectrum of dark matter substructure using strong gravitational lensing," J. Cosmol. Astropart. Phys **1611**, 048 (2016).

[81] Denis Erkal and Vasily Belokurov, "Forensics of subhalo–stream encounters: the three phases of gap growth," Mon. Not. Roy. Astron. Soc. **450**, 1136–1149 (2015).

[82] Denis Erkal, Vasily Belokurov, Jo Bovy, and Jason L. Sanders, "The number and size of subhalo-induced gaps in stellar streams," Mon. Not. Roy. Astron. Soc. **463**, 102–119 (2016).

[83] Ken Van Tilburg, Anna-Maria Taki, and Neal Weiner, "Halometry from Astrometry," J. Cosmol. Astropart. Phys. **1807**, 041 (2018).

[84] Gianfranco Bertone, Dan Hooper, and Joseph Silk, "Particle dark matter: evidence, candidates and constraints," Phys. Rep. **405**, 279–390 (2005).

[85] Sean Tulin and Hai-Bo Yu, "Dark matter self-interactions and small scale structure," Phys. Rep. **730**, 1–57 (2018).

[86] Luca Amendola, Amedeo Balbi, and Claudia Quercellini, "Peculiar acceleration," Phys. Lett. B **660**, 81–86 (2008).

[87] C. Quercellini, L. Amendola, and A. Balbi, "Mapping the galactic gravitational potential with peculiar acceleration," Mon. Not. Roy. Astron. Soc. **391**, 1308–1314 (2008).

[88] H. Silverwood and R. Easther, "Stellar Accelerations and the Galactic Gravitational Field," arXiv:1812.07581 .

[89] Jo Bovy, "Galactic rotation in *Gaia* DR1," Mon. Not. Roy. Astron. Soc. **468**, L63–L67 (2017).

[90] David G. Koch, William J. Borucki, Gibor Basri, Natalie M. Batalha, Timothy M. Brown, Douglas Caldwell, Jørgen Christensen-Dalsgaard, William D. Cochran, Edna DeVore, Edward W. Dunham, *et al.*, "Kepler mission design, realized photometric performance, and early science," Astrophys. J. **713**, L79–L86 (2010).

[91] Andrew Szentgyorgyi, Daniel Baldwin, Stuart Barnes, Jacob Bean, Sagi Ben-Ami, Patricia Brennan, Jamie Budynkiewicz, Daniel Catropa, Moo-Young Chun, Charlie Conroy, *et al.*, "The GMT-consortium large earth finder (G-CLEF): an optical echelle spectrograph for the

Giant Magellan Telescope (GMT)," Proc. SPIE **10702**, 63 (2018).

[92] P. Spanò, D. Mégevand, J. M. Herreros, F. M. Zerbi, A. Cabral, P. Di Marcantonio, C. Lovis, S. Cristiani, R. Rebolo, N. Santos, *et al.*, "Optical design of the ESPRESSO spectrograph at VLT," Proc. SPIE **7735**, 77350K (2010).

[93] A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Eyer, F. Jansen, *et al.*, "*Gaia* Data Release 2," Astron. Astrophys. **616**, A1 (2018).

[94] Roberto Gilmozzi, Bernard Delabre, Philippe Dierickx, Norbert N. Hubin, Franz Koch, Guy J. Monnet, Marco Quattri, Francois J. Rigaut, and Raymond N. Wilson, "Future of filled aperture telescopes: is a 100-m feasible?" Proc. SPIE **3352**, 778–791 (1998).

[95] Glen LeDrew, "The Real Starry Sky," J. Royal Astron. Soc. Can. **95**, 32 (2001).

[96] Wesley A. Traub, "Terrestrial, Habitable-Zone Exoplanet Frequency from *(*Kepler)," Astrophys. J. **745**, 20 (2012).

[97] Steven H. Saar, R. Paul Butler, and Geoffrey W. Marcy, "Magnetic Activity-related Radial Velocity Variations in Cool Stars: First Results from the Lick Extrasolar Planet Survey," Astrophys. J. **498**, L153–L157 (1998).

[98] Deepak Raghavan, Harold A. McAlister, Todd J. Henry, David W. Latham, Geoffrey W. Marcy, Brian D. Mason, Douglas R. Gies, Russel J. White, and Theo A. ten Brummelaar, "A survey of stellar families: multiplicity of solar-type stars," Astrophys. J. Suppl. Ser. **190**, 1–42 (2010).

