PLANAR-WAVEGUIDE FOURIER-TRANSFORM SPECTROMETERS FOR SPACE EXPLORATION

HUGH DENNIS PODMORE

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN PHYSICS AND ASTRONOMY
YORK UNIVERSITY
TORONTO, ONTARIO
DECEMBER 2019

Abstract

Optical spectroscopy is a powerful, fundamental analytical technique that is deployed across many domains of scientific research. Due to the widespread use of this technique, the development of novel methods and instruments for collecting and analyzing spectral data provides new foundations and bridges by which the frontier of human inquiry may be explored. Miniature spectrometers, in particular, lead to advancements in space exploration by enabling advanced analytical techniques to be deployed on small rovers and satellites. One promising platform for realizing miniature spectrometers with high resolving power are integrated optical circuits consisting of many hundreds of discrete optical components implemented via planar-waveguides printed lithographically on wafers.

This dissertation presents the design, fabrication, and characterization of two miniature spectrometers based on planar-waveguide technology. The miniature spectrometers are based on a spatial-heterodyne Fourier-transform spectrometer architecture that results in very high spectral resolution over a limited spectral range.

The first device presents an advanced method for mitigating temperature sensitivity of the device—an enabling step towards deployment of the sensor in harsh environments. The device is also used to demonstrate, for the first time, the retrieval of a complex spectrum consisting of multiple gas absorption features using on-chip Fourier-transform spectroscopy. The second device advances the state of the art by deploying novel spectral retrieval techniques from the field of compressive sensing in order to substantially increase the spectral range of the instrument while maintaining high spectral resolution.

This dissertation motivates planar-waveguide Fourier-transform spectrometers as the ideal vehicle for the adoption of this technique, which ultimately results in a significant increase in capability for the spectrometer. Ultimately, a four-fold increase in spectral range over conventional systems is achieved for the compressive-sensing instrument, and spectral retrieval of complex spectra is achieved despite significant under-sampling with respect to a conventional instrument architecture. The invention and successful demonstration of this novel instrument architecture, and the development of techniques for fabricating such devices, creates an avenue through which high-performing on-chip spectrometers may be deployed to advance scientific inquiry across a multitude of fields including space applications.

"Astronomy is useful because it raises us above ourselves; it is useful because it is grand. It shows us how small is man's body, how great his mind, since his intelligence can embrace the whole of this dazzling immensity, where his body is only an obscure point, and enjoy its silent harmony." - Henri Poincaré

"I don't know where I'm going from here, but I promise it won't be boring." – David Robert Jones

Acknowledgements

Throughout the course of my education there have been many friends, mentors, and family members who have—in ways great and small—helped to see me over the finish line. Newton famously claimed: “If I have seen further, it is by standing on the shoulders of Giants”, and while I cannot claim to have seen as far as Newton, it is only fitting to recognize those giants in my own life who have raised me up, given me the benefit of their wisdom, and the (many) blessings of their patience.

First, to my supervisor, Dr. Regina Lee, I am overwhelmed by your tireless commitment to my personal and scientific improvement over the past six years. You have made it your mission to create for me the educational environment, growth opportunities, and valuable experiences that have made me the scientist that I am today. I will be forever grateful for the guidance and support that you have provided, and consider myself exceptionally fortunate to have been selected by a supervisor who takes such pride in her students. I am not certain whether I will return to academia, but if I do, I cannot think of a supervisor whom I would be better-served to emulate.

To my dear friend and colleague Dr. Alan Scott, space lord, I continue to be amazed by and grateful for your wisdom and experience in optics, and the physics of space. Your contributions towards this thesis were, of course, profound, and I greatly appreciate the continual efforts that you have made on behalf of my education and growth as a scientist. Every day I continue to feel that I am learning through osmosis by working alongside you, and in addition to being appreciative of your great sense of humour, it is my aspiration to develop into as knowledgeable and meticulous a scientist as you.

To my great friend and mentor, Dr. Pavel Cheben, it has been my privilege to be welcomed by you into your home and your friendship. I have benefited immensely from your tutelage and perspective in the art of photonic design, as well as culture, politics, and the wider world. I am constantly impressed with your breadth of knowledge and depth of insight and consider you an exemplary role model as both a scientist and human being. I have greatly valued our far-ranging discussions over dinners, and continue to be grateful for your wisdom, support, and your patience.

To Dr. Michael Daly and Dr. A Kumarkrishnan, I am grateful for your support over the years as members of my research evaluation committee. In addition to your valuable reviews of my research progress, I am grateful to each of you for your provision of material support for this work in the form of equipment access and loans—without which this work would not have been possible.

I would also like to thank Andrew Lohmann, Zachary Corriveau, Martin Vachon, Thong Thai, Juan Guzman, and Siddarth Dave for their various contributions to the test setups and electronics that were required to achieve the results presented in this thesis. I am particularly grateful to Andrew and Zachary for their extended support in assembling and characterizing the experimental setup at Com Dev.

To my labmates and friends over the years, Patrick, Thomas, Andrew, Catherine, Sid, Juan, Thong, Konstantin, Guy, KJ, Tiffany, and Akash, thank you for your support in sounding off crazy ideas, and for making the lab a great place to work.

To Marlene Caplan, Janaki DeCamillis, and Cristalina Del Biondo, thank you for your administrative support over the years.

To Emily Yorke, thank you for being my first friend in a new city—twice!

To my best friends, Ben and Dan, thank you for keeping me saddled and upright.

To Dr. Hermina, my partner in life and peste dulce, words can only scratch the surface of my gratitude for your kindness, understanding, and love throughout all the trials and tribulations of these past years. It has been my privilege to walk this journey at your side.

Finally, to my parents Dorothy and Neil, and my sisters Alice, Olivia, Louise, and Laura, thank you for your decades of love and support. You have all contributed positively to make me the person that I am today, and I am grateful to have each of you in my life.

Table of Contents

List of Tables

List of Figures

Abbreviations

Abbreviation	Long form
ASE	Amplified spontaneous emission
BPM	Beam propagation method
CS	Compressive sensing
CMP	Chemical mechanical polishing
DFT	Discrete Fourier transform
DIAL	Differential absorption LIDAR
EIM	Effective index method
EME	Eigenmode expansion
FD	Finite differences
FDTD	Finite-difference time-domain
FFT	Fast Fourier-transform
FT	Fourier transform
FTS	Fourier-transform spectrometer
FWHM	Full width at half-maximum
IFFT	Inverse fast Fourier-transform
ILS	Instrument line shape
LIBS	Laser-induced breakdown spectroscopy
LSSA	Least-squares spectral analysis
MFD	Mode-field diameter
MMI	Multi-mode interference device
MZI	Mach-Zehnder interferometer
OPD	Optical path delay
PIC	Photonic integrated circuit
PSF	Point-spread function
RIP	Restricted isometry property
SHFTS	Spatial-heterodyne Fourier-transform spectrometer
SHS	Spatial-heterodyne spectrometer
SSC	Spot-size converter
SWG	Subwavelength grating
TIR	Total internal reflection
TOC	Thermooptic coefficient
ZPD	Zero path difference

Symbol	Description
P_{avg}	Average photon arrival rate
$\bar{N}$	Average number of photons
$\bar{I}$	Average photocurrent
η	External quantum efficiency
Δf	Detector bandwidth
τ	Time interval recorded by photodiode
α_G, β_G	Grating incidence and exit angle for dispersive instrument
m	Diffraction order
d	Diffraction grating spacing
D_v	Combined grating angle
W_{ent}, W_{ext}	Grating spectrometer entrance and exit pupil widths
f_{ext}	Exit focal length
$\delta\lambda, \Delta\lambda$	Spectral resolution and range
$h_{slit}, A_{grating}$	Slit height and illuminated area of grating
G_{spec}	étendue of dispersive spectrometer
$\mathrm{III}_T(t)$	Dirac comb with periodicity T
F_{III}	Frequency of Dirac comb based on comb periodicity
σ	Optical wavenumber
$\delta\sigma, \Delta\sigma$	Spectral resolution and range in wavenumbers
$\delta t, \Delta t$	FTS sampling period and length (time)
Δx	Optical path delay (spatial)
$I_0(\sigma), I_{det}(\sigma)$	Input spectral intensity, output intensity at detector
$\mathbf{A}$	Spectral sampling matrix
$\mathbf{x}, \hat{\mathbf{x}}$	Spectrum of interest and estimate of spectrum
$\mathbf{y}$	Output interference pattern (measurement)
$\vec{n}_{12}$	Unit normal vector from medium 1 to medium 2
$\sigma_s, \vec{J}_s$	Surface charge and surface current density
n_i, n_t	Refractive index of incident and transmitted media
θ_i, θ_t	Incident and transmitted ray angles
n_{eff}, n_{wg}	Effective index of waveguide, waveguide index
β_i	Propagation constant for mode i
L_π	Modal beat-length
S_{ij}	Scattering (amplitude, phase) from port i to port j
E^{in}, E^{out}	Input and output electric field for fundamental mode
α_{eff}	Effective thermooptic coefficient of waveguide
$P_0(\sigma)$	Spectral power (intensity) of input source

Preface

Preface

This thesis consists of an introduction, a theoretical discussion, three published, peer-reviewed articles, and a concluding statement. In chapter 1, the state of microphotonic spectroscopy and recent advantages and open challenges in the development of miniaturized spectrometers are reviewed.

The thesis builds towards the eventual design, fabrication and test of three distinct Fourier-transform waveguide spectrometers. Accordingly, much of the front matter of the thesis deals with the theoretical background and design considerations that must be understood before delving into the specifics of waveguide spectrometer design. In chapter 2, the underlying theory of Fourier-transform spectroscopy including sampling theory, the processing of uniform and non-uniform Fourier-transforms, and compressive sensing is presented. In chapter 3, the general design and layout principles for an on-chip Fourier-transform spectrometer are presented. In the chapters 4–6, the three waveguide spectrometer devices, each addressing different applications and manifesting different advancements of the underlying technology are presented. Finally, in Chapter 7 the contributions of each of these three devices are detailed, and a concluding statement summarizing the overall findings and impact of the thesis is presented.

Objective

The objectives of this thesis and this research are as follows:

1. Develop, characterize, and validate, through spectral retrieval of gas absorption features, a thermally-insensitive on-chip Fourier-transform spectrometer for greenhouse gas detection.

2. Investigate and demonstrate, through spectral retrieval of complex spectra, the potential improvement of on-chip Fourier-transform spectrometers via the application of compressive sensing.

1 Introduction

Spectroscopy is a fundamental scientific technique with applications spanning the fields of astronomy [1], atomic physics [2], biology [3], chemistry [4], dentistry [5], earth sciences [6], and even entomology [7]. Fundamentally, spectroscopy is a technique permitting quantitative analysis of the spectral (*i.e.* colour) content of electromagnetic radiation—typically in the ultraviolet, visible, and infrared—that is transmitted or reflected by an object of interest. Spectroscopy may be performed in a laboratory environment or as a field experiment [8], however, because of the ubiquity of electromagnetic radiation throughout the universe, its most powerful application is in the analysis of objects at standoff distances.

Spectroscopy is equally suitable for investigating geological samples from the vantage point of a rover arm meters above the ground [9], as well as for analyzing the atomic content of stars billions of lightyears away [1]. Indeed, astronomy is a particularly instructive case study in the history of spectroscopy: in 800 BC Babylonian astronomers differentiated stars by colour, as did the Greeks, Romans, and renaissance scientists in their time [10]. As spectroscopic technology has improved the technique has been integral to many disruptive discoveries and observations including: exoplanets [11], pop III stars [12], liquid water beyond earth [13], the expansion of the universe [14], and many more. Closer to home, spectroscopy has been a key driver of our understanding of the greenhouse-gas effect [15], the structure of atoms [16], and human biology [17]. In even more general terms, spectroscopy is such a fundamentally powerful measurement technique that colour analysis has developed naturally across much of the animal kingdom [18, 19].

The above breadth of applications of spectroscopy are presented in order to provide context for the subject of this dissertation, which is the development of a specific pair of spectrometers for comparatively focused applications. Many of the tradeoffs and decisions presented in this dissertation are driven by the specific applications of these devices, however, as should now be apparent from the previous discussion, the spectrometers and concepts presented herein may trace a lineage to an enormous diversity of scientific inquiry.

This dissertation, however, focuses in particular on the development of miniature spectrometers for remote sensing and space exploration. In the last half-century, since the advent of the space age, spectroscopy has been increasingly used to understand the Earth system via remote observation from satellite platforms [15, 20]. The richness of spectroscopic data collected from space platforms enables a diverse array of analyses. In the ultraviolet and visible domain, spaceborne spectroscopy may be used to characterize aurorae to better understand Earth's magnetosphere [21]. Spaceborne spectrometers may also be used to quantify ocean colour to identify algae blooms or to map ocean turbidity [22]. They may also be used to measure plant colour to evaluate crop health [23], chlorophyll content [24], and identify dead and dying forests [25]. In the near-infrared regime, spectroscopy is used to study atmospheric chemistry [26], enabling quantification of greenhouse-gas species present in the upper atmosphere, and identify emission sources on the ground [27]. Finally, in the submillimetre regime, remote sensing spectometers are used for observation of soil moisture content, ocean salinity, snowpack structure and chemistry [28], and atmospheric chemical species in the presence of clouds and other aerosols [29].

Traditionally this kind of earth-monitoring spectrometers have been large, multimillion dollar projects sponsored by national space agencies such as the National Aeronautics and Space Administration, the European Space Agency, or the Canadian Space Agency. Increasingly, however, the private sector is driving a demand for miniaturized spectroscopic systems that could be deployed on low-cost satellite platforms such as nanosatellites [30] and microsatellites [31]. This trend in remote-sensing towards miniaturization is also reflected in other fields such as planetary exploration where small, lightweight spectrometers may be used for in-situ sample analysis on rover platforms. Rover-based spectroscopy has been utilized in order to better understand the Martian, lunar, and Titanic systems, and the ever decreasing form-factors of these rovers, following the micro and nanosatellite revolution, contributes to the need for miniaturized high-performance spectrometers. Of particular interest, for planetary exploration, are infrared spectroscopic systems for gas detection, and UV-visible spectrometers for biological and geological applications—particularly through Raman and laser-induced breakdown spectroscopy (LIBS) [9].

1.1 Waveguide spectrometers

In the pursuit of ever more miniaturized spectrometers to address the applications described in the previous section, some fields have moved away from traditional free-space optical assemblies consisting of gratings, lenses, prisms, apertures and mirrors. Spectroscopic devices may also be realized in highly compact arrangements of optical waveguides that are inscribed on wafers lithographically. Such devices, referred to as "photonic integrated circuits" enable highly miniaturized implementations of complex optical devices—including high-performance spectrometers.

Photonic integrated circuits (PICs) were initially developed in the 1990s, as an extension of burgeoning research and commercial interest in fibre-optic communications. The advantage of PIC devices, particularly for commercial applications, is that the ability to manufacture complex optical interconnections and circuits in high-volume and at extremely low per-unit cost without the need for element-by-element optical alignment.

As an example, one common PIC device is a distributed Bragg reflector semiconductor laser diode. This device consists of an active gain medium forming the lasing component, and a waveguide Bragg reflector to form a high-Q cavity surrounding the lasing region. This device, implemented on an indium phosphide (InP) chip, replicates the functionality of a traditional free-space lasers such as the helium neon (HeNe) laser, which also consists of a lasing medium (helium neon gas) and a high-Q cavity formed by two partially silvered mirrors. Unlike the HeNe device, further improvements to this device may be implemented lithographically in order to extend its capability at little additional cost and size.

For instance, a reflective diffraction grating may be used in place of the Bragg reflector to enable wavelength tuning of the laser via optical feedback from the grating into the laser cavity. In this system, referred to as a distributed feedback laser, the temperature of the grating may be changed in order to alter the diffraction angle for a particular wavelength, thereby altering the resonant wavelengths within the lasing medium. In addition to enabling wavelength tunability, on-chip laser devices may be further improved by line-narrowing through feedback from an external high-Q cavity, implemented directly on the chip. Such devices using ring-resonators, and optical whispering mode gallery resonators have been demonstrated to achieve better than 25 Hz laser linewidth on an integrated chip [32].

In this manner, PICs may be used to realize complex optical circuits using simple building blocks implemented directly on a semiconductor chip. In the telecommunications industry, in particular, it is not uncommon to develop optical circuits consisting of hundreds of discrete components. PIC coherent optical transceivers, for instance, may combine optical sources, phase and amplitude modulators, wavelength multiplexers, splitters, filters, amplifiers, delay interferometers, phase analyzers, and balanced photodiodes all on a single device no larger than a small coin [33]. Aside from telecommunications, PICs have also been used to implement complex optical circuits for fibre-optic gyroscopes [34], quantum information systems and entangled photon sources [35], and autonomous vehicle navigation [36].

Despite their potential applications within spectroscopy (*e.g.*, biomedical, chemical and gas sensing), PICs are generally used in a laboratory setting for in-situ analysis. Indeed, despite many demonstrations as a spectroscopic tool [37–41], PICs are rarely used for the interrogation of signals from standoff distances [41]. The dearth of remote-sensing systems employing PIC technology may be partially attributed to the reduced aperture available to PIC devices. Remote sensing systems require a high degree of light-gathering ability in order to ensure a precise measurement, however the étendue of many of the earliest PIC spectrometer designs, based on arrayed waveguide gratings or echelle gratings, is limited to that of one single-mode access waveguide [42]. In addition to throughput limitations, PICs are typically very sensitive to their thermal environment which limits their application in non-laboratory settings. These limitations are sufficient to prevent adoption of PIC spectrometers in remote distributed sensor network applications—such as pipeline monitoring—even in the face of their considerable advantages in manufacturability, cost, and size, weight and power (SWaP).

One potential solution to the limited étendue afforded by single-mode access waveguides is to adopt a Fourier-transform architecture, thereby enabling an expansion of the aperture to multiple waveguides and a corresponding increase in étendue by multiple orders of magnitude. On-chip FTS may be been realized using conventional Michelson interferometry with multi-mode waveguide channels and scanning mirrors propelled by micro opto-electromechanical systems actuators [43]. In this system, étendue is enhanced through the use of a large multi-mode waveguide cavity, which has significantly more light gathering area than a single-mode waveguide. PIC FTS may also be realized as a waveguide interferometer using thermal heaters to vary the optical path delay [38], however this architecture does not offer any advantage in terms of étendue. Finally, PIC FTS has been realized as a spatial heterodyne Fourier-transform spectrometer (SHFTS) consisting of an array of interferometers. This implementation enables a significant increase in étendue over the single-mode access waveguide solution, and does not require the use of moving parts, thereby simplifying the fabrication process and providing an additional measure of robustness to vibration.

It is this spectrometer architecture that has been pursued throughout this thesis. In this work the SHFTS architecture is preferred for its mechanical and electrical simplicity; in latter sections further improvements to the SHFTS architecture are realized through the use of compressive sensing—an improvement that would carry less advantages for non-SHFTS architectures. In Chapters 2 and 3, the theory of these devices is explored in greater detail as well as the design and fabrication of the two devices developed in this thesis. First, however, the following sections discuss general design principles of spectroscopic systems for various applications, trade-offs and figures of merit for various architectures, and summarize the implications for high-performance miniaturized spectroscopy.

1.2 Design of spectroscopic systems

In principle, spectroscopic instruments may be divided into four subcategories based on their method of spectral interrogation. These subcategories are listed below, and diagrammatic illustrations of these categories are provided throughout the chapter:

1. *Dispersive spectrometers* are designed using a dispersive element (*e.g.*, a prism or diffraction grating) to spatially separate wavelengths of light, which are then

captured using detectors placed at various positions within the dispersed beam.

2. *Filter-based spectrometers* incorporate a set of optical filters (*e.g.*, thin-film [44], quantum dots [45]), or a single variable filter (*e.g.*, a Fabry-Perot) to transmit narrow selections of wavelengths, each of which are then captured by a detector or set of detectors.

3. *Fourier-transform spectrometers* map the spectral content of incident light into the time domain by separating and re-combining incoming light to form a spatial interference pattern (from which spectral content is retrieved) which is captured by an array of detectors or a single moving detector.

4. *Active spectrometers* utilize a light source with known spectral content (*e.g.*, a tunable laser) to interrogate the spectral content of a scene by illuminating the scene and capturing the intensity of light that is reflected or transmitted at a particular wavelength.

From an designer's perspective, there is a substantial trade space that exists between these four instrument architectures including (but not limited to) instrument size, mechanical complexity, alignment tolerance, input spectrum, measurement of interest, optical throughput, and operational environment. From a scientific perspective, however, the most important performance parameters of spectroscopic systems are the measurement signal to noise ratio (SNR), and the system spectral resolution and range. In this regard, some typical applications of each archetype are presented in Table 1.1. The SNR drives the minimum uncertainty with which a measurement can be made, regardless of whether that measurement is the wavelength of a laser, the chemical content of a star, or the temperature of the ocean. The spectral resolution of the system denotes the ability of the system to differentiate between two spectrally-similar sources—which may have significant implications for the SNR of a measurement. While the quantity of "signal" in a particular measurement may be dependent on the instrument choice and the type of measurement being performed, the noise mechanisms in a spectroscopic signal are consistent across instrument architectures. Since these noise factors are relatively constant, and since the noise will drive SNR—the primary basis for comparison between instruments—any discussion of the advantages and disadvantages of different instrument types should begin with a discussion of noise in optical systems.

1.2.1 Noise in optical systems

Noise observed in detecting an optical signal is governed by a combination of external and internal sources including: fluctuations in the amount of signal and background photons entering the optical system (shot noise), the bulk motion of electrons inside the detector (Johnson noise), voltage noise in the electrical pre-amplifier (readout noise), and spontaneous emission from any optical pre-amplifying compo-

Table 1.1 Table of spectrometer architectures, advantages and disadvantages, and typical application fields

Architecture	**Advantages**	**Disadvantages**	**Applications**
Dispersive spectrometers	Simplicity, large spectral range, mechanically robust	Low resolution, reduced étendue	Astronomy [46], Raman and laser-induced breakdown spectroscopy [9, 47, 48], geology [6], soil analysis [23], mining and drilling [49], security [50], biology [51]
Filter-based spectrometers	Simplicity, enables imaging, mechanically robust	Low spectral range, limited spectral resolution	Imaging spectroscopy [13], gas leak detection [27, 52, 53], fluorescence spectroscopy [54], archaeology and art history [55], food inspection [56]
Fourier-transform spectrometers	High spectral resolution, large spectral bandwidth, large étendue	High mechanical and electrical complexity	Astronomy [57], atmospheric chemistry [26], foundry and industrial processing [58], analytical chemistry [4]
Active spectrometers	Very high resolution, high étendue	Limited spectral range, electrically complex	Atmospheric chemistry [59], gas leak detection [60]

nents (amplified spontaneous emission noise), and digitization noise in the case of an analog to digital converter (ADC).

A diagram of a generalized single-element photodetector with optical pre-amplification is presented in Figure 1.1. The diagram shows a series of photons passing through an optical gain medium (*e.g.*, a thin-film semiconductor optical amplifier [61], doped-fiber

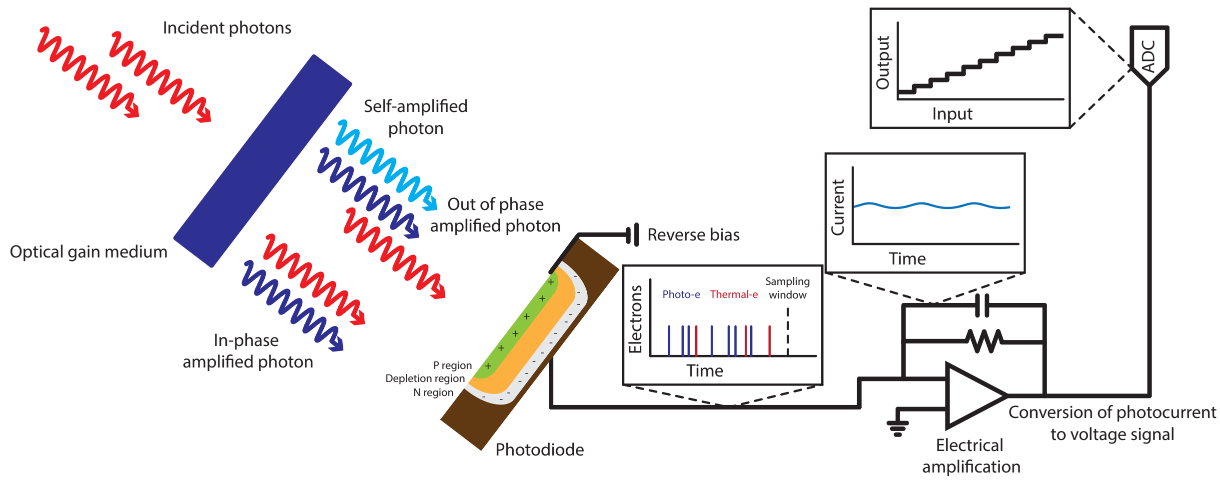

Fig. 1.1 Diagram detailing various sources of optical noise inherent to a generalized photo-detection scheme. Incoming photons at a stable average incidence rate are amplified by an optical gain medium, captured by a PIN photodiode and converted to photoelectrons. The photons are then converted to a voltage signal via a transimpedence amplifier, and converted to a digital value by an analog to digital converter. Various sources of noise throughout this process are detailed.

amplifier [62], or phosphor intensifiers [63]), striking a PIN photodiode and generating a photocurrent. This current is then pre-amplified and captured by an ADC. In a noiseless, perfectly linear system, the number of photons per second which enter the system would directly map to the voltage level collected at the ADC (the "signal"). In a real system, however, each element of this basic diagram will contribute some non-zero, fluctuating current, resulting in an imperfect voltage count at the ADC.

In the photon-counting regime, each element would contribute to spurious counts or non-counts over a given time-period. The constituent noise components in Figure 1.1 are described in the sections below, following their derivation in *Infrared Detectors and Systems* [64] and *Fiber-Optic Communication Systems* [65].

1.2.1.1 Shot noise

The most important source of noise to consider in Figure 1.1 is the shot noise or "photon noise". This source of noise derives from the quantized nature of light arriving at the photodetector; as such, the shot-noise places a fundamental limit on the SNR that may be achieved by any optical measurement scheme, hence its primacy over other sources of noise which may be mitigated through improved technologies or techniques. In Figure 1.1 a stream of photons is directed towards the photodetector, where they are converted to photoelectrons with probability η, corresponding to the external quantum efficiency of the detector.

Assuming that the average rate of arrival of these photons, P_{avg}, is constant, and the time of arrival of each photon is independent of the time of arrival of each other photon, then the number of photons captured by the detector is governed by Poisson statistics. Thus the probability of $N = k$ photons arriving at the detector over a fixed time interval t is given by

$$P(N = k) = \frac{e^{-P_{avg}t}(P_{avg}t)^k}{k!}. \tag{1.1}$$

In this probability distribution both the expectation value E[N] $\approx \bar{N}$, and variance Var[N] equal the average photon arrival rate multiplied by the time interval: $P_{avg}t$. The average current flow (*i.e.* roughly the expectation value) from the system in Figure 1.1 is given by

$$\bar{I} = \frac{q\eta\bar{N}}{\tau} \tag{1.2}$$

where q is the electron charge, and τ is the time interval recorded by the photodiode (*e.g.*, the integration time). Since the expectation value is set by the arrival rate of the photons, the shot variance ${I_s}^2 =$ Var[I] will be governed by the variance of the number of photons Var[N] $= \bar{N}$; therefore the variance may be written:

$$\begin{aligned} {I_s}^2 &= \frac{q^2\eta}{\tau^2}\bar{N} \\ &= \frac{q}{\tau}\bar{I}. \end{aligned} \tag{1.3}$$

Substituting the bandwidth of the detector $\Delta f = 1/(2\tau)$ and combining Equations 1.2 and 1.3, the SNR in the shot noise regime (*e.g.*, when shot noise is the dominant noise term) may be written as the average current (the signal, $\bar{I}$) divided by the standard deviation of the current due to shot noise, $\sigma_s = \sqrt{{I_s}^2}$

$$
\begin{aligned}
SNR_{shot} &= \frac{\bar{I}}{\sqrt{2q\bar{I}\Delta f}} \\
&= \frac{q\eta\bar{N}\Delta f}{\sqrt{q^2\eta\bar{N}\Delta f^2}} \\
&= \sqrt{\eta\bar{N}}
\end{aligned}
\tag{1.4}
$$

Since shot noise is intrinsic to the act of counting randomly-occurring events it can never be removed from an optical detection system. Indeed, since shot noise increases with the intensity of illumination, shot noise tends to dominate observations of highly luminescent sources. Shot noise can be mitigated in certain experiments by using squeezed photon states [66]. In these states the phase or frequency of the photons are tightly constrained, however these setups are generally not practical for general remote-sensing spectroscopy.

Note that this description of shot noise assumes a perfect detector with zero dark current. In reality, all detectors under a bias voltage produce a non-zero current even in the absence of any incident electrons. This resting current, termed “dark current”, is produced by the random generation of electrons and holes in the depletion region of the photodiode—as shown in Figure 1.1. These randomly generated electrons and holes then traverse the depletion region, facilitating a small current across the negative and positive contacts of the photodiode. Hole and carrier generation arises from the presence of crystal lattice defects in the detector material, diffusion from adjacent pixels, or from the detector substrate [64]. In each of these cases, the rate of electron-hole pair generation may be reduced by cooling the detector. This current, and its intrinsic shot variance, are indistinguishable from ordinary photocurrent and therefore sets a "baseline" photocurrent level that must be removed from any subsequent measurement. This correction is known as "dark frame subtraction", however since the variance of this dark current cannot be removed, the dark current must be accounted as a contributor to the shot noise in Equation 1.9.

1.2.1.2 Johnson noise

The source of noise to consider in spectroscopy are Johnson noise. Where the shot noise typically represents the limiting noise case for a bright source, Johnson noise and readout noise represent the limiting cases for dim observations. In Figure 1.1 Johnson noise manifests in the photodetector and associated electronics through thermal agitation of charge carriers inside the circuit. The magnitude of the Johnson current in the photodiode is determined by its temperature, electrical resistance, and its electrical bandwidth. Ordinarily, the photons incident on the photodetector should produce a current equal to:

$$\bar{I} = \frac{q\eta\bar{N}}{\tau}. \tag{1.5}$$

Superimposed on this current are small charge gradients caused by localized thermal carrier motion; the Johnson current variance, ${I_J}^2$, produced by these gradients is given by:

$$I_J{}^2 = 4k_BT/R_{PIN}\Delta f \tag{1.6}$$

where k_B is Boltzmann's constant, T is the temperature of the diode, and R_{PIN} is the load resistance applied to the photodetector. For dim light sources, in which the measurement is limited by Johnson noise, the SNR can be written

$$SNR_{Johnson} = \frac{q\eta\bar{N}/\tau}{\sqrt{k_BT/R_{PIN}\Delta f}} \tag{1.7}$$

Johnson noise is a common problem for infrared (IR), particularly far-infrared, spectroscopy as many light sources (*e.g.*, stars, galaxies, light from our own sun) are weakly-emitting in the NIR. Accordingly, for instruments operating in the Johnson noise limit, a common strategy to improve SNR is to decrease the temperature of the detector [64].

1.2.1.3 Readout noise

Another important source of noise in NIR spectroscopy (or, more generally, spectroscopy of any dim source) is readout noise. Readout noise is independent of the intensity of the source and, generally, the temperature of the detector therefore readout noise typically represents the fundamental sensitivity limit of the detector. The term "Readout noise" is a catch-all term that encompasses noise from amplification of the photoelectric current, ADC conversion of the current signal, and electronic switching noise picked up from the surrounding circuitry and power supply. In general, it is rare for this term to dominate the the noise of the measurement.

1.2.1.4 Amplified spontaneous emission

Amplified spontaneous emission (ASE) noise is a source of noise specific to instruments involving optical amplification. Such noise is common in optical communication systems [65], streak cameras employed in time-domain Raman spectroscopy [67], and phosphor-intensified spectrometers [68]. In each of these detection schemes, incoming signals are either amplified by a gain medium, or converted into photoelectrons, which are subsequently re-converted to photons and captured by a detector. In either case, photons pass through a medium of excited potential charge carriers which may be released downstream through their interactions with signal photons. Since the gain medium is excited, however, charge carriers will be spontaneously released (or de-excited via radiative recombination) on a stochastic basis. These spontaneously emitted charge carriers contribute to the noise when they are captured at the detector, but may also cause the release of further charge carriers from the gain medium—each of which also contributes to the noise.

1.2.1.5 Combining noise sources

Each of the above noise sources combine to reduce the SNR of the instrument. All noise sources are assumed to be independent meaning that an increase in shot noise does not necessitate increased Johnson noise and vice versa. Accordingly, the combined noise is the root sum of squares (RSS) of the individual components:

$$\sigma = \sqrt{{I_s}^2 + {I_J}^2 + {I_{RO}}^2 + {I_{ASE}}^2}. \tag{1.8}$$

The total SNR of the generalized detection scheme shown in Figure 1.1 is then:

$$SNR = \frac{q\eta\bar{N}}{\sqrt{2q(\eta\bar{N} + I_{dark})\Delta f + k_B T/R_{PIN}\Delta f + {I_{RO}}^2 + {I_{ASE}}^2}} \tag{1.9}$$

where the term I_{dark} is included to account for the shot noise of the intrinsic dark current of the detector under a bias voltage. The relative magnitude of each noise term will differ between different combinations of optical signals and detector schemes. For instance, the SNR from a bright source observed with a high-quality detector will likely be shot-limited, whereas the SNR from a dim source observed with a low-quality detector will likely be readout or dark-current limited. In the mid-infrared and far-infrared detectors are often Johnson-limited [64], and optically-amplified systems are almost always ASE-limited [65]. From the instrumentalist's perspective, then, it is important to understand the characteristics of the measurement of interest (*e.g.*, brightness of signal, wavelength of interest, temporal characteristics) as well as the characteristics of the available spectrometers and detectors (*e.g.*, étendue, quantum efficiency, detector noise), and to select a measurement scheme that maximizes overall SNR.

1.2.2 Spectrometer architectures

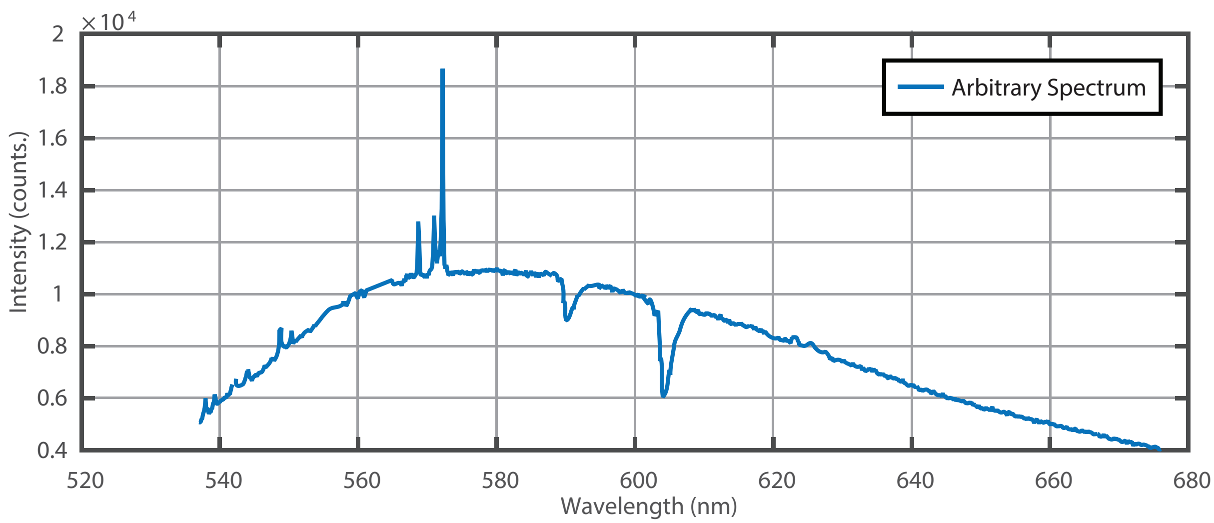

Fig. 1.2 Artificial spectrum containing emission and absorption features drawn for illustration purposes

Having discussed the mechanisms by which noise is generated in a spectroscopic system, it remains to address the meaning of the term "signal" and to differentiate between optical signal (*i.e.* flux incident on a detector) and the measurement signal—which is typically the width or depth of a particular spectral absorption or emission feature. Just as the noise in a spectroscopic system may be determined by detector selection, the signal at the detector and the signal of the measurement depends on the architecture of the spectrometer and the spectral characteristics of the measurement of interest. The spectrum in Figure 1.2 consists of a variety of unique features, each of which may have associated measurements of interest. For instance, the difference in intensity between the peak near 572 nm, and the background level near 580 nm may be used to infer the strength of an emission mechanism. The difference in intensity between the broad continuum (the smooth, slowly-varying curve) near 580 nm and the valleys (sharp declines in intensity) near 590 nm and near 605 nm may indicate absorption by a molecular species, and in this case could be used to determine the concentration of the absorbing species. The width of these features could also be used to determine the temperature profile of the absorbing species (*e.g.*, by Doppler broadening). Finally, the shape of the broadband background emission, spanning the full spectrum, may be compared to the theoretical emission curve of a blackbody to determine the temperature of the emission source.

In the first of these measurements, the signal of the measurement is the difference in intensity between two spectral points in Figure 1.2, and the noise is the noise associated with each of those points. The SNR of this example measurement may then be written:

$$SNR_{peak} = \frac{I_1 - I_2}{\sqrt{{\sigma_1}^2 + {\sigma_2}^2}} \tag{1.10}$$

where I_1 is the photocurrent observed at the "peak" of the emission line, I_2 is a measurement obtained of the background level, and σ_1 and σ_2 are the noise terms associated with each measurement. In this case, it would be natural to assume that the spectral points collected at the emission peak, I_1,would be shot noise limited, whereas the points at the base of the peak, I_2, would be dark noise or readout noise limited. In fact, the limiting noise of the measurement in each case may also be determined by the architecture and performance of the spectrometer.

1.2.2.1 Dispersive spectrometers

The most straightforward and common spectrometer design is the dispersive spectrometer. This design consists of three elements: (1) an entrance slit to constrain the spatial (angular) extent of the source (2) a grating or prism to disperse the source beam, introducing a mapping between wavelength and angle, and (3) an exit slit, or series of exit slits, to capture a small subset of angles (wavelengths) emanating from the dispersive element. Two representative diagrams of this system are shown in Figure 1.3.

The first design, a Czerny-Turner monochromator [69], uses a single entrance and exit slit and rotates the dispersive element in order to vary the angles that are

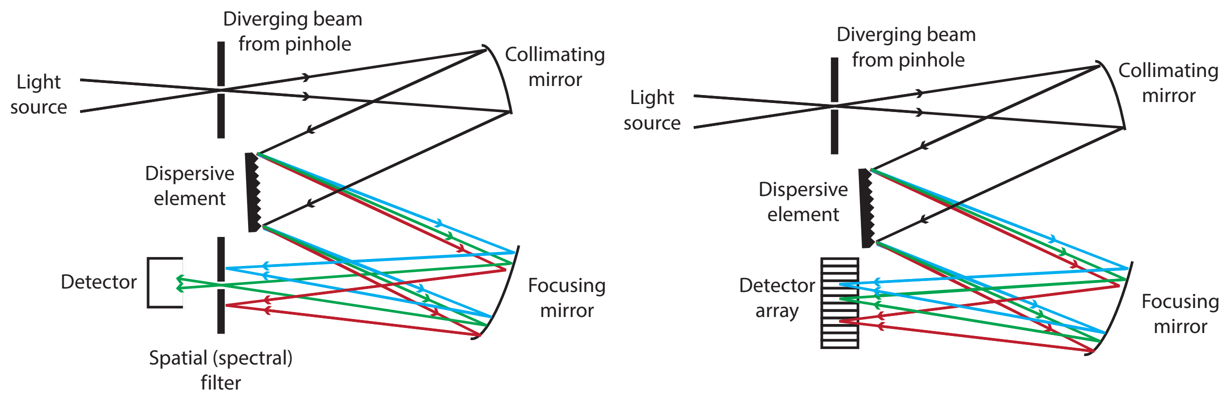

Fig. 1.3 Diagram of typical Czerny-Turner configurations. A monochromator (left) consists of an entrance slit, collimating mirror, dispersive element, focusing mirror, exit slit, and a single photodetector. Czerny-Turner spectrometers are also commonly implemented with a linear photodetector array (right) in which the spatial extent of the pixels in the array define a plurality of exit slits.

selected by the exit slit. Monochromators are, generally, impractical outside of a laboratory setting due to their large size, moving parts, and poor optical throughput, however as a very early version of a dispersive element spectrometer their underlying design is instructive and motivates the language used to describe more conventional spectrometers. The second design shown in Figure 1.3 is also a Czerny-Turner design, however a linear detector array is used at the exit plane rather than a single slit. In this case, the individual pixels of the linear detector array—having finite width—effect an array of exit slits.

The spectral resolution of a dispersive spectrometer is intrinsically limited by the f-number of the imaging optics at the exit of the instrument, as well as the physical size of the entrance aperture of the system. It is a property of imaging systems that their angular resolution is determined by their f-number ($F/\#$)—*i.e.* the focal length divided by the diameter of the aperture $F/\# = f/D$. Since the dispersive spectrometer is effectively a system that maps separation in the spectral domain to separation in the angular domain, the limiting angular resolution of the optical system necessarily impacts the limiting spectral resolution. The limiting angular resolution of an optical system (*i.e.* the ability to differentiate between two co-located point sources) is known as the Rayleigh criterion [70] and arises through the appearance of the Airy disk [71] diffraction pattern at the image plane. The angular function of the Airy disk pattern (an example is shown in Figure 3.10) is defined by a large central peak, surrounded by infinitely repeating maxima and minima. Two such Airy disk functions at the image plane are said to be resolved when the central maxima of the second pattern overlaps the first minima of the original pattern.

More commonly, for non point-source images, the spectral resolution of a dispersive instrument is limited by the physical size of the entrance and exit apertures along the dispersion axis of the instrument–with higher-resolution systems requiring smaller input slits. The angular extent of the entrance aperture, as seen from the image plane, is mapped to a spectral range by the angular dispersion of the grating or prism. Therefore, for a particular point on the image plane, and for each point across the entrance aperture, there exists some set of wavelengths such that each point in the entrance aperture is mapped to that same point on the image plane. Similarly, for a particular point on the entrance aperture, and for each position across the exit aperture, there exists some set of wavelengths such that the point at the entrance aperture may be mapped to each position on the exit aperture. The spectral range of these two sets of wavelengths therefore determines the minimum spectral resolution of the instrument.

The noise characteristics for a Czerny-Turner system are straightforward. In the case of the emission feature shown Figure 1.2, the signal and noise associated with the peak and background measurements will be determined by the integrated spectral flux over the aperture-limited resolution of the instrument. It should be noted that as the spectral resolution of the instrument is reduced, and each sampling point integrates a greater range of spectral flux, the noise value of each sample will increase with the shot noise, but the differential term in Equation 1.10 will remain constant. For this reason, it is generally preferred to achieve spectral resolution and sample density greater than the intrinsic line width of the measurement of interest.

Assuming a dispersive system with high spectral resolution, the noise statistics will proceed as previously discussed (*i.e.* the "peak" is shot-limited, and the "background" is dark or readout-limited). This may be inferred by the linear array instrument diagram shown in Figure 1.3. The photodiode in the detector array corresponding to the peak wavelength will be strongly illuminated, whereas the surrounding diodes will be weakly illuminated. This spectrometer configuration enables a high SNR measurement for emission lines, particularly if the background level is averaged across multiple pixels. Conversely, dispersive systems may be poorly-suited to capturing the absorption features in the far wavelengths of Figure 1.2. In this case, it is possible that the readout and dark current noise define a noise floor below which the depth of an absorption feature cannot be determined.

1.2.2.2 Filter-based spectrometers

Another common spectrometer architecture, illustrated in Figure 1.4, uses a spectral filter—or multiple filters—to prevent unwanted wavelengths from reaching the photodetector while admitting only a specific range of wavelengths ("bandpass"). The spectral sorting element itself may consist of a thin-film interference filter, fiber Bragg-grating, tunable acousto-optical element, or Fabry-Perot cavity, among others. To some extent, a dispersive spectrometer may also be considered a filter-based spectrometer in that the linear detector array performs spatial filtering of the incoming light (which has undergone spectral—spatial mapping via the dispersive element). For the purpose of this discussion, however, only spectrometers performing direct spectral (*i.e.* non-spatial) filtering are considered.

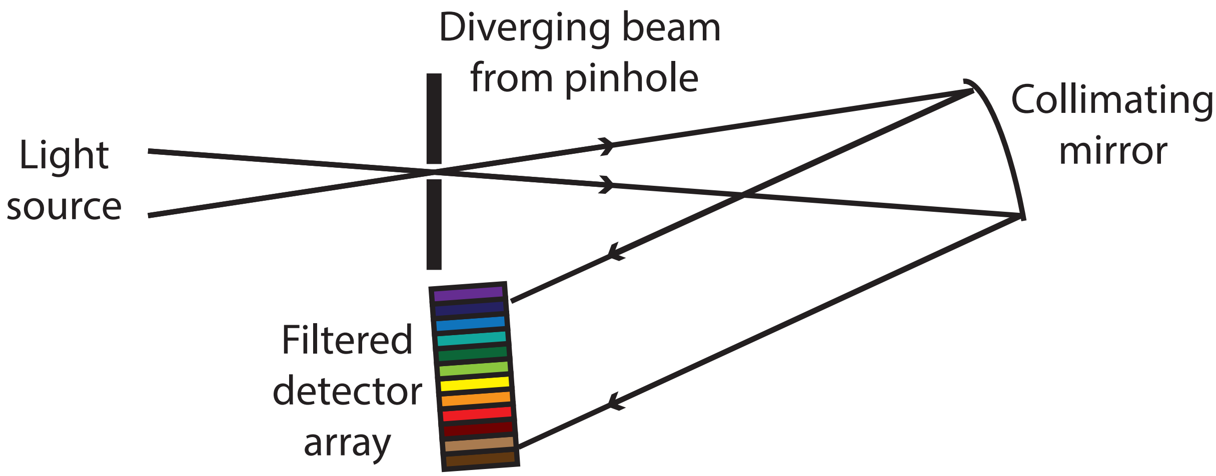

Fig. 1.4 Diagram of a simple filter-based spectrometer. The spectrometer consists of an entrance slit, collimating mirror, and a linear photodetector array, each element of which is preceeded by a unique spectral bandpass filter.

Filter-based spectrometers have the advantage that, unlike dispersive systems, the spectral resolution of their measurement scheme is not tied to the entrance aperture of the instrument. This design freedom may enable systems with substantial entrance apertures and high spectral resolution [72]. Unfortunately, filter-based spectrometers may not achieve continuous, high spectral resolution measurements across a wide bandwidth without either breaking the instrument into several subapertures (each with its own filter), or by time-multiplexing measurements using several filters (*e.g.*, a tunable Fabry-Perot or filter wheel [72]).

