POLITECHNIKA WROCŁAWSKA
WYDZIAŁ ELEKTRONIKI

Raport nr W04/P-013/18

Katedra Telekomunikacji i Teleinformatyki

Signal processing in terahertz and mid-infrared spectroscopy with frequency combs

ŁUKASZ ANTONI STERCZEWSKI

Rozprawa doktorska

Promotor:
prof. dr. Edward Pliński
Politechnika Wrocławska

Ko-promotor:
Prof. dr. Gerard Wysocki
Princeton University

WROCŁAW 2018

WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF ELECTRONICS

Telecommunications and Teleinformatics Department

Signal processing in terahertz and mid-infrared spectroscopy with frequency combs

ŁUKASZ ANTONI STERCZEWSKI

Doctoral thesis

Supervisor:
prof. dr. Edward Pliński
Wroclaw University of Science and Technology

Co-supervisor:
Prof. dr. Gerard Wysocki
Princeton University

WROCLAW 2018

If you want to find the secrets of the universe,

think in terms of energy,

frequency

and vibration.

— Nikola Tesla

Abstract

Optical frequency combs have established their position in precision laser spectroscopy over the last two decades thanks to the brilliant concept of a hybrid between narrow line and broadband laser emission. In an optical frequency comb, hundreds to hundreds of thousands of equidistant optical modes are emitted simultaneously, enabling to perform spectroscopic experiments in a massively parallel fashion: each comb tooth probes a different part of the optical spectrum at the same time, which is in stark contrast to single-mode laser solutions. To fully exploit the spectroscopic potential of frequency combs, including those highly-miniaturized or even unstable, a novel signal processing framework has been developed and described in this thesis. The main motivation was the necessity of coherent averaging over extended timescales for low-uncertainty broadband optical characterization of the sample. This dissertation considers two classes of spectroscopic experiments utilizing frequency combs. In the first, referred to as homodyne detection mode, a single mode-locked femtosecond laser in the near-infrared pumps photoconductive antennas to generate and detect terahertz radiation. Since the nature of the measurement scan is mechanical, irreproducibility of the moving mirror position introduces considerable amounts of multiplicative phase noise. It is shown, how cross-correlation based alignment of time-domain terahertz pulses enables to suppress the noise for long term averaging, and thereby improve the spectroscopic coverage, as well as lower the uncertainty of the measurement, which is proven on measurements of gaseous and solid samples. In the second class of experiments, two highly miniaturized, chip-scale frequency combs are set up in a multi-heterodyne (dual-comb) arrangement without any moving parts. Unfortunately, since each laser comes from an independent cavity, fluctuations of their offset frequencies and repetition rates are unsynchronized. This is turn produces a photodetector signal rich in phase noise with non-resolvable radio frequency comb lines in extreme conditions. To overcome this issue, it is shown, how broadly understood signal processing helps to synchronize two quantum cascade or interband cascade laser frequency combs via hardware loops, which eventually can be replaced with a near real-time purely computational solution enabling to perform dual comb spectroscopy with completely free-running sources. A mathematical derivation of the digital phase correction formulas, as well as supporting experimental data of spectroscopy on molecular absorbers in the mid-infrared and terahertz range are provided. The proposed computational solution can find application in highly-miniaturized battery-operated portable spectrometers without any moving parts for space and environmental applications. With the help of the algorithm, the first demonstrations of molecular dual comb spectroscopy (DCS) with noisy but coherent interband cascade lasers, as well as DCS with terahertz quantum cascade lasers suffering from large instabilities caused by a vibrating pulse tube cryostat have been performed.

Streszczenie

Optyczne grzebienie częstotliwości mają od dwóch dekad ugruntowaną pozycję w precyzyjnej spektroskopii laserowej dzięki genialnej idei połączenia szerokopasmowej emisji z wąską linią emisyjną lasera. W optycznym grzebieniu częstotliwości jednocześnie emitowane jest kilkaset do setek tysięcy równoodległych modów, które pozwalają na przeprowadzanie eksperymentów spektroskopowych w sposób masywnie równoległy: każdy mod optyczny próbkuje inną część widma optycznego w tej samej chwili, co odróżnia go od lasera jednomodowego. Aby w pełni skorzystać z możliwości oferowanych przez optyczne grzebienie częstotliwości, włączywszy w to miniaturowe, a nawet niestabilne źródła, potrzebne było rozwinąć stosowne algorytmy przetwarzania sygnałów, które są opisane w niniejszej pracy. Motywacją do jej powstania był wymóg koherentnego uśredniania sygnałów celem przeprowadzania szerokopasmowej charakteryzacji optycznej badanego medium o małej niepewności pomiaru. W rozprawie wzięto pod uwagę dwie klasy pomiarów spektroskopowych na bazie optycznych grzebieni częstotliwości. Pierwszy, zwany układem detekcji homodynowej, wykorzystuje pojedynczy laser femtosekundowy bliskiej podczerwieni do optycznego pompowania terahercowych anten fotoprzewodzących na potrzeby generacji i detekcji fal terahercowych. Mechaniczny sposób pomiaru wiąże się z niepowtarzalnością położenia ruchomego lustra w optycznej linii opóźniającej, co skutkuje dużą ilością multiplikatywnego szumu fazowego w sygnale spektroskopowym. W pracy pokazane jest, jak wykorzystując korelację wzajemną przeciwdziała się niekorzystnemu wpływowi szumu fazowego, uniemożliwiającego pierwotnie długofalowe uśrednianie. Dzięki korekcji fazowej, poprawiono pasmo sygnału terahercowego, jak również obniżona została niepewność pomiarowa, co pokazano na przykładzie próbek gazowych i stałych. W drugiej klasie eksperymentów, dwa wysoce zminiaturyzowane grzebienie częstotliwości o rozmiarach mikroukładów elektronicznych zostają zestawione w układzie multi-heterodyny (spektrometru dwugrzebieniowego) pozbawionej jakichkolwiek elementów ruchomych. Owe rozwiązanie ma jednak wadę związaną z faktem, iż dwa grzebienie częstotliwości pochodzące z różnych źródeł będą miały nieskorelowane fluktuacje częstotliwości offsetu oraz powtarzania. To z kolei przekłada się na bogaty w szum fazowy sygnał na fotodetektorze z nierozróżnialnymi liniami grzebienia w zakresie radiowym w najbardziej skrajnych przypadkach. W rozprawie zaprezentowane jest, jak szeroko rozumiane przetwarzanie sygnałów pomaga zsynchronizować pracę dwóch grzebieni częstotliwości wytwarzanych przez kwantowe lasery kaskadowe lub międzypasmowe lasery kaskadowe dzięki sprzętowym pętlom synchronizacji, które mogą być zastąpione cyfrową pętlą korekcji fazowej z możliwością sprzętowej implementacji w czasie rzeczywistym. Cyfrowy algorytm posłużył ostatecznie do pomiarów spektroskopowych z całkowicie niezsynchronizowanymi laserami. W rozprawie przedstawione jest matematyczne wyprowadzenie procedur korekcji fazowej, które wykorzystano w licznych eksperymentach spektroskopowych absorberów molekularnych, dowodzące użyteczności rozwiniętych metod. Proponowane rozwiązanie cyfrowej korekcji fazowej może znaleźć zastosowanie w wysoce zminiaturyzowanych spektrometrach z zasilaniem bateryjnym pozbawionymi jakichkolwiek elementów ruchomych na potrzeby aplikacji kosmicznych i ochrony środowiska. Z pomocą algorytmu, przeprowadzono pierwszą na świecie demonstrację molekularnej spektroskopii na grzebieniach optycznych pochodzących z lekko zaszumionych, lecz koherentnych międzypasmowych laserów kaskadowych, jak również kwantowych laserów kaskadowych w zakresie terahercowym, których praca była mocno zakłócona wibrującym kriostatem utrzymującym kriogeniczne warunki pracy.

Preface

This doctoral thesis is the result of my research I have been involved in different laboratories across the world. My research started in the fall of 2013, when I proudly joined prof. Edward Plinski's Terahertz Laboratory at Wroclaw University of Science and Technology to unlock the potential of the weakly explored terahertz regime in application to study the thermodynamics of pharmaceutical ingredients in collaboration with the Department of Chemistry of Drugs at Wroclaw Medical University. Initially, I was involved in the characterization of the optical properties of new derivatives of an anti-inflammatory drug Piroxicam used in colon cancer treatment in the terahertz regime using femtosecond lasers and photoconductive optical switches. At that time, I embedded a unique system to control the temperature of pharmaceutical samples with high accuracy into the terahertz spectrometer that additionally enabled us to analyze the dehydration kinetics of the two of the most popular pharmaceutical ingredients: α-D-glucose and α-D-lactose monohydrate with the help of a custom advanced statistical signal processing framework I developed at that time.

In May of 2015, after a spectroscopy lecture of prof. Gerard Wysocki who was visiting his *Alma Mater*, I was invited to join his laboratory at Princeton University in order to explore the fascinating world of small-footprint yet high power terahertz (THz) quantum cascade laser (QCL) frequency combs with a large potential for miniaturization into portable devices. The use of THz QCL combs in spectroscopy at that time had been unexplored. Since I was eager to face the challenge, I joined the overseas laboratory in September 2015, where I had access to the cutting-edge mid-infrared and terahertz lasers from numerous collaborators across the country. What was promised to be a challenge, turned out to be such. In contrast to the mode-locked femtosecond oscillators that produced a perfectly stable optical spectrum with the turn of a key I was used to, the first QCLs required luck and patience to find a suitable regime where they behaved like frequency combs. Not only was it required to find one laser emitting a stable comb. For our spectroscopic experiments we needed two lasers matched in parameters and "lucky" regimes. I started my first experiments with the well-established QCLs in the mid-infrared range operating at near room temperature, while getting ready to face the cryogenic equipment needed to run the THz QCLs.

As we know, engineering helps us to deal with numerous imperfections of Mother Nature. The initially chaotic behavior of the mid-IR QCLs was mastered with the help of different servo loops we developed, which enabled us to perform all kinds of broadband and high resolution spectroscopic experiments with gaseous samples. A year later, I harnessed my signal processing skills to deal with an even more chaotic behavior of a new kind of semiconductor light source for use in a dual-comb spectroscopy configuration: the interband cascade laser (ICL). The first devices showed a featureless spectral blob, which motivated me to invent a new kind of near real-time signal processing algorithm for dual comb spectroscopy after being inspired by earlier works on the powerful yet computationally demanding Kalman filter. This

allowed us to perform the first demonstration of broadband and high resolution multiheterodyne spectroscopy using ICLs in a comb-like regime with a current consumption comparable to three light emitting diodes such as those in a smartphone.

The usefulness of the ICL coupled with its low power consumption was reflected in the successful detection of methane on Mars with its use several years earlier as part of the Mars Curiosity mission, however the space ICL operated in a spectrally pure single mode regime. Between June and August 2017 I had the unprecedented opportunity to work on ICL frequency combs with a saturable absorber that acted like hundred phase-locked single mode lasers in a small footprint moving-parts free mid-infrared spectrometer. I joined the same division at NASA Jet Propulsion Laboratory in Pasadena, California, that developed the space ICL – the Microdevices Laboratory. Similar as in the QCLs, we faced some difficulties with the dispersion that perturbed the phase-locking mechanism in the ICLs, which I characterized throughout the duration of my internship. Notably, exactly the same issues needed to be tackled at the time of first femtosecond fiber lasers.

Surprisingly, the terahertz project that was supposed to be the main reason for my 2.5-year research stay in the United States turned out to be a successful by-product of the numerous solutions we developed to deal with noise and optical feedback issues in the mid-infrared regime. Without the experience gathered in systems running on a daily basis at room temperature, our liquid-helium-dependent terahertz measurements would probably have failed. An additional effect was the enrichment of my thesis including results of spectroscopy in numerous configurations and the associated signal processing path.

For convenience, this thesis is divided into four chapters and an appendix. In the first chapter, I introduce the concept of optical frequency combs and their generation mechanisms. In the second chapter, I discuss the signal processing aspects of spectroscopy utilizing a tabletop commercially available stabilized near-infrared mode-locked femtosecond laser pumping a terahertz photoconductive antenna arranged in homodyne detection mode. In this technique, referred to as terahertz time domain spectroscopy (THz-TDS) only one frequency comb is used, but the mechanical nature of the acquisition precludes simple point-wise averaging for characterization of samples with low uncertainty. In the third chapter, the homodyne technique and bulky optical setup is replaced with chip-scale direct emitters of optical frequency combs with current and temperature tunability arranged in a dual-comb (multiheterodyne) setup. Unfortunately, it comes at the cost of high requirements of mutual stability between the lasers, which in heterodyne detection mode gives rise to substantial amounts of phase noise, making spectroscopic assessments nearly impossible unless analog and digital lock loops as well as advanced phase-correction algorithms are used for synchronization and/or phase correction, and those have been proposed and proven highly effective in this thesis. In the fourth chapter, I summarize the obtained spectroscopic results and provide an outlook for future experiments.

Łukasz Sterczewski
Princeton, 17.12.2017

Acknowledgements

First of all, I would like to thank my supervisors: prof. Edward Plinski and prof. Gerard Wysocki for enabling me to purse research on spectroscopy using frequency combs. They seeded in me the curiosity to explore new areas of science, as well as the courage to combine various scientific disciplines in interdisciplinary research. The PhD adventure I experienced thanks to their courtesy shaped the way I perceive science communication and collaboration with other researchers.

My gratitude goes to Dr. Jonas Westberg of Princeton University, who inspired me with his consistency and persistency in achieving scientific goals. Our seamless collaboration and inspiring discussions have led to many interesting discoveries and improvements in the field of multiheterodyne spectroscopy based on QCL and ICL frequency combs. He introduced me to the world of laser spectroscopy, which I approached from a different perspective, as well as he commented in detail on this work. I cannot forget about all the members of prof. Wysocki's laboratory, whose help and fruitful discussions have also helped improve this dissertation: Yifeng Chen, Cheyenne Teng, Link Patrick, Michael Soskind, Dr. Genevieve Plant, and Dr. Eric Zhang.

I would like also to thank MSc. Michal Grzelczak, my laboratory partner at Wroclaw Unviersity of Science and Technology for his willingness to risk in attempt to make broadband terahertz spectroscopy of drugs more accurate and less susceptible to noise.

Special thanks go to Dr. David Burghoff and MSc. Yang Yang of MIT for providing us with the excellent dispersion-compensated terahertz quantum cascade lasers. Their commitment to the project often involved laborious fabrication and characterization of devices, which after weekend tests, were shipped to Princeton on Monday.

I would like to extend my sincerest thanks to the employees of the Microdevices Laboratory at NASA Jet Propulsion Laboratory who enabled me to enjoy an unprecedented adventure with interband cascade laser (ICL) frequency combs. I want to thank Dr. Mahmood Bagheri for his mentorship and believing in my rapid adaptation skills in order to analyze complex problems of the dispersion in the ICLs.

I owe special thanks to Dr. Clifford Frez who eagerly stayed until late evening hours in order to introduce me to SEM and show me the defects we had in one of the batches of ICLs. My thanks also go to Mr. Mathieu Fradet who helped me with the characterization of optical coatings and guided me how to use a wire bonder.

I cannot forget about prof. Krzysztof Abramski of Wroclaw University of Technology who introduced me to the world of lasers and optoelectronics as first. Without his mentorship and guidance I would not have been able to participate in this scientific adventure I have had for the last couple of years.

Most importantly, I would like to cordially thank my wife who supported me every day, and my parents who arouse in me the passion of science and engineering.

Finally, I would like to acknowledge support from the Kosciuszko Foundation Grant for research in the United States I was awarded, which enabled to extend my stay for an additional half a year and perform the nearly forgotten terahertz experiments.

List of publications

This thesis is based on the following publications:

L. A. Sterczewski, M. P. Grzelczak, K. Nowak, B. Szlachetko, and E. F. Plinski, "Bayesian separation algorithm of THz spectral sources applied to D-glucose monohydrate dehydration kinetics," *Chemical Physics Letters*, vol. 644, pp. 45–50, Jan. 2016 DOI: 10.1016/j.cplett.2015.11.031.

L. A. Sterczewski, M. P. Grzelczak, and E. F. Plinski, "Heating system of pellet samples integrated with terahertz spectrometer," *Review of Scientific Instruments*, vol. 87, no. 1, p. 013106, Jan. 2016 DOI: 10.1063/1.4939755.

K. Nowak, E. F. Pliński, B. Karolewicz, P. P. Jarząb, S. Plińska, B. Fuglewicz, M. J. Walczakowski, Ł. Augustyn, **Ł. A. Sterczewski**, M. P. Grzelczak, M. Hruszowiec, G. Beziuk, M. Mikulic, N. Pałka, and M. Szustakowski, "Selected aspects of terahertz spectroscopy in pharmaceutical sciences," *Acta poloniae pharmaceutica*, vol. 72, no. 5, p. 851—866, 2015.

L. A. Sterczewski, M. P. Grzelczak, and E. F. Plinski, "Terahertz antenna electronic chopper," *Review of Scientific Instruments*, vol. 87, no. 1, p. 014702, Jan. 2016 DOI: 10.1063/1.4939461.

L. A. Sterczewski, J. Westberg, and G. Wysocki, "Tuning properties of mid-infrared Fabry-Pérot quantum cascade lasers for multiheterodyne spectroscopy," *Photonics Letters of Poland*, vol. 8, no. 4, pp. 113–115, 2016 DOI: 10.4302/plp.2016.4.08.

J. Westberg*, **L. A. Sterczewski***, and G. Wysocki, "Mid-infrared multiheterodyne spectroscopy with phase-locked quantum cascade lasers," *Applied Physics Letters*, vol. 110, no. 14, p. 141108, Apr. 2017 DOI: 10.1063/1.4979825.

L. A. Sterczewski*, J. Westberg*, and G. Wysocki, "Molecular dispersion spectroscopy based on Fabry–Perot quantum cascade lasers," *Optics Letters*, vol. 42, no. 2, p. 243, Jan. 2017 DOI: 10.1364/OL.42.000243.

L. A. Sterczewski*, J. Westberg*, L. Patrick, C. Soo Kim, M. Kim, C. L. Canedy, W. W. Bewley, C. D. Merritt, I. Vurgaftman, J. R. Meyer, and G. Wysocki, "Multiheterodyne spectroscopy using interband cascade lasers," *Optical Engineering*, vol. 75, no. 1, 2017 DOI: 10.1117/1.OE.57.1.011014.

J. Westberg*, **L. A. Sterczewski***, F. Kapsalidis, Y. Bidaux, J. Wolf, M. Beck, J. Faist, and G. Wysocki "Dual-comb spectroscopy using plasmon-enhanced waveguide dispersion compensated quantum cascade lasers," *in manuscript*, 2018.

Other publications by the author, not included in the thesis:

L. A. Sterczewski, M. P. Grzelczak, K. Nowak, and E. F. Plinski, "Cast terahertz lenses made of caramelized sucrose," *Optical Engineering*, vol. 55, no. 9, p. 090505, Sep. 2016 DOI: 10.1117/1.OE.55.9.090505.

L. A. Sterczewski, K. Nowak, B. Szlachetko, M. P. Grzelczak, B. Szczesniak-Siega, S. Plinska, W. Malinka, and E. F. Plinski, "Chemometric Evaluation of THz Spectral Similarity for the Selection of Early Drug Candidates," *Scientific Reports*, vol. 7, no. 1, Dec. 2017 DOI: 10.1038/s41598-017-14819-6.

M. Bagheri, C. Frez, **L. A. Sterczewski**, I. Gruidin, M. Fradet, I. Vurgaftman, C. L. Canedy, W. W. Bewley, C. D. Merritt, C. S. Kim, M. Kim, and J. R. Meyer, "Passively mode-locked interband cascade

optical frequency combs," *Scientific Reports*, vol. 8, no. 1, p. 3322, Feb. 2018 DOI: 10.1038/s41598-018-21504-9.

---

*These authors contributed equally to this work

Contents

Chapter 1

Introduction

According to Albert Einstein “Everything in Life is Vibration”. If I had to shortly summarize what this thesis is about, I would say that it is how to harness the vibrations of light to probe the characteristic vibrations of a molecule. In a slightly more detailed view, it is how to synchronize multiple concurrent vibrations within a coherent light source, and then to synchronize it with another coherent light source with a similar structure to probe the intra-molecular or inter-molecular vibrations of a sample.

A question that naturally arises is how signal processing, being part of the dissertation title, is related to vibration. Here the answer is also pretty straightforward. Given a system to characterize, we provide a stimulus and observe a response to it. The stimulus is a physical signal that has a time and frequency domain representation. In the latter, everything can be seen as an infinite superposition of oscillations (vibrations) at different frequencies with varying intensities. Even if our stimulus is non-stationary (changes its statistics over time), we can find short time intervals when its frequency domain picture is arbitrarily accurate. Signal processing establishes procedures how to properly acquire the response, and analyze the input-output relationship to extract information about the sample or system under test to broaden our knowledge about the surrounding world. Intuitively, an object is perceived as richly characterized if its response is probed at the widest possible range of frequencies, equivalent to probing it sequentially or concurrently with a large number of single-frequency oscillations.

After this slightly philosophical introduction, we will be more accurate regarding the coverage of this work. This dissertation studies how to retrieve the optical parameters of a sample illuminated by an optical frequency comb, which consists of multiple synchronized oscillations capable of simultaneous multi-frequency yet high resolution stimulation of the medium. While this concept is brilliant, a true frequency comb jitters in frequency: the spacing between its teeth as well as the absolute position (offset) fluctuates in time. In homodyne detection mode, those fluctuations affect the frequency accuracy of the measurement, albeit predominantly they are negligible compared to the width of the probed absorption features. Also in the studied technique: terahertz time-domain spectroscopy (THz-TDS), only the repetition rate fluctuations persist due to its differential frequency generation (DFG) nature. What really justifies the necessity of studying signal processing aspects of spectroscopy with a single frequency comb (where fluctuations are common in nature: the same laser is used to pump and probe the sample) is the mechanical nature of the scan. Variability in shot-to-shot acquisition of a terahertz pulse precludes

long term averaging, which is crucial for broadening the spectroscopic coverage and lowering the uncertainty of the phase and amplitude measurements.

The homodyne case due to its common-mode nature camouflaged some of the imperfections of the source. In heterodyne mode, wherein one optically beats two independent comb sources, jitter in parameters is more pronounced, making spectroscopic assessments in extreme cases nearly impossible: the lines of a radio frequency comb mapping the optical domain mutually overlap. This is often observed in dual-comb spectroscopy highly miniaturized chip-scale semiconductor quantum cascade and interband cascade laser frequency combs. Dedicated signal processing tools to combat this issue are of large importance due to numerous advantages offered by this technique, i.e. all solid-state implementation, microsecond scale acquisition, and unprecedented scale of miniaturization. It turned out that slightly noisy but coherent semiconductor devices initially considered incompatible with dual comb spectroscopy can be successfully harnessed for probing molecular vibrations with the help of quite sophisticated phase correction algorithms. It is likely that further developments in this field will enable the broad application of miniaturized frequency combs in portable devices for environmental and space applications. Let us start with an overview of frequency combs and their generation mechanisms followed by formulating the thesis statement that we will be proving in the subsequent chapters.

1.1. Frequency combs

The genius idea of the frequency comb led to the Nobel Prize in Physics awarded to John L. Hall and Theodor W. Hänsch in 2005, which was shared with Roy J. Glauber. While a frequency comb can be defined in many ways, some authors propose to limit its definition to optical sources only, even though many electronic (i.e. microwave) sources are consistent with its model as well. In general, a frequency comb is just an array of evenly spaced frequencies (teeth) or equivalently a series of delta functions. In the frequency domain, it is described by a simple equation

$$f(n) = f_{\mathrm{ceo}} + n f_{\mathrm{rep}}, \qquad (1.1)$$

where n denotes the n-th tooth of a frequency comb, f_{ceo} is the carrier-envelope offset frequency, and f_{rep} is the repetition rate (frequency). From this definition it follows that there is no restriction regarding the frequency range. This is important, since historically one of the widespread experimental techniques – dual comb arrangement – used widely nowadays for precision laser spectroscopy was first developed in the microwave range by Keilmann *et al.* in 1998 [1]. Nevertheless, in this work we will focus on laser-based combs operating in the mid- and far-infrared frequency range.

In the frequency domain, frequency combs act as rulers or gears often referenced to ultra-stable atomic clocks enabling to transpose the stability of a microwave atomic transition to high optical frequencies or even X-rays. Many impressive results have been obtained in prof. John L. Hall's group at National Institute of Standards and Technology (NIST) in USA regarding absolute stabilization to a metrological standard [2]. The frequency comb can be also seen as an array of multiple phase-locked single mode

sources centered at different frequencies, which would be experimentally difficult to obtain in a conventional way with optical phase locked loops. What is their generation mechanism? To draw a simple picture, let us move to the time domain by means of the Fourier Transform.

One of the Fourier Transform properties is that a periodic time domain signal yields a discrete frequency spectrum. This addresses the question of the origin of comb's repetition rate. In order to create an array of evenly spaced frequencies, one has ensure a mechanism that locks the frequencies to one another, which can be obtained generally in two ways. One requires strong amplitude modulation, whereas the other relies on frequency modulation. The first mechanism known as mode locking utilizes active (periodic acousto- or electro-optic modulation) or passive (saturable absorption) techniques to synchronize the optical phases of optical modes within a laser cavity. A mode-locked laser produces ultrafast (femtosecond to picosecond) pulses with high peak powers with identical optical phases, as illustrated in Figure 1.1.

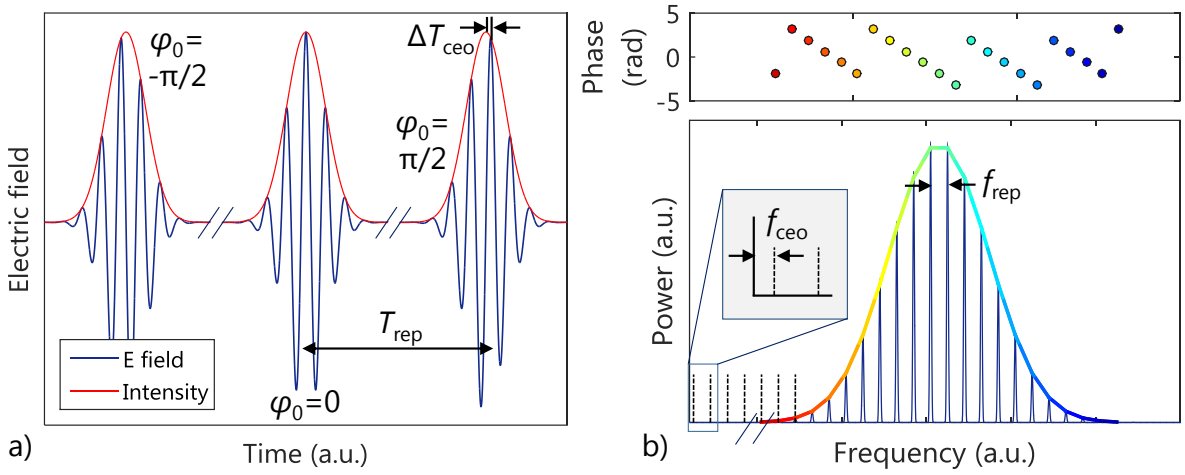

Figure 1.1. a) Time-domain picture of a mode-locked laser (zoom). Periodic pulses with a Gaussian envelope are the result of a superposition of cosine waves with identical phases. In this simple model, the amplitudes of the cosines follow the Gaussian distribution centered around an arbitrary frequency with an arbitrary width. The spacing between the pulses and the time delay due to the carrier-envelope offset is also presented. b) Frequency spectrum of the time-domain trace in a. The carrier envelope offset frequency is shown with arrows in inset. Note the phase spectrum is linear (after unwrapping) which is expected to grow linearly as a function of frequency. After a roundtrip, all phases are equal hence interfering constructively, yielding intensity modulation.

In an optical cavity, the repetition rate f_{rep} will be related to its length L and the mean group velocity v_g, which in turn depends on the refractive index

$$\frac{1}{f_{\mathrm{rep}}} = T_{\mathrm{rep}} = \frac{2L}{v_g}. \qquad (1.2)$$

Notably, light travelling in the medium will encounter different phase and group velocities, which causes the carrier-envelope phase φ_{ceo} to increase from pulse to pulse by

$$\Delta\varphi_{\mathrm{ceo}} = 2\pi f_0 \Delta T_{\mathrm{ceo}} \quad (1.3)$$

where f_0 is the carrier frequency, and ΔT_{ceo} is the time delay. In a mode locked frequency comb, this incremental phase reaches 2π after T_{ceo}, whose reciprocal is known as the carrier-envelope offset frequency f_{ceo}. Since this happens after *n* roundtrips in the resonator, one can conveniently write

$$\Delta\varphi_{\mathrm{ceo}} = 2\pi \frac{f_{\mathrm{ceo}}}{f_{\mathrm{rep}}}. \quad (1.4)$$

This type of frequency combs is predominantly produced by mode locked fiber combs, which are now commercially available, yet their footprints exceed those of millimeter-scale semiconductor laser combs. Furthermore to produce light in the spectroscopically-important mid-infrared or terahertz regime, they require nonlinear conversion techniques for difference frequency generation (down-conversion) to a lower wavelength range from the most popular telecom wavelengths around 1-1.5 μm. However, recently first passively mode-locked semiconductor sources emitting light in the mid-infrared have been demonstrated and will be presented in the third chapter of this dissertation.

An alternative solution to active or passive mode-locking is provided by frequency modulation (FM). In fact, the frequency comb model requires that the teeth are only equidistant rather than of identical phase. While one of the side effects is the loss of pulsed operation, from a spectroscopic standpoint such continuous wave combs are perfectly suitable for probing molecular transitions. To analyze how they operate, let us analyze a sinusoidally frequency modulated (FM) comb. A perfect single-tone FM signal is given by

$$E(t) = E_0 \cos\left(2\pi f_c t + m \cos\left(2\pi f_{\mathrm{rep}} t\right)\right). \quad (1.5)$$

The modulation index m is the ratio of the peak frequency deviation f_Δ to the modulation frequency, which here is simply the repetition rate, therefore $f_\Delta = m f_{\mathrm{rep}}$. Intuitively, a large swing of the modulation frequency will correspond to a large occupied bandwidth. As a rule of thumb, it is often approximated by $B = 2f_\Delta$. The time domain profile of such an FM signal can be expressed as [3], [4]:

$$E(t) = E_0 \sum_{n=-\infty}^{\infty} J_n(m) \cos(2\pi(f_{\mathrm{c}} + n f_{\mathrm{rep}})t - |n|\frac{\pi}{2} \quad (1.6)$$

where $J_n(m)$ is the Bessel function of the first kind. A simulation of the time-domain profile, and frequency spectrum for m=10 is shown in Figure 1.2. As expected, the power spectrum consists of evenly spaced discrete frequencies (sidebands) with most of the spectral energy contained within $\pm 10 f_{\mathrm{rep}}$. What should be noted here is the lack of amplitude modulation – the instantaneous intensity of the electric field is constant because the phases of the sidebands are not synchronized.

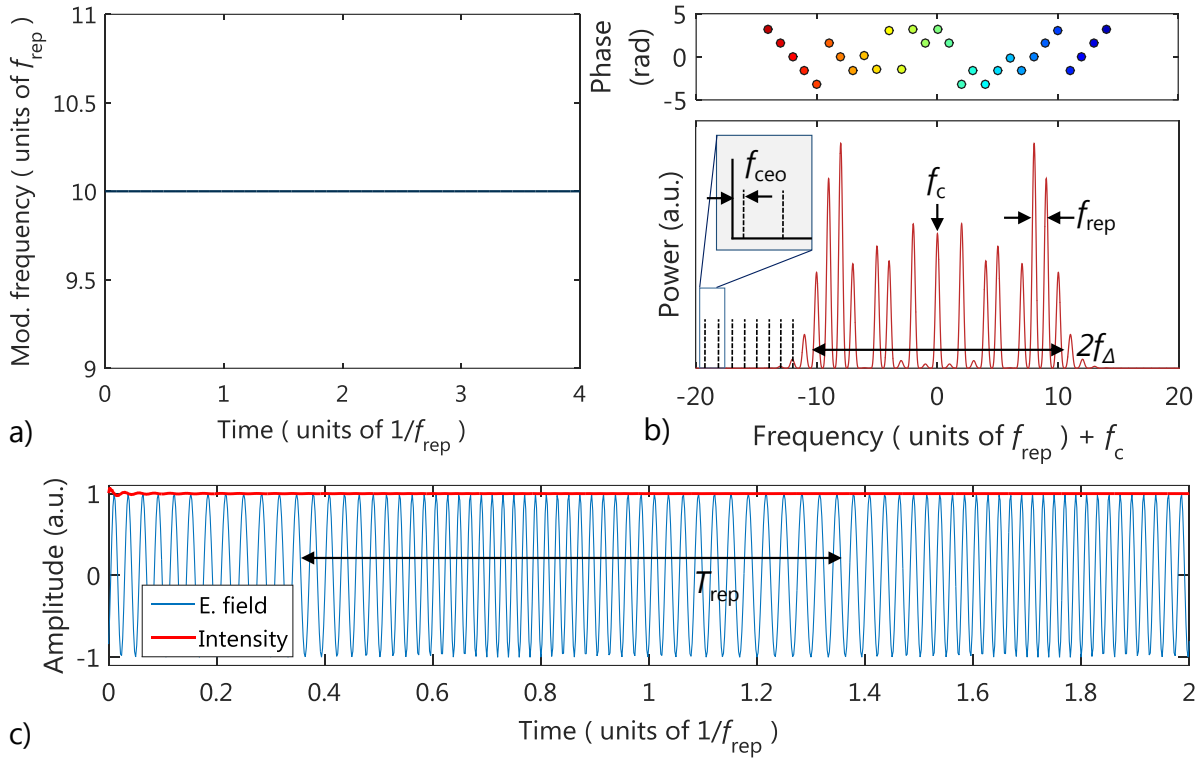

Figure 1.2. Perfect sinusoidal frequency modulation of a carrier wave. a) instantaneous frequency of the modulating signal remains constant b) Symmetric FM spectrum with numerous sidebands and a symmetric but non-linear phase (after unwrapping). Compared to the mode-locked comb, the spectrum shows a large modulation of the envelope. Some tooth almost do not exists, which has many practical implications for spectroscopy. c) The amplitude and intensity of the electric field as a function of time. Note the constant intensity operation. Figure inspired by Ref. [4].

This simple picture serves as an introduction to a more complex model of an FM comb developed within the cavity of a quantum cascade laser. Rather than being modulated by pure sine wave, the carrier will be modulated now by a random-like yet periodic signal. This was proposed and modeled by Khurgin et al. [5] to describe the intrinsic comb operation of a Fabry-Pérot quantum cascade laser. In this model, the time domain profile follows the equation

$$E(t) = E_0 \cos\left(2\pi f_c t + 2\pi f_\Delta \int_0^t x_m(\tau)\mathrm{d}\tau\right), \qquad (1.7)$$

where the shape of the random-like modulation is included in x_m. Figure 1.3 shows a simulation of such an FM comb, where the instantaneous frequency of the modulating signal plotted in a) shows a random-like character, but it is periodic. The Fourier Transform of a single period defines the envelope of the optical spectrum, whereas the repetition frequency defines the comb tooth spacing. The optical spectrum resembles that encountered in semiconductor combs: certain teeth are of high intensity compared to the average, whereas some almost do not exist or are of low intensities. There are many implications of this

undesired behavior in spectroscopy. As in the previous sine-wave modulated case, the phases are not synchronized leading to CW operation. In fact, a true QCL will possess both FM, and residual AM rather than being purely frequency modulated, hence this model should serve for illustrative purposes. Nevertheless, it enables to get a rough picture of the fundamental difference between the two modes of comb operation.