[99] Gaspard Duchêne and Adam Kraus, "Stellar Multiplicity," Annu. Rev. Astron. Astrophys. **51**, 269–310 (2013).

[100] A. Cassan, D. Kubas, J.-P. Beaulieu, M. Dominik, K. Horne, J. Greenhill, J. Wambsganss, J. Menzies, A. Williams, U. G. Jørgensen, *et al.*, "One or more bound planets per Milky Way star from microlensing observations," Nature **481**, 167–169 (2012).

[101] Li Zeng, Stein B. Jacobsen, Dimitar D. Sasselov, and Andrew Vanderburg, "Survival Function Analysis of Planet Orbit Distribution and Occurrence Rate Estimate," (2018), arXiv:1801.03994 .

[102] Li Zeng, Stein B. Jacobsen, Dimitar D. Sasselov, and Andrew Vanderburg, "Survival Function Analysis of Planet Size Distribution," (2018), arXiv:1801.03993 .

[103] Dolev Bashi, Ravit Helled, Shay Zucker, and Christoph Mordasini, "Two empirical regimes of the planetary mass-radius relation," Astron. Astrophys. **604**, A83 (2017).

[104] S Roberts, M Osborne, M Ebden, S Reece, N Gibson, and S Aigrain, "Gaussian processes for time-series modelling," Philos. Trans. Royal Soc. A **371**, 20110550 (2013).

[105] V. Rajpaul, S. Aigrain, M. A. Osborne, S. Reece, and S. Roberts, "A Gaussian process framework for modelling stellar activity signals in radial velocity data," Mon. Not. Roy. Astron.

Soc. **452**, 2269–2291 (2015).

[106] N. Meunier, A.-M. Lagrange, and M. Desort, "Reconstructing the solar integrated radial velocity using MDI/SOHO," Astron. Astrophys. **519**, A66 (2010).

[107] Carl Edward Rasmussen and Christopher K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, 2006).

[108] R. D. Haywood, A. Collier Cameron, D. Queloz, S. C. C. Barros, M. Deleuil, R. Fares, M. Gillon, A. F. Lanza, C. Lovis, C. Moutou, *et al.*, "Planets and stellar activity: hide and seek in the CoRoT-7 system," Mon. Not. Roy. Astron. Soc. **443**, 2517–2531 (2014).

[109] Andrew Lyne and Francis Graham-Smith, *Pulsar Astronomy*, 4th ed. (Cambridge University Press, 2012).

[110] John G Hartnett and Andre N Luiten, "*Colloquium*: Comparison of astrophysical and terrestrial frequency standards," Review of Modern Physics **83**, 1–9 (2011).

[111] Andrea N Lommen, "Pulsar timing arrays: the promise of gravitational wave detection," Reports on Progress in Physics **78**, 124901 (2015).

[112] Chiara M F Mingarelli, Lauren Anderson, Megan Bedell, and David N Spergel, "Improving Binary Millisecond Pulsar Distances with *Gaia*," (2018), arXiv:1812.06262v1 .

[113] Ross J. Jennings, David L. Kaplan, Shami Chatterjee, James M. Cordes, and Adam T. Deller, "Binary Pulsar Distances and Velocities from *Gaia* Data Release 2," The Astrophysical Journal **864**, 26 (2018).

[114] Hans J. Deeg and Roi Alonso, "Transit Photometry as an Exoplanet Discovery Method," (2018), 10.1007/978-3-319-55333-7_117, arXiv:1803.07867 .

[115] I. A. G. Snellen, R. J. de Kok, R. le Poole, M. Brogi, and J. Birkby, "Finding extraterrestrial life using ground-based high-resolution spectroscopy," The Astrophysical Journal **764**, 182 (2013).

[116] Florian Rodler and Mercedes López-Morales, "Feasibility Studies for the Detection of O_2 in an Earth-like Exoplanet," The Astrophysical Journal **781**, 54 (2014).

[117] Sagi Ben-Ami, Mercedes López-Morales, Juliana Garcia-Mejia, Gonzalo Gonzalez Abad, and Andrew Szentgyorgyi, "High Resolution Spectroscopy Using Fabry Perot Interferometer Arrays: An Application to Searches for O_2 in Exoplanetary Atmospheres," (2018), arXiv:1805.12011v1 .