For this reason, though filter-based spectrometers may possess physically larger entrance apertures than dispersive systems, the effective aperture is limited by the number of sub-apertures (spatial or temporal) required to cover the bandwidth of the measurement. The SNR mechanisms of a measurement collected with a filter-based spectrometer are similar to that of a dispersive spectrometer. In the emission-line measurement example above, the signal and noise terms of the emission peak and background level will both proceed identically to the dispersive spectrometer.

As discussed, filter spectrometers may achieve substantially larger entrance apertures than dispersive systems, at the cost of measurement density. This restriction makes filter-based spectrometers a poor choice for interrogation of a large bandwidth signal such as the full spectrum presented in Figure 1.2. However, for investigation of predictably-situated spectral features (*e.g.*, molecular absorption or emission lines [13, 73]), filter spectrometers offer a compelling architecture.

Further architectures based on filter spectrometers are beyond the scope of this thesis, however many interesting variants exist including gas filter correlation spectroscopy [74], and laser heterodyne spectroscopy with radio-frequency filtering [75].

1.2.2.3 Fourier-transform spectrometers

An alternate spectrometer architecture, substantially different to dispersive and filter-type systems is the Fourier-transform spectrometer (FTS). In a general FTS system, the incoming light is separated into two beams, which are offset by a relative phase delay, then re-combined. Spectral components of the recombined beam then

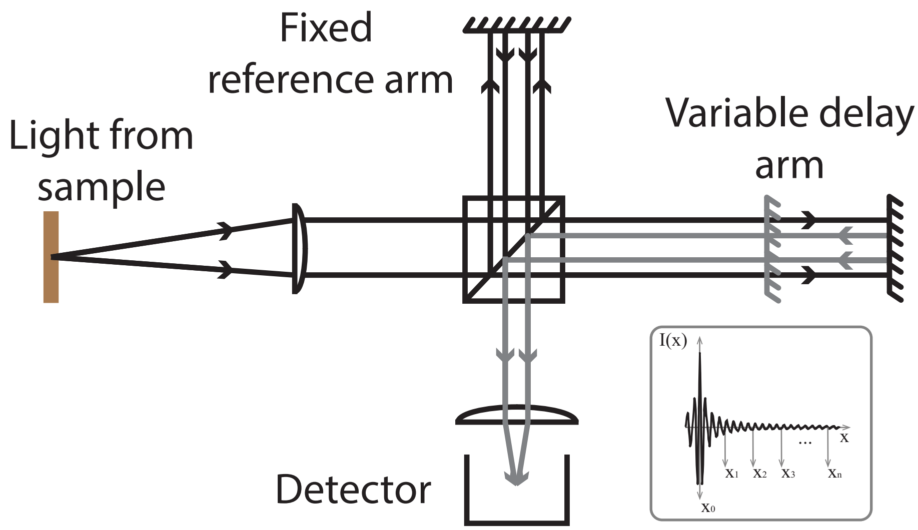

Fig. 1.5 Diagram of a Fourier-transform spectrometer based on a Michelson interferometer architecture. Light from the sample enters a collimating lens, and is separated into a reference and delay arm before being recombined to form an interference pattern (inset) at a single photodetector.

interfere constructively and destructively in accordance with the product of their wavelength and relative phase delay. The integrated intensity of these interfering spectral components then corresponds to a single component of the Fourier transform of the input spectrum.

In the system shown in Figure 1.5, which is described in greater detail in Chapter 2, the photodiode acts as a spectral integrator, collecting the absolute magnitude of the Fourier component corresponding to the particular phase delay. In a traditional FTS system, a linearly translating mirror is used to vary the phase delay of the interferometer, and the photodiode collects a set of Fourier components which effect the discrete Fourier-transform of the input spectrum.

Like the filter-based spectrometer, an FTS is not restricted in terms of its input aperture; this enables high-resolution FTS systems to achieve significant étendue over comparable dispersive spectrometers. The freedom to increase the input aperture of an FTS instrument is referred to as the "Jacquinot" advantage [76]. Also like the filter-based spectrometer, an FTS must use spatial or temporal sub-aperturing in order to collect a spectrum over a broad range of wavelengths—which may ultimately restrict the integration time available for each measurement. Unlike the filter-based and dispersive spectrometers, however, an FTS collects substantially more optical signal in each sampling point since the photodiode integrates over the full spectral range of the input signal. This advantage of the FTS scheme is referred to as Fellgett's advantage, or the "multiplex" advantage, and may considerably improve the SNR for optically faint input signals [77].

Historically, FTS is deployed in the signal-starved IR regime, where detectors tend towards worse performance in terms of dark current. IR detector arrays are also often more expensive, and offer reduced pixel density and performance than their UV-visible counterparts. Given the poor availability of IR detector arrays, then, the use of a Michelson architecture allows the instrument designer to make use of a single IR photodiode, which may be of higher quality than the diodes in an array.

Even if the FTS uses a low-quality photodetector, the SNR may still be improved through the multiplex advantage, since each point in the reconstructed spectrum is generated through measurement of the the total integrated power of the spectrum. By collecting the integrated power, an FTS may ensure that each measurement in the reconstructed spectrum is effectively shot-noise limited—even in the case of significant Johnson noise in the detector.

In the case of the emission feature in Figure 1.2, the noise associated with each reconstructed point will be equivalent to the shot-noise of the average optical power in the spectrum. For an FTS spectrum consisting of N measurement points, then, the SNR at the peak of the emission line is increased by a factor of $\sqrt{N}$ over the same measurement collected by a dispersive spectrometer.

The advantages of FTS make it a compelling architecture—especially for measurements of faint signals where high-quality detectors are not readily available. However, in addition to the optical noise sources discussed here, there are secondary noise factors in FTS, including phase-noise, and noise due to sampling non-uniformity, that are not present in dispersive or filter-based instruments. As will be discussed further in Chapter 2, the measurements obtained by an FTS must obey strict sampling criteria in order to produce an accurate reconstruction of the original optical spectrum. Noise arising from imperfect sampling—which is unavoidable in a real system—must therefore be considered along with shot, dark, Johnson, and readout noise.

1.2.2.4 Active spectrometers

Finally, for the purpose of this dissertation, active spectroscopy refers to a family of techniques in which a narrowband laser source, or multiple sources, are used to probe the spectral properties of a sample at a particular wavelength by analyzing the reflected or transmitted signal. One common implementation of this technique is differential absorption LIDAR spectroscopy (DIAL) in which a single photodetector, or a pair of photodetectors is used to interrogate the relative intensities of optical pulses that are emitted by a laser and reflected by the sample of interest. Such systems typically target a known spectral feature such as the absorption line shown in Figure 1.2 as the "on-line" measurement, and a nearby spectral point as the "off-line" reference measurement. Another common architecture for active spectroscopy is tunable laser-diode absorption spectroscopy (TLDAS), in which the wavelength of a single laser diode is continuously swept across a single spectroscopic feature, and the shape and depth of that feature is determined by repeated sampling of the transmitted or reflected optical signal as a function of wavelength.

Note that active spectroscopy may also include Raman and LIBS techniques. In Raman and LIBS spectroscopy, however, the optical detection system does not differ greatly from the afore-mentioned architectures. Raman and LIBS signals are commonly recorded by dispersive [48] and FTS [78]; and the SNR of the measurement is therefore well-described by the noise characteristics of these receivers. DIAL and TLDAS systems, by contrast, typically consist of a single photoreceiver with an associated bandpass filter for background rejection—and are more sensitive to noise in the excitation source.

In the case of a DIAL system, the optical SNR may be equivalent to an ordinary filter-based spectrometer, however, depending on the molecular or atomic species of interest, the measurement SNR may be closely related to the wavelength stability and repeatability of the excitation laser, or lasers. Since DIAL collects point measurements rather than a full spectrum, any wavelength drift in the "on-line" signal away from the peak of the absorption signal may have a profound effect on the intensity of the collected measurement. Similarly, if the "off-line" signal is not suitably separated from the absorption feature of interest, then wavelength variation in the excitation source may cause undesirable drift in the intensity of the background reference. Finally, variations in the intensity of the excitation source (*i.e.* due to temperature drifts and mode-hops in the laser cavity, or due to amplifier noise in the case of amplified DIAL systems) may also contribute to decreased measurement SNR. Intensity variations may be mitigated by monitoring the intensity of the excitation source and normalizing the return signal—however, the measurement is still constrained by the uncertainty of the monitoring scheme. Note that mode-hops in the excitation wavelength of a Raman system may also affect the measurement by introducing an equal spectral shift in the emission spectrum. This spectral shift, however, may be well below the inherent resolution of the measuring spectrometer.

DIAL systems may also be subject to increased levels of amplification noise relative to other filter-based spectrometers. Atmospheric DIAL systems typically deploy highly-sensitive "photon-counting" detectors such as avalanche photodiodes [59] or photo-multiplier tubes (PMTs) [79] that provide high-resolution timing information on the return signal. Such detectors—particularly infrared avalanche photodiodes—may inject significant amplification noise into the measurement [80].

Despite these additional sources of noise, active spectrometers may achieve substantially improved sensitivity over passive systems thanks to their significantly improved spectral resolution and—in the case of TLDAS—increased optical intensity. Active spectrometers may also offer additional system-level advantages, such as ranging information in the case of DIAL.

1.2.3 Throughput and étendue

As discussed, the traditional advantage of FTS over dispersive element spectroscopy is that an FTS may achieve greater optical throughput (étendue) than a dispersive grating instrument for the same spectral resolution. The throughput advantage of an FTS arises from the relationship between the spectral resolution of a dispersive grating instrument and the width of the entrance and exit slits of the spectrometer. Specifically, as the spectral resolution of a dispersive instrument increases, the width of the entrance slit must decrease in order to avoid spectral overlap at the pixel array (exit slit). By contrast, the throughput of the FTS is limited only by the diameter of the collimated ray bundle within the FTS—which is typically much larger than the entrance slit of the dispersive instrument. Since the spectral resolution of the FTS is limited by the maximum optical path delay (OPD) in the interferometer, the FTS may, in theory, achieve arbitrarily high spectral resolution without affecting optical throughput. In reality, the diameter of the collimated beam (the Jacquinot stop) and the spectral resolution may be limited at large apertures by chromatic aberrations that lead to imperfect collimation over long OPDs, however this effect is negligible in comparison with the throughput advantage gained by eliminating the entrance slit.

The throughput for an arbitrary diffraction grating spectrometer consisting of an input slit, dispersive element, focusing optics, and a linear detector array may be calculated for a particular spectral resolution as follows [81]. First the orientation of the grating may be described by the difference between the incidence angle, α_G, and the diffraction angle, β_G, as $D_v = \beta_G + \alpha_G$, where α_G and β_G are related through the diffraction equation:

$$\sin \alpha_G + \sin \beta_G = m\lambda/d \tag{1.11}$$

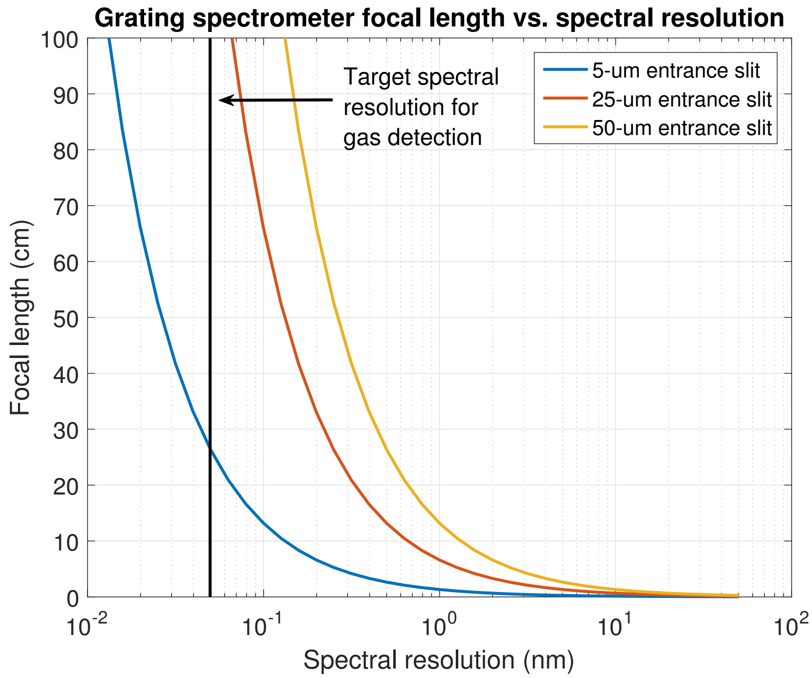

Fig. 1.6 Diagram showing the significant increases in focal length (e.g. spatial offset) required for a Czerny-Turner spectrometer to achieve sub-nm spectral resolution. For gas detection via identification of molecular absorption features, resolution < 0.05 nm is desired, which would require focal lengths in excess of 20 cm for achievable slit sizes. The calculation assumes 3-μm groove spacing (333 lines / mm), a grating angle of 24°, and that the grating selects the first diffracted order.

where m is the diffraction order (typically 1), λ is the central wavelength of the spectrometer, and d is the spacing of the diffraction grating. Assuming a monochromator configuration featuring a concave aberration-corrected grating, the incidence angle may be calculated as:

$$\alpha_G = \sin^{-1}\left(\frac{m\lambda}{2d\cos(D_v/2)}\right) - D_v/2. \tag{1.12}$$

Assuming a pixel width of $W_{pix} = 25\ \mu$m—a common value for infrared detectors—and that focal length of the entrance optics equals the focal length of the exit optics, then in order to achieve $\delta\lambda = 0.050\ nm$ the exit focal length of the spectrometer, f_{ext}, and entrance slit width, W_{ent}, may be determined as:

$$f_{ext} = \frac{d\cos\beta_G\, W_{pix}}{m\delta\lambda} \tag{1.13}$$

$$W_{ent} = \frac{m f_{ext}\delta\lambda}{d\cos\alpha_G}. \tag{1.14}$$

In Figure 1.6, the relationship between spectral resolution, slit width and focal length is shown. As discussed, at higher spectral resolution the focal length increases and slit width decreases—both of which constrain the optical throughput of the system by reducing light gathering. The total étendue of a diffraction grating spectrometer may be estimated as:

$$G_{spec} = \frac{h_{slit}A_{grating}m\delta\lambda}{d\, f_{ext}} \tag{1.15}$$

where $A_{grating}$ is the total illuminated area of the grating itself and h_{slit} is the height of the entrance slit.

Taking these relationships together, a grating spectrometer achieving 50 pm spectral resolution using an InGaAs detector array with 25 μm pixels at $D_v = 24$ degrees would require an exit (and entrance) focal length of 1.3 m—an impractical length for a miniaturized instrument. The substantial focal length required to achieve this spectral resolution will also reduce the optical throughput, which may be re-written in terms of the entrance slit width as:

$$G_{spec} = \frac{h_{slit} W_{ent} A_{grating} \cos \alpha_G}{{f_{ext}}^2}. \tag{1.16}$$

By contrast, the light-gathering power of the methane spectrometer presented in Chapter 4 may be determined by the sum of the individual étendue values for each of the 100 waveguides

$$G_{wg} = \pi A_{mode} {NA_{wg}}^2 \tag{1.17}$$

where A_{mode} is the effective area of the waveguide mode, and NA_{wg} is the numerical aperture of the waveguide. The numerical aperture of the waveguide spectrometer in Chapter 4 was measured to be 0.58, and the effective area of the waveguide mode is $A_{mode} = 1.4 \times 10^{-6}$ mm^2. Over the full set of 100 waveguides, this yields étendue $G_{chip} = 1.5 \times 10^{-4}$mm^2. Per Equation 1.15, a dispersive grating spectrometer achieving the same throughput and spectral resolution would require an entrance slit measuring 0.022 mm $\times$ 23 mm, a grating measuring 23 mm $\times$ 23 mm, and exit focal length $f_{ext} = 1.32$ m; an impractical set of dimensions for a miniaturized instrument.

2 Principles of Fourier-transform spectroscopy

In this chapter the principles of Fourier-transform spectroscopy are discussed, with particular focus on their application to on-chip and compressive sensing Fourier-transform devices. The background information provided in this chapter is generally derived from a few excellent references on the subject, particularly *Understanding FT-IR data processing* by Herres [77] and *Spectral imaging of the atmosphere* by Shepherd [82].

2.1 Principles of Fourier transform spectroscopy

To understand the implementation of an on-chip FTS it is first necessary to understand the Fourier-transform and its realization in optical systems. The Fourier-transform (FT) and its inverse are the pair of transformations that relate a signal in the time domain to its equivalent expression in the frequency domain and vice-versa. The FT and the inverse FT are given by:

$$F(\nu) = \int_{-\infty}^{\infty} f(t) e^{-i2\pi t\nu} dt \tag{2.1}$$

$$f(t) = \int_{-\infty}^{\infty} F(\nu) e^{i2\pi t\nu} d\nu, \tag{2.2}$$

where, $f(t)$, is the function expressed in the time domain, and $F(\nu)$ is its equivalent in the frequency domain. At each frequency, ν, the FT expression in Equation 2.1 computes the magnitude of the projection of an infinitely long time-domain signal along a complex sinusoid with frequency ν. There are a few important properties of this transform that are relevant to this work. These are the properties of incoherence, convolution, and the preservation of the Dirac comb.

1. *Incoherence*

   In the first special case it may be intuited that a signal that is well-confined in the time-domain will be poorly confined in the frequency-domain, and vice-versa. For instance, consider a delta function in the frequency domain, $F(\nu) = \delta\nu - \nu_0$. In this case the inverse FT of $F(\nu)$ results in a single sinusoid with frequency ν_0:

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} \delta(\nu - \nu_0) e^{i2\pi t\nu} d\nu \\ &= e^{i2\pi\nu_0 t}. \end{aligned} \tag{2.3}$$

This property is sometimes referred to as the uncertainty principle, or as the property of incoherence in the field of compressive sensing [83, 84].

2. *Preservation of the Dirac comb*

The second important property of the FT is the invariance of a Dirac comb under transformation. The Dirac comb consists of an infinite series of uniformly-spaced Dirac delta functions. In the time domain the Dirac comb may be written:

$$\mathrm{III}_T(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT), \tag{2.4}$$

where T is the periodicity of the comb or the time between adjacent delta functions.

From Equation 2.3, the FT of the Dirac comb will be described by an infinite sum of complex sinusoids:

$$F(\nu) = \sum_{k=-\infty}^{\infty} e^{i2\pi\nu kT}. \tag{2.5}$$

Considering $F_{III} = 1/T$, then Equation 2.5 can be written:

$$F(\nu) = \sum_{k=-\infty}^{\infty} e^{i2\pi\nu k/F_{III}}, \tag{2.6}$$

in which case, if νk is any integer multiple of F_{III}, then $e^{i2\pi\nu k/F_{III}} = 1$, and Equation 2.6 will sum to infinity. In the case where νk is not an integer multiple of F_{III}, then Equation 2.6 may be re-written as a combination of geometric series of the form $(e^{i2\pi\nu/F_{III}})^k$:

$$F(\nu) = \sum_{k=0}^{\infty}(e^{i2\pi\nu/F_{III}})^k + \sum_{k=0}^{\infty}(e^{-i2\pi\nu k/F_{III}})^k - (e^{i2\pi\nu/F_{III}})^0, \tag{2.7}$$

Since we are only concerned with cases where νk is not an integer multiple of F_{III}, then we can say that the radius of the geometric series, $|r| = |(e^{i2\pi\nu/F_{III}})|$ is $|r| < 1$, hence the sum of each geometric series is reduces to $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$, or

$$\begin{aligned} F(\nu) &= \frac{1}{1 - e^{i2\pi\nu/F_{III}}} + \frac{1}{1 - e^{-i2\pi\nu/F_{III}}} - 1 \\ &= \frac{2 - \left(e^{i2\pi\nu/F_{III}} + e^{-i2\pi\nu/F_{III}}\right)}{2 - \left(e^{i2\pi\nu/F_{III}} + e^{-i2\pi\nu/F_{III}}\right)} - 1 \\ &= 0. \end{aligned} \tag{2.8}$$

Hence, the sum of complex sinusoids in Equation 2.5 produces a function that is zero at all points, save for integer multiples of $F_{III} = 1/T$, at which point the function reaches infinity. This is the definition of a Dirac comb with spacing

$1/T$, thus the FT of a Dirac comb with periodicity T is another Dirac comb with periodicity $1/T$.

3. *The convolution theorem*

A third important property of the FT is summarized by the convolution theorem. While addition and scalar multiplication operations are invariant under FT (*e.g.*, $FT[a \cdot f(t) + b \cdot g(t)] = a \cdot F(\nu) + b \cdot G(\nu)$), the convolution theorem states that when transforming between the time and frequency domains, a convolution in one domain is equal to pointwise multiplication in the other. (*e.g.*, $FT[f(t) \cdot g(t)] = F(\nu) * G(\nu)$) The convolution of one function with respect to another, $f * g$ may be written:

$$f * g = \int_{-\infty}^{\infty} g(T) f(t - T) dT. \tag{2.9}$$

Taking the FT of $f(t - T)$, this expression becomes:

$$\begin{aligned} f * g &= \int_{-\infty}^{\infty} g(T)[\int_{-\infty}^{\infty} F(\nu) e^{i2\pi\nu(t-T)} dv] dT \\ &= \int_{-\infty}^{\infty} F(\nu)[\int_{-\infty}^{\infty} g(T) e^{-i2\pi\nu T} dT] e^{-i2\pi\nu t} d\nu \\ &= \int_{-\infty}^{\infty} F(\nu) G(\nu) e^{-i2\pi\nu t} d\nu \\ &= FT^{-1}[F(\nu) G(\nu)]. \end{aligned} \tag{2.10}$$

The expression in Equation 2.10 is the mathematical equivalent to the previous statement that a convolution in one domain (in this case, time) is equivalent to performing a multiplication in the other domain (frequency).

These three properties will ultimately impart an effect on the performance of a FTS instrument, as detailed in Section 2.1.2.

2.1.1 Optical Fourier-transform via interferometry

Although the FT presented above is a mathematical operation, the use of an FTS requires the implementation of an *optical* Fourier-transform. An optical FT may be performed using an interferometric system such as the Michelson interferometer shown diagrammatically in Figure 1.5. In this system, the 1-D travelling wave solution to Maxwell's equations may be used to describe the evolution of the electric field of a photon with wavelength λ:

$$\vec{E}(x,t) = \vec{E}_0 e^{i2\pi(\sigma x - \nu t)}, \tag{2.11}$$

where x is the linear position in cm, and $\sigma = 1/\lambda$ is the optical wavenumber in cm^{-1}. Assuming a perfect beam splitter, the electric field at the detector in Figure 1.5 may then be described as the addition of the electric field from the two paths of the Michelson, one of which has experienced a time delay Δt:

$$\vec{E}_{det}(t) = \frac{1}{\sqrt{2}}\vec{E}_0(e^{i2\pi(\sigma x - \nu t)} + e^{i2\pi(\sigma x - \nu(t+\Delta t))}). \tag{2.12}$$

The intensity captured at the detector is given by $I_{det}(t) = |\vec{E}_{det}(t)\vec{E}^*_{det}(t)|$. Expanding the even terms in Equation 2.12 will amount to a zero-frequency (DC) offset, while the cross terms produce a sinusoidal modulation:

$$\begin{aligned} I_{det}(\Delta t) &= \frac{1}{2}\vec{E}_0\vec{E}^*_0(2 + e^{i2\pi(\sigma x - \nu t)}e^{-i2\pi(\sigma x - \nu(t+\Delta t))} + e^{i2\pi(\sigma x - \nu(t+\Delta t))}e^{-i2\pi(\sigma x - \nu t)}) \\ &= \frac{1}{2}\vec{E}_0\vec{E}^*_0(2 + e^{i2\pi\nu\Delta t} + e^{-i2\pi\nu\Delta t}). \end{aligned} \tag{2.13}$$

Recalling Euler's identity, $e^{i\phi} = \cos\phi + i\sin\phi$, Equation 2.13 may be written:

$$I_{det}(\Delta t) = I_0(1 + cos(2\pi\nu\Delta t)). \tag{2.14}$$

If I_0 is some polychromatic spectrum $I_0(\sigma)$, and recalling that $\sigma = \nu/c$, and making a substitution for the time delay such that $\Delta t = \Delta x/c$, then the photodetector, which serves as an optical integrator, will record intensity:

$$I_{det}(\Delta x) = \int_{\sigma_0}^{\sigma_N} I_0(\sigma)(1 + cos(2\pi\sigma\Delta x))d\sigma, \tag{2.15}$$

where $\sigma_0 < \sigma < \sigma_N$ defines the optical bandwidth of the detector. Since σ is a unit of frequency, the second term in Equation 2.15 is equivalent to the real part of the inverse FT of $I_0(\sigma)$. It follows that if the continuous inverse FT of the spectrum as a function of distance (the interferogram $I_{det}(x)$) is known, then the spectrum may be recovered via FT:

$$I_0(\sigma) = \int_{-\infty}^{\infty} (I_{det}(x) - I_{det}(0)/2) e^{-i2\pi\sigma x} dx. \tag{2.16}$$

For reference, these operations are illustrated in Figure 2.1. It should be noted that the real component of the solution to Equation 2.16, which is governed by the cosine modulation, is equally valid at both positive and negative frequencies. As a result, the power spectrum generated by the FFT for a given interferogram will be "mirrored" about zero. The values of the imaginary (sinusoidal) components of the solution to Equation 2.16 will also be mirrored about zero, however the imaginary values will take their conjugate form since $\sin(x)$ is an odd function. The negative frequency solutions are assumed not to add any new information and are often disregarded. Notwithstanding, the mirrored spectrum is a very real solution to Equation 2.16 and as such will have a direct impact on the operation of the instrument, as described in the following section.

To summarize: the basis of the Fourier-transform spectrometer is an optical arrangement designed to render optical interference at a detector, generating a sinusoidal modulation of the input spectrum governed by the power spectrum of incoming light and time-delay applied by the interferometer. The resulting interference intensity, as a function of the time-delay, is then equivalent to the inverse FT of the input spectrum, which may be re-generated via forward FT.

2.1.2 Sampling, aliasing and the discrete Fourier-transform

Having established a theoretical understanding of the FT and its optical equivalent it remains to discuss the practical implementation of such a system, and the effect of the afore-mentioned theories on the operation of a practical system.

2.1.2.1 Sampling and spectral bandwidth

One practical consideration for a real FTS is the discontinuous nature of any interferogram that is collected by digitized readout electronics. Equation 2.16 generally describes an interferogram that is collected over the continuous interval $-\infty < x < \infty$. However, the Michelson pictured in Figure 1.5 will collect its samples at some finite step resolution $x_{i+1} - x_i \approx \delta x$; moreover, it will necessarily operate over some range $x_{min} < x < x_{max}$. These two considerations: (1) discontinuous sampling and (2) truncation of the interferogram, will independently affect the FT of the interferogram.

In the case of discontinuous sampling, the interferogram is not collected as a smooth function. Instead, owing to the finite integration and readout times of the photo detector, as well as the smoothness of the motion, individual points in the interferogram $(I_{det}(x_0), I_{det}(x_1), ..., I_{det}(x_{N-1}), I_{det}(x_N))$ are read out at regularly-spaced intervals such that $x_{n+1} = x_n + \delta x$. Mathematically, this is equivalent to multiplying the continuous interferogram $I_{det}(x)$ by a Dirac comb:

$$I_{disc}(x) = \text{III}_{\delta x}(x) \cdot I_{det}(x). \tag{2.17}$$

From the convolution theorem, the multiplication of the continuous interferogram by a Dirac comb in the spatial (time) domain is equivalent to a convolution of the FT of the spectrum and the FT of the Dirac comb in the spectral (frequency) domain. As the Dirac comb is preserved under transformation, the resulting spectrum may be written:

$$FT[I_{disc}(x)](\sigma) = \int_{-\infty}^{\infty} I_0(\sigma')\text{III}_{1/\delta x}(\sigma - \sigma')d\sigma'. \tag{2.18}$$

This operation is shown diagrammatically in Figure 2.1. A basic property of the convolution operation is the convolution of some function $f(\sigma)$ with the shifted Dirac delta $\delta(\sigma - \sigma')$ yields the original function shifted by the location of the delta: $f(\sigma)*\delta(\sigma-\sigma') = f(\sigma-\sigma')$. Thus the convolution of the spectrum $I_0(\sigma)$ with an infinite train of delta functions in Equation 2.18 yields an infinite train of reproductions of $I_0(\sigma)$ shifted in frequency by $k/\delta x$. This effect is called "Aliasing", and ultimately defines the spectral bandwidth ($\Delta\sigma$) of the FT.

At each Dirac delta in the infinitely repeating train, the input spectrum and its negative frequency solution, will be faithfully reproduced. This is shown diagram-

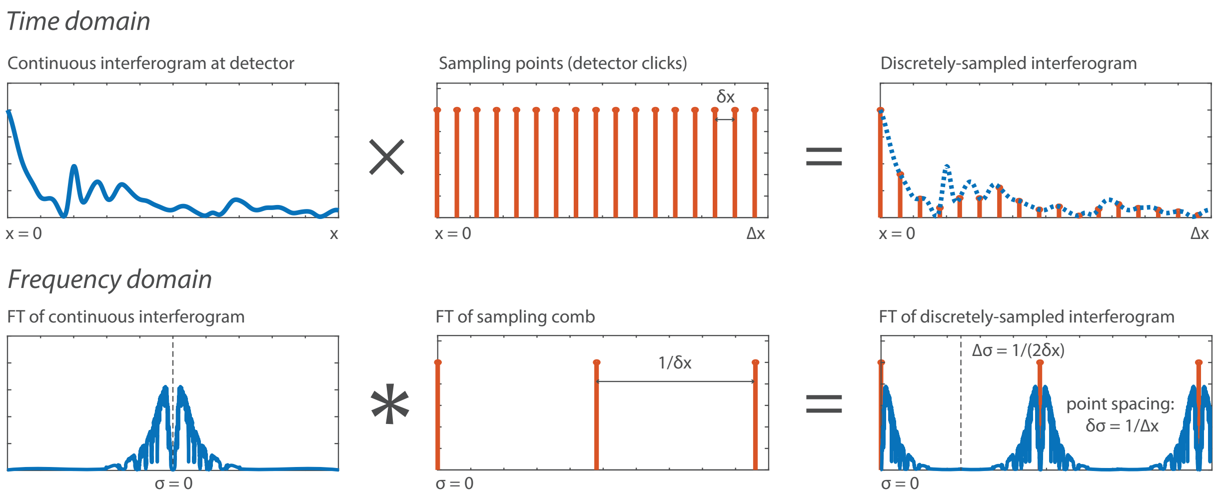

Fig. 2.1 Diagram illustrating the convolution theorem in time and frequency domains. In the time domain (top) a continuous interferogram is defined by the photodiode signal as a function of the displacement of the delay arm. This continuous function is sampled (multiplied) by a Dirac comb with minimum step-size δx resulting in a discretely-sampled interferogram. In the frequency domain (bottom) the input spectrum and its negative solution are determined by the transform of the continuous interferogram. This solution is then convolved with the transform of the sampling comb, resulting in aliasing or "folding" of the original spectrum.

matically in Figure 2.1. Thus, if the double-sided bandwidth of the input spectrum exceeds the periodicity of the Dirac comb ($\Delta\sigma > 1/(2\delta x)$), then the positive fold of the spectra at σ_i will overlap with the negative fold at σ_{i+1}. Overlapping orders in an FT spectrum results in ambiguity in the retrieved spectrum, as it is not possible to differentiate between the contributions of the spectrum and its alias. For this reason, the bandwidth of the single-sided input spectrum is limited to less than half the Dirac comb spacing (e.g. $\Delta\sigma < 1/(2\delta x)$).

The frequency at which the positive-sided and negative-sided spectrum meet is the "Nyquist" or "folding" frequency, and an FTS designed such that $\Delta\sigma = 1/(2\delta x)$ is said to be sampling at the Nyquist limit [77,85]. Typically, the spectral bandwidth of an FTS will be kept below the Nyquist limit in order to aid the retrieval process; such a system is said to be "oversampled" with respect to the Nyquist limit, where the oversampling factor is given by $c = (2\delta x\Delta\sigma)^{-1}$. In a Michelson interferometer, the bandwidth of the spectrometer is naturally limited by the spectral response function of the detector, however the bandwidth may be further limited through the use of optical filters.

2.1.2.2 Sampling and spectral resolution

As discussed earlier, when the Dirac comb is transformed, each shifted Dirac delta in the time domain maps to a cosine function in the frequency domain. In a uniformly-sampled system, cosines with linearly increasing frequency will add constructively at integer periods of the lowest-frequency cosine and add destructively at each other wavelength; this process is illustrated in Figure 2.2. The sum of an infinite series of such cosines yields a Dirac comb as expected, however if the interferogram is truncated

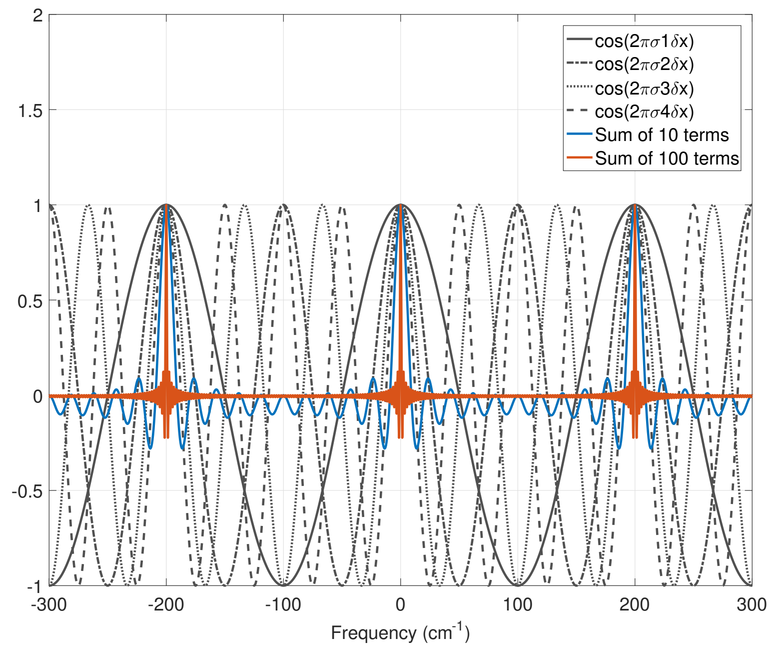

Fig. 2.2 Frequency-domain diagram illustrating how the constituent sampling points of a time-domain sampling comb (Dirac comb) are super-imposed under Fourier-transformation in order to produce a comb-like function in the frequency domain. For clarity only the first four sampling terms are shown along with a 10-term superposition, and 100-term superposition. With only 10 samples, significant ripple in the frequency domain is apparent, however this effect is mitigated significantly for the 100-sample comb.

in the time domain, then the series of cosines is non-infinite and the sum of the cosine terms will yield a comb of sinc functions.

The spectral width of each of the sinc functions is given by the inverse of the longest sampling point in the time series, $FWHM = \frac{1.207}{2\Delta x_{max}}$. In the optical FT this means that the sampling function of a Michelson with with maximum mirror stroke Δx_{max} will produce a comb of sinc functions in the frequency domain such that the i-th sinc function has the form $\text{sinc}_i(2\pi\Delta x_{max}\sigma - i\delta x^{-1})$. The sinc function, which is unique to each interferometer, is often referred to as the "instrument line shape" (ILS).

The result that truncated sampling of the interferogram yields a comb of sinc functions—each with a certain bandwidth—has implications for the ability of the discretely sampled FT of a time signal to resolve two similar frequencies. The FT of a sinusoid, as previously discussed, is a delta function. By the convolution theorem, then, the FT of a discretely sampled time series of a sinusoid results in a convolution of the FT of the sinusoid (a Dirac delta) and the FT of the sampling function (a comb of sinc functions). In other words, even a perfectly monochromatic signal, sampled discretely, will manifest as a sinc function (the ILS) following spectral retrieval.

In the case of two monochromatic signals, each delta function will be convolved with the ILS to produce a pair of sinc functions. As the spectral separation of the two monochromatic signals (*e.g.*, two lasers, a doublet, or absorption lines) approaches the bandwidth of the ILS, it becomes impossible to distinguish between the two lines. The Fourier-transform of various doublets with decreasing wavelength separation is shown in Figure 2.3. The smallest spectral separation at which these signals may

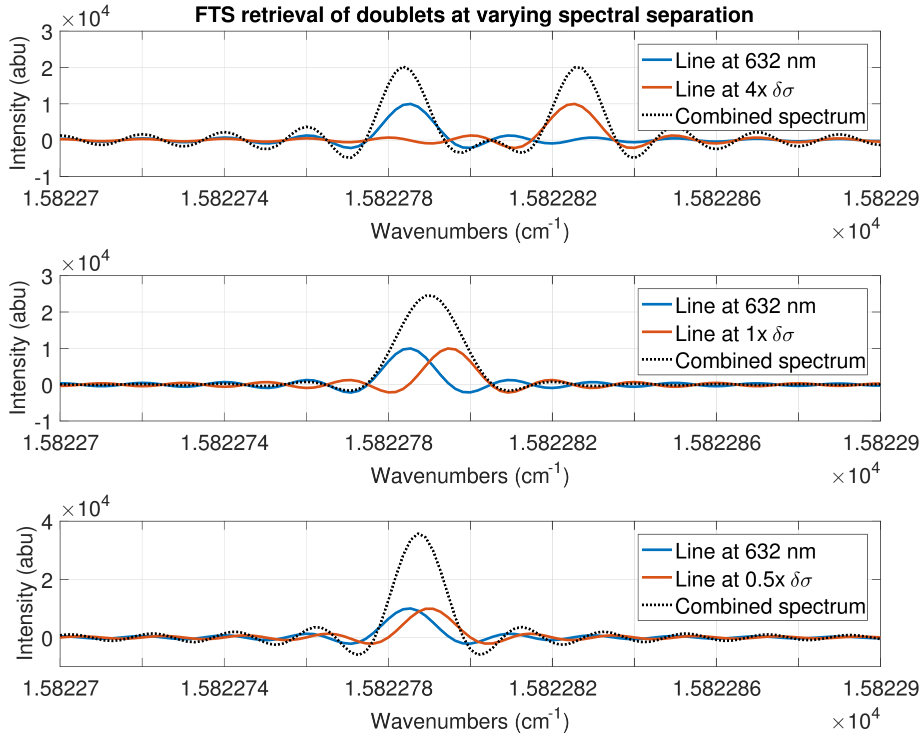

Fig. 2.3 Diagram showing simulated FTS spectra obtained for doublets with varying spectral separation. In the upper plot, the doublets are separated by four times the spectral resolution of the instrument, and are well-resolved. In the middle plot, the doublets are at the spectral resolution limit, and some overlap between the doublets in the combined spectrum may be observed. In the bottom plot the doublets are separated by half the spectral resolution of the instrument, and the two lines cannot be resolved.

be distinguished is given by the Rayleigh criterion. The criterion states that the smallest separation at which two sinc functions may be differentiated is the distance at which the peak of the second spectral component is co-located with the first zero-crossing of the primary peak [77]. This separation distance may be written as a function of the FWHM of the sinc function as $\delta\sigma = 1.207 \cdot FWHM$, or in terms of the maximum OPD: $\delta\sigma = \frac{1}{2\Delta x_{max}}$. Thus, the longest OPD in the FTS defines the "resolution" of the spectrometer. Note that the resolution limit for an FTS is often quoted as $\frac{1}{\Delta x_{max}}$, which is a common misconception deriving from the use of "apodized" interferograms which have been post-processed prior to transformation. Apodization of an interferogram results in under-weighting of the longest OPDs in the interferogram which, in turn, reduces the spectral resolution of the retrieval. A comprehensive discussion of apodization functions is presented by Herres [77], and an intuitive derivation of the loss of spectral resolution is presented by Shepherd [82].

It is interesting, at this juncture, to note the symmetry between the definition of *spectral* resolution in Fourier-transform spectroscopy (and, indeed, in dispersive spectroscopy) and the definition of *spatial* resolution in astronomical (*i.e.* imaging) systems [70]. This symmetry exists because imaging systems are also Fourier-transform instruments, albeit ones which use the Fourier transform to relate the electric field distribution at the instrument pupil plane to the field distribution at the focal plane. In this case, the frequencies that are interrogated by the imaging system are *spatial* frequencies, which are truncated by the clear aperture of the system. The clear aperture of the instrument defines a spatial function, in which all points within the clear aperture are sampled ($f(x_{in}) = 1$), and all spatial points outside the clear aperture are not sampled ($f(x_{out}) = 0$). Since this spatial sampling function is a boxcar function, its Fourier-transform also defines a sinc function—the "point-spread function" (PSF) for the imaging system. The PSF is the imaging analogue of the ILS in an FTS, and like the ILS it may be used to define the spatial resolution of the system in accordance with the Rayleigh criterion, thereby eliciting a satisfying commonality between the two systems.

2.1.2.3 Phase errors and non-uniform sampling

From the prior discussions, it is clear that there are some subtle differences between the continuous and discrete Fourier-transform, and that these differences may have significant consequences for the design of an FT spectrometer. The impact of discretized sampling on the optical FT becomes even more pronounced with the further assumption that samples are not collected with perfect uniformity.

As previously discussed, the effect of discrete sampling in the time domain is to produce a comb upon transformation to the frequency domain. As shown in Figure 2.2 the comb is produced by the summation of a number of sinusoids in the frequency domain, with all sinusoids adding constructively at integer multiples of the

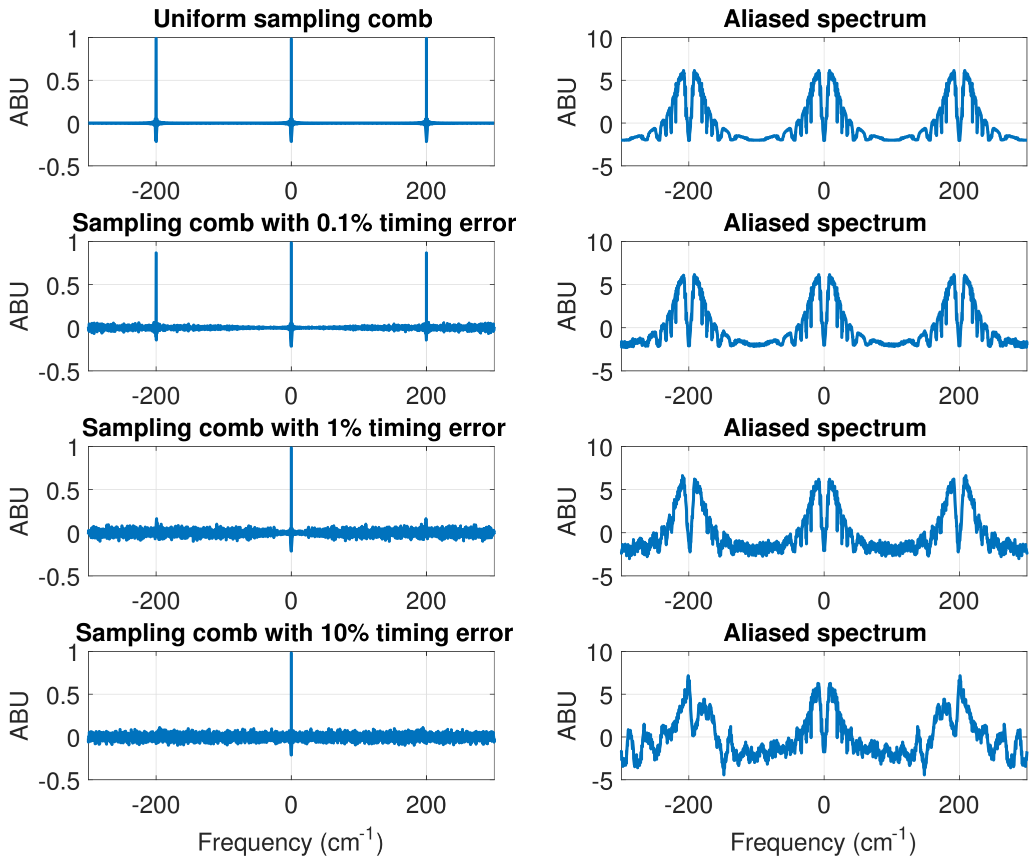

Fig. 2.4 Diagram demonstrating the effects of timing errors (i.e. non-uniform sampling in the time-domain) on the Dirac-comb (left), and the resulting aliased spectra that are produced by convolution with the distorted comb functions (right). It can be seen that the distortion to the signal is minimal near the zero (DC) frequency point, and increases at higher frequencies.

longest-period sinusoid. In the case where the frequency of each sinusoid is not a strict integer multiple of the fundamental frequency, perfect constructive interference cannot occur. Instead, each higher-order sinusoid will be out-of-phase with respect to the fundamental sinusoid, as illustrated in Figure 2.4.

In Figure 2.4, the degree of non-uniformity between sampling points in time is shown to affect the quality of the ILS. At the zero point each cosine is still in-phase, and the function is still "clean", however the cosines de-phase as they extend away from the zero-point, resulting in significant distortion. The practical effect of this distortion is to further truncate the bandwidth of the FTS. As previously discussed, the bandwidth of the FTS arises from the convolution of the Dirac comb with the input spectrum, resulting in overlapping between the the spectrum and its alias—this concept is illustrated in Figure 2.1). In the case of a distorted comb, the input spectrum is convolved with the underlying comb structure, as well as the phase-error artifacts. This results in spectral broadening of the higher-order aliases, which ultimately limits the effective bandwidth of the first order.

In order to avoid these effects considerable design effort is expended to ensure that Michelson-type FTS sample the interferogram at regular intervals. Uniformity of sampling in a Michelson is ensured using a well-calibrated on-board HeNe laser—the transform of which is a sinusoid—to trigger the detector at regular intervals (*e.g.*, using zero-crossings). If uniform sampling cannot be ensured, however, then the bandwidth of the input spectrum must be restricted in order to avoid overlap between the first and second aliases. Provided that the bandwidth is properly reduced, a variety of non-uniform FT techniques including fast-Gaussian gridding and least-squares spectral analysis (LSSA) may be employed to recover the input spectrum from a distorted interferogram [86, 87].

2.1.3 Spectral retrieval from an optical Fourier-transform

Generally speaking, retrieval of spectra, $I_0(\sigma)$, from a continuous interferogram, $I_{det}(x)$, may be accomplished by way of the FT, as in Equation 2.16. However, in the case of a discretely sampled interferogram, the DFT may be used for signal recovery as

$$I(k \cdot \sigma) = \sum_{n=0}^{N-1} I(n\Delta x) e^{i2\pi nk/N} \tag{2.19}$$

Note that the DFT has no explicit dependence on wavelength. Instead, the interferogram and its associated spectrum are treated as dimensionless vectors $k = 1,\ 2,\ 3,\ ...,\ K$ and $n = 1,\ 2,\ 3,\ ...,\ N$. Typically, the fast Fourier-transform (FFT) is used to perform the DFT operation in Equation 2.19. In Michelson interferometry, in which interferograms may run to many thousands of samples, the FFT is preferred to the DFT as the former runs in $O(n \log n)$ time, whereas the latter runs in $O(n^2)$.