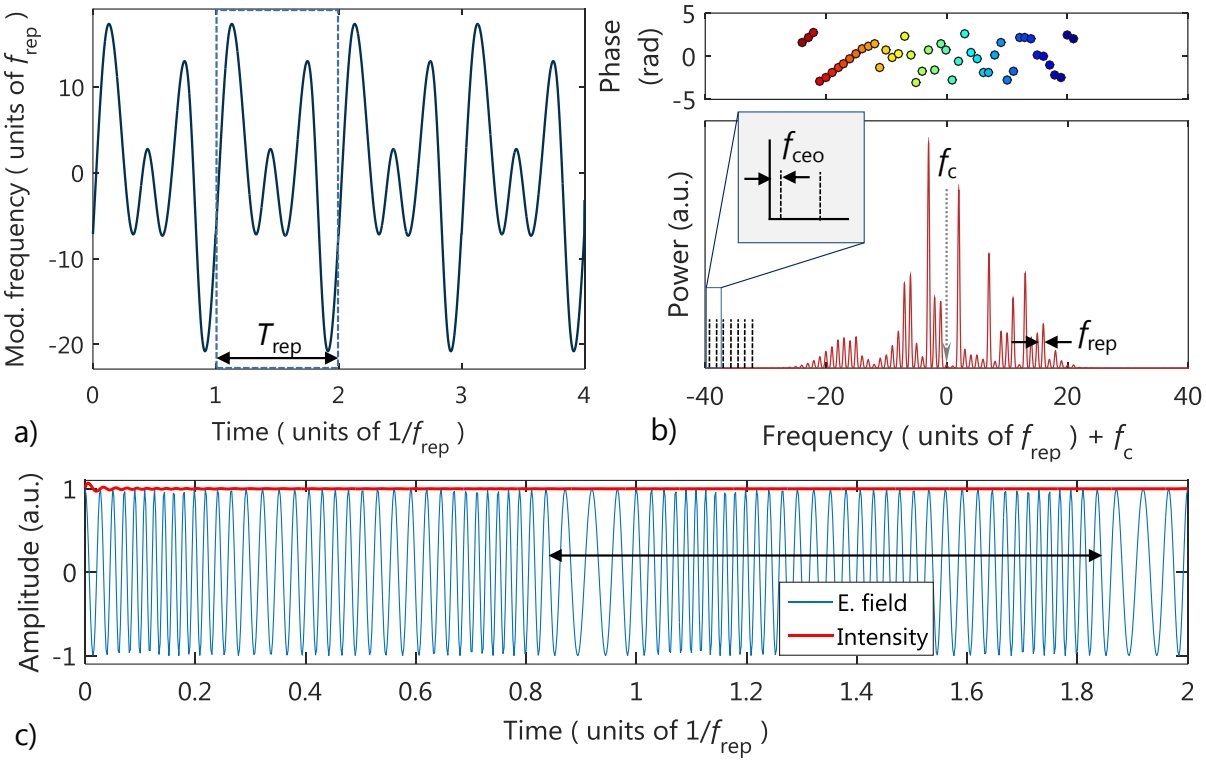

Figure 1.3. Simulation of a periodic random-like FM modulated carrier. a) instantaneous frequency of the modulating signal varies but repeats periodically b) Asymmetric FM spectrum with the numerous sidebands and an asymmetric but non-linear phase (after unwrapping). c) The amplitude and intensity of the electric field as a function of time. Again, one can observe constant intensity operation.

1.2. Thesis statement, goals and organization

This thesis aims to prove the following statement:

Coherent averaging techniques are crucial for reliable characterization of optical parameters of a medium probed spectroscopically with frequency combs.

It is divided into chapters proving the validity of the statement supported by experimental data acquired in two distinct configurations:

- Chapter 2 discusses the use of a single stabilized fiber frequency comb to pump terahertz photoconductive antennas in a THz time domain spectroscopy configuration for gas sensing, and drug spectroscopy applications. It provides a recipe coherent averaging of multiple

acquisitions, while comparing it with the case when signals are simply point-wise averaged. This part was performed at Wroclaw University of Technology in prof. Plinski's Terahertz Laboratory, Wroclaw, Poland.

- Chapter 3 includes the results of research on multiheterodyne and dual comb spectroscopy with small-footprint quantum cascade lasers in the mid-infrared and terahertz regime, as well as the first demonstration of dual comb spectroscopy in the mid-infrared using interband cascade lasers. The newly developed fast computational and hardware solutions to account for the non-stationarity of the sources are presented in details. Without these procedures, dual comb spectroscopic measurements are virtually impossible, and hence the optical parameters of the probed medium. This part of research was conducted in prof. Wysocki's Princeton University Laser Sensing Laboratory (PULSe) at Princeton University, NJ, USA with some elements at NASA Jet Propulsion Laboratory in the Microdevices Laboratory in Dr. Mahmood Bagheri's laboratory in Pasadena, CA, USA.

Chapter 2

THz time-domain spectroscopy using femtosecond lasers and photoconductive antennas

Terahertz time domain spectroscopy (THz-TDS) is one of the most wide-spread spectroscopic techniques in the far infrared region, which gains popularity in the first industrial applications, like detection of foreign bodies in food, characterization of plastic welding joints, and detection of narcotics. Its unique capabilities are broadband operation, high dynamic signal-to-noise ratio, and cryogen-free operation. Another strength of THz-TDS is the possibility of simultaneous measurements of absorption and dispersion spectra without using the Kramers-Kronig relations. Because of its numerous advantages, and the intensive use of THz-TDS to study the spectral behavior of pharmaceuticals and their excipients, we will analyze this technique with a special focus on signal processing path. This is dictated by the fact, that signal levels in terahertz spectroscopy are predominantly weak and require averaging over extended timescales to obtain reliable spectroscopic data. Such an operation, however, may create spectral artifacts and narrow the bandwidth of the terahertz spectrum, if performed incorrectly. Also the spectral phase may be subject to erroneous retrieval. Consequently, it is of utmost importance to establish a reliable way of signal processing that avoids these undesired effects. In this chapter we will first discuss how THz waves are generated in the terahertz spectrometer and what factors affect its performance. Next, we will characterize coherent and incoherent averaging in the time and frequency domain, to finally demonstrate the relevance of the investigated techniques on spectroscopic examples.

2.1. THz generation and detection principles

The generation of terahertz radiation is attainable in many ways depending on the requirements of application regarding power, bandwidth, spectral purity, coherence, and operation temperature. Due to the fact that terahertz waves fill the gap between the optical and radio frequency domains, it is possible to approach them in three different ways: by down-conversion from the optical domain, up-conversion from the microwaves, and direct generation in the terahertz. In this section, we will start by describing the first technique used in all experiments in this chapter, briefly touching the up-conversion approach, while devoting a whole chapter in this thesis to the direct generation technique based on QCLs.

One of the wide-spread optical down-conversion techniques to generate terahertz radiation is the use of femtosecond lasers illuminating ultrafast optical switches. They harness photoconductivity to generate a transient electrical pulse with a large spectral coverage in the terahertz but low power per spectral

element. Due to its cryogen-free operation, this method has gained popularity in the terahertz world, however it is difficult to implement in out-of-laboratory environments. Similar techniques utilize the interaction between high-peak intensity optical pulses and nonlinear crystals to radiate the terahertz envelope carried by the pulse. In linear optics, the electric field **E** induces a polarization **P** in a linear medium

$$\mathbf{P} = \varepsilon_0 \boldsymbol{\chi} \mathbf{E}\,, \tag{2.1}$$

where ε_0 is the electric permittivity of vacuum, and $\boldsymbol{\chi}$ is the electrical susceptibility. On the contrary, in a nonlinear medium the relationship is a series of terms in polynomial expansion [6] with *n*-order electric susceptibility $\boldsymbol{\chi}^{(n)}$ tensors of *n*+1 order

$$\begin{aligned} \mathbf{P} = \mathbf{P}^{(1)} + \mathbf{P}^{(2)} + \mathbf{P}^{(3)} + \cdots &= \varepsilon_0\left(\boldsymbol{\chi} \cdot \mathbf{E} + \boldsymbol{\chi}^{(2)} : \mathbf{EE} + \boldsymbol{\chi}^{(3)} \vdots \mathbf{EEE} + \cdots\right) \\ &= \varepsilon_0 \boldsymbol{\chi}^{(1)} \mathbf{E} + \mathbf{P}_{\mathrm{NL}}, \end{aligned} \tag{2.2}$$

In the above equation, the first term (linear) describes optical properties such as refraction and absorption, whereas higher order nonlinear terms denoted for simplicity as $\mathbf{P}_{\mathrm{NL}}$ are the domain of nonlinear optics, responsible for instance for frequency mixing processes or optical rectification, all of wide use in THz spectroscopy.

In terahertz generation, one takes advantage of difference frequency generation (DFG), which is a second order nonlinear effect. A broadband ultrafast pulse is focused onto a nonlinear crystal that mixes all frequency components within the optical spectrum. The theory of mode-locked lasers states that an optical pulse with a duration $\Delta\tau$ and optical bandwidth $\Delta\omega$ obeys the relation:

$$\Delta\omega\Delta\tau \geq \frac{1}{2}. \tag{2.3}$$

It is clear from the above equation that short pulses have a broad optical bandwidth. In particular, sub-picosecond pulses obtainable with mode-locked near-infrared lasers have THz-wide bandwidths, hence it would be desired to convert that spectral envelope into a THz field. To explain this phenomenon in terms of nonlinear optics, consider an electric field composed of two arbitrary frequency components of the near-infrared laser oscillating with frequencies ω_1 and ω_2. i.e. $\mathbf{E}(t) = E_1 \sin\omega_1 t + E_2 \sin\omega_2 t$. The second order nonlinear polarizability in general is given by [7]

$$\begin{aligned} \mathbf{P}_{\mathbf{NL}}^{(\mathbf{2})} = \varepsilon_0 \boldsymbol{\chi}^{(\mathbf{2})} \Big[&\frac{1}{2}E_1^2(1 - 2\cos 2\omega_1 t) + \frac{1}{2}E_2^2(1 - 2\cos 2\omega_2 t) + E_1E_2\cos(\omega_1 + \omega_2)t \\ &- E_1E_2\cos(\omega_1 - \omega_2)t\Big]. \end{aligned} \tag{2.4}$$

In the above we can see several terms. The first two second harmonic components $2\omega_i$ are responsible for frequency doubling, the next one describes sum frequency generation $\omega_1 + \omega_2$, and the last one stands for difference frequency generation $\omega_1 - \omega_2$ responsible for the capability of producing THz radiation. The DFG term is next radiated with a field intensity dependent on the geometrical properties

of the crystal and the effective second order susceptibility Notably, there are also two DC terms that constitute optical rectification a closely related effect inducing a bifringence in a non-linear crystal by an external THz electric field, referred to as the Pockels effect (linear electro-optic effect), which is used for detection of ultrafast terahertz pulses through balanced detection. In other words, the same nonlinear crystal can be used to coherently detect THz radiation too but using a different term of the nonlinear polarizability.

2.1.1. Photoconductive generation

Due to the importance of photoconductive generation in terahertz spectroscopy, this technique will be discussed in more detail here.

A photoconductive antenna is essentially an ultrafast optically-activated switch with an antenna structure (usually a Hertzian dipole) to radiate the generated electromagnetic pulse into the far field, which determines its resonant frequency and quality factor. It comprises a semiconductor substrate with a metallic antenna pattern deposited on its surface (Figure 2.1). There are two major semiconductor materials covering the ~750-850 nm, and the most popular telecom wavelength range of 1550 nm, being the low temperature-grown gallium arsenide (LT-GaAs), and low-temperature grown indium gallium arsenide (LT-InGaAs), respectively. The reason for their growth at low temperatures is to introduce defects in the crystal lattice to improve their carrier dynamics. For instance, regular temperature GaAs (RT-GaAs) grown at around 600°C has a carrier lifetime on the order of nanoseconds [8] in contrast to LT-GaAs grown at 200-300°C with a picosecond lifetime.

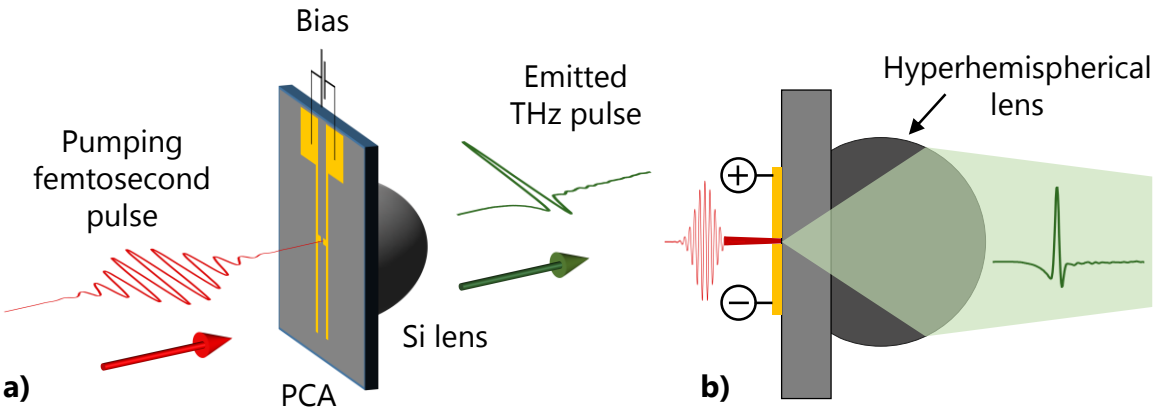

Figure 2.1. Generation of terahertz radiation using a photoconductive antenna. (a) isometric view (b) top view. Figure inspired by Ref. [9].

The previously introduced equation for the pulse width – spectral width product implies that terahertz waves correspond to pulses with picosecond and sub-picosecond timescales. Other materials for substrates are also in use: semi-insulating GaAs (SI-GaAs), indium phosphide (InP), and radiation-damaged silicon-on-sapphire (RD-SOS) to name a few.

To reduce parasitic effects, the antenna's arms acting also as biasing pads are kept close to each other, usually separated by micrometers. The tight gap formed between them is irradiated with a tightly focused femtosecond pulse whose wavelength matches the bandgap of the semiconductor. Ultrafast illumination gives rise to a rapid increase in the number of electron-hole pairs, which rapidly recombine. The time-dependent current density $J(t)$ in the semiconductor can be expressed as

$$J(t) = N(t)\mu E_B \,, \tag{2.5}$$

where μ is the mobility of charge carriers, E_B is the applied bias, and $N(t)$ is the number of charge carriers. Since holes in the material of interest have low mobility, the dynamics of the electrons is sufficient to explain the generation of the ultrafast transient. The electric field of the emitted pulse depends on the rate of change in the current density $J(t)$, which should be as steep as possible

$$\boldsymbol{E}_{\mathrm{THz}} \propto \frac{\mathrm{d}J(t)}{\mathrm{d}t}. \tag{2.6}$$

From the above equation, it is also straightforward that the electric field amplitude of generated pulses increases linearly with applied bias, which is constrained by the voltage withstand of the antenna and self-heating effects caused by the high optical fluence (laser energy per focal spot area). The relationship between the optical power and the number of carrier carriers is more complex. Although initially linear, it saturates due to screening and temperature-related effects. Radiation from a large-aperture photoconductive antenna is modeled using the scaled optical intensity F/F_0 [10].

$$\boldsymbol{E}_{\mathrm{THz}} \propto \frac{F/F_0}{1 + F/F_0}. \tag{2.7}$$

The dynamics of the photoconductive THz generation process is in general complex. The photocurrent density in the semiconductor does not follow exactly the envelope of the stimulating pulse (which is modeled using the Drude-Lorentz theory) but has a delayed and time-stretched response. An example THz pulse together with a numerically evaluated photocurrent is plotted in panel (a) of Figure 2.2, whereas the corresponding power spectrum of the THz pulse proportional to the derivative of the photocurrent is plotted in panel (b).

The PCA can be alternatively seen as a multi-frequency photomixer. When stimulated by a broadband femtosecond pulse, the antenna radiates a signal being the result of mixing between all frequencies under the spectral envelope. If the same device is irradiated with two high-spectral-purity single mode DFB lasers detuned in frequency by several nanometers, it will radiate the difference frequency based on the same photomixing effect belonging to the domain of nonlinear optics (difference frequency generation – DFG).

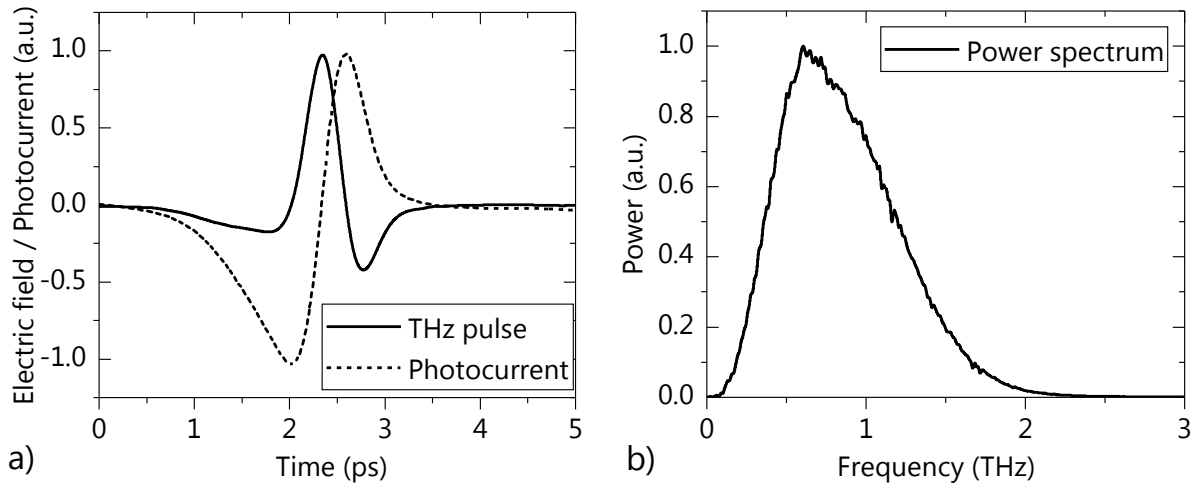

Figure 2.2. (a) THz pulse in a THz-TDS spectrometer together with a numerically evaluated photocurrent. (b) Power spectrum of the THz pulse in (a).

This principle is used in CW generation of terahertz waves and allows for fully tunable terahertz sources suitable for high resolution spectroscopy, however the efficiency of photoconductive switches in photomixing is usually lower [11]. Two example PCA geometries suitable for photomixing applications are shown in Figure 2.3.

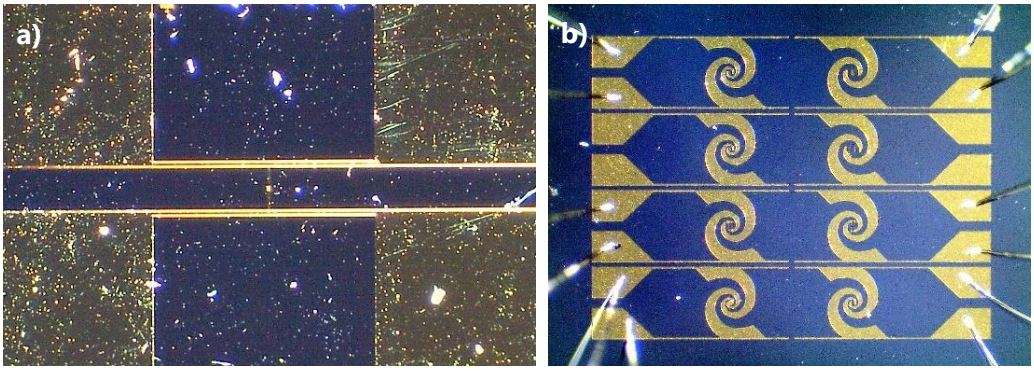

Figure 2.3. Different geometries of THz-PCA. (a) Hertzian dipole fabricated by Dr. Martin Miculics of Forschungszentrum Jülich (b) Log-spiral antenna with plasmonic-enhancement provided by prof. Mona Jarrahi's group of University of California Los Angeles (UCLA).

What is also interesting is the highly polarized emission of the THz pulse, which is parallel to the flow of carriers in the bias field [12]. This feature allows for spectroscopic studies with polarization-sensitive techniques [13], such as Faraday rotation spectroscopy (FRS). Among its many advantages, the sensitivity only to paramagnetic molecules allows to deal with selected spectrally overlapping species and eliminates the need for reference measurement thanks to its absolute nature.

2.1.2. Photoconductive detection

Terahertz photoconductive antennas are reciprocal devices, meaning that the same physical phenomena responsible for the generation of THz signal can be used for its detection. The probing femtosecond laser pulse generates photocarriers, which are modulated by the external terahertz field, allowing for coherent detection of the incident THz pulse. The main difference between the two modes of operation is that the antenna is left unbiased (see Figure 2.4). To measure the photocurrent proportional to the amplitude of the external electric field, one connects external circuitry, i.e. transimpedance amplifier (TIA). In the detection mode, the probed THz pulse is temporally overlapped with the near infrared femtosecond pulse and only a small part of the time-domain waveform is detected in a "single shot". To reconstruct the entire shape of the THz electric field, one needs to vary the delay between the probed and probing pulse, which is conventionally implemented using a linear stage and a retroreflector, similar to a typical interferometer. This process is shown schematically in Figure 2.5.

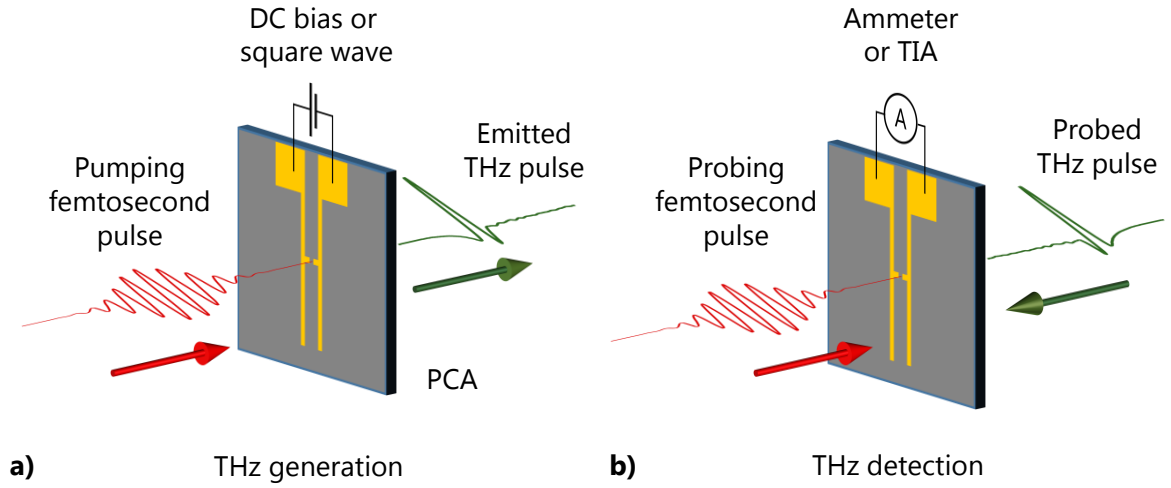

Figure 2.4. Side-by-side comparison of terahertz generation (a) and detection (b) using photoconductive antennas. The width of the femtosecond pulses is not to scale, while the polarization of the generated THz radiation is preserved (parallel to the bias field).

The top panel of Figure 2.5 shows a single THz pulse with an arbitrarily chosen time instant τ_0, which denotes the beginning of the waveform. Due to the ultrashort duration of the probing pulse (single femtoseconds to tens of femtoseconds) compared to the generated THz one (near-picosecond), the reconstruction process can be described in a simple model through the sifting (sampling) property of a time-delayed Dirac delta (shown as dashed line in the figure):

$$E_{\mathrm{THz}}(\tau) = \int_{-\infty}^{\infty} E_{\mathrm{THz}}(t)\delta(\tau - t)\mathrm{d}t. \qquad (2.8)$$

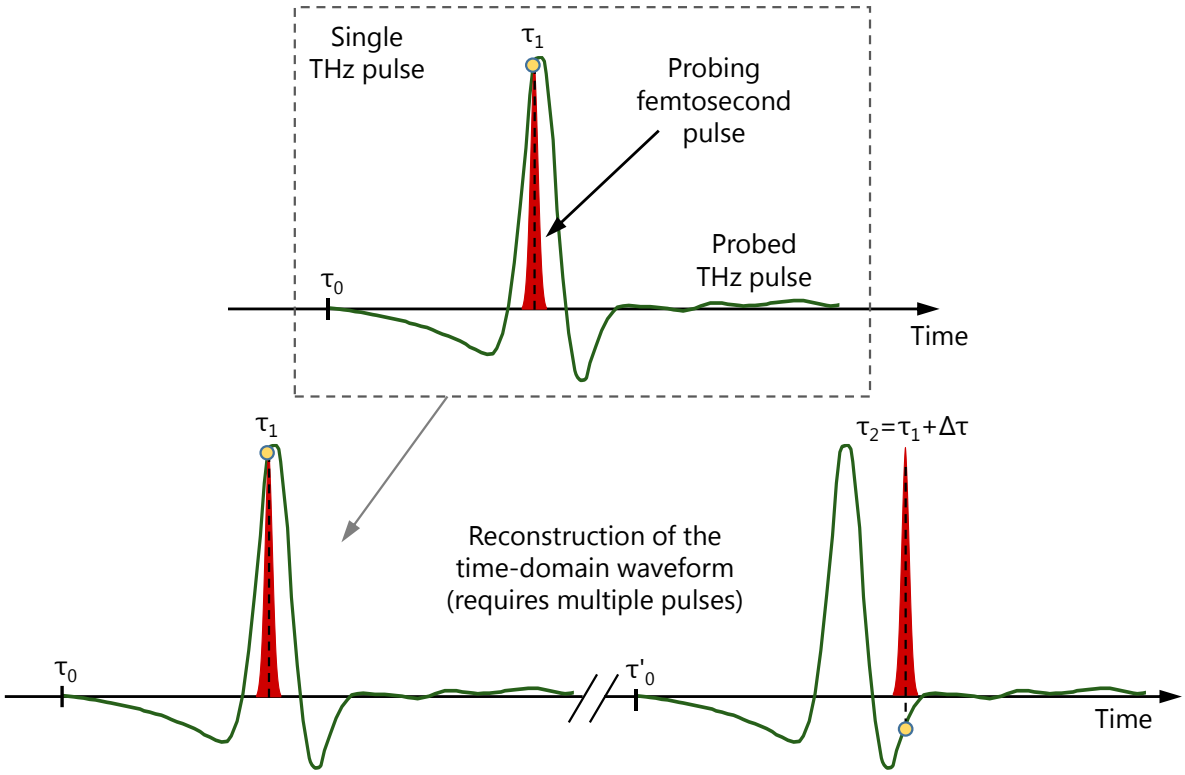

Figure 2.5. Reconstruction of the time-domain pulse obtained through varying the delay between the probing NIR pulse and the detected THz pulse. The top part of the figure shows that within a "single shot" only a small part of the waveform is probed. To reconstruct the entire waveform, the pulse is delayed in steps to repeatedly probe multiple THz pulses arriving every roundtrip time of the fs laser.

In other words, the Dirac delta samples a single point from the THz electric field. By varying the relative delay between the Dirac delta and the THz field, we can reconstruct its time-domain profile. In practice, however, the Dirac delta is physically not realizable therefore to describe the detection process, we need to express the measured photocurrent in terms of non-ultrafast dynamics of photoconductivity σ_{PC} by employing a convolution integral [9], [12]:

$$J_{\mathrm{THz}}(\tau) = \int_{-\infty}^{\tau} E_{\mathrm{THz}}(t)\sigma_{\mathrm{PC}}(\tau - t)\mathrm{d}t, \qquad (2.9)$$

where J_{THz} denotes the induced photocurrent at time delay τ. Since convolution in the time domain is equivalent to multiplication in the spectral domain, the true spectrum $\tilde{E}_{\mathrm{THz-tr}}(\omega)$ is multiplied by the Fourier transform of the photoconductivity:

$$\tilde{E}_{\mathrm{THz-m}}(\omega) = \tilde{E}_{\mathrm{THz-tr}}(\omega)\mathcal{F}[\sigma_{\mathrm{PC}}(\omega)]. \qquad (2.10)$$

Intuitively, this will have a tremendous effect on the detected signal if a photoconductive antenna detects pulses much shorter than those created by the antenna itself in generation mode. Such can be obtained for instance by inducing a mJ-laser-driven electrical breakdown in the air to create ultrafast plasma, with bandwidths exceeding 10 THz, far beyond the capabilities of most PCAs. Another implication of the

convolution in the equation is that the spectral performance of any THz-TDS system will be limited by the performance of the slower of the generating and receiving element.

During the detection process, multiple pulses are measured, all with the assumption of almost perfect reproducibility. The bottom panel in Figure 2.5 shows the mechanical scan reconstruction process of the THz pulse. For instance, 100 million pulses contribute to a time domain trace captured within one second. This process relies on the high coherence of the source and the stability of the repetition rate, and naturally rejects any background THz radiation incoherent with the femtosecond laser (optical gating), which allows for room-temperature acquisitions with a high dynamic range and signal-to-noise ratio.

Recently, a moving-parts-free implementation of the above scheme has emerged, known as Asynchronous Optical Sampling (ASOPS) [14], [15]. It is a configuration that utilizes two femtosecond lasers (one for THz generation and one for detection) with the possibility of millisecond time acquisition and insensitivity to errors introduced by the mechanical stage. This technique will be described in detail in the next chapter, since it belongs to the family of dual-comb techniques, but shortly speaking in such systems the delay between the generated and probing pulse is a consequence of different repetition rates of the lasers.

2.1.3. Terahertz time-domain spectrometer

Equipped with the basic knowledge of how THz generation and detection works, we can study the operation principle of a THz time domain spectrometer, shown in the figure below (Figure 2.6). This will help us better understand the origin and nature of errors we will try to eliminate through signal processing. Since the same laser is used for detection and generation of the THz radiation, and the optical phase delay between the transmitter and the receiver is effectively set to utilize the same laser pulse,

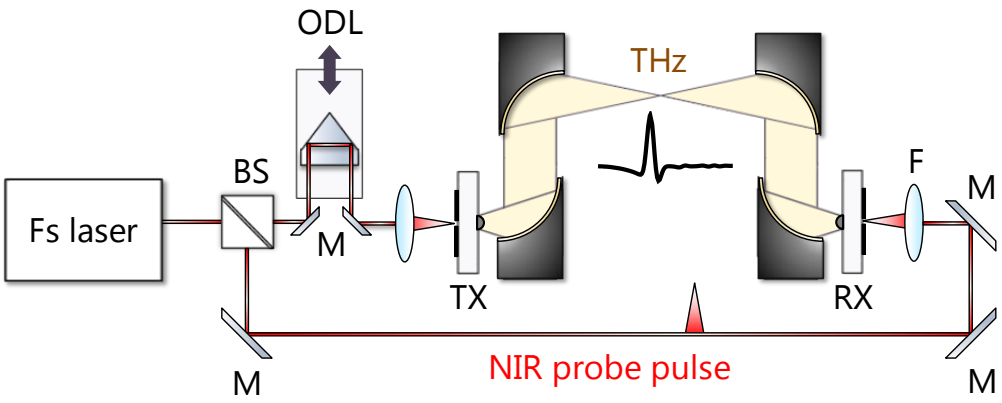

Figure 2.6. Schematic of a terahertz time domain spectrometer. BS – beam splitter, M – mirror, ODL – optical delay line, TX – transmitting photoconductive antenna, RX – receiving photoconductive antenna, F – focusing microscope objective.

the laser's pulse-to-pulse instabilities will not corrupt spectroscopic signal as much as in the case of two unsynchronized continuous-wave lasers discussed in the next chapter. Nevertheless, some amounts of both amplitude and phase noise persist in the acquired signals, both requiring efficient suppression.

A near-infrared mode-locked femtosecond laser (Fs laser), which in the spectral domain is a highly stable frequency comb, produces ultrafast pulses split into a pumping and probing arm using a beam splitter (BS). The length of one of the arms is changed using an optical delay line (ODL) by varying the position of a retro reflector mounted on a precision linear stage. Two microscope objectives (F) focus the ultrafast light onto the gap between the dipole of the transceiver (TX) and receiver (RX) antennas, which are mounted on high-resistivity float zone hyper-hemispherical lenses (HRFZ-Si). Next, four off-axis parabolic mirrors guide the THz light produced by the TX, and focus it onto the back side of the RX antenna. To detect the peak of the THz pulse, the total optical path on the transceiver side (NIR + THz) must be equal to the receiver side (NIR), hence the refractive indices and the beam propagation distances within the THz lenses, NIR lenses, and the PCA substrates must be taken into account as well compared to just ensuring geometrically-equal paths.

Since the received THz signal in a TDS spectrometer is extremely weak, and yields a photocurrent on the order of nanoamperes, it is very susceptible to electromagnetic noise. Therefore, a lock-in detection scheme is used, thereby turning TDS into a phase-sensitive homodyne receiver. The principle of its operation is shown in Figure 2.7. The transceiver is modulated either through an optical chopper or an electronic biasing circuit (described further) with the chopping frequency acting as a local oscillator (LO). When the detected signal is mixed in quadrature with the LO using a lock-in amplifier, it produces a complex signal with the two components: X and Y, or more commonly denoted as I (in-phase) and Q (quadrature). This homodyne detection scheme utilizing the LO from the same source as the signal is in stark contrast to the heterodyne receiver, where the LO comes from an external source.

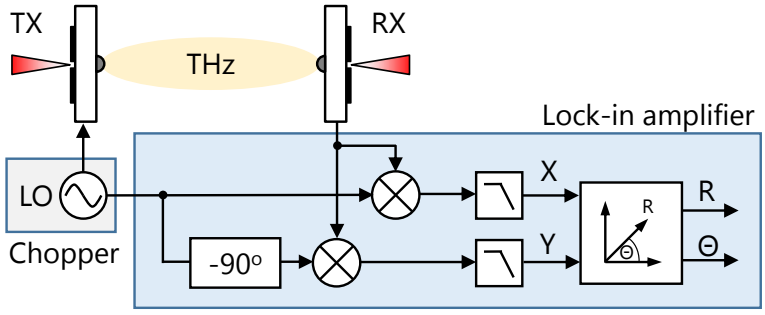

Figure 2.7. Simplified scheme of a THz homodyne receiver. The reference signal from the chopper driving the TX antenna is connected to the two-channel lock-in amplifier, which mixes the detected signal from the receiving antenna with an in-phase and in-quadrature version of the reference signal. This results in a complex signal, further transformed into polar form.

The use of a two-channel lock-in amplifier eliminates the sensitivity of the receiver to reference phase misalignment, which in extreme cases (90-degree shift) leads to zero output. To retrieve the time domain profile of the electric field, one transforms the complex vector [X,Y] measured for each position of the delay line into polar coordinates, yielding a signal magnitude

$$R = \sqrt{X^2 + Y^2}, \tag{2.11}$$

and phase

$$\Theta = \tan^{-1}\left(\frac{Y}{X}\right). \tag{2.12}$$

Finally, to obtain the time-domain trace, the two entities are plugged to the formula

$$x(t) = R\cos(\Theta), \tag{2.13}$$

which represents the profile of the THz electric field, albeit accurate to an arbitrary constant depending on many experimental factors, like detector responsivity, coupling efficiency, internal settings of the amplifier etc.

2.1.4. Experimental implementation

Nearly all THz-TDS experiments in this dissertation were performed using a home-built terahertz spectrometer. Two high-bandwidth photoconductive antennas (Fraunhofer, Tera8-1) were biased at 30 V using a square wave biasing circuit flipping the polarity with a rate of 20 kHz (described further). The antennas were pumped using a commercially-available C-Fiber 780 Erbium-based femtosecond laser with ~200 mW of output power equipped with a second harmonic generation crystal to produce a 780 nm output attenuated to ~10 mW before shining onto the photoconductive antennas. The average photocurrent on the generating antenna at 30 V bias did not exceed 50 μA. The repetition rate of 85 fs wide pulses was internally stabilized using the laser's control electronics at 100 MHz. The THz-modulated photocurrent in the receiving antenna was fed into a Stanford Research Systems lock-in amplifier (SRS830), operating in transimpedance mode followed by numerous built-in low pass filters. The samples were acquired from the lock-in's output buffer via a GPIB interface card while the optical delay line comprising a precision linear stage (Zaber T-LSM200A) stepped the retroreflector.