[118] Mercedes Lopez-Morales, Sagi Ben-Ami, Gonzalo Gonzalez-Abad, Juliana Garcia-Mejia, Jeremy Dietrich, and Andrew Szentgyorgyi, "Optimizing Ground-based Observations of O_2 in Earth Analogs," (2019), arXiv:1905.05862v1 .

[119] Dilovan B Serindag and Ignas A G Snellen, "Testing the detectability of extraterrestrial O_2 with the ELTs using real data with real noise," (2019), arXiv:1901.02469v1 .

[120] V. S. Meadows *et al.*, "Exoplanet Biosignatures: Understanding Oxygen as a Biosignature in the Context of Its Environment," (2017), arXiv:1705.07560 .

[121] L.R. Brown and C. Plymate, "Experimental Line Parameters of the Oxygen A Band at 760 nm," Journal of Molecular Spectroscopy **199**, 166–179 (2000).

[122] Benjamin Scherer, Jürgen Wöllenstein, Matthias Weidemüller, Wenzel Salzmann, Johannes Michael Ostermann, Fernando Rinaldi, and Rainer Michalzik, "Measurement of the pressure broadening coefficients of the oxygen A-band using a low cost, polarization stabilized, widely tunable vertical-cavity surface-emitting laser," Microsystem Technologies **14**, 607–614 (2008).

[123] Brian J. Drouin, D. Chris Benner, Linda R. Brown, Matthew J. Cich, Timothy J. Crawford, V. Malathy Devi, Alexander Guillaume, Joseph T. Hodges, Eli J. Mlawer, David J. Robichaud, Fabiano Oyafuso, Vivienne H. Payne, Keeyoon Sung, Edward H. Wishnow, and Shanshan Yu, "Multispectrum analysis of the oxygen A-band," Journal of Quantitative Spectroscopy and Radiative Transfer **186**, 118–138 (2017).

[124] Z. Yang, P.O. Wennberg, R.P. Cageao, T.J. Pongetti, G.C. Toon, and S.P. Sander, "Ground-based photon path measurements from solar absorption spectra of the O_2 A-band," Journal of Quantitative Spectroscopy and Radiative Transfer **90**, 309–321 (2005).

[125] Charles E. Miller and Debra Wunch, "Fourier transform spectrometer remote sensing of O_2 A-band electric quadrupole transitions," Journal of Quantitative Spectroscopy and Radiative Transfer **113**, 1043–1050 (2012).

[126] David A. Long and Joseph T. Hodges, "On spectroscopic models of the O_2 A-band and their impact upon atmospheric retrievals," Journal of Geophysical Research: Atmospheres **117**, D12309 (2012).

[127] Joseph Mendonca, Kimberly Strong, Debra Wunch, Geoffrey C. Toon, David A. Long, Joseph T. Hodges, Vincent T. Sironneau, and Jonathan E. Franklin, "Using a speed-dependent Voigt line shape to retrieve O_2 from Total Carbon Column Observing Network solar spectra to improve measurements of XCO_2," Atmospheric Measurement Techniques **12**, 35–50 (2019).

[128] Peter F. Bernath, *Spectra of Atoms and Molecules*, 2nd ed. (Oxford University Press, New York, 2005).

[129] Jon T. Hougen, "Classification of Rotational Energy Levels for Symmetric-Top Molecules," The Journal of Chemical Physics **37**, 1433–1441 (1962).

[130] Jon T. Hougen, "Classification of Rotational Energy Levels. II," The Journal of Chemical Physics **39**, 358–365 (1963).

[131] S.M. Abrarov and B.M. Quine, "Efficient algorithmic implementation of the Voigt/complex error function based on exponential series approximation," Applied Mathematics and Computation **218**, 1894–1902 (2011).

[132] Ibrahim Reda and Afshin Andreas, *Solar Position Algorithm for Solar Radiation Applications*, Tech. Rep. (National Renewable Energy Laboratory, 2008) NREL/TP-560-34302.

[133] R. A. Washenfelder, G. C. Toon, J.-F. Blavier, Z. Yang, N. T. Allen, P. O. Wennberg, S. A. Vay, D. M. Matross, and B. C. Daube, "Carbon dioxide column abundances at the Wisconsin Tall Tower site," Journal of Geophysical Research **111**, D22305 (2006).