The DFT may also be expressed as a matrix operation $\mathbf{x} = \mathbf{B}\mathbf{y}$ where $\mathbf{y}$ is the interferogram, $\mathbf{x}$ is the spectrum of interest, and $\mathbf{B}$ is a $K \times N$ matrix defined by:

$$\mathbf{B(k,n)} = \mathbf{e^{i2\pi nk/N}}. \tag{2.20}$$

We may assume that the matrix $\mathbf{B}$ has some generalized inverse, $\mathbf{B^{-1}} = \mathbf{A}$, such that:

$$\mathbf{BAB} = \mathbf{B}, \tag{2.21}$$

and since $\mathbf{y} = \mathbf{ABy}$, then

$$\mathbf{y} = \mathbf{Ax}.. \tag{2.22}$$

Following Equation 2.22, the operation of the optical FT spectrometer itself may be described as a matrix operation. In fact, at the limit that $K \longrightarrow \infty$, the summation operation in Equation 2.19 is equivalent to a continuous integral; therefore, for sufficiently well-defined matrix $\mathbf{A}$, Equation 2.22 is equivalent to the optical FT in Equation 2.15.

The matrix $\mathbf{A}$, like the continuous FT, maps signals in the spectral domain to the time domain. For an FTS device, the matrix contains a numerical description of the instrument response for a given spectral input $\mathbf{x}$. This response may be determined experimentally through calibration, in which case Equation 2.22 may be solved by inversion of the experimentally determined matrix $\mathbf{A}$ using the Moore-Penrose pseudoinverse [88].

While the advantages of retrieval via matrix inversion may not be immediately apparent, they become so in consideration of non-uniform sampling of the interferogram—as previously discussed. As may be seen in Equation 2.19, the DFT (and FFT) assume perfectly uniform sampling of the interferogram; when the DFT is applied to a non-uniformly sampled interferogram, this assumption results in significant distortion to the retrieved spectrum. As discussed previously, a common solution to ensure uniform sampling in Michelson interferometry is to employ a fringe-counting HeNe laser as a position reference. Despite this mitigation, the interferogram will not be perfectly uniform; in addition to its effects on bandwidth, the non-uniform sampling will affect the shape of the imaginary solution to Equation 2.19.

To compensate, Michelson interferometers may initiate their mirror stroke at some negative displacement, $\Delta x_{min} < 0cm$, and pass through the zero path difference (ZPD) before reaching maximum displacement $\Delta x_{max} > 0cm$. Passing through the ZPD in this manner allows for collection of a short double-sided interferogram, from which low-resolution envelopes of the amplitude (real) and phase (imaginary) solutions may be recovered. The low-resolution imaginary envelope may then be subtracted from a single-sided interferogram to produce a phase-corrected interferogram that may be transformed into a final spectrum. This procedure is developed in greater detail in the literature [77, 89].

In the case where a HeNe laser is not available, matrix inversion using Equation 2.22 may compensate directly for non-uniform sampling. Applying the Moore-Penrose pseudoinverse to interferogram $\mathbf{y}$ is equivalent to performing a simultaneous row-wise least-squares fit to Equation 2.22. The general term for this operation is LSSA, which is a technique that was first popularly deployed by the astronomical community. LSSA has been used to conduct time-series analysis and frequency decomposition of the light curves of variable stars—the time-samples of which could not be collected on a uniform basis due to weather and observatory schedules [90]. The technique was later refined by Petr Vanicek in order to analyze irregularly-sampled geological time-series [86], and by Jeffrey Scargle, who formalized the technique for the then-popular field of astroseismology [91]. In each of these algorithms it is assumed that the interferogram $\mathbf{y}$ is the product of a series of projections of the spectrum $\mathbf{x}$ along a set of well-characterized sinusoids in the frequency domain. Estimates of the spectral signal $\hat{\mathbf{x}}$ are then generated computationally, and the estimate that minimizes the squared difference between the interferogram and the projections of the estimate ($min\ \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|^2$), is identified using a search routine. This estimate that minimizes the difference squared is then selected as the best estimate of the true spectrum $\mathbf{x}$. The primary advantage of this method is that, compared to the DFT, LSSA is much more agnostic as to the spectral content of the matrix $\mathbf{A}$, and can therefore tolerate a matrix featuring non-uniform sampling.

2.1.4 Compressive sensing and FTIR

Compressive sensing (CS) generally refers to a group of signal processing methods related to the capture and retrieval of signals that are "sparse". The technique is commonly used in image compression and reconstruction, however it may also be applied to spectroscopy. The unique advantage of compressive sensing spectroscopy is that in the case of sparse signal detection, a compressive sensing spectrometer may require substantially fewer sampling points than a non-compressive system. Indeed, in image reconstruction and in spectroscopy compressive sensing has been shown to enable a greater-than four-fold reduction in sampling without compromising data quality [92]. In order to achieve these reductions in measurement density, three preconditions must be established:

1. The signal of interest must be sparse in some domain, termed the "signal domain".

2. Measurements of the signal should be collected in a second domain, the "sensing domain", that is maximally incoherent with respect to the signal domain.

3. The measurement points in the sensing domain, when expressed as basis vectors in the signal domain, should be nearly orthogonal and unitary.

The first requirement is that the signal be sparse in the signal domain; the concept of a sparse signal is illustrated in Figure 2.6. This requirement is a restriction on the measurement of interest: a sparse signal is one in which, in some domain, the majority of the information in the signal is contained in a few elements. Many naturally occurring spectra which are of interest to optical physicists are sparse. This includes: laser lines, doublets, molecular and atomic emission lines (*e.g.*, from atomic transitions), and Raman emission. In each of these cases, the spectra of these emission sources will be mostly dark, with a few bright spots indicating the spectral peak of an emission source. Conversely, the spectra of a non-sparse (*i.e.* broadband) signal, such as the solar spectrum, would be bright throughout the spectrum, less any absorption features. It should be noted that "sparse" emission sources such as laser lines and atomic transitions actually have well-defined structure and line width at sufficiently high resolution scales. Therefore the distinction must be made that "sparsity" in a spectral system implies only that the structure and spectral width of that signal is smaller than the inherent resolution of the instrument.

The second requirement, that the measurements be collected in an incoherent basis, is a restriction on the measurement method for collecting the signal. As described in Section 2.1, incoherence refers to the notion that a signal that is narrowly confined (high power-density) in one domain is widely dispersed in another domain (low power-density). The time and frequency domains, especially as applied to FT spectroscopy, are a particularly good example of this property (as illustrated in Figure 6.1). In the case of a spectrally sparse input signal, such as a HeNe laser, many of the samples in the interferogram are redundant; for instance, if the operator knows that the signal is monochromatic they may forgo FT retrieval entirely and simply fit parts of the interferogram to a sinusoid—using far fewer measurements in the process. This principle holds even in the case of a sparse, polychromatic, signal. Since each spectral component in the polychromatic signal will be well-dispersed in the time domain, the total number of measurements required to extract the non-redundant information is far fewer than the full set of $N = 2\Delta\sigma/\delta\sigma$ samples for given spectral range $\Delta\sigma$, and resolution $\delta\sigma$. Note that the freedom to collect fewer than N samples does not emancipate the spectroscopist from the limitations of the Nyquist-Shannon sampling theorem. A large maximal OPD is still required in order to distinguish between two signals that are closely separated in frequency space. Similarly, the minimum separation between sampling points in the time domain still defines the spectral bandwidth of the instrument. For this reason, compressive-sensing is not particularly practical in Michelson interferometry. A Michelson interferometer necessarily collects the full suite of N measurements as the mirror OPD translates from the ZPD to Δx_{max}. Therefore, to form an undersampled measurement set, data must be wastefully discarded from the full interferogram. Accordingly, compressive sensing is best deployed in FTS systems in which the spectroscopist has complete and independent control over collection of each sample in the interferogram.

The final requirement, that the measurement matrix vectors describe a nearly

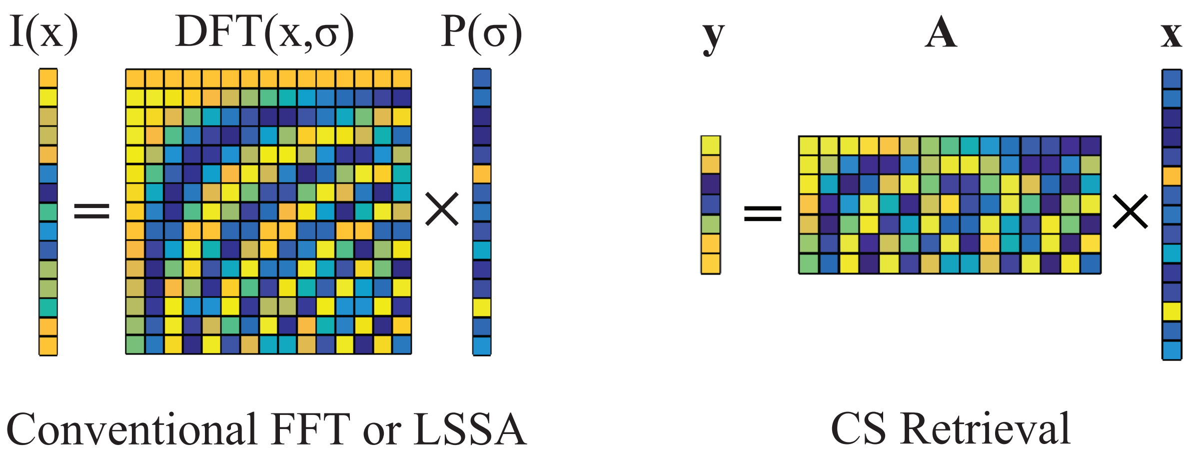

Fig. 2.5 Diagram illustrating the difference between the measurement matrix of a fully-sampled FTS instrument and the measurement matrix of a compressive-sensing FTS using a partial Fourier matrix.

orthogonal basis, is another restriction on the measurement method. It specifies that the measurement points—when expressed in the signal domain—should be independent or nearly independent, and that they should be unitary when acting on a sparse signal. In other words, when the measurement vectors are collected into a sensing matrix, as in Equation 2.22, then this sensing matrix, $\mathbf{A}$, should preserve the length of the sparse input vector $\mathbf{x}$. A matrix, or set of measurements, satisfying this criteria is said to satisfy the "restricted isometry property" (RIP), meaning that the matrix defines an isometric transform for a restricted set of input vectors (*i.e.* sparse vectors). Generating large matrices that satisfy this property is an NP-hard computational task, however certain families of matrices can be shown to satisfy the RIP with overwhelming probability [93, 94]. In particular, partial Bernoulli, Gaussian, and Fourier matrices (see Figure 2.5) have been shown to satisfy the RIP with an exponentially high degree of probability. Of these matrices, partial Fourier matrices should be of greatest interest to FT spectroscopists, as a partial Fourier matrix is facilely constructed by a random sampling of points in an interferogram. This implies that all that is required for compressive sensing retrieval from an FTS is (1) a spectrally sparse signal, and (2) a subset of interferogram measurements randomly selected from a complete interferogram.

Provided that the above criteria are satisfied, a compressive sensing retrieval may be performed using a very similar formalism to the LSSA retrieval. As discussed in Section-2.1.3, LSSA retrieval solves an equation of the form $\mathbf{y} = \mathbf{Ax}$ by using the pseudoinverse of the measurement matrix, $\mathbf{A}$. As the name implies, this process is equivalent to solving a least-squares optimization problem of the form:

$$min\ ||\mathbf{x}||_2\ s.t.\ \mathbf{y} = \mathbf{Ax}. \tag{2.23}$$

where $||\mathbf{x}||_2$ denotes the $l2$ norm (Euclidean norm) of the vector $\mathbf{x}$. The objective of the problem in Equation 2.23 is to obtain the lowest energy estimate of $\mathbf{x}$ that satisfies $\mathbf{y} = \mathbf{Ax}$. Equation 2.23 describes an optimization problem which may be solved by any family of suitably robust optimization algorithms (*e.g.*, gradient descent, Levenberg-Marquardt, etc.). This problem may be re-cast as a compressive sensing problem merely by changing the objective function to find the *sparsest* solution to $\mathbf{y} = \mathbf{A}\mathbf{x}$. The search for a sparse solution may be performed by optimizing for the solution that minimizes the $l0$ or $l1$-norm of $\mathbf{x}$. Whereas the $l2$-norm is simply the Euclidean norm of the vector, the $l1$-norm refers to the sum of the absolute values of elements of $\mathbf{x}$. Finally, the $l0$-norm denotes the number of non-zero elements in $\mathbf{x}$, and, adopting the definition that $0^0 = 0$, the $l0$-norm may be expressed as:

$$||\mathbf{x}||_2 = \sqrt{{x_1}^2 + {x_2}^2 + ... + {x_{n-1}}^2 + {x_n}^2} \tag{2.24}$$

$$||\mathbf{x}||_1 = |x_1| + |x_2| + ... + |x_{n-1}| + |x_n| \tag{2.25}$$

$$||\mathbf{x}||_0 = |x_1|^0 + |x_2|^0 + ... + |x_{n-1}|^0 + |x_n|^0. \tag{2.26}$$

True optimization for the sparsest solution would require optimization with respect to the $l0$-norm. Unfortunately, the search space for the $l0$-norm is not a smooth space, and therefore $l0$-norm minimization is an NP-hard computational task that requires a brute-force search of all possible solutions to $\mathbf{y} = \mathbf{A}\mathbf{x}$. However, provided that the matrix $\mathbf{A}$ satisfies the RIP, then with high probability, the solution obtained via $l1$-norm minimization will match the $l0$-norm solution [95, 96]. In contrast to minimizing the $l0$-norm, $l1$-norm minimization is a convex optimization problem that may be solved by a variety of algorithms (*e.g.*, basis-pursuit [95], LASSO [92], etc.). Finally, then, for a sparse spectral signal measured using random samples of a complete interferogram, the solution may be retrieved as:

$$min\ ||\mathbf{x}||_1\ s.t.\ \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{2.27}$$

2.1.4.1 Compressive-sensing FTS system modelling

To demonstrate the concept of FTS using CS, two instrument models were developed to investigate CS retrieval for real and simulated FTS data. Using these models, the CS retrieval process for spectrally sparse and spectrally non-sparse data was verified using trial spectra obtained from dispersive and FTS instruments.

A basic model for an optical FTS is presented in Section 2.1.1 and Section 2.1.3. Following these examples, the interferogram produced by sample spectra $\mathbf{x}$ may be approximated by $\mathbf{y} = DFT(\mathbf{x})$, or $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is the matrix form of the DFT. The resulting interferogram, $\mathbf{y}$ may then be inverted using the DFT, or using the pseudoinverse of the matrix $\mathbf{A}$. In the case of a CS FTS, only a partial sampling of the interferogram is collected. Partial sampling may be achieved by constructing matrix $\mathbf{A}'$ consisting of randomly selected rows of the DFT matrix $\mathbf{A}$. Alternatively, the interferogram for a CS FTS may be generated by collecting the full interferogram corresponding to the input spectrum, and digitally down-sampling the interferogram $\mathbf{y}$ to produce the CS interferogram $\mathbf{y}'$ with associated measurement matrix $\mathbf{A}'$. In each case, the interferogram is inverted using the l1-norm minimization approach detailed in Section 2.1.4.

This basic instrument model was used to generate reference interferograms from a trial dataset consisting of visible spectra collected via a dispersive Raman spectrometer. Raman spectroscopy was selected as the first example case because the spectra of interest (the Raman emission) is partially sparse in the spectral domain—and therefore satisfies the core requirement for CS. Raman spectroscopy is also an interesting example case because a typical Raman spectra may contain significant non-sparse spectral power due to the presence of broadband background fluorescence.

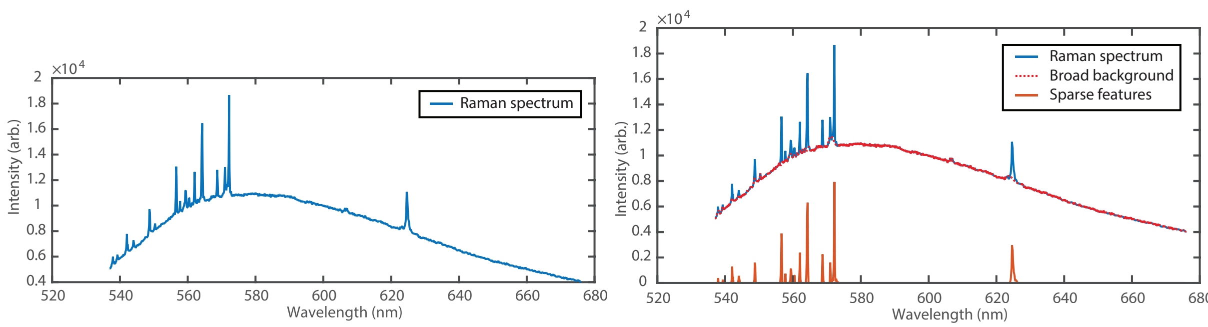

Fig. 2.6 Diagram showing an example Raman spectrum (left) provided by Ed Cloutis at the University of Winnipeg, and a breakdown of that spectrum (right) into sparse and non-sparse spectral components.

Raman spectra collected from synthetic organic samples using a B & W iRaman dispersive spectrometer were provided by Ed Cloutis at the University of Manitoba. The Raman instrument operates over the wavelength range 537 nm to 676 nm with a wavelength resolution of ≈ 0.14 nm. A comparable Nyquist-limited on-chip SHFTS would consist of $N = 2\Delta\lambda/\delta\lambda = 1868$ MZIs, with with evenly spaced linearly increasing OPDs from $\Delta L_1 = 1.306$ μm to $\Delta L_{max} = 2440$ μm. These OPDs may be used to generate a DFT matrix, **A**, wherein each row corresponds to the transmittance function of a given MZI. This matrix may be used to approximate the interferogram that would be produced by the SHFTS, for a given input spectrum. Following the CS formulation presented in Section 2.1.4, a similar compressive sensing SHFTS (CS SHFTS) instrument model would deploy a randomly-selected subset of the rows of matrix **A** to produce a partial interferogram for the same input spectrum.

The Raman spectra shown in Figure 2.6 consists of a broadband background signal punctuated by high-intensity, narrow-bandwidth peaks. This spectra can be separated into a coarse background spectrum (red dotted curve, Figure 2.6), and a sparse Raman signal with a smaller number of non-zero components (orange curve Figure 2.6). The sparse signal is a suitable candidate for CS reconstruction, however it must first be separated from the coarse background.

The background signal forms a slowly-varying envelope, accordingly, it is well-described by the first few components of its Fourier transform (the first few components in the interferogram), which describe the rough outline of the signal. Overall, in the Raman dataset that was provided, even highly complex background spectra could be well described using only the 40 shortest MZIs—corresponding to the 40 first components in the interferogram. This implies that, a CS SHFTS may be able to perform background estimation by measuring the shortest time-delay components using a dedicated set of MZIs in a Nyquist-sampled configuration.

In this case, an initial estimate of the background $\hat{\mathbf{b}}$ may be obtained from the first 40 elements of the signal, $\mathbf{y_{40}}$, and sensing matrix $\mathbf{A_{40}}$, via conventional LSSA retrieval of the sensing equation $\mathbf{y_{40}} = \mathbf{A_{40}}\mathbf{b}$. The estimated background may then be used along with the original sensing matrix to produce a new interferogram, $\mathbf{y_{sparse}}$ in which non-sparse contributions have been removed:

$$\mathbf{y_{sparse}} = \mathbf{y} - \mathbf{A}\hat{\mathbf{b}}. \tag{2.28}$$

Following background removal in Equation 2.28, the remaining sparse components, $\mathbf{x_{sparse}}$, may be retrieved via $l1$-norm minimization of the residual interferogram $\mathbf{y_{sparse}}$, following Equation 2.27. These sparse components may then be combined with the earlier background estimate to produce the measurement of the complete spectrum $\hat{\mathbf{x}} = \hat{\mathbf{x}}_{\mathbf{sparse}} + \hat{\mathbf{b}}$.

Using the instrument model described above, an on-chip CS SHFTS corresponding to the BW-Tek Raman instrument was simulated. This model used a set of 400 MZIs comprising the 40 shortest MZIs, and 360 additional MZIs randomly selected from the set of $N = 1868$. This instrument model was used to analyze each of the 69 Raman datasets provided, yielding accurate retrievals in nearly all cases. Of the 69-element set, CS retrievals for two different types of Raman spectra are highlighted

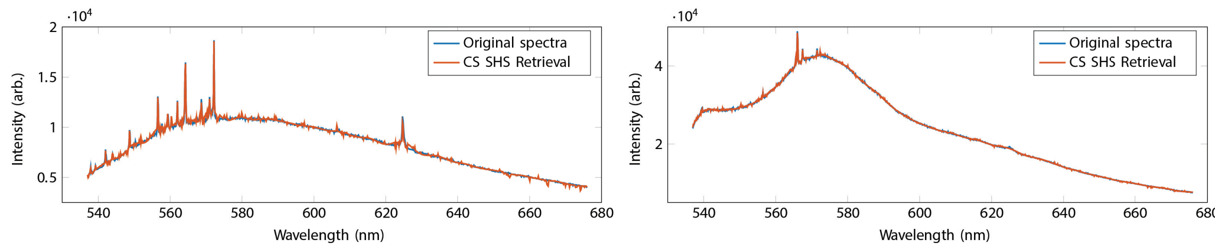

Fig. 2.7 Raman spectra retrieved using CS instrument model. The CS SHFTS uses 400 samples out of an interferogram with $N = 1868$ total samples. In both retrievals, 40 low-frequency samples (uniformly spaced) are retained for background estimation, while the remaining 360 samples are allocated randomly from the set of 1868. The first spectrum (left) has very strong sparse features and a smoothly-varying background spectrum; the second spectrum (right) has a more complex background spectrum and relies more heavily on background estimation and removal for CS retrieval.

in Figure 2.7. The first retrieval in Figure 2.7 shows the reconstruction of a highly sparse spectra with simple background fluorescence. The root-mean-square error (RMSE) of the CS retrieval with respect to the measured data in Figure 2.7 is near to the RMS difference between the smooth background spectra and a polynomial fit. Assuming that the broadband background emission is a smoothly-varying function, then this inherent RMSE should reflect the inherent photon and readout noise in the detection system. The second retrieval in Figure 2.7 shows the reconstruction of a more complex spectra, in which the non-sparse components (background) are a significant contributor to the overall flux. However, using the background compensation scheme described above it is possible to produce a residual spectrum with enforced sparsity; at which point spectral retrieval CS retrieval may be performed. Again, the reconstruction error is on the order of the signal noise.

As described above, the Raman analysis was performed using an instrument model that generates a discretely-sampled continuous interferogram from pre-discretized spectra. Therefore there remains the possibility that a real interferogram—generated by optical Fourier-transform of a continuous spectra—will be less amenable to background removal using low-order retrieval of the background spectrum. In order to verify that this is not the case, and to further validate the proposed concept, compressive sensing retrievals were also performed using archived data from the Fourier-transform spectrometer onboard the Herschel space observatory.

The Spectral and Photonic Imaging Receiver (SPIRE) instrument on the Herschel space observatory is an imaging Michelson interferometer with 12.56 cm maximum optical path displacement and three imaging bands centered at 250 μm, 360 μm, and 520 μm [57]. Data are collected as single-sided (high-resolution) and dual-sided (low-resolution) interferograms and are made available in raw and pre-processed format through the Herschel portal. Various astronomical targets with spectrally-sparse emission features were observed throughout the telescope's four-year life span, each of which may be a suitable candidate to validate CS retrieval from an FTS.

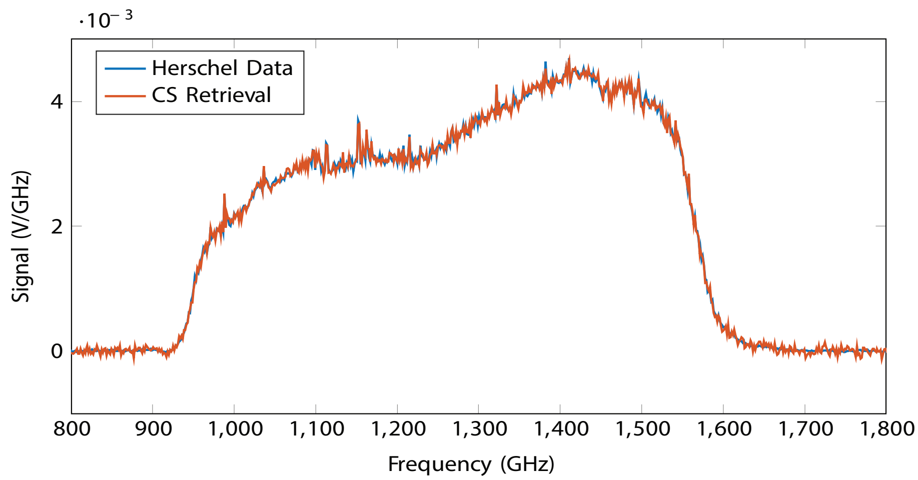

Fig. 2.8 CS retrieval of emission spectrum from VY Canis Majoris reconstructed from raw interferogram obtained by SPIRE. The original spectrum is reconstructed by FFT of the full interferogram consisting of 8192 samples. The CS retrieval, which retains the full bandwidth of the original, is obtained by using the first 128 samples to perform background estimation, and the remaining sparse emission features are retrieved from 2500 randomly-selected interferogram points.

A high-resolution single-sided interferogram from an observation of VY Canis Majoris was used to validate the CS background compensation and retrieval scheme. The spectrum contains a variety of high-intensity emission lines, which constitute the spectrally sparse signal of interest for the CS retrieval. The measurement bandwidth of the interferometer exceeds the spectral range of the detector, accordingly, the retrieved spectrum is superimposed on the responsivity curve of the detector itself—

thereby ensuring the presence of a broadband non-sparse background.

A processed high-resolution measurement from the SPIRE instrument consists of 8192 samples of the interferogram—equivalent to sampling 8192 independent MZIs in an on-chip SHFTS. The first 128 samples are kept for the purpose of performing background estimation according to Equation 2.28, and a further 2500 samples are randomly selected from the full set in order to form the partial interferogram for the CS retrieval. The rows of a DFT matrix corresponding to each randomly selected sample are generated numerically and are concatenated to form the sensing matrix, $\mathbf{A}$. This matrix, and the associated set of samples, $\mathbf{y}$, are inverted using l1-minimization and the background estimation technique in order to retrieve the sparse spectral features. The result of this CS retrieval is shown in Figure 2.8 alongside the conventional Fourier-transform retrieval using the fully-sampled interferogram. Once again, the compressive sensing retrieval successfully reproduces the sparse emission features of the spectrum, as well as the broadband background components corresponding to the detector responsivity curve.

These two successful demonstrations validate the concept of compressive-sensing Fourier-transform spectroscopy by demonstrating that l1-norm minimization can be used to retrieve discrete samples of continuous interferograms, even in the case of significant non-sparse spectral features. Therefore, it is likely that spectrally-sparse measurements obtained using an on-chip SHFTS (which also discretely samples a continuous interferogram) should also be amenable to reconstruction via CS. Moreover, the use of a small subset of Nyquist-sampled MZIs, with short path delays, should suffice to eliminate any contributions from spectrally broad background signals.

2.1.5 Free-space implementations of Fourier-transform spectroscopy

While the prior discussion of the principles of FTS focuses on the Michelson interferometer, there are many alternative implementations of FTS—each with advantages and disadvantages. In this section, a selection of common FTS architectures is presented, and the various advantages and limitations of each architecture are discussed. Note that this discussion is limited to spectroscopic implementations of FTS; many further implementations of interferometry exist for various applications including optical metrology [97], angular rate-sensing [34], gravimetry [98], and so on. A more comprehensive list of FTS architectures is available in [99].

2.1.5.1 Michelson interferometer

By far the most common implementation of FTS is the Michelson interferometer and its variants (*e.g.*, the Sagnac interferometer). Michelson interferometers, as previously described in Section 2.1.1, split light into two separate arms which are then re-combined at a single detector. The signal at the detector corresponds to a single point in the continuous interferogram of the input optical spectrum (the inverse

Fourier-transform of the spectrum). To capture a sampling of the interferogram the displacement of one or both arms in the interferometer is modified in order to vary the OPD between the two arms. As the OPD of the interferometer is varied, producing the interferogram, the detector captures a discrete set of measurements from which the input spectrum is recovered using the Fourier-transform. This concept of operations is shown diagrammaticality in Figures 1.5 and 2.1.

In a conventional Michelson interferometer, shown in Figure 1.5, the interference cavity is formed using a beam splitter and two reflective elements (*e.g.*, corner-cubes, well-aligned mirrors). The OPD of the cavity is then varied by linearly translating one of the two retro-reflectors. In principle, there is no restriction on the diameter of the beam within the Michelson interferometer, therefore Michelson interferometers may be constructed to achieve very high numerical aperture (and correspondingly higher étendue) on the collimating and focusing assemblies. In practice, the spherical and chromatic aberrations associated with collimating highly-diverging beams will limit the achievable throughput in these systems.

The advantage of a Michelson interferometer, relative to alternative FTIR techniques is that the spectral resolution and bandwidth of the spectrometer are limited only by the mechanical resolution, range of travel, and stability of the travelling arm. As such, state-of-the-art Michelson interferometers are capable of unmatched spectral bandwidth and resolution [100]. For this reason, Michelson interferometers are commonly deployed in gas-sensing applications in which the concentrations of a broad family of gas-species must be evaluated (*e.g.*, semiconductor foundry process-monitoring [58], atmospheric chemistry stations [101]). Despite these advantages, high-performance Michelson interferometers are difficult to miniaturize, and are very sensitive to mechanical vibrations. For this reason, Michelson interferometers are rarely implemented in remote-sensing systems which operate in highly dynamic environments.

2.1.5.2 Spatial Heterodyne Spectrometer

Another very common implementation of FTIR, illustrated in Figure 2.9, is the spatial heterodyne spectrometer (SHS), also sometimes called the spatial heterodyne Fourier-transform spectrometer (SHFTS). Like the Michelson interferometer, SHFTS devices split light into two separate arms and recombine light at a detector while introducing a varying OPD between arms. In each arm of the SHFTS the input beam is dispersed by a diffraction grating, and the resulting beams are re-combined at the detector plane to produce a Fizeau fringe pattern with periodicity (angle) determined by the wavelength of the input beam. This fringe pattern (the interferogram) may be captured using a linear detector array or 2-D image array and the original input spectrum is then reconstructed using the Fourier transform.

The SHFTS, like the Michelson interferometer, has no restriction on the diameter

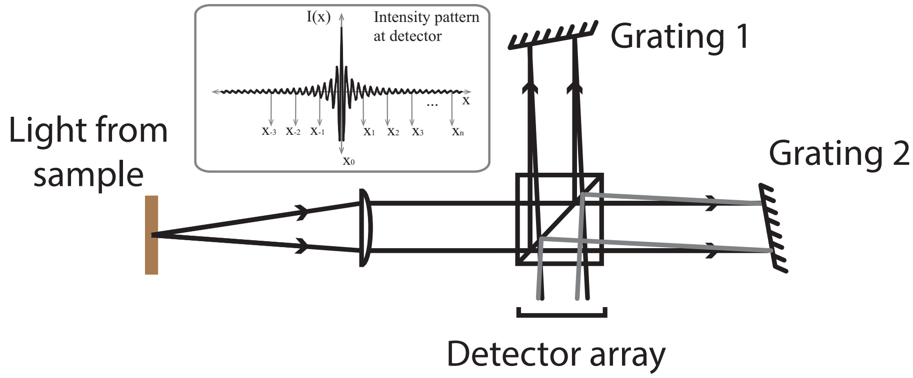

Fig. 2.9 Diagram of an FTS based on a spatial-heterodyne architecture. Light from the sample enters a collimating lens, and is separated into two arms whereupon the beams are diffracted by gratings back into the beamsplitter and recombined at the detector. The difference in angle between incoming diffracted beams of a given wavelength in each arm produce a spatial fringe (interference) pattern at the plane of the detector array. Note that unlike the Michelson interferometer, which typically images the source onto the detector, an SHS collects the interferogram in collimated space.

of the beam within the interferometer, enabling it to maximize étendue [102]. In addition, since the two arms of the SHFTS are aligned and fixed in place with no moving parts, the entire interferometric cavity can be formed from a monolithic block of glass—producing a mechanically and thermally robust system suitable for dynamic operational environments. Unfortunately, since the SHFTS produces a *spatial* interference pattern, the spectral bandwidth and resolution that may be interrogated by the instrument is limited by the spatial extent and pixel density of the detector array. For this reason, most SHFTS devices target very high spectral resolution over a narrow waveband. While this approach is suitable for targeted detection of a specific gas species with high spectral resolution, it is not suitable for multi-gas detection.

Nonetheless, SHFTS-type devices have been successfully implemented for gas-detection [102], Raman spectroscopy [78], and optical metrology [103]. The SHFTS architecture is also well suited for adaptation to planar-waveguides, as described in Section 3.1. Note also that the spatial heterodyne architecture does not specifically require path delays to be introduced through grating dispersion; indeed, SHFTS devices have been demonstrated using OPDs generated by mechanical displacement [104].

2.1.5.3 Fabry-Perot FTIR

One less-common implementation of FTS uses a low-finesse interferometric cavity (such as a Fabry-Perot etalon) to produce an interference pattern approximating the inverse fast-Fourier transform (IFFT) of the input spectrum. In this case the fringe pattern is produced by interference between multiple reflection paths within the etalon, and the OPD of the interferometer is determined by the physical gap of the etalon as well as the number of reflections (order) within the cavity. Provided that the physical gap of the etalon can be adjusted as the instrument collects data, then a full interferogram may be collected just as in the case of the Michelson interferometer.

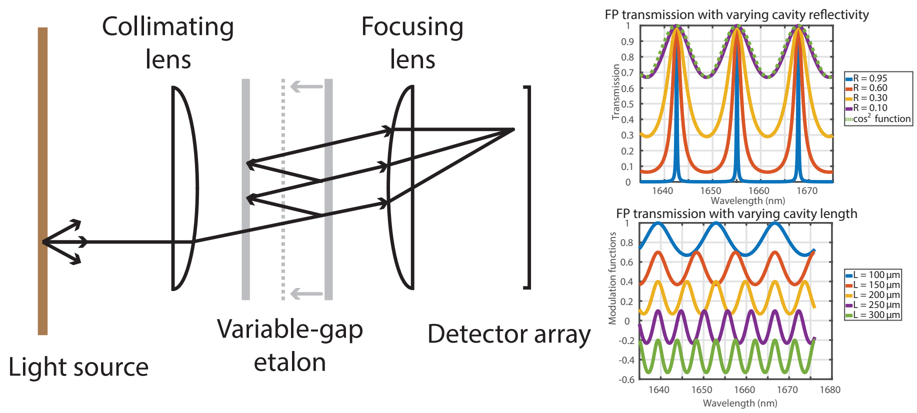

Fig. 2.10 Diagram of a basic Fabry-Perot interferomter, as well as transmittance functions for various surface reflectivities and cavity spacings. Light from the sample enters a collimating lens, and interferes with reflections within the primary cavity (*i.e.* the etalon). Reflections within the cavity generate a transmission profile as a function of wavelength, angle, and cavity length and reflectivity. In the inset (upper right) it can be seen that for low values of reflectivity the transmission profile approximates the $\cos^2$ profile of a traditional Michelson interferometer. By controlling the spacing of the cavity (lower right), the periodicity of the transmission profile may be increased or decreased, similar to scanning the delay arm in a Michelson interferometer.

The use of a low-finesse cavity for FTIR is somewhat counter-intuitive since Fabry-Perot interferometers (FPIs) more commonly use high finesse (*i.e.* highly reflective) surfaces in order to produce a very spectrally-selective transmission function—as shown in Figure 2.10. However, in the low-finesse regime, the transmission function of an etalon approximates that of a Michelson, implying that the resulting intensity values may be used to reconstruct the input spectrum using the FFT. The transmittance of an etalon, as a function of surface reflectivity, R, is:

$$T(x, \sigma) = \frac{{T_{surf}}^2}{1 + R^2 - 2R\cos(2\pi\sigma x)}. \tag{2.29}$$

As R falls to $0.1 < R < 0.3$ Equation 2.1.5.3 may be expanded into a Fourier series such that $T(x, \sigma) = \sum_{n=0}^{\infty} A_n \cos(\pi n \sigma x)$; this enables the data collected from a scanning FPI to be processed in a similar manner to conventional FTIR in order to produce the same spectral reconstruction. This configuration is sometimes referred to as a multiplex Fabry-Perot interferometer [105].

The primary advantages of an FPI FTS implementation are compactness and ruggedness relative to comparable Michelson implementations. The compact common-path nature of the FPI FTIR makes it a suitable choice for dynamic environments in which size is a driving system requirement. One key difference between an FPI FTS system and conventional Michelson interferometry is that FPI FTS cannot practically achieve zero or near-zero OPDs, which may ultimately limit its spectral bandwidth [105].

3 Design and fabrication of on-chip spectrometers

3.1 Waveguide theory

3.1.1 Waveguide modes

Optical waveguides, which form the basis for photonic integrated circuits, are structures that confine propagating light, enforcing particular directions or modes of propagation. Waveguides may be constructed of metal [106], dielectric [107], or even organic materials [108], however, in this work only waveguide structures based on dielectric materials are considered. The propagation of light through waveguides (or any other medium) is governed by Maxwell's equations, shown in their complex, time-independent (*i.e.*, assuming harmonic time-dependence) form in Equations 3.1-3.4.

$$\nabla \cdot \vec{E} = \frac{\rho}{\eta} \tag{3.1}$$

$$\nabla \cdot \vec{H} = 0 \tag{3.2}$$

$$\nabla \times \vec{E} = i\omega\mu\vec{H} \tag{3.3}$$

$$\nabla \times \vec{H} = \vec{J} - i\omega\eta\vec{E} \tag{3.4}$$

Mathematical descriptions of waves which satisfy these equations ("solutions to Maxwell's equations") describe any possible configuration of propagating light. This is true of incoherent light emitted from the photosphere of a star, for the propagation of a laser beam through free-space, and for the propagation of light through optical fibers and waveguides.

In waveguide theory Maxwell's equations are used to find an electric field distribution for a propagating wave that satisfies the boundary conditions at the edges of the waveguide structure. These solutions are termed the "modes" of the waveguide. The modes of a waveguide provide crucial insight into the characteristics of an optical waveguide. Among other things they govern the speed of light propagating through the waveguide, the interaction between light in the waveguide and its surrounding environment, and any interference effects that may occur within the waveguide. The solutions to Maxwell's equations for waveguide structures are strongly influenced by the boundary conditions of Maxwell's equations:

$$\vec{n}_{12} \times (\vec{E_2} - \vec{E_1}) = 0 \tag{3.5}$$

$$\vec{n}_{12} \cdot (\epsilon_2 \vec{E_2} - \epsilon_1 \vec{E_1}) = \sigma_s \tag{3.6}$$

$$\vec{n}_{12} \cdot (\vec{B_2} - \vec{B_1}) = 0 \tag{3.7}$$

$$\vec{n}_{12} \times (\vec{B_2}/\mu_2 - \vec{B_1}/\mu_1) = \vec{J_s} \tag{3.8}$$

where $\vec{n}_{12}$ is the unit normal vector travelling from medium 1 into medium 2, σ_s is the surface charge (usually zero for dielectric waveguides), $\vec{J_s}$ is the surface current density (also usually zero for dielectric waveguides), and ϵ_i and μ_i are the electric permittivity and magnetic permeability, respectively, of material i. Qualitatively, the boundary conditions state that the tangential component of the electric field must be continuous across the interface, and the normal component of the electric field will have a discontinuity that is governed by the difference in permittivity between the two media. In the case of the magnetic field, the normal component is continuous across the interface and the tangential component will have a discontinuity governed by the difference in permittivity. These discontinuities may be seen in the plots of waveguide electric field amplitude and intensity shown in Figures 3.2 and 3.4.

Though Maxwell's equations are exact, finding solutions to the equations can be challenging—for this reason a variety of numerical methods exist for determining the optical characteristics of waveguides. In this section an overview of techniques for waveguide analysis and their applications to waveguide design are discussed. A detailed comparison of computational techniques for waveguide design is beyond the scope of this thesis, however more comprehensive discussion on this subject can be found in the literature [109, 110].

3.1.1.1 The infinite dielectric slab

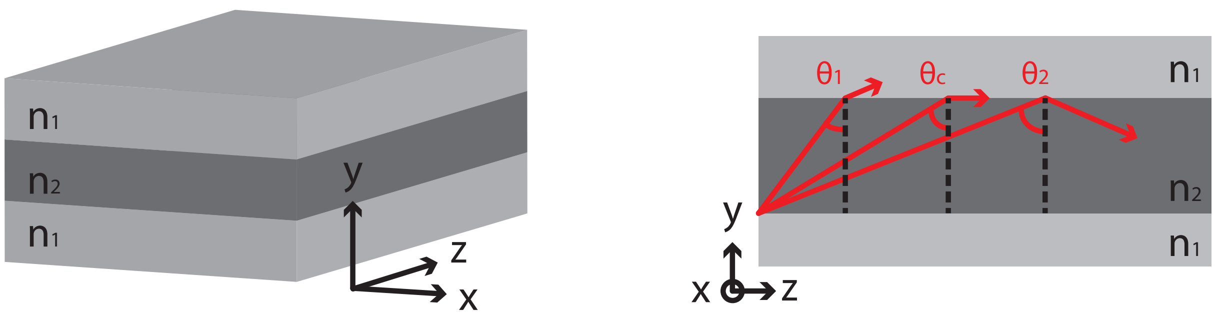

Fig. 3.1 The infinite dielectric slab model (left) consists of an infinitely wide, infinitely long slab of high-index material and finite height sandwiched between two slabs of lower-index material. Light traveling within the high index core (right) will escape or remain in the slab depending on the angle of incidence of that light with respect to the vertical axis.

The simplest model to describe dielectric waveguides is the slab waveguide model, illustrated in Figure 3.1. The model consists of three infinite dielectric slabs of defined thickness, with refractive indices n_1, n_2, and n_3, where $n_2 > n_1$, and $n_2 > n_3$. Reflection and transmission at each of the upper and lower interfaces are governed by Snell's law, whereby a wave with incidence angle θ_i with respect to the surface normal will be transmitted at angle θ_t according to the expression:

$$\sin\theta_t = \sin\theta_i \frac{n_i}{n_t}. \tag{3.9}$$

Where n_i is the refractive index of the incident medium, n_2 is the refractive index of the transmission medium. Note that for the case where $n_i > n_t$ there exists some incidence angle θ_c such that $\theta_t = \pi/2$. At this angle—termed the "critical angle"—the transmitted wave propagates directly parallel to the interface between the two slabs. Waves that arrive at incidence angles greater than the critical angle are then fully reflected back into the waveguide in a process known as "total internal reflection" (TIR). This condition may be verified by examination of the Fresnel reflection coefficients for polarized light oriented in parallel or perpendicular to the surface. The Fresnel intensity coefficients are given as:

$$R_{\parallel} = \left|\frac{n_i\cos\theta_t - n_t\cos\theta_i}{n_i\cos\theta_t + n_t\cos\theta_i}\right|^2 \tag{3.10}$$

$$R_{\perp} = \left|\frac{n_i\cos\theta_i - n_t\cos\theta_t}{n_i\cos\theta_i + n_t\cos\theta_t}\right|^2. \tag{3.11}$$

Taking the case where the polarization is perpendicular to the interface, $R_\perp$ the cosine terms $\cos\theta_t$ may be re-written as $\cos\theta_t = \sqrt{1-\sin^2\theta_t}$ and by Snell's law, $\sin^2\theta_t = \sin^2\theta_i \frac{{n_i}^2}{{n_t}^2}$. For angles of incidence greater than the critical angle, the resulting expression gives $\sin^2\theta_t > 1$, hence the cosine term $\cos\theta_t = \sqrt{1-\sin^2\theta_t}$ must be equal to some imaginary coefficient $\cos\theta_t = \pm j\alpha$. The Fresnel intensity coefficient may then be re-written as:

$$\begin{aligned} R_{\perp} &= \frac{n_i\cos\theta_i \mp jn_t\alpha}{n_i\cos\theta_i \pm jn_t\alpha}\cdot\frac{n_i\cos\theta_i \pm jn_t\alpha}{n_i\cos\theta_i \mp jn_t\alpha} \\ &= \frac{{n_i}^2\cos^2\theta_i + {n_t}^2(1-\sin^2\theta_t)}{{n_i}^2\cos^2\theta_i + {n_t}^2(1-\sin^2\theta_t)}. \end{aligned} \tag{3.12}$$

TIR is thus demonstrated by Equation 3.12 which reduces to $R = 1$ in all cases. Note that this expression is only achieved through the imaginary nature of the cosine term—for all cases in which $\theta_i < \theta_c$ the Fresnel reflection coefficients will be less than unity. Since light propagating through the waveguide at an angle less than the critical angle is perfectly reflected at each interface, and since the angle of reflection is equal to the angle of incidence, TIR light will remain within the center slab until it is either absorbed by the medium or scattered out of the core.

The TIR model of waveguide propagation appears to imply that the light (*e.g.*, the total field amplitude of all photons) is perfectly and completely confined to the center medium. In actuality, it can be demonstrated that there exists some portion of the field that is propagating in the upper and lower materials. This important result is the "evanescent field" that propagates outside of the waveguide core, and its derivation is shown below.

The time-invariant electric field of light propagating in a planar wavefront, with the electric field oriented along $\hat{y}$, is described by the planar wave solution to Maxwell's equations: $\vec{E} = \hat{y}E_0 e^{-j\vec{k}\cdot\vec{r}}$, where $\vec{k} = 2\pi n/\lambda$ is the wavevector of light propagating in a medium with refractive index n, and $\vec{r}$ is the position vector. Considering TIR light incident at the interface in Figure 3.1 with polarization oriented out of the page, the electric field of the transmitted wave may be written as

$$\begin{aligned} \vec{E_t} &= \hat{y}E_{yt}e^{-j\vec{k_t}\cdot\vec{r}} \\ &= \hat{y}E_{yt}e^{-jk_t(cos\theta_t x + sin\theta_t z)} \end{aligned} \tag{3.13}$$

where $\vec{k_t}$ is the wavevector transmitted through the upper cladding. $E_{yt} = t_{\parallel}E_{yi}$ is described by the Fresnel amplitude transmission coefficient $t_{\parallel} = \frac{2n_i\cos\theta_i}{n_i\cos\theta_i + n_t\cos\theta_t}$. Recalling from Snell's law, $\cos\theta_t = \pm j\alpha$, and $k_t\sin\theta_t = k_i\sin\theta_i$, Equation 3.13 may be rewritten:

$$\begin{aligned} \vec{E_t} &= \hat{y}E_{yi}\left(\frac{2n_i\cos\theta_i}{n_i\cos\theta_i \pm jn_t\alpha}\right)e^{-jk_t\sin\theta_t z}e^{-j(\pm jk_t\alpha)x} \\ &= \hat{y}E_{yi}\left(\frac{2n_i\cos\theta_i}{n_i\cos\theta_i \pm jn_t\alpha}\right)e^{-jk_i\sin\theta_i z}e^{\pm(k_t\alpha)x}. \end{aligned} \tag{3.14}$$

Noting that the solution $\cos\theta_t = +j\alpha$ results in the last exponential term increasing infinitely as $x \to \infty$, the only physical solution to Equation 3.14 is:

$$\vec{E_t} = \hat{y}E_{yi}\left(\frac{2n_i\cos\theta_i}{n_i\cos\theta_i - jn_t\alpha}\right)e^{-jk_i\sin\theta_i z}e^{-k_t\alpha x}. \tag{3.15}$$

Equation 3.15 thus describes a field with exponentially falling amplitude in the $\hat{x}$ direction. This component of the TIR light is referred to as the evanescent wave. Though the evanescent wave exists outside of the core of the waveguide it does not carry any power away from the waveguide. This can be verified by examining the Poynting vector of the evanescent wave $\langle \vec{S_t} \rangle = \langle E_t \times H_t \rangle$ which reveals that the evanescent wave carries power in the z (forward) direction only.