2.1.5. Biasing techniques

Due to the low separation distance of the electrodes on an LT-GaAs substrate on the order of micrometers, the maximum applied voltage across the electrodes of the PCA, as well as the amount optical power is limited, and if exceeded, it can cause a severe electrical breakdown, as shown in Figure 2.8. Consequently, special care must be taken when working with photoconductive switches: the voltage and optical power must always lie in the safe operating area (SOA), mainly determined by the substrate's thermal properties. Furthermore, static electricity and transient voltages can contribute to a failure of a photoconductive switch too. Such states can arise dynamically, so to account for them an isolated

electronic chopper circuit has been designed by the author [16]. Its full implementation details can be found in the referenced publication, but in short it is a stable square wave voltage generator with an adjustable frequency up to 200 kHz, and a fixed galvanically-separated high voltage amplifier following the generation stage with a maximally fast rise time and almost no output voltage overshoot.

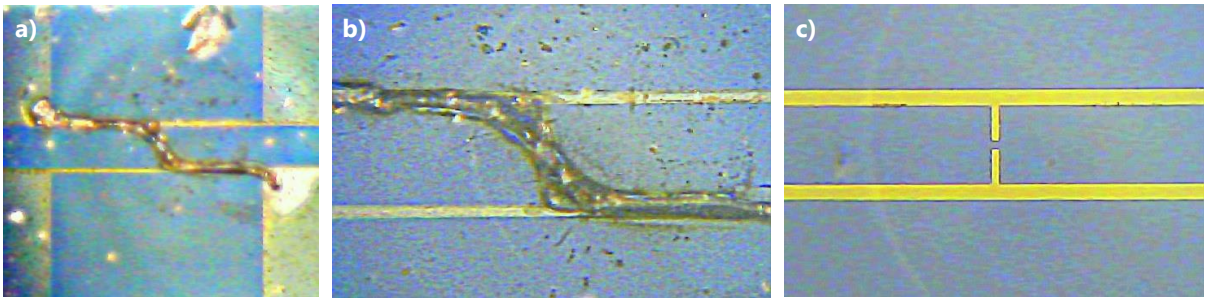

Figure 2.8. Electric breakdown across the arms of the photoconductive antenna caused by overvoltage, likely to be attributed to transient spikes during switching. a) Overview photo. The surface discharge shorted the dipole diagonally. b) zoom into the central part showing the path of the breakdown. c) non-damaged device.

The DC shorted low-resistance secondary side of the transformer protects the antenna against electrostatic discharge, as long as reliable contact between the device and the antenna is provided. In addition to providing the fast chopping speed (ten to hundred times faster), the chopping circuit has several significant advantages over the conventional mechanical chopping wheel. Not only does it not create any mechanical vibrations that would misalign the sensitive optical elements causing a modulation of the photocurrent, but also it boosts the signal-to-noise ratio virtually by a factor of two. With the mechanical chopper, half of the femtosecond pulses is skipped due to the blocking of the near-infrared laser. This contrasts with the electronic chopper, which just flips the bias field, and thereby

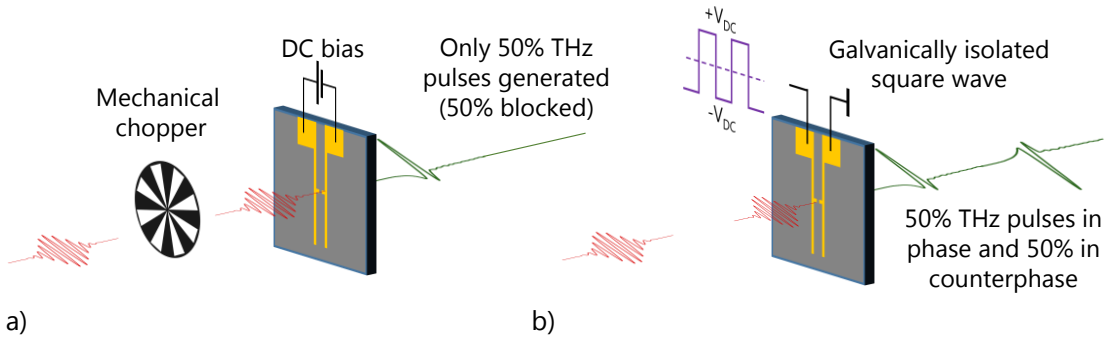

Figure 2.9. Comparison of different biasing techniques for the generation of THz radiation using a photoconductive antenna. a) Unipolar DC bias with a mechanical chopper, which skips 50% of the femtosecond pulses. b) bipolar high voltage square wave generator with galvanic isolation (coupled through a transformer).

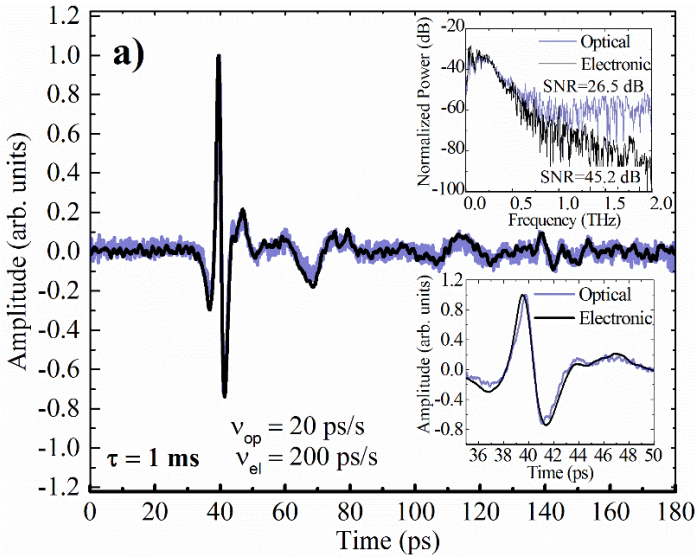

Figure 2.10. Comparison of the two chopping schemes: mechanical (optical), and electronic at 1 ms of time constant. An improvement in scanning velocity, and SNR is visible. Results published in [16].

the phase of the THz pulses without blocking the laser. A comparison of the two techniques is shown in Figure 2.9 and Figure 2.10.

The non-mechanical chopping additionally allows for faster and spectrally broader acquisitions. By increasing the chopping speed, using all pulses, and eliminating mechanical vibrations, we could increase the optical delay line speed from the initial 20 ps/s to 200 ps/s (maximum for the used linear stage), which was advantageous in our time-consuming pulsed imaging experiments [17]. Also the signal to noise increased by almost 20 dB, enabling to higher frequencies above 1 THz, previously buried in noise. Why is the use of the right biasing scheme important? In the stability analysis section we will show that THz pulses can be averaged over relatively short timescales due to a gradual drift of the setup. By using the bipolar square wave high frequency chopping, we get considerably better SNR within the same acquisition time, hence the optimally averaged spectrum will have lower uncertainty, which translates into more reliable characterization of the optical medium.

2.2. Acquired signal processing path

So far, we have investigated the effects of antenna biasing on the quality of THz signals. Unfortunately, THz pulses acquired in a terahertz time-domain spectrometer are prone to a larger number of quality-deteriorating factors. Some of the most important are related to the mechanical nature of the acquisition: slight jitter in absolute position on the pulse, and a variation of the pulse shape from scan to scan. While single-shot acquisition of terahertz pulses is possible [18], for practical reasons a reliable signal processing path must have been developed that enables averaging of multiple traces acquired over longer timescales to improve the spectral signal-to-noise ratio, especially at high frequencies.

2.2.1. Coherent averaging of time-domain THz pulses

It has been observed that one of the dominant problems with THz-TDS is the non-reproducibility of the position of the time-domain trace [12] introducing an error in the spectral phase. This of course precludes reliable estimation of the refractive index and absorption coefficient. The well-known shift property of the Fourier Transform relates the two domains, namely:

$$\mathcal{F}[x(t-t_0)](\mathrm{j}\omega) = \mathrm{e}^{-\mathrm{j}2\pi\omega t_0}F(\mathrm{j}\omega), \qquad (2.14)$$

where j is the imaginary unit such that $\mathrm{j}^2 = -1$. Simply speaking, a shift in time corresponds to a shift in spectral phase with a multiplicative character. Multiplicative noise cannot be eliminated through simple point-wise averaging, therefore special procedures must be applied to account for it, and will be described here. Furthermore, together with the multiplicative phase noise, additive amplitude noise coexists, which can be effectively suppressed by averaging either in the time, or spectral domain, as long as all signal-processing steps are linear (*i.e.* the signal is not squared or log-transformed, which would rectify the noise).

For the sake of simplicity, and the insignificant effect of the delay line nonlinearity in our system, let us consider just the numerical compensation of the time-domain trace position uncertainty here. The next step to improve the accuracy of a sampled THz pulse is to implement much more complex acquisition schemes using a stabilized reference HeNe laser and a fringe-triggered acquisition system [19].

To align similar signals with a timing jitter, the well-known cross-correlation technique was used. First, a reference signal must be defined. It can be for instance either the first acquired pulse, or a member of a set of pulses that had the largest cross-correlation coefficients within the dataset. To find the time delay between two signals, we search for the maximum of the cross-correlation function

$$\tau_\mathrm{d} = \underset{t}{\mathrm{argmax}}\big((f \star g)(\tau)\big), \qquad (2.15)$$

where the cross-correlation function as a function of lag τ is given by

$$(f \star g)(\tau) = \int_{-\infty}^{\infty} f^*(t)g(t-\tau)\mathrm{d}t. \qquad (2.16)$$

Obviously, in the discrete case integration is replaced by summation over samples. Having found the time delays with respect to the reference, the signals are shifted and cropped to the common overlap limits.

To quantify the severity of jitter and amplitude fluctuations, we will use the well-known Allan-Werle deviation [20], [21], used initially to characterize the stability of atomic clocks, or frequency of oscillators in general, but easily extendable to any physical quantity like amplitude or position. The Allan deviation has a direct link with the power spectral density of noise characteristics [22] and is commonly employed to detect the limits of averaging. In a typical spectroscopic system, samples are

acquired within discrete time intervals, equal to the sampling period. Having acquired *N* discrete samples, we can divide them into *M* subsets with *k=N/M* samples in each a multiple of the sampling period τ_s. Next, for each averaging period $k\tau_s$ we can calculate the Allan variance [23]:

$$\sigma_y^2(k\tau_s) = \frac{1}{2(M-1)} \sum_{i=0}^{M-2} (\bar{y}_{i+1} - \bar{y}_i)^2, \quad (2.17)$$

where the mean of each subset containing measured samples x is calculated according to

$$\bar{y}_i = \frac{1}{k} \sum_{j=0}^{k} x_{i \cdot k+j}. \quad (2.18)$$

Finally, the Allan deviation is derived from the Allan variance by taking its square root, by the same way as in the ordinary standard deviation.

$$\sigma_y(k\tau_s) = \sqrt{\sigma_y^2(k\tau_s)}. \quad (2.19)$$

An Allan plot is a log-log representation of the Allan deviation, where the averaging time $k\tau_s$ is the independent variable. In a typical Allan plot, three distinct regions can be identified: one that is characterized by a random (white-like) nature of fluctuations, one that shows oscillatory behavior, and another that describes a drift. Most importantly, the minimum of this plot determines the optimal conditions for acquisition, which is critical for reliable retrieval of the optical parameters of the probed medium. We will identify and describe those regions on experimental THz-TDS data of a 10% α-D-glucose dehydrate pellet acquired over 800 s with a continuously running optical delay stage, and characterize the effects of averaging on the quality of spectroscopic signals in the time and frequency domains.

Figure 2.11 plots the data in two columns: left, where the signals have not been aligned, and right with the cross-correlation-based alignment. Prior to analysis, the time domain traces were normalized by the peak value of the dataset, corresponding to scaling by an arbitrary factor, yet identical for all the traces to allow for non-biased amplitude stability studies. The top panels include a fan chart of the normalized TDS data with the mean trace plotted with a black solid line. We can clearly observe that without alignment, the mean trace loses its peak amplitude to approximately 60% and broadens its width by more than 40%.

The latter is strongly undesired, since it will limit the bandwidth of the THz system and will smear out the oscillations in the signal tail following the main pulse, thereby giving rise to unwanted artifacts. In contrast to the ordinary point-wise averaging, the aligned traces have a low spread in the position, and the outline of the stacked traces is very thin, hence the coherently-averaged trace follows the individual traces accurately. This will obviously facilitate more reliable characterization of the sample. To compare the two cases in a more analytical way, we will harness the aforementioned Allan-Werle deviation,

starting with the analysis of the peak amplitude, but in principle any arbitrary point of the trace can be selected. The non-aligned case shows a decreasing trend with a global peak amplitude deviation minimum of 2×10^{-2} at 400 s, albeit with some oscillations.

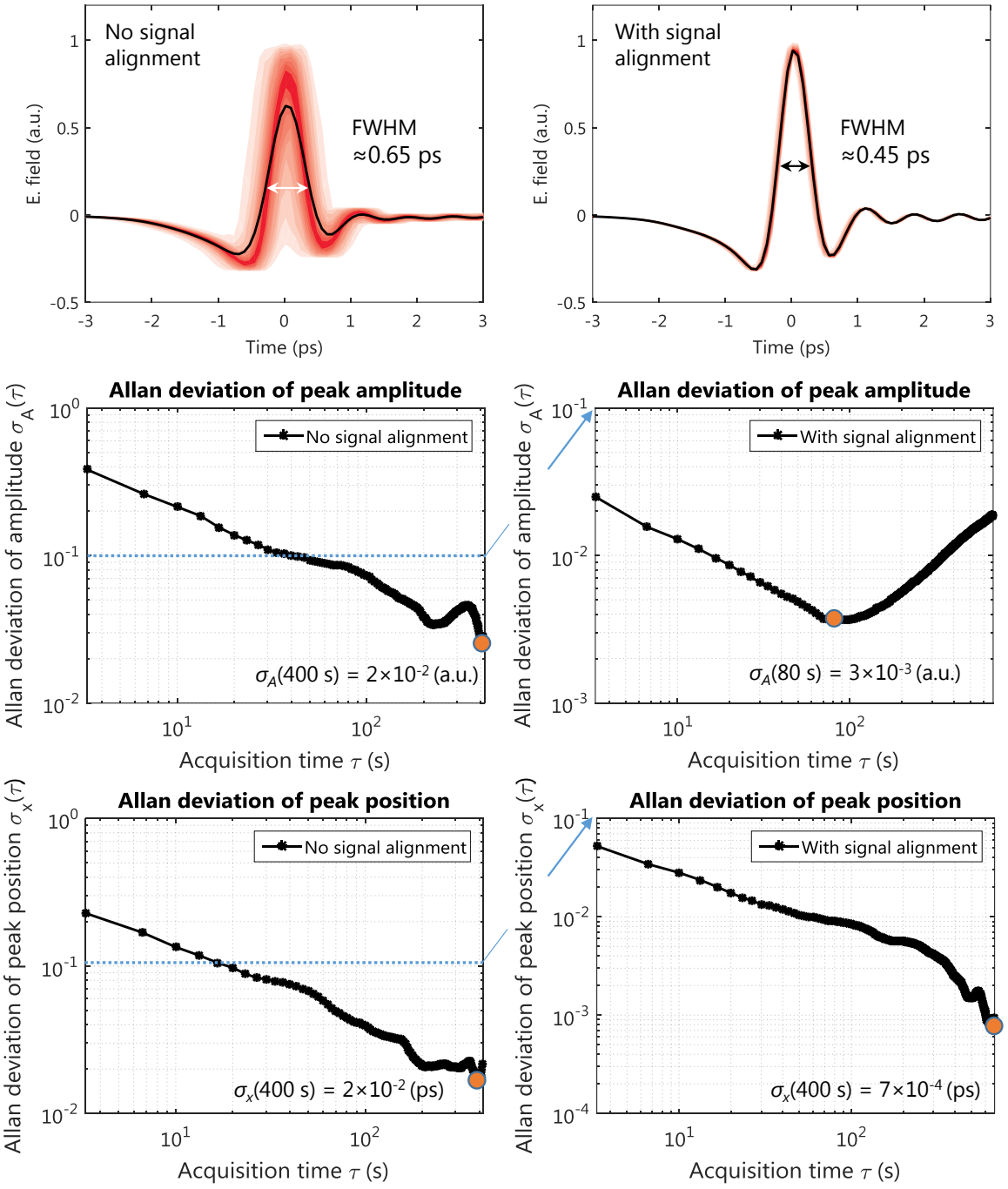

Figure 2.11. Effect of incoherent (left) and coherent (right) averaging in the time domain. The top panels show time-domain traces together with the mean trace plotted with a solid black line. The middle panels show the Allan deviation of the peak amplitude, whereas the bottom ones deviation of the peak position.

The coherently-averaged case on the other hand shows a considerable improvement in that respect, lowering the peak amplitude deviation by a factor of 15 within a 5 times shorter time. What could not

be seen before, is the limit of averaging defined by the global minimum followed by a linear slope after 80 s. We clearly see that further averaging will not lower the deviation in amplitude, revealing the system drift on a minute scale.

The reason of that is closely related to the free-space nature of the system: the femtosecond laser is focused on the antennas using free-space lenses and precision mechanical stages with a micrometer accuracy. A change in the position of the focused beam by a micrometer will noticeably affect the power produced by the THz emitting antenna, further worsened by an identical phenomenon on the detection side. Such drift time scales usually correspond to temperature fluctuations, and to minimize their effects, it would be desired to fiber-couple the antennas, use opto-mechanical elements with low temperature coefficients (invar or similar alloys or zero-dur for optics), and stabilize the laboratory temperature with greater accuracy.

An alternative explanation involves fluctuations of the water vapor concentration in the measurement chamber, which is one of the strongest THz gas absorbers. Throughout the acquisition, the system was constantly purged with dry air, however no feedback loop was used to keep the humidity at a constant level. Its slightest change will affect the amplitude of the THz signal, thereby limiting long-term averaging capabilities.

The bottom panels plot the results of an identical analysis performed for the peak position deviation, which translates into phase errors. In the non-aligned case, we observe a global minimum of the peak position deviation equal to 20 fs after 400 s, again preceded by some oscillations on a minute scale. Thanks to coherent averaging, the peak position deviation can be lowered to 0.7 fs or even more over extended timescales, however the previously observed amplitude drift suggests acquisitions no longer than 80 s, corresponding to an error in the peak position of 9 fs.

The above considerations provide an insight into the stability of a THz-TDS system in the time domain and help identify the possible sources of limited averaging capabilities. In the next subsection, we will see how averaging both in the time and frequency domain affects the resulting frequency spectrum – a preferred representation from the spectroscopic standpoint. Let us start with several important definitions of spectral transformations.

2.2.2. Fourier Transform of time-domain trace

To obtain a complex frequency spectrum $X(\omega)$ from a continuous time domain pulse $x(t)$, one applies the Fourier transform

$$\mathcal{F}[x(t)] = X(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t, \qquad (2.20)$$

where the angular frequency ω is related to the ordinary frequency by $\omega = 2\pi f$. In practice, however, signal $x(t)$ is sampled at discrete time intervals as $x_1, x_2, \ldots x_N$ with a sampling frequency f_s and the

continuous transform does not apply here. Consequently, a discrete Fourier Transform (DFT) is used instead

$$X_k = X(\omega_k) = \sum_{n=0}^{N-1} x_n \cdot \mathrm{e}^{-\mathrm{j}2\pi kn/N}, \qquad k = 0,1,\dots,N-1 \tag{2.21}$$

where k is the DFT bin number spaced from the neighbor by $\Delta\omega = f_s/N$. Obviously, the signal must be sampled at least twice as fast as its highest frequency component (Nyquist criterion) to avoid spectral aliasing. Otherwise, spectral components located at frequencies greater than half of the sampling frequency will overlap with the lower-frequency ones, contaminating the resulting spectrum.

For formal reasons, we need to underline that the Fourier transform (FT) is not the best choice for analysis of non-stationary, pulse-like signals, such as those encountered in THz-TDS. For such, the Fourier spectrum of is an average representation of the true characteristics within an analysis window, which instantaneously change as a function of time delay. A remedy to this is a time-frequency representation like S-transform [24] or wavelets [25], however less-intuitive to interpret. On the contrary, results yielded by the FT are easy to interpret making the FT representation a widely accepted standard and a preferred choice for THz spectroscopy.

Due to the large number of acquired samples in spectroscopic experiments, a fast implementation of the DFT is in common use, known as the Fast Fourier Transform (FFT), which is usually optimized for $N = 2^p$, where p is an integer, however it is not a stringent requirement due to the existence of implementations like FFTW (Fastest Fourier Transform in the West) [26]. The latter computes the FFT efficiently for any positive integer in the form $N = 2^a 3^b 5^c 7^d 11^e 13^f$, with arbitrary exponents except for the constraint that $e + f = 0$ or 1. Once the transformation has been performed, we obtain a series of complex numbers half of which is unique. Each number can be represented in polar coordinates with an amplitude and phase:

$$X(\omega_k) = |X(\omega_k)|\mathrm{e}^{\mathrm{j}\varphi(\omega_k)}. \tag{2.22}$$

A plot of the two terms as a function of frequency for the previously discussed coherently and non-coherently averaged signals is shown in Figure 2.12. Solid line represents the power (a) and phase (b) spectrum of the aligned (coherently averaged) case, which shows a better high-frequency response, more uniform phase, and does not have a spectral artifact around 2.2 THz compared to the incoherently averaged spectrum plotted with dashed line. The reason of the spectral narrowing was mentioned before: incoherent averaging temporally broadens the pulse, effectively reducing the bandwidth. Finally, the errors in the position of averaged pulses in the non-aligned case corrupt the spectral phase, in particular above the spectral dip at 2.2 THz.

So far we assumed that averaging is a necessary step in the signal processing path but we have not shown the amplitude and phase spectrum of a non-averaged trace. Note that incorrectly averaged data can

introduce severe artifacts, thus the gain of this step must be quantified to justify the added complexity. This will be analyzed in the following subsection, starting with a mathematical background of averaging.

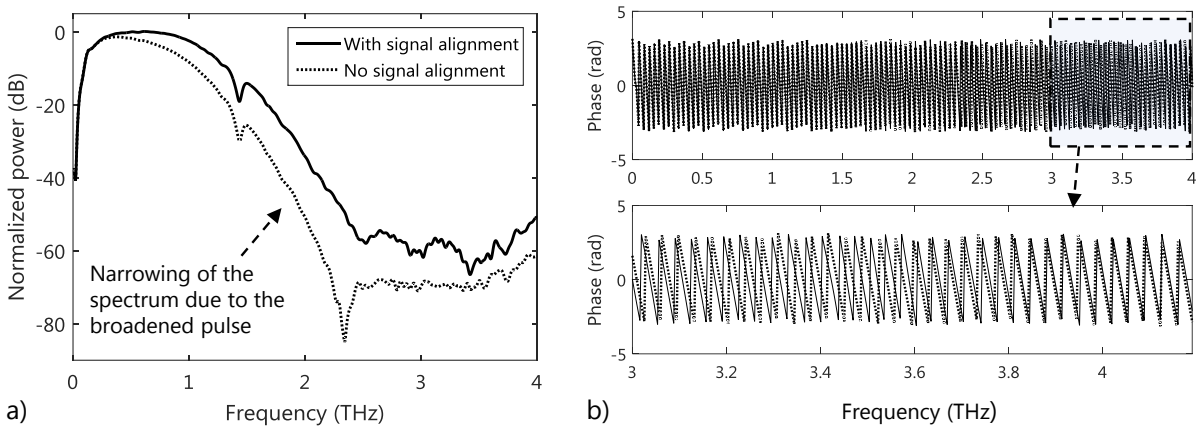

Figure 2.12. Power (a) and phase (b) spectra of the coherently and incoherently averaged time-domain pulses propagating through an α-D-glucose dehydrate sample discussed in the previous sections. Note the narrowed spectral bandwidth of the incoherently averaged pulses plotted with dashed line.

2.2.3. Coherent averaging in the frequency domain

Assuming the time domain (TD) signals are aligned, we can express each measurement $x_m(t)$ as a sum of a noise-free (deterministic) signal and a random noise $n_m(t)$ [12] (with a stochastic character)

$$x_m(t) = x_0(t) + n_m(t). \quad (2.23)$$

The aim of averaging is to cancel-out the noise term by assuming that it is white and zero mean. If we denote an expectation operator by $\langle\cdot\rangle$, we can write that $\langle n_m(t)\rangle = 0$, hence

$$\langle x_m(t)\rangle = x_0(t). \quad (2.24)$$

The expected value of a signal corrupted by white noise is the noise-free signal itself, implying that averaging of time-domain data decreases the noise, as long as its characteristics are white. Now, by transforming $x_m(t)$ into the spectral domain, we write

$$X_m(\omega) = X_0(\omega) + N_m(\omega). \quad (2.25)$$

Linearity of the Fourier Transform allows to decompose the spectrum into two additive terms, again yielding an expected noise-free spectrum since $\langle X_m(\omega)\rangle = X_0(\omega)$. It must be underlined, however, that such averaging either in the time or spectral domain will work only if the deterministic part of the signals preserves the phase (the signals are aligned in time). As an example, one can easily imagine a set of complex vectors of a constant amplitude but random phase. After averaging in the complex domain, the magnitude of the mean vector will be zeroed. This is the main motivation for signal alignment, a vital part of coherent averaging.

If the phase term is not necessary or cannot be corrected, one can average just the magnitude spectrum, which is

$$|X_m(\omega)| = |X_0(\omega) + N_m(\omega)|, \qquad (2.26)$$

In contrast to the previous linear operation formulas, the above uses a nonlinear operator of the absolute value, which precludes noise reduction. Mathematically the expected value of the magnitude is

$$\langle |X_m(\omega)| \rangle = \langle |X_0(\omega) + N_m(\omega)| \rangle. \qquad (2.27)$$

It is evident that the noise term does not vanish no matter how many averages are performed, yet the result in most cases will be more accurate than when averaging phase-incoherent complex spectra (FT of non-aligned signals) followed by taking the absolute value.

The same property of noise bias holds for entities dependent upon squared magnitude, such as power spectrum or power spectral density. Consequently, the need for coherent averaging to suppress noise contribution is of utmost importance. Let us support this statement by analyzing the Fourier spectra of the aligned time-domain data, plotted in Figure 2.13.

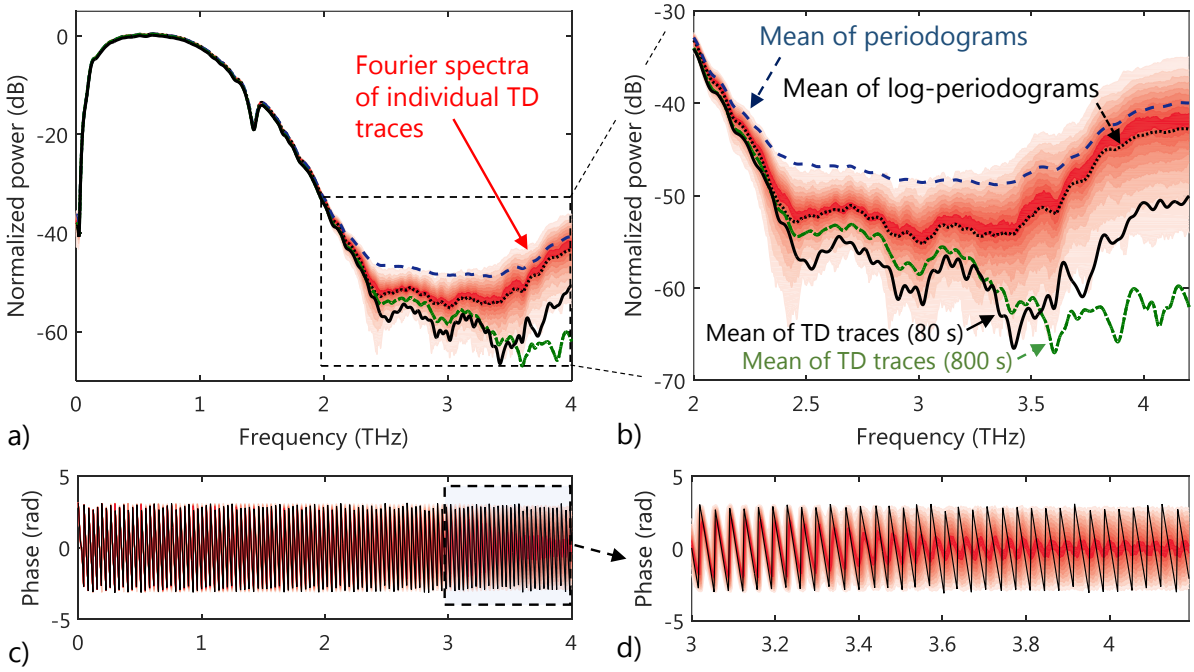

Figure 2.13. a) power spectra of all the aligned time-domain pulses b) zoom of the power spectra above 2 THz together with the average spectra obtained using different averaging techniques. c) phase spectra d) zoom of the phase spectra above 3 THz showing the smearing out of the spectral phase without coherent averaging together with the coherently-averaged trace (solid black line).

Panel a) shows a fan chart of the magnitude spectra of the individual TD traces in dB scale plotted with red lines, where darker regions correspond to traces with a larger number of occurrences. Together with the fan chart, the average spectra are plotted with solid, dashed, and dotted lines, calculated using

different averaging techniques. First of all, the low-frequency part below 2 THz has a predominantly deterministic character with a low spread of data (low noise) and insignificant differences between the magnitudes. Above 2 THz (zoom shown in panel b)) we can observe much larger differences in spectral amplitudes of up to 20 dB between the minimum and maximum. Also the average curves are located at significantly different levels. This is the region where averaging has the most noticeable effects, and the choice of the optimal averaging procedure is a central requirement for reliable retrieval of spectral data.

The top average power spectrum (APS) curve is plotted with dashed blue line and stands for averaging of the periodograms – estimates of the power spectral density (PSD) here scaled by the bandwidth to represent the power spectrum (squared magnitude), followed by taking the logarithm. From all the average curves, it is the best estimator of the PSD with the smallest variance, yet it does not reduce the noise, hence the spectral magnitude is biased. This effect is the result of noise rectification, and is commonly modeled using a Rice distribution [27]. We will discuss the usefulness of the periodogram in further sections of this work. Although, it is not the optimal choice for noise suppression in a TDS system, the lowest variance is useful in some applications.

The second from top APS curve (plotted with black dotted line) is calculated using the mean of log-periodograms. Essentially, it follows the darkest region of the red traces, but again, it is not the optimal choice for averaging due to the noise bias, yet lower than before. The second to last APS curve is plotted with green dash-dotted line and corresponds to the mean of all time-domain traces acquired within 800 s. Compared to the periodogram-based averages, we clearly see a decrease in bias and noise level in general but the averaging time is not chosen optimally, therefore the noise in the useful region below 3.5 THz is not optimally suppressed. The origin of the bumpy behavior of the spectrum above that frequency is not well known but it can be attributed to periodic TD signal sampling errors. They yield spectral artifacts that should not be taken into consideration in spectroscopic assessments [19].

Finally, the black APS curve is calculated from the coherently averaged time-domain traces over the optimal averaging time of 80 s. It has the lowest root mean square (RMS) noise floor within the usable bandwidth of the spectrometer, and the highest dynamic range (DR), defined in the spectral domain with respect to amplitude as [28]

$$\mathrm{DR} = \frac{\text{maximum magnitude of amplitude}}{\text{RMS of noise floor}}. \qquad (2.28)$$

Finally, it matches the literature data [29] much better than the other curves. This analysis proves that the same terahertz signal processed using slightly different techniques leads to differences in noise level as large as 20 dB, which in excessive amounts may hinder reliable characterization of the sample.

The spectral phase (wrapped) has also been studied and is plotted using a fan chart in panel c) together with a zoom on higher frequencies in panel d). It is evident that above 3 THz fluctuations of the phase

gradually increase and the envelope of the dark red region decays to fractions of a radian. By calculating the phase spectrum of the optimally-averaged time-domain pulse this effect is eliminated.

In concluding remarks, it should be underlined that the choice of integration time using peak amplitude deviation is just one of the many criterions. An alternative choice would be to analyze the behavior of the spectral amplitude deviation at any arbitrary frequency ω_a, which is likely to provide different optimal integration times depending on the choice of ω_a.

2.2.4. Nomenclature of spectral data

In the literature, one can find several definitions of amplitude spectrum, power spectrum, and power spectral density obtained from DFT with various normalization factors, which make a direct comparison of spectral data difficult, and can be a source of significant errors in the retrieval of sample's optical constants. For instance, it is in common practice to call power spectral density simply a power spectrum, even though they are not the same. To avoid such ambiguities, we will explicitly define the quantities used within this work. We also assume that signal is generated across a resistor, yielding a voltage.

For the purpose of time-domain data windowing used for instance to remove spectral echoes (and further justified in the next subsections), we will define the following normalization factors:

$$S_1 = \sum_{j=0}^{N-1} w_j \qquad (2.29)$$

and

$$S_2 = \sum_{j=0}^{N-1} w_j^2, \qquad (2.30)$$

where w_j are the weights of the used window function.

The first quantity is called the amplitude spectrum and in the discrete case is given by

$$\mathrm{AS} = |X(k\Delta\omega)| = |X(\omega_k)| = \frac{|X_k|}{S_1}. \qquad (2.31)$$

The unit of amplitude spectrum is Volt [V]. A closely related entity is the power spectrum calculated in two ways: from the magnitude of the squared complex spectrum or by multiplying the complex spectrum by its complex conjugate

$$\mathrm{PS} = |X(k\Delta\omega)|^2 = |X(\omega_k)|^2 = \frac{|X_k|^2}{S_1^2} = \frac{X_k \cdot X_k^*}{S_1^2}. \qquad (2.32)$$

Due to the squaring operation, the units are now voltage squared [V^2]. Finally, the power spectral density (PSD) is a bandwidth-normalized power spectrum. It allows for estimating the power of signal within a given frequency bandwidth

$$\mathrm{PSD} = \hat{S}_{xx}(\omega_k) = \frac{|X(\omega_k)|^2}{f_s} = \frac{X_k \cdot X_k^*}{f_s \cdot S_2}. \quad (2.33)$$

Here, the units are Volts^2/Hertz [V^2/Hz] or equivalently Volts^2·second [V^{2}s].

It is also worth noting, that the large dynamic range of signals in THz-TDS implies that frequency spectrum is commonly presented in decibels (log-scale). Since that uses power, amplitude and power spectra in dB scale are synonyms, as below:

$$\mathrm{PS_{dB}} = 10\log_{10}|\mathrm{AS}|^2 = 10\log_{10}|\mathrm{PS}|. \quad (2.34)$$

Another important term in spectral data nomenclature is the wavenumber, which is widely used in spectroscopy. It measures the frequency of a wave in cycles per unit distance (x-axis in the spectrum). The spectroscopic wavenumber is simply

$$\tilde{\nu} = \frac{1}{\lambda}. \quad (2.35)$$

The units of wavenumber are typically reciprocal centimeters (cm^{-1}) or simply wavenumbers. Note that this is different from the angular wavenumber k by a factor of 2π. What is of larger interest, is how to convert the ordinary frequency into wavenumber. This is given by

$$\tilde{\nu} = \frac{f}{c}. \quad (2.36)$$

We clearly see that the above relationship is linear and corresponds to ordinary scaling by the speed of light. To be compatible with the spectroscopic unit of cm^{-1}, *c* must be expressed in units of cm/s. As a rule of thumb, 1 cm^{-1} equals approximately 30 GHz.