[134] Peter R. Griffiths, *Fourier Transform Spectrometry*, 2nd ed. (John Wiley & Sons, Inc., 2007).

[135] M. Brogi, I. A. G. Snellen, R. J. de Kok, S. Albrecht, J. L. Birkby, and E. J. W. de Mooij, "Detection of molecular absorption in the dayside of exoplanet 51 Pegasi b?" The Astrophysical Journal **767**, 27 (2013).

[136] Y. Takeda and S. Ueno, "Detection of Gravitational Redshift on the Solar Disk by Using Iodine-Cell Technique," Solar Physics **281**, 551–575 (2012).

[137] I.E. Gordon *et al.*, "The HITRAN2016 molecular spectroscopic database," Journal of Quantitative Spectroscopy and Radiative Transfer **203**, 3–69 (2017).

[138] Johan Hult, "A Fourth-Order Runge-Kutta in the Interaction Picture Method for Simulating Supercontinuum Generation in Optical Fibers," Journal of Lightwave Technology **25**, 3770–3775 (2007).

[139] G.P. Agrawal, *Nonlinear Fiber Optics*, 5th ed. (Academic Press, 2012).

[140] J. C. Travers and J. R. Taylor, "Soliton trapping of dispersive waves in tapered optical fibers," Opt. Lett. **34**, 115–117 (2009), doi: 10.1364/OL.34.000115.

[141] A. C. Judge *et al.*, "Optimization of the soliton self-frequency shift in a tapered photonic crystal fiber," J. Opt. Soc. Am. B **26**, 2064–2071 (2009), doi: 10.1364/JOSAB.26.002064.

[142] O. Vanvincq, J. C. Travers, and A. Kudlinski, "Conservation of the photon number in the generalized nonlinear Schrödinger equation in axially varying optical fibers," Phys. Rev. A **84**, 063820 (2011), doi: 10.1103/PhysRevA.84.063820.

[143] K. L. Corwin, N. R. Newbury, J. M. Dudley, S. Coen, S. A. Diddams, K. Weber, and R. S. Windeler, "Fundamental Noise Limitations to Supercontinuum Generation in Microstructure Fiber," Physical Review Letters **90**, 113904 (2003).

[144] R. Paschotta, "Noise of mode-locked lasers (Part I): numerical model," Appl. Phys. B **79**, 153–162 (2004), doi: 10.1007/s00340-004-1547-x.

[145] Axel Ruehl *et al.*, "Ultrabroadband coherent supercontinuum frequency comb," Phys. Rev. A **84**, 011806 (2011), doi: 10.1103/PhysRevA.84.011806.

[146] G.P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. (Academic Press, 2001).

[147] Mark A. Foster and Alexander L. Gaeta, "Ultra-low threshold supercontinuum generation in sub-wavelength waveguides," Optics Express **12**, 3137–3143 (2004).

[148] M. A. Foster, K. D. Moll, and Alexander L. Gaeta, "Optimal waveguide dimensions for nonlinear interactions," Optics Express **12**, 2880 (2004).

[149] Allan W. Snyder and John D. Love, *Optical Waveguide Theory* (Chapman and Hall, 1983).

[150] B. Kibler, J. M. Dudley, and S. Coen, "Supercontinuum generation and nonlinear pulse propagation in photonic crystal fiber: Influence of the frequency-dependent effective mode area," Applied Physics B: Lasers and Optics **81**, 337–342 (2005).

[151] Aakash Ravi, Nicholas Langellier, David F. Phillips, Malte Buschmann, Benjamin R. Safdi, and Ronald L. Walsworth, "Probing Dark Matter Using Precision Measurements of Stellar Accelerations," Physical Review Letters **123**, 091101 (2019).

[152] Jorge Nocedal and Stephen J. Wright, *Numerical Optimization*, edited by Peter Glynn and Stephen M. Robinson (Springer, 1999) pp. 136–138.

[153] G. van Rossum, *Python tutorial*, Tech. Rep. CS-R9526 (Centrum voor Wiskunde en Informatica (CWI), Amsterdam, 1995).

[154] Travis Oliphant, "NumPy: A guide to NumPy," USA: Trelgol Publishing (2006).

[155] Eric Jones, Travis Oliphant, Pearu Peterson, *et al.*, "SciPy: Open source scientific tools for Python," (2001).

[156] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, and M. O'Neil, "Fast direct methods for gaussian processes," arXiv:1403.6015 .