3.1.1.2 The dielectric box waveguide

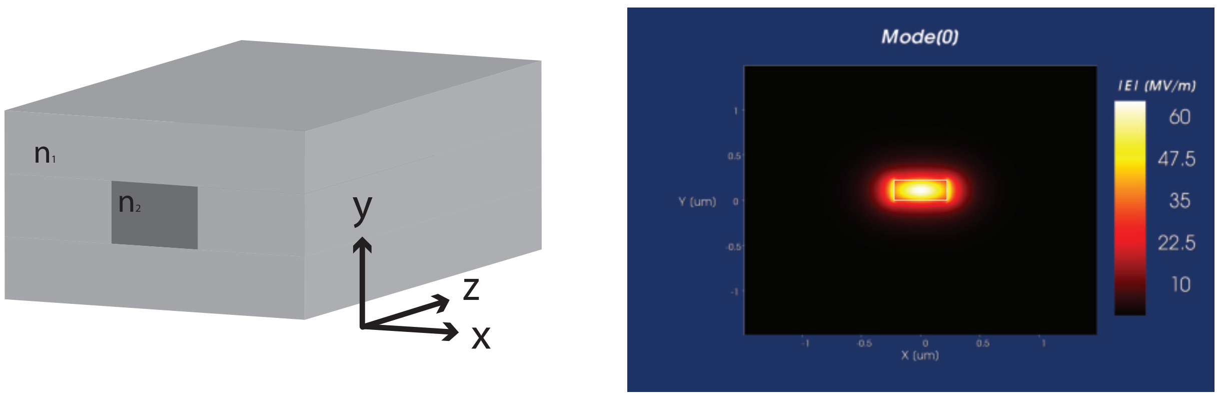

Fig. 3.2 A dielectric box waveguide (left) consists of a high-index infinitely long box of finite width and height encased in lower-index material (cladding). Light traveling within the high index core (right) will take on an intensity profile in the form of a spatial "mode" which may be evaluated numerically.

While the slab waveguide model presents an intuitive picture for understanding waveguide propagation, the infinite dielectric slab itself is not particularly useful waveguide structure as light cannot be bent or manipulated. Conventional optical waveguide devices, therefore, typically consist of a rectangular box structures, as shown in Figure 3.2, which permit optical routing over a printed photonic circuit due to confinement in both axes. This waveguide structure is similar to the infinite dielectric slab in that the waveguide core has refractive index greater than the surrounding upper and lower cladding materials. As previously discussed, the propagation of light through the waveguide is best described through the use of electric and magnetic field modes which satisfy Maxwell's equations.

Returning, for a moment, to the infinite dielectric slab model, it is important to note that not every incidence angle θ_i between 90° and the critical angle is supported. Consider a wave propagating in the dielectric slab with electric field oriented transverse to the plane of incidence (out of the page), and time-invariant electric field $\vec{E_i} = \hat{y}E_i e^{-j(k_x x + k_z z)}$. Upon reflection at the interface, this wave produces a reflected field $\vec{E_r} = -\hat{y}E_r e^{-j(-k_x x + k_z z)}$.

Assuming that $E_r = E_i = E_0$ these incident and reflected waves will combine constructively and destructively, resulting in a standing wave with the form:

$$\begin{aligned} \vec{E} &= \vec{E_i} + \vec{E_r} \\ &= \hat{y}E_0 e^{-jk_z z}(e^{jk_x x} - e^{-jk_x x}) \\ &= \hat{y}E_0 e^{-jk_z z}\sin(k_x x). \end{aligned} \tag{3.16}$$

The standing wave described in Equation 3.16 considers only the reflection off of the upper surface of the slab waveguide, however as the wave propagates in $\hat{z}$ it must reflect off of both surfaces, therefore the solution to Equation 3.16 must avoid self-interference between upward propagating and downward propagating standing waves. For a slab waveguide of thickness a this condition is achieved when a is equal to an integer number of characteristic wavelengths $a = m\frac{L_x}{2}$, where $L_x = \frac{2\pi}{k_x}$ and $m = 1, 2, \cdots$.

Since the wavevector in the $\hat{x}$ direction is given by $k_x = \frac{2\pi n_i}{\lambda}\cos\theta_i$, the allowed ("supported") values of m such that $a = m\frac{L_x}{2}$ define a specific set of angles at which a wave can propagate through the slab waveguide without encountering destructive self-interference. Each of these angles describes a particular solution to the wave equation in Equation 3.16, *i.e.* a waveguide mode.

This model also reveals that, for suitably small waveguides, there exists a wavelength above which no modes are supported. The condition for supporting a waveguide mode may be re-written as $a = m\frac{\lambda}{2n_i\cos\theta_i}$; taking the limit as $\cos\theta_i \to 1$, and $m = 1$, then the longest wavelength (the "cutoff" wavelength) supported by the waveguide is $\lambda_c = 2n_i a$.

Returning to the finite box waveguide shown in Figure 3.2, the waveguide modes must comprise electric and magnetic fields in all three dimensions $\vec{E}(x, y, z)$, $\vec{H}(x, y, z)$ that provide a solution to Maxwell's equations satisfying the boundary conditions at the edges of the waveguide core. An analytical solution to this waveguide is not straightforward, however with appropriate simplifications an approximate solution may be obtained [111]. Considering the electric and magnetic field distributions of the above waveguide, assuming that the mode propagates as a sinusoid in the $\hat{z}$ direction, with waveguide propagation constant $\beta = \frac{2\pi n}{\lambda}$:

$$\vec{E}(x,y,z) = [\hat{x}E_x(x,y) + \hat{y}E_y(x,y) + \hat{z}E_z(x,y)]e^{i\beta z} \tag{3.17}$$

$$\vec{H}(x,y,z) = [\hat{x}H_x(x,y) + \hat{y}H_y(x,y) + \hat{z}H_z(x,y)]e^{i\beta z}. \tag{3.18}$$

The transverse components of the fields may be collected as:

$$\vec{E}(x,y) = \hat{x}E_x(x,y) + \hat{y}E_y(x,y) \tag{3.19}$$

$$\vec{H}(x,y) = \hat{x}H_x(x,y) + \hat{y}H_y(x,y), \tag{3.20}$$

and the waveguide mode may be determined as a solution to the Helmholtz equation:

$$\nabla \times \nabla \times \vec{E}(x,y,z) = \frac{\omega^2}{c^2}n^2(x,y)\vec{E}(x,y) \tag{3.21}$$

where $n(x,y) = \sqrt{\eta(x,y)\mu(x,y)}$ is the local refractive index distribution that defines the waveguide structure. Breaking out the transverse terms of the gradient operator as $\nabla_t = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y}$ then from $\nabla \times \vec{E} = i\omega\mu\vec{H}$, and using the property of the cross-product that the $\hat{z}$ component of $\vec{E}$ cannot contribute to the $\hat{z}$ component of $\vec{H}$, H_z may be written:

$$(\nabla_t \times \vec{E}_t) \cdot \hat{z} = i\omega\mu H_z \tag{3.22}$$

$$H_z = \frac{(\nabla_t \times \vec{E}_t) \cdot \hat{z}}{i\omega\mu}. \tag{3.23}$$

Similarly, it can be shown that the $\hat{z}$ component of the electric field may be determined using only the transverse components of the $\vec{H}$ field:

$$E_z = \frac{(\nabla_t \times \vec{H}_t) \cdot \hat{z}}{-i\omega\eta n^2}. \tag{3.24}$$

This indicates that knowledge of the transverse electric and magnetic fields is sufficient to determine the behaviour of the waveguide in the z-dimension (*i.e.* the direction of propagation). Simplifying the problem of obtaining a solution to Equations 3.21 to consider only transverse terms, the waveguide mode may be obtained as a solution to the equation:

$$\nabla \times \nabla \times \vec{E} = \nabla(\nabla \cdot \vec{E}) - \nabla^2\vec{E} = \frac{\omega^2}{c^2}n^2(x, y)\vec{E}. \tag{3.25}$$

Equation 3.25 may be written in terms of transverse components as:

$$\nabla_t(\nabla \cdot \vec{E}) - \nabla^2 \vec{E}_t e^{i\beta z} - \frac{\omega^2}{c^2} n^2(x,y) \vec{E}_t e^{i\beta z} = 0. \tag{3.26}$$

Writing the gradient of the electric field as a combination of transverse and $\hat{z}$ components $\nabla \cdot \vec{E} = \nabla_t \cdot \vec{E}_t e^{i\beta z} + i\beta E_z e^{i\beta z}$ and using Equation 3.1 with $\rho = 0$ it is possible to obtain an alternate expression for the first term in Equation 3.26

$$\nabla_t \cdot (n^2 \vec{E}_t) + i\beta n^2 E_z = 0 \tag{3.27}$$

$$\nabla \cdot \vec{E} = \nabla_t \cdot \vec{E}_t e^{i\beta z} + i\beta E_z e^{i\beta z} \tag{3.28}$$

$$\nabla \cdot \vec{E} = \nabla_t \cdot \vec{E}_t e^{i\beta z} - \tfrac{\nabla_t \cdot (n^2 \vec{E}_t)}{n^2} e^{i\beta z} \tag{3.29}$$

$$\nabla_t(\nabla \cdot \vec{E}) = \nabla_t(\nabla_t \cdot \vec{E}_t) e^{i\beta z} - \nabla_t \left[\tfrac{\nabla_t \cdot (n^2 \vec{E}_t)}{n^2} \right] e^{i\beta z}. \tag{3.30}$$

The second term in Equation 3.26 may also be expanded as

$$\nabla^2 \vec{E}_t e^{i\beta z} = {\nabla_t}^2 \vec{E}_t e^{i\beta z} - \beta^2 \vec{E}_t e^{i\beta z}, \tag{3.31}$$

leading to the re-written form of Equation 3.26:

$$-{\nabla_t}^2 \vec{E}_t + \nabla_t \left[\nabla_t \cdot \vec{E}_t - \frac{\nabla_t \cdot (n^2 \vec{E}_t)}{n^2} \right] - \frac{\omega^2}{c^2} n^2 \vec{E}_t = -\beta^2 \vec{E}_t. \tag{3.32}$$

Finally, taking advantage of the linearity of the gradient operator, Equation 3.32 can be re-written in the form of an eigenvalue problem:

$$\begin{bmatrix} \hat{A}_{xx} & \hat{A}_{xy} \\ \hat{A}_{yx} & \hat{A}_{yy} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} = -\beta^2 \begin{bmatrix} E_x \\ E_y \end{bmatrix}. \tag{3.33}$$

The four terms on the left hand side of Equation 3.33 are:

$$\begin{aligned}
\hat{A}_{xx}E_x &= -\frac{\partial}{\partial x}\left[\frac{\partial(n^2E_x)}{n^2\partial x}\right] - \frac{\partial^2 E_x}{\partial y^2} - \frac{\omega^2}{c^2}n^2E_x & (3.34)\\
\hat{A}_{xy}E_y &= -\frac{\partial}{\partial x}\left[\frac{\partial(n^2E_y)}{n^2\partial y}\right] - \frac{\partial^2 E_y}{\partial x \partial y} & (3.35)\\
\hat{A}_{yx}E_x &= -\frac{\partial}{\partial y}\left[\frac{\partial(n^2E_x)}{n^2\partial x}\right] + \frac{\partial^2 E_x}{\partial y \partial x} & (3.36)\\
\hat{A}_{yy}E_y &= -\frac{\partial}{\partial y}\left[\frac{\partial(n^2E_y)}{n^2\partial y}\right] - \frac{\partial^2 E_y}{\partial x^2} - \frac{\omega^2}{c^2}n^2E_y & (3.37)
\end{aligned}$$

Equation 3.33 can then be solved numerically to provide the transverse components of $\vec{E}$. The transverse components of the mode define the electric field, from which the propagation constant β may be derived. The z-component of the field $E_z(x, y)$ can then be obtained from $\vec{E}_t(x, y)$ using Equation 3.27, and the corresponding magnetic field may be obtained using Equation 3.3 to complete the solution of the waveguide mode.

From this example, and the previous slab waveguide model above it is evident that it may be possible to obtain analytical expressions for guided modes for simple waveguide structures, but 2D and 3D problems rapidly scale in complexity—even for relatively simple structures.

For more complex waveguide structures, such as multi-core slabs, ridge waveguides, or slot waveguides it is more convenient to determine the guided modes using computational methods. Though these complex modal fields do not have convenient analytical expressions they are nonetheless still subject to Maxwell's equations and the field-matching conditions at the boundaries of the waveguides. With these starting points, and enough computational effort, the modal fields may then be determined. The numerical methods that have been developed to solve this problem are generally referred to as "mode solvers". A brief summary of common mode solver algorithms is presented in the following sections, a more comprehensive overview is available in the literature. [109, 110]

An interesting additional case of the dielectric box waveguide are subwavelength waveguides which are used to develop the device presented in Chapter 5. A subwavelength waveguide consists of a periodic structure of "gaps" imposed on a dielectric core waveguide, producing a waveguide structure that resembles a dashed line or a diffraction grating (see Figure 5.1). In subwavelength waveguides, the spacing of the gaps and dielectric cores which produce the waveguide structure is kept well below the effective wavelength of light in the waveguide material; in this case the light propagates through the waveguide in a Floquet-Bloch mode, wherein the underlying plane-wave function is modulated in amplitude by the periodic structure of the high and low index regions of the subwavelength structure. Since the periodicity of this structure is lower than the diffraction-limit, the propagating wave does not experience this structure as a set of scattering sites, but rather as a medium with an effective index governed by a combination of the two refractive indices comprising the periodic structure. Further details on these complex structures and examples of their uses may be found in the literature. [112]

3.1.1.3 Effective index method

The effective index method (EIM) [113] is a common approximation to Maxwell's equations. It is commonly used to reduce the dimensionality of a complex waveguide structure into a set of slab waveguides which have simple solutions. Consider the waveguide structure shown in Figure 3.3: an EIM calculation proceeds by breaking the structure into three separate vertical stacks, two identical stacks on either side of the waveguide core, and one stack containing the core itself.

For each of these stacks, TM and TE modes can be computed straightforwardly to yield the propagation constant of the slab waveguide mode β_i (as done by hand in Section 3.1.1.2). The propagation constant of each slab defines an effective material refractive index as $\beta = \frac{2\pi n_{eff}}{\lambda}$. The three effective indices are then collected as a horizontal stack of materials with indices n_{eff1}, n_{eff2}, n_{eff1} and width equal to the width of the waveguide core. These slabs define another dielectric slab albeit one that is rotated by 90° with respect to the first three. The propagation constant of this slab of effective materials is then—to a good approximation—equivalent to the propagation constant of the original waveguide.

Though the effective index method does not provide the modal field distribution for the waveguide, the propagation constant itself is an important parameter for waveguide design. A priori knowledge of the propagation constant can be used as a starting point for more advanced numerical mode solvers, shortening the overall computation time. In addition, the EIM is often used in mode propagators to reduce a 3D waveguide layout to a 2D layout with effective "core" and "cladding" indices. This transformation substantially simplifies the propagation routine—enabling larger or more complex layouts to be simulated with reduced computational overhead.

3.1.1.4 Finite difference

The most rigorous computational algorithm is the method of finite differences (FD) [114]. In this scheme, the waveguide core is subdivided into a high-density grid or mesh as shown in Figure 3.3. This grid extends from the core of the waveguide out into the cladding and substrate material; the size of the grid is selected such that the field conditions at the edge of the grid can be reasonably assumed to be zero (*i.e.* the evanescent field of the guided mode has become negligible). Often this field condition at the boundary is enforced by the solver using a "perfectly matched layer".

The grid is defined with sufficient density that the refractive index in each cell is constant (trivial for rectangular waveguides but potentially more complex for rounded

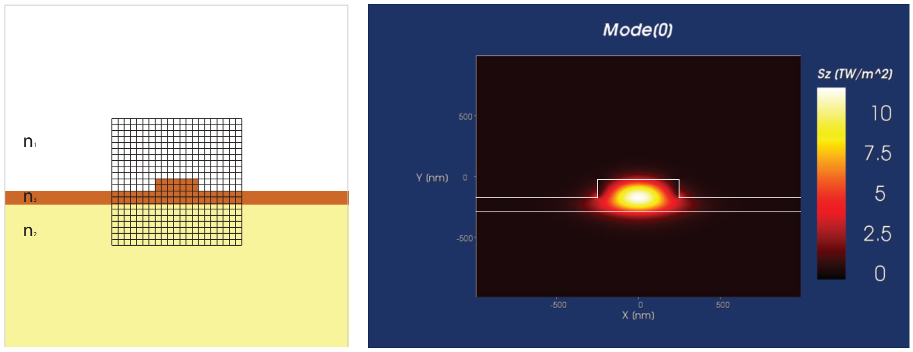

Fig. 3.3 A Silicon "ridge" waveguide without upper cladding (left) is shown along with a representative grid which would be used to discretize the waveguide structure in order to solve for the waveguide mode via the finite-differences method. The waveguide mode (right) as obtained using an FD solver, has a "bell" shape that would be difficult to describe via an analytic solution.

waveguides). Rather than describing the fields in the waveguide structure using partial differential equations, the FD method instead takes the difference in field amplitudes between surrounding grid points to construct the first and second order field derivatives at each grid point. The mode solver then iterates the field amplitude at each point in the grid in order to find a distribution of field amplitudes that satisfies the Helmholtz equation. The method of implementing the iterative search may vary from solver to solver, however the underlying principle remains consistent.

The advantage of numerical methods, such as FD, over approximate solvers, such as EIM, is that the accuracy of the mode solution provided by FD can always be improved simply by adding more grid points (at the expense of computation time and efficiency). Numerical methods such as FD directly attack Maxwell's equations, which provides a rigorous solution to the waveguide mode. By contrast, approximate methods such as EIM provide exact solutions to approximations of Maxwell's equations—and therefore may not be accurate in all cases.

3.1.2 Propagation of guided modes

Having established the methods for computing the guided electric field modes of a particular waveguide, it remains to discuss the propagation of these modes through the waveguide structure. The waveguide propagation constant β describes the speed at which different modes travel through a given waveguide structure. This behaviour is mathematically captured in the time-invariant mode Equation 3.18, in which the time-invariant transverse field is multiplied by the propagation factor $e^{i\beta z}$. Incorporating time-evolution into this equation is achieved by multiplying by the time-dependent propagation factor $e^{-i\omega t}$.

As with mode solvers, propagation of modes through waveguide structures is another problem that is best-addressed through the use of computational methods. In free-space optical systems, an entire optical system may be facilely simulated using simple ray-tracing, or transfer-matrix, or Fresnel propagation software (*e.g.*, Zemax, Code V). However, in waveguide optics, the computational overhead associated with mode propagation is so significant that mode behaviour is only evaluated for particular sub-systems within a waveguide circuit (*e.g.*, splitters, couplers, filters, etc.). Moreover, as all waveguide propagators are computationally intensive—varying in magnitude but not in kind—it is important for a waveguide designer to select the correct solver for a given subsystem within the circuit.

In this section the advantages, disadvantages, and typical use-scenarios of common "mode propagators" are presented; a more comprehensive treatment is available in the literature. [109, 110]

3.1.2.1 Beam propagation method

The simplest and most common mode propagation method is the aptly-named beam propagation method (BPM). BPM is implemented using a finite difference approximation, beginning with the Helmholtz scalar wave equation:

$$\nabla^2\Psi(x,y,z) + \beta^2\Psi(x,y,z) = 0 \tag{3.38}$$

and making the paraxial approximation that the amplitude of the propagating wave $\psi(x,y,z)$ varies slowly in $\hat{z}$ compared with the length-scale of the wavelength of interest ($2\pi\lambda/n_{eff}$).

BPM can be implemented as a 3D problem or as a 2D problem using refractive index values supplied by EIM. The typical advantage of BPM, particularly 2D BPM, is that it is a computationally fast algorithm that delivers accurate results for modelling the performance of single-mode waveguide splitters, directional couplers, and Mach-Zehnder interferometers. Typically BPM is used for the simulation of large, relatively simple waveguide structures with low-index contrast. Thanks to its low computational overhead it is possible to use BPM to simulate much larger waveguide devices than would be possible using other methods; however the paraxial approximation is not suitable for the simulation of more complex structures with significant variance and structure in the refractive index.

One additional disadvantage of BPM is that since BPM computes the scalar field it is not possible to distinguish between incoming and outgoing light. While this drawback may not impact simulations of certain structures (*e.g.*, directional couplers), the inability to interrogate the directionality of the propagating mode makes BPM unsuitable for the design of more complex structures in which reflections must be identified and understood.

3.1.2.2 Eigenmode expansion

In the case where both computational efficiency and bi-directionality are desired (e.g. multi-mode interference devices), the eigenmode expansion method (EME) may be used. In EME, the waveguide structure is discretized along the z-axis into a series of non-uniform slices, each with its own refractive index. In each slice in the discretized structure, the waveguide modes may be determined using any of the mode-solver techniques listed above. Each optical mode supported by the slice is an eigenfunction of the Helmholtz equation with eigenvalue β_i defining the propagation constant of the mode:

$$\vec{E}(x, y, z) = \vec{E}_i(x, y) \cdot e^{i\beta_i z}. \tag{3.39}$$

Differing propagation constants between waveguide modes (eigenmodes) set up harmonic z-dependencies for the electric field within the waveguide. In aggregate, the eigenmodes for each slice define a basis set for all possible solutions to Maxwell's equations that are supported by the slice. The eigenmodes of each slice may propagate either forwards or backwards through the structure, allowing the set of eigenmodes to describe any possible solution to Maxwell's equations in the slice.

The field in each slice in the discretized structure may then be expanded into its constituent basis modes, and the modes may be propagated through each slice (with minimal computational effort) according to Equation 3.39. At the end of the slice, the basis modes may be added together (considering their accumulated phase) to produce the overall output field distribution. At the interface between two slices, the mode overlap coefficients for each mode in the bases of each slice may be used to define a scattering matrix for the interface. This scattering matrix determines the fraction of light that is transmitted and reflected for each mode in the first slice, as well as the resultant weighting of propagating modes in the second slice. In an EME propagator, then, scattering matrices are generated for each interface in the waveguide structure, and the overall propagation of light through the structure may be determined by progressive multiplication of these scattering matrices. Since an interface need only be defined where there is a change in the index of a slice, individual slices may be very large without affecting accuracy (*e.g.*, the central cavity of an MMI). For this reason, it is often best to avoid the use of tapering or bending structures in EME simulations.

As mentioned, the EME method has the advantage of providing a robust solution to the propagation problem while avoiding much of the discretization effort involved in more computationally-intensive methods. Another advantage of the EME method is that the phase and amplitude of each mode within each slice of the structure is recorded throughout the calculation and may be preserved for post-processing. This facility is particularly useful in the design of multi-mode devices since it enables the waveguide designer to optimize the design to produce particular modal behaviours or enforce a desired interaction between modes (*e.g.*, to combine in-phase or out-of-phase at a particular location). Finally, the EME method enables efficient modelling of periodic or symmetric structures since the eigenmode expansion and scattering matrices of these structures need only be calculated for the first period in the repeating structure and the scattering matrix may be re-used thereafter.

A disadvantage of the EME method is in simulating optical devices with very large cross sections. As the size of the waveguide cross section increases relative to the optical wavelength, the number of basis modes in each slice increases and the corresponding scattering matrices may become extremely large. For this reason, EME is best restricted to cases involving smaller waveguides—or large structures with very few interfaces (slices).

3.1.2.3 Finite-Difference Time-Domain

The most accurate electromagnetic wave propagation method is the finite-difference time-domain method (FDTD), which is based on a discrete (i.e. finite-difference) representation of the time-dependent Maxwell's equations. As in the case of the finite-difference mode-solver, the FDTD method relies on discretizing the waveguide structure into a grid. Unlike the FD mode-solver, the FDTD method demands the presence of an additional dimension of discretization: time. At each time-step in the computation, the electromagnetic field amplitude in each spatial grid point is determined by the field amplitude at the current time propagated forward by one step. Despite its limitations in terms of computational power, the FDTD method confers significant advantages to the waveguide designer. The exactness of the solution means that FDTD methods are ideal for the simulation of complex photonic structures such as photonic crystals, ring-resonators, and high-index contrast waveguides. Moreover, since the output of the FDTD simulation is a time-series, the FFT of the time-dependent electromagnetic field may be used to reveal wavelength-dependent properties of the photonic structure (e.g. bandgap). This powerful property of the FDTD method makes it an extremely versatile tool for waveguide and photonic crystal design.

As noted, the introduction of an additional dimension to the simulation (time) makes FDTD highly computationally intensive—especially for 3-D structures. As such, FDTD is most commonly used for the simulation of small, complex structures, or for sub-samples of infinitely repeating structures (e.g. subwavelength gratings, grating couplers, photonic crystals, etc.). Even for small structures, however, this computational overhead is significant enough that FDTD is often not used as a tool to optimize waveguide structures; rather, FDTD is best employed as a "final check" on the performance of a structure that has already been optimized through computationally cheap methods such as BPM or EME.

3.2 Design of on-chip FTS

Having discussed the computational tools used for waveguide design, the remainder of this chapter presents the details of the design process that was followed throughout this work in order to produce the waveguide devices described in Chapters 4, and 6. The discussion primarily focuses on the design of the athermal waveguide spectrometer described in Chapter 4, the layout of which is shown in Figure A.3. In the following subsections the design of sub-elements within this layout (waveguides, splitters, Mach-Zehnder interferometers) is presented; most of this work was performed using the waveguide simulation tool OptoDesigner [115].

Additional information regarding the overall spectrometer layout, including layout algorithms, and fabrication processes themselves are captured in Appendix A.

3.2.1 Waveguide core design

For different reasons, the design of both the methane spectrometer in Chapter 4 and visible CS spectrometer in Chapter 6 required the development of a custom waveguide core geometry. Custom waveguide development is not necessary for all PIC spectrometers. Indeed, for applications in optical communication waveguide designers may take advantage of many standardized waveguide platforms in which the waveguide core geometry is pre-defined [107]. In the case of the methane spectrometer, however, a custom core geometry was pursued in order to reduce the thermal sensitivity of the SHFTS. In the CS spectrometer a custom waveguide geometry was required to accommodate the large operational bandwidth of the spectrometer.

Development of a custom waveguide geometry has impacts on the optical layout that must be assessed by the optical designer. In particular, the core geometry will determine the waveguide propagation constant (*i.e.* the mode index), propagation loss, minimum bending radius, and minimum waveguide separation. In the following subsection the development of the core geometry for both the CS and methane instrument is discussed, as well as the impacts of those geometries on the optical layout.

1. *Waveguide confinement*

   Broadly speaking, single-mode waveguide cores may be distinguished as low-confinement or high-confinement cores. In a high-confinement core, much of the mode-field amplitude distribution is contained within the high-index core of the waveguide, whereas in a low-confinement core the majority of the mode-field amplitude is contained within the cladding material.

   High-confinement cores are advantageous because they may achieve very small bending radii, resulting in high-density on-chip arrangements. A high-degree of core confinement, however, may also result in increased propagation loss due to scattering of the intense field at the rough waveguide sidewalls. In any waveguide core, there exists a discontinuity in the mode-field amplitude at the edges of the waveguide. This mismatch results from the requirement, imposed by Maxwell's equations boundary conditions, that the field amplitude be equal on either side of the boundary; this mismatch is evident in the mode field shown in Figure 3.2. In a high-confinement core this discontinuity is significant, resulting in a sharp change in field amplitude, which is easily scattered by the unavoidable surface roughness of the waveguide.

By contrast, low-confinement cores experience less disruption at the interface between the waveguide core and the cladding. As a greater amount of the mode intensity propagates through the cladding material there is less disruption due to scattering, so very-low propagation losses may be achieved [107]. However, since the mode is not as well-confined to the core, the waveguide mode will have a greater propensity to leak into the cladding material—especially if the waveguide is bent. Low-confinement cores have the additional advantage that they may achieve single-mode operation over a greater range of wavelengths, thereby enabling devices with large optical bandwidths.

2. *Waveguide separation and crosstalk*

The development of a novel waveguide core requires investigation of the minimum waveguide separation—meaning the minimum separation distance between waveguides in order to avoid crosstalk. The confinement of the fundamental waveguide mode sets a lower bound on the minimum separation between waveguides, with lower-confinement cores experiencing greater crosstalk than higher-confinement cores. Crosstalk between two closely-separated waveguides occurs when the evanescent mode of one waveguide (*i.e.* the portion of the fundamental mode propagating in the cladding—see Section 3.1) overlaps

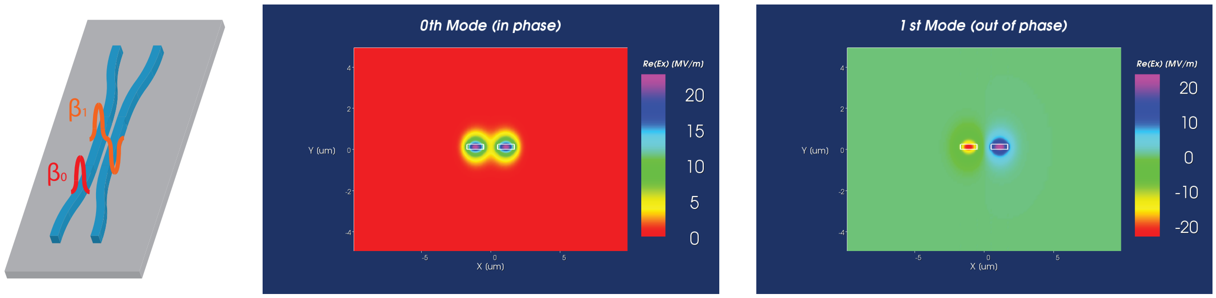

Fig. 3.4 Two waveguides in close proximity (left) may sustain hybrid or "super" modes (right) in which the electric field is shared between the two waveguide cores. These higher order modes propagate at a different rate than the fundamental waveguide mode (center), and will interfere constructively and destructively as they co-propogate.

with the evanescent mode of the second waveguide—resulting in excitation of a "super-mode". Excitation of the second waveguide results in equal suppression of the first mode, eventually leading to a complete transfer of the fundamental mode from the first waveguide to the second. The transfer of modes between two adjacent waveguides may be used to excite a resonant cavity such as a ring-resonator, or may be used as a waveguide splitter or combiner. In general, however, unplanned waveguide coupling is undesirable. A common mitigation strategy, therefore, is to set a design rule separating all waveguides on a chip by some minimum amount.

For two adjacent waveguides, the "coupling length" is the interaction length required to generate a complete transfer of the waveguide mode. This length may be determined by simulating the propagation of light through two waveguides, or by calculation of the fundamental and super-modes of an adjacent waveguide structure.

As shown in Figure 3.4, solving for the confined modes of a system with two adjacent waveguides produces super-modes in which the mode amplitude is simultaneously distributed between the two waveguide cores. The first two super-modes are the in-phase (even) and out-of-phase (odd) modes. The first mode describes propagation of the fundamental modes of each waveguide, whereas the second mode describes a mode in which the field amplitude is shared between the waveguides in such a way that the field amplitude in the first waveguide is 180° out-of-phase with the field amplitude of the second. When two waveguides are placed close together, both modes will propagate, and the difference in propagation constants between the modes will determine the coupling length as $L_\pi = \pi/(\beta_0 - \beta_1)$.

3. *Waveguide bend radius*

In addition to waveguide separation, the layout rules governing minimum bend radius must be established before proceeding to optical layout. In both the methane and Raman spectrometers the minimum waveguide bend radius was

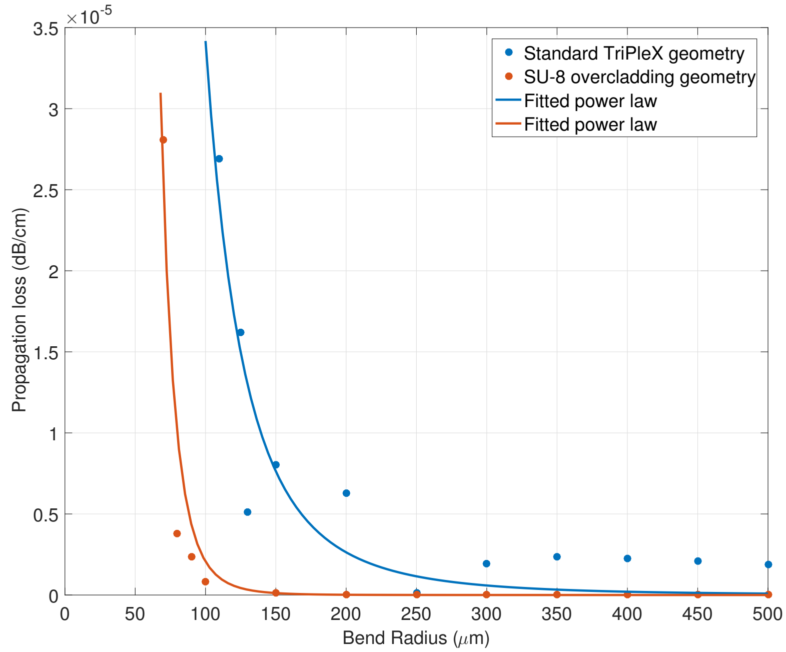

Fig. 3.5 Simulated waveguide bend loss, determined using the complex component of the refractive index of the fundamental waveguide mode, at decreasing bend radii for standard TriPleX [107] geometry, as well as the athermalized core geometry. Trendline added as a visual aid. A minimum bend radius of 250 μm was selected as a design rule for conservatism.

assessed by examining the complex component of the refractive index of the waveguide mode at differing bend radii. In OptoDesigner [115] this was accomplished using the finite difference mode solver. The complex component of refractive index governs the propagation loss of the waveguides and can be expressed in dB of loss per centimetre of propagation. The relationship between the bend radius and propagation loss of the waveguide mode for the methane chip is shown in Figure 3.5. At large bend radii there is very little impact on the propagation loss, however, it may be seen that as the bend radius is decreased, the complex value of the refractive index rapidly increases—indicating greater propagation loss through leakage of the mode into the cladding.

3.2.2 Splitter design

As illustrated in Figure 3.6, optical splitting from one to two (or more) waveguides is typically performed using either (1) Y-splitter devices (2) directional couplers or (3) multimode interference (MMI) devices.

Y-splitter devices, in which the waveguide is physically separated into two waveguides by a junction, are straightforward to design and implement, however they may suffer from substantial reflections and scattering at the junction point, and may exhibit imbalance between arm outputs. A perfectly adiabatic Y-junction would achieve low reflection and insertion loss, however in practice an adiabatic junction cannot be achieved due to the finite minimum feature size of fabricated waveguides—resulting in a "bunt" at the center of the junction as shown in Figure 3.6.

Directional couplers operate on the principle of evanescent mode coupling, whereby the electric field propagating outside the core of the first waveguide gradually excites the fundamental mode of the adjacent waveguide. As the two modes propagate, light

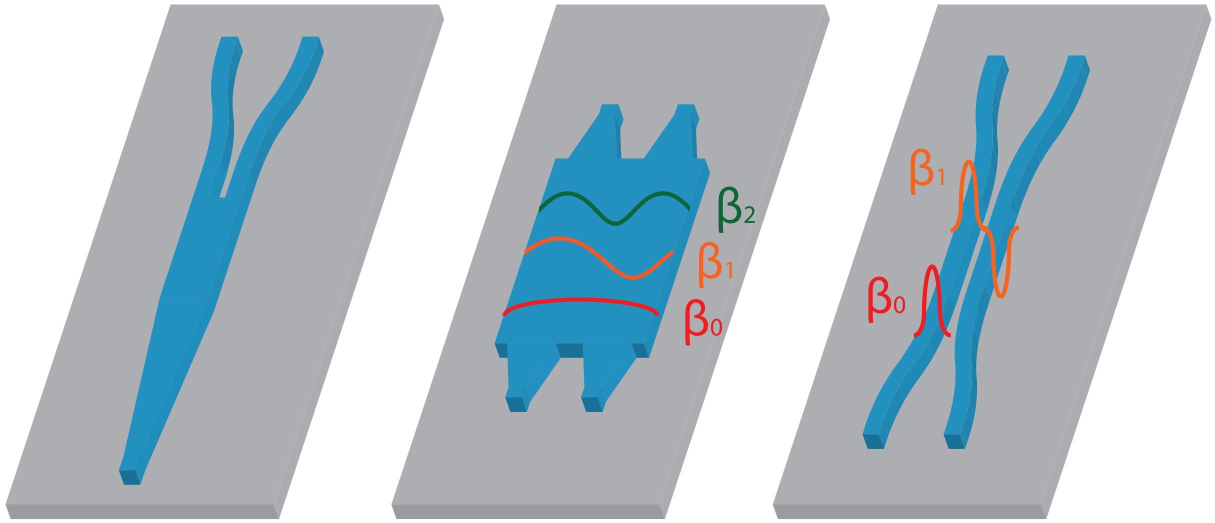

Fig. 3.6 Diagram illustrating the three most common varieties of 50:50 splitters. The Y-splitter (left) consists of a single entrance port which widens to accomodate two exit ports; due to minimum feature-size constraints, there is often a lossy "bunt" at the join of the two exit ports. The MMI splitter (center) relies on the interference of multiple modes propagating within the central cavity to produce and even division of light at the outputs. The directional coupler (right) takes advantage of the evanescent field effect described in Section 3.2.1.

from the first mode is gradually siphoned into the adjacent waveguide until the only the second mode remains—at which point the process repeats. If the mode-transition is interrupted by separating the waveguides at the point where the power is split evenly between waveguides, then 50 : 50 splitting may be achieved. Directional couplers achieve low insertion loss relative to Y-splitters, however the propagation distance required to achieve 50 : 50 separation may be significant.

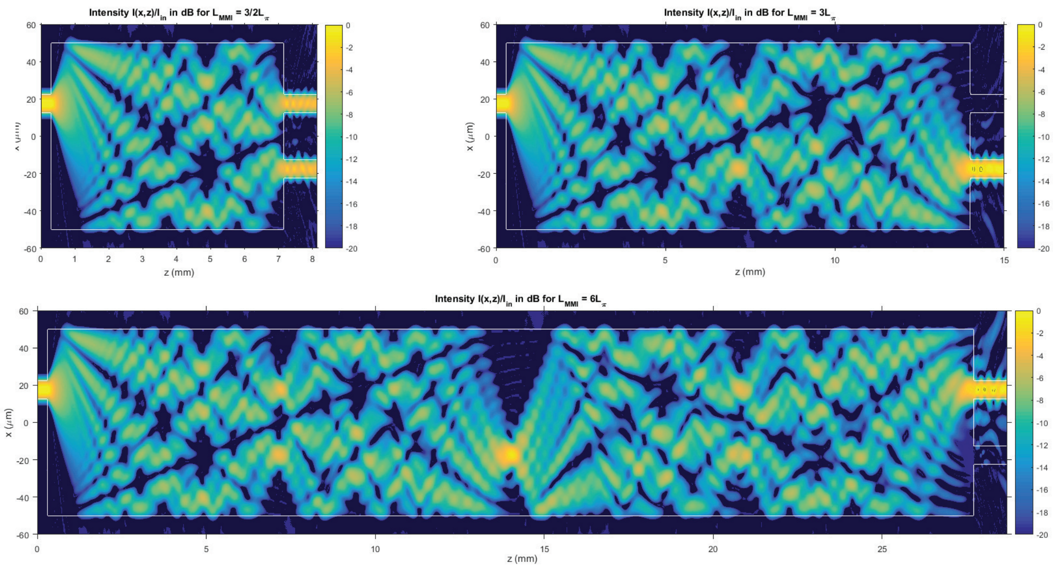

Fig. 3.7 Beam propagation showing evolution of field intensity (colour map) and mode re-imaging within MMI cavities of varying lengths. The first device (upper left) shows an MMI cavity at $L_{MMI} = 3/2L_{\pi}$, resulting in 50:50 splitting. The second device (upper right) shows an MMI cavity with $L_{MMI} = 3L_{\pi}$ resulting in complete transfer of the input mode to the opposing exit port. The third device (bottom) shows an MMI cavity with $L_{MMI} = 6L_{\pi}$ in which the input mode is re-imaged at the same exit port.

MMI devices achieve splitting by exciting multiple modes in an enlarged central cavity. These modes then interfere constructively at the exit ports producing a split of the input power. Each of the modes within the cavity propagates independently according to their modal propagation constants, $\beta_i = \frac{2\pi n_i}{\lambda}$, where n_i is the mode index of the i-th mode. The difference in modal propagation constants $\Delta\beta = \beta_1 - \beta_2$ is consistent between modes, and defines a specific length (the "beat length") at which all modes will combine in-phase:

$$L_\pi = \frac{\pi}{\beta_0 - \beta_1}. \tag{3.40}$$

By changing the length of the central cavity of the MMI, constructive and destructive interference between modes can be designed to split or re-image the input field. At $L_{MMI} = 3L_\pi$, for instance, all modes are in-phase at the output plane of the MMI, and light from the input waveguide is re-imaged, as shown in Figure 3.7. Alternatively, at $L_{MMI} = \frac{3}{2}L_\pi$ half the modes are in-phase and half the modes are out-of-phase, resulting in $50:50$ splitting at the exit plane. The two outputs of the $50:50$ splitter shown in the upper left of Figure 3.7 are referred to as the in-phase (I) and quadrature (Q) signals, indicating that the phase-difference between the I and Q waveguides is $\Delta\phi = 90°$ ($e.g.$ $\vec{E}_I = i\vec{E}_Q$).

One disadvantage of MMI splitters relative to Y-splitters or directional couplers is that the design effort to produce a functional MMI may be significant—especially for rib-waveguides. However, the advantage of MMI splitting in SHFTS over Y-splitters is that the the phase relationship between I and Q outputs of MMI devices permits signal normalization and, more generally, MZI phase correction through FTS oversampling [116,117]. The derivation of this effect is presented formally in Section 3.2.3. An additional advantage of MMI splitters is that the design of the splitter is typically not limited by fabrication constraints. In the case of the Y-splitter, the bunt produced by the minimum lithographic feature size results in an insertion loss penalty; in the case of the directional coupler, the minimum feature size governs minimum waveguide separation–which may in turn produce very large photonic devices. Since MMI devices are relatively insensitive to the minimum feature size they may achieve highly efficient splitting in a very compact form.

MMIs for the athermal spectrometer presented in Chapter 4 were developed using the OptoDesigner software. The overall MMI design, and the general design flow are shown in Figure 3.8. Like many aspects of waveguide design, there are no hard "rules" in the design of these cavities, and the design flow is not a strictly linear process. Rather, MMI design is an iterative process in which the designer explores a multi-dimensional trade space, repeatedly evaluating the impact of design parameters (cavity width, waveguide separation, etc.) on overall system performance. With that said, one straightforward method for the design of these devices proceeds as follows.

First, the width of the MMI cavity is selected such that multiple modes may be supported in the central cavity. As the mode from the input waveguide enters the MMI cavity, it is projected onto each of the modes supported by the central cavity (as in a Fourier decomposition). These modes then propagate through the MMI and

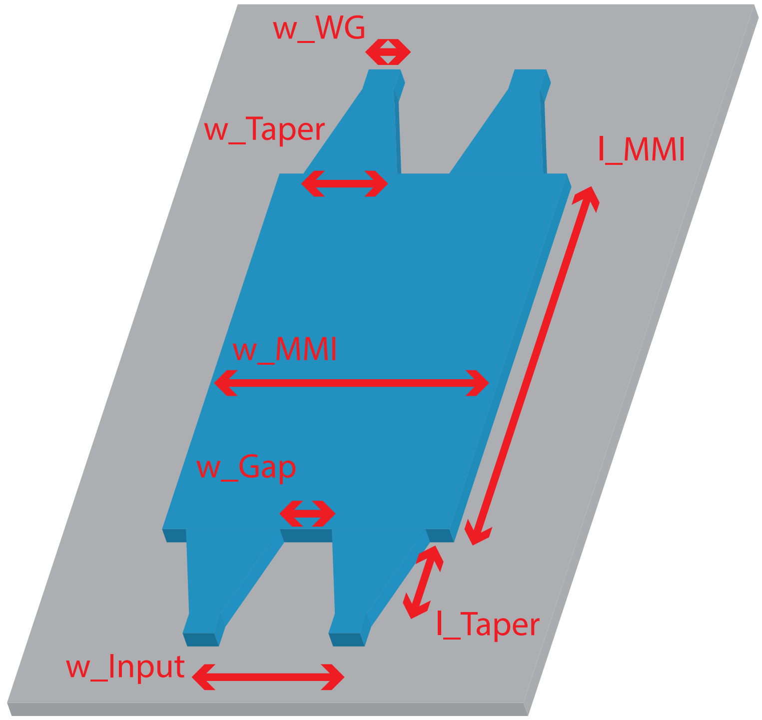

Fig. 3.8 A 2x2 MMI coupler depicting the key parameters of an MMI splitter. A typical design flow begins with the width of the MMI cavity, which is designed to support 9 or fewer modes. The locations of the input ports are offset symmetrically from the centerline by approximately $\pm w_{MMI}/6$, and the width of the access tapers are maximized until the point where cross-talk occurs. Finally, L_{MMI} is optimized in the neighbourhood of $L_{MMI} = 3/2L_{\pi}$ to achieve an even coupling ratio.

are re-imaged on the exit plane according to the beat length of the cavity. This modal decomposition becomes more accurate with an increasing number of modes in the central cavity—similar to a Fourier decomposition with a greater number of samples. One would naively intuit, that a wider MMI device, supporting a greater number of modes, would yield improved image quality at the exit plane. In reality, image quality is controlled not only by the number of modes in the decomposition, but also by their phase at the image plane. In other words, the beat-length between modes i, and $i+1$ should be identical for all modes—which is rarely the case. In practice, the beat length relationship begins to deteriorate as the number of modes increases. Many MMI designs, therefore, select a cavity width that supports 9 or fewer modes [118]. Narrowing the MMI cavity has the further advantage of reducing overall cavity length—further reducing the impact of beat-length non-uniformity. Narrower MMIs supporting fewer modes exhibit greater differences between mode indices, thus yielding a shorter beat length via Equation 3.40.

Secondly, once the cavity width has been selected, the separation between input couplers, and the width of the access ports must be selected. A common value for the offset of the access waveguides from the centerline is one-sixth of the effective width of the MMI cavity [118, 119], or approximately $\frac{1}{6}w_{MMI}$. From this starting point, the width of the tapers may be adjusted to produce the best coupling to the MMI cavity while minimizing crosstalk. As described previously, directional coupling may occur when two waveguides are placed in close proximity; left un-checked in an MMI device, directional coupling between the input waveguides leads to crosstalk and imbalance at the MMI outputs. Accordingly, it is important to separate input waveguides with sufficient margin to avoid cross-talk. Since the baseline separation between the waveguides is related to the cavity width, there may be a trade-off between cavity size (*i.e.*, number of modes) and crosstalk.