2.2.5. Reduction of uncertainty in power spectral density

It is very tempting to say that having calculated $|X(\omega_\mathrm{k})|$ and $\varphi(\omega_\mathrm{k})$ we have retrieved the amplitude (power when squared), and phase spectrum – key factors in characterization of optical properties of a medium. Unfortunately, compared to the initial continuous Fourier Transform case with an infinite duration of the signal, the definition of DFT includes a truncation to only *N* samples, which has serious practical implications: we cannot obtain the full knowledge about the stochastic (random noise dominated) part of the process yielding the samples x_n, and we can only *estimate* the frequency and phase spectra. Even though we studied coherent averaging techniques to cancel out the noise in detail, some amount of randomness is still present in the measured spectrum. Hence, to minimize uncertainties in $|X(\omega_k)|$ for reliable characterization of the sample. we need to determine the optimal processing steps by recalling the spectral estimation theory. Note that the phase spectrum $\varphi(\omega_k)$ cannot improved using the techniques below. Only sufficiently long coherent averaging or the use of a two-detector balanced detection system can minimize uncertainties in phase.

Let us start with the estimation of power versus frequency distribution, known as the power spectral density (PSD) denoted by S_xx, which is the base for retrieval of multiple intensity-related spectroscopic

parameters. This is obtained by using the periodogram spectral estimator, which is essentially the squared magnitude of the DFT divided by the number of samples:

$$\hat{S}_{xx}(\omega_k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x_n \cdot \mathrm{e}^{-\frac{\mathrm{j}2\pi kn}{N}}\right|^2, \qquad k = 0,1,\dots,N-1 \tag{2.37}$$

One can ask, whether the periodogram is a consistent estimate of the true power spectral density S_{xx}, that is both asymptotically unbiased and with a variance that goes to zero? Unfortunately, not. Although the estimator is asymptotically unbiased, since converges to the true power spectral density S_{xx} for an infinite number of samples [30], its variance (understood intuitively as uncertainty) does not decrease as the number of samples *N* is increased for a random process realization.

$$var\{\hat{S}_{xx}(\omega_k)\} = S_{xx}^2(\omega_k). \tag{2.38}$$

In other words, the variance of the PSD estimate of a *random process* is as large as the true squared PSD itself (100%!) no matter how many samples are taken into account. This is caused by the lack of spectral averaging. To obtain low-variance estimates of the PSD, we need to average multiple periodograms. While in the low-frequency part the deterministic character dominates, and the experimentally derived variance is usually low, at higher frequencies, even in the case of a trace averaged over 80 s, the spectrum (dominated by random noise) significantly oscillates. In summary, if we just coherently average multiple time domain traces, we can lower the noise floor, yet at lower-SNR conditions the power spectral density will oscillate since the periodogram is calculated from just one averaged time-domain (TD) trace. Conversely, if we average multiple periodograms calculated from single non-averaged TD traces, we lower the uncertainty in amplitude but no noise is suppressed. It is evident that ideally both techniques must be used simultaneously to obtain a smooth THz spectrum with a maximal dynamic range.

In general there are two classes of averaging used to reduce the variance of the PSD referred to as *power averaging*: first that utilizes a single TDS trace but sacrifices the frequency resolution, and a second that requires multiple TDS traces while keeping the resolution defined by a single one. Let us start with the first technique, known as Welch's method [31], where a TDS trace is divided into *L* segments of length *M* with *D* overlapping points. Each segment is windowed, and Fourier-transformed followed by squaring the spectral magnitude. This operation uses a modified periodogram which is essentially

$$\hat{S}_{xx}^{\mathrm{w}}(\omega_k) = \frac{1}{S_2}\left|\sum_{n=0}^{N-1} x_n \cdot w_n \cdot \mathrm{e}^{-\frac{\mathrm{j}2\pi kn}{N}}\right|^2, \qquad k = 0,1,\dots,N-1 \tag{2.39}$$

Finally, the modified periodograms are averaged. Because of this operation, the variance in the averaged periodogram is reduced, especially in the proximity of the noise floor, yet the cost of greater confidence in amplitude estimate is a lowered frequency resolution. Notably, only squared magnitudes yield an improvement. If amplitude spectrum is a desired format, it is of utmost importance to take the square root of the averaged power spectra (periodograms) instead of averaging the raw amplitude.

To mitigate the lowering of the spectral resolution, in a second class of power averaging one averages the periodograms of multiple TDS traces in contrast to the previously processed one. It can be equivalently summarized as averaging of the squared magnitudes of windowed pulses. This is one of the efficient *incoherent averaging* techniques that is phase-insensitive, and allows for reducing the uncertainty in power spectrum in the presence of large errors in position of the delay line, yet it does not suppress the noise. Recall the top dashed line in the plot of different spectral averaging techniques (Figure 2.13). The curve was smooth but with the largest noise bias.

A completely different approach that can also reduce the uncertainty in the power spectrum involves point-wise spectral smoothing. Several neighboring points are averaged into one at the expense of frequency resolution. However, for practical reasons, rather than averaging in the spectral domain, a windowing technique is used in the time domain, which can additionally suppress slight ringing artifacts caused by a finite acquisition. The price, however, is high, since the frequency resolution is lowered again. In THz-TDS this step is of high importance, therefore we will review it in the next subsection.

2.2.6. Importance of windowing and zero padding

The finite number of samples creates bias in the PSD estimate related to the convolution of S_{xx} with a Barlett (triangular) window, which gives rise to spectral leakage causing the power of the signal to "leak" into neighboring frequency bins. Furthermore, the discrete grid of the Fourier Transform can affect the amplitude estimation of narrowband frequency components (contained within one bin), known as the scalloping loss. The spectral magnitude of a sine wave which lies anywhere between two discrete frequency bins with a frequency equal to a non-integer multiple of f_s/N will be incorrectly retrieved. This effect is the most severe for discrete-like spectral components, and less pronounced for broad spectral peaks. The severity of this phenomenon is visualized in Figure 2.14, showing the magnitude response of different window functions with a zoom into a bin-wide spectral amplitude. If no window is used, equivalent to the rectangular window, the true amplitude can be underestimated by 3.9 dB, which corresponds to a relative error of

$$\frac{\Delta A}{A} = \left(1 - 10^{-\frac{P_{\mathrm{EdB}}}{20}}\right) \cdot 100\% \approx 36\%. \qquad (2.40)$$

Such an underestimation is not acceptable in any spectroscopic application; therefore, we need to apply some compensation techniques. One of the well-known techniques to alleviate scalloping effects is to use windowing. Not only does it flatten the magnitude response within one bin, but also suppresses broadband spectral leakage visible as sidelobes in Figure 2.14, however at the expense of spectral resolution. In the case of real world signals, however, the scalloping loss is not so severe due to their usually broadband nature [32] and can be easily lowered by increasing the number of FFT points known as zero-padding. While it does not improve the spectral resolution, it provides a smoother estimate of the spectrum in both amplitude and phase, and is commonly used to correctly retrieve the correct

position and amplitude of spectral peaks. From a practical standpoint, zero padding always follows windowing.

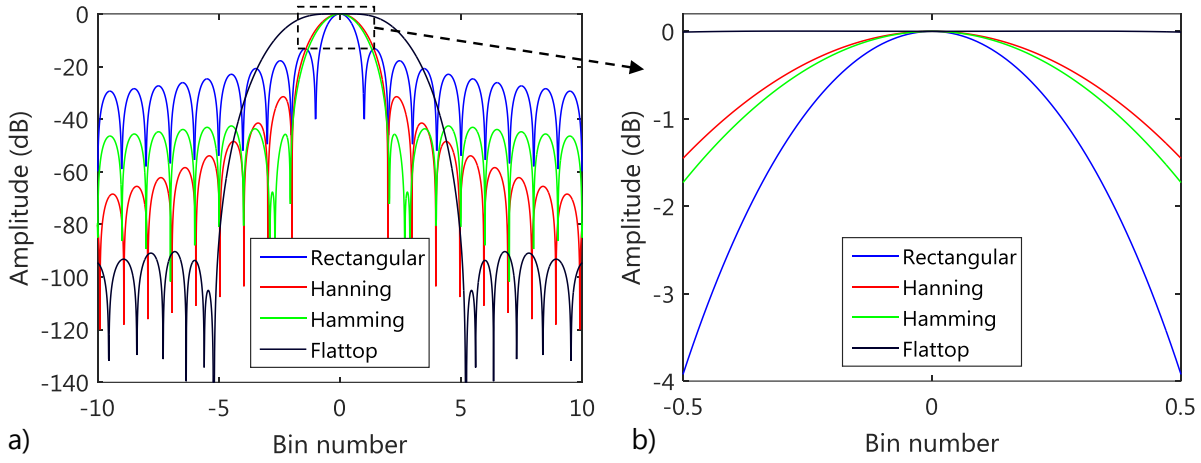

Figure 2.14. Log-magnitude response of four common windows used in FFT. a) plot illustrating spectral leakage b) plot illustrating scalloping loss within one FFT bin. The flattop window shows the lowest error in amplitude, albeit at the expense of reduced frequency resolution.

Let us show how the two techniques used simultaneously can suppress unwanted reflections in solid THz TDS spectroscopy, where one has to deal with refractive index mismatch between the air and sample causing time-domain echoes. They translate into the etalon effect – a severe modulation of the Fourier spectrum, which precludes reliable spectroscopic assessments. While one way to partially account for it is to increase the sample thickness, it is always accompanied by an increase in absorption, which in turn lowers the signal to noise ratio and the usable spectral bandwidth. The optimal thickness of pellets for THz-TDS has been found to be approximately 3 mm [33], again severely limiting the spectral resolution to tens of GHz. An example time-domain measurement of a 3.1 mm thick pure polyethylene (PE) pellet used as a host matrix for solid samples is shown in Figure 2.15, together with a pulse propagating through dry nitrogen.

In the time domain, we can observe three general effects. The first is one is obvious: the pulse propagating through the PE sample is partially attenuated, visible as the lowered peak amplitude caused by absorption. The second one is related to the delayed arrival of the pulse propagating through the sample due to a lower phase velocity in the medium. Assuming a nearly constant group refractive index $n_g = n_m + \omega(\partial n_m/\partial\omega)$ of the medium with a thickness d, we can express the time delay of the reflected pulse with respect to the dry-air pulse

$$\tau_{\text{delay}} = \frac{d\Delta n_g}{c}, \qquad (2.41)$$

where Δn_g is the difference in the group refractive indices between the air and the sample pellet.

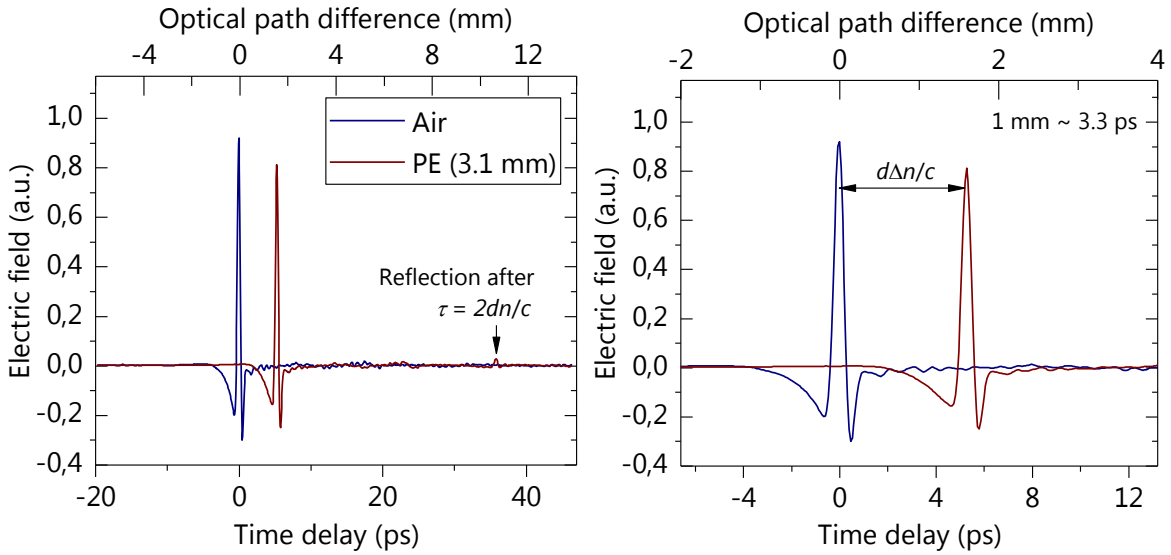

Figure 2.15. Time-domain measurement of dry air (blue) and polyethylene sample (brown), wherein an echo caused by the refractive index mismatch can be observed. The delay in the arrival time of the pulse propagating through the polyethylene pellet can be used to roughly estimate the mean refractive index.

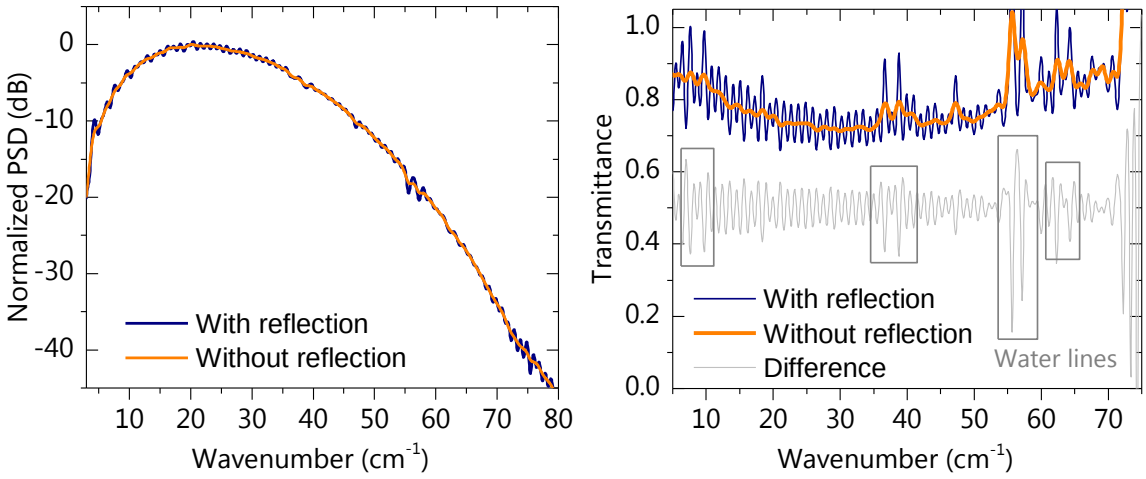

Figure 2.16. Normalized power spectral densities (left) and transmittance (right) of the PE pellet. The blue line plots data without windowing, hence the spectrum is fringy. By numerically removing the echo, the oscillations are removed while preserving weak absorption features of water vapor.

Finally, the third effect is visible as a reflected echo after propagating twice within the polyethylene sample

$$\tau_{\mathrm{et}} = \frac{2dn_g}{c}. \qquad (2.42)$$

The spectrum of the Fourier-transformed signal including a time-domain echo (plotted with blue line in Figure 2.16) is modulated with a period equal to the roundtrip time in the sample, known as the etalon effect

$$f_{\mathrm{et}} = \frac{1}{\tau_{\mathrm{et}}}. \quad (2.43)$$

The peaks (in a dispersionless medium where $n_g = n$) occur at frequencies

$$f_N = \frac{c}{2nd} N \quad (2.44)$$

and throughs at

$$f_{N+1/2} = \frac{c}{2nd}\left(N + \frac{1}{2}\right), \quad (2.45)$$

where *N* is an integer describing the order of the peak. Also the amplitude of the peaks can be described mathematically. In a lossless medium, the power transmission of the Fabry-Pérot etalon is given by

$$T_p(f) = \left[1 + \mathcal{F} \sin^2\left(\frac{2\pi f}{c} nd\right)\right]^{-1}, \quad (2.46)$$

where the etalon finesse $\mathcal{F}$ is

$$\mathcal{F} = \frac{4R}{(1-R)^2}, \quad (2.47)$$

and the Fresnel reflectivity R is

$$R = \left(\frac{n-1}{n+1}\right)^2. \quad (2.48)$$

For the example shown in Figure 2.16, the roundtrip time delay in the PE sample is approximately 30.7 ps, which corresponds to ~32.6 GHz or 1.09 cm^{-1} of the spectrum modulation period. Now, if we apply windowing to suppress out the echo (a 60 ps-long Hanning window centered at the maximum energy of the signal multiplied by the TDS traces), we can numerically remove the etalon at the expense of frequency resolution (shown with orange trace in Figure 2.16). The residual of transmittance plotted with gray line is almost sinusoidal however some large amplitude oscillations can be observed too, caused by water vapor absorption: the major curse of spectroscopy in the THz.

An alternative way to tackle the reflection is to calculate the autocorrelation function in the search of peaks other than zero delay. By performing deconvolution, the echoes can be numerically removed, while preserving the full duration of the time-domain signal, which is in stark contrast to the windowing. However, to avoid introducing spectral artifacts, such attempts require an accurate model [34].

Even though windowing does not belong to coherent averaging techniques, it enables spectral averaging for etalon removal at the expense of frequency resolution. Its use is of utmost importance in the characterization of solid samples, enabling to probe only sample interface and retrieve the optical constants in an undistorted way.

2.2.7. Effects of windowing on spectral resolution of amplitude and phase

So far, we have supported the use of windowing on a solid spectroscopy example. At room temperature and pressure, linewidths of solids in the terahertz regime are on the order of hundreds of GHz, therefore the numerical removal of echoes does not have a negative effect on the accuracy of the measured spectrum and is widely used in practice. As we will see in this subsection, in gas spectroscopy requirements regarding echo-free delay line scans are much more stringent. Insufficiently long acquisitions, equivalent to applying a narrow time-domain window, broaden the absorption lines in the measured spectrum far beyond their true widths at given conditions.

Let us demonstrate the implications of the narrowing of the acquisition time on the spectrum of water vapor, which is one of the strongest and richest THz absorbers. In the first case we will consider a 120 ps coherently averaged long acquisition, where the first ~20 ps preceding the pulse does not carry any spectroscopic information. No windowing is used here (except for rectangular which is caused by the finite measurement itself) since the oscillations in the tail following the main pulse gradually decay and eventually reach the noise level. As a result, we expect to obtain a frequency spectrum not severely limited by the instrumental line shape: the resolution limit of the spectrometer is comparable with the line width of water vapor lines at atmospheric pressure and room temperature.

The time-domain pulses of dry and water-vapor rich air measured at the end of October in central Europe (Poland, Wroclaw, laboratory conditions) are plotted in top left panel of Figure 2.17. The two pulses differ significantly. First, the peak amplitude of the electric field for the humid air is lowered to ~86% compared to the dry air case.

Second, the humid air pulse is followed by rapid high amplitude oscillations, which arise due to the dynamics of the water vapor resonating at terahertz frequencies with multiple rotational and vibrational transitions [35], [36]. When the two pulses are Fourier-transformed, they reveal their completely different spectral characteristics. The oscillations in the humid air trace translate into numerous sharp peaks in the frequency spectrum, whereas the dry air spectrum is featureless. Also note that the natural decay of the water vapor oscillations' amplitude relates with the width of the absorption lines. At lower pressures or temperatures, the lines would narrow, hence producing time-domain oscillations with a longer persistence time.

For illustrative purposes, the power spectrum is plotted in linear and logarithmic (log) scale. It is evident that frequencies above 70 cm^{-1} have much lower intensities and can be analyzed only in log scale. What is also visible in linear scale and less pronounced in log scale are slight oscillations of the envelopes of the two curves. It can be attributed to echoes in the substrate of the photoconductive antenna yielding optical fringes, as well as instrumental noise still persistent in the system.

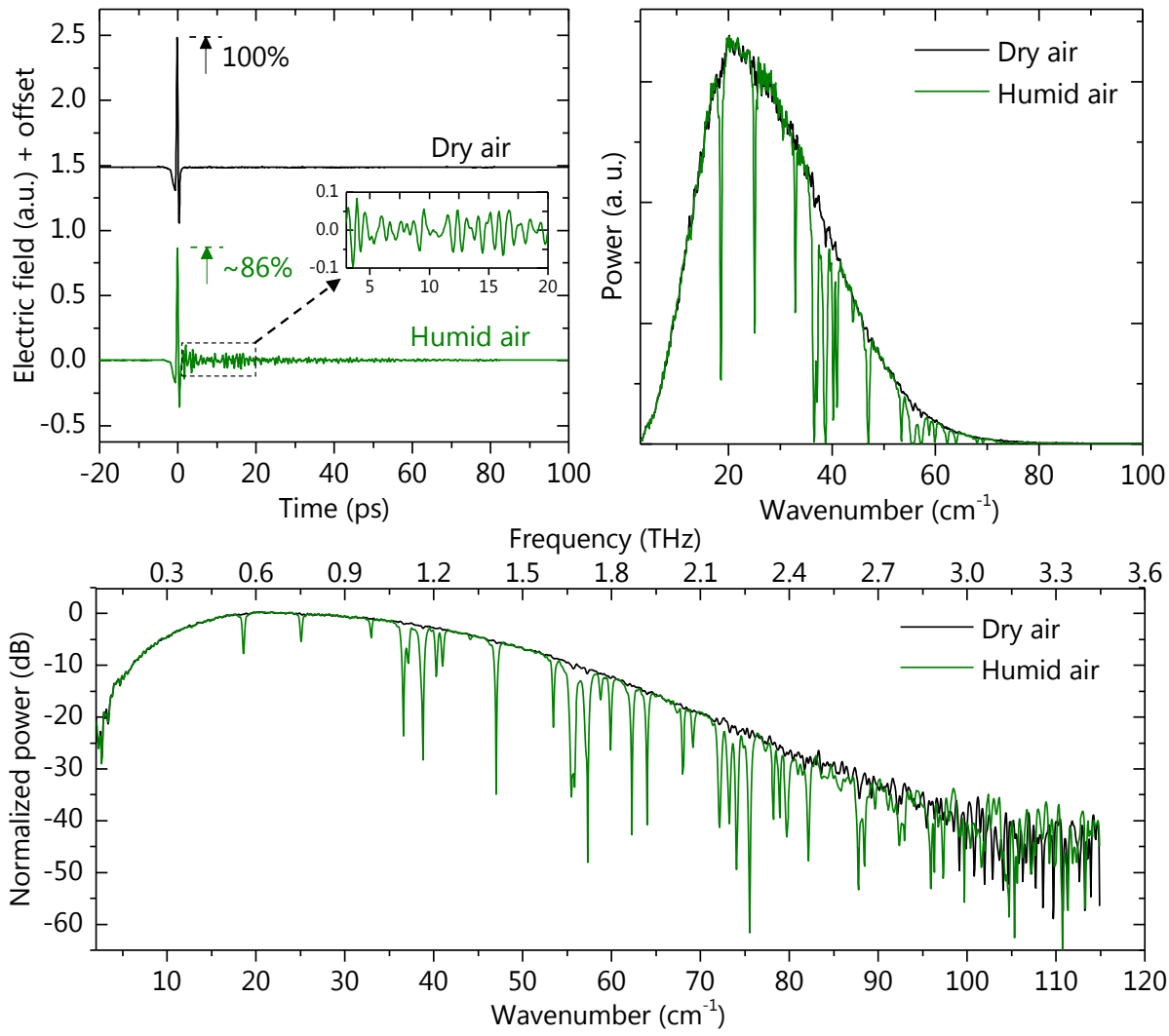

Figure 2.17. Broadband terahertz time domain spectroscopy of water vapor. Top left: time-domain traces of dry and humid air. The amplitude of the humid air time-domain pulse decreases down to 86% of the dry.

To represent the measurement in a more conventional way that suppresses the instrumental characteristics, we will perform spectral deconvolution. This operation divides the complex sample spectrum by the reference spectrum, yielding a complex sample transmittance:

$$H(\omega_k) = \frac{\tilde{X}_{\mathrm{smp}}(\omega_k)}{\tilde{X}_{\mathrm{ref}}(\omega_k)} = \rho(\omega_k) \cdot \mathrm{e}^{-\mathrm{j}\phi(\omega_k)}. \qquad (2.49)$$

The amplitude term $\rho(\omega_k)$ is called the amplitude transmission, amplitude transmittance or field transmittance. When squared, it represents power transmission, sometimes simply referred to as transmittance.

$$T_p(\omega_k) = \rho(\omega_k)^2. \qquad (2.50)$$

Also the phase (dispersion) term is of interest for characterization of dispersive properties

$$\phi(\omega_k) = \angle H(\omega_k) - 2\pi M(\omega_k). \quad (2.51)$$

Since the phase of DFT is wrapped, we introduce an integer $M(\omega_k)$ to account for phase discontinuities. The importance of this operation will be analyzed in the next subsection, together with possible spectral artifacts. We will also discuss how these entities are used to retrieve the intrinsic optical parameters of a medium in further sections. For now, let us focus on the windowing effects on the transmittance and phase spectrum which did not undergo any unwrapping due to the phase shifts lower than $\pm\pi$.

The optical transmission and optical phase spectra of humid air are plotted together in Figure 2.18. We clearly see that whenever an absorption peak appears, it is followed by rapid oscillations in phase. The stronger the peak, the greater the phase deviation. In general, the phase spectrum fluctuates less than the transmission, and its shape looks approximately like the derivative. We will describe the latter phenomenon more accurately in the following section using the Kramers-Kronig relations between the real and complex part of the fundamental optical constant: the complex refractive index, which can be obtained from the THz-TDS measurement.

A zoom into some of the absorption lines of water vapor plotted in the lower panels of the figure allows to observe good agreement with the HITRAN spectroscopic database. In the simulation, we assumed the USA high latitude summer model, atmospheric pressure, temperature of T=293 K, and the propagation distance of 0.45 m. The measured full-width at half maximum (FWHM) of the absorption lines is ~15 GHz, being equal to that in the database. This additionally proves the validity of the coherent averaging operations performed on the signal.

Now, if we window the signal using a 60 ps wide Hanning window, centered at the maximum of energy, and calculate the frequency spectrum, we can observe a tremendous difference in the resolution capabilities of the system coupled with a large decrease of fast frequency oscillations in the power spectrum, as shown in Figure 2.19. For comparative purposes, the windowed and non-windowed traces are shown in inset. The tail of the numerically windowed pulse gradually decays towards zero, which is the reason of the broadened spectral features.

Windowing, however, has some positive effects (see Figure 2.20). Being one of the spectral averaging techniques, it reduced the variance of the power spectrum, especially at frequencies exceeding 70 cm^{-1}. As a result, the transmission spectrum does not exceed one up to 100 cm^{-1}, allowing for more broadband acquisitions, whereas in the non-windowed case such artifacts limited the coverage to 85 cm^{-1}.

Now, if we compare the windowed and the non-windowed transmission and phase spectra, we can observe the convolution of the absorption lines with the instrumental line function (ILS). Since the width of the ILS is inversely proportional to the scan range (window length) of the spectrometer, by cropping the TDS data to approximately one third of the full length, we decreased the resolution by a factor of three. Absorption lines lying in close proximity are no longer resolvable in both transmission and phase.

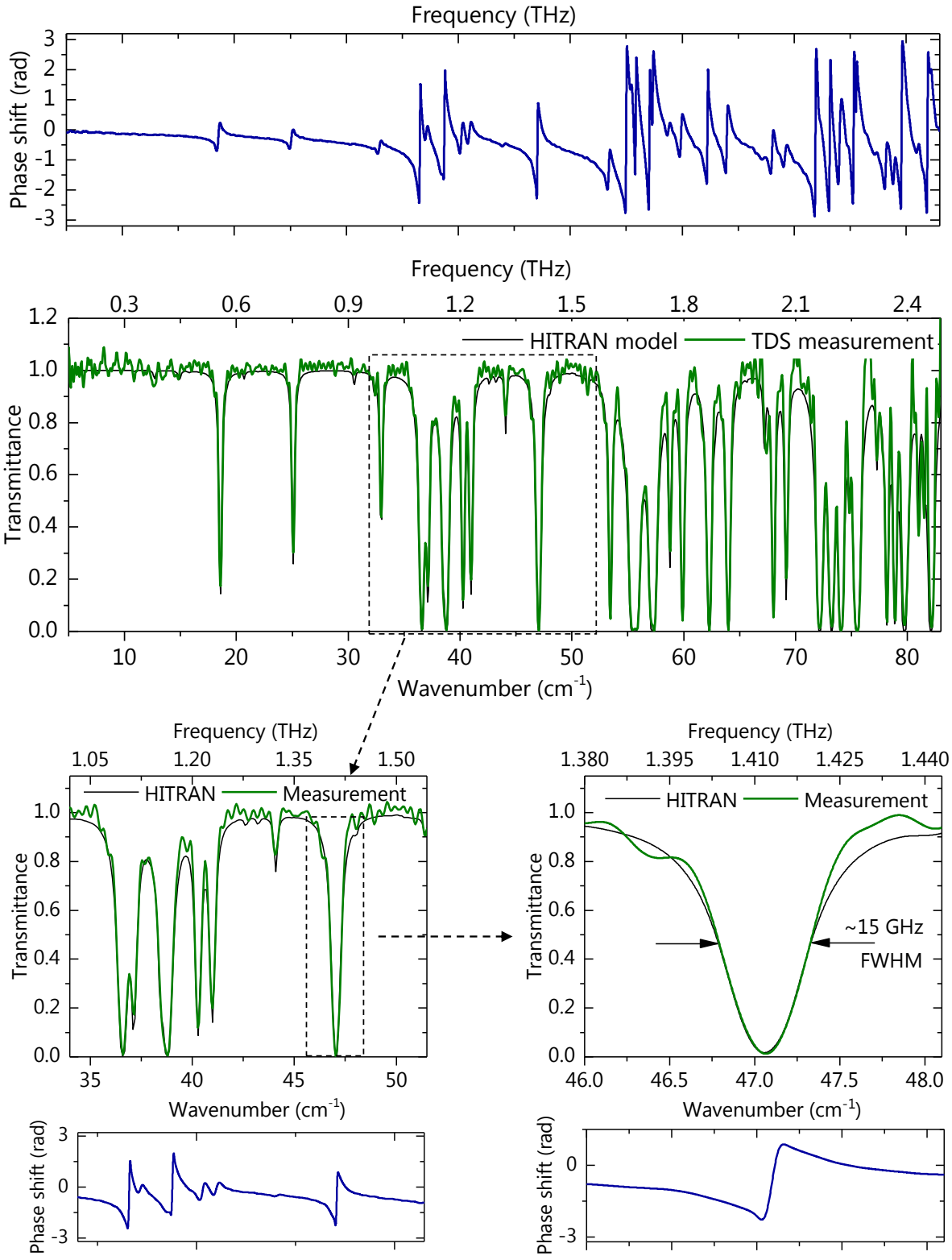

Figure 2.18. Broadband THz Transmission (green), and phase (blue) spectrum of water vapor without windowing plotted together with a HITRAN fit (black). Note the larger oscillations of the transmission spectrum.

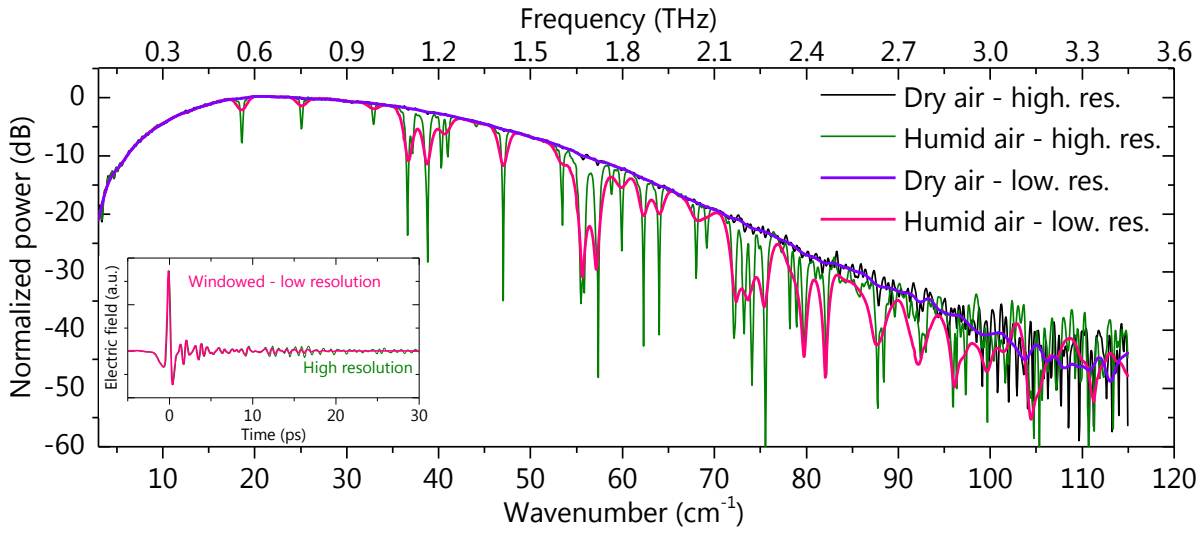

Figure 2.19. Comparison of the windowed (thick pink and violet traces) and non-windowed (thin green and black traces) power spectra. Note the reduction in variance of the spectrum, especially at frequencies greater than 70 cm^{-1}.

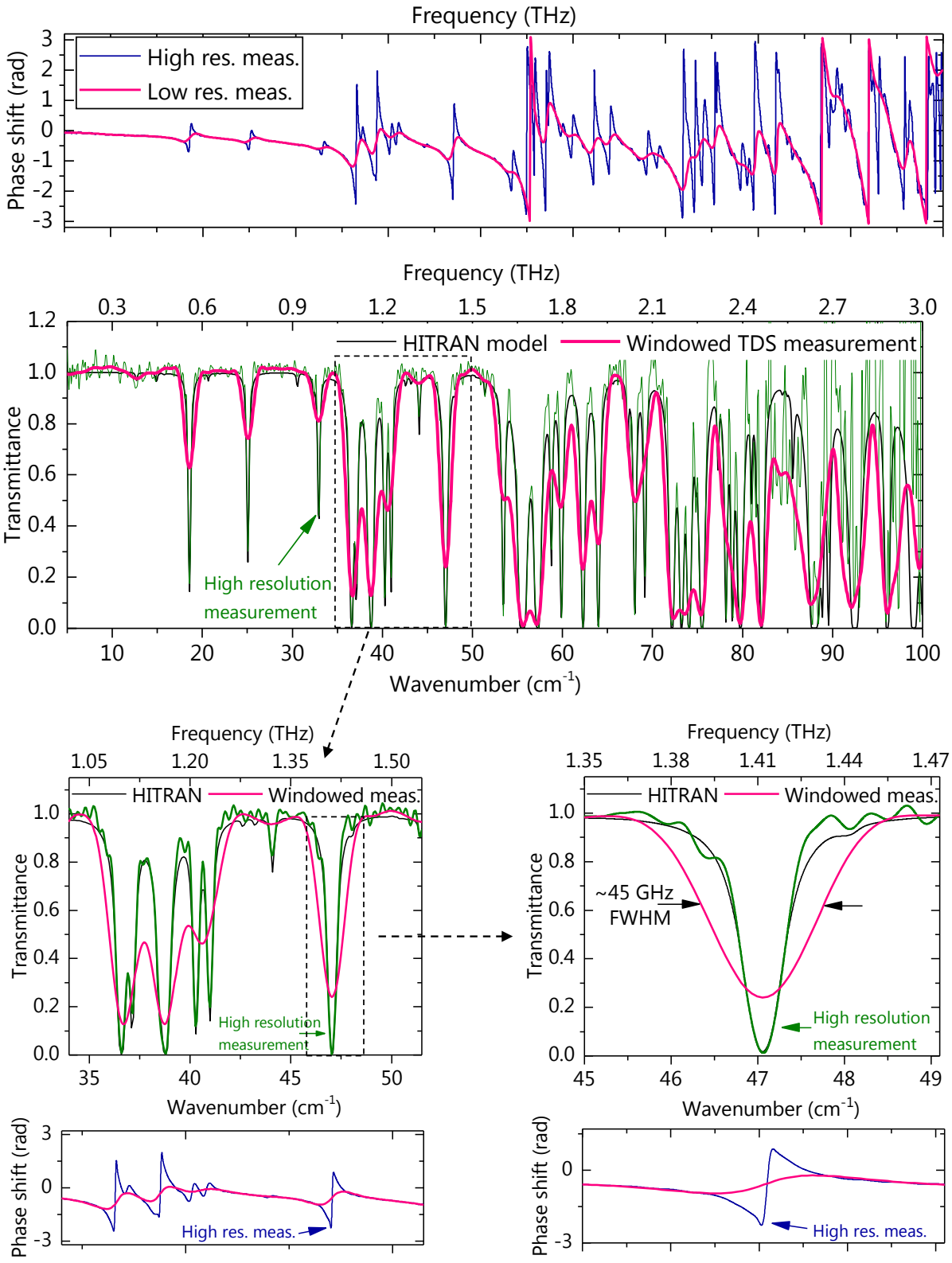

Figure 2.20. Broadband THz transmission and phase spectrum of water vapor obtained from the windowed pulse plotted with pink line. The colors of the non-windowed, high-resolution measurements have been preserved for comparative analysis.