Waveguide crosstalk may also be mitigated by reducing the length over which access waveguides are in close proximity, which may be achieved y reducing the waveguide taper. While the decision to narrow the access waveguides may improve system cross-talk, it may adversely impact system throughput and image quality at the exit plane. To understand the impact of this trade, consider the toy model of modal decomposition in the MMI cavity discussed previously. A narrower input waveguide has a correspondingly smaller fundamental mode, and therefore requires more higher-order modes in the MMI cavity in order to be fully-resolved by the modal decomposition. A narrower input waveguide, then, requires a wider MMI cavity to be fully-resolved, which implies loss of image quality at the exit plane due to beat-length non-uniformity. Furthermore, the mode-index mismatch between the access waveguide and the MMI cavity is greater for a narrower access wavguide—which may result in greater losses due to reflection.

Finally, once the width, waveguide separation, and access waveguide width have been selected, the length of the MMI cavity should be optimized for overall throughput. The beat-length formula in Equation 3.40 may be used as a starting point for the MMI cavity length, as $L_{MMI} = 3/2L_{\pi}$, however the length should be further optimized to account for unavoidable phase-errors between modes. MMI length may be optimized by simulating device performance; in the case of the methane spectrometer, the MMIs were evaluated using EME, resulting in a cavity length of 199.3 μm.

Table 3.1 MMI design tradespace

Variable	**Increase**	**Decrease**
MMI cavity width	+ Resolution of input mode - Interference quality - Longer device	+ Interference quality + Shorter device - Greater crosstalk
Waveguide separation	+ Lower crosstalk + Wider access waveguide - Wider MMI	+ Narrower MMI - Greater crosstalk - Narrower access waveguide
Access waveguide width	+ Resolution of input mode + Lower reflection	+ Shorter taper - Greater reflection

The MMI design trade-space is summarized in Table 3.1, from this table it may be seen that the optimal MMI design should possess wide access waveguides, greatly separated, that feed into a narrow MMI cavity. Plainly, these characteristics are in conflict—hence the iterative approach to MMI design.

3.2.3 Waveguide Mach-Zehnder Interferometers

The Mach-Zehnder interferometer is the equivalent of a Michelson interferometer in a fixed position. Light enters the Mach-Zehnder interferometer (shown in Figure 3.9) at an input port, and is divided into two optical paths: the reference arm, and the delay arm.

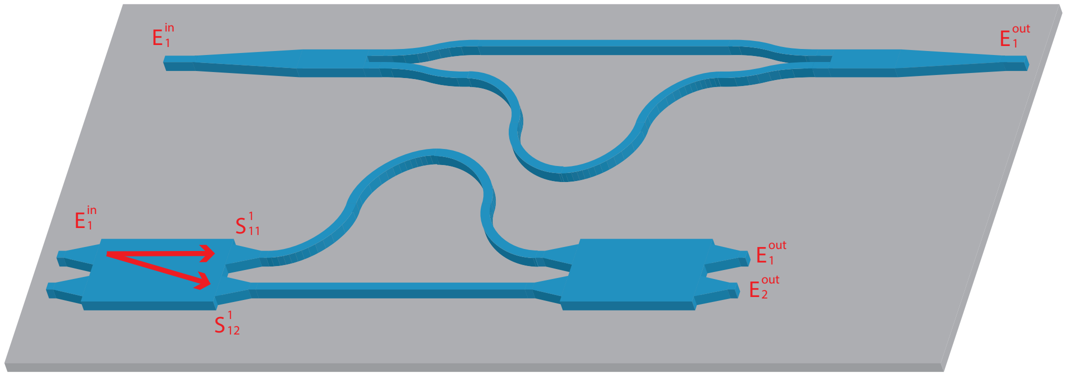

Fig. 3.9 (Top) a diagram of a waveguide MZI is shown consisting of a Y-splitter, delay arm, reference arm, and combiner, alongside a two-port waveguide MZI which uses 2x2 MMIs in order to divide light evenly between the delay and reference arms (bottom).

The physical length of the reference arm is denoted L_R, and the physical length of the delay arm is denoted L_D. Assuming that the incident field is evenly divided by the input beam splitter, then the optical fields (modes) propagating through the two paths may be described as:

$$E^R = \frac{1}{\sqrt{2}} E_0 e^{-i(\beta L_R - \omega t)} \tag{3.41}$$

$$E^D = \frac{1}{\sqrt{2}} E_0 e^{-i(\beta L_D - \omega t)}. \tag{3.42}$$

$$\tag{3.43}$$

At the recombination point—and ignoring the time-dependent term—these optical fields are superimposed, resulting in interference between the reference arm and the delay arm:

$$E^{out} = \frac{1}{\sqrt{2}} E_0 e^{-i(\beta L_R)} + \frac{1}{\sqrt{2}} E_0 e^{-i(\beta L_D)} \tag{3.44}$$

Following the derivation in Section 2.1, the intensity of this output signal—determined by the complex conjugate of Equation 3.44—is a cosine with modulation frequency determined by the propagation constant and the difference in path lengths ΔL

$$I^{out} = I_0(1 + \cos(\beta \Delta L)). \tag{3.45}$$

Equation 3.45 is exact in the case where the input power is evenly divided between the two arms, with no additional phase-shift applied during the splitting and recombination of the fields. In reality, any fabricated splitter (Y-splitter, directional coupler, MMI, etc.) will imperfectly divide the input power, resulting in a power imbalance between the reference arm and the delay arm. Moreover, devices such as 2x2 MMI splitters and directional couplers have non-equal path lengths within the splitter, leading to a static phase offset even in the absence of a delay arm. In the case of an MMI splitter in Figure 3.9, for instance, light exiting the lower port is 90° out-of-phase with respect to light exiting the upper port. In an arbitrary splitting device, the phase delay ($\Delta\phi_{xy}$) and transmission coefficient (S_{xy}) relating input port x to output port y may be captured in the form of a scattering matrix

$$\begin{bmatrix} S_{11}e^{-i\Delta\phi_{11}} & S_{12}e^{-i\Delta\phi_{12}} \\ S_{21}e^{-i\Delta\phi_{21}} & S_{22}e^{-i\Delta\phi_{22}} \end{bmatrix}. \tag{3.46}$$

Since the MZI itself is a multi-port device, the optical properties of the MZI may also be expressed in matrix form, leading to a joint transfer function describing the operation of both the MMI and MZI:

$$\begin{bmatrix} E_1^{out} \\ E_2^{out} \end{bmatrix} = \begin{bmatrix} S_{11}^2 e^{-i\Delta\phi_{11}} & S_{12}^2 e^{-i\Delta\phi_{12}} \\ S_{21}^2 e^{-i\Delta\phi_{21}} & S_{22}^2 e^{-i\Delta\phi_{22}} \end{bmatrix} \begin{bmatrix} S_{11}^D e^{-i\Delta\phi_D} & 0 \\ 0 & S_{22}^R e^{-i\Delta\phi_R} \end{bmatrix} \begin{bmatrix} S_{11}^1 e^{-i\Delta\phi_{11}} & S_{12}^1 e^{-i\Delta\phi_{12}} \\ S_{21}^1 e^{-i\Delta\phi_{21}} & S_{22}^1 e^{-i\Delta\phi_{22}} \end{bmatrix} \begin{bmatrix} E_1^{in} \\ E_2^{in} \end{bmatrix} \tag{3.47}$$

One key advantage of using an MMI splitter over a Y-splitter, is that the phase difference imparted by the MMI may be used to generate a normalization signal for the MZI. Considering the case of the above MZI illuminated from a single port (*i.e.*, $E_2^{in} = 0$), and using perfectly-balanced 50-50 MMI splitters with a 90° phase shift between exit ports, then the MMI transfer function may be written as:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \tag{3.48}$$

Additionally, the phase terms in the MZI scattering matrix may be combined to form a single phase shift term, $e^{-i\Delta\phi}$. Finally, the output fields E_1^{out}, and E_2^{out} can then be written as:

$$\begin{bmatrix} E_1^{out} \\ E_2^{out} \end{bmatrix} = \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} e^{-i\Delta\phi} & 0 \\ 0 & 1 \end{bmatrix} \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix} \begin{bmatrix} E_1^{in} \\ 0 \end{bmatrix} \tag{3.49}$$

$$\begin{bmatrix} E_1^{out} \\ E_2^{out} \end{bmatrix} = \tfrac{1}{2} \begin{bmatrix} E_1^{in}(e^{-i\Delta\phi} - 1) \\ iE_1^{in}(e^{-i\Delta\phi} + 1) \end{bmatrix}. \tag{3.50}$$

The output intensities $I_1^{out} = E_1^{out} E_1^{out*}$ are then:

$$\begin{aligned} \begin{bmatrix} I_1^{out} \\ I_2^{out} \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} |E_1^{in}|^2(e^{-i\Delta\phi} - 1)(e^{i\Delta\phi} - 1) \\ i(-i)|E_1^{in}|^2(e^{-i\Delta\phi} + 1)(e^{i\Delta\phi} + 1) \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} |E_1^{in}|^2(1 - cos(\Delta\phi)) \\ |E_1^{in}|^2(1 + cos(\Delta\phi)) \end{bmatrix} |E_1^{in}|^2 \end{aligned} \tag{3.51}$$

Leading to the normalization properties:

$$I_1^{out} + I_2^{out} = |E_1^{in}|^2 \tag{3.52}$$

$$I_2^{out} - I_1^{out} = |E_1^{in}|^2 \cos(\Delta\phi). \tag{3.53}$$

Equation 3.53 may be used to evaluate the intensity of light that is inserted into each MZI. In the case where an array of MZIs with independent input apertures, is used, Equation 3.53 may be used to pre-process the resultant interferogram in order to correct for variations in waveguide insertion loss, and non-uniform illumination across the input aperture [117].

Since integrated optical Fourier transform spectrometers formed from arrays of MZIs may be prone to such non-uniformities (e.g. defects introduced during fabrication), the use of 2x2 MMI splitters in the construction of the MZIs is key to deployment of these devices in practical applications outside of the laboratory. Recent work on on-chip Fourier-transform spectrometers has further extended this concept using 3x3 MMIs and 4x4 MMIs ("90° optical hybrids") in order to capture additional phase information from each interferogram sample.

Considering a generalized sampling point in the interferogram, non-uniformities in the fabrication process may result in both phase, ϵ_i, and amplitude, A_i, errors

$$I_i = A_i|E^{in}|^2 \cos\left(\beta \Delta L[1 + \epsilon_i]\right). \tag{3.54}$$

If not corrected, the above phase and amplitude errors may result in a non-uniform interferogram, which must be processed using the appropriate non-uniform Fourier-transform techniques (see Section 2.1). In 2018 Uda et al [116] demonstrated that the above amplitude and phase errors may be determined as the unique solution to a system of equations relating the three output intensities of an MZI using 3x3 MMIs. In this case, provided that the path length of each MZI is known exactly, then a phase and amplitude corrected interferogram may be produced, and the spectrum may be recovered using the conventional Fourier-transform. This approach yields an elegant solution to the problem of non-uniformities introduced during fabrication. Future developments or iterations of on-chip Fourier-transform spectrometers would benefit from adopting this approach, or similar approaches using 4x4 MMIs.

3.2.4 Input and output coupling

The inputs to each MZI are bundled together and collected at an input aperture—which will connect the spectrometer to its fore-optics system. The physical input aperture of a waveguide system may be formed in many ways, and the selection of a particular input coupling scheme is a critical component of waveguide design. Coupling to waveguides may be accomplished by insertion into the evanescent wave (*e.g.*, similar to the cross-talk phenomenon), direct excitation of the fundamental mode by focusing a planar beam onto the edge of the chip, or diffractive coupling into the plane of the waveguide using gratings [120].

In any of these schemes, achieving high-efficiency input coupling relies upon matching the intensity distribution of the incoming field to the field intensity of the fundamental waveguide mode, as well as minimizing reflections. In an ideal coupling scenario, the input field would be a mirror image of the waveguide mode in size and intensity, and would be perfectly spatially overlapped with the waveguide mode. In practice, when coupling from free-space, a perfect overlap is impossible to achieve due to the mis-matched appearances of the mode-field in a waveguide, and the PSF of a diffraction-limited telescope. One-dimensional slices of a waveguide mode and a diffraction-limited PSF are shown in Figure 3.10, where it can be seen that—although

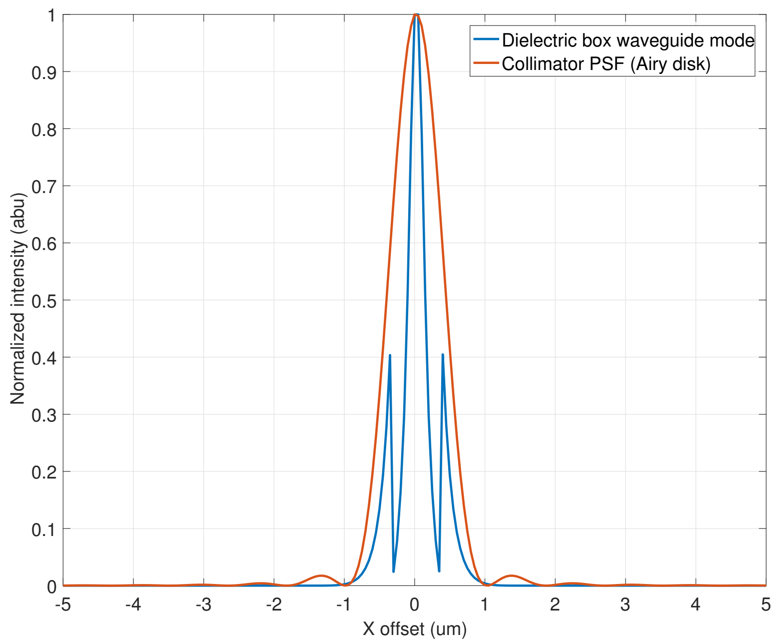

Fig. 3.10 1D slice of a fundamental waveguide mode for a 450 nm wide Si-wire waveguide overlaid with the Airy disk function for a plane-wave focused at the waveguide facet. The focal length corresponding to the Airy disk is optimized to produce maximal overlap with the fundamental waveguide mode.

there is good overlap in the core of the fiber—the evanescent field is not well-matched to the shape of the Airy disk.

A portion of the incoming light, the part that is mismatched with the fiber field, radiates into the cladding material and is "lost" in the input coupling process. Another portion of the incoming light may be lost to reflection if there is a significant mismatch between the index of the incident medium and the effective index of the waveguide mode. Losses due to reflections in input coupling arise at the chip-to-free-space interface. The portion of light that is transmitted through the interface, T_F, may be approximated via the Fresnel equation for reflection at normal incidence

$$T_F = 1 - R = 1 - \left| \frac{n_1 - n_2}{n_1 + n_2} \right|^2 . \tag{3.55}$$

Where n_1 is the refractive index of the incident medium, and n_t the refractive index of the transmitted medium. For complex interfaces, the Fresnel equation may not be fully sufficient to describe the reflectance—particularly in the case of exotic waveguide structures such as subwavelength gratings [121]—however they are sufficient for the purpose of estimating input losses due to reflection at the waveguide facet. For a discussion on the finer points of reflection at irregular interfaces, a thorough treatment may be found in *Subwavelength Gratings For Space Applications* [122].

The expected coupling loss at a waveguide interface, neglecting scattering losses, may be calculated as the product of the Fresnel transmission, T_F, and the overlap transmission T_O

$$T_{input} = T_F T_O, \tag{3.56}$$

with the overlap transmission given by the overlap integral between the excitation field $E_1(x, y)$ and the waveguide mode $E_2(x, y)$.

$$T_O = \frac{\int\int (E_1(x,y)E_2(x,y))^2 dxdy}{\int\int ({E_1}^2(x,y){E_2}^2(x,y))dxdy}. \tag{3.57}$$

3.2.4.1 Butt coupling

The most straightforward method of coupling light to a waveguide is direct illumination of the fact edge. Analogous to coupling light to a fiber through the use of a collimating lens, coupling light in this manner relies on the focusing of a planar wavefront into a small spot—of comparable size to the mode-field diameter of the waveguide—directly onto the edge of the waveguide. Focusing of light onto the waveguide aperture may be accomplished in a variety of ways including free-space lenses, or lensed and ball-tipped fibers. In each case, alignment of the focused beam to the facet of the waveguide is critical to the coupling efficiency; misalignment between the modal field and focused spot will reduce the overlap between the two fields, limiting excitation of the waveguide mode. Despite its sensitivity to misalignment, direct illumination may be attractive for its ease of implementation, especially in research environments.

In addition to matching the size of the focused spot with the MFD, and minimizing horizontal and vertical offsets, it is important to ensure that the incoming focused wavefront underfills the NA of the waveguide. Any light entering the waveguide at an angle greater than that supported by the NA will pass through the waveguide into the substrate—thereby failing to couple to the fundamental mode. To conserve étendue, the NA of a waveguide increases as the mode-field diameter decreases; thus, the selection of a well-confined waveguide core (*e.g.*, to achieve a small bend radius) implies a large acceptance angle at the input facet and more stringent alignment between focused spot and waveguide facet. Conversely, by increasing the MFD the alignment tolerance at the input may be improved, at the cost of a reduced acceptance angle.

3.2.4.2 Spot-size converters

Increasing the MFD may be accomplished through the use of spot-size converters (SSCs), devices in which the geometry or index of the waveguide is gradually tapered to a fine point in order to produce a significantly enlarged MFD at the waveguide output. As the geometry of the waveguide, or the waveguide refractive index is gradually decreased, the waveguide mode becomes less and less well-confined to the core, and expands into the cladding. To avoid leakage of the mode into the surrounding cladding or adjacent waveguides, SSCs are typically straight waveguides without bends, and are placed further apart on the wafer than standard waveguides.

Expansion of the mode via SSCs is highly advantageous in terms of coupling to waveguides from optical fibers, wherein the mode fields may be substantially mismatched. For instance, an Si-wire waveguide may have a MFD of $\approx 0.5\ \mu m$ at

$\lambda = 1550\ nm$, whereas the MFD of standard single-mode fiber (SMF) is 10.4 μm. By tapering the waveguide geometry, to expand the MFD of the waveguide, the overlap between the two may be substantially improved—leading to greater overall coupling efficiency [123]. A secondary advantage of expanded mode-size is that spatial misalignments between the input field and the waveguide mode are more tolerable—allowing for ease of manufacture during high-volume production involving high-density v-groove arrays. One caveat of expanding the MFD, however, is that the NA of the waveguide will correspondingly decrease to preserve the étendue of the waveguide.

3.2.4.3 Grating couplers

A more recent form of input coupling involves redirecting incoming light into the plane of the waveguide by means of a diffraction grating [120]. Light of wavelength λ_0 incident at angle θ_{inc} upon a diffraction grating of period Λ, and refractive index n_1 is deflected by angle θ_{out} according to the grating equation

$$\sin(\theta_{out}) = (\sin(\theta_{inc}) - m\lambda_0/\Lambda), \tag{3.58}$$

where m refers to a particular diffracted order. This effect may be leveraged in waveguide design, to produce a diffraction grating that deflects light directly into a waveguide aperture. In this case, input light is directed downwards towards the surface of the wafer—rather than the edge—either by a telescope, or by a fiber fixed above the surface of the wafer. Grating couplers (and vertical coupling schemes in general) are an attractive input coupling scheme as they allow the formation of high-density arrangements of input apertures in a 2-D plane, enabling a greater number of interconnects on a single optical chip. Furthermore, the ability to insert and receive light out-of-plane permits large volume chip manufacturers to perform wafer-scale automated chip testing, without the need for individual chip dicing and characterization. Unfortunately, grating couplers exhibit substantially higher losses than SSCs, primarily due to inefficient diffraction into the first diffracted order. Unlike in an SSC, where substantially all of the light may be captured, much of the light incident upon the grating will pass through the grating entirely, having failed to diffract into the first order. These losses may be mitigated by using reflections from the lower cladding to result in a two-pass diffraction scheme [120]. Unlike SSCs, grating couplers are also highly wavelength selective which may make them undesirable as a solution for high-bandwidth applications.

In the methane spectrometer, described in Chapter 4, straightforward butt-coupling from free-space was selected as the input coupling scheme. Gratings were deemed unsuitable due to their expected stray-light generation, as well as the difficulty in ensuring uniform operation in a custom design with uneven polymer overcladding.

Similarly, SSCs were dismissed as the waveguide manufacturer was not able to ensure close physical separation (*i.e.*, “bundling”) of SSCs at the input facet of the spectrometer—which would severely degrade the fill-factor of the spectroemter input aperture. The resulting waveguide bundle consists of 100 waveguide apertures at the edge facet of the chip, with each facet separated by 5 μm.

In the CS spectrometer, described in Chapter 6, a hybrid vertical coupling scheme was developed based on butt-coupled facets matched to mirror surfaces etched at 45° to the wafer surface. This solution achieves the broadband input coupling capabilities of direct facet illumination techniques, while also enabling the creation of a high-density 2-D input aperture. The resulting waveguide bundle consists of 83 waveguide apertures, arranged on an 8×10 grid, with 500 μm periodicity.

3.2.5 Athermality

A recurring concern with waveguide devices, and particularly devices reliant on specific optical path lengths (such as MZIs) is their sensitivity to temperature. As with any refractive medium, the indices of refraction of the waveguide core and cladding are temperature dependent; consequently, the index of the fundamental waveguide mode is also temperature dependent. In high-confinement Si-wire waveguides, for instance, the mode index will scale roughly according to the refractive index of the silicon core: $n_{Si}(T) = n_{Si}(273.15K)(1 + dn_{Si}/dT\Delta T)$, where $dn_{Si}/dT =$

$1.9 \times 10^{-4} K^{-1}$ is the thermo-optic coefficient (TOC) of silicon. In the case of an extended MZI, with an OPD on the order of centimetres, the TOC may have a significant effect on the intensity output of the MZI.

To eliminate this effect, a few pathways may be pursued.

1. *Variable waveguide core geometries*

   In the first case, the thermal response of an MZI may be attenuated by using distinct waveguide core geometries in the delay arm and the reference arm [124]. Assuming the use of a cladding with a lower TOC than the core material (as is the case for Si-wire waveguides), the geometry of the delay arm core may be designed such that a greater portion of the mode propagates in the cladding (*i.e.*, a low-confinement core), thereby reducing the effective TOC of the waveguide core. At the same time, the reference arm may be designed to concentrate power in the waveguide core, thus raising the TOC of the reference arm. By balancing the relative TOCs of the delay and reference arms, as well as the relative path lengths of each arm, a passively athermal MZI may be realized. Unfortunately, the use of a low-confinement core in the delay arm increases the footprint of each MZI (as the bend-radius increases), and the the physical length of the MZI increases to accommodate the lower waveguide index. In addition, the use of multiple core geometries may introduce additional losses in the MZI due to reflections and scattering at the transition points.

2. *Active phase control*

   In the second case, thermal compensation may be achieved using active phase-shifting devices, such as thermal heaters or electro-optical shifters on-chip [124]. In this case, the OPD of each MZI may be adjusted independently to offset any errors introduced by thermal gradients within the chip. The advantage of active phase control is that, in addition with thermal compensation, phase shifters enable simultaneous compensation for OPD non-uniformity. To correctly implement phase correction, all MZIs should be "locked" to a single reference frequency (*e.g.*, a laser diode) in order to ensure that the corrected OPD of each MZI is an integer multiple of the first. Independent phase control may be difficult to implement, however, requiring additional chip footprint allocations for each MZI in the array (to accommodate the heating pads) as well as requiring additional chip footprint at the edges of the device. Active phase control may also introduce packaging concerns regarding integration of the chip to the detector, as each active phase shifter will require a wire-bonded connection to the driving electronics.

3. *Passive waveguide athermalization*

   Finally, the thermal drift of an MZI may be attenuated by directly reducing the TOC of the waveguides to produce a "passively athermal" waveguide core.

Athermal waveguides may be designed by taking advantage of the fraction of modal power propagating in the upper and lower cladding media [124]. If a material with a negative TOC (*e.g.*, SU-8, TiO_2) is used to form the cladding, then the fraction of modal power propagating in the cladding will experience a negative TOC, offsetting the positive TOC of the waveguide core and thereby reducing the overall TOC of the waveguide. By adjusting the fraction of power propagating in the waveguide core versus the cladding, the modal TOC may be biased towards the cladding material (in a low-confinement geometry), or towards the core material (in a high-confinement geometry). Passively athermal waveguides designed in this manner may achieve substantially lower TOCs than conventional Si-wire waveguides, without sacrificing chip footprint or greatly complicating device integration and packaging. Despite these advantages, passively athermal waveguide design may reduce wafer yield through variations in core-confinement due to manufacturing tolerances on the core geometry; for this reason, a waveguide geometry that is robust to fabrication defects must be developed. Passively athermal waveguide designs may also drive further compromises in the device, as the waveguide geometry is fixed by the thermal requirement and cannot be further optimized for reduced propagation loss, small bending radii, or low crosstalk.

In the methane spectrometer presented in Chapter 4 a passively athermal waveguide geometry is developed. Design robustness is ensured via Monte-Carlo analysis of the expected waveguide manufacturing tolerances, resulting in a waveguide with design TOC bounded to $\alpha_{wg} < 5 \times 10^{-6} K^{-1}$; further details on the selection of the waveguide geometry are presented in Chapter 4. In the CS spectrometer presented in Chapter 6 the requirements on thermal stability are substantially lower than the methane device as (1) the longest OPD is two orders of magnitude shorter than the methane spectrometer, and (2) the TOC of the waveguide is already effectively reduced to the TOC of deposited SiO_2, $\alpha_{SiO_2} = 9.6 \times 10^{-6}$ by its extremely low-confinement core geometry.

4 Athermal planar-waveguide Fourier-transform spectrometer for methane detection

This chapter is based on a paper (Podmore *et al.* [125]) which describes a planar-waveguide Fourier-transform spectrometer developed for the detection of atmospheric methane. During the course of this body of work I designed a custom, temperature-insensitive waveguide geometry, all waveguide components, and wrote the adaptive layout scripts to generate the highly-compact FTIR waveguide spectrometer. The waveguide device was fabricated externally, with some post-processing (SU-8 deposition, wafer dicing) performed by myself with the assistance of colleagues from Honeywell and the University of Waterloo under my supervision. The device was characterized using a measurement setup and data-processing pipelines that I designed in the engineering laboratory at Honeywell Aerospace. Data collection using this setup was performed in tandem with my co-authors, all data analysis was performed by myself. Ultimately in this work we demonstrate a reduction in thermal sensitivity by two orders of magnitude over competing technologies, and demonstrate the first ever retrieval of atmospheric absorption features using a Fourier-transform waveguide spectrometer. I performed all of the calculations, prepared all of the figures, and wrote the manuscript with revisions provided by my co-authors.

4.1 Introduction

Methane is a greenhouse gas that contributes significantly to global climate change through its absorption of outgoing IR radiation and interaction with atmospheric aerosols [126]. Methane is increasingly recognized as a significant but poorly-monitored contributor to global climate change [127]. Consequently, there exists a recognized need to develop new methods to accurately evaluate the global methane distribution. Many high-density sources of methane emission (fracking sites, pipelines, mines, agribusiness, etc.) have been identified, however monitoring these sites remains a technical challenge due to their extensive geographic distribution [127, 128].

One promising technique for inexpensively monitoring sparsely distributed sources of methane is remote detection of atmospheric methane via observation of molecular absorption lines using high-resolution IR spectrometers [129]. Such spectrometers may be housed on highly mobile platforms such as unmanned aerial vehicles, and micro and nanosatellites in order to provide global coverage of methane emission sources. A key technological challenge to utilizing these platforms lies in the development of small, low-mass spectrometers with sufficient spectral resolution to distinguish molecular absorption lines separated by tens of picometers [31].

High-resolution spectrometers usually fall into one of two categories: dispersive element spectrometers, and FTS. Of these two, FTS devices, in which the input light undergoes sinusoidal modulation via self-interference, are preferred for their throughput advantage in high-resolution applications (the Jacquinot advantage [76]). FTS instruments designed as SHFTS are well suited to remote-sensing from aircraft as they may be implemented as a monolithic glass block with no moving parts [102]. In an SHFTS, light is spectrally dispersed and recombined to produce a spatially-modulated interference pattern that may be read out using an array of detectors. SHFTS devices achieve high spectral resolution at the expense of bandwidth, which limits their utility for simultaneous detection of multiple atmospheric species. Ultimately, the integration of SHFTS devices onto small mobile platforms may be limited by the size and mass of the monolithic interferometer.

One promising avenue for miniaturization of high-resolution spectrometers lies in planar waveguide photonic chip technology [130]. Waveguides can be fabricated on silicon substrates in high-density arrangements, enabling the integration of complex optical circuits, including SHFTS, into centimeter-sized chips, or smaller [37].

The first on-chip SHFTS was proposed in Michelson configuration [131] and was subsequently implemented as an array of MZIs with linearly increasing OPDs [117, 132–134]. These MZIs constitute a discrete Fourier-transform in which the output of each MZI corresponds to a point in the spatial interferogram. These points may be sampled independently via photodiodes placed on the chip or by an external linear detector array. The interferogram may then be inverted using the FFT to retrieve the input spectrum. The simultaneous spatial sampling scheme in an SHFTS allows for instantaneous acquisition of an interferogram in a single shot, unlike a scanning FTS where the interferogram is collected over time. Single-shot acquisition enables signal enhancement through extended exposure, as well as observation at high readout rates—suitable, for example, for fast-moving satellites. The small size of these chips may also enable the detection of multiple atmospheric species from a single integrated platform using vertically stacked on-chip spectrometers, each designed for a separate atmospheric species.

One challenge yet to be addressed in the implementation of planar waveguide SHFTS technology lies in compensation for rapid thermal drifts, which result in non-uniform sampling of the interferogram. Spectral inversion of an interferogram via the FFT requires the sampling locations to be evenly separated, *i.e.*, each OPD must be an exact integer multiple of the first [57]. In regular operation, however, the OPDs will vary due to imperfections introduced during the fabrication process as well as fluctuations in chip temperature. As with many dispersive materials, planar waveguides experience temperature-dependent shifts in their refractive index which may be significant over long optical distances. Therefore, thermal compensation in

SHFTS devices is of critical importance [135].

In this chapter we present a miniaturized high-resolution planar waveguide SHFTS for methane detection with many innovative aspects. We use passive thermal stabilization of the device—achieved through waveguide refractive index engineering—to stabilize the interferogram and reduce temperature sensitivity. We use a multi-aperture input array to maximize optical throughput, and sample MZI outputs in quadrature to compensate for irregular input coupling. Finally, we use the waveguide SHFTS and a reference gas cell to demonstrate on-chip SHFTS detection of atmospheric absorption features for the first time.

4.2 Theory and simulation

Our SHFTS device is designed to target the Q-branch absorption features of methane, consisting of strong, closely-spaced molecular lines covering a spectral range of 2.2 nm centered at $\lambda_0 = 1666.3$ nm. The Q-branch features are selected for this experiment as they do not overlap with other atmospheric species; however, in principle, any atmospheric species may be selected. The planar waveguide spectrometer is formed by 100 MZIs arrayed in a high-density configuration on a single chip. Our device targets a spectral resolution of $\delta\lambda = 0.050$ nm and bandwidth of $\Delta\lambda = 2.5$ nm in order to resolve multiple absorption features within the Q-branch. The OPDs of the longest and shortest MZIs set the resolution ($\delta\sigma$) and bandwidth ($\Delta\sigma$) in wavenum-

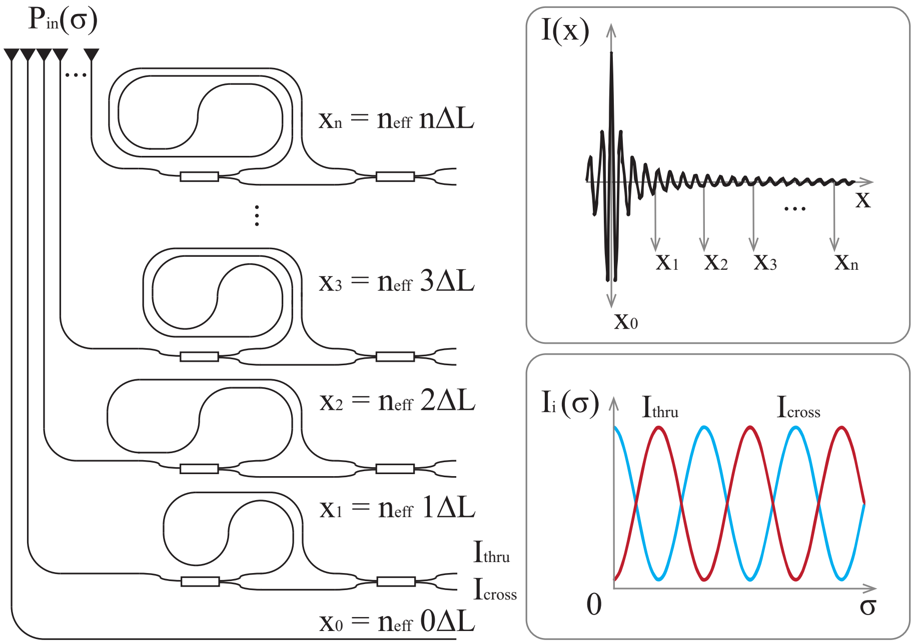

Fig. 4.1 Diagram showing the operating principle for an on-chip SHFTS. MZIs are implemented as spirals, enabling high-density packing. Each MZI has an independent aperture and splits its outputs into two ports which are 180° out-of-phase as shown in the lower inset. The continuous interferogram, $I(x)$, is sampled at locations defined by the OPDs of each MZI, as shown in the upper inset.

bers as $\delta\sigma = 1/(n_{eff}\Delta L_{max})$, and $\Delta\sigma = 1/(2n_{eff}\Delta L)$. In our device $\Delta L = 0.0326$ cm is the physical path difference of the least unbalanced MZI, n_{eff} is the effective index of the waveguide, and $\Delta L_{max} = 3.26$ cm is the maximum physical path difference. The OPD of the i-th MZI is then given by $x_i = n_{eff} i\Delta L$, as shown in Fig. 4.1.

The SHFTS device is implemented on a silicon wafer with an 8 μm thick silicon dioxide (SiO_2) layer forming the lower cladding. The waveguide layer consists of two 300 nm thick layers of silicon nitride (Si_3N_4) separated by a 90 nm layer of strain-relieving SiO_2. The upper cladding of the waveguide structure is formed by a 6 μm layer of SU-8 polymer. This waveguide structure is designed to produce an effective index $n_{eff} = 1.7$ at $\lambda_0 = 1666.3$ nm for the fundamental TE mode.

The first innovative aspect of this device is the use of a multi-aperture input scheme coupled with MZI signal normalization. This concept was proposed in [117]; in this chapter we demonstrate it experimentally for the first time. Each MZI in our device possesses an independent aperture (input coupler), and splits its interferometric signal between two waveguide outputs to provide signal normalization. Independent input coupling—rather than splitting a single input signal multiple times—increases the optical throughput of each MZI, improving the SNR at the pixel array [134]. In a multi-aperture input scheme, non-uniform illumination of the aperture and input-to-input variations in coupling efficiency may distort the interferogram. Therefore, the interferometric signal of each MZI must be normalized against its input intensity [117].

In our device, splitting and recombining in each MZI is accomplished through the use of 2×2 MMIs. MMIs are preferred over Y-splitters as MMIs introduce a 90° relative phase shift between the cross-port and through-port signals; this phase shift may be used as the basis for signal normalization. Light entering the MZI is split at the first MMI where the reference arm (cross-port) receives a +90° phase delay relative to the delay arm (through-port). When the two arms are joined at the through-port of the second MMI, the delay arm and reference arm will be 180° out-of-phase in addition to any OPD-induced phase shift. By contrast, at the cross-port of the second MMI, the two arms will each have received a 90° phase shift from one of the two MMIs and will combine in-phase. This process is shown diagrammatically in Fig. 4.1. The resulting pair of outputs for the i-th MZI, and the associated interferogram, may then be described using the following modulation functions:

$$P_{thru}(\sigma, i) = \frac{P_0(\sigma)}{2}(1 + cos(2\pi n_{eff} i \Delta L \sigma)) \tag{4.1}$$

$$P_{cross}(\sigma, i) = \frac{P_0(\sigma)}{2}(1 - cos(2\pi n_{eff} i \Delta L \sigma)) \tag{4.2}$$

$$I(x_i) = \int P_0(\sigma) cos(2\pi n_{eff} i \Delta L \sigma) d\sigma, \tag{4.3}$$

where $P_0(\sigma)$ is the optical input power as a function of wavenumber. The through-port and cross-port signals may be added to retrieve the normalization factor $P_0 = P_{thru} + P_{cross}$, or subtracted to produce a modulation function, $P_{thru} - P_{cross} = P_0 cos(2\pi n_{eff} i \cdot \Delta L \sigma)$. The normalization factor may then be used to correct for irregular illumination across the input aperture. Such irregularities may be caused by misalignment of the frontend optics, or chip fabrication imperfections including damage to the facet.

The second innovation achieved in this device is to reduce the thermal sensitivity of the SHFTS. Spectral retrieval from a planar waveguide SHFTS relies on accurate knowledge of the OPD for each MZI, however, the OPDs of each MZI may fluctuate with temperature. Thermal drifts may be compensated provided that the temperature of the spectrometer is known to a sufficiently precise degree [135]. In this case, the required level of precision of the temperature measurement will be set by the thermal sensitivity of the chip. The response of each MZI to an uncompensated (*i.e.*, unknown) temperature change of magnitude ΔT may be derived from the change in the phase delay, $\phi(\Delta T) = 2\pi\sigma(n_{eff}\Delta L + d/dT(n_{eff}\Delta L)\Delta T)$. We collect the temperature dependent terms into an effective TOC, $\alpha_{eff} = 1/n_{eff}dn_{eff}/dT + d/dT(\Delta L)$ and write the MZI response as:

$$I_i(\sigma, \Delta T) \approx P_0(\sigma)cos(2\pi\sigma i\Delta L n_{eff}(1 + \alpha_{eff}\Delta T)), \tag{4.4}$$

where α_{eff} is an equivalent TOC of the waveguide combining the coefficient of thermal expansion of the substrate, and refractive index changes in the cladding and waveguide material. Equation (4.4) can be used to evaluate the impact of a particular TOC on the interferometric output, and determine the temperature knowledge required to limit variability in the interferogram to less than 5%. Were the spectrometer to be realized using standard silicon wire waveguides ($n_{eff} = 2.5$, $\Delta L_{max} = 2.22$ cm), the TOC of silicon ($\alpha_{Si} = 1.8 \times 10^{-4} K^{-1}$) at the central wavenumber $\sigma_0 = 6001 cm^{-1}$

would require that the temperature be monitored to an accuracy of 0.1 mK. In a dynamic environment it is unlikely that such a measurement could be practically achieved. However, if the waveguide TOC is substantially reduced, then temperature measurements may be obtained with less difficulty.

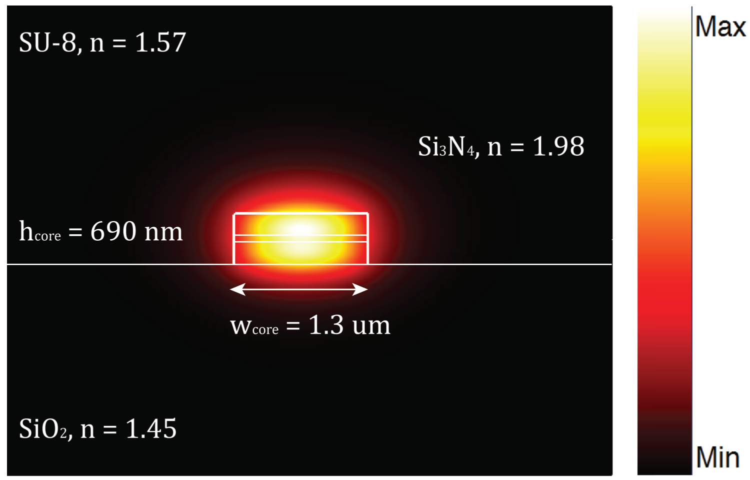

Fig. 4.2 Waveguide cross-section showing modal power distribution and waveguide dimensions. The waveguides are formed on an 8 μm thick layer of SiO_2, and the waveguide core consists of two 300 nm thick layers of PECVD Si_3N_4 separated by a 90 nm thick SiO_2 layer. The upper cladding is formed by a 7 μm thick layer of SU-8 polymer.

The waveguide geometry, shown in Fig. 4.2, is designed to passively reduce the equivalent TOC, thereby attenuating the thermal sensitivity (Eq. (4.4)). The TOC of the effective mode index is engineered by controlling the fraction of modal power that propagates within the SU-8 upper cladding—which has a negative refractive index coefficient—versus the Si_3N_4 core and SiO_2 lower cladding [136]. We calculated the equivalent TOC of the mode effective index for the fundamental TE mode by weighted sum of the material coefficients, $\alpha_{eff} = 1/n_{eff} \sum \Gamma_i c_i + d\Delta L/dT$, as described in Table 4.1. In this sum the optical terms for each material are weighted by their modal power fraction (*i.e.*, the percentage of modal power propagating in the respective medium), and the thermal expansion of the substrate is weighted by the effective index of the waveguide.

Table 4.1 Thermal coefficients, c_i, with associated weighting factors, Γ_i, and contributions to equivalent TOC, α_i.

Material	Coefficient (K^{-1})	Weight (Γ_i)	α_i (10^{-6})
SU-8	-1.2×10^{-4}	0.21	-25.2
Si_3N_4	2.45×10^{-5}	0.72	17.6
SiO_2	0.95×10^{-5}	0.07	0.7
$d\Delta L/dT$	4.4×10^{-6}	–	4.4
Total			-2.5

For our chosen waveguide materials and core geometry we calculate an expected TOC of $\alpha_{eff} = -2.5 \times 10^{-6} K^{-1}$, corresponding to a temperature knowledge requirement of $\Delta T = 12$ mK. The magnitude of the TOC of SU-8 relative to those of the waveguide core and lower cladding necessitates a high degree of waveguide core confinement to produce an athermal waveguide: $\Gamma_{core} = 0.7$. In actuality, the core confinement parameters in Table 4.1 are subject to some uncertainty due to expected variations in the thickness and width of the waveguide core during fabrication.

We quantify the effect of this uncertainty by Monte-Carlo analysis: accounting for expected fabrication tolerances ($\pm 1\%$ on all cross-section dimensions) the expected upper bound of the TOC is found to be —$\alpha_{eff}| < 5 \times 10^{-6} K^{-1}$. This corresponds to a temperature knowledge requirement of $\Delta T = 6$ mK, a significant improvement over prior implementations of the technology in Si-wire waveguides [135].

4.3 Experimental results and discussion

The wafers and waveguides were fabricated by LioniX International, and the SU-8 upper-cladding and wafer dicing were performed as a post-process. The lower cladding was formed by a thick thermal oxide layer grown on a Si substrate. The three core waveguide layers were deposited by plasma-enhanced chemical vapour deposition (PECVD), patterned with an i-line stepper and defined by reactive ion etching (RIE). The wafers were further processed and the SU-8 was deposited via spin-coating to form the upper cladding. Finally, the wafers were diced into individual chips at the Toronto Nanofabrication Center. An image of the processed spectrometer chip next to a Canadian 25-cent coin is shown in Fig. 4.3(c).

We built a benchtop setup to characterize, calibrate, and demonstrate the functionality of the spectrometer chip. Light from a tunable laser source is coupled to a single-mode fiber, and routed to a polarizing beam splitter (PBS). One arm of the PBS monitors the wavelength and power of the laser output; in the second, light is

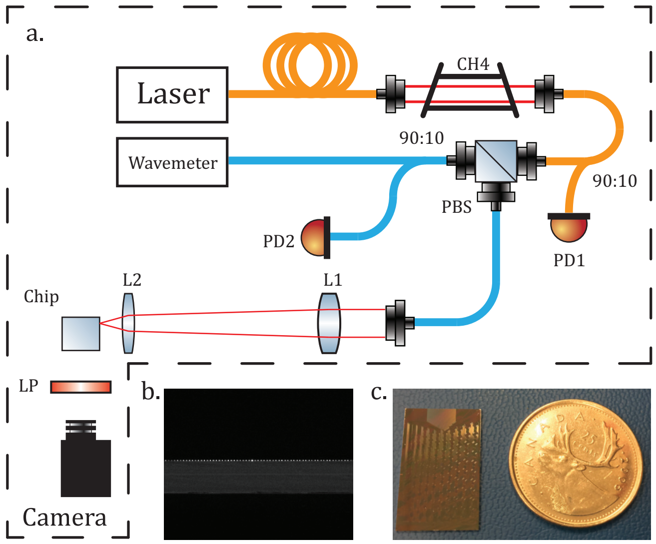

Fig. 4.3 (a) Light from a tunable laser is coupled into a single-mode fiber (yellow), and passes through a 7.5 cm methane cell. The optical power is tapped at 10% and monitored using an InGaAs photodiode (PD1). The single mode fiber is split by a polarizing beam splitter (PBS) into two polarization-maintaining fibers (blue). One arm leaving the PBS is directed to a wavemeter and photodiode (PD2). The second arm is directed to the chip via a fiber collimator and focused in the horizontal and vertical axes by two cylindrical lenses (L1, L2). The output signal is passed through a linear polarizer (LP) and captured by an InGaAs camera. (b) Output signal captured by camera. (c) Photograph of FTS chip next to a 25-cent Canadian coin.

passed to a fiber collimator directed at the input aperture of the chip. The splitting ratio is monitored by a pair of photodiodes located before and after the PBS. Light from the fiber collimator is focused onto the waveguide facets via two cylindrical lenses placed in front of the aperture of the photonic chip. The output signal passes through a linear polarizer where transverse-magnetic (TM)-polarized light is rejected, and the TE-polarized light is imaged by an IR camera. The experimental setup is shown schematically in Fig. 4.3(a). The chip itself is placed on a 6-axis mechanical stage and its temperature is controlled by a Peltier module.

The wavelength of the tunable laser is swept across the operating range of the spectrometer, and the output signals from each MZI are captured as point-source intensities in the IR camera image. The IR camera images are referenced to the wavelength of the laser, and the background-subtracted intensity values are used to assemble the calibration matrix shown in Figure 4.4. When this procedure is performed with the gas cell removed, and with a laser linewidth substantially less than the SHFTS spectral resolution, the product is an $M \times N$ transformation matrix, $\mathbf{A}$, in which M is the number of MZIs, and N is the number of measured wavelength points. In this transformation matrix, also called a spectral (or calibration) map, each row contains the modulation response for a single MZI output as a function of wavelength. The experimental matrix obtained from the methane spectrometer chip is shown in Fig. 4.4, in this matrix the outputs I_1 and I_2 of the MZIs were used to

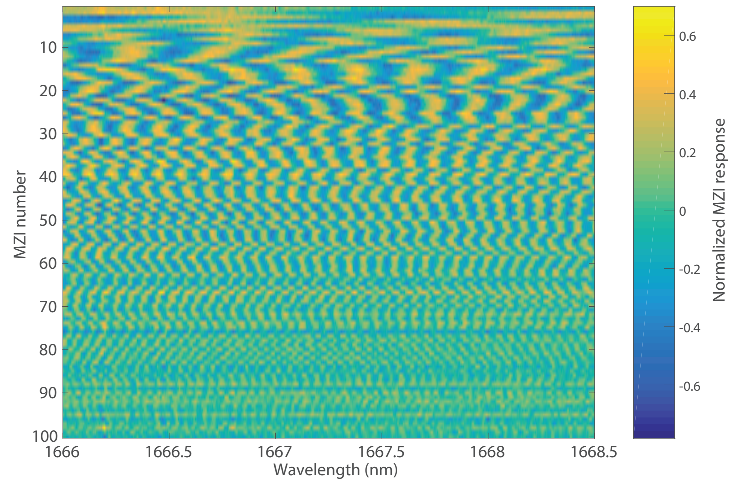

Fig. 4.4 Measured calibration matrix of the SHS chip normalized using the difference of the MZI outputs divided by the sum according to Equations 4.1 and 4.2. The map shows the expected wavelength-dependent response of each of the 100 MZIs, which is a good match to MZI theory. The spectral responses were obtained during a high-resolution scan over a 2.5 nm bandwidth between 1666 nm and 1668.5 nm. MZIs are ordered sequentially according to their OPDs, with MZI 1 having the shortest OPD and MZI 100 having the longest.

normalize the transform matrix as $A(\sigma) = (I_1(\sigma)) - I_2(\sigma))/(I_1(\sigma) + I_2(\sigma))$. It can be seen from this matrix that the fringe contrast, defined as the amplitude of the normalized MZI transfer function in matrix $\mathbf{A}$, decreases with increasing OPD. The decreasing fringe contrast is an expected consequence of the waveguide propagation loss, which may be modelled by adding a real loss term to the MZI delay term in

Equation 3.47. Using the data from the experimentally observed calibration matrix, and fitting the decreasing fringe contrast to the MZI transfer function, the waveguide propagation loss is determined to be 2.1 dB/cm.

The tunable laser is used to record the interferogram corresponding to a specific optical bandpass. With the gas cell in the optical path, an interferogram, $\mathbf{y}$, is obtained by performing laser sweeps across a selected spectral range and time-averaging the waveguide outputs (Fig. 4.3(b)). The spectral range covered by the laser during sampling simulates a flat-topped anti-aliasing bandpass filter. In an operational scenario the bandpass filter would be placed at the telescope input, and a broadband light source (such as the sun) would illuminate the SHFTS. We utilize the tunable laser source to have precise control over the oversampling factor for the interferogram, noting that this may overestimate the coupling efficiency compared to realistic incoherent broadband input. The oversampling factor for a particular bandpass with spectral range $\Delta\lambda$, and spectral resolution $\delta\lambda$ is given by $c = M\delta\lambda/(2\Delta\lambda)$, where $c = 1$ corresponds to critical (Nyquist) sampling. The input spectrum, $\mathbf{x}$, can be retrieved as the FFT of a critically-sampled interferogram, or by non-uniform Fourier-transform of an oversampled interferogram [57, 86]. Without loss of information, the number of spectral points in $\mathbf{x}$ is necessarily half the number of MZIs since the FFT requires a solution to both the amplitude and phase of each spectral component.

For this analysis we found that an LSSA retrieval using the pseudoinverse of the

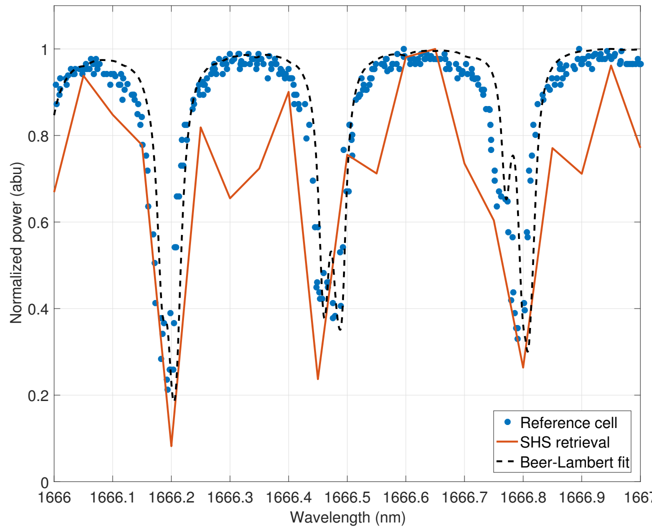

Fig. 4.5 Experimental observation of methane absorption features using on-chip SHFTS over 1 nm bandwidth with 50 pm resolution. The spectrum is obtained from the interferogram via pseudoinverse of the calibration matrix shown in Fig. 4.4. The signal collected by the reference photodiode (PD1 in Fig. 4.3(a)) is also displayed along with the fitted Beer-Lambert spectrum. The methane absorption features in the reference cell are retrieved with SNR $> 3:1$.

transformation matrix [86, 132, 135] with an oversampling factor of $c = 2.5$ provided the best SNR (see also the discussion in Section 2.1.3). The pseudoinverse of $\mathbf{A}$ is applied to the interferogram, $\mathbf{y}$, to retrieve an estimate, $\hat{\mathbf{x}}$, of the input spectrum, where $\mathbf{y} \approx \mathbf{A}\mathbf{x}$. This process may be summarized as:

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{4.5}$$

In Fig. 4.5 we present a retrieval of the methane transmission spectrum obtained from the gas cell by LSSA. The tunable laser source is used to effect a boxcar band-pass filter from 1666 nm to 1667 nm, providing an oversampling factor of $c = 2.5$. The interferogram is inverted using Eq. (4.5), where A is the calibration matrix shown in Fig. 4.4. To obtain an estimate of the methane concentration in the gas-column, the SHS spectrum is fit to a Beer-Lambert absorption model using absorption coefficients obtained from HITRAN [73]. Using SHFTS measurements the volume mixing ratio along the optical path including the gas reference cell is determined to be 0.40 ± 0.11. For the purpose of comparison, a reference volume mixing ratio value of 0.370 ± 0.009 is obtained by performing the same fitting routine on the high-resolution photodiode signal shown in Fig. 4.5. The SHFTS and reference values are in agreement, with the reference signal achieving greater precision due to its improved spectral resolution. The precision of the SHFTS retrieval may be improved by increasing the input coupling efficiency of the SHFTS using spot-size converters, surface grating couplers or micro-lens arrays. The precision may be further improved by bringing a linear detector directly to the edge of the chip in order to increase light capture; an SHFTS mated with commercial detectors would achieve 41 dB dynamic range.

The setup shown in Fig. 4.3 was also used to characterize the thermal sensitiv-

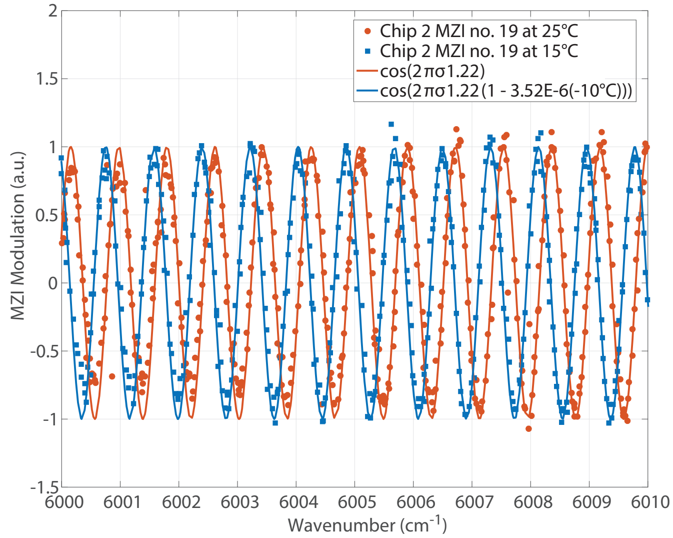

Fig. 4.6 Modulation function of a high-sensitivity MZI as a function of wavenumber at $T_1 = 25$ °C and $T_2 = 15$ °C. The modulation function is obtained as the difference in intensity between the two out-of-phase MZI outputs normalized by their sum. The OPD of the device (1.22 cm) is first constrained by a cosine fit to the modulation function at 25 °C. The TOC is then determined by fitting the modulation function at 15 °C (*i.e.*, $\Delta T = -10$ °C) to Eq. (4.4), yielding the TOC $\alpha_{eff} = -(3.52 \pm 0.05) \times 10^{-6} K^{-1}$.

ity of the chips. The TOC of the waveguides is determined using two calibration maps obtained at $T_1 = 25$ °C and $T_2 = 15$ °C. The difference in modulation period observed in a single modulation function at T_1 and at T_2 is governed by the TOC acting on the OPD. The modulation functions of MZIs with long OPDs—which exhibit greater sensitivity to temperature changes—were measured and used to determine the value of the waveguide TOC by least-squares fitting to Eq. (4.4). This procedure is shown in Fig. 4.6. The value obtained from the high-sensitivity MZIs is cross-validated against the response of low-sensitivity MZIs to eliminate the possibility of aliasing (*i.e.*, underestimation due to an apparent 2π shift in the sinusoid). The resulting TOC is $\alpha_{eff} = -3.52 \times 10^{-6} K^{-1}$, corresponding to a temperature knowledge requirement of $\Delta T < 9$ mK. This value agrees with the Monte-Carlo envelope, and represents an order-of-magnitude improvement over uncompensated Si_3N_4/SiO_2 waveguides ($\alpha_{eff} = 2.5 \times 10^{-5} K^{-1}$) and a two-order improvement over silicon waveguides ($\alpha_{eff} = 1.8 \times 10^{-4} K^{-1}$).

These results demonstrate significant advances in SHFTS on-chip capability. Through passive thermal management we demonstrate that SHFTS devices may be substantially insulated against their environment, enabling SHFTS deployment on low-power miniaturized remote sensing platforms. Secondly, through selective oversampling we demonstrate, for the first time, that SHFTS on-chip technology may be used to retrieve complex, broadband spectra, and that these spectra may be used to correctly predict the column-length of an atmospheric species.

4.4 Conclusion

In this chapter we implement a high-resolution spatially-heterodyned Fourier-transform on-chip spectrometer for remote detection of methane. In doing so, we address a number of outstanding challenges to the realization of miniaturized spectrometers for remote sensing in dynamic environments. Waveguide interferometers are lithographically defined as squared-spirals enabling a high-density layout of 100 interferometers on a device measuring 12 mm by 22 mm. The SHFTS comprises an array of 100 waveguide interferometers with independent input apertures. This design is used to enlarge the total input aperture of our waveguide spectrometer over dispersive-element systems, enhancing optical throughput. High-resolution (50 pm) spectroscopy in the presence of non-uniform spatial sampling of the interferogram is achieved using LSSA over a restricted optical bandwidth of 1 nm. The thermal sensitivity of the waveguide device is substantially moderated through the use of a passively athermal waveguide geometry, enabling the device to operate in a dynamic environment with reduced active thermal compensation. Finally, we use this spectrometer to demonstrate, for the first time, standoff detection of atmospheric absorption features using an on-chip spectrometer. The technical advances presented here extend the capability of the on-chip SHFTS architecture for high resolution spectroscopy, opening the door for future realization and deployment of integrated SHFTS devices on miniaturized remote-sensing platforms in dynamic environments.

5 Demonstration of a compressive-sensing Fourier-transform on-chip spectrometer

This chapter is based on a paper (Podmore *et al.* [132]) which was based on a collaborative discussion with Dr. Alan Scott of Honeywell, that occurred during the design of the methane spectrometer. Dr. Scott suggested that we consider the use of compressive-sensing techniques as a temperature gauge for the methane spectrometer. Following this discussion, and some independent research, I determined that a compressive-sensing waveguide spectrometer would have significant advantages over conventional designs, and that planar-waveguide Fourier-transform spectrometry is inherently well suited to compressive-sensing. Following a discussion with Dr. Pavel Cheben, of the National Research Council, it was determined that this principle could be demonstrated using an existing spectrometer developed at the National Research Council. In collaboration with Drs. Cheben, Velasco, and Schmid, as well as Martin Vachon, I was able to use this device to demonstrate compressive-sensing retrievals of monochromatic and polychromatic spectra. I designed the project and collected the data on-site at the National Research Council in collaboration with Martin Vachon. Ultimately in this chapter we demonstrate for the first time that Fourier-transform waveguide spectrometers are applicable as a vehicle for compressive-sensing and that, in doing so, they achieve a four-fold increase in free-spectral range. I performed all of the calculations, prepared all of the figures, and wrote the manuscript with revisions provided by my co-authors.

5.1 Introduction

Miniature spectrometers are an invaluable tool in many applications including environmental and biological sensing, medical diagnostics, geology, security, and planetary science, to name a few [8, 31]. They are particularly sought after in space instrumentation and planetary exploration where it is desirable to minimize the mass and volume of the instruments without compromising performance. Miniature Raman spectrometers are also important for applications in exominerology and exobiology [9]. A particularly promising platform for the implementation of miniature spectrometers is planar waveguide photonic chip technology [137]. In planar waveguide spectrometers, light is collected and routed through either a dispersive grating element [138] or a Fourier-transform interferometer array [139].

Of these two architectures, FTS are preferred for their higher optical throughput compared with dispersive grating devices [117, 137] (see earlier discussion in Chapter 1). FTS can be implemented in planar waveguides as an array of MZIs with linearly increasing OPDs [117, 133], also called an SHFTS. In an SHFTS instrument, each point in the spatial interferogram (corresponding to the output of a particular MZI) is captured independently by a linear detector array, allowing for acquisition of the entire interferogram in a single capture. However, limits on waveguide minimum bend radius, detector size and fabrication capabilities place constraints on the number of MZIs that may be integrated on a photonic chip, ultimately limiting the bandwidth and the resolution of an SHFTS device. The SHFTS is based on spatial sampling, unlike a scanning FTS [140], where the interferogram is scanned in the temporal domain and captured by a single detector operating at a high readout rate.

CS techniques can be advantageously used in SHFTS devices, providing a compelling path towards reducing the number of individual interferometers on a planar-waveguide chip and device footprint reduction (see earlier discussion in Chapter 2. In a CS scheme, the input signal is assumed to be sparse, and this assumption is used to significantly reduce the number of sampling points--or MZIs in SHFTS—to retrieve the input signal. CS schemes have already been applied to free space dispersive-element type systems, but require coded-aperture masks which further reduce optical throughput with respect to FTS [141]. Fourier-transform devices circumvent mask requirements since the sensing basis (frequency) and the measurement basis (time) are maximally incoherent, *i.e.*, a sparse signal in the frequency domain produces a non-zero signal level at all points in the time domain [95, 142]. CS schemes are particularly well suited to SHFTS devices as the sampling points in the interferogram are collected independently. Reducing the number of samples (individual interferometers) in an SHFTS directly reduces chip footprint. A discretely sampled interferogram consisting of M data points—corresponding to M individual MZIs—can be reduced to a subset of $K < M$ sampling points with no loss in spectral information. This advantage cannot be realized in a scanning FTS, which necessarily collects the full set of sampling points in the time domain.

In this chapter we demonstrate for the first time CS spectral retrieval of sparse spectra in a waveguide spatial heterodyne FTS device, implemented as an array of Mach-Zehnder interferometers. Our CS SHFTS concept is particularly promising for miniaturized Raman spectrometers, as Raman spectra exhibit a high degree of sparsity [143].

5.2 Spatial heterodyne Fourier-transform spectrometer chip

The planar-waveguide spectrometer used in this work comprises an array of 32 unbalanced MZIs with linearly increasing OPDs, as schematically shown in figure 5.1. The OPDs are produced via inducing a change in the optical propagation constant β between the two arms of the MZI by a subwavelength grating (SWG), rather than by introducing a geometrical path difference as in conventional devices. One arm of each

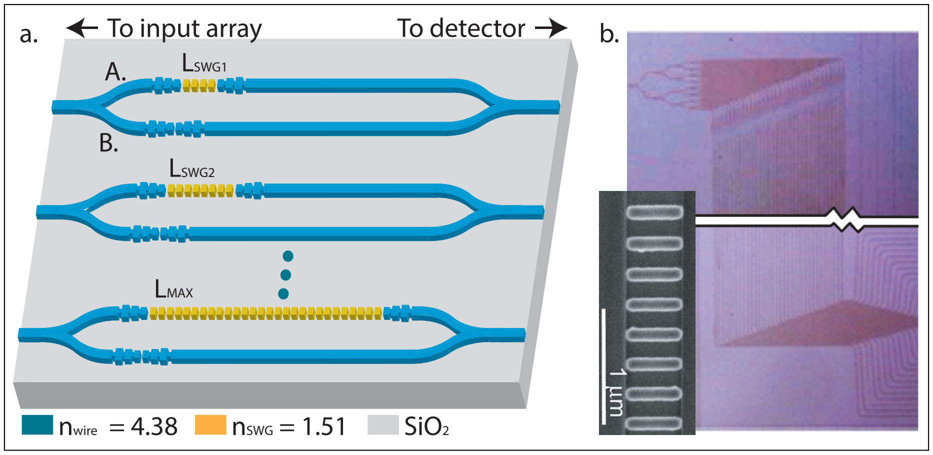

Fig. 5.1 a) Light from the spectrometer aperture is directed to an array of interferometers each consisting of a 50:50 splitter, a delay arm (A), and reference arm (B). The delay arm in the $i-$th interferometer consists of a pair of subwavelength mode-converters and a SWG region characterized by length L_{SWGi}. The reference arm consists of a 1.5-cm-long Si-wire waveguide and it also includes a compensatory set of mode-converters connected back-to-back (to balance the loss between the two arms). b) Fabricated SHFTS chip. Inset: scanning electron microscope image of an SWG waveguide section.

MZI consists of standard Si-wire waveguide, while the other arm is an SWG waveguide [121], *i.e.*, a non-diffractive, linear periodic array of Si segments with a period of less than one half of the effective wavelength to suppress diffraction effects [121]. The SWG structure acts as an optical metamaterial with decreased confinement of the propagating mode, resulting in a reduced group index that can be adjusted by selecting the periodicity and duty-cycle of the grating. Further information regarding the fabrication and mode propagation of these waveguides can be found in Section 3.1.1.2 and in the literature [121, 144].

The MZIs are fabricated on a silicon-on-insulator wafer with a 260-nm-thick Si waveguide layer and a 2-μm-thick buried oxide. A 2-μm-thick SU-8 polymer layer is used as the upper cladding. The Si-wire waveguide nominal width is 450 nm, while the SWG grating waveguide has a width of 300 nm, grating periodicity of 400 nm, and a 50% fill-factor. This results in a group index of $n_{SWG} = 1.51$ for the SWG waveguide and $n_{wire} = 4.38$ for the photonic wire waveguide [121, 144]. The corresponding difference in the group index between the two MZI arms results in the required phase imbalance, controlled across the array by changing the length of the subwavelength region in the delay arm.

The spatial light distribution at the MZI array output ports forms a discretely-sampled inverse Fourier-transform of the input spectra. The MZI number (M) and OPD physical lengths ($L_{SWGi} = i \times L_{SWG1}$) are selected in order to set the resolution, ($\delta\lambda$), and bandwidth, ($\Delta\lambda$), of the device according to

$$\delta\lambda = \frac{{\lambda_0}^2}{L_{max}(n_{wire} - n_{SWG})} \tag{5.1}$$

$$\Delta\lambda = \delta\lambda\frac{M}{2}. \tag{5.2}$$

Where λ_0 is the central wavelength of the device (1550.5 nm) and $L_{max} = M \times$

L_{SWG1} [133]. In the device under analysis, OPD lengths increase linearly from $L_{SWG1} = 470$ μm, up to a maximum of $L_{max} = 1.5$ cm for $M = 32$. Experimentally, the SHFTS resolution and bandwidth were determined as $\delta\lambda = 48$ pm, and $\Delta\lambda = 0.78$ nm, in good agreement with the theoretical values ($\delta\lambda = 55$ pm and $\Delta\lambda = 0.89$ nm, respectively).

An important advantage of inducing the optical path delays through a change in the mode propagation constant (effective index) rather than a change in the physical path length is that the fringe visibility of the MZIs, and hence the signal quality, is improved. The visibility factor, $V = (I_{max} - I_{min})/(I_{max} + I_{min})$, where I_{max}, and I_{min} are the maximum and minimum MZI output signals, increases as the loss imbalance between the two MZI arms decreases.

In general, the observed fringe maxima and minima in a specific MZI are determined by the propagation loss, α, and the physical path lengths of the two arms, z. Since in our FTS with SWG delay lines there is no geometrical path length difference between the two MZI arms, the fringe visibility corresponds to the difference in propagation loss between the SWG and wire waveguides. In our device the losses are nearly identical for SWG and wire waveguides, $\alpha_{SWG} = -3.0$ dB/cm and $\alpha_{wire} = -3.1$ dB/cm, resulting in a measured visibility factor $V = 0.96$.

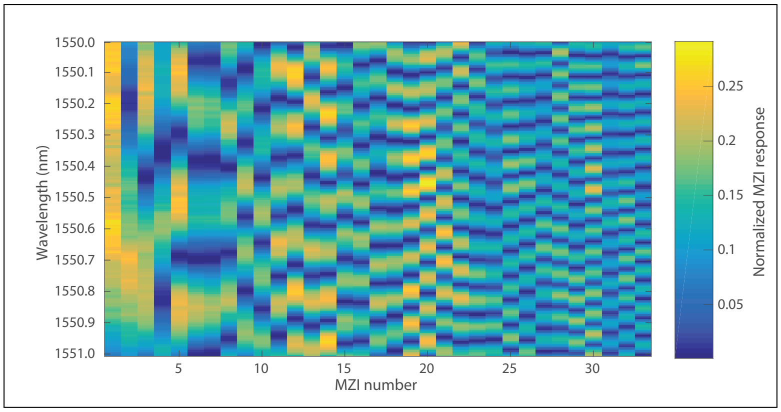

Fig. 5.2 Measured calibration matrix of the spectrometer chip, showing the wavelength-dependent output power of each of the 32 MZIs as obtained in a high-resolution wavelength scan for a 1 nm spectral range centered at 1550.5 nm.

5.3 Spectral retrieval experiment

A high-resolution C band tunable laser (Agilent 81682A) was used as an input source to characterize the SHFTS chip. The wavelength was scanned in 5 pm steps over a 1 nm spectral range (encompassing the full bandwidth of the device) for the input power of 1 mW. A polarization controller was used to set the input polarization state to transverse electric (TE), and the light was coupled to the chip via a lensed fiber and on-chip mode transformer [145]. The input power was split on-chip and routed to the individual MZIs using cascaded Y-splitters. The output signal from each output was collimated using a high numerical aperture microscope objective (NA = 0.4), and this collimated beam was captured using a calibrated InGaAs camera (Sensors Unlimited SU320M-1.7RT). Throughout this procedure, the chip temperature was stabilized to within 0.1 °C with a Peltier stage. The product of this characterization (calibration) procedure is an $M \times N$ transformation matrix, $\mathbf{\Phi}$, in which M is the number of waveguide outputs, and N is the number of measured wavelength points. The experimental matrix obtained from our SHFTS chip is shown in figure 5.2, demonstrating good fringe visibility and signal-to-noise ratio for the chosen input power. In this transformation matrix (also called a spectral or calibration map, as described in Section 2.1.3) each row contains the phase and frequency modulation information for a single MZI in the spectrometer. The transformation matrix can be directly used to retrieve an input spectrum via its pseudoinverse [133].

The transformation matrix can be used to implement a CS scheme, wherein an unknown, sparse input spectrum, x, produces an output signal, y, via the sensing scheme:

$$y = \mathbf{A}x. \tag{5.3}$$

In this formulation, $\mathbf{A}$ is a $K \times N$ matrix with $K < M$ that satisfies the RIP; *i.e.*, the columns in A are approximately orthogonal. Randomly selected rows of a Fourier-transform matrix are known to satisfy the RIP, hence a set of K randomly selected MZIs will suffice to define $\mathbf{A}$ while satisfying the RIP [146]. Selecting a subset of K rows from the calibration matrix is equivalent to measuring the outputs of a restricted set (K) of MZIs, without considering the remaining ($M - K$) MZIs. Once the output signals of the selected (K) MZIs have been collected, equation 5.3 is solved via l_1-norm minimization, basis-pursuit, to calculate the input spectrum x. In this work, the full set of MZIs is included on the chip for the purposes of establishing a baseline comparison, but only restricted set (K) of MZIs is used to demonstrate the CS scheme spectral retrieval.

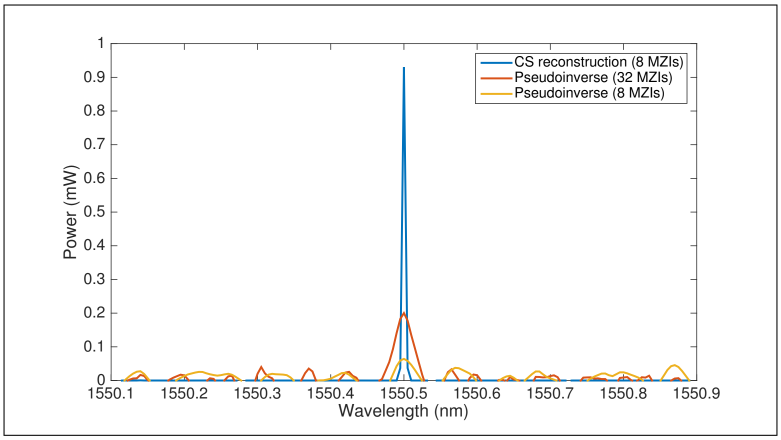

Fig. 5.3 Spectra from narrow-band laser source with 1 mW input power retrieved using un-apodized interferometric measurements on the MZI array. A subset of 8 MZIs is randomly selected and their output intensities are used to retrieve the laser spectrum via l_1-norm minimization (blue). For comparison, the spectrum is also reconstructed via the conventional pseudoinverse method using the full set of 32 MZIs (orange), as well as for the subset of 8 MZIs (yellow).

First, we consider a monochromatic input spectrum defined by a narrow-band laser source at $\lambda_0 = 1550.5$ nm (linewidth and frequency excursions less than 0.5 pm) and 1 mW input power. 8 MZIs are selected randomly from the interferometer array, their output values are recorded, and the input spectrum is retrieved through l_1-norm minimization via basis-pursuit de-noising [142]. This spectrum is shown in figure 5.3 alongside a retrieval obtained using the pseudoinverse of the calibration map, $\mathbf{\Phi}$, as well as the pseudoinverse of $\mathbf{A}$. The retrieval obtained via l_1-norm minimization lacks the hyperbolic sinusoid ripple associated with an FTS device, which can be observed in the result obtained using the pseudoinverse of the full set of MZIs. The instrument line shape produced by this ripple has a full width at half-maximum (FWHM) of 0.029 pm. The lack of ripple in a CS FTS device is a consequence of *a priori* assumption of sparse spectrum in the CS schemes. We find that the spectra retrieved via l_1-norm minimization successfully reconstructs the frequency of the laser line despite substantial undersampling of the interferogram, with an undersampling factor of $c = 0.25$. By contrast, it is not possible to retrieve a meaningful spectra from the same 8 MZIs using the pseudoinverse of the $\mathbf{A}$ matrix, as the interferogram is too greatly undersampled for non-CS methods. We note that the total power retrieved in all three methods is preserved ($\sim$ 1mW) when integrated within the instrument spectral bandwidth. There is no observable numerical loss in the pseudoinverse retrieval, with the input light intensity being spectrally redistributed due to the line widening, sidelobes, and ripple effects.

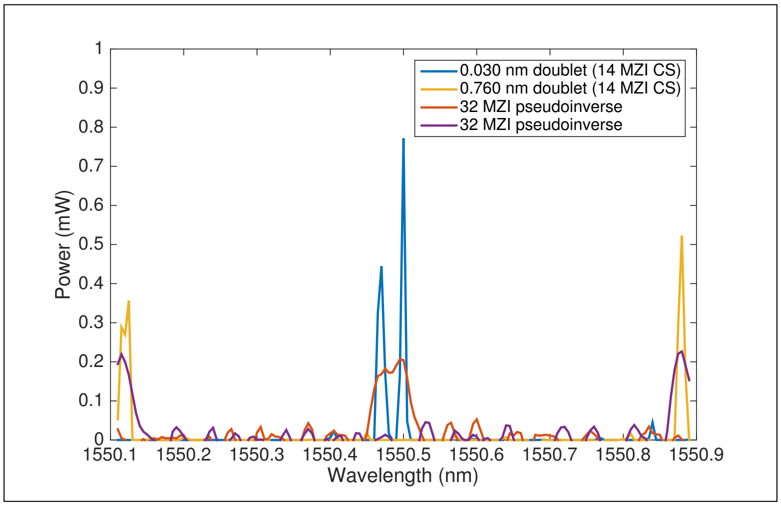

Fig. 5.4 Spectra from two doublets, with peak-to-peak separation of 0.030 nm and 0.76 nm, retrieved using un-apodized interferometric measurements of the MZI array. A subset of 14 MZIs is randomly selected and their output intensities are used to retrieve the spectra via l_1-norm minimization (blue, yellow). The spectra reconstructed via pseudoinverse methods using the full set of 32 MZIs (orange, purple) are included for comparison.

In figure 5.4 we present the retrieval of two polychromatic signals (doublets) representing limiting cases of narrowband and broadband signals. The first doublet is spaced at 0.030 nm, closely matching the FWHM of the instrument lineshape. The sharp spectral features of the CS retrieval, produced by the *a priori* assumption of sparsity, allow for clear resolution of the finely spaced doublet. By contrast, the pseudoinverse retrieval is under-resolved with respect to the Rayleigh criterion [147]; this effect is analogous to the spectral broadening of an apodized interferogram, except that the interferogram has been apodized by setting random points in the interferogram to zero (see Section 2.1.2). A second doublet retrieval with separation 0.76 nm, near to the bandwidth of the device, is also shown. As in the case of the monochromatic source, it was not possible to retrieve the spectra of either doublet using the pseudoinverse of the matrix **A**, as the interferogram produced by the 14 MZIs is excessively undersampled.

The results presented in figures 5.3 and 5.4 demonstrate that the bandwidth and resolution of the CS spectrometer are similar to the conventional non-CS spectrometer, but with a significantly reduced number of interferometers for the CS FTS. The latter requires only 14 MZIs compared to 32 MZIs for the non-CS FTS in the case of a doublet (undersampling factor $c = 0.44$), and only 8 MZIs to retrieve a monochromatic source (undersampling factor $c = 0.25$). The capability of the CS-spectrometer to operate with a reduced number of intermediate MZIs also offers a promising path for spectral range expansion at a given resolution, as CS techniques can directly offset the MZI number increase required by equation 5.2. The substantial reduction in the number of sampling interferometers comes with a compromise that the CS spectrometer can only be applied to sparse signals. There are, however, a large number of important naturally occurring sparse spectra. For example, both Raman and LIBS techniques produce sparse signals. Strong spectral emission or absorption lines of interest *e.g.*, to atmospheric science and astronomy may also be considered sparse if a suitable background removal scheme is applied.

5.4 Conclusions

In this chapter we demonstrate that the spatial heterodyne configuration of on-chip FTS devices, consisting of an array of independent interferometers, is uniquely well-suited to CS spectroscopy. We use an SHFTS on a photonic chip with an array of Mach-Zehnder interferometers to experimentally demonstrate this technique. Subwavelength grating engineered optical delay lines are used in the interferometer arms to minimize the propagation loss imbalance and maximize the fringe visibility of the interferogram. We take advantage of the spatial-heterodyne configuration to independently sample the MZIs, and selectively down-sample the spatial interferogram. Using restricted sets of MZIs and l_1-norm minimization we retrieve spectra with full resolution and bandwidth, specifically a singlet (with 8 MZIs) and a doublet (with 14 MZIs), while a conventional non-CS retrieval requires a substantially larger interferometer array (32 MZIs). These results demonstrate that compressed-sensing can be effectively implemented on an FTS chip, allowing a substantial reduction in the number of individual interferometers required for a given resolution and bandwidth. These results open excellent prospects for further miniaturization of integrated Fourier-transform spectrometer devices.

6 On-Chip Compressed Sensing Fourier-Transform Visible Spectrometer

This chapter is based on a paper (Podmore *et al.* [148]) in which the advances presented in chapter 4 are tied to the novel developments in chapter 5 to present a Fourier-transform waveguide spectrometer that leverages compressive-sensing to expand its bandwidth by a factor of four. I developed an instrument model and retrieval algorithms to demonstrate that Raman spectra are suitable for detection by compressive-sensing, and verified this algorithm using representative datasets obtained from Professor Ed Cloutis at the University of Winnipeg, and the Herschel space telescope archive. I designed a visible-wavelength compressive-sensing waveguide spectrometer using the scripts described in Appendix A, and developed a novel broadband input coupling scheme and waveguide geometry in order to support the expanded bandwidth of the device. I developed a novel characterization scheme using hyperspectral imagery for snapshot waveguide characterization, and collected and analyzed the characterization data. This chapter demonstrates the first ground-up compressive-sensing waveguide spectrometer. This chapter also validates the algorithm developed to retrieve data in the presence of significant background contamination—a known challenge for compressive-sensing instruments. I performed all of the calculations, prepared all of the figures, and wrote the manuscript with revisions provided by my co-authors.

6.1 Introduction

The widespread commercial availability of miniaturized spectrometers that may operate outside of a laboratory environment has been a boon to many scientific disciplines including greenhouse gas detection, oceanography, and biomedical sciences [51]. In the field of planetary science, the need for low-mass, low-volume rover platforms has driven considerable interest in miniature spectrometers for Raman spectroscopy [9]. Similarly, in the space sector, following the advent of the micro and nanosatellite revolution, there has been increased interest in miniature spectrometers with regards to remote-sensing [31].

One promising platform for the development of these miniaturized spectrometers is integrated optics. In the commercial sector, the substantial investment in optical communications technology has lead to the development of highly-miniaturized PICs which accomplish many of the primary functions of a spectrometer. PICs consisting of high-density arrangements of waveguides, splitters, couplers, gratings and photodetectors printed on silicon wafers may represent a potentially advantageous technology for the development of miniaturized spectrometers [37]. With PICs, it is possible to realize a complex optical layout lithographically, thereby reducing the need for the assembly and alignment of individual optical components. In addition, the size and mass of such optical chips may be substantially lower than comparable assemblies of free-space beam splitters, gratings, and lenses.

PIC spectrometers have been realized as both dispersive and FTS either using dispersive gratings [37], filter-arrays [149] or scanning MZIs (see also Chapter 1). Such devices may achieve high spectral resolution and large optical bandwidths on a single photonic chip measuring a few hundred square millimetres [150]. In the case of each of these designs, the optical throughput of the system is limited by the étendue of the access waveguide defining the entrance aperture to the spectrometer. Throughput may be improved by using multimode waveguides, however, for dispersive spectrometers, high spectral resolution may only be achieved by limiting the entrance aperture to one single-mode waveguide.

An attractive method for achieving increased optical throughput in high-resolution on-chip spectrometers is to form a SHFTS. An SHFTS consists of an array of MZIs which, in aggregate, collect a discrete Fourier-transform of the input spectrum [117, 125, 132, 133]. The advantage of the SHFTS architecture is that light may be coupled independently into each MZI, greatly expanding the total input aperture of the instrument over dispersive spectrometers—which must be fed by a single access waveguide (see discussion in Section 1.2.3). On-chip SHFTS devices may achieve picometer spectral resolution, however their bandwidth is determined by the number of MZIs in the array and is therefore limited by the physical size of the photonic chip. Recently we have demonstrated that this bandwidth restriction can be mitigated using CS methods. In particular, using CS retrieval methods the number of MZIs required to retrieve a spectral signature over a specific bandwidth may be reduced by a factor of four when detecting spectrally sparse signals [132].

In this chapter we present an innovative multi-aperture visible-wavelength on-chip CS SHFTS for sparse signal characterization achieving high spectral resolution over an extended bandwidth. The SHFTS leverages CS retrieval methods to under-sample its interferogram by a factor of $c = 1/4$, expanding the operational bandwidth without sacrificing resolution. Using a fully-sampled subset of low-resolution Mach-Zehnder Interferometers (MZIs) to estimate non-sparse background signals, we demonstrate CS retrieval of sparse spectra in the presence of a broadband contaminating background signal. We implement a novel multi-aperture vertical coupling scheme to achieve broadband input coupling to 83 independent apertures. Finally, we demonstrate a novel method for snapshot waveguide characterization via white-light hyperspectral imaging at the waveguide outputs.

6.2 Theory

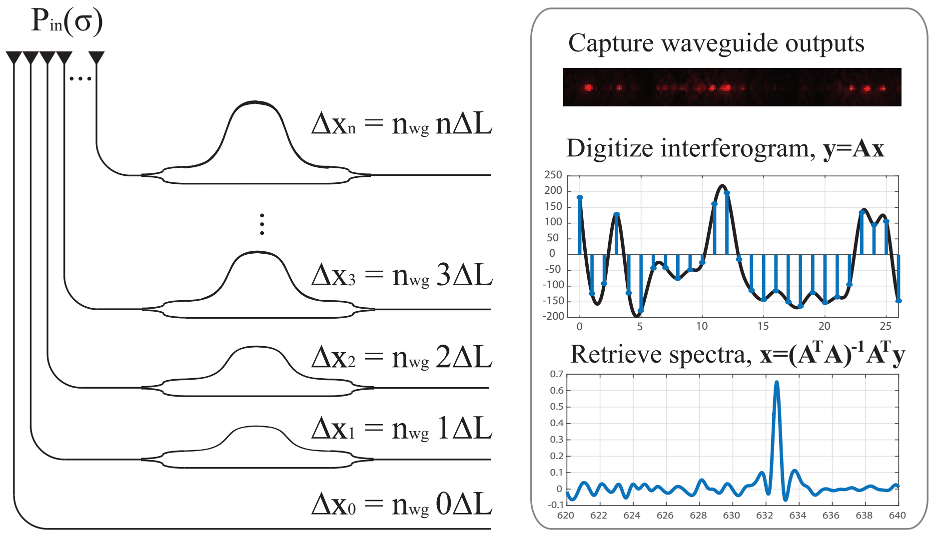

Fig. 6.1 Diagram describing SHFTS principle of operation: MZIs implemented as U-bends (left) collect a discrete sampling (interferogram) of the inverse Fourier-transform of an input spectrum. The temporal sampling point of each MZI, as in Michelson interferometry, is defined by the MZI OPD $\Delta x_i = n_{eff} \cdot i\Delta L$. The interferogram is then defined by the intensity of the waveguide outputs (right), which may be imaged, digitized, and inverted to retrieve the input spectrum.

In an on-chip SHFTS, light is coupled to an array of planar-waveguide MZIs which output a spatial interference pattern constituting a discrete sampling of a continuous interferogram—equivalent to the output of a triggered Michelson interferometer. This concept is illustrated in Figure 6.1.

The output intensity of each MZI in the SHFTS, with fixed OPD Δx, is given by the IFFT:

$$I(\Delta x) = \frac{1}{2}P_0 + \int_0^{\Delta\sigma} \frac{1}{2}P_0 cos(2\pi\Delta x\sigma)d\sigma \tag{6.1}$$

Where σ is optical wavenumber, and $\Delta\sigma$ the free spectral range of the input spectrum, P_0. The continuous IFFT in Equation 6.2, when sampled discretely at fixed path increments, δx, may be inverted using the discrete Fourier-transform (DFT):

$$P(k\cdot\delta\sigma) = \sum_{i=0}^{N-1} I(i\cdot\delta x)cos(2\pi ik/N) \tag{6.2}$$

where $N = 2\Delta\sigma/\delta\sigma$ is the number of MZIs (sampling points) in the array, and $\delta\sigma$ is the spectral resolution.

Each MZI in the array has an associated OPD $\Delta x = n_{wg}\Delta L$, determined by the waveguide refractive index, n_{wg}, and the physical path difference of the MZI, ΔL. The OPDs of the individual MZIs increase linearly along the array such that the OPD of the i-th MZI is determined by $\Delta x_i = i\cdot n_{wg}\Delta L$, where ΔL is a fixed length increment.

The bandwidth (spectral range), and resolution of the FTS are set by the sampling frequency of the interferogram, δx, as well as the length of the greatest OPD, Δx_{max}. The bandwidth in wavenumbers is given as $\Delta\sigma = 1/(2\delta x)$ where a factor of two is imposed by the aliasing limit of the Nyqyuist sampling theorem. The spectral resolution is determined by the greatest OPD as $\delta\sigma = 1/(\Delta x_{max})$. Thus, an SHFTS chip consisting of $N = \Delta L_{max}/\delta L$ MZIs will capture a spectrum with bandwidth $\Delta\sigma = 1/(2n_{wg}\delta L)$ and resolution $\delta\sigma = 1/(n_{wg}\Delta L_{max})$. In an on-chip SHFTS, OPDs are inscribed into the chip lithographically as in Figure 6.1.

Our SHFTS device is designed to operate over a spectral range of $\Delta\sigma = 2850\ cm^{-1}$, corresponding to an average waveguide index of $n_{wg} = 1.5$ and physical path length increment of $\delta L = 1.077\ \mu m$. The bandwidth spans the shifted wavenumbers 2970 cm^{-1} to 120 cm^{-1}, where shifted wavenumber refers to the offset between the measured wavenumber and the excitation wavenumber of a typical green (532 nm) Raman system: $\sigma_0 = 18797\ cm^{-1}$. The spectral resolution is $\delta\sigma = 9\ cm^{-1}$ corresponding to the maximum physical path delay of $\Delta L = 0.074\ cm$. Using conventional retrieval methods, therefore, this device should consist of 633 MZIs, however using CS we are able to reduce the number of MZIs to $N = 160$.

6.3 Spectral retrieval in on-chip SHFTS

Spectra may be recovered from an SHFTS using the DFT in Equation 6.2; more commonly, LSSA is used [86, 125, 133]. LSSA, described also as pseudoinverse retrieval, may be used to calibrate and compensate for waveguide non-uniformities as part of the retrieval process. The LSSA method describes the interferogram, $\mathbf{y}$, as a linear product of the the input spectrum, $\mathbf{x}$, and the discrete cosine transform matrix $\mathbf{A}$

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{6.3}$$

For a perfect cosine matrix, Equation 6.3 is equivalent to the DFT equation. However, due to variations in waveguide dimensions during fabrication, the transformation matrix will exhibit defects and non-uniformities unique to each chip. In a practical scenario, $\mathbf{A}$ will be an experimentally-determined transform matrix in which each row captures the spectral response of a particular MZI in the SHFTS—including spectral non-uniformities, coupling efficiency and phase errors. By using the pseudoinverse of this "calibration matrix" to invert Equation 6.3 the phase errors and waveguide defects may be mitigated, and the input spectrum recovered as:

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-\mathbf{1}}\mathbf{A}^{\mathbf{T}}\mathbf{y}, \tag{6.4}$$

where $\hat{\mathbf{x}}$ denotes an estimate of the true spectrum $\mathbf{x}$.

The formulation in Equation 6.3 may also be used as the basis for a CS retrieval [125] in which the sampling rules may be relaxed, and the input spectrum retrieved as the solution to an optimization problem minimizing the l1-norm:

$$min||\hat{\mathbf{x}}||_{l1} \; subject\, to\, \mathbf{y} = \mathbf{A}\hat{\mathbf{x}}. \tag{6.5}$$

Here $||\hat{\mathbf{x}}||_{l1} = \sum_{i=0}^{n} |x_i|$ denotes the l1-norm of the vector $\hat{\mathbf{x}}$ [95, 142]. Minimizing the l1-norm, provides the "sparsest" solution to Equation 6.3, meaning a solution wherein the power is concentrated in the least number of spectral components. The advantage of a CS retrieval is that, for sparse input spectra, accurate reconstruction may be achieved using far fewer samples than the in the case of LSSA or DFT. The corresponding disadvantage of CS retrieval methods is that the input signal must be spectrally sparse in order to achieve accurate retrieval. This assumption of sparsity is not necessarily an impediment as there are many spectrally sparse signals of interest to spectroscopists including laser lines, atomic emission lines, Raman emission and LIBS.

In addition to the assumption of sparsity, there exist two conditions for under-sampled retrieval via CS: first, that the measurement basis and signal basis are mutually incoherent, and second, that $\mathbf{A}$ be a unitary operator when applied to sparse vectors—the RIP. Many matrices satisfying the RIP exist, including partial Gaussian, Bernoulli and partial Fourier matrices in which each row in $\mathbf{A}$ is randomly selected from a larger DFT matrix [93, 94].

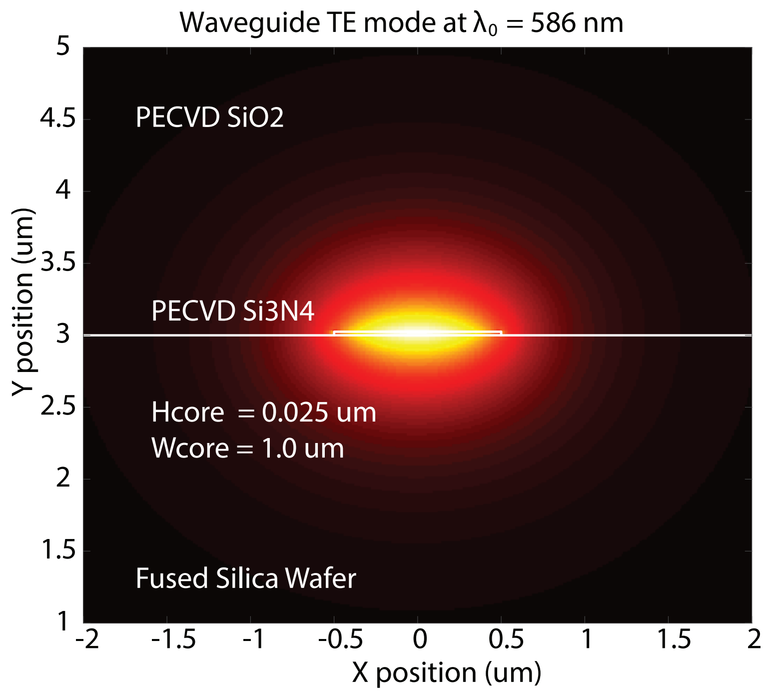

Fig. 6.2 Cross-section and mode-field distribution (MFD) for the low-confinement high aspect-ratio single-mode silicon nitride waveguides. The low profile of the waveguide core, measuring 1 $\mu m \times 25\ nm$, produces a loosely-confined waveguide mode with broadband operation encompassing 532 nm to 640 nm.

In an SHFTS, the first condition is satisfied by the natural incoherence between the time (measurement) and frequency (signal) bases [95]. A properly designed SHFTS may also satisfy the second condition by forming a partial Fourier matrix using a randomly selected subset of MZI OPDs drawn from the full set defined by the desired bandwidth and resolution. This analog downsampling operation is accomplished facilely in an on-chip device, where the designer has complete control over the OPDs of each sampling point in the interferogram. By contrast, free-space SHFTS and Michelson interferometers, which do not have independent control of sampling points, must record full interferograms and synthesize a partial Fourier measurement by digital downsampling. We have previously demonstrated, using digital downsampling on a fully-populated SHFTS chip, that CS may be used to reduce the number of MZIs required to achieve a given bandwidth by a factor of four [132]. In this new spectrometer we undersample in analog by fabricating an under-populated SHFTS containing only 160 MZIs of the 633 MZIs required for LSSA retrieval.