2.2.8. Retrieval of sample's optical parameters

Optical transmission and phase are relative, non-normalized entities that depend upon sample thickness, concentration in addition to the optical constants of the medium itself. In material characterization, the complex index of refraction $\tilde{n}$ is a more appropriate representation. It can be expressed in terms of dispersion (refractive index n) and absorption (extinction coefficient κ) as:

$$\tilde{n}(\omega) = n(\omega) - \mathrm{j}\kappa(\omega). \tag{2.52}$$

The two parts of $\tilde{n}$ are related by the well-known Kramers-Kronig (K-K) relations, responsible for the derivative-like shape of the optical phase

$$n(\omega) = 1 + \frac{2\omega}{\pi}\mathcal{P}\int_0^\infty \frac{\kappa(\omega')}{\omega'^2 - \omega^2}\mathrm{d}\omega', \tag{2.53}$$

$$\kappa(\omega) = \frac{2\omega}{\pi}\mathcal{P}\int_0^\infty \frac{\omega'(n(\omega') - 1)}{\omega'^2 - \omega^2}\mathrm{d}\omega', \tag{2.54}$$

where $\mathcal{P}$ is the Cauchy principal value due to the singularity in the integral denominator. While the above formulas can be employed to transform the absorption (transmission) spectrum into dispersion and vice versa, in THz-TDS one can retrieve both parameters directly from the complex transmittance $H(\omega_k)$ being the result of spectral deconvolution:

$$H(\omega) = \frac{\tilde{X}_{\mathrm{smp}}(\omega)}{\tilde{X}_{\mathrm{ref}}(\omega)} = \rho(\omega)\cdot e^{-\mathrm{j}\phi(\omega)}. \tag{2.55}$$

This is advantageous because the K-K formulas are quite difficult to evaluate numerically. If we assume incidence at 0 degrees, the refractive index $n(\omega)$ can be conveniently derived from the optical phase,

$$n(\omega) = 1 + \frac{c\left(\phi_{\mathrm{smp}}(\omega) - \phi_{\mathrm{ref}}(\omega)\right)}{\omega d} = 1 + \frac{c\phi(\omega)}{\omega d}, \tag{2.56}$$

where d is the sample thickness, and c is the speed of light. Next, the refractive index is plugged into a formula for extinction coefficient $\kappa(\omega)$ [37]:

$$\kappa(\omega) = \frac{c}{d\omega}\ln\left[\frac{4n(\omega)}{\rho(\omega)(1 + n(\omega))^2}\right], \tag{2.57}$$

Recall that $\rho(\omega)$ is the field (amplitude) transmittance.

An alternative way of characterizing the sample's optical parameters common in solid spectroscopy is inspired by spectrophotometry, wherein one employs the well-known Beer-Lambert law. It states that the intensity of light propagating in an absorbing medium decays exponentially through

$$I(\omega) = I_0 \mathrm{e}^{-\alpha(\omega)\cdot L}. \tag{2.58}$$

The absorption coefficient $\alpha(\omega)$ (in units of cm^{-1}) in the formula can be conveniently obtained from the extinction coefficient $\kappa(\omega)$ by using a relation:

$$\kappa(\omega) = \frac{c\alpha(\omega)}{2\omega}, \tag{2.59}$$

which after rearrangement and substitution yields the expression for $\alpha(\omega)$:

$$\alpha(\omega) = \frac{2\kappa(\omega)\omega}{c} = \frac{2}{d}\ln\left[\frac{4n(\omega)}{\rho(\omega)\left(1+n(\omega)\right)^2}\right]. \tag{2.60}$$

All the above formulas require the knowledge of the sample thickness. Because in some applications the shape of the spectrum is sufficient, we can define absorbance, which is the negative decimal (common) logarithm of the power transmittance

$$A(\omega) = -\log_{10}(T(\omega)). \tag{2.61}$$

The absorbance has no unit, and it does not account for the Fresnel losses at the air-sample interface $R = (n-1)^2/(n+1)^2$. This is because the measured quantity on the photodetector is light intensity proportional to the squared amplitude of the electric field. The spectrophotometric power transmission is defined as

$$T(\omega) = \frac{I_T(\omega)}{I_0(\omega)}, \tag{2.62}$$

where $I_T(\omega)$ is the intensity of light transmitted through a sample, and $I_0(\omega)$ is the intensity of light illuminating the sample. The lack of phase measurement implies that in absorbance one cannot distinguish between material absorption and losses caused by the refractive index mismatch, hence there is no simple formula converting absorbance into absorption coefficient.

Under what conditions will absorbance sufficiently explain just absorption in the sample? The answer is that we need to ensure that the reference pulse propagates through a medium with a similar refractive index and thickness as the sample. In solid spectroscopy, this is obtained through measuring an optically-transparent pure host matrix reference pellet followed by an absorbing sample pellet with several per cent (up to ~20%) of a solid mixed with the same transparent host matrix. Small concentration of the absorber implies that the Fresnel losses (calculated using the mean refractive index theory) will be low due to the high refractive index similarity of the two samples.

Absorbance has another important property: it is proportional to the molar concentration c (in units of mol/dm^3 or equivalently mol·L^{-1}), path length l (in cm), and molar absorptivity ε (in units of m^2/mol or in practice L·mol^{-1}·cm^{-1}):

$$A(\omega) = \varepsilon(\omega)lc. \tag{2.63}$$

The absorbance is an attractive choice for characterizing spectral mixtures. The absorbance of a linear mixture is the sum of the individual absorbances

$$A_{\mathrm{mix}}(\omega) = \sum_{i=1}^{N} A_i = l \sum_{i=1}^{N} \varepsilon(\omega) c. \qquad (2.64)$$

We will show in this chapter that by employing blind source separation techniques, we can decompose the absorbance of a THz spectral mixture into individual components together with their concentrations.

To make the retrieval of the optical constants complete, we can also characterize the sample in terms of the dielectric constants by utilizing the relation between the complex relative permittivity and the complex refractive index: $\tilde{\varepsilon}_r(\omega) = (\tilde{n}(\omega))^2$. Note that this relation is valid only for dielectric materials. The relative permittivity is a complex number $\tilde{\varepsilon}(\omega) = \varepsilon'(\omega) - \mathrm{j} \cdot \varepsilon''(\omega)$ with the real part:

$$\varepsilon'(\omega) = [n(\omega)]^2 - [\kappa(\omega)]^2 = [n(\omega)]^2 - \left[\frac{c\alpha(\omega)}{2\omega}\right]^2, \qquad (2.65)$$

which reflects the stored energy, and the complex (dissipative) part given by

$$\varepsilon''(\omega) = 2n(\omega)\kappa(\omega) = \frac{cn(\omega)\alpha(\omega)}{\omega}. \qquad (2.66)$$

2.2.9. Phase unwrapping artifacts and the influence on refractive index calculation

Now that we know how to retrieve the refractive index of a sample, we can study the undesired effect of incorrect phase unwrapping at low THz frequencies on the estimation of this important optical parameter. Not only is it necessary to describe dispersive properties of the medium, but also it is used in the calculation of absorption coefficient. The assumption in phase unwrapping is that the emitted and measured spectrum starts at DC, which is not true for THz-TDS systems. The resonant frequency of a photoconductive antenna is on the order of hundreds of GHz, and at frequencies below 2 cm^{-1} almost no power is emitted. If phase unwrapping is performed on raw phase in the low-frequency, weak-signal region, it takes unexpected oscillatory shapes, as shown with solid lines in Figure 2.21 in panels a) and b). After unwrapping such oscillations translate into a large phase offset even close to DC.

What should we expect from the phase for a medium with a nearly constant refractive index with minor local features, such as polyethylene? The answer is that $c\phi(\omega)/\omega d$ must remain approximately constant, which corresponds to a linearly increasing phase difference versus frequency. If the phase difference is offset at low frequency (does not start at zero), we will observe a hyperbolic-like shape of the refractive index (Figure 2.21c, red trace). This is an artifact, however possible to correct using signal processing techniques.

To alleviate this undesired effects, the unwrapped phase is first extrapolated down to DC (for instance using a linear fit [12]), and then it is wrapped and unwrapped again to force zero phase at DC frequency, as shown in Figure 2.21b. As a result, the refractive index (Figure 2.21c, blue trace) reaches its desired flat profile with the correct value of ~1.53 over a wide frequency range.

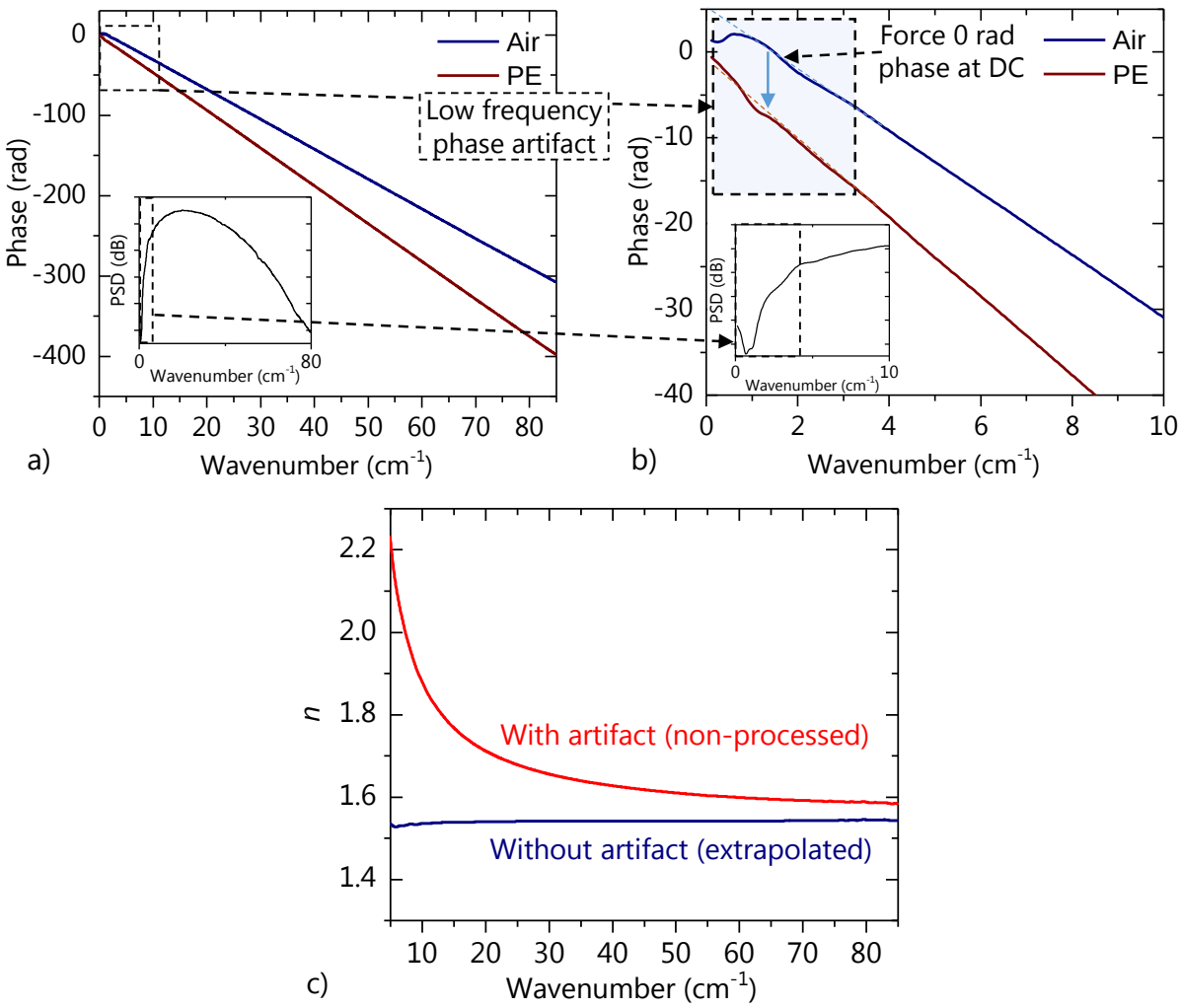

Figure 2.21. Phase unwrapping artifacts. In the low frequency region (b), the phase is incorrectly unwrapped with a significant offset persistent in the entire phase spectrum (a). By extrapolating (dashed line) and forcing the unwrapped phase to start at zero, we can retrieve the correct dispersion spectrum (c).

2.2.10. Summary of digital signal processing path in THz-TDS

So far we have discussed the effects of multiple signal processing operations influencing the quality of THz spectroscopic measurements and proved the validity of the proposed processing path on spectroscopic examples. For clarity, let us summarize the proposed digital signal processing path in the algorithm below.

Digital signal processing path in THz-TDS
Coherent averaging and time-domain signal preparation
1: Find/define a reference THz pulse
2: Reject measurements with a correlation coefficient below a given threshold (outliers)

3: Align the correct signals and calculate the mean time-domain trace (given the optimal averaging time)

4: Find any spectral echoes (reflections) and determine the length of the processed signal

5: Remove the mean value (DC offset) within the analysis window

6: Window the data of interest to suppress the echoes. A Hanning window is a good choice for start

Fourier Transform:

7: Calculate the Fast Fourier Transform after zero padding, at least 4 times the length of the TDS signal

8: Correct the phase error in the low frequency part of the spectrum after unwrapping

9: Repeat the steps 1-8 for the pulse propagating through the sample (so far it was the zero-gas pulse)

10: Calculate the complex transmittance function being the ratio of the sample and reference spectra

11: Determine the usable bandwidth of the spectrometer using the noise floor in the sample measurement

Retrieval of sample's optical parameters:

12: Calculate the refractive index from the spectral phase, used further to account for Fresnel losses

13: Calculate the absorbance/absorption coefficient from the amplitude/power transmittance

14: Compare the data with a spectroscopic database, if exists

End of algorithm

2.3. Application to pharmaceuticals

One of the main reasons for the popularity of terahertz waves is their ability to probe fingerprints of solid state pharmaceuticals. In particular, terahertz waves enable to detect and characterize a major concern concert in pharmaceutical industry: the existence of an active pharmaceutical ingredient (API) in more than one crystal lattice configuration known as *polymorphism*. Even though polymorphs are chemically identical compounds, their physical properties, i.e. melting point, and solubility vary from one form to another, causing a varying oral bioavailability understood as a fraction of a drug reaching the bloodstream after oral administration. Having established the signal processing path in THz-TDS, we can demonstrate the usefulness of pulsed terahertz spectroscopy in monitoring of such phenomena.

Environmental conditions such as temperature, humidity or pressure can cause a polymorphic transition, wherein a therapeutically active form of drug changes spontaneously into the inactive one [38], or even toxic. For instance, a well-known antibacterial antibiotic – *chloramphenicol palmitate* – loses its therapeutic properties when transforms from A to B polymorph [39] induced by improper storage conditions. Now, government drug approval agencies require to test a drug candidate for the existence of all polymorphic forms and thoroughly characterize their stability in drug storage conditions [40].

Conventionally, Differential Scanning Calorimetry (DSC) coupled with X-ray powder diffraction (XRD) pattern method is used to detect and characterize these destructive changes in drugs [41].

Regrettably, not only are these techniques invasive, but also require the use of ionizing radiation to detect the polymorphic change. This has changed with the advent of terahertz spectroscopy. In the last decade numerous demonstrations of non-ionizing THz radiation in application to pharmaceuticals have been presented, for instance the analysis of thermal decomposition of drugs [42], imaging of tablet coatings [43], and detection of polymorphic forms [44]. An example spectrum of two polymorphic forms of the same drug compound (here a Piroxicam analogue) is shown in Figure 2.22. Clearly, the two spectra are distinguishable without harnessing sophisticated calibration models or machine learning techniques.

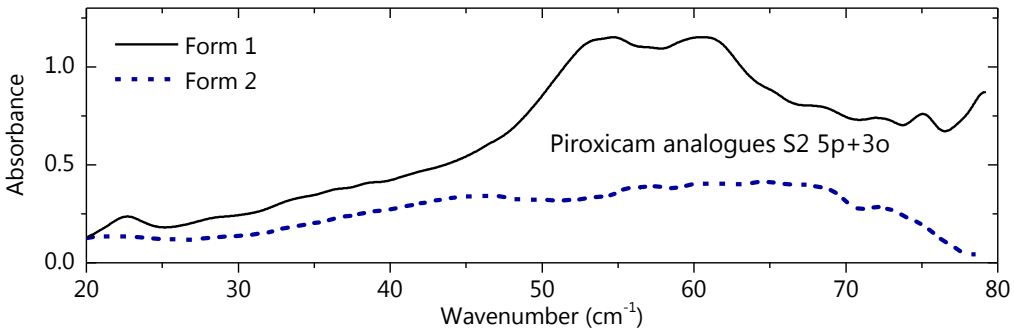

Figure 2.22. Two polymorphs of a chemically identical drug candidate (Piroxicam analogue No. 2 5p+3o). Even though the polymorphs share the same structural formula, their bio-availability and therapeutic activity can be completely different.

Furthermore, in addition to polymorphism, a related phenomenon of *pseudopolymorphism* is also a subject of intensive research due to similar reasons as for the first one: there might be a difference in bioavailabilities between different pseudopolymorphic states of a drug. One example of a pseudopolymorphic transition is dehydration, wherein a crystalline substance loses water molecules stimulated by the heat. The presence of hydration water embedded in the crystal lattice strongly influences the shape of the terahertz spectrum due to the existence of hydrogen bonds, which give rise to multiple resonances at terahertz frequencies. The importance of hydration and dehydration studies is additionally motivated by a drug's shelf life: hydration water can translate into an increased level of humidity in the packaging, which in turn is non-negligible for other ingredients and can lead to their molecular decomposition.

The two most famous examples with pharmaceutical importance and distinct features in the THz regime are α-D-lactose and α-D-glucose monohydrate, whose transformations will be analyzed in this work using advanced signal processing tools. Their common feature is that they possess multiple temperature-dependent absorption peaks. Conventionally, to characterize the kinetics of the dehydration reaction, only one arbitrarily chosen peak was used to retrieve the time-dependent abundance of the hydrate and dehydrate, unfortunately with some cross-talk from the neighboring peaks. To overcome this issue, in the next subsections we will show how to apply the Blind Source Separation algorithm to include the

contribution of all absorption features and enable the characterization in a way that is not biased by the choice of the peak.

2.3.1. Importance of sample preparation and storage

Before we proceed to an actual spectroscopic experiment, let us highlight several important aspects in the preparation of a solid-state sample for spectroscopic studies. Firstly, compared to most gaseous species, solid state drug ingredients are usually very strong absorbers. A thin sub-millimeter layer of powder can attenuate the beam by two or even three orders of magnitude in power. In a THz-TDS system this effectively limits the coverage by excluding the high frequency range from analysis, where the signal-to-noise ratio of an empty sample holder in purged air is weak (10 dB or less). A natural way to overcome this limitation is to use a stronger THz source, or coherently average multiple scans, however the latter will limit the temporal resolution in studies on transformation dynamics of a molecule.

Secondly, most powders for spectroscopic studies have a sub-millimeter particle size, which is comparable to the wavelength of the probing beam. This will have a considerable effect on the scattering properties of the sample, giving rise to a strong baseline, especially at higher frequencies. To account for it, one grinds the sample in a crystalline form into fine particles in a pestle and mortar [29]. Unfortunately, even such a procedure does not suppress scattering completely. It is a widely accepted practice to remove the baseline in numerical post-processing [45] through polynomial regression or more sophisticated algorithms, which we will shortly describe in one of the next subsections. For now, let us focus on the preparation of a pellet.

Having obtained a fine-ground powder for spectroscopic studies, we need to dilute it, mix thoroughly, and press into a pellet. For pharmaceuticals, a widely accepted standard is a pellet of 13 mm in diameter, and a thickness between 1.5 [29] and 3.5 mm [33], obtained by pressing a mixture of a host matrix and the powdered sample at one [45], two [46], four [29], or even five tons [33] for two [33] to three [45] minutes. In particular, Ref. [33] studies the effect of all those parameters in details on spectroscopic performance and concludes with the selection of the optimal sample of 300 mg and a thickness of approximately 3 mm, as a trade-off between the resolution (thinner samples cause more severe etalon oscillations) and non-linear effect due to a strong absorption, however the pressure was chosen arbitrarily without varying it in the comparative studies. The percent by weight (wt-% or wt%) of the absorber varies in most studies between 40% [29], and 5%, with a typical value of 10%.

Another important aspect of the pellet preparation is to allow the discs to rest once they have been prepared for 30 minutes [33]. Intuitively, their thickness measured immediately after pressing will differ from that evaluated after resting. When incorrectly measured, it will translate into inaccurate values of the refractive index and absorption coefficient, which depend on the difference in thickness between the reference and the sample. In binary mixtures like that in the sample pellet, the thickness is a linear combination of the compressed densities of the ingredients [33].

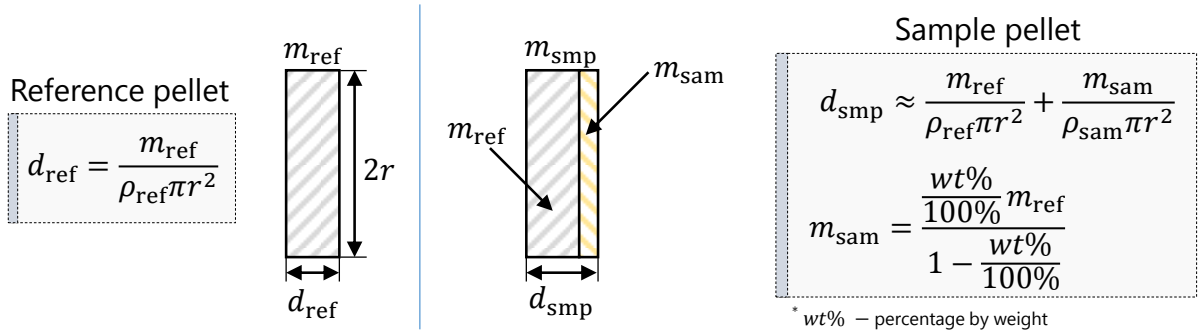

Figure 2.23. Summary of the pellet parameters given the mass of the reference pellet containing pure host matrix, and the percent by weight (wt%) corresponding to the concentration of the absorbing sample in the sample pellet.

Figure 2.23 illustrates how the reference and sample pellet should be prepared, and how their parameters vary depending upon the input parameters: percent by weight (wt%), mass of the powdered absorber m_{sam}, and reference pellet m_{ref}. From the above figure, it is also straightforward that the measured absorber has a thickness equal to the difference between the sample and reference thicknesses, which is further plugged into the equations to characterize the complex refractive index.

The choice of the host (diluting) matrix introduces yet another degree of freedom. In most studies, the gold standard is spectroscopic grade pulverized high density polyethylene (HDPE) matrix with a particle size of 10 μm [46], however there are experiments using Teflon (PTFE – Polytetrafluoroethylene) [46], and adamantane [47], [48]. They show different physical properties and their choice depends on the application. PE pellets were first proposed in 1955 by Hiroshi Yoshinaga and Robert A. Oetjen [49] as an alternative to the well-established KBr standard in the mid-infrared. Although it performs well in terms of THz transmission, it can be heated up to only 100°C in contrast to PTFE with a melting point of 327°C. For this reason, temperature-induced studies of drugs prefer the choice of Teflon, however at the expense of spectral coverage. A summary of the most popular pelleting materials is given in Table 2.1.

An interesting alternative for the aforementioned host matrix candidates in studies on exceptionally valuable samples is adamantane. Because it is completely featureless up to 442 cm^{-1} [50], it is perfectly suitable for far-infrared spectroscopy. So far, very few experiments with this matrix material have been presented for the reason that it slowly sublimes even at room temperature, which necessitates measuring the spectrum within a few hours after pressing the pellet. On the other hand, the preciously synthesized drug ingredient can be easily recovered from the mixture. When left in the air or vacuum, the pellet will become porous and eventually the adamantane will completely sublime.

The aforementioned guidelines find application in a vast majority of THz characterization experiments. However, for polymorphic or pseudopolymorphic studies additional requirements may apply. First, if one wants to measure a dehydrated form of a compound, it ought to be stored over silica gel in a dry air

environment [46]. Second, in (pseudo)polymorphic studies it is not always desired to have a host matrix, which in fact changes the environment of the analyzed compound. For instance, pressed polyethylene has water-sealing properties that can influence the kinetics of the phase transition. To address this issue, Liu and Zhang proposed to press a layered (non-mixed) pellet in order to ensure that the solid-state reaction took place in pure D-glucose monohydrate layer rather than in a PE-glucose mixture [29], whereas Zeitler et al. measured a heat-treated disc of 250 mg pure Carbamazepine without diluting [51].

Table 2.1. Summary of pelleting materials in the 0.05-2.50 THz range.

Material	Mean refractive index	Max. absorption coefficient (cm^{-1})	Melting point (°C)	Visible transparency
Adamantane*	~1.6	–	205-210*	×
PTFE	1.40	2.6	327	×
HDPE	1.54	0.35	~120	×
TPX†	1.46	0.8	235	✓
COC‡	1.53	1.0	n/a (75-170)*	✓

*Adamantane sublimes slightly at room temperature and even more at elevated temperatures. No accurate absorption coefficient was available but the transmission spectrum is comparable to HDPE. †Polymethylpentene (PMP). TPX is a trade name. Data source [52]. ‡ COC – cyclic olefin copolymer with a trade name TOPAS. Data from [53]. *Melting point does not exist. Instead, a heat deflection temperature (HDT) is defined, varying between 75°C and 170°C depending on the grade [54].

Another issue that cannot be overlooked is anisotropy of the crystals. In particular, crystalized sucrose [55]–[57] and oxalic acid dehydrate [57] show a significant birefringence in the THz range, which is exceedingly high Δn=0.6 for the acid. The THz spectrum of those samples aligned at different angles looks completely different, where certain absorption features appear only in the vertical or horizontal (0/90 degree) orientation. Similar effects have been observed in protein crystals [58]. Even though in a powdered sample the orientation of the crystals is random, it might be advantageous to test how the rotation of a sample changes the spectrum.

In conclusion, special care must be taken in studies on solid state samples to avoid misleading results, and a description of sample preparation must always accompany the obtained spectrum.

2.3.2. Active pharmaceutical ingredients

The origin of spectral features of active pharmaceutical ingredients in the THz region is related to the fact that they are heavy, complex organic molecules in crystalline form. It is the long-range interactions in the crystal lattice that are responsible for distinct THz spectral features [59]. For instance, the same spectrally-rich crystalline API measured in amorphous form turns featureless with a continuous density of states. Given that THz vibrational spectrum is determined by phonon modes, it is possible to predict with Density Functional Theory (DFT) [59], [60], albeit challenging for complex molecules.

Figure 2.24 plots the absorbance spectra of three popular APIs as an example. First is a zoledronic acid analogue used in the treatment of bone diseases (provided by prof. Matczak-Jon of Wroclaw University of Science and Technology, Poland). This drug has a broad, 15-cm^{-1}-wide peak located at approximately 50 cm^{-1} (~1.5 THz). Second is anti-inflammatory nonsteroidal drug piroxicam (provided by Dr. Berenika Szczesniak-Siega of Wroclaw Medical University, Poland) with a rich THz spectrum including 5 closely spaced peaks, two of which slightly overlap. Next in order is acyclovir hydrate used for the treatment of *herpes simplex* virus infections. It shows 4 weak, broad peaks with a large spacing and a considerable baseline slope. Clearly, the drugs are easily distinguishable in the THz domain. A more exhaustive study on THz spectroscopy of different APIs can be found in [61], in which the author has been involved.

For comparison, the last spectrum shows a drug excipient – a substance without a therapeutical effect, yet important from a drug formulation standpoint. α-D-lactose monohydrate shown there is one of the richest THz absorbers with distinct spectral features even at sub-terahertz frequencies. Not only is it widely encountered in pharmaceutical, but also in food industry. We will study its thermally-induced phase transition in the following subsection.

2.3.3. Pseudopolymorphic transitions of drug excipients

In the recent decades THz-TDS gained popularity in characterization of popular drug excipients. Inspired by earlier works on pseudopolymorphs, we studied the temperature-induced dehydration of α-D-lactose monohydrate using three different rate models to estimate the kinetics of the reaction [62]. Such parameters are currently subject to intensive characterization since hydration water is an important factor influencing the shelf-life of a drug dosage form. To estimate the rate of hydration/dehydration, one stores the sample at elevated temperatures (kept constant thorough the acquisition), while probing its optical response. Over timescales exceeding tens of minutes, one form gradually transforms into another. Having ensured a complete transformation, one extracts the normalized hydrate/dehydrate concentration profile α as a function of time from spectroscopic data.

In case multiple overlapping peaks coexist in the spectrum, there is no easy way to quantify the abundance evolution of each form. In the literature the most common technique is peak area integration, as for instance in [29], [51]. In [62] and [63] we proposed to employ a more sophisticated statistical tool for this purpose that does not require the choice of an arbitrary peak for integration, and shows immunity to overlapping peaks: the Bayesian blind source separation (BSS) algorithm [64].

In a simple picture it can decompose a linear mixture of a hydrate and dehydrate form with non-negativity constraints (absorption and concentration is never negative) into statistically independent components. Due to the immense complexity of the procedure, the reader is encouraged to consult the appropriate references for mathematical details [63], [64].

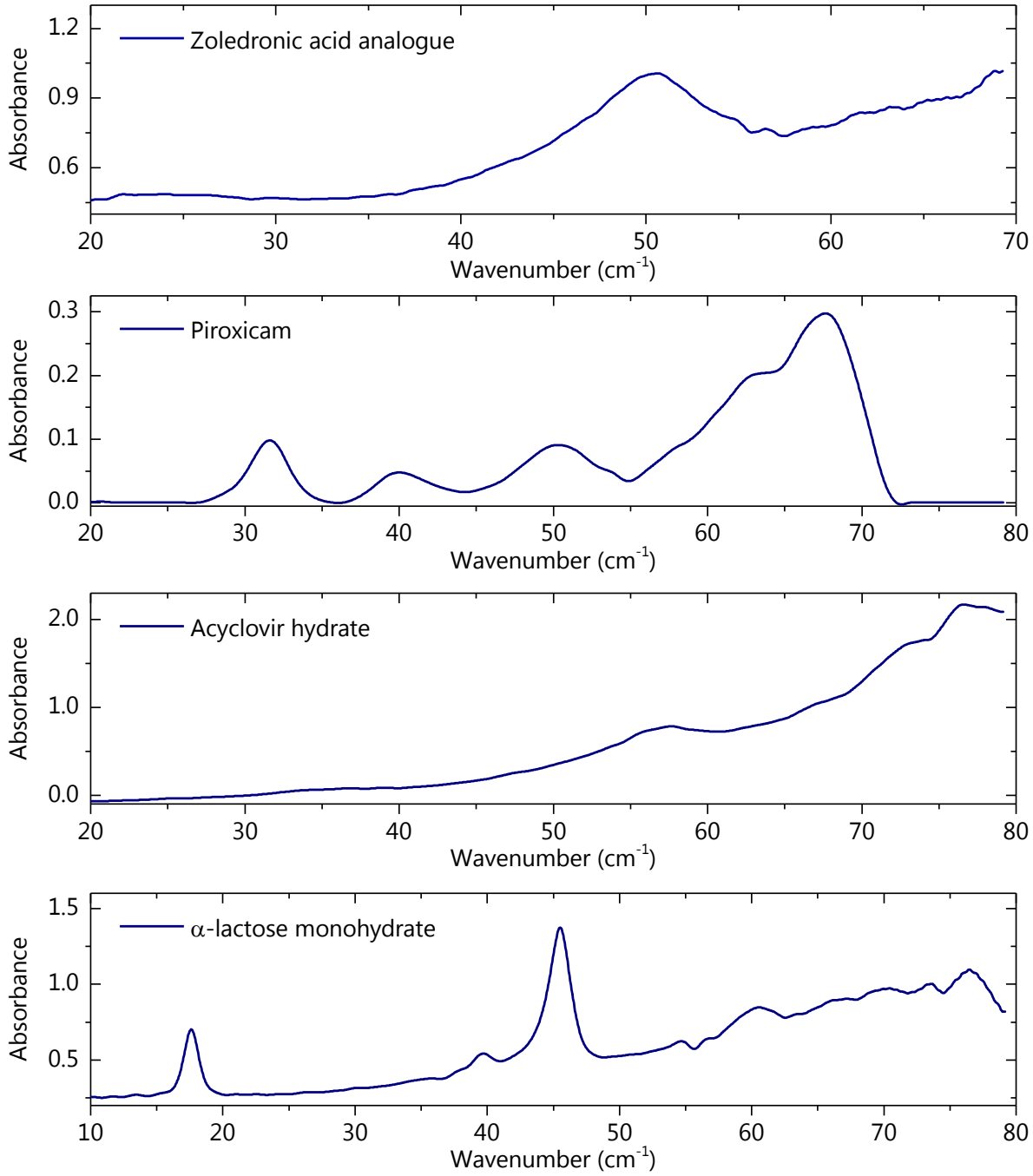

Figure 2.24. Example spectra of solid state pharmaceuticals and pharmaceutical excipients. From the top: zoledronic acid analogue used for the treatment of bone diseases, piroxicam – a nonsteroidal anti-inflammatory drug used for arthritis treatment, acyclovir – used for the treatment of herpes simplex virus infections, and α-D-lactose monohydrate – a popular pharmaceutical excipient with distinctively sharp features in the terahertz range.

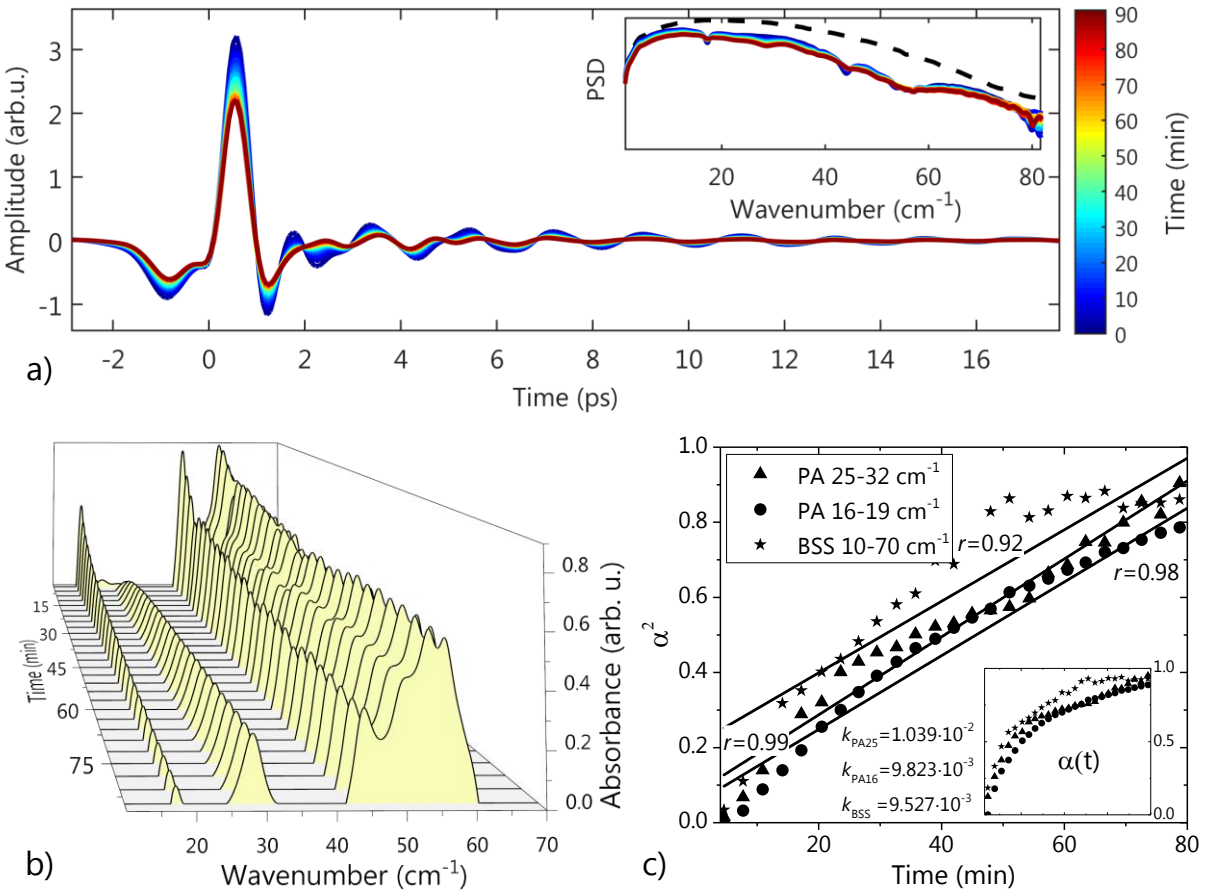

Figure 2.25. Dehydration of α-D-lactose monohydrate studied in the terahertz regime. a) evolution of the time domain waveform at a constant temperature of 100°C. b) THz spectrum after baseline compensation at different time instances c) Estimation of the dehydrate abundance in the spectrum with three different techniques: the peak-area method (PA) and the Blind Source Separation (BSS). Results published in [62].

After extracting the normalized concentration profile evolution from the terahertz spectrum $\alpha(t)$ using an arbitrary method, one needs to fit a rate model governing the transition. In solid state reactions, some of the most popular models are one-dimensional diffusion (D1)

$$\alpha^2 = k \cdot t, \tag{2.67}$$

contracting area [29],

$$1 - (1 - \alpha)^{1/2} = k \cdot t, \tag{2.68}$$

and one-dimensional contraction equation [65]

$$1 - (1 - \alpha)^{2/3} = k \cdot t, \tag{2.69}$$

where k is the rate constant. Figure 2.25 illustrates the estimation procedure in application to α-D-lactose monohydrate heated at 100°C. In the top panel (a) one can observe the evolution of the time-domain waveform. The amplitude of the transmitted terahertz pulse gradually decreases accompanied by a

suppression of the lowest frequency vibration responsible for the rapidly oscillating response after the main pulse, which is additionally visible in the spectral domain (b). By extracting the abundance of the dehydrate from the area under peaks located at 16 cm^{-1}, and 25 cm^{-1}, as well as using the aforementioned BSS approach, we obtained the dehydration rates after fitting the one-dimensional diffusion model to the retrieved concentration profiles.

Due to the temperature limitations of the heating system, the full capability of probing the reaction kinetics was demonstrated in our work on glucose monohydrate [63]. We performed six dehydration experiments at temperatures ranging from 40 to 50℃ in steps of 2℃, and estimated the abundance of the dehydrate using the novel BSS algorithm for comparison with the previously used peak area method [29]. We found that the best fit was obtained with the one-dimensional contraction equation. The spectral evolution together with the extracted abundances and fits are plotted in Figure 2.26.

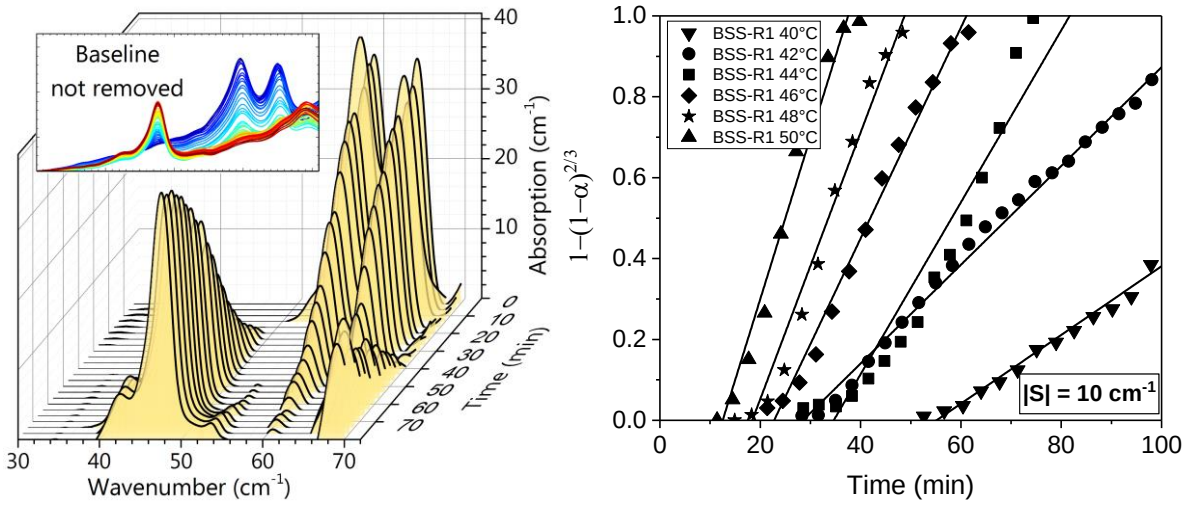

Figure 2.26. Dehydration of α-D-glucose monohydrate studied in the terahertz regime. (Left) Spectral evolution of the sample heated at 46℃ revealing the appearance of an intensive absorption peak located at ~45 cm^{-1}. (Right) Abundance of the dehydrate together with the fits to the model. Results published in [63].

To claim the characterization of the reaction kinetics, it is desired to fit the retrieved rate constants $k(T)$ to the Arrhenius equation

$$k(T) = A \cdot \exp(-E_A/RT), \qquad (2.70)$$

which consists of the pre-exponential factor A, activation energy E_A, gas constant R, and absolute temperature T. For convenience, the above equation is often written in a form facilitating linear fit

$$\ln\big(k(T)\big) = -\frac{E_A}{R} \cdot \frac{1}{T} + \ln(A). \qquad (2.71)$$

In the final step of our analysis, we employed this equation to generate the Arrhenius plots using different baseline removal techniques (see [63] for details), as shown in Figure 2.27. From the slope of

the fitted curves we estimated the activation energy to be 131.6 kJ/mol. Given this value, one can determine the conditions allowing for long-term storage of a drug enriched with this excipient.

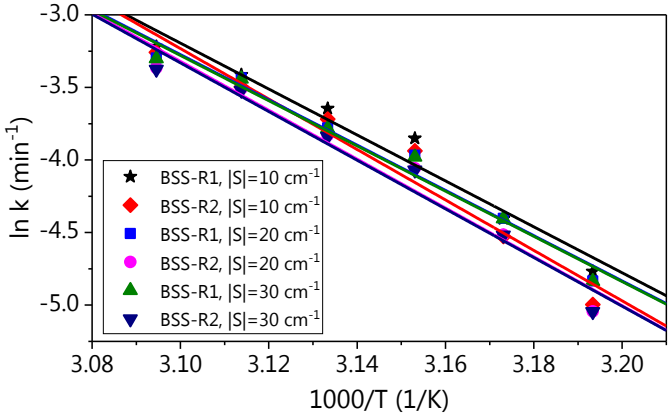

Figure 2.27. The Arrhenius plot of the α-D-glucose monohydrate dehydration generated with two *α*-evaluation methods. Results published in [63].

Chapter 3

Dual comb spectroscopy

During the recent decade, dual comb spectroscopy (DCS) has emerged as a powerful technique to probe the optical spectrum in a simultaneously broadband, and high-resolution fashion. In contrast to the well-established Fourier Transform Spectrometers (FTS) using broadband incoherent emitters, the resolution of DCS measurements is limited by the laser used rather than the instrument itself. By replacing the slow-moving mirror acquisition with optical multiheterodyning on a fast photodetector, one can perform measurements at unprecedented timescales of nano- or microseconds, which finds application in diagnostics of rapid processes such as combustion. Spectroscopic information in DCS is conveniently down-converted to the radio frequency (RF) domain, well suited for further digital post-processing. Another advantage of DCS lies in its high optical power per spectral element, which translates into greater signal-to-noise figures and better sensitivity for a given averaging time. Finally, the comb sources can be synchronized to stable oscillators such as atomic clocks, thus enabling fully-referenced measurements with a well-defined frequency axis.

Most of the above was initially demonstrated in the near-infrared regime due to the availability of optical components for the most popular telecom wavelengths of 1.2-1.6 μm and their second harmonics. To access the mid-infrared region (3-15 μm) or far-infrared (15-1000 μm), more complication in the setup is needed. Techniques utilizing nonlinear crystals like optical parametric oscillation (OPO) or difference frequency generation (DFG) are used for the down-conversion to longer wavelengths, however such systems become bulky and expensive.

A remedy to the above is to use frequency combs generated intrinsically in the cavity of a quantum cascade laser (QCL) through four wave mixing (FWM). As an electrically-pumped direct emitter of coherent mid- or far-infrared radiation with a millimeter-scale footprint, the QCL has already shown to be compatible with DCS [66]. Soon, the QCL technology will enable the construction of miniaturized, battery-operated broadband spectrometers for threat or pollution detection in both the mid-infrared and terahertz regions. As shown in this chapter, a closely related device known as an interband cascade laser (ICL) can also emit stable frequency combs suitable for DCS in the mid-infrared with unprecedentedly low power consumption comparable to that of three miniature light emitting diodes.

Unfortunately, miniaturization has its price. The relative stability of the repetition rate over ~1 s in a free-running QCL or ICL comb (10^3 Hz/10^{10} Hz = 10^{-7} Hz) [67], [68] is worse by approximately two to

three orders of magnitude compared to mature femtosecond oscillators [69]. Combined with MHz-wide common offset frequency fluctuations, this translates into difficulties with spectroscopic assessments using such noisy but still coherent sources. The measured radio frequency spectrum often turns into a featureless blob without any distinguishable multiheterodyne lines. To address these issues, this chapter provides several novel hardware and software techniques that enabled us to perform spectroscopy under such high phase noise conditions with a focus on the relevance of the techniques to molecular spectroscopy of gaseous species.

3.1. Quantum cascade laser

So far, we used the term of quantum cascade laser (QCL) without its explicit introduction. Due to the fact that it is one of the most intensively employed devices for comb generation in the mid infrared and terahertz range in this dissertation, it would be beneficial to familiarize the reader with the basic idea behind its operation. Other techniques for comb generation, such as femtosecond illumination of a photoconductive antenna for generation of a terahertz comb were intensively described in the previous chapter.

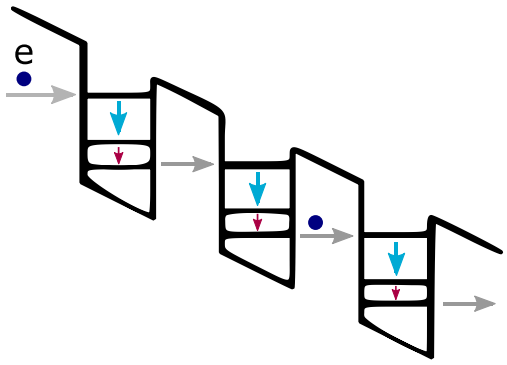

Figure 3.1. Simplified model of a quantum cascade laser. Figure inspired by [70].

In a QCL, rather than relying on a band gap characteristic of a semiconductor material as in a typical semiconductor laser, the band gap is engineered by carefully growing alternating layers of different materials on a substrate through molecular beam epitaxy (MBE), thus forming a stack of quantum well heterostructures – an artificial band structure. Some of the most popular materials are GaAs, $Al_xGa_{1-x}As$, $In_xGa_{1-x}As$, and $Al_xIn_{1-x}As$. Because their thickness can be varied, the operation wavelength of the laser can be tailored for specific application. In a QCL, one forms a cascade for electrically-injected electrons, which undergo numerous intersubband transitions. This is in contrast to most semiconductor lasers relying on interband transitions that require an electron-hole pair for photon generation. In other words, QCL is a unipolar laser capable of generating multiple photons from a single electron. In a highly simplified model, thanks to the external bias the electrons behave like droplets in a waterfall, cascading from one quantum well to another while emitting photons. Of course, it is imperative to ensure population inversion. The electron shown in Figure 3.1 (blue dots) undergoes a slow radiative transition

first (blue), followed by a fast non-radiative (red), eventually tunneling to the next stage. In fact, the gain region of a true QCL consists of a separate active region and an injector. Finally, the cleaved facets of the semiconductor material provide a sufficient reflectivity to form the optical cavity of the laser.

This brilliant idea demonstrated in 1994 by Jérôme Faist, Federico Capasso, Deborah Sivco, Carlo Sirtori, Albert Hutchinson, and Alfred Cho at Bell Laboratories in 1994 [71] revolutionized the world of mid-infrared lasers. Next, in 2012 Hugi et al. showed that optimized QCLs are capable of producing frequency combs [72], which sparked the era of intensive research on their application in dual-comb spectroscopy. This chapter provides novel software and hardware solutions to deal with some of the stability imperfections of QCL-based frequency combs, as well as a QCL derivative – the interband cadcade laser, which utilizes a cascaded structure for both electrons and holes.

3.2. Principles of dual comb spectroscopy

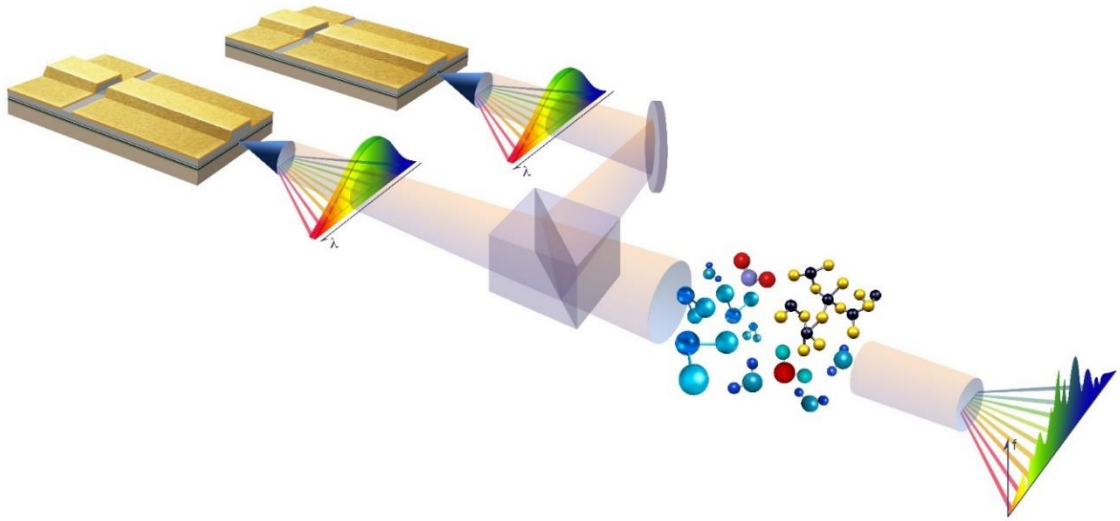

Figure 3.2. Artistic interpretation of dual comb spectroscopy. Two chip-scaled sources are combined with a beam splitter to probe the optical spectrum of a molecular species mixture. Reproduced with permission from Dr. Mahmood Bagheri of NASA Jet Propulsion Laboratory.

3.2.1. Optical multiheterodyne detection

The heart of dual comb spectroscopy lies in a simultaneous down-conversion of all optical modes of a frequency comb to the radio frequency domain. This is obtained through parallel optical multiheterodyne detection at multiple frequencies, known as multiheterodyne detection (Figure 3.2). The process takes advantage of the electrical nonlinearity of a photodetector, which outputs a voltage proportional to the optical intensity rather than the electric field, thus enabling optical heterodyning. Note that this holds for THz systems driven by two femtosecond lasers as well, where the nonlinearity arises from the optical gating on the near-infrared detection side, even though in TDS one can probe the THz electric field.

To draw a simple picture of the optical heterodyning process, we will start by combining two electric fields $\boldsymbol{E}_{\mathrm{sig}}$ and $\boldsymbol{E}_{\mathrm{LO}}$ on a photodetector, oscillating at angular optical frequencies: $\omega_{\mathrm{LO}} = 2\pi f_{\mathrm{LO}}$, and $\omega_{\mathrm{sig}} = 2\pi f_{\mathrm{sig}}$, which stand for the local oscillator (LO), and signal (sig), respectively. Assuming the two fields differ in phase by φ, we can write the intensity-proportional output of the photodetector V_{PD} as:

$$V_{\mathrm{PD}} \propto I \propto \left[E_{\mathrm{sig}}\cos(\omega_{\mathrm{sig}}t + \varphi) + E_{\mathrm{LO}}\cos(\omega_{\mathrm{LO}}t)\right]^2 \propto \frac{E_{\mathrm{sig}}^2}{2} + \frac{E_{\mathrm{LO}}^2}{2} + \\ + \frac{E_{\mathrm{sig}}^2}{2}\cos(2\omega_{\mathrm{sig}}t + 2\varphi) + \frac{E_{\mathrm{LO}}^2}{2}\cos(2\omega_{\mathrm{LO}}t) + 2E_{\mathrm{sig}}E_{\mathrm{LO}}\cos(\omega_{\mathrm{sig}}t + \varphi)\cos(\omega_{\mathrm{LO}}t). \quad (3.1)$$

Intuitively, the photodetector will respond with a constant DC term proportional to the sum of the intensities of the LO and signal, but more importantly, numerous frequency mixing terms appear. The two terms following the DC terms are second harmonic products with optical frequencies, therefore they do not appear in the bandwidth-limited down-converted signal. What is of largest interest, is the last term. From the product-to-sum formula for the cosine, we can write it as

$$2\cos(\omega_{\mathrm{sig}}t + \varphi)\cos(\omega_{\mathrm{LO}}t) = \underbrace{\cos\left((\omega_{\mathrm{LO}} + \omega_{\mathrm{sig}})t + \varphi\right)}_{\text{Sum frequency}} + \underbrace{\cos\left((\omega_{\mathrm{sig}} - \omega_{\mathrm{LO}})t + \varphi\right)}_{\text{Difference Frequency}}. \quad (3.2)$$

The sum frequency usually lies beyond the electrical bandwidth of the photodetector, in contrast to the difference frequency term, which explains the down-conversion of both the amplitude, and optical phase $\varphi = \varphi(\omega_{\mathrm{sig}})$ to the RF domain. Therefore, the AC response of the photodetector is

$$V_{\mathrm{AC}} \propto \frac{1}{2}E_{\mathrm{sig}}E_{\mathrm{LO}}\cos\left((\omega_{\mathrm{sig}} - \omega_{\mathrm{LO}})t + \varphi\right). \quad (3.3)$$

Now, let us consider optical heterodyning utilizing multiple lines, also known as multiheterodyne beating. Since the optical frequency of each comb tooth (mode) can be described using just two parameters: the optical offset frequency f_0, and the optical repetition frequency f_{rep}, the optical frequency of the n-th comb line can be expressed as

$$f_n = f_0 + nf_{\mathrm{rep}}. \quad (3.4)$$

The electric field of such a comb consisting of N lines with electric field amplitudes A_n and phases $\varphi_n = \varphi(f_n)$ is given by

$$E_{\mathrm{comb}}(t) = \mathrm{Re}\left\{\sum_{n=1}^{N} A_n e^{\mathrm{i}\varphi_n} e^{\mathrm{i}2\pi(f_0 + nf_{\mathrm{rep}})t}\right\}. \quad (3.5)$$

To enable multiheterodyne beating at different frequencies for down-converting the optical spectrum of the signal comb, we will mismatch the repetition rate of the LO comb setting $f_{\mathrm{repLO}} = f_{\mathrm{rep}} - \Delta f_{\mathrm{rep}}$. If we combine the two sources as described on the photodetector, after neglecting the DC, second harmonic, and sum frequency terms, we arrive at

$$V_{\mathrm{DCS}}(t) \propto \mathrm{Re}\left\{\sum_{n=1}^{N} A_n B_n e^{\mathrm{i}\varphi_n} e^{\mathrm{i}2\pi(\Delta f_0 + n\Delta f_{\mathrm{rep}})t}\right\}, \quad (3.6)$$

where B_n denotes the amplitude of the n-th LO comb tooth. The down-converted spectrum is essentially a radio-frequency comb with the mixing products appearing at frequencies equal to multiples of the repetition rate difference, offset by the difference in carrier-envelope offset frequencies. The n-th RF comb line lies at a frequency

$$f_{\mathrm{RF},n} = \Delta f_0 + n\Delta f_{\mathrm{rep}}. \quad (3.7)$$

A simplified illustration of the process is shown in Figure 3.3. Note that by convention, the repetition rate of the signal comb is greater than that of the LO.

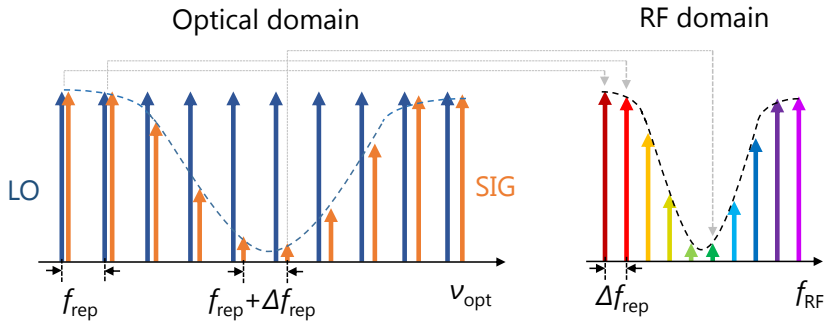

Figure 3.3. Illustration of the optical down-conversion process in dual comb spectroscopy. The two optical frequency combs with different repetition rates combined on a fast photodetector map the optical spectrum to the radio frequency (RF) domain, creating a one-to-one link between the two domains.

3.2.2. Spectral aliasing and mapping direction

One of the many risks associated with the selection of Δf_0 and Δf_{rep} in DCS is caused by spectral aliasing. If we optically beat two combs with n lines, we can obtain a situation as shown in Figure 3.4,

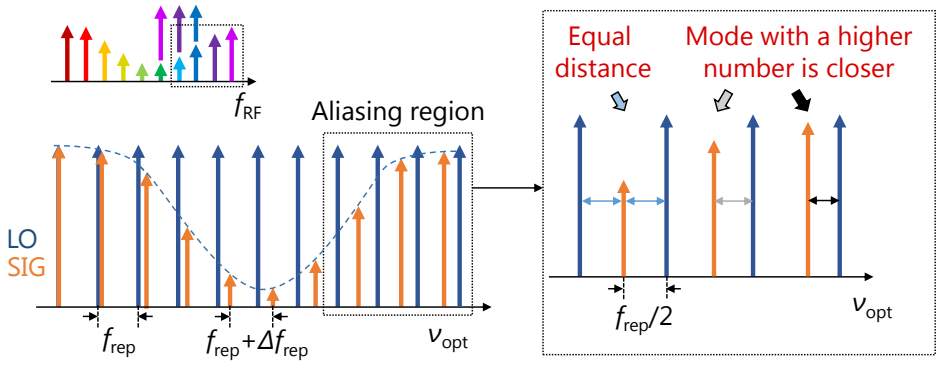

Figure 3.4. Aliasing in DCS assuming $\Delta f_0 \approx 0$. Above the mode beating at equal frequencies (whose amplitude will be doubled), the mapping direction flips. Therefore the RF beat notes mapping higher optical frequencies overlap with the lower frequency ones.

where the cumulative repetition frequency difference after m lines ($m\Delta f_{\mathrm{rep}}$) is equal to half of the repetition rate of the LO comb. In other words, the m-th line of the signal laser will optically beat with the m-th and $(m+1)$-th lines of the LO at the same (absolute) RF frequency. For simplicity assume now that $\Delta f_0 \approx 0$. Above that equidistance point, the signal comb tooth will be down-converted by higher order lines of the LO, thus mapping higher optical frequency in reverse order with decreasing RF frequencies.

In DCS we have to deal with one more issue caused by the choice of the optical offset frequency. The optical modes of the signal comb can be placed on higher frequency side of the LO so that the optical frequency of the m-th signal comb line is always greater than that of the LO. This maps the optical spectrum to the RF in a forward direction (Figure 3.5a). The opposite situation causes the optical spectrum to be mapped in a reverse direction, as shown in Figure 3.5b. Interestingly, a two-step acquisition in both mapping directions can be used for the suppression of photodetector nonlinearities, as in Ref. [73].

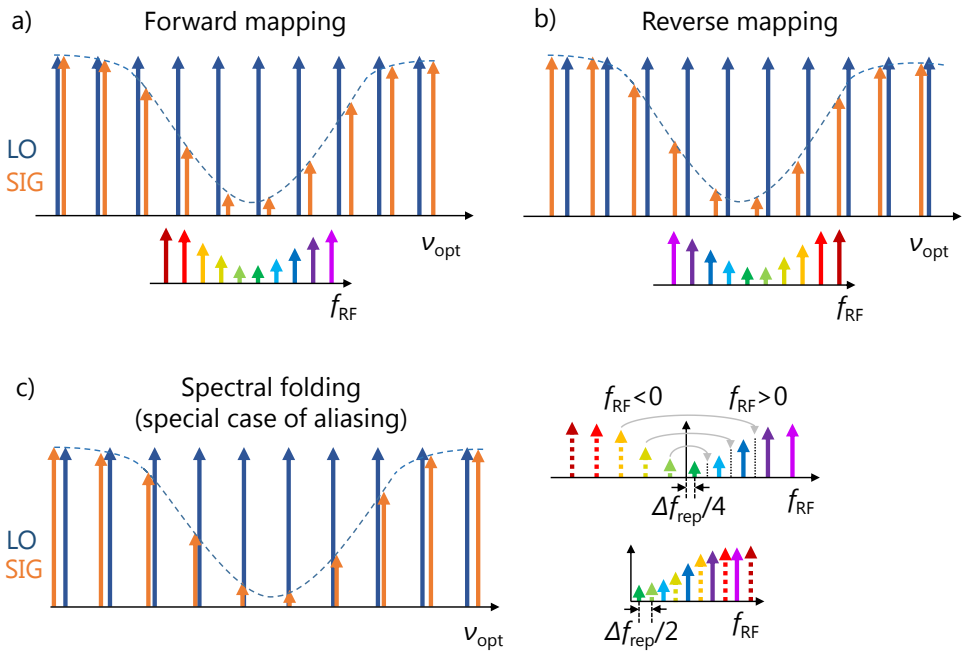

Figure 3.5. Forward (a) and reverse (b) mapping of the optical to the RF domain. Spectral folding to double the number of down-converted modes within a given bandwidth is shown in (c).

A commonly encountered case is when some of the signal comb teeth lie on the right, and some on the left of the LO, thus resulting in an aliased spectrum, as plotted in Figure 3.5c. Negative frequencies in a real baseband signal from the photodetector are indistinguishable from the positive (spectrum is symmetric), therefore they get mapped to the positive frequency side. While this effect in most cases results in a spectral congestion or even beat note overlap, there exists a special case that enables a doubling of the number of the mapped optical modes within a given photodetector bandwidth, originally

proposed by Schiller [74], albeit at the expense of more complicated analysis. By placing the lowest positive frequency RF beat note at $\Delta f_{\mathrm{rep}}/4$, obtained by tuning the offset frequency difference, we will reverse-map every other optical mode. As a result, the RF spectrum consists of beat notes evenly spaced by $\Delta f_{\mathrm{rep}}/2$, half of which is forward mapped, and half is reversed. This procedure was used in our early multiheterodyne experiments due to the large mismatch between the repetition rates of the sources compared to the electrical bandwidth of the photodetector.

3.2.3. Spectral replicas

So far, we assumed that the electrical bandwidth of our photodetector is relatively low compared to the repetition rate and the number of comb lines, thus only the low frequency RF beat notes are accessible. This is generally true for DCS using QCLs in the mid-infrared, heterodyned on a HgCdTe (MCT) detector. However ultrafast near-infrared photodetectors, quantum well infrared photodetectors (QWIPs), and THz hot electron bolometers allow to observe numerous spectral replicas (also known as images) originating from beating with comb lines lying more than a repetition rate apart. It is easy to imagine that each signal comb tooth will be down converted by a respective LO comb tooth together with a mode that lies one repetition rate further, two repetition rates, and so on. Additionally, lower order LO modes will down-convert that tooth as well. This is similar to periodicity of the Discrete Fourier Transform, but here it is caused by the discrete spacing of comb lines in the optical domain. Mathematically, the frequency of each beat note detected with an ultrafast photodetector can be expressed as

$$f_{\mathrm{RF},n,k} = \left|\Delta f_0 + n\Delta f_{\mathrm{rep}} + k f_{\mathrm{rep}}\right|, \qquad (3.8)$$

where k is an integer denoting the difference in mode numbers between the LO and signal teeth. Again the real nature of the beating signal implies that negative frequencies are folded onto the positive frequency side. Note also that the repetition rate beat note can be optically detected too. For illustrative purposes, the RF spectrum of a fast photodetector is plotted in Figure 3.6.

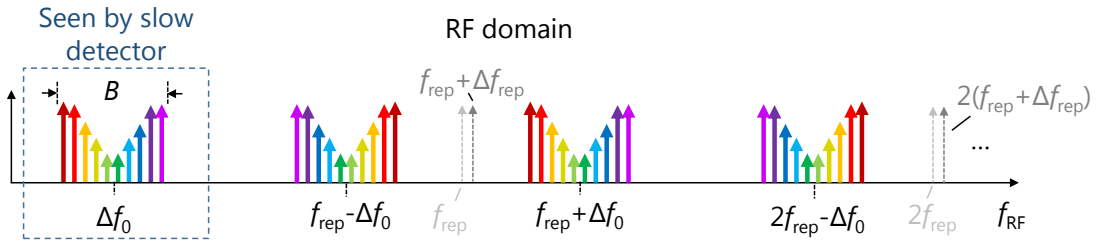

Figure 3.6. Spectral replicas visible in the RF spectrum obtained by beating two combs on a fast photodetector. Note that the forward- and reverse-mapped beat notes move in opposite directions when Δf_0 is varied.

To give a real-world example, in our THz QCL experiments, a forward mapped, 500 MHz-wide multiheterodyne spectrum was offset from DC by Δf_0=2 GHz. Given that the repetition rate of the THz

combs was $f_{\mathrm{rep}} \approx 9$ GHz, a reverse mapped (mirrored) spectrum appeared centered at $f_{\mathrm{rep}} - \Delta f_0 =$ 7 GHz, which lied within the bandwidth of the fast bolometer. By changing the injection current of the QCLs, one can tune the offset frequency, which causes the forward and backward mapped spectra to move in opposite directions. In extreme cases, both can perfectly overlap if $\Delta f_0 = f_{\mathrm{rep}}/2$.

We clearly see that the optical bandwidth of the combs, repetition rate difference, and offset frequencies must be carefully chosen with the fundamental requirement: the number of comb lines times Δf_{rep} must be lower than half of f_{rep}, which can be doubled for complex acquisitions [75].

3.2.4. Experimental configurations

Dual comb spectroscopy can be performed in three general configurations, shown schematically below.

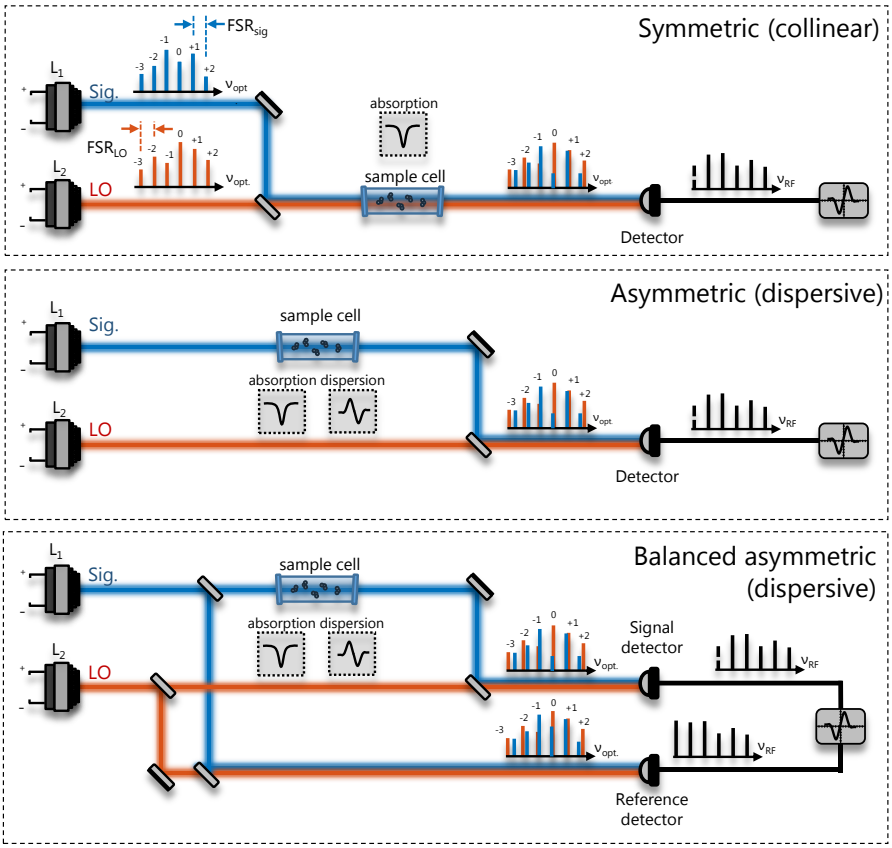

Figure 3.7. Three experimental configurations for dual comb spectroscopy: symmetric, asymmetric, and balanced asymmetric.

The first is known as symmetric (collinear), wherein the LO and signal combs are combined before passing through an absorber. This configuration is analogous to a Fourier Transform Spectrometer and is more suitable for turbulent environments [75]. Unfortunately, it allows only for absorption measurements, which is mitigated in the asymmetric (dispersive) configuration by placing the absorber in the signal comb beam path before the beams are combined. Each beat note now carries information about both amplitude, and optical phase. The last configuration – balanced asymmetric – is used for

amplitude- and phase-noisy combs, as often generated from QCLs or ICLs. It employs a pair of balanced detectors to help suppress common intensity and frequency fluctuations, unfortunately at the expense of additional optical components.

Initially, all our experiments employed the balanced asymmetric configuration due to significant fluctuations in mode intensities and optical frequencies. In the THz setup on the contrary, we had access to only one superconducting bolometer, which limited us to the asymmetric arrangement. To enable accurate spectroscopic measurements with one detector, we have developed analog and digital phase stabilization techniques to suppress frequency jitter effects, which allows for averaging of noisy multiheterodyne signals over extended acquisition times. These developments, which were implemented in all setups, eventually paved the way for the demonstration of single-detector DCS with free-running interband cascade laser frequency combs, i.e. when no feedback loops were used but just digital phase correction.