As noted, the disadvantage of CS is that it may only be applied to spectra that are known to be sparse. Though naturally-occurring sparse signals do exist, this restriction precludes the presence of any broadband background light in the measurement. This requirement is easily satisfied in the case of a laser line or rare-gas spectrum, but is less straightforward in the case of Raman and LIBS measurements—which can contain significant amounts of background light from laser-induced fluorescence [48]. To overcome this limitation, we select 40 of the 160 MZIs to be non-random, fully-sampled MZIs that may provide a low-resolution LSSA estimate of any broadband background fluence. An observed interferogram may be considered as a combination of its sparse and non-sparse components; we can retrieve these components indepen-

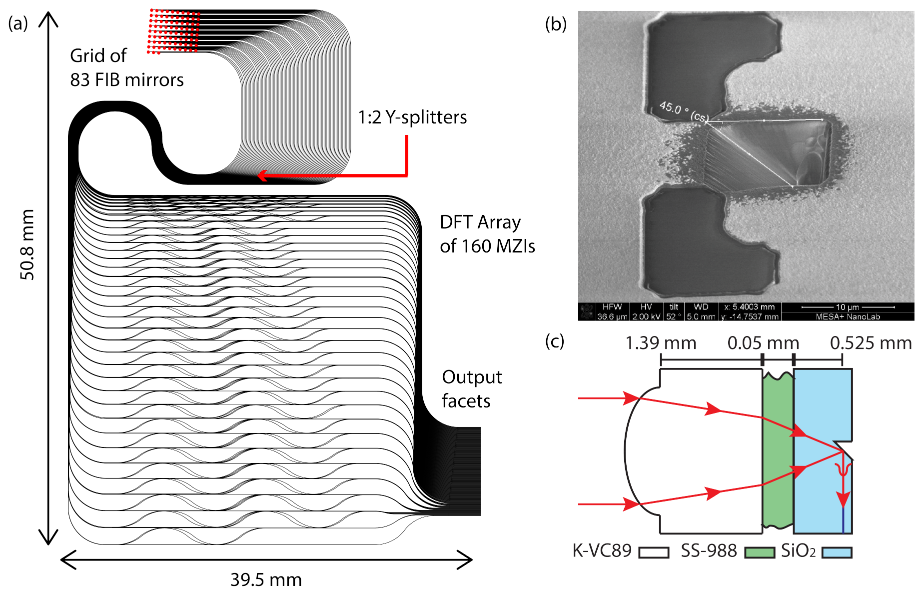

Fig. 6.3 (a) Chip layout of the SHFTS design with FIB micro-mirrors highlighted in red; the dimensions of the layout are 39.5 mm × 50.8 mm. (b) SEM image of an FIB mirror etched at 45 from the wafer surface. (c) Diagram of FIB mirror input coupling scheme. A collimated beam is focused by a high-index (n=1.72) lens, and passes through an index-matching optical gel (n=1.47) into the substrate of the SHFTS chip. The focused spot is focused onto the FIB mirror, and light is deflected into the waveguides by internal reflection.

dently by using the low-resolution MZIs to estimate the background signal, and the high-resolution MZIs to retrieve spectrally-sparse components via l1-norm minimization. In this manner, we may perform a CS retrieval of the sparse components even in the presence of non-sparse background contamination.

6.4 Experiment

SHFTS chips were fabricated by LioniX International. The substrate is a 500 μm thick fused silica wafer, the waveguide core consists of a thin stripe of ultra-high aspect ratio silicon nitride, and the upper cladding consists of an 8 μm thick layer of silicon dioxide deposited by plasma-enhanced chemical vapour deposition. The low mode-confinement high aspect-ratio core, measuring 1 $\mu m \times 25$ nm, ensures single-mode operation over the selected bandwidth of the spectrometer; the waveguide core dimensions and MFD are shown in Figure 6.2. The device footprint encompasses a 39.5 $mm \times 50.8$ mm rectangle as shown in Figure 6.3. This large footprint, a consequence of the low-confinement high-bandwidth waveguides, demonstrates the advantage of the CS architecture. A non-compressive device with the full complement of MZIs would require a significantly larger wafer. The waveguide index is determined to be $n_{wg} = 1.504$, and the propagation loss is $\alpha = -0.61$ dB/cm.

To support multi-aperture input coupling over the extended bandwidth of this device, we implement a novel input coupling scheme as shown diagrammatically in

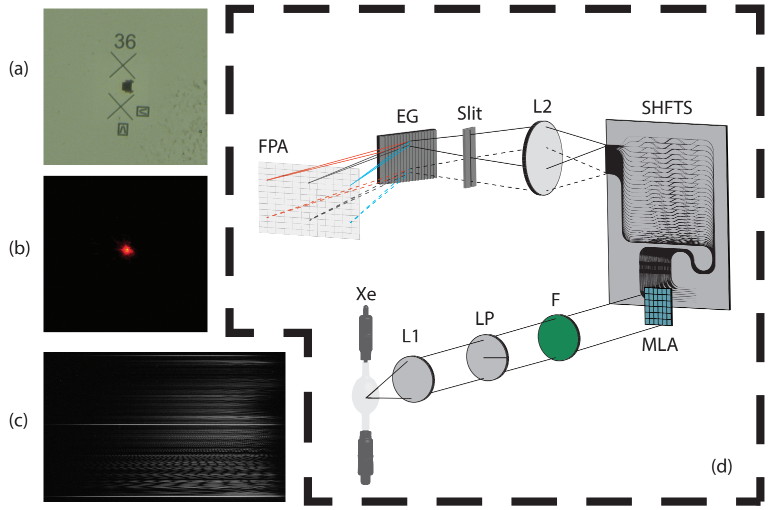

Fig. 6.4 (a) Illuminated backside microscope image of 45° micro-mirror showing alignment markings. (b) Backside microscope image of 45° micro-mirror showing focused spot from the MLA. (c) Image obtained from hyperspectral slit imager; waveguides are focused in the vertical (spatial) axis, and dispersed along the horizontal (spectral) axis. (d) Diagram of experimental setup: collimated light from a broadband source (Xe) passes through a collimating lens (L1), linear polarizer (LP), and a bandpass filter (F) to the optical chip. Light is focused to individual waveguides via the microlens array (MLA) and coupled at the waveguide core via 45° FIB mirrors. The MLA focal points are aligned using an imaging microscope focused at the backside of the chip. A telecentric lens (L2) images the waveguide outputs onto the slit of a hyperspectral imager, the full spectral output of each waveguide may then be captured by the CCD in a single frame.

Figure 6.3. Multi-aperture input coupling, the traditional advantage of SHFTS implementations, yields an increased light-collection area over similar devices employing a single-access waveguide with cascaded y-splitters [125]. We select a free-space vertical-coupling scheme, based on 45° micro-mirrors etched into the waveguide layer, in order to eliminate the need for a many-element fiber bundle or v-groove array. In this scheme, light from a collimated beam is focused by the MLA and directed upwards through the substrate into the waveguide core via 45° mirrors etched by focused ion beam (FIB). Reflective coupling via mirrors (rather than gratings) is pursued in order to minimize wavelength selectivity.

Waveguide facets are aligned simultaneously to the focal points of the MLA using an imaging microscope positioned behind the chip; once the alignment is complete, the MLA is bonded to the substrate. An SEM micrograph of an FIB micro-mirror is shown in Figure 6.3, and a visual microscope image demonstrating backside alignment of the focused spots is shown in Figure 6.4.

We characterize the SHFTS by examining its spectral response from a white-light source. We use a Xenon arc-lamp to provide a broadband input spectrum including the expected wavelengths of operation for a Raman system with excitation wavelength 532 nm and instrument bandpass over 120 cm^{-1} to 2970 cm^{-1} shifted wavenumbers. We record the SHFTS response to the broadband input using a hyperspectral slit-imaging system. Collimated light from the arc-lamp is passed to the

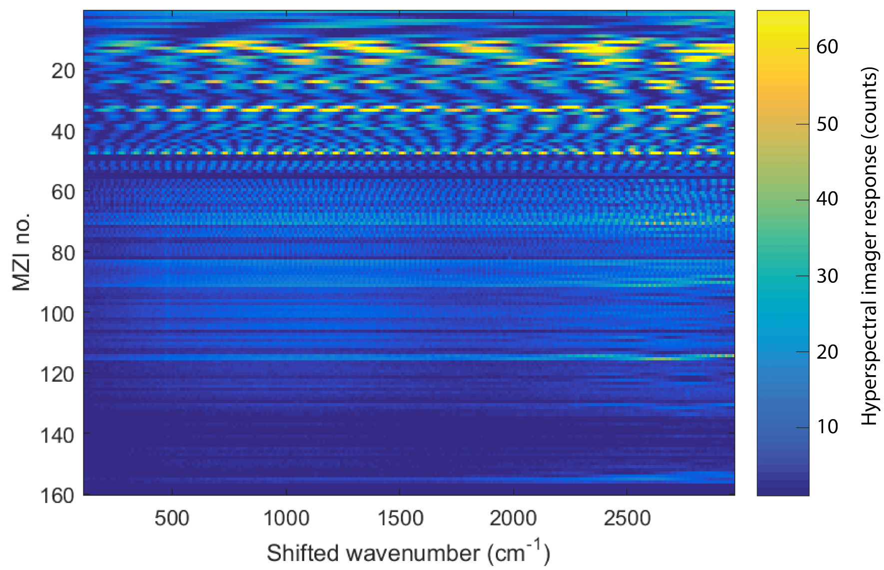

Fig. 6.5 Calibration matrix of the SHFTS chip showing the wavelength-dependent response of each MZI across a 132 nm bandwidth. The optical output of the MZIs is a sinusoidal pattern corresponding to wavelength points with high transmission (yellow) and low transmission (blue). MZIs are ordered by increasing OPD, with the shortest MZIs at the top and the longest at the bottom. OPDS increase in linear increments of 1.077 μm up to MZI no. 40, thereafter increasingly randomly up to a maximum path delay of 0.074 cm.

MLA and waveguide facets through a linear polarizer, isolating the transverse-electric mode. Polarized light propagates through the array of MZIs to the waveguide outputs, which are imaged onto the slit of the hyperspectral instrument using a telecentric lens. The wavelength-dependent response of each MZI is then dispersed by a blazed grating and captured by a CCD. A diagram of the setup is shown in Figure 6.4, and the calibration matrix obtained is shown in Figure 6.5. This novel waveguide characterization scheme was adapted from similar methods used to determine the group-index of bulk materials [151]. One advantage of hyperspectral characterization—rather than characterization using a tunable laser—is that the calibration matrix is collected instantaneously, and thermal drift during calibration is therefore minimized.

As the setup shown in Figure 6.4 is intended for characterization in a laboratory setting, it is not designed for compactness. A miniaturized SHFTS package consisting of a fore-optics unit (such as a telescope or fiber probe), SHFTS and MLA, and a linear detector array bonded to the edge of the chip would be substantially smaller.

6.5 Results

In Figure 6.6 we show retrievals of a near-monochromatic signal centered at 690 shifted wavenumbers from the nominal Littrow wavelength of $532nm$. We position a narrowband (50 cm^{-1} full-width at half-maximum) optical filter in the optical path, as shown in Figure 6.4. An interferogram is obtained from the hyperspectral

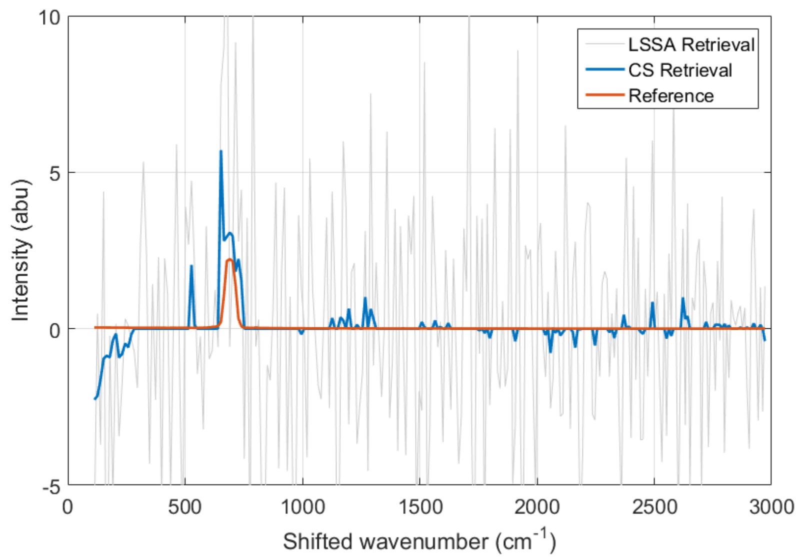

Fig. 6.6 Experimental retrieval of a near-monochromatic sparse signal, produced by a narrowband optical filter at 690 shifted wavenumbers (552 nm). An l1-norm minimization retrieval (blue) is presented alongside an LSSA retrieval (grey), and the true input spectrum (orange). Both retrievals retrieve the narrowband filter peak at 690 cm^{-1}; the SNR of the CS retrieval is 5.8, whereas the SNR of the LSSA retrieval is 3.1

image by integrating along each waveguide row, and spectra are retrieved via LSSA, l1-norm minimization, and via direct readout from the hyperspectral system. The measurement of interest in this retrieval is the difference in intensity between the retrieved filter line and the background, versus the noise floor. Using a CS retrieval a SNR of 5.8 is achieved, whereas the LSSA retrieval achieves SNR 3.1.

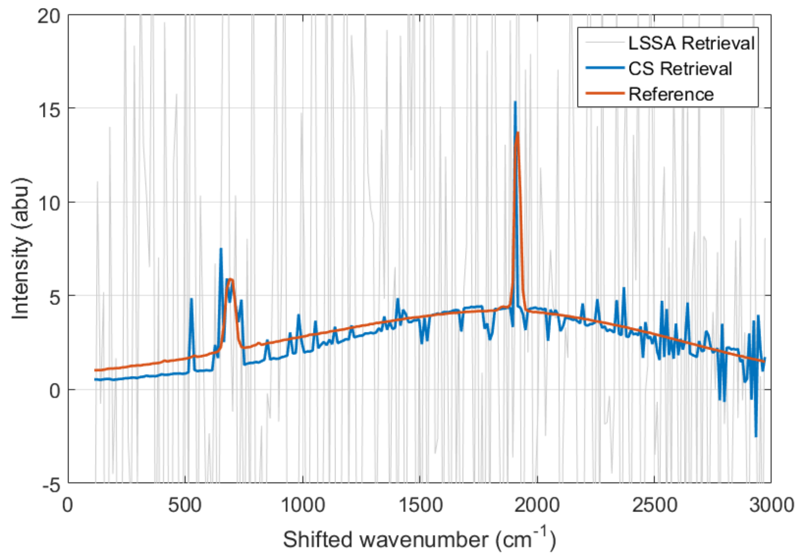

Fig. 6.7 Experimental retrieval of sparse spectra in the presence of non-sparse background signal. The sparse spectra consists of two narrowband filter lines located at 690 cm^{-1} and 1918 cm^{-1} (552 nm and 592 nm, respectively), while the background light spans the full bandwidth of the SHFTS. A CS retrieval (blue) is plotted against an LSSA retrieval (grey), as well as the input spectrum obtained by the hyperspectral imager (orange). The CS retrieval positively identifies the filter lines with SNR 4.3 at 690 cm^{-1}, and 9.9 at 1918 cm^{-1}, the LSSA retrieval fails to differentiate either line from the noise floor.

In Figure 6.7 we show retrievals of a complex spectrum consisting of both sparse and broadband components. The interferogram for this spectrum is obtained by co-adding interferograms captured from multiple narrowband optical filters as well as attenuated broadband light from the Xenon arc lamp. An estimate of the background spectrum and interferogram is retrieved from the combined interferogram using the 40 low-resolution MZIs. The estimate is subtracted from the combined interferogram, and the residual interferogram is inverted using l1-norm minimization; the full input spectrum is then produced as a combination of the residual, retrieved via CS, and the background retrieved via LSSA. The SNR of the retrieved filter lines at 690 cm^{-1} and 1918 cm^{-1} are 4.3 and 9.9 for the CS retrieval. By contrast, the LSSA retrieval fails to differentiate either line from the noise.

The results presented in Figure 6.6 and Figure 6.7 illustrate that, despite under-sampling the interferogram by a factor of $c = 4.0$, the CS SHFTS is able to successfully retrieve sparse spectra even in the presence of background light. In this case, a spectral retrieval corresponds to a retrieval that returns the optical frequencies and line shapes for the sparse features–as compared with the "truth" data reported by the hyperspectral instrument. In the case of the near-monochromatic signal, both CS and LSSA retrievals are successful, with the CS retrieval achieving superior SNR. The absence of background light in the monochromatic case results in a highly overdetermined system of equations in Equation 6.3, allowing LSSA retrieval to succeed despite using an undersampled the interferogram. As expected, in the case of the polychromatic signal, wherein the system is highly underdetermined, only the CS retrieval produces an interpretable spectra. This result demonstrates that background removal via low-resolution LSSA may be deployed in order to enable CS retrieval of sparse components. In the polychromatic retrieval the filter line at 690 cm^{-1} is retrieved with lower SNR than in the monochromatic case; the reduction in SNR is attributed to imperfect background estimation and subtraction as can be seen at lower shifted wavenumbers in Figure 6.7. The second filter line, at 1918 cm^{-1}, has higher peak transmission than the first, and accordingly is retrieved with greater SNR. The SNR of these retrievals, and the presence of "speckle", could be improved by averaging multiple measurements, or by improving the optical efficiency of the SHFTS (*e.g.*, by reducing waveguide propagation loss and further minimizing reflections at MZI Y-splitters).

Note that the application of CS principles to Raman and LIBS spectroscopy is always subject to an assumption of sparsity. For any given spectrum and a given amount of spectral sampling (*i.e.* number of MZIs) there exists a maximum number of non-sparse components that may be retrieved by the instrument. Therefore, it is imperative that a CS spectrometer be used only in instances in which it may be guaranteed that the number of non-sparse components is within the sparsity limit imposed by the retrieval technique and the number of MZIs on the chip [95, 142].

This successful demonstration validates the capabilities of SHFTS as a pure-play CS spectroscopy platform (*e.g.*, a CS FTS which can only be usedd with CS retrieval techniques). CS techniques, including the background estimation scheme demonstrated herein, are well-suited to on-chip SHFTS since: (1) the signal and detection bases are incoherent, (2) arrays of MZIs (i.e. partial Fourier matrices) strongly satisfy the RIP, and (3) unlike Michelson interferometry, on-chip SHFTS may undersample the interferogram without digitally downsampling (*i.e.*, discarding data once it has already been collected).

6.6 Conclusions

In this chapter we have presented the first CS Fourier-transform spectrometer to directly undersample the interferogram without discarding data. We implement this spectrometer as a planar-waveguide spatial heterodyne spectrometer consisting of array of Mach-Zehnder interferometers. Waveguides are coupled to free space via a novel combination of FIB micro-mirrors aligned to a 2-D MLA—resulting in simultaneous, broadband coupling to 83 independent waveguide apertures. The spectrometer achieves an undersampling factor of $c = 4.0$, reducing the number of MZIs by a factor of four over conventional SHFTS devices. We develop a novel white-light characterization scheme for snapshot spectral characterization using hyperspectral imaging of waveguide outputs. We demonstrate a novel compensation scheme to permit CS retrieval even in the presence of broadband (non-sparse) background light, and finally we validate the SHFTS by obtaining retrievals of sparse signals both with and without background light. This technology demonstration, and in particular the demonstration of a background compensation scheme validates the instrument concept and underlying measurement technique, and represents an enabling step towards the development of high-throughput, chip-based miniature spectrometers with expanded capabilities using CS. Building on this work, the objective of which was to de-risk the instrument concept, further steps towards an elegant breadboard will focus on improvements to system étendue in order to maximize SNR. These improvements should be implemented at the waveguide fabrication level and may include increasing the waveguide numerical aperture, reducing insertion loss into the MZI Y-splitters, and increasing the number of input apertures.

7 Summary and conclusions

In this chapter the principal contribution, major accomplishments and results of this dissertation are summarized and discussed along with potential future directions for the research and technologies developed herein.

7.1 Major contributions

In this thesis the design, fabrication, and characterization of two on-chip Fourier-transform spectrometers was presented.

The first device, which forms the basis of the paper presented in Chapter 4, is a narrowband FTIR chip for methane detection. This device is based on an instrument architecture that was pre-existing in the literature, however the device expanded on the previous literature by demonstrating an approach to mitigating the temperature sensitivity of the instrument through passive waveguide athermalization. A novel waveguide geometry, consisting of a silicon nitride core, SiO_2 undercladding and SU-8 over-cladding was designed in order to reduce the effective thermo-optic coefficient of the fundamental waveguide mode. Athermalization was achieved by balancing modal power fractions between the core and cladding materials, with fractional weighting based on the relative strength of the thermo-optic coefficients of each material. The approach to athermalization guided significant waveguide design work, detailed in Section 3.2 including setting minimum bend radius, minimum waveguide separation, as well as 2x2 multi-mode interference couplers, and layout scripts. The temperature sensitivity of this device was investigated as part of the characterization activities presented in Chapter 4, where it was demonstrated that the passive athermalization technique successfully achieved an effective thermo-optic coefficient of $\alpha_{eff} = -3.52 \times 10^{-6} K^{-1}$—two orders of magnitude better than previous devices. Temperature sensitivity of on-chip Fourier-transform devices is a known limitation of the technology; therefore, the successful development of a passively athermal waveguide geometry demonstrates a path towards practical commercial development and deployment of these systems in harsh environments including space.

The first device, presented in Section 4, was also used to demonstrate the retrieval of complex spectra. Prior implementations of on-chip SHFTS had performed only monochromatic retrievals of a single laser line spectrum—a simplistic case in which the spectral retrieval problem is highly overdetermined. In this work, we performed the retrieval of a spectrum including multiple molecular absorption lines from a methane gas cell. Through this investigation, it was revealed that the commonly-implemented "pseudoinverse" technique (which is rightfully termed least-squares spectral analysis) is unsuitable for the retrieval of a critically-sampled complex spectra. In this work, it was determined that this limitation is due to unavoidable imperfections in waveguide fabrication which prevent the formation of an array of MZIs with uniformly increasing OPDs. In the case of a non-uniform array of MZIs, the spectral aliasing problem common to all interferometers is significantly worsened. The practical result is that on-chip FTS devices such as the methane chip presented herein, as well as similar devices in the literature, may achieve less than their full design spectral bandwidth when retrieving polychromatic signals. This is an important result that furthers the existing body of knowledge for the design of on-chip FTS devices; future system designers would be well-served to consider the impact of MZI non-uniformity on their spectral retrievals.

In the second device, powerful techniques from the recently-developed field of compressive sensing are applied to the pre-existing on-chip FTS architecture to create a first-of-its-kind, compressive-sensing FTS. The initial design concept of this device forms the basis for the paper presented in Chapter 5. This paper, as well as the relevant discussions in Section 2.1, present the theoretical motivation for the application of compressive-sensing techniques to on-chip FTS devices. For sparse input spectra—such as monochromatic sources, or multi-species emission lines—compressive sensing techniques may be used to perform accurate spectral retrieval using a small subset of randomly selected MZIs. These retrievals may be achieved despite significant under-sampling of the interferogram—in Chapter 5 undersampling of $c = 0.25$ is achieved. A theoretical investigation of the applicability of these techniques towards complex spectra is carried out using Raman spectra provided by Professor Ed Cloutis at the University of Winnipeg. Through this investigation, a method for enforcing sparsity using a subset of fully-sampled low-frequency MZIs was developed in order to address the applicability of compressive-sensing to practical spectra; this method is summarized in Section 2.1.4.1. This theoretical development, which was captured and published in Optics Letters [132], and patented in [152], is an important contribution to the emerging field of compressive-sensing. The paper details an entirely new architecture for the practical implementation of compressive-sensing spectrometers, which has subsequently been adopted by other researchers in the field of compressive-sensing spectroscopy and integrated photonics [38, 153, 154].

The fabrication and characterization of the first-of-its kind compressive-sensing FTS is presented in Chapter 6. This device incorporates a number of novel techniques, including a broadband multi-aperture input coupling scheme based on vertically-emitting FIB-etched mirrors aligned to a micro-lens array. The on-chip FTS is designed to operate over a broad range of visible wavelengths in order to demonstrate the applicability of the technique towards Raman spectroscopy. The spectral bandwidth of the device drove the development of an ultra-broadband waveguide core, associated design rules, and waveguide components. In the case of the input coupling array, the examination of broadband operation resulted in the development of the all-reflective vertical input coupling array. The successful demonstration of this multi-aperture input coupling technique is presented in Chapter 6. Though this input coupling technique is less practical than commonly-used alternatives (*e.g.*, vertical grating couplers), the technique successfully demonstrated a substantial 3 dB bandwidth of operation.

The characterization of the second device also involved the development of a novel waveguide interrogation technique, summarized in Chapter 6. Conventionally, waveguide devices are characterized using swept-wavelength sources (*e.g.*, tunable lasers). In this experiment, the uniaxial arrangement of waveguide outputs along the chip facet is utilized to perform spectral interrogation of the device using a visible-wavelength hyperspectral imaging system capturing the waveguide response to incoherent broadband excitation. This technique is similar to the use of optical spectrum analyzers, however the use of a hyperspectral system offers considerable advantages over an optical spectrum analyzer in terms of the number of channels that may be simultaneously addressed. The paper presented in Chapter 6 is the first time that such a technique has been proposed or deployed. In the paper the hyperspectral imaging system is used to simultaneously characterize 163 waveguide output channels—a milestone for high-throughput device characterization.

Finally, Chapter 6 demonstrates a compressive sensing retrieval of complex spectra consisting of both narrowband and broadband features despite significant undersampling of the interferogram. These retrievals provide experimental validation of the methods that are developed from a theoretical standpoint in Section 2.1. In the realized device, 160 randomly-selected MZIs are fabricated out of a "complete" set of 633 MZIs, representing the first implementation of a true compressive-sensing Fourier-transform device that directly undersamples the interferogram. Prior demonstrations of compressive-sensing Fourier-transform devices, including the paper presented in Chapter 5, have relied on digital downsampling of an interferogram that is collected in its entirety. By contrast, the fabricated device collects only a partial interferogram at the outset—meaning that it cannot be used as a conventional on-chip FTS. For this reason, the successful demonstration of a polychromatic retrieval using the low-frequency MZIs is an important step towards a practical realization of compressive-sensing Fourier-transform spectroscopy. The successful implementation of this technique, on a platform that directly undersamples the interferogram, therefore validates the architecture of the pure-play compressive-sensing Fourier-transform spectrometer and advances the state of the art through the provision of a novel compressive-sensing spectroscopy platform.

7.2 Future directions

The major contribution of this research presented in Chapters 5 and 6 is in the development of a novel instrument architecture for compressive-sensing spectroscopy. This architecture has since been replicated by other researchers [38, 153, 154], and represents a potentially rich vein of applications. Spectrally-sparse signals that are well-suited to this instrument architecture are commonly found in nature (*e.g.*, atomic emission lines), and in laboratory research (*e.g.*, laser lines).

A potential avenue of future research would be in the applicability of these techniques to precision metrology and sensing applications. For instance, given a stable, calibrated monochromatic input of known frequency (*e.g.*, a laser locked to an atomic transition), a compressive-sensing on-chip FTS could be used as a precision temperature sensor by performing retrievals on a partial output interferogram and evaluating the temperature-dependent "drift" in wavelength versus the known frequency of the source. This effect could be enhanced by inverting the approach taken in Chapter 4 in order to produce highly temperature-sensitive waveguides. Alternately, an appropriately temperature-stabilized waveguide chip could be used to effect a high-precision optical wavemeter.

In the development of on-chip FTS devices for greenhouse gas detection, further attention should be paid to mitigating the problem of non-uniform Fourier-transforms.

This issue could be addressed through the use of on-chip heaters or electro-optic modulators, or by using multiple passes through a common-path delay cavity to ensure that OPDs are uniformly increasing. Some interesting work on this subject has already been pursued by Uda et al. [116], who approximate non-uniformities in waveguide fabrication as a phase error, however further study is needed in order to determine the best method to address these challenges.

7.3 Closing statement

In this dissertation, the design, fabrication, and characterization of on-chip Fourier-transform spectrometers are discussed. The rationale behind an on-chip implementation of Fourier-transform spectroscopy is motivated, and an conventional on-chip Fourier-transform spectrometer—with specific enhancements for durability—is presented. The potential of the on-chip platform to provide a basis for compressive-sensing is discussed and motivated through a theoretical investigation of the system model. This capability is realized experimentally using a custom-built on-chip compressive-sensing Fourier-transform spectrometer. This dissertation, and the papers on which it is based, represent the invention of a unique and novel instrument architecture with rich potential for future applications in precision sensing.

A Spectrometer layout and fabrication

A.0.1 Spectrometer layout

Once the design rules for a waveguide spectrometer (bend radius, waveguide separation, MMI design, etc.) have been established, the layout of the main body of the spectrometer must be constructed in a computer-aided design format (DXF, GDSII, etc.) for delivery to the wafer foundry. The waveguide layout will be converted into a lithographic mask, which will then be used to realize the waveguides on the surface of the wafer. In principle, the layout may be generated by hand—however in practice it is preferable to generate layouts via script.

Layouts for the CS and methane spectrometers were prepared in OptoDesigner as adaptive scripts that may be used to realize any arbitrary spectrometer, provided that a set of design inputs are provided. The overall flow of the scripting algorithm is shown in Figure A.1; the output of this process is the final waveguide layout file, the inputs are summarized in Table A.1.

The individual MZIs may be implemented as one of three distinct types. These

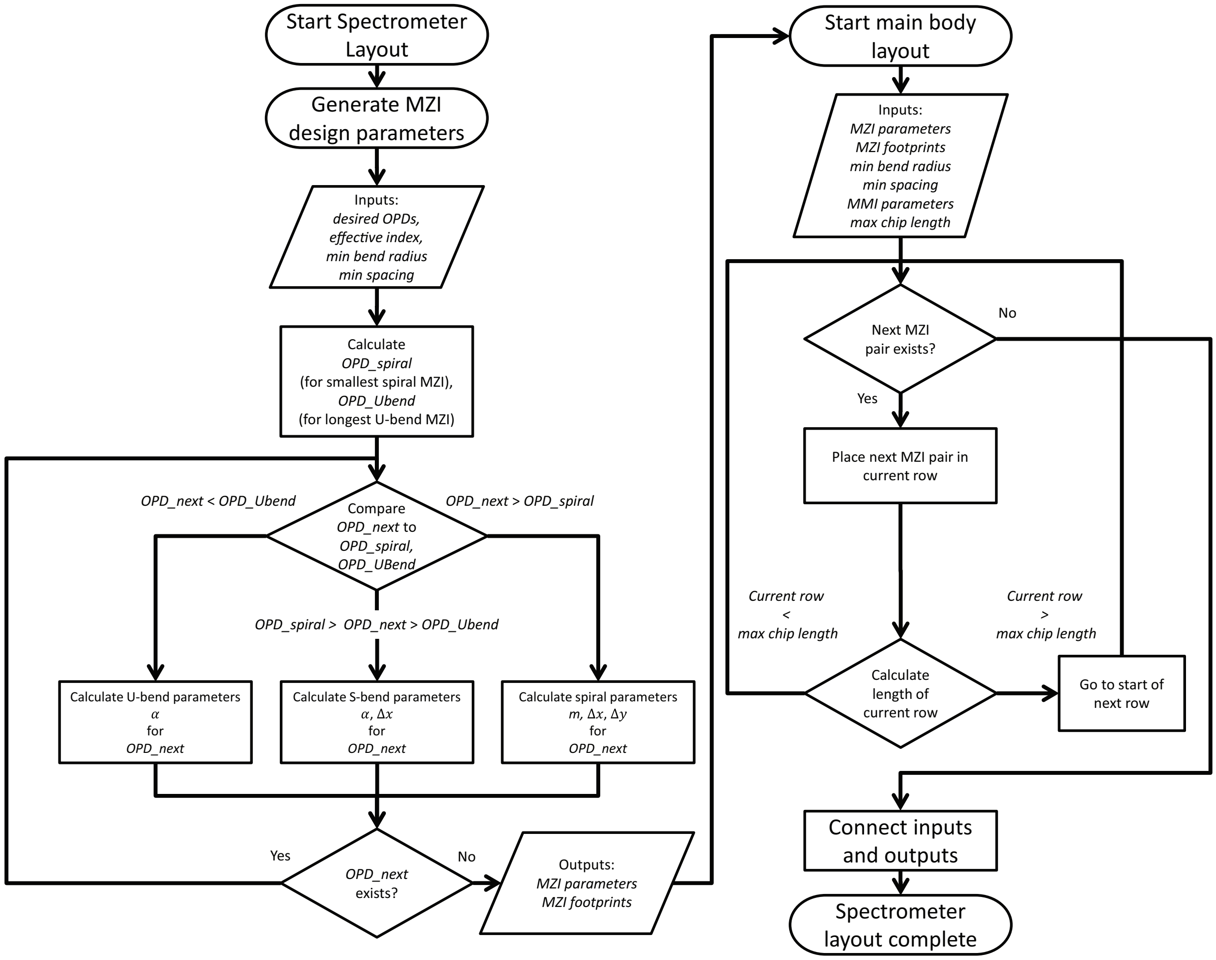

Fig. A.1 Flowchart capturing the design and layout process for the adaptive script developed to produce the SHFTS device. The parameters used to generate the MZI structures are illustrated in Figure A.2

three MZI architectures, U-bend, S-bend, and spiral, are shown in Figure A.2, along with an illustration of their design parameters. In each of these three MZIs, the OPD may be determined analytically as a function of the design parameters, as well as the overall waveguide index, width, radius of curvature, and separation. The first

Table A.1 Variables used to define spectrometer layout

Variable name	**Variable description**
N_{MZI}	Number of MZIs
$[OPD]$	Array of desired OPDs
n_{wg}	Waveguide refractive index
W_{wg}	Waveguide width
R_{min}	Minimum bend radius
W_{ws}	Waveguide separation
L_{MMI}	MMI length
W_{MMI}	MMI width
L_{taper}	MMI taper length
W_{access}	MMI access waveguide width
W_{gap}	MMI access waveguide separation
L_{Block}	Desired length of central MZI block

step in the waveguide layout process, which is carried out in MATLAB, is to map the waveguide refractive index, width, bend radius, separation, and desired OPDs to MZI layout parameters by solving one of three analytical functions. First, the desired OPD is used to select one of the three MZI architectures. The desired OPD is compared against the maximum OPD supported by a U-bend, which is the OPD given by a U-bend with 90° turns. If the desired OPD is greater than this value, then it may be compared against the maximum OPD that should be supported by an S-bend, which—for compactness—is the minimum OPD supported by a spiral. Evaluating the OPD of each MZI against these threshold values thus defines which designs are used for which MZI in the array.

Once a particular MZI design has been selected, the OPD equation must be solved for the correct design parameters. This may be straightforwardly implemented as an

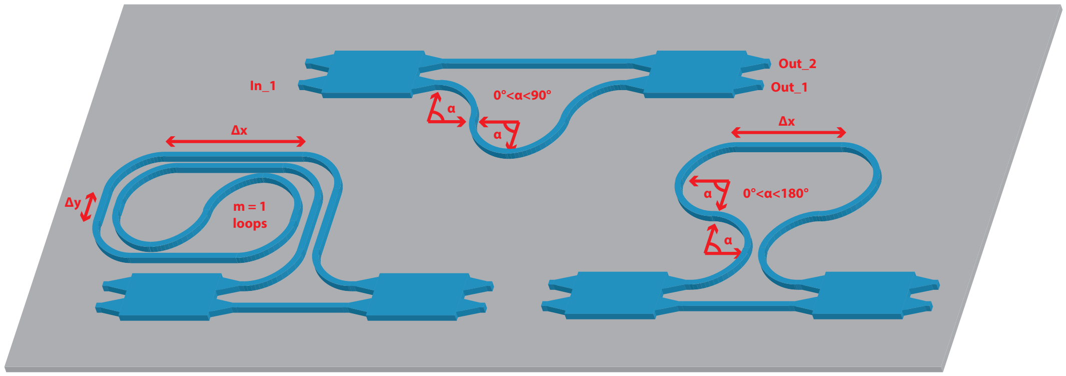

Fig. A.2 The three MZI types used in the SHFTS design are shown along with their design parameters. U-bend MZIs (top) are used to achieve the shortest OPDs and are controlled only by the angle of the S-bend, α. S-bend MZIs (bottom right) are used to achieve OPDs greater than the longest U-bend but shorter than the shortest spiral; the OPD is controlled by the angle of the S-bend, α, and the length of the offset, Δx. Spiral MZIs (bottom left) are used to achieve the longest OPDs; their length is controlled by the number of loops, m, and the horizontal offset Δx.

optimization problem using the bisection algorithm. In the case of a U-bend MZI, there exists a unique bend angle, θ_U, between 0° and 90° such that $F_U(\theta_U) = OPD$, where $F_U(\theta_U$ is a function relating the bend angle tot he OPD for a U-bend MZI. The bisection algorithm is then implemented according to the description in this section.

In the case of an S-bend MZI a similar procedure is followed, with an additional decision point in place to determine whether the bisection algorithm should be applied

Algorithm 1: Bisection method as used to solve for angle of U-bend MZI

input : $F(\theta_U)$, OPD

output: Values of θ_U s.t. $F(\theta_U) = OPD$

$\theta_0 = 0° \; OPD_0 = F(\theta_0)$;

$\theta_1 = 90°$;

$OPD_1 = F(\theta_1)$;

$\mu = \theta_1 - \theta_0$;

while $|\mu| \geq 10^{-6}$ **do**

  $\theta_2 = 0.5 * (\theta_0 + \theta_1)$;

  if $sign(F(\theta_2)) == sign(F(\theta_0))$ **then**

    $\theta_0 = \theta_2$

  else

    $\theta_1 = \theta_2$

  end

  $\mu = \theta_1 - \theta_0$

end

return θ_1

to the linear separation at the top of the S-bend, or to the angle. If $G(\theta_S, \Delta x)$ is a function mapping the S-bend angle, θ_S, and the linear separation, Δx, to the MZI OPD then the MZI may be defined by solving for θ_S provided that the OPD is bounded such that $G(90°, 0) < OPD \wedge G(180°, 0) > OPD$. If the MZI is not bounded by the

S-bend angle alone, then the MZI parameters are defined by fixing $\theta_S = 180°$, and solving for Δx within the range $0 < \Delta x < OPD$.

In the case of a spiral MZI the variable parameters are the number of "turns" in the spiral, m, and the linear separation Δx. Given a function mapping the number of turns and the linear separation to the OPD, $H(n, \Delta x)$, the parameters of the squared spiral may be solved following the steps in Algorithm 2:

Once the individual MZI parameters have been determined, the spectrometer layout is performed in OptoDesigner. The MZIs are arrayed in "daisy-chains" of pairs, in linearly increasing OPD. Arranging MZIs in linearly increasing pairs ensures a compact layout by minimizing the amount of excess white space required to maintain minimum waveguide separation margins between the larger and smaller MZIs. The layout script proceeds placing the first such daisy-chain at the origin and building out sequentially from there. Subsequent daisy-chains are placed next to the original chain, with the minimum waveguide separation distance defining the vertical offset. Daisy chains continue to nest in this fashion until the chain reaches sufficient length that it must be restarted to preserve the horizontal chip footprint. Throughout this process the length of the daisy chain is monitored by recording the length of each pair of MZIs in the chain; this may be accomplished by recording the dimensions of the "bounding box" (*i.e.* reserved footprint) of each MZI pair. The dimensions of each bounding box may be determined uniquely from the design parameters and the type

Algorithm 2: Algorithm used to find design parameters of squared spiral. The line $\Delta x_2 = BISECT(H(m, \Delta_x), OPD, \Delta x_0, \Delta x_1)$ indicates that the algorithm calls a bisection routine to find a solution to $H(m, \Delta x) = OPD$ on the interval $\Delta x_0 < \Delta x_2 < \Delta x_1$.

input : $H(m, \Delta x)$, OPD

output: Values of m, Δx s.t. $H(m, \Delta x) = OPD$

$m_0 = 0$;

$\Delta x_0 = 0$;

$m_1 = 1$;

$\Delta x_1 = 0$;

$\theta_1 = 90°$;

$OPD_1 = F(\theta_1)$;

$\mu = \theta_1 - \theta_0$;

while $H(m_1, \Delta x_1 < OPD$ **do**

  $m_0 = m_0 + 1$;

  $m_1 = m_1 + 1$;

end

$\Delta x_1 = OPD$;

$\Delta x_2 = BISECT(H(m, \Delta_x), OPD, \Delta x_0, \Delta x_1)$;

return $m_0, \Delta x_2$

of MZI, and are used both to monitor the length of each chain and to determine the appropriate vertical offset between the final MZI pair in one chain and the start of the next. These arrangements are shown in Figure A.3.

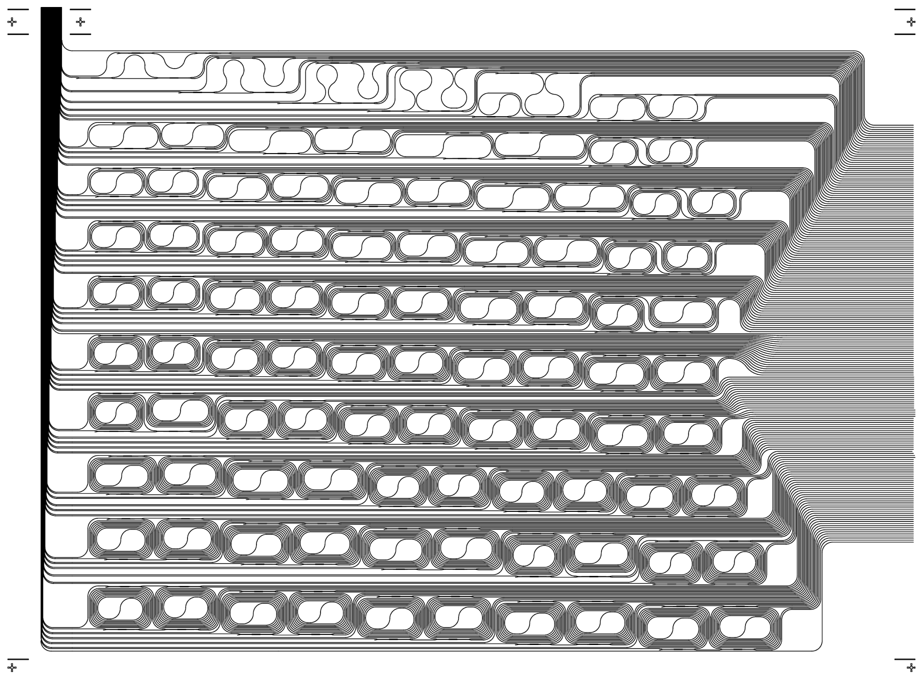

Fig. A.3 Layout of athermal methane spectrometer produced using adaptive scripting routine. MZIs are implemented in inverted pairs, which are joined together to form "daisy chains" of MZIs; these chains are arranged in parallel resulting in a dense arrangement of MZIs.

Once the central block of daisy chains has been assembled, the outputs of each of the MZIs are collected together in a bundle and fanned-out to the edge of the chip—where their outputs may be captured by a linear detector array, or imaged by a camera. If a linear detector array is used to collect the output signals then the pitch of the waveguide outputs on the chip should be set as an integer multiple of the pixel pitch. If the edge of the chip is to be imaged using a 2-D detector, then the waveguide pitch should be selected to ensure that the full array can be imaged by the selected detector, while maintaining at least one-pixel separation between adjacent waveguides. The size or resolution of the 2-D detector, as well as the numerical aperture (NA) of the waveguides, may therefore drive some tradeoffs in waveguide pitch at the output. For instance, using a high-magnification lens in order to ensure a large number of dark pixels between waveguide outputs may prevent simultaneous imaging of the full array of waveguides, or may preclude the use of a high-NA imaging system (preventing the imaging system from capturing all of the light emitted by each waveguide.

A.0.2 Waveguide fabrication

Waveguides for both the methane and CS spectrometer were fabricated by LioniX International. In the case of the CS spectrometer both the silicon nitride waveguide layer, and the silicon dioxide upper cladding, were deposited, via plasma-enhanced chemical vapour deposition (PECVD), directly onto the fused-silica wafer which formed the substrate and lower cladding. The waveguide layer was patterned by contact lithography, and the core itself was formed by RIE. The wafers were diced into chips by LioniX, requiring no further processing steps.

In the case of the methane spectrometer, the silicon nitride layers were deposited via PECVD onto a silicon dioxide lower-cladding layer formed by thermal oxidization of the underlying silicon substrate. The silicon nitride layers were then patterned by i-line lithography, and the waveguide cross-section formed by RIE. Unlike the CS spectrometer, some further post-processing of the wafers was performed including deposition of the SU-8 upper cladding and dicing of the wafer to prepare the waveguide facets.

To realize the athermal waveguide cross-section shown in Figure 4.2, a process for deposition of the SU-8 layer onto the un-clad methane spectrometer was developed and employed to coat each of the wafers. The process is as follows:

1. The wafer surface is cleaned and made hydrophilic by a 30 second reactive-ion oxygen etch.

2. The wafer is placed on a vacuum chuck in the spin-coating station, and SU-8 is deposited onto the surface of the wafer. The wafer is spun at 2000 RPM for 40 seconds to evenly coat the surface of the wafer with an $\approx 7\ \mu m$ thick layer of SU-8.

3. The wafer and SU-8 coating are baked on a hot plate for a period of 4 minutes at $95^\circ\ C$ in order to harden the SU-8 and improve adhesion to the wafer.

4. The SU-8 resist is developed by exposure to UV light for a period of 3 seconds. UV exposure in this manner promotes cross-linking of the SU-8, further improving adhesion to the substrate.

5. Finally, the wafer undergoes a post-bake at 95° C for a further 4 minutes in order to harden the SU-8.

Following deposition of the SU-8 upper cladding, the wafers were separated into individual chips by dicing. To ensure efficient coupling to waveguide facets, care must be taken during the dicing process to ensure that the cladding is not damaged and that surface roughness is minimized. Damage to the cladding near input and output waveguides substantially reduces input coupling efficiency due to reflections at the waveguide-cladding interface, and may even result in broken or cracked waveguides. Examples of damage to the waveguide facets obtained by visual microscopy are shown in Figure A.4. Likewise, increased surface roughness at the waveguide facet may result in coupling losses due to scattering. Finally, significant surface roughness resulting in a partially-angled waveguide facets may affect the co-alignment of waveguide input acceptance “cones”.