3.2.5. Calculation of transmission and phase

For clarity, let us explain how the optical transmission is calculated in the three configurations. First, note that the MH beat note voltage at a given frequency is proportional to the product of the electric field amplitudes of the LO and signal and lines.

$$V_{\mathrm{AC}}(\omega_n) \propto E_{\mathrm{sig}}(\omega_{\mathrm{sig},n}) E_{\mathrm{LO}}(\omega_{\mathrm{LO},n})\,. \tag{3.9}$$

The transmission is of our interest; therefore, we will replace the voltage with an electrical power P to obtain an intensity-dependent quantity:

$$P_{\mathrm{AC}}(\omega_n) \propto V_{\mathrm{AC}}^2(\omega_n) \propto I_{\mathrm{sig}}(\omega_{\mathrm{sig},n}) I_{\mathrm{LO}}(\omega_{\mathrm{LO},n})\,. \tag{3.10}$$

Since in each configuration we perform two measurements: an empty cell (zero gas Z), and a sample gas (G), we can denote them simply as P_{Z}, and P_{G}, respectively, keeping in mind that they vary with ω. By definition, the transmission is simply

$$T = \frac{P_{\mathrm{G}}}{P_{\mathrm{Z}}}\,. \tag{3.11}$$

The above, however, does not include the reference branch in the balanced asymmetric configuration, where we measure the signal from the sample (S), and the reference (R) detector simultaneously. To account for fluctuations between the zero gas, and absorber measurements, we will include the four measured quantities in the transmission formula:

$$T = \frac{\dfrac{P_{\mathrm{GS}}}{P_{\mathrm{GR}}}}{\dfrac{P_{\mathrm{ZS}}}{P_{\mathrm{ZR}}}} = \frac{P_{\mathrm{GS}}}{P_{\mathrm{GR}}} \cdot \frac{P_{\mathrm{ZR}}}{P_{\mathrm{ZS}}}. \tag{3.12}$$

where, the first letter in the subscript of P denotes the acquisition (zero-gas/gas), and the second stands for the detector (sample/reference).

The optical phase is calculated in a similar way. In the first two cases, we calculate the difference between the zero-gas and sample phase at a given frequency, whereas in the balanced configuration we calculate the difference between the two detectors with an absorber, and without, which is finally subtracted for comparison. The derivation of optical constants follows the formulas from the previous chapter.

3.3. Frequency and phase locking

3.3.1. Instabilities of frequency in dual-comb spectroscopy

While spectroscopy performed in a dual-comb configuration offers many advantages such as moving parts free acquisition of simultaneously broadband and high-resolution spectra in microsecond timescales, it puts sharp requirements on the relative stability between the sources. For instance, two free-running (unsynchronized) frequency combs will typically drift in carrier-envelope offset (CEO) frequencies f_0, causing the individual RF beat notes to be dispersed over a wider bandwidth if averaged over longer timescales. In the near-infrared, where commercially available fiber-based femtosecond oscillators are technologically mature, dedicated locking circuits can be embedded into the control electronics for mutual synchronization, however for QCL- and ICL-based dual-comb spectroscopy such procedures have not yet been well established and needed to be developed.

Let us recall the frequency comb model after down-conversion to the radio frequency (RF) domain. Rather than being stationary, the instantaneous radio frequency of the *n*-th comb line varies with time, and can be expressed by

$$f_{\mathrm{RF,n}}(t) = \Delta f_0(t) + n\Delta f_{\mathrm{rep}}(t). \qquad (3.13)$$

The above equation can be divided into two terms. The first one is called an RF offset frequency Δf_0 and represents the difference in the optical carrier envelope offset frequencies f_{01} and f_{02} of the combs. Since in the optical domain, fluctuations of f_{0n} shift the optical modes equally, the same property is preserved in the RF domain, giving rise to an equal shift in the position of multiheterodyne beat notes. In other words, it causes common mode multiplicative phase noise in the multiheterodyne signal, which broadens the beat notes or causes overlap. This effect is illustrated in the top panel of Figure 3.8.

The second term of Eq. 3.13 corresponds to non-stationarity in the repetition rates of the combs. In the optical domain, changes in f_{rep} cause a mode-number-dependent shift of the optical frequency and resemble "spectral breathing": the central part of the spectrum remains at a nearly constant position, while the shift for modes positioned further from the center gradually increases proportional to the distance, as shown schematically in the bottom panel of Figure 3.8.

One can ask: which effect is typically more severe? In a QCL- or ICL-based multiheterodyne system, fluctuations of the repetition rate are approximately two orders of magnitude lower than those of the offset frequency [76], [77], which reach MHz level compared to tens of kHz of the repetition rate.

Indeed, if we consider the number of modes and beat note spacing in our initial experiments, it was always the offset fluctuations that predominantly limited the performance of the spectroscopic system.

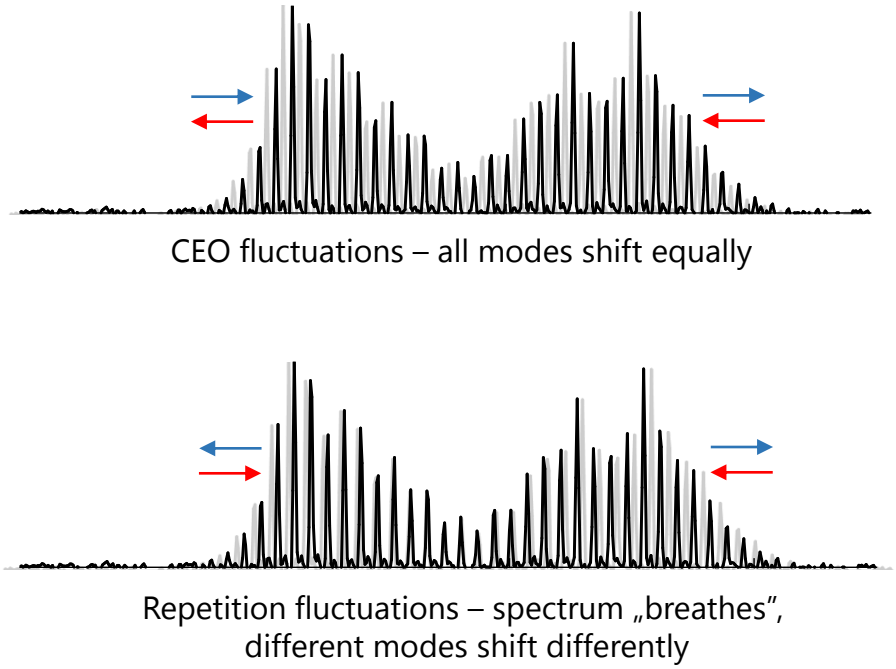

Figure 3.8. Two kinds of frequency instabilities of frequency combs. (Top) fluctuations of the optical offset frequency. (Bottom) fluctuations of the repetition rate.

However, as the number of beat notes in the RF spectrum exceeds a hundred, the effect of f_{rep} instabilities become comparable to f_{ceo} due to its cumulative behavior. In femtosecond laser based systems, the number of beat notes *n* can reach tens or hundreds of thousands, and the instability of f_{rep} is a major contributor to the broadening of the beat notes.

A convenient advantage offered by sources for DCS relying on difference frequency generation (DFG) such as previously shown THz photoconductive antennas in the THz, or nonlinear crystals in the mid-IR [78], is that the generated radiation is offset-free (harmonic) after down-conversion. For a DCS system it means that out of the initial four, only two variables are to be corrected for their noisy behavior. As DFG sources show an increasing trend in efficiency (nearly 15% for NIR-THz conversion using plasmonic photoconductive antennas), they are likely to be widely used in future DCS system due to the fact that the repetition rate is much easier to measure optically or electrically extract than the offset, and can be accounted for in a much easier way. On the other hand, they do not offer the tunability through offset frequency change while preserving a constant repetition rate [79].

A natural question that arises, is whether apparently unstable comb sources can be harnessed for spectroscopy? It turns out that even commonly encountered megahertz fluctuations in the QCLs are small compared to the mid-IR and THz spectral linewidths of absorbing gas species at atmospheric pressure lying in the gigahertz range, not to mention solids or liquids with linewidths of even tens or

hundreds of GHz. Although a QCL can be stabilized for repetition rate through injection locking [80] or an active PID loop, followed by locking a laser to a narrow molecular line or an additional DFB laser, such schemes introduce additional complexity to the spectroscopic system, which may not be necessary for the desired application. In this work, rather than focusing on the spectroscopy of hyperfine transitions with sub-hertz resolution reserved for the domain of femtosecond lasers referenced to an atomic clock, we will put stress on the development of *relative frequency stabilization* techniques to enable prolonged coherent averaging of small-footprint QCL/ICL MHS, thus accepting MHz drifts throughout the acquisition. This is also dictated by the hierarchy of problems to tackle in QCL MHS: two free running lasers with a close beat note spacing of single MHz usually do not have resolvable discrete lines even within tens of microseconds until they are mutually locked. Consequently, we need to keep the RF beat notes stable first to observe their discrete nature, and account for absolute frequency drifts later.

In this chapter three different approaches for mutual locking will be presented in incremental steps. The first two modulate the laser current and can be seen as active techniques: a frequency locked loop (FLL), an optical phase locked loop (OPLL). The third one – a digital phase locked loop – is a coherent averaging scheme that corrects for frequency instabilities in post-processing but shows a potential for real-time FPGA implementation. We will start by describing the active techniques that pre-lock two QCLs and illustrate their relevance for spectroscopy of molecular absorbers.

3.3.2. Frequency locked loop

To illustrate the scale of drift of a multiheterodyne spectrum obtained by beating two 1 mm free-running Fabry-Pérot (FP) QCLs centered at ~8.5 μm, we characterized the position of the lowest frequency RF beat note from Figure 3.9 over the course of approximately 16 minutes, plotted in Figure 3.10. What should be noted is that the lasers were mounted on the same chip and in principle share the same thermal drift. Thermal instabilities can severely corrupt any MHS experiment, therefore a high precision (10 mK) thermoelectric cooling system was used with a continuous output PID control loop to operate the lasers at ~0°C while injecting ~250 mA of current at ~8 V to each.

To track the peak, we fitted a Gaussian to spectral snapshots acquired every few seconds using a spectrum analyzer whose observation window was set to 50 MHz. The beat note's center shifted ~20 MHz within the observation time by in a nearly linear fashion, which we can extrapolate to a rate of ~70 MHz/h. After more than an hour, the long-term averaged spectrum would not show resolvable frequency components. This is a simple consequence of the Fourier Transform's properties: it is not suitable for non-stationary signals since the amplitude would be distributed over a wide range of frequencies. In the next subsections, we will show under optical feedback conditions, such effects are obtainable in just 500 μs.

The drift observable in the RF domain is a result of changes in CEO of either one or both lasers. Without an octave-spanning source for f_{ceo} retrieval or a single mode CW laser reference, it is nearly impossible

to tell which one drifted, since the observed drift stems from the difference in f_{ceo}. What also deserves attention is that the shot-to-shot variation in the peak position is on average 1 MHz plus some larger oscillations occurring after 10 and 13 minutes.

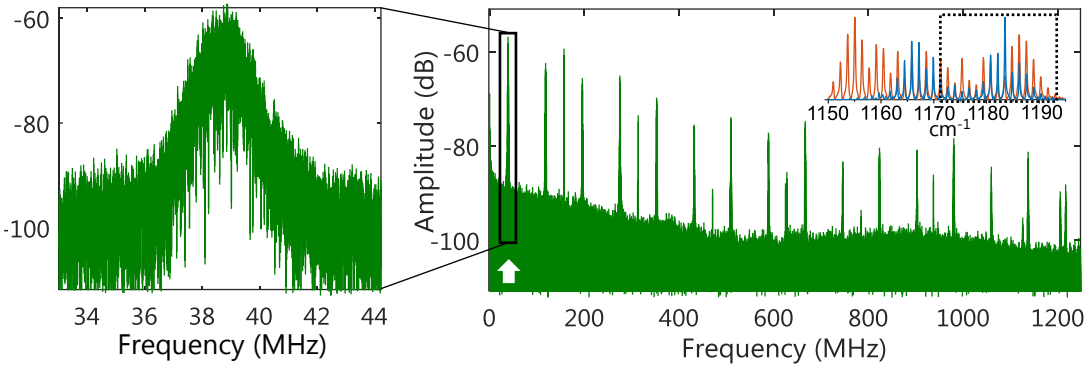

Figure 3.9. Multiheterodyne spectrum of two free-running 1 mm FP-QCLs centered at ~8.5 μm, which were operated at ~250 mA and ~8 V (FTIR spectra in top corner of the right panel) and detected with a fast photodetector within 500 μs. Though the RF spectrum plotted over a gigahertz span seems to show multiple narrow beat notes, a closer look at the strongest one (left panel) reveals the width of the noise pedestal on the order of ~6 MHz despite the short acquisition time. This peak was tracked for drift over extended timescales. The lasers for this experiment were provided by Professor Federico Capasso at Harvard University and his collaborators Dr. Laurent Diehl and Dr. Mariano Troccoli.

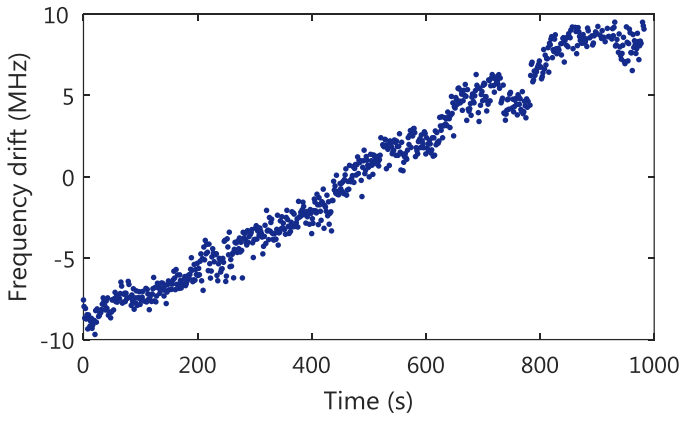

Figure 3.10. Drift of the lowest frequency beat note evaluated over ~16 minutes.

We proposed to stabilize the position of the beat note in the RF domain by inter-locking the two lasers denoted as local oscillator (LO) and signal (SIG). A pair of optical modes or equivalently comb tooth (one from each source) is ensured to beat in the RF domain at a constant frequency, effectively meaning that the f_{ceo} of the lasers becomes synchronized, allowing one laser to track the other. To experimentally realize this, we proposed a home-made Frequency Locked Loop (FLL), shown schematically in Figure 3.11. To understand how it works, one has to recall that the optical frequency of semiconductor lasers tunes with temperature and current. Both tuning mechanisms are closely related since tuning is obtained

predominantly through nonlinear changes of the temperature-dependent refractive index of the medium, which in turn can be caused by either a change in the temperature of the submount or a change of the laser's internal temperature due to its thermal and electrical resistance, and the amount of the supplied electrical power (voltage-current product) [81]. It is estimated that other effects, such as an increase of the physical cavity length is responsible only for 10% of the tuning. We analyzed the tuning properties of QCLs in the previous section but the main point here is that by changing the current, we can change the lasing frequency in both directions. From the electrical engineering standpoint, the laser can be seen as a current (voltage) controlled oscillator: a key element of a frequency stabilization feedback loop.

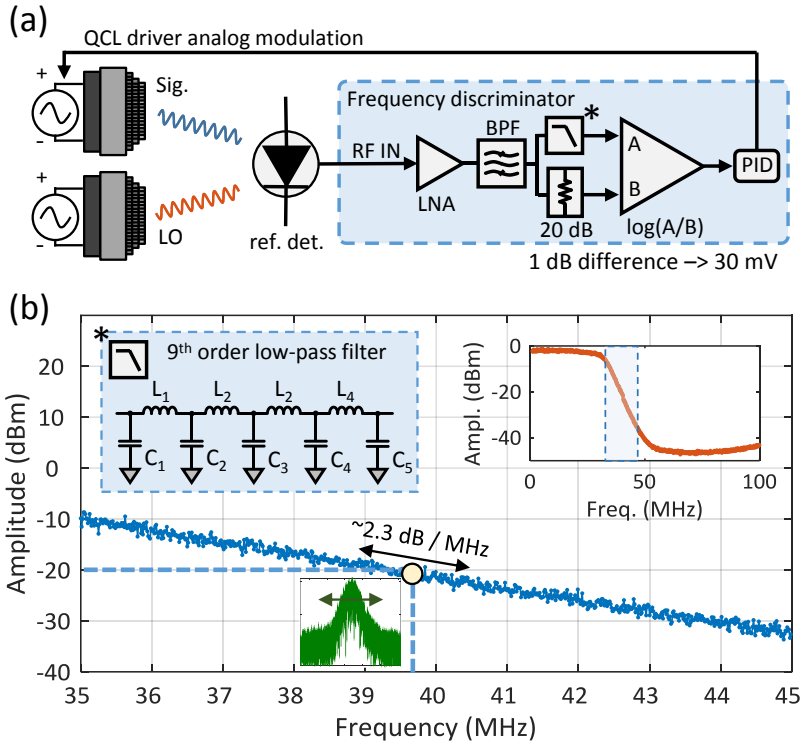

Figure 3.11. (a) The principle of operation of the frequency locked loop (FLL) in a multiheterodyne system. Two QCLs beating optically on a fast photodetector create an RF signal amplified by a low noise amplifier (LNA) followed by a band-pass filter (BPF) that isolates a single beat note from the multiheterodyne spectrum. Next, the signal is split into two branches: a frequency-dependent slope acting as a frequency discriminator, and a fixed value attenuator. Both are fed to a log-amplifier comparing the frequency-dependent power difference, thus converting a shift in frequency into a voltage suitable for the PID controller, which modulates one of the lasers. (b) Frequency characteristics of the custom-made RF low-pass filter with a sharp cut-off to maximize the frequency-to-voltage conversion slope while preserving a wide locking range. Results published in [82].

Equipped with a mechanism to control the process value – optical frequency of one of the lasers – we can formulate the task in the language of control engineering feedback loops: minimize the deviation of beat note's frequency through a control signal. In the first step, we need to convert the frequency

deviation into an error signal. The solution implemented in our systems proved to be robust and universal and is known in the field of radio engineering as frequency discriminator, drawn schematically in the top panel of Figure 3.11.

A frequency discriminator was used in early FM radio receivers as part of slope detection and essentially consists of a tuned RF filter with a cut-off frequency close to the received frequency with maximally flat roll-off characteristics. A frequency modulated signal fed into such a filter converts changes in frequency into amplitude, thereby being demodulated. Unfortunately, the range of slope's linearity in general was limited, and today more efficient ways of FM demodulation are used (and employed in the signal processing routines in this work too). Nevertheless, the value of slope detection has been revisited in our system due its robustness and ease of implementation. In our circuit, this slope is realized by a custom-built 9^{th} order Butterworth filter tuned at approximately 35 MHz, whose measured low-pass characteristics are plotted in Figure 3.11b. For convenience, the filter design is described in detail in the Appendix.

Since a multiheterodyne signal consists of multiple beat notes compared to only one required for the locking, it needs to be pre-conditioned before being fed to the discriminator. To match the required power level, it is first amplified by a low-noise RF amplifier (LNA), followed by band-pass filtering to isolate the beat note we want to lock to, and eventually fed to the two branches of the log-gain comparator. The last block is just a fast analog PID controller (SRS SIM960, 100 kHz bandwidth) which drives the modulation input of the laser current driver. For practical reasons, it is also desirable to operate on large control signals (on the order of volts), which are attenuated right before being fed to the laser current driver.

The characterization of the locking effectiveness is summarized in Figure 3.12. The top panel shows the Allan-Werle deviation of the free running and locked system. A multiheterodyne beat note between two free-running QCLs undergoes a drift of 320 kHz after just 10 seconds, while in the locked case it is decreased by two orders of magnitude. Panel (b) plots the frequency detuning data over 1000 seconds. Within that time, the multiheterodyne beat note remained at a constant position and no RF drift could be observed.

The locking loop is of large importance, because in addition to making the two lasers to follow each other, it enables to take full spectroscopic advantage of the narrow linewidth of the QCL modes (comb teeth) and the phase coherence between them. All the modes can be simultaneously swept through an absorber and probe different parts of the optical spectrum at the same time, which is in stark contrast to a single-mode laser. In the following subsection, we will provide a proof-of concept demonstration.

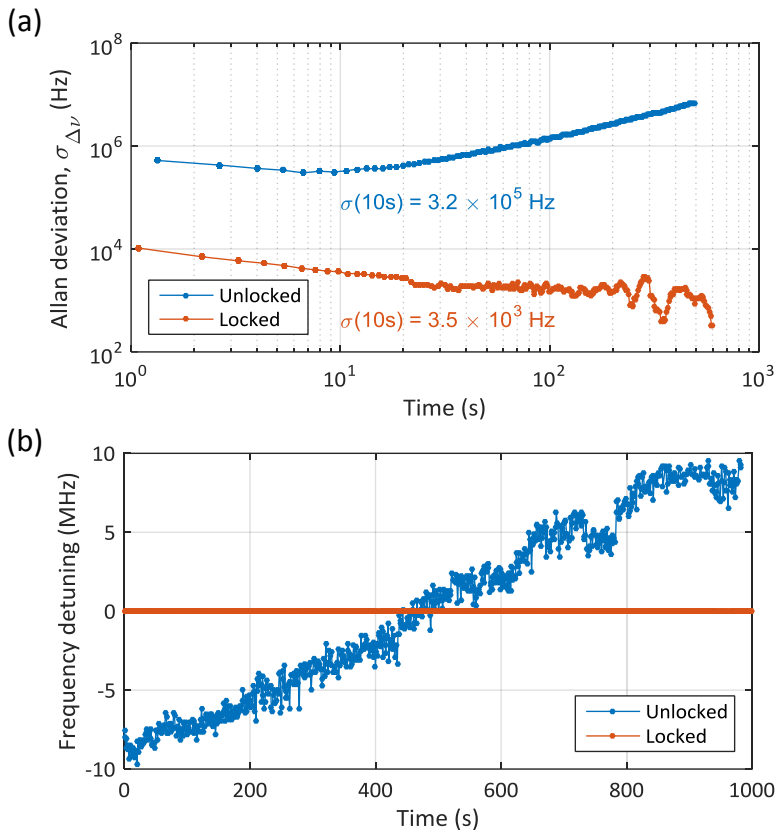

Figure 3.12. Comparison of the unlocked and locked beat note case. (a) Allan-Werle plot of the peak position. (b) Frequency detuning versus time. Clearly, the locking circuit enables to keep the beat note at a constant position over extended time scales. Results published in [82].

3.3.3. Application of FLL to swept high resolution absorption and dispersion spectroscopy of low pressure nitrous oxide in the 8.5 µm region

The dynamic performance of the locking circuitry is reflected in the wide frequency detuning range it can handle. The two QCLs could be detuned by more than 700 MHz in less than a millisecond while keeping the RF spectrum at a constant position. To get a better picture of the swept absorption and dispersion experiment, one can imagine that each beat note acts as a radio station at its own frequency. In the multiheterodyne spectrum we can tune to each of them independently to listen to the instantaneous amplitude, hence obtaining spectroscopic information. Additionally, the phase can be retrieved too, however the latter works the most effectively in balanced configuration when comparing it with a reference. Of course, all of our "radio stations" can be accessed simultaneously. Figure 3.13 shows the experimental setup used for swept acquisitions. In short, it is a balanced asymmetric configuration with two band-pass filters following each detector to isolate the lowest frequency beat note. The signal filtered from the sample and reference detector is phase-compared in an Analog Devices AD8302 board, while simultaneously measuring the amplitude. This hardware implementation can be replaced with computational demodulation of the phase and amplitude, as was done in our more recent setups.

To demonstrate the spectroscopic usefulness of our system, we measured the R35e transition of nitrous oxide (N_2O) at 1199.3576 cm^{-1}, accessible at 40 MHz in the multiheterodyne spectrum. The pure gas in a 100-mm absorption cell was kept at low pressure of 10 Torr in order to narrow its linewidth. The transmission spectrum together with a HITRAN fit based on a Voigt absorption profile, as well as dispersion spectrum obtained through K-K relations is plotted in Figure 3.14. Overall, the measurement shows good agreement with the fit, however slight oscillations are visible in the residuals of both transmission and phase.

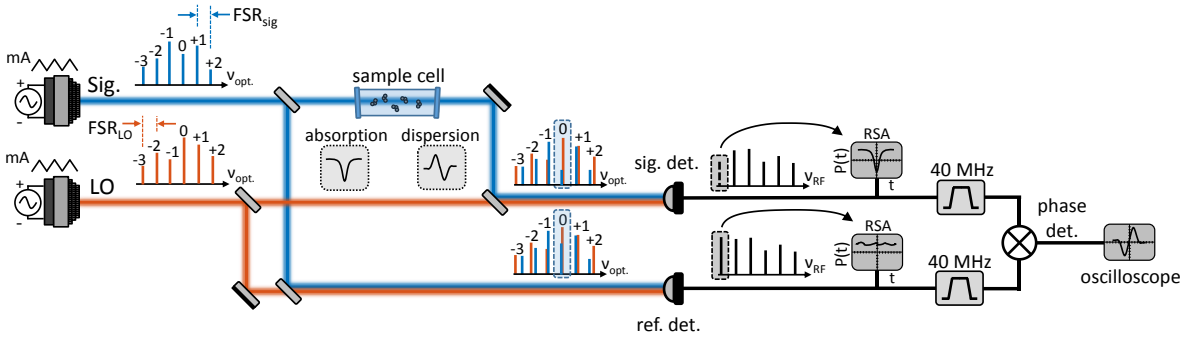

Figure 3.13. Experimental setup for performing swept absorption and dispersion spectroscopy in a dual comb configuration with frequency locked QCLs. The lowest frequency beat note is isolated using a band-pass filter, and phase-compared in an Analog Devices AD8302 board to get the dispersion spectrum. Results published in [82].

The standard deviation of residuals from the fit σ_0 is included in an important figure of merit in spectroscopy to characterize the system performance: the noise equivalent absorption (NEA). It is simply σ_0 multiplied by the square root of the acquisition time $\sqrt{\tau}$:

$$\mathrm{NEA} = \sigma_0\sqrt{\tau}. \qquad (3.14)$$

The noise equivalent absorption is in units of $Hz^{-1/2}$. For our system, the absorption can be deduced from both transmission and dispersion spectrum, equal to approximately $3\times10^{-4}/\sqrt{\mathrm{Hz}}$. NEA describes what changes of intensity $\Delta I/I$ can be detected within a given measurement time. Conventionally to convert this figure into high spectroscopic sensitivity (ppm/ppb level), one needs to cause the laser light to propagate through a low-concentration c absorber over a long (meters to kilometers) distance l to cause a large change in optical intensity ΔI, which by Beer-Lambert law is proportional to a product of the two.

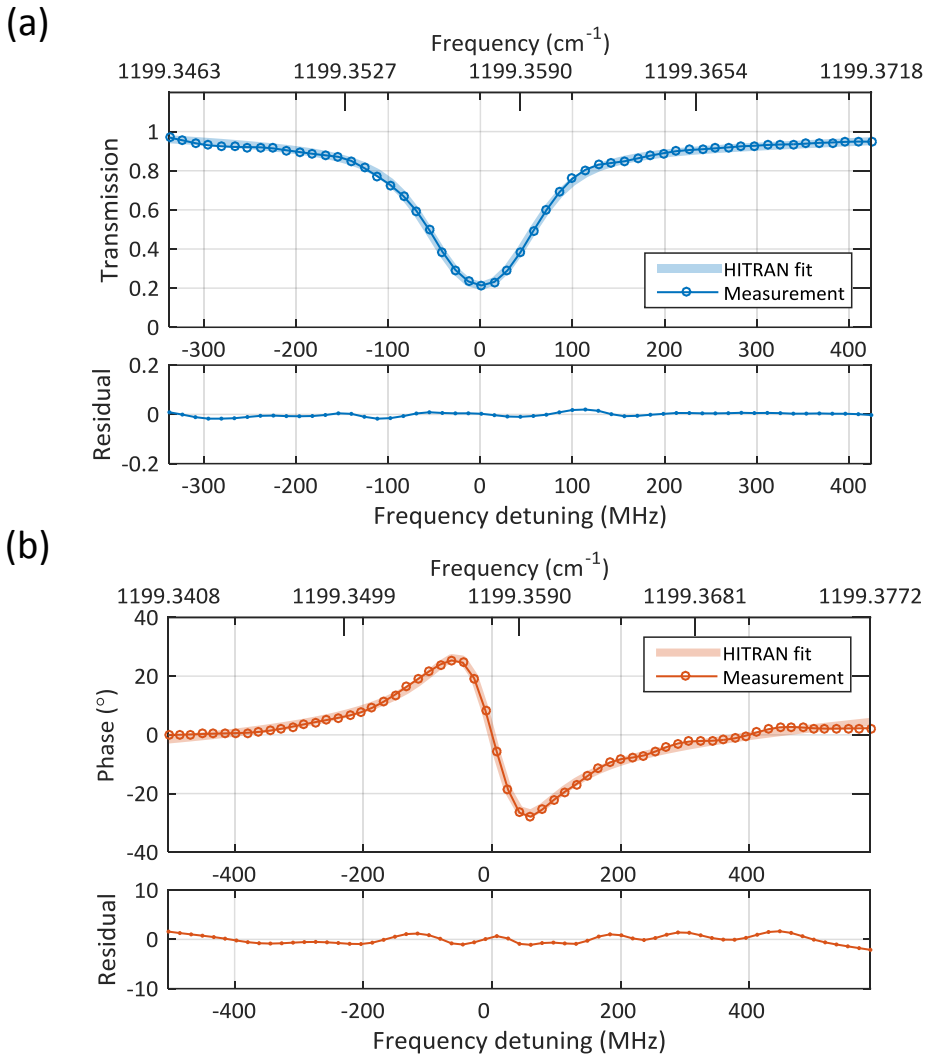

Figure 3.14. Absorption (a) and dispersion (b) spectrum of the R35e transition of N_2O at 1199.3576 cm^{-1}. Each data point represents a time average over 8 μs. The panels below each plot show a residual from the fit. Results published in [82].

3.3.4. Optical phase locked loop

So far, we have discussed the possibility of relative stabilization of two lasers using a frequency locked loop (FLL). A much better scheme employed in a majority of modern electronic devices is a phase locked loop (PLL). Its block diagram is plotted below.

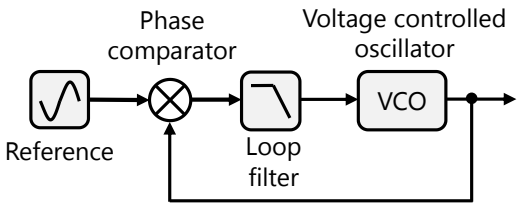

Figure 3.15. Block diagram of a phase locked loop.

A PLL is a closed-loop system, but unlike the FFL, it needs a reference frequency source. To oscillate in phase, the output signal is compared with the reference (using for instance a mixer), further filtered, and eventually fed to a voltage controlled oscillator (VCO). It is easy to extrapolate this electronic scheme to the optical domain. As we mentioned before, a QCL can be seen as a voltage (current) controlled oscillator. By comparing one of the beat notes from the photodetector against a fixed phase oscillator, we can derive an error signal suitable after some filtering to dynamically modulate one of the lasers, thus keeping both phase-locked. Clearly, we have all necessary elements to set up an optical phase locked loop (OPLL).

Due to the high bandwidth requirements of the OPLL, we employed in our experiments a commercially available module from Toptica (mFALC110), which shows a 90-degree phase delay at 45 MHz. Despite such wideband operation, it often unlocked due to its narrow capture range (<100 kHz). This was circumvented by combining the FLL and OPLL together to pre-lock the beat note first with FLL. A hybrid of the two enabled a locked operation over hours. An experimental schematic of the hybrid locking system together with a demonstration of its effect on the RF spectrum is shown in Figure 3.16.

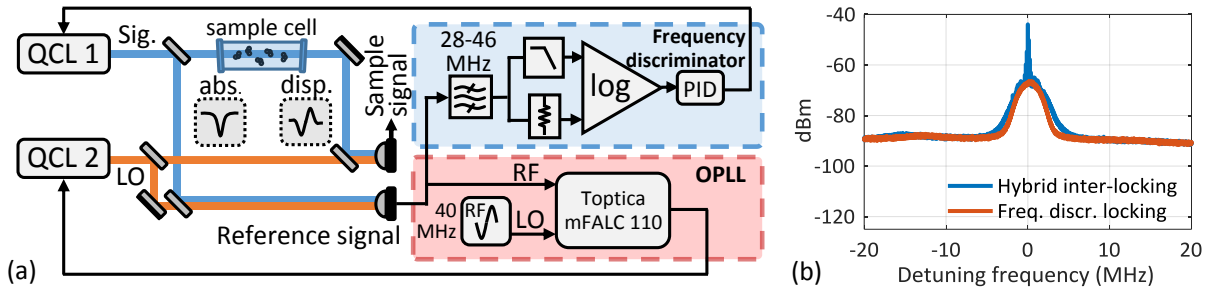

Figure 3.16. Schematic of the hybrid inter-locking system and the effect of its operation on the shape of the beat note. (a) The RF signal is split into two parts, where the upper part feeds the FLL, whereas the lower is compared with a reference 40 MHz oscillator. The error signal is processed internally by a PID controller embedded in the Toptica mFALC 110 module, which modulates one of the lasers. Note that the laser driver limits the overall bandwidth of the loop to 3 MHz. (b) The narrowing of the beat note, characteristic for phase locking. Results published in [83].

Compared to the FLL, the OPLL enables a compression of most of the RF power within a narrow bandwidth, which gives a significant improvement of signal-to-noise (here ~20 dB). 3 dB line widths ranging from 11 to 220 kHz have been observed, where the broadest lines lie far away from the locked beat note, which is caused by the cumulative effect of the repetition rate fluctuations. In other words, the OPLL can only account for the offset frequency fluctuations without stabilizing the repetition rates. Finally, one must remember that this loop stabilizes the lasers only relatively, hence the optical linewidth is not reduced. Despite this notable imperfection, the relevance of the inter-locking system over extended time scales is evident in the frequency stability plot in Figure 3.17.

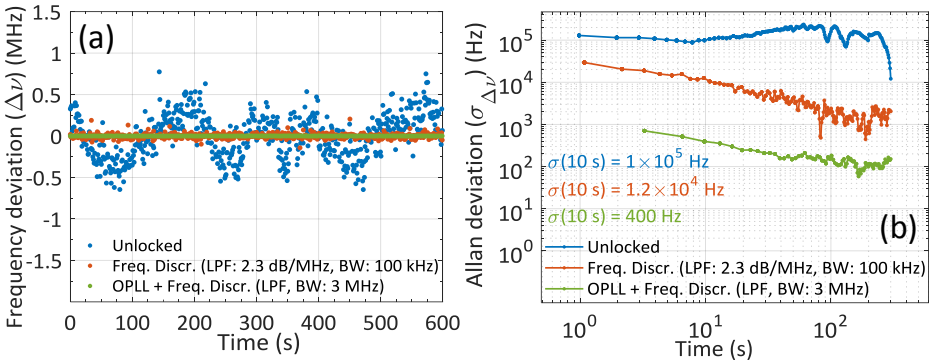

Figure 3.17. Comparison of the beat note frequency stability of the three modes of operation: free-running, frequency-locked, and phase-locked. (a) Time series measurement. (b) Allan deviation of data in (a). Results published in [83].

The superiority of the OPLL is clearly visible in the time series plot. From all measurements, its shape is the closest to a straight line, which is the expected for a tight lock. This observation is additionally supported by an Allan deviation analysis. The OPLL improves the frequency stability by almost two orders of magnitude compared to the FLL, and almost three compared to the free-running case, when evaluated after 10 seconds. Let us see, how the improvement in signal-to-noise and frequency stability translates into spectroscopic performance.

3.3.5. Application of OPLL to broadband high-resolution absorption spectroscopy of ammonia and isobutane

Using the same pair of lasers as before, we performed spectroscopic measurements of 50% ammonia at atmospheric pressure, and 15% isobutane (C_4H_{10}) together with HITRAN and PNNL database fits, as shown in Figure 3.18. Whereas the first is a rich spectral absorber with numerous absorption features across the entire frequency range, the second is broadband and spectrally-smooth. Compared to the first generation of the system [84], the OPLL improved the NEA by a factor of 5 for the ammonia measurement down to $3\times10^{-4}/\sqrt{\text{Hz}}$, which was not the ultimate limit of the system due to interference from optical fringes. Improvements in the alignment of the absorption cell in the isobutane case allowed us to reach an unprecedented NEA of $\sim2.8\times10^{-5}/\sqrt{\text{Hz}}$, which corresponds to approximately 2% error in transmission. What should be underlined that the optical spectra were measured in just 10 microseconds per mode. The non-simultaneous measurement of the spectrum was due to a bandwidth limitation of our spectrum analyzer, and in principle the same performance can be reached by acquiring all beat notes at once. At this stage, we demodulated the instantaneous amplitude around each beat note and no Fourier Transform was used for power estimation. The origin of the low NEA can be deducted from panels (c) and (d) of Figure 3.18. The stability of the beat note amplitude in the OPLL case outperforms the FLL and free-running case, equal to approximately 2×10^{-2} after 10 μs.