Typically, surface roughness of diced chips may be reduced by chemical mechanical polishing (CMP) of the edge-facets of diced chips. Alternatively, a higher-quality facet may be prepared by cleaving the substrate directly, thereby taking advantage

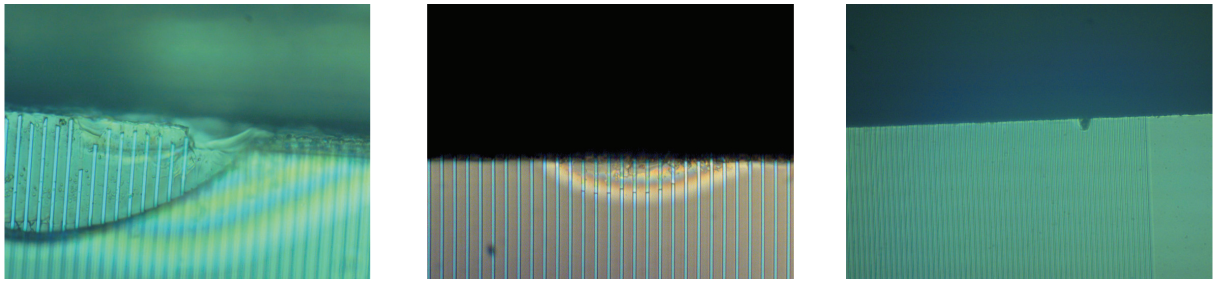

Fig. A.4 Microscope images of three varieties of damaged 5-μm spaced waveguide facets obtained from wafer dicing. First, the dicing saw may produce complete liftoff of the SU-8 cladding layer (left), resulting in damage to the waveguides, and excessive surface roughness. Second, the dicing saw may produce partial lifting and stressing of the SU-8 layer (center), resulting in visible Newton's rings and cracking of waveguides. Finally, the dicing saw may "catch" on the wafer itself resulting in pitting of the facet surface (right).

of the underlying crystal structure of silicon to produce an extremely smooth facet. Compared to dicing and polishing, cleaving is a low-yield process since it cannot be assured that the wafer will cleave directly along the desired plane (*i.e.*, the wafer may cleave through the center of a chip). Wafer cleaving may be preferred, therefore, in applications where high-yield is unnecessary, but coupling efficiency is of critical importance.

Unfortunately, neither wafer cleaving nor polishing were suitable processes for

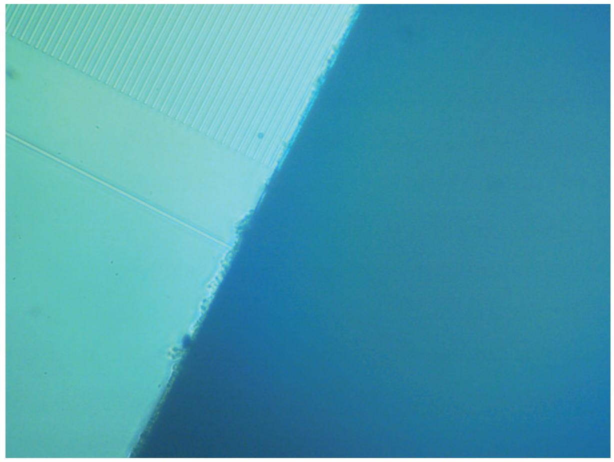

Fig. A.5 Microscope image of cleaved wafer facet, showing surface roughness of SU-8 polymer, and recession of polymer from facet edge.

the preparation of waveguide facets with the SU-8 upper cladding. Since SU-8 is a polymer, it does not repeatably cleave along a single plane, resulting in chipping or "ripping" of the SU-8. An example of an SU-8 facet prepared by cleaving is shown in Figure A.5, where it may be seen that edge of the SU-8 layer is recessed from the edge of the substrate and is pitted due to tearing of the SU-8.

In the case of dicing and polishing, while it is possible to dice chips clad with SU-8, it is not possible to polish their facets using CMP. The CMP process involves a fast-spinning chuck that grinds the surface of the chip using a slurry containing very fine grit. When applied to a polymer surface, grit from the slurry may rip the surface of the or may even become embedded in the polymer. CMP, therefore, has the opposite of its desired effect on a polymer surface—increasing surface roughness

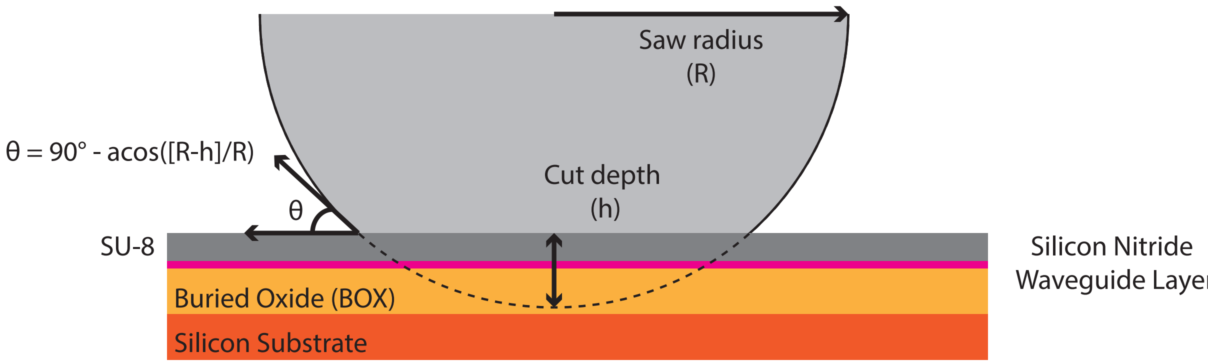

Fig. A.6 Diagram of wafer dicing geometry indicating the relationship betwene the cut depth and the angle of the blade at the wafer surface. By reducing the depth of the cut, the angle of attack of the blade may be reduced with respect to the SU-8 polymer layer.

and causing damage to the cladding.

The dicing process itself may result in polymer lift-off due to kickback at the trailing edge of the dicing saw. This process is shown diagrammatically in Figure A.6, and an example of polymer lift-off from a single-pass dicing run is shown in Figure A.4. To prepare a suitable waveguide facet, a double-pass dicing technique was developed in which the chips are first polished by a shallow cut of the dicing saw, and then diced by a second, deeper cut that is horizontally offset from the first. During this process, a fine-grained dicing saw is used in order to minimize the roughness of the blade surface in contact with the chip edge. The advantage of this two-step process is that, in the first pass of the blade, the angle of the trailing edge with respect to the wafer surface is substantially reduced. The smaller angle of attack reduces the upward friction force on the polymer layer—resulting in fewer instances of polymer lift-off. A microscope image of surfaces prepared by the double-pass technique, is shown in Figure A.7. It is qualitatively apparent, that the double-pass technique results in reduced "pitting" of the polymer and fewer instances of lift-off; this was also confirmed quantitatively by greater yield of chips compared with the single-pass technique.

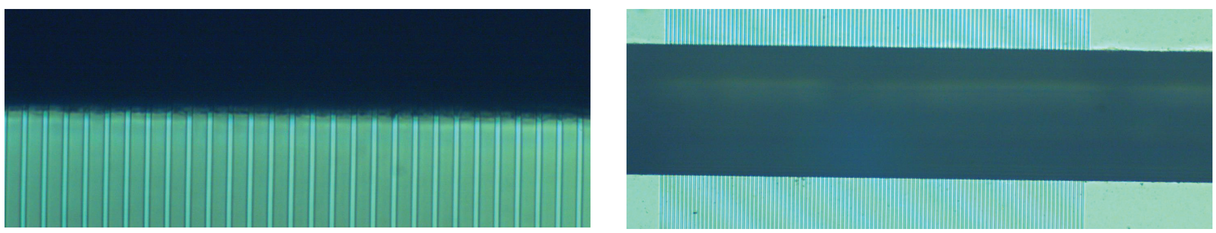

Fig. A.7 Micrograph images showing improved surface roughness and lack of SU-8 liftoff following double-pass dicing of wafers. A close up image (left) indicates the improved surface roughness, while the wide-field image (right) shows the absence of damage over the full input coupling bundle.

Bibliography

[1] Eisenhauer, F. and Raab, W., "Visible/Infrared Imaging Spectroscopy and Energy-Resolving Detectors," *Annual Review of Astronomy and Astrophysics*, Vol. 53, No. 1, 2015, pp. 155–197.

[2] Band, Y. B., *Light and Matter: Electromagnetism, Optics, Spectroscopy and Lasers*, Light and Matter, John Wiley & Sons, 2006.

[3] Stuart, B. H., *Infrared Spectroscopy of Biological Applications: An Overview*, American Cancer Society, 2012.

[4] Haas, J. and Mizaikoff, B., "Advances in Mid-Infrared Spectroscopy for Chemical Analysis," *Annual Review of Analytical Chemistry*, Vol. 9, No. 1, 2016, pp. 45–68.

[5] Tsuda, H. and Arends, J., "Raman Spectroscopy in Dental Research: A Short Review of Recent Studies," *Advances in Dental Research*, Vol. 11, No. 4, 1997, pp. 539–547.

[6] Hawthorne, F. C., *Spectroscopic methods in mineralogy and geology*, Vol. 18, Walter de Gruyter GmbH & Co KG, 2018.

[7] Christensen, D., Rüther, A., Kochan, K., Pérez-Guaita, D., and Wood, B., "Whole-Organism Analysis by Vibrational Spectroscopy," *Annual Review of Analytical Chemistry*, Vol. 12, No. 1, 2019, pp. 89–108.

[8] Bacon, C. P., Mattley, Y., and DeFrece, R., "Miniature spectroscopic instrumentation: Applications to biology and chemistry," *Review of Scientific Instruments*, Vol. 75, No. 1, 2004, pp. 1–16.

[9] Blacksberg, J., Alerstam, E., Maruyama, Y., Cochrane, C. J., and Rossman, G. R., "Miniaturized time-resolved Raman spectrometer for planetary science based on a fast single photon avalanche diode detector array," *Applied Optics*, Vol. 55, No. 4, feb 2016, pp. 739–748.

[10] Kelley, D. H., Aveni, A. F., and Milone, E. F., *Exploring Ancient Skies: A Survey of Ancient and Cultural Astronomy*, Springer New York, 2011.

[11] Marcy, G. and Butler, R., "Extrasolar Planets Detected by the Doppler Technique (invited review)," *Brown Dwarfs and Extrasolar Planets*, edited by R. Rebolo, E. L. Martin, and M. R. Zapatero Osorio, Vol. 134 of *Astronomical Society of the Pacific Conference Series*, jan 1998, p. 128.

[12] Sobral, D., Matthee, J., Darvish, B., Schaerer, D., Mobasher, B., Röttgering, H., Santos, S., and Hemmati, S., "Evidence for PopIII-like Stellar Populations in the Most Luminous Lyman-$\{\alpha\}$ Emitters at the Epoch of Reionization: Spectroscopic Confirmation," *ApJ*, Vol. 808, aug 2015, pp. 139.

[13] Roth, L., Saur, J., Retherford, K. D., Strobel, D. F., Feldman, P. D., McGrath, M. A., and Nimmo, F., "Transient Water Vapor at Europa's South Pole," *Science*, Vol. 343, No. 6167, 2014, pp. 171–174.

[14] Hubble, E., "A relation between distance and radial velocity among extra-galactic nebulae," *Proceedings of the National Academy of Sciences*, Vol. 15, No. 3, 1929, pp. 168–173.

[15] Crisp, D., Pollock, R., Rosenberg, R., Chapsky, L., Lee, R., Oyafuso, F., Frankenberg, C., O'Dell, C., Bruegge, C., Doran, G., Eldering, A., Fisher, B., Fu, D., Gunson, M., Mandrake, L., Osterman, G., Schwandner, F., Sun, K., Taylor, T., and Wunch, D., "The On-Orbit Performance of the Orbiting Carbon Observatory-2 (OCO-2) Instrument and its Radiometrically Calibrated Products," *Atmospheric Measurement Techniques Discussions*, 2016, pp. 1–45.

[16] Balmer, J. J., *Notiz über die spektrallinien des wasserstoffs*, J.A. Barth, 1885.

[17] Diem, M., Romeo, M., Boydston-White, S., Miljkovic, M., and Matthäus, C., "A decade of vibrational micro-spectroscopy of human cells and tissue (1994–2004)," *Analyst*, Vol. 129, 2004, pp. 880–885.

[18] Osorio, D. and Vorobyev, M., "A review of the evolution of animal colour vision and visual communication signals," *Vision Research*, Vol. 48, No. 20, 2008, pp. 2042–2051.

[19] Bennett, A. T. D., Cuthill, I. C., and Norris, K. J., "Sexual Selection and the Mismeasure of Color," *The American Naturalist*, Vol. 144, No. 5, 1994, pp. 848–860.

[20] Transon, J., D'Andrimont, R., Maugnard, A., and Defourny, P., "Survey of current hyperspectral Earth observation applications from space and synergies with Sentinel-2," *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, jun 2017, pp. 1–8.

[21] Chen, M. W., Schulz, M., Anderson, P. C., Lu, G., Germany, G., and West, M., "Storm time distributions of diffuse auroral electron energy and X-ray flux: Comparison of drift-loss simulations with observations," *Journal of Geophysical Research: Space Physics*, Vol. 110, No. A3, 2005.

[22] Blondeau-Patissier, D., Gower, J. F., Dekker, A. G., Phinn, S. R., and Brando, V. E., "A review of ocean color remote sensing methods and statistical tech-

niques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans," *Progress in Oceanography*, Vol. 123, 2014, pp. 123 – 144.

[23] Odi-Lara, M., Campos, I., Neale, C. M. U., Ortega-Faras, S., Poblete-Echeverra, C., Balbontn, C., and Calera, A., "Estimating Evapotranspiration of an Apple Orchard Using a Remote Sensing-Based Soil Water Balance," *Remote Sensing*, Vol. 8, No. 3, 2016.

[24] Gitelson, A. A. and Merzlyak, M. N., "Remote sensing of chlorophyll concentration in higher plant leaves," *Advances in Space Research*, Vol. 22, No. 5, 1998, pp. 689 – 692, Synergistic Use of Multisensor Data for Land Processes.

[25] White, J., Wulder, M., and Grills, D., "Detecting and mapping mountain pine beetle red-attack damage with SPOT5 10- m multispectral imagery," *BC Journal of Ecosystems and Management*, Vol. 7, 07 2006.

[26] Bernath, P. and Fernando, A. M., "Trends in stratospheric HCl from the ACE satellite mission," *Journal of Quantitative Spectroscopy and Radiative Transfer*, Vol. 217, May 2018, pp. 126–129.

[27] Varon, D. J., Jacob, D. J., McKeever, J., Jervis, D., Durak, B. O., Xia, Y., and Huang, Y., "Quantifying methane point sources from fine-scale (GHGSat) satel-

lite observations of atmospheric methane plumes," *Atmospheric Measurement Techniques Discussions*, 2018, pp. 1–25.

[28] Pardo, J. R., Encrenaz, P. J., and Breton, D., "Utilization of the Radiofrequency Spectrum above 1 GHz by Passive Services." *Preserving the Astronomical Sky*, edited by R. J. Cohen and W. T. Sullivan, Vol. 196 of *IAU Symposium*, Jan 2001, p. 255.

[29] Turner, E. C., Withington, S., Newnham, D. A., Wadhams, P., Jones, A. E., and Clancy, R., "Simulation of submillimetre atmospheric spectra for characterising potential ground-based remote sensing observations," *Atmospheric Measurement Techniques*, Vol. 9, No. 11, 2016, pp. 5461–5485.

[30] Swartwout, M. A., "Reliving 24 Years in the next 12 Minutes: A Statistical and Personal History of University-Class Satellites," *32nd AIAA/USU Conference on Small Satellites*, August 2018, pp. SSC18–WKVIII–3.

[31] Scott, A. D., Rowlands, N., and Bell, A., "Miniature spectrometers for planetary remote sensing," *Proceedings of SPIE*, Vol. 5660, 2004, pp. 78–87.

[32] Zheng, Y., Wu, Z., Shum, P., Xu, Z., Humbert, G., Zhang, H., Zeng, S., and Dinh, X.-Q., "Sensing and lasing applications of whispering gallery mode microresonators," *Opto-Electronics Review*, Vol. 1, 2018, pp. 1–15.

[33] Doerr, C., Chen, L., Vermeulen, D., Nielsen, T., Azemati, S., Stulz, S., McBrien, G., Xu, X.-M., Mikkelsen, B., Givehchi, M., Rasmussen, C., and Park, S.-Y., "Single-Chip Silicon Photonics 100-Gb/s Coherent Transceiver," *Optical Fiber Communication Conference: Postdeadline Papers*, Optical Society of America, 2014, p. Th5C.1.

[34] Ciminelli, C., Dell'Olio, F., Campanella, C. E., and Armenise, M. N., "Photonic technologies for angular velocity sensing," *Advances in Optics and Photonics*, Vol. 2, No. 3, 2010, pp. 370.

[35] Guo, X., Zou, C.-l., Schuck, C., Jung, H., Cheng, R., and Tang, H. X., "Parametric down-conversion photon-pair source on a nanophotonic chip," *Light: Science & Applications*, Vol. 6, No. 5, 2017, pp. e16249–e16249.

[36] Poulton, C. V., Yaacobi, A., Cole, D. B., Byrd, M. J., Raval, M., Vermeulen, D., and Watts, M. R., "Coherent solid-state LIDAR with silicon photonic optical phased arrays," *Opt. Lett.*, Vol. 42, No. 20, oct 2017, pp. 4091–4094.

[37] Cheben, P., Schmid, J. H., Delâge, A., Densmore, A., Janz, S., Lamontagne, B., Lapointe, J., Post, E., Waldron, P., and Xu, D.-X., "A high-resolution silicon-on-insulator arrayed waveguide grating microspectrometer with sub-micrometer aperture waveguides," *Opt. Express*, Vol. 15, No. 5, mar 2007, pp. 2299–2306.

[38] Souza, M. C. M. M., Grieco, A., Frateschi, N. C., and Fainman, Y., "Fourier transform spectrometer on silicon with thermo-optic non-linearity and dispersion correction," *Nature Communications*, Vol. 9, No. 1, 2018, pp. 665.

[39] Wang, R., Vasiliev, A., Muneeb, M., Malik, A., Sprengel, S., Böhm, G., Amann, M.-C., Simonyte, I., Vizbaras, A., Vizbaras, K., Baets, R., and Roelkens, G., "IIIV-on-Silicon Photonic Integrated Circuits for Spectroscopic Sensing in the 24 m Wavelength Range," *Sensors*, 2017.

[40] Nishi, H., Kim, J., Fukuda, H., Song, H., Tsuchizawa, T., and Yamada, K., "Si photonic integrated circuit for a compact and stable CW-THz spectrometer," *11th International Conference on Group IV Photonics (GFP)*, Aug 2014, pp. 223–224.

[41] Roelkens, G., Wang, R., Vasiliev, A., Radosavljevic, S., Pavanello, F., Malik, A., Muneeb, M., Baets, R., Sprengel, S., Boehm, G., Amann, M.-C., Šimonytė, I., Vizbaras, A., and Vizbaras, K., "III-V-on-silicon Photonic Integrated Circuits for Spectroscopic Sensing in the Mid-Infrared," *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*, Optical Society of America, 2018, p. IM4B.1.

[42] Gatkine, P., Veilleux, S., Hu, Y., Bland-Hawthorn, J., and Dagenais, M., "Arrayed waveguide grating spectrometers for astronomical applications: new re-

sults," *Opt. Express*, Vol. 25, No. 15, Jul 2017, pp. 17918–17935.

[43] Manzardo, O., Herzig, H. P., Marxer, C. R., and de Rooij, N. F., "Miniaturized time-scanning Fourier transform spectrometer based on silicon technology." *Optics letters*, Vol. 24 23, 1999, pp. 1705–1707.

[44] Skauli, T., Torkildsen, H. E., Nicolas, S., Opsahl, T., Haavardsholm, T., Kåsen, I., and Rognmo, A., "Compact camera for multispectral and conventional imaging based on patterned filters," *Appl. Opt.*, Vol. 53, No. 13, May 2014, pp. C64–C71.

[45] Bao, J. and Bawendi, M. G., "A colloidal quantum dot spectrometer," *Nature*, Vol. 523, 2015, pp. 67–70.

[46] McLean, I. S., Steidel, C. C., Epps, H. W., Konidaris, N., Matthews, K. Y., Adkins, S., Aliado, T., Brims, G., Canfield, J. M., Cromer, J. L., Fucik, J., Kulas, K., Mace, G., Magnone, K., Rodriguez, H., Rudie, G., Trainor, R., Wang, E., Weber, B., and Weiss, J., "MOSFIRE, the multi-object spectrometer for infra-red exploration at the Keck Observatory," *Ground-based and Airborne Instrumentation for Astronomy IV*, edited by I. S. McLean, S. K. Ramsay, and H. Takami, Vol. 8446, International Society for Optics and Photonics, SPIE, 2012, pp. 149 – 163.

[47] Angel, S. M., Gomer, N. R., Sharma, S. K., and McKay, C., "Remote Raman Spectroscopy for Planetary Exploration: A Review," *Applied Spectroscopy*, Vol. 66, No. 2, 2012, pp. 137–150.

[48] Eshelman, E., Daly, M. G., Slater, G., and Cloutis, E., "Time-resolved detection of aromatic compounds on planetary surfaces by ultraviolet laser induced fluorescence and Raman spectroscopy," *Planetary and Space Science*, Vol. 119, No. Supplement C, 2015, pp. 200–207.

[49] Crocombe, R. A., "Portable Spectroscopy," *Applied Spectroscopy*, Vol. 72, No. 12, 2018, pp. 1701–1751.

[50] Gares, K. L., Hufziger, K. T., Bykov, S. V., and Asher, S. A., "Review of explosive detection methodologies and the emergence of standoff deep UV resonance Raman," *Journal of Raman Spectroscopy*, Vol. 47, No. 1, 2016, pp. 124–141.

[51] Bacon, C. P., Mattley, Y., and Defrece, R., "Miniature spectroscopic instrumentation : Applications to biology and chemistry," *Review of Scientific instruments*, Vol. 1, No. 2004, 2012.

[52] Naranjo, E., Baliga, S., and Bernascolle, P., "IR gas imaging in an industrial setting," *Thermosense XXXII*, edited by R. B. Dinwiddie and M. Safai, Vol. 7661, International Society for Optics and Photonics, SPIE, 2010, pp. 160 – 167.

[53] Sandsten, J., Edner, H., and Svanberg, S., "Gas imaging by infrared gas-correlation spectrometry," *Opt. Lett.*, Vol. 21, No. 23, Dec 1996, pp. 1945–1947.

[54] Borotto, E., Englender, J., Pourny, J.-C., Naveau, S., Chaput, J.-C., and Lecarpentier, Y., "Detection of the fluorescence of GI vessels in rats using a CCD camera or a near-infrared video endoscope," *Gastrointestinal Endoscopy*, Vol. 50, No. 5, 1999, pp. 684 – 688, 2000 ASGE Meeting.

[55] Mansfield, J. R., Attas, M., Majzels, C., Cloutis, E., Collins, C., and Mantsch, H. H., "Near infrared spectroscopic reflectance imaging: a new tool in art conservation," *Vibrational Spectroscopy*, Vol. 28, No. 1, 2002, pp. 59 – 66, A Collection of Papers Presented at SHEDDING NEW LIGHT ON DISEASE: Optical Diagnostics for the New Millennium, Winnipeg, Canada, June 25-30, 2000.

[56] Qin, J., Chao, K., Kim, M. S., Lu, R., and Burks, T. F., "Hyperspectral and multispectral imaging for evaluating food safety and quality," *Journal of Food Engineering*, Vol. 118, No. 2, 2013, pp. 157 – 171.

[57] *Data processing pipeline for a time-sampled imaging Fourier transform spectrometer*, Vol. 5546, 2004.

[58] Tsao, Y.-C., Chang, P.-E., Chen, S.-Y., and Hwang, Y.-H., "Real-Time Fab-Wise Airborne Molecular Contaminant (AMC) Monitoring System Using Mul-

tiple Fourier Transform Infrared (FTIR) Spectrometers in a Semiconductor Plant," *Aerosol and Air Quality Research*, Vol. 15, 2015, pp. 1640–1651.

[59] Spuler, S. M., Repasky, K. S., Morley, B., Moen, D., Hayman, M., and Nehrir, A. R., "Field-deployable diode-laser-based differential absorption lidar (DIAL) for profiling water vapor," *Atmospheric Measurement Techniques*, Vol. 8, No. 3, 2015, pp. 1073–1087.

[60] Werle, P., Mücke, R., and Slemr, F., "The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (TDLAS)," *Applied Physics B*, Vol. 57, No. 2, Aug 1993, pp. 131–139.

[61] Bjorlin, E. S., Kimura, T., and Bowers, J. E., "Carrier-confined vertical-cavity semiconductor optical amplifiers for higher gain and efficiency," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 9, No. 5, 2003, pp. 1374–1385.

[62] Harrington, J., *Infrared Fibers and Their Applications*, SPIE Press, 2004.

[63] National Research Council, *Seeing Photons: Progress and Limits of Visible and Infrared Sensor Arrays*, 2010.

[64] Dereniak, E. L. and Boreman, G. D., *Infrared detectors and systems*, Wiley series in pure and applied optics, Wiley, 1996.

[65] Agrawal, G. P., *Fiber-Optic Communication Systems*, Wiley Series in Microwave and Optical Engineering, Wiley, 2012.

[66] Andersen, U. L., Gehring, T., Marquardt, C., and Leuchs, G., "30 years of squeezed light generation," *Physica Scripta*, Vol. 91, No. 5, apr 2016, pp. 53001.

[67] Williamson, S. and Mourou, G. A., "Picosecond Electro-Electron Optic Oscilloscope," *Picosecond Electronics and Optoelectronics*, edited by G. A. Mourou, D. M. Bloom, and C.-H. Lee, Springer Berlin Heidelberg, Berlin, Heidelberg, 1985, pp. 58–61.

[68] Willetts, D. V., Freedman, P. A., and Jones, W. J., "A Czerny-Turner spectrograph incorporating a 2-stage image intensifier for the study of very weak Raman spectra," *Journal of Raman Spectroscopy*, Vol. 2, No. 3, 1974, pp. 249–256.

[69] Czerny, M. and Turner, A. F., "Über den Astigmatismus bei Spiegelspektrometern," *Zeitschrift für Physik*, Vol. 61, No. 11, Nov 1930, pp. 792–797.

[70] Rayleigh, L., "Investigations in optics, with special reference to the spectroscope," *Mon. Notices Royal Astron. Soc*, Vol. 40, feb 1880, pp. 254.

[71] Airy, G., "On the Diffraction of an Object-glass with Circular Aperture," *Transactions of the Cambridge Philosophical Society*, Vol. 5, 1835, pp. 283.

[72] Coffelt, E. L. and Martella, M. A., "Optical filters on board the Space Telescope Imaging Spectrograph (STIS)," *Current Developments in Optical Design and Engineering VI*, edited by R. E. Fischer and W. J. Smith, Vol. 2863, Nov. 1996, pp. 324–336.

[73] Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J.-M., Gamache, R. R., Harrison, J. J., Hartmann, J.-M., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Roy, R. J. L., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenko, S., Müller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V., and Wagner, G., "The HITRAN2012 molecular spectroscopic database," *Journal of Quantitative Spectroscopy and Radiative Transfer*, Vol. 130, 2013, pp. 4–50.

[74] Sebacher, D. I., Aeronautics, U. S. N., Scientific, S. A., and Office, T. I., *A gas filter correlation monitor for CO, CH4, and HCl*, NASA technical paper, National Aeronautics and Space Administration, Scientific and Technical Information Office, 1977.

[75] Kostiuk, T. and Mumma, M. J., “Remote sensing by IR heterodyne spectroscopy,” *Appl. Opt.*, Vol. 22, No. 17, sep 1983, pp. 2644–2654.

[76] Jacquinot, P., “The Luminosity of Spectrometers with Prisms, Gratings, or Fabry-Perot Etalons,” *J. Opt. Soc. Am.*, Vol. 44, No. 10, 1954, pp. 761—-765.

[77] Herres, W. and Gronholz, J., “Understanding FT-IR data processing, Part 1-3,” *Comp. Appl. Lab.*, Vol. 2, No. 216, 1984.

[78] Gomer, N. R., Gordon, C. M., Lucey, P., Sharma, S. K., Carter, J. C., and Angel, S. M., “Raman Spectroscopy Using a Spatial Heterodyne Spectrometer: Proof of Concept,” *Applied Spectroscopy*, Vol. 65, No. 8, 2011, pp. 849–857.

[79] Wagner, G. A. and Plusquellic, D. F., “Multi-frequency differential absorption LIDAR system for remote sensing of CO2 and H2O near 1.6 μm,” *Opt. Express*, Vol. 26, No. 15, jul 2018, pp. 19420–19434.

[80] Hakim, N. Z., Saleh, B. E. A., and Teich, M. C., “Generalized excess noise factor for avalanche photodiodes of arbitrary structure,” *IEEE Transactions on Electron Devices*, Vol. 37, No. 3, mar 1990, pp. 599–610.

[81] Lerner, J. and Thevenon, A., *The optics of spectroscopy*, Jobin-Yvon Optical Systems/Instrumentss SA, 1988.

[82] "6 The michelson interferometer," *Spectral Imaging of the Atmosphere*, edited by G. G. Shepherd, Vol. 82 of *International Geophysics*, Academic Press, 2003, pp. 129 – 150.

[83] Folland, G. B. and Sitaram, A., "The uncertainty principle: A mathematical survey," *Journal of Fourier Analysis and Applications*, Vol. 3, No. 3, may 1997, pp. 207–238.

[84] Candès, E. and Romberg, J., "Sparsity and incoherence in compressive sampling," *Inverse Problems*, Vol. 23, No. 3, apr 2007, pp. 969–985.

[85] Nyquist, H., "Certain Topics in Telegraph Transmission Theory," *Transactions of the American Institute of Electrical Engineers*, Vol. 47, No. 2, apr 1928, pp. 617–644.

[86] Vaníček, P., "Approximate spectral analysis by least-squares fit," *Astrophysics and Space Science*, Vol. 4, No. 4, aug 1969, pp. 387–391.

[87] Greengard, L. and Lee, J.-Y., "Accelerating the Nonuniform Fast Fourier Transform," *SIAM Review*, Vol. 46, No. 3, 2004, pp. 443–454.

[88] Moore, E. H., "On the reciprocal of the general algebraic matrix," *Bull. Am. Math. Soc.*, Vol. 26, 1920, pp. 394–395.

[89] Chase, D. B., "Phase Correction In Fourier Transform Infrared (FTIR)," *1981 Intl Conf on Fourier Transform Infrared Spectroscopy*, edited by H. Sakai, Vol. 0289, International Society for Optics and Photonics, SPIE, 1981, pp. 252 – 252.

[90] Barning, F., "The numerical analysis of the light-curve of 12 Lacertae," *Bulletin of the Astronomical Institutes of the Netherlands*, Vol. 17, aug 1963, pp. 22.

[91] Scargle, J., "Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data," *ApJ*, Vol. 263, dec 1982, pp. 835–853.

[92] Draganic, A., Orovic, I., and Stankovic, S., "On some common compressive sensing recovery algorithms and applications - Review paper," *Facta universitatis - series: Electronics and Energetics*, Vol. 30, 04 2017, pp. 477–510.

[93] Candes, E. J., Romberg, J., and Tao, T., "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, Vol. 52, No. 2, feb 2006, pp. 489–509.

[94] Rudelson, M. and Vershynin, R., "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, Vol. 61, No. 8, 2008, pp. 1025–1045.

[95] Candes, E. and Wakin, M., "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, Vol. 25, No. 2, 2008, pp. 21–30.

[96] Davenport, M. A., Duarte, M. F., Eldar, Y. C., and Kutyniok, G., *Introduction to compressed sensing*, Cambridge University Press, 2012, p. 164.

[97] Monchalin, J.-P., Kelly, M. J., Thomas, J. E., Kurnit, N. A., Szöke, A., Zernike, F., Lee, P. H., and Javan, A., "Accurate laser wavelength measurement with a precision two-beam scanning Michelson interferometer," *Appl. Opt.*, Vol. 20, No. 5, mar 1981, pp. 736–757.

[98] Niebauer, T. M., Sasagawa, G. S., Faller, J. E., Hilt, R., and Klopping, F., "A new generation of absolute gravimeters," *Metrologia*, Vol. 32, No. 3, jan 1995, pp. 159–180.

[99] Persky, M. J., "A review of spaceborne infrared Fourier transform spectrometers for remote sensing," *Review of Scientific Instruments*, Vol. 66, No. 10, 1995, pp. 4763–4797.

[100] Davis, S. P., Abrams, M. C., and Brault, J. W., "2 - Why chose a fourier transform spectrometer?" *Fourier Transform Spectrometry*, edited by S. P. Davis, M. C. Abrams, and J. W. Brault, Academic Press, San Diego, 2001, pp. 17 – 27.

[101] Brachet, F., Hbert, P.-J., Cansot, E., Buil, C., Lacan, A., Roucayrol, L., Courau, E., Bernard, F., Casteras, C., Loesel, J., and Pierangelo, C., "Static Fourier transform spectroscopy breadboards for atmospheric chemistry and climate," *Optical Design and Engineering III*, edited by L. Mazuray, R. Wartmann, A. Wood, J.-L. Tissot, and J. M. Raynor, Vol. 7100, International Society for Optics and Photonics, SPIE, 2008, pp. 430 – 440.

[102] Bourassa, A. E., Langille, J., Solheim, B., Degenstein, D., and Dupont, F., "The Spatial Heterodyne Observations of Water (SHOW) Instrument for High Resolution Profiling in the Upper Troposphere and Lower Stratosphere," *Light, Energy and the Environment*, Optical Society of America, 2016, p. FM3E.1.

[103] Faust, B. and Klynning, L., "Low-cost wavemeter with a solid Fizeau interferometer and fiber-optic input," *Appl. Opt.*, Vol. 30, No. 36, dec 1991, pp. 5254–5259.

[104] Lacan, A., Bréon, F.-M., Rosak, A., Brachet, F., Roucayrol, L., Etcheto, P., Casteras, C., and Salaün, Y., "A static Fourier transform spectrometer for atmospheric sounding: concept and experimental implementation," *Opt. Express*, Vol. 18, No. 8, apr 2010, pp. 8311–8331.

[105] Hays, P. B. and Snell, H. E., "Multiplex Fabry–Perot interferometer," *Appl. Opt.*, Vol. 30, No. 22, Aug 1991, pp. 3108–3113.

[106] Kaminow, I. P., Mammel, W. L., and Weber, H. P., "Metal-Clad Optical Waveguides: Analytical and Experimental Study," *Appl. Opt.*, Vol. 13, No. 2, feb 1974, pp. 396–405.

[107] Kerstin, W., G., H. R., Arne, L., and Marcel, H., "TriPleX: a versatile dielectric photonic platform," *Advanced Optical Technologies*, Vol. 4, 2015, pp. 189.

[108] Imamura, S., "Polymeric optical waveguides [materials, packaging and applications]," *1998 IEEE/LEOS Summer Topical Meeting. Digest. Broadband Optical Networks and Technologies: An Emerging Reality. Optical MEMS. Smart Pixels. Organic Optics and Optoelectronics (Cat. No.98TH8369)*, jul 1998, pp. III/35–III/36.

[109] Kawano, K. and Kitoh, T., *Introduction to Optical Waveguide Analysis: Solving Maxwell's Equation and the Schrödinger Equation*, Wiley, 2004.

[110] Wächter, C., "Integrated optics design: software tools and diversified applications," *Frontiers in Planar Lightwave Circuit Technology*, edited by S. Janz, J. Ctyroky, and S. Tanev, Springer Netherlands, Dordrecht, 2006, pp. 245–280.

[111] Rana, F., "Chapter 8 Integrated Optical Waveguides," *ECE 5330 Semiconductor Optoelectronics*, Cornell University, 2003.

[112] Cheben, P., Halir, R., Schmid, J., Atwater, H., and Smith, D., "Subwavelength integrated photonics," *Nature*, Vol. 560, 08 2018.

[113] Kin Seng Chiang, "Analysis of the effective-index method for the vector modes of rectangular-core dielectric waveguides," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 44, No. 5, may 1996, pp. 692–700.

[114] Bienstman, P., Selleri, S., Rosa, L., Uranus, H. P., Hopman, W. C. L., Costa, R., Melloni, A., Andreani, L. C., Hugonin, J. P., Lalanne, P., Pinto, D., Obayya, S. S. A., Dems, M., and Panajotov, K., "Modelling leaky photonic wires: A mode solver comparison," *Optical and Quantum Electronics*, Vol. 38, No. 9, jul 2006, pp. 731–759.

[115] Ghillino, E., Virgillito, E., Mena, P. V., Scarmozzino, R., Stoffer, R., Richards, D., Ghiasi, A., Ferrari, A., Cantono, M., Carena, A., and Curri, V., "The Synopsys Software Environment to Design and Simulate Photonic Integrated Circuits: A Case Study for 400G Transmission," *2018 20th International Conference on Transparent Optical Networks (ICTON)*, July 2018, pp. 1–4.

[116] Uda, R., Yamaguchi, K., Takada, K., and Okamoto, K., "Fabrication of a silica-based complex Fourier-transform integrated-optic spatial heterodyne spectrometer incorporating 120°; optical hybrid couplers," *Appl. Opt.*, Vol. 57, No. 14, may 2018, pp. 3781–3787.

[117] Florjańczyk, M., Cheben, P., Janz, S., Scott, A., Solheim, B., and Xu, D.-X., "Multiaperture planar waveguide spectrometer formed by arrayed Mach-Zehnder interferometers," *Opt. Express*, Vol. 15, No. 26, dec 2007, pp. 18176–18189.

[118] Soldano, L. B. and Pennings, E. C. M., "Optical multi-mode interference devices based on self-imaging: principles and applications," *Journal of Lightwave Technology*, Vol. 13, No. 4, 1995, pp. 615–627.

[119] Halir, R., Molina-Fernandez, I., Ortega-Monux, A., Wanguemert-Perez, J. G., Xu, D.-X., Cheben, P., and Janz, S., "A Design Procedure for High-Performance, Rib-Waveguide-Based Multimode Interference Couplers in Silicon-on-Insulator," *Journal of Lightwave Technology*, Vol. 26, No. 16, 2008, pp. 2928–2936.

[120] Benedikovic, D., Cheben, P., Schmid, J. H., Xu, D.-X., Lamontagne, B., Wang, S., Lapointe, J., Halir, R., Ortega-Moñux, A., Janz, S., and Dado, M., "Subwavelength index engineered surface grating coupler with sub-decibel efficiency for 220-nm silicon-on-insulator waveguides," *Opt. Express*, Vol. 23, No. 17, aug 2015, pp. 22628–22635.

[121] Bock, P. J., Cheben, P., Schmid, J. H., Lapointe, J., Delâge, A., Janz, S., Aers, G. C., Xu, D.-X., Densmore, A., and Hall, T. J., "Subwavelength grating peri-

odic structures in silicon-on-insulator: a new type of microphotonic waveguide," *Opt. Express*, Vol. 18, No. 19, sep 2010, pp. 20251–20262.

[122] Podmore, H., *Subwavelength antireflection for space applications*, Master's thesis, York University, 2015.

[123] Cheben, P., Xu, D.-X., Janz, S., and Densmore, A., "Subwavelength waveguide grating for mode conversion and light coupling in integrated optics," *Opt. Express*, Vol. 14, No. 11, May 2006, pp. 4695–4702.

[124] Padmaraju, K. and Bergman, K., "Resolving the thermal challenges for silicon microring resonator devices," *Nanophotonics*, Vol. 3, 2013, pp. 269 – 281.

[125] Podmore, H., Scott, A., Cheben, P., Sioris, C., Cameron, P., Schmid, J., Lohmann, A., Corriveau, Z., and Lee, R., "Athermal planar-waveguide Fourier-transform spectrometer for methane detection," *Optics Express*, Vol. 25, No. 26, 2017.

[126] Shindell, D. T., Faluvegi, G., Koch, D. M., Schmidt, G. A., Unger, N., and Bauer, S. E., "Improved Attribution of Climate Forcing to Emissions," *Science*, Vol. 326, No. 5953, oct 2009, pp. 716 LP – 718.

[127] Howarth, R. W., Santoro, R., and Ingraffea, A., "Methane and the greenhouse-gas footprint of natural gas from shale formations," *Climatic Change*, Vol. 106, No. 4, apr 2011, pp. 679.

[128] Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quere, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E., and Zeng, G., "Three decades of global methane sources and sinks," *Nature Geosci*, Vol. 6, No. 10, oct 2013, pp. 813–823.

[129] Harig, R., Matz, G., Rusch, P., Gerhard, J.-H., Schaefer, K. P., Jahn, C., Schwengler, P., and Beil, A., "Remote detection of methane by infrared spectrometry for airborne pipeline surveillance: first results of ground-based measurements," 2004.

[130] Cheben, P., "Wavelength Dispersive Planar Waveguide Devices: Echelle and

Arrayed Waveguide Gratings," *Optical Waveguides: From Theory to Applied Technologies*, edited by M. L. Calvo and V. Lakshminarayanan, chap. 5, London, 1st ed., 2007, pp. 173—-217.

[131] Cheben, P., Powell, I., Janz, S., and Xu, D.-X., "Wavelength-dispersive device based on a Fourier-transform Michelson-type arrayed waveguide grating," *Opt. Lett.*, Vol. 30, No. 14, 2005, pp. 1824—-1826.

[132] Podmore, H., Scott, A., Cheben, P., Velasco, A., Schmid, J., Vachon, M., and Lee, R., "Demonstration of a compressive-sensing Fourier-transform on-chip spectrometer," *Optics Letters*, Vol. 42, No. 7, 2017.

[133] Velasco, A. V., Cheben, P., Bock, P. J., Delâge, A., Schmid, J. H., Lapointe, J., Janz, S., Calvo, M. L., Xu, D.-X., Florjańczyk, M., and Vachon, M., "High-resolution Fourier-transform spectrometer chip with microphotonic silicon spiral waveguides," *Opt. Lett.*, Vol. 38, No. 5, mar 2013, pp. 706–708.

[134] Yang, M., Li, M., and He, J.-J., "Static FT imaging spectrometer based on a modified waveguide MZI array," *Opt. Lett.*, Vol. 42, No. 14, jul 2017, pp. 2675–2678.

[135] Herrero-Bermello, A., Velasco, A. V., Podmore, H., Cheben, P., Schmid, J. H., Janz, S., Calvo, M. L., Xu, D.-X., Scott, A., and Corredera, P., "Temperature

dependence mitigation in stationary Fourier-transform on-chip spectrometers," *Opt. Lett.*, Vol. 42, No. 11, jun 2017, pp. 2239–2242.

[136] Schmid, J. H., Ibrahim, M., Cheben, P., Lapointe, J., Janz, S., Bock, P. J., Densmore, A., Lamontagne, B., Ma, R., Ye, W. N., and Xu, D.-X., "Temperature-independent silicon subwavelength grating waveguides," *Opt. Lett.*, Vol. 36, No. 11, jun 2011, pp. 2110–2112.

[137] Florjaczyk, M., Cheben, P., Janz, S., Lamontagne, B., Lapointe, J., Scott, A., Solheim, B., and Xu, D.-X., "Development of a slab waveguide spatial heterodyne spectrometer for remote sensing," *Proceedings of SPIE*, Vol. 7594, 2010, pp. 75940R–75940R–9.

[138] Ma, X., Li, M., and He, J. J., "CMOS-Compatible Integrated Spectrometer Based on Echelle Diffraction Grating and MSM Photodetector Array," *IEEE Photonics Journal*, Vol. 5, No. 2, apr 2013, pp. 6600807.

[139] Okamoto, K., Aoyagi, H., and Takada, K., "Fabrication of Fourier-transform, integrated-optic spatial heterodyne spectrometer on silica-based planar waveguide," *Optics Letters*, Vol. 35, No. 12, jun 2010, pp. 2103–2105.

[140] Erfan, M., Sabry, Y. M., Sakr, M., Mortada, B., Medhat, M., and Khalil, D., "On-Chip Micro–Electro–Mechanical System Fourier Transform Infrared

(MEMS FT-IR) Spectrometer-Based Gas Sensing," *Applied Spectroscopy*, Vol. 70, No. 5, may 2016, pp. 897–904.

[141] Starling, D. J., Storer, I., and Howland, G. A., "Compressive sensing spectroscopy with a single pixel camera," *Appl. Opt.*, Vol. 55, No. 19, Jul 2016, pp. 5198–5202.

[142] van den Berg, E. and Friedlander, M. P., "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, Vol. 31, No. 2, 2008, pp. 890–912.

[143] Katz, O., Levitt, J. M., and Silberberg, Y., "Compressive Fourier Transform Spectroscopy," *Frontiers in Optics 2010/Laser Science XXVI*, 2010, pp. FTuE3.

[144] Bock, P. J., Cheben, P., Velasco, A. V., Schmid, J. H., Delâge, A., Florjaczyk, M., Lapointe, J., Xu, D.-X., Vachon, M., Janz, S., and Calvo, M. L., "Subwavelength grating Fourier-transform interferometer array in silicon-on-insulator," *Laser & Photonics Reviews*, Vol. 7, No. 6, 2013, pp. L67–L70.

[145] Cheben, P., Bock, P. J., Schmid, J. H., Lapointe, J., Janz, S., Xu, D.-X., Densmore, A., Delâge, A., Lamontagne, B., and Hall, T. J., "Refractive index engineering with subwavelength gratings for efficient microphotonic couplers

and planar waveguide multiplexers," *Optics Letters*, Vol. 35, No. 15, aug 2010, pp. 2526–2528.

[146] Fornasier, M., *Theoretical foundations and numerical methods for sparse recovery*, Vol. 9, Walter de Gruyter, 2010.

[147] Jones, A., Bland-Hawthorn, J., and Shopbell, P., "Towards a General Definition for Spectroscopic Resolution," *Astronomical Data Analysis Software and Systems IV, Astronomical Society of the Pacific Conference Series*, Vol. 77, 1995, pp. 503.

[148] Podmore, H., Scott, A., and Lee, R., "On-Chip Compressed Sensing Fourier-Transform Visible Spectrometer," *IEEE Photonics Journal*, Vol. 10, dec 2018, pp. 2876657.

[149] Xia, Z., Eftekhar, A. A., Soltani, M., Momeni, B., Li, Q., Chamanzar, M., Yegnanarayanan, S., and Adibi, A., "High resolution on-chip spectroscopy based on miniaturized microdonut resonators," *Opt. Express*, Vol. 19, No. 13, 2011, pp. 12356–12364.

[150] Nitkowski, A., Preston, K., Sherwood-Droz, N., Schmid, B. S., and Hajian, A. R., "On-chip spectrometer for low-cost optical coherence tomography," *Proc. SPIE*, Vol. 8934, 2014, pp. 8934 – 8934 – 8.

[151] Sáinz, C., Jourdain, P., Escalona, R., and Calatroni, J., "Real time interferometric measurements of dispersion curves," *Optics Communications*, Vol. 111, No. 5, 1994, pp. 632–641.

[152] Scott, A. and Podmore, H., *Integrated fourier transform optical spectrometer*, US Patent No. US9964396B1, 2018.

[153] Pavanello, F., Vasiliev, A., Muneeb, M., and Roelkens, G., "Broadband Digital Fourier Transform Spectrometer for On-Chip Wavelength Monitoring in the 2.3-μm Wavelength Range," *IEEE Photonics Journal*, Vol. 11, No. 3, jun 2019, pp. 1–9.

[154] Kita, D. M., Miranda, B., Favela, D., Bono, D., Michon, J., Lin, H., Gu, T., and Hu, J., "High-performance and scalable on-chip digital Fourier transform spectroscopy," *Nature Communications*, Vol. 9, No. 1, 2018, pp. 4405.