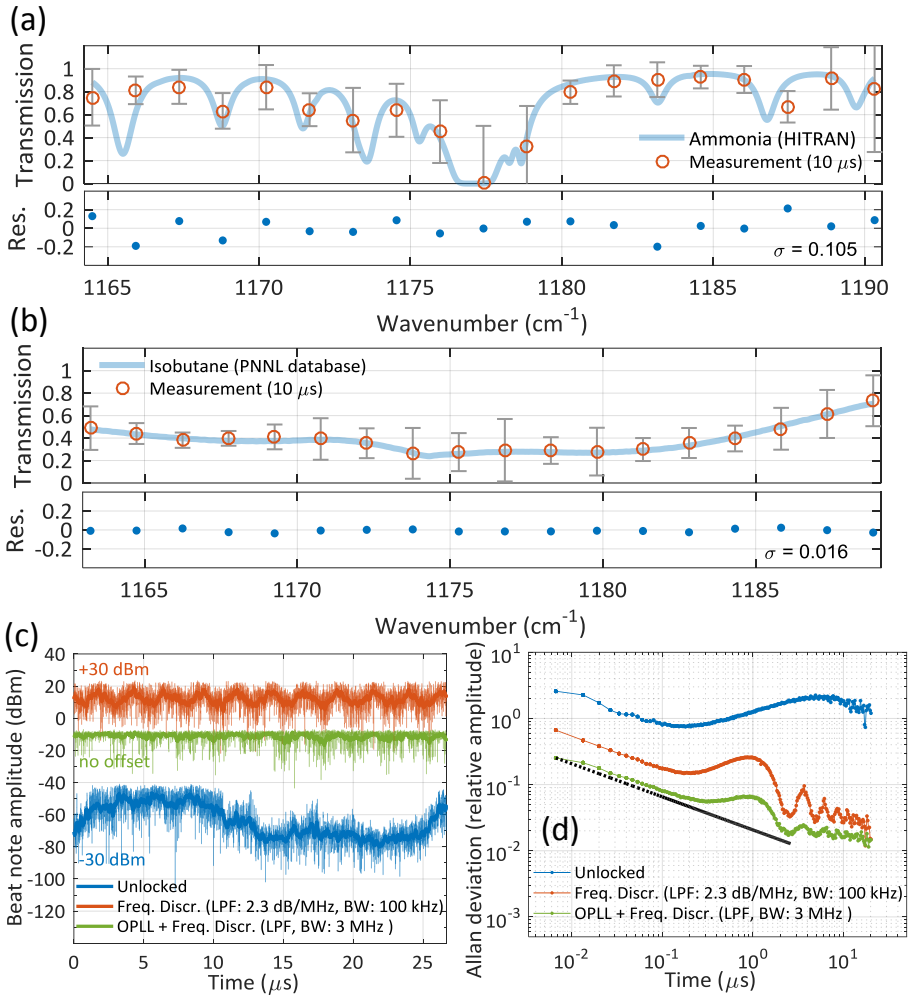

Figure 3.18. Spectroscopic measurements using the OPLL inter-locking system. (a) 50% ammonia at atmospheric pressure in a 10-cm absorption cell. (b) 15% Isobutane at atmospheric pressure. (c) Lowest frequency beat note amplitude versus time. (d) Allan deviation of data in (c). Black curve corresponds to detector noise floor. Results published in [83].

3.4. Multiheterodyne spectroscopy with FP-ICLs in the 3.5 µm region

In the previous sections we discussed the various novel relative stabilization techniques of FP-QCLs for DCS, but historically, the first demonstration of suitability of bulk FP-QCLs for multiheterodyne spectroscopy (MHS) in a dual-comb arrangement was performed by Villares *et al.* [66] in 2014. QCLs are unipolar devices that cascade electrons to generate mid-infrared or terahertz radiation, and their comb operation was somewhat difficult to explain: there was no saturable absorption within the cavity that could cause the generation of optical pulses. Instead, QCLs are modeled as devices with a saturable gain [5], which forces the instantaneous optical frequency to change in a random-like fashion, thus producing a frequency-modulated-like light. Since this behavior is periodic with the roundtrip time, a QCL can possess a comb spectrum even though no pulses are produced. All this is enabled thanks to the

ultrafast, picosecond dynamics of the gain. Note that comb generation takes place in a bulk Fabry-Pérot chip, however with very stringent requirements regarding the intracavity dispersion.

One can ask, whether bipolar devices, i.e. such that use electron-hole recombination can also produce such a comb-like optical spectrum without any saturable absorbers. The interband cascade laser belongs to this family of devices, but its dynamics is much slower: the interband active transition lifetime is greater by several orders of magnitude (~ns). We challenged this question and provided unexpectedly the first experimental demonstration of multiheterodyne spectroscopy with Fabry-Pérot interband cascade lasers, as described in details in [77].

Why is it beneficial to explore the usefulness of ICLs in MHS? An undoubted advantage is their low power consumption and high wall-plug efficiency in the 3-5 micron range [85]. In particular, the spectroscopic region around 3 microns is of interest due to a large number of organic molecules with characteristic C-H bond stretches. This is where the performance of QCLs tends to degrade, and it would be extremely desired to have a semiconductor source of comb-like laser radiation for small portable sensors. For instance, an ICL consumes approximately an order of magnitude less electrical power in continuous wave (cw) operation than a QCL operating at the same wavelength [86].

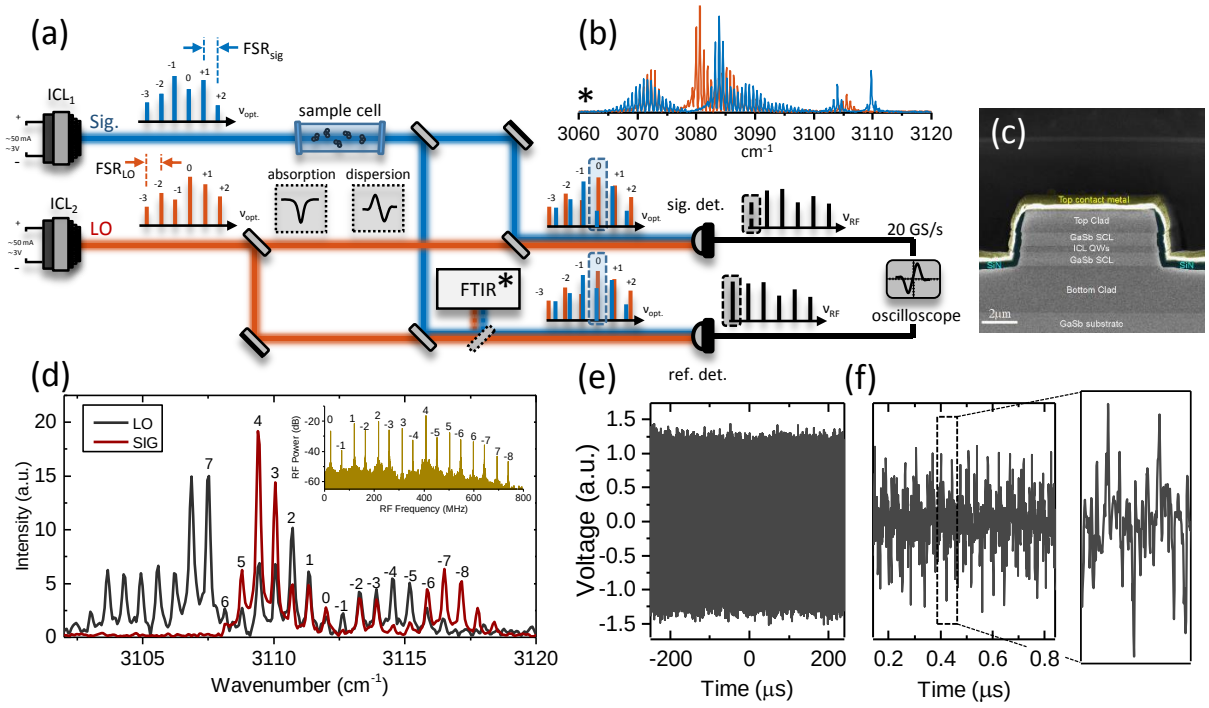

Figure 3.19. Multiheterodyne spectroscopy with FP-ICLs. (a) Experimental setup. (b) Optical spectrum at high currents showing broadband (50 cm^{-1}) operation of the ICLs. (c) SEM picture of the ICL (d) Optical and multiheterodyne spectra of the ICLs running in low-phase noise regime. (e) Time domain interferogram showing that no pulses are observed. (f) Zoom into interferogram in (e). Results published in [77].

The successful detection of methane on Mars using a single-mode ICL as part of NASA's Mars Curiosity Mission in 2012 [87] proved that these sources are well-suited for solar- or battery-powered applications. Multimode ICL comb-like sources to simultaneously probe multiple spectral bands are in this context a natural step of evolution. A graphical summary of our FP-ICL experiment is shown in Figure 3.19.

In the experiment, we employed two closely matched (ΔFSR=96 MHz) 2 mm FP-ICLs provided by Dr. Jerry Meyer from the U.S. Naval Research Laboratory (NRL). The optical modes were centered at approximately 3100 cm^{-1} (~3.2 μm) with free-spectral ranges of 0.640 and 0.643 cm^{-1} (one laser was shorter by ~15 μm than the other). The lasers emitted multimode laser radiation with comb-like properties at as low as 40 mA and 3 V of bias, however for maximizing the optical coverage, we increased it to approximately 60 mA. The temperature of the chips was kept at a constant value using a thermoelectric cooler around 4-6°C. Temperature tuning enabled to shift the laser gain to allow the modes to develop in regions that promote low-phase noise operation, as at higher currents and temperatures the multiheterodyne spectrum turned into an incoherent blob. This aspect of phase-noisy operation of ICLs, which is predominantly caused by the existence of significant dispersion regions, will be discussed in the next chapter of this work.

Panel (a) shows the well-known asymmetric balanced configuration. Its choice was dictated by the fact that the first ICL sources showed less stable behavior compared to the QCLs, and we needed to decouple changes introduced by an absorber from those induced by random intensity fluctuations. For characterization purposes, we added a Fourier transform spectrometer (FTIR) to the experimental setup. Also in contrast to the previous acquisition schemes, we collected spectroscopic data with a fast oscilloscope, which allowed us to analyze data using our novel phase-correction procedure described in detail in this chapter.

Panel (b) plots the optical spectra of the two ICLs in a high current and temperature regime (17.32°C and 11.84°C, and currents of 105.3 and 127.8 mA, respectively). The optical spectra span more than 50 cm^{-1}, but unfortunately such conditions were not suitable for multiheterodyne measurements. We were constrained by the intrinsic low phase noise regimes, which were in general much narrower compared to the previously used QCLs. An example of such conditions is shown in panel (d) together with a phase-corrected RF spectrum, showing a comb of equidistant multiheterodyne lines. In (c) one can see the SEM micrograph of an ICL narrow ridge output facet. At that time, the technological limit of 8 μm for the ridge width did not allow the laser to operate in a single transverse mode. Consequently, we observed noticeable instantaneous changes in mode intensities, possibly attributed to mode competition. This issue was mitigated in the next generation of ICL combs, which allowed for single-detector free running measurements, which we will describe further.

Finally, panels (e) and (f) plot a time-domain interferogram. The waveform resembles that of beating between two QCL combs, i.e. with a nearly constant amplitude where no pulses can be observed, however the origin of the phase-locked operation of ICLs is yet to be theoretically explained.

Initially we tried to phase-lock the ICLs to maximize the signal-to-noise ratio in the RF domain, but it turned out to be difficult due to the broad noise pedestal of the beat note, which was too wide even for the fast OPLL. Seeing that a large number of beat notes took an asymmetric or dual-peak shape, and their signal-to-noise ratio was approximately 20 dB lower than for the previously discussed QCLs, we had to develop a novel analysis technique of spectroscopic data. We took advantage of the high phase coherence between the modes to computationally turn a phase-noisy multiheterodyne spectrum into a radio frequency comb, as plotted in Figure 3.20.

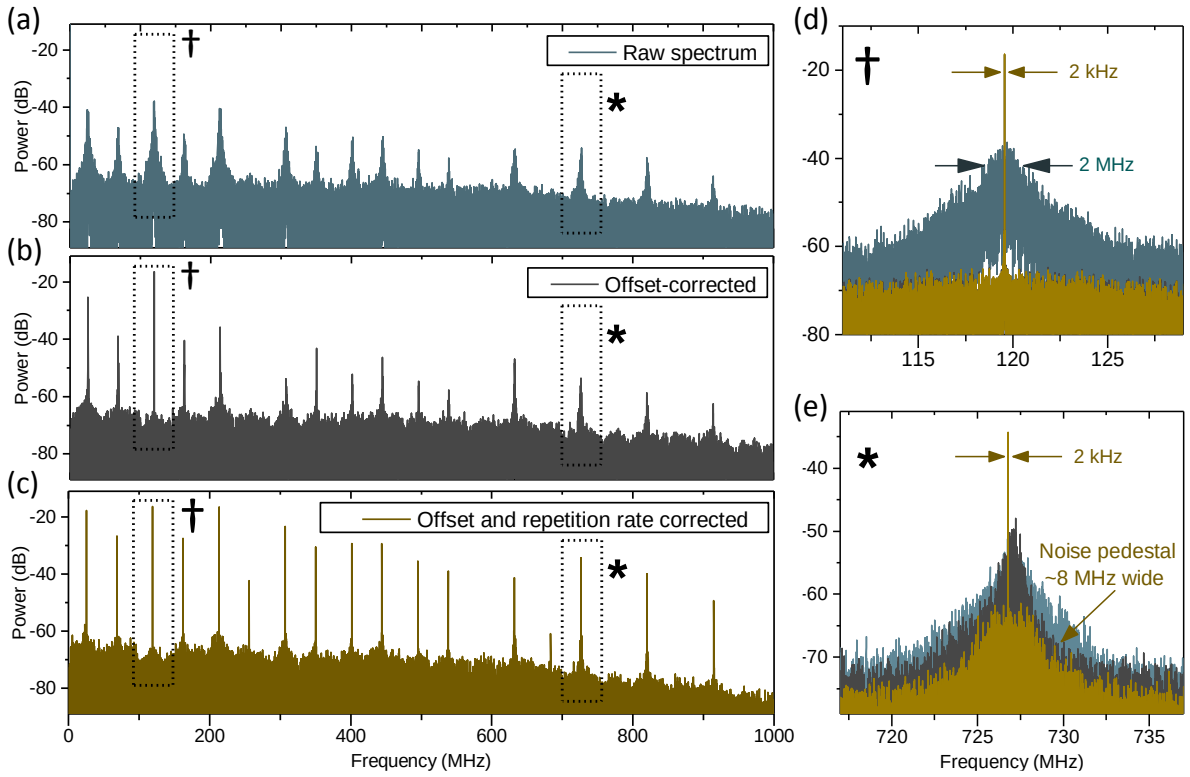

Figure 3.20. Computational phase and repetition rate correction of a phase-noisy multiheterodyne spectrum of two optically beaten FP-ICLs. (a) Raw, uncorrected spectrum. (b) Spectrum corrected just for offset fluctuations. Although the correction is efficient close to the correction beat note pair (d), a broadening of the beat notes at higher frequencies occurs (e) due to the cumulative non-compensated effect of repetition rate fluctuations. (c) Spectrum corrected for both repetition rate and offset fluctuations. Now the beat note is almost of the same shape as at lower frequencies, however a slight noise pedestal can be observed. It was mitigated in further revisions of the algorithm by resampling using higher order harmonics of the repetition rate difference. Results published in [77].

This procedure will be subject to a deeper analysis later, but shortly speaking by filtering two neighboring RF beat notes, we extracted the instantaneous repetition rate difference $\Delta f_{\mathrm{rep}}(t)$, followed by extracting the instantaneous offset frequency difference $\Delta f_0(t)$ from one of them. After resampling the time-domain trace, and phase shifting to counteract the global fluctuations of Δf_0, we reduced the RF linewidth from the MHz to kHz level, limited here by the acquisition time of 500 μs. Note the boost in signal-to-noise ratio by approximately 20 dB. Obviously, over extended time scales this gain will improve even more.

Having developed this computational procedure, we performed numerous spectroscopic experiments, which we will describe now.

3.4.1. Spectroscopy with interband cascade lasers

The ICLs have shown to be compatible with all the previously demonstrated spectroscopic techniques: broadband "snapshot" static acquisitions, further extended here by temperature tuning to fill the gaps caused by the discrete nature of the sampling modes, as well as swept high-resolution absorption and dispersion measurements at multiple frequencies simultaneously. To understand how the multiheterodyne measurement works, we will illustrate the mapping of the optical spectrum to the RF domain on a molecular absorber example, which is shown in Figure 3.21.

The top left panel plots the optical spectrum of the two ICLs measured with a Fourier transform interferometer (FTIR) with a 0.125 cm^{-1} resolution, which precludes any identification of the difference in the FSRs. We can clearly see that the LO comb emits a much broader optical spectrum than the signal, but unfortunately the electrical bandwidth limited our measurements to approximately a dozen of simultaneously-measured beat notes, hence its full potential could not be fully exploited.

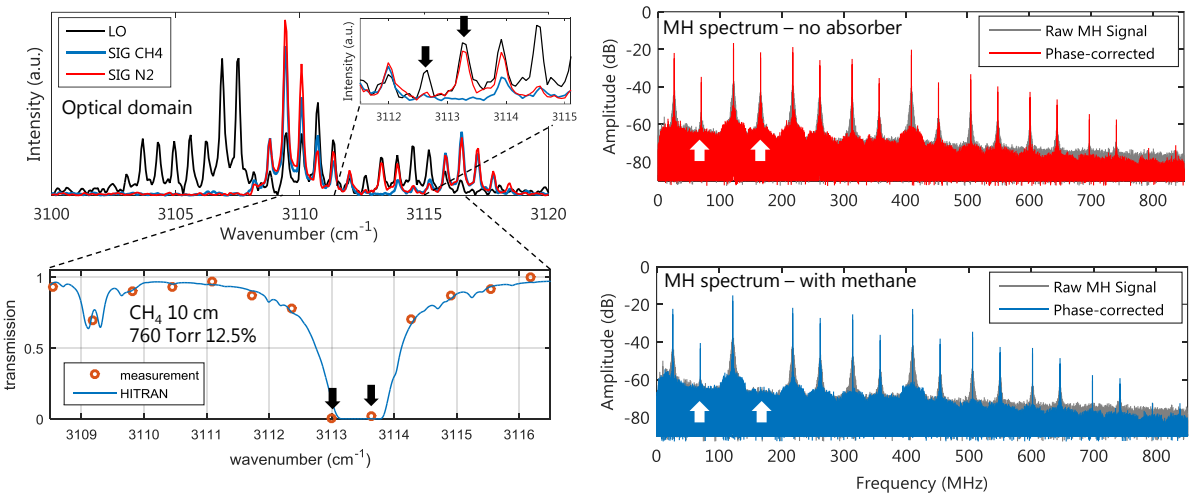

Figure 3.21. Illustration of the direct optical-radio frequency domain link in multiheterodyne spectroscopy. Left panels show the optical domain, whereas the right show the radio frequency domain.

More importantly, the shape of the spectral envelope is not flat as in our previous assumptions, which is the reason of the uneven multiheterodyne beat notes' amplitudes. The red multiheterodyne spectrum (right top) corresponds to a zero-gas measurement taken within 500 μs, which changes significantly in regions marked with white arrows, when 12.5% methane at atmospheric pressure is pumped to a 10-cm absorption cell, as visible in the blue MH spectrum (bottom right). To help understand the link between the optical and RF domains, the optical spectrum of the signal laser was measured before and after introducing the absorber. The absorption of methane around 3113-3114 cm^{-1} makes the optical modes in the FTIR spectrum almost completely disappear, which is also reflected in the blue RF spectrum, in particular at ~150 MHz. Note that every other beat note starting from that at 50 MHz is aliased to double the number of beat notes in the electrical bandwidth of the photodetector (Vigo PVI-4TE-10.6, 1 GHz). In the final step, the two spectra are compared to yield the transmission, whose profile is subject to a HITRAN database fit (left bottom). We can clearly see that the measured transmission plot shows overall good agreement with the fit.

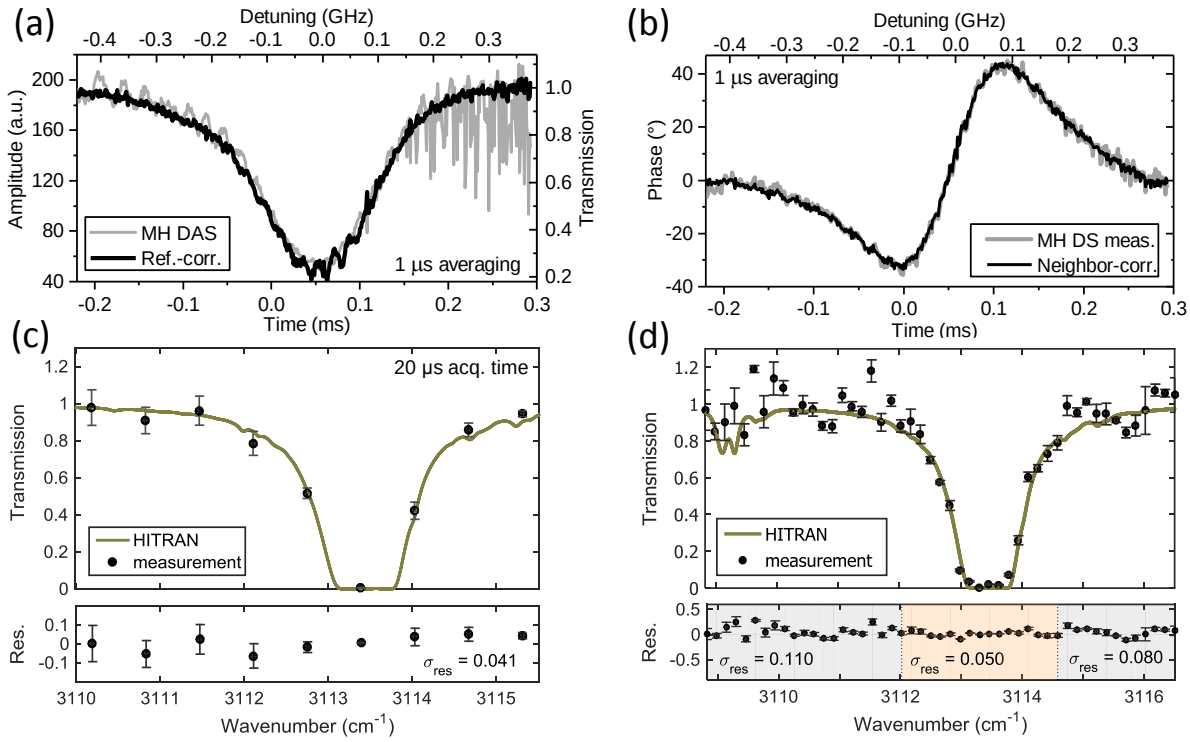

Figure 3.22. Multiheterodyne spectroscopy with FP-ICL. (a) Swept direct absorption and (b) dispersion spectroscopy of 1.6 kPa ethylene at 3112.6 cm^{-1}. (c) Rapid broadband spectroscopy of methane. (d) Interleaved measurement of methane obtained by stacking multiple measurements taken at different temperatures of the lasers. Results published in [77].

Without any changes in the optical setup, numerous spectroscopic techniques can be implemented and seamlessly switched between one another (Figure 3.22). First is swept direct absorption (a) and dispersion spectroscopy (b), here demonstrated on 1.6 kPa ethylene at 3112.6 cm^{-1}. Next is rapid

broadband spectroscopy of 99 kPa pure methane within 20 μs (c). Even at such low time scales, the residual is just ~4%. Finally, by changing the temperature of the lasers, we can tune their position to interleave the discretely sampled spectrum of ~19.2 GHz, thereby filling the gaps as shown in (d).

3.5. Multiheterodyne spectroscopy with QCL combs in the terahertz region

In the previous chapter we discussed numerous applications of terahertz spectroscopy in particular to solid state pharmaceuticals. One can ask, whether the advantages that made DCS so popular, such as ultrafast microsecond acquisition time, high resolution, and high power per spectral element can be extrapolated to the terahertz range. For instance, real-time drug production monitoring puts sharp requirements on high optical power and speed, and it would be challenging to probe hundreds of samples per second with a mechanical nature of the spectroscopic scan as in THz-TDS, whereas the moving parts-free DCS solution does not have such restrictions.

The answer to the extrapolation of DCS to the THz is generally positive. One way to perform DCS in the far-infrared is to employ two femtosecond lasers with different repetition rates illuminating two photoconductive antennas (PCAs), known as asynchronous optical sampling (ASOPS) THz-TDS [79]. Obvious advantages of photoconductive switches are broadband, room temperature operation, and MHz sampling resolution defined by the near-infrared lasers, however at higher frequencies (> 1 THz) even the most recent generations of PCAs tend to have power levels in the microwatt and lower range [11]. This is the gap filled by terahertz QCLs emitting milliwatts of optical power between 1 and 5 THz. As direct emitters of coherent terahertz light, they are fully tunable in offset, and slightly in repetition rate just like mid-infrared QCLs, which in principle allows to perform interleaved (gapless) measurements. While they possess repetition rates on the order of GHz, they are well suited for solid samples and atmospheric pressure absorbers.

These advantages were the driving force for research on terahertz QCL frequency combs. The first THz QCL comb was demonstrated by Burghoff *et al*. from MIT [88] under the supervision of prof. Qing Hu. The researchers had to overcome the exceedingly high group velocity dispersion (GVD) of GaAs at ~3.5 THz, which is greater than that in the mid-infrared (~7 μm) even 250 times. The GVD essentially causes light at different wavelengths to travel with different group velocities, and if its value is too high, it effectively precludes comb formation. The MIT solution was to etch a corrugation into the QCL facet to act as a double-chirped mirror: it reflects shorter wavelengths earlier, whereas longer travel to the end. As a result, the high positive dispersion of the material could be compensated. We had the privilege to collaborate with prof. Qing Hu's group to evaluate the spectroscopic performance of these THz QCL combs. Figure 3.23 shows a microscope picture of these devices.

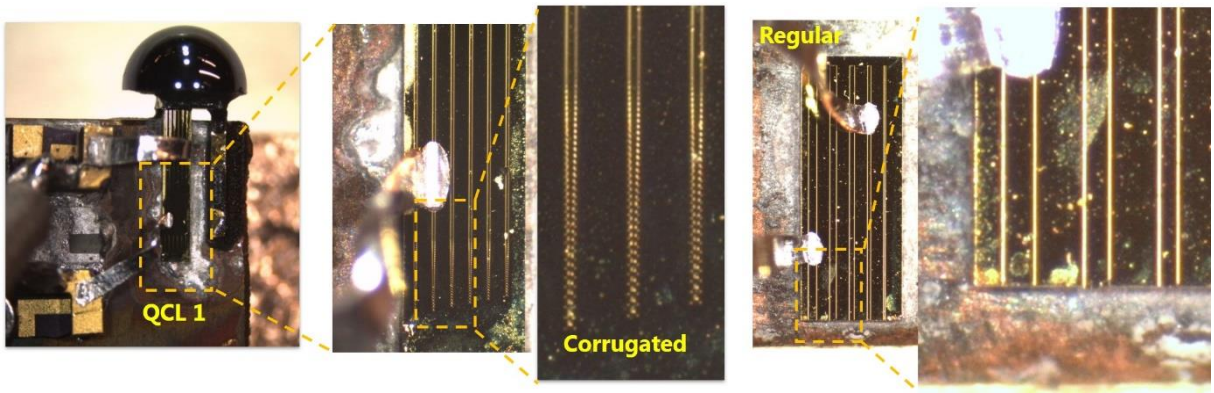

Figure 3.23. THz QCL frequency combs. Left: multi-chip device with a silicon lens glued to the facet. A zoom into the back faces reveals the corrugated nature of the waveguide. Right: regular device.

Due to the less mature QCL technology in the terahertz (the first THz QCL was demonstrated in 2001 [89]), operation of these sources is currently restricted to cryogenic temperatures. Obviously, this makes the DCS setup considerably more complex. To operate the lasers at ~30 K, we housed the QCLs in a vacuum-isolated pulsed tube cryo-cooler capable of maintainig 40 K at 40 watts of heat load, as shown in Figure 3.24c. To collect THz light from the facets, each laser was equipped with a High Resistivity Float Zone Silicon (HRFZ-Si) hyper-hemispherical lens, as visible in Figure 3.24a. Otherwise, the distance from the TPX windows to the facets would not allow for efficient collection of the THz light. The lasers were attached to a copper submount, while their electrodes were soldered to microwave pigtails terminated with snap-on microwave MCX connectors, which seemed to withstand harsh cryogenic conditions and thermal cycling. This allowed us to simultaneously DC-bias the lasers, and extract their ~9 GHz repetition rate beat notes generated intrinsically through a microwave bias tee. Strong microwave beat notes appeared only when they operated as combs, hence served as an important diagnostic tool. What deserves attention here is the choice of the interconnecting microwave coaxial cables in the cryostat that guides these signals. One cannot use a regular copper-based coax, since it would provide a severe thermal link between room temperature and the cold finger. Instead, a dedicated cryogenic coax rated for DC-61 GHz with a center conductor made of BeCu, and an outer conductor made of low thermal conductivity stainless steel (SS) (UT-085B-SS from Micro-Coax), as visible in Figure 3.24b. We observed that the SS-shielded coax did not affect the cooling capacity of our system.

The last element of the QCL-based MHS system is a photodetector. In the terahertz, one can typically use a room-temperature Schottky mixer, or an ultra-sensitive liquid-helium (LHe) cooled superconducting hot electron bolometer (HEB). To obtain the best perfomance possible, the latter was used here, albeit it required to have the LHe bath refilled on a daily basis. This device is significantly larger than a typical mid-infrared photodetector but allows to detect modulated light with a remarkably low noise-equivalent power (NEP) of $3\text{-}5\times10^{-13}$ W/Hz$^{1/2}$ even at 7 GHz.

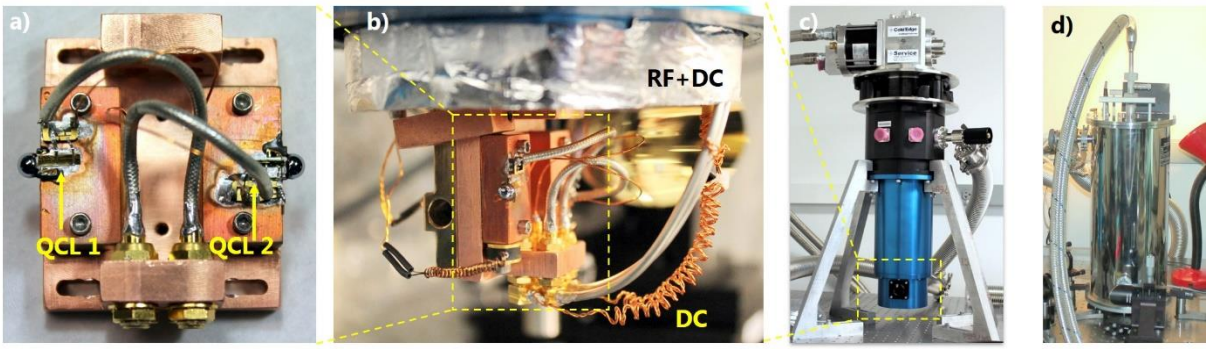

Figure 3.24. THz QCLs equipped with HRFZ-Si lenses mounted on a submount with their electrodes soldered to microwave panel-mount coax connectors (a) and their placement on the cold finger (b). The SS coax is connected from the right together with auxiliary sensors. (c) PTC Cooler with a cap attached that covers the cold finger in (b). (d) LHe superconducting HEB during helium transfer.

The HEB itself is an interesting device. It uses electrical nonlinearities caused by superconductivity of niobium nitrate (NbN) to multiheterodyne THz light emitted by the two THz combs. From a radio frequency engineering perspective, it can be pictured as a frequency mixer with a broad intermediate frequency (IF) range from kHz to approximately 7 GHz. In contrast to the conventional HgCdTe mid-infrared photodetector, the HEB requires a careful user choice of the operating point, therefore it may be beneficial to discuss its characteristics in this respect, as provided in the Appendix. Having understood what risks are associated with the improper operation of the mixer, we can discuss the spectroscopic results obtained using our THz DCS system.

3.5.1. Spectroscopic measurements

The heart of our dual comb system (Figure 3.25a) was the previously introduced pulsed tube cryocooler housing two lensed QCLs emitting sub-milliwatt output powers of ~250 GHz-broad terahertz radiation centered at 2.8 GHz, as in Ref. [90] . The lasers were biased using microwave bias tees, which allowed to extract their instantaneous repetition rate (intermode) beat notes, and identify comb operation visible as kHz-wide lines with high SNR at microwave frequencies. The nominal repetition rates were roughly 9.1 GHz (Figure 3.25a), and were separated by a tunable difference Δf_{rep} ranging from 13 MHz to 36 MHz (Figure 3.25c,d). The emitted THz light escaped through two TPX windows in the cryostat, and was collimated using two external off-axis parabolic mirrors. The setup was built around an asymmetric DCS configuration, where only one beam probed the absorber. The two beams were attenuated with apertures to adjust the amount of HEB pumping, and further combined with a HRFZ-Si beam splitter. Finally, they were coupled to the HEB with a long focal length OAPM. To minimize the plague of all THz systems – water vapor interference – the system was enclosed in a purging box with a continuous supply of dry nitrogen. An example 512 MHz-wide multiheterodyne beat note spectrum centered at the difference in offset frequencies between the two QCLs (Δf_0=1.85 GHz) was acquired with a 43 GHz

real time spectrum analyzer R&S FSW43 within 1 ms, and post-processed by our computational phase correction algorithm, as shown in Figure 3.25e.

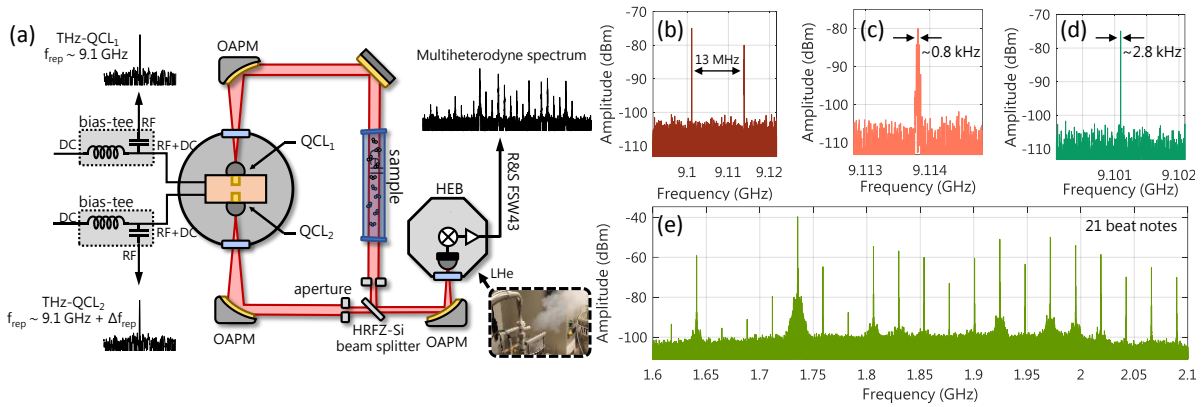

Figure 3.25. (a) Experimental setup of the THz QCL DCS experiment. (b) Two simultaneously measured intermode beat notes from the QCLs. (c), (d) Zoom into the intermodes showing their kHz linewidths. (e) Phase-corrected multiheterodyne beat note spectrum acquired within 1 ms. Results published in [91].

What is not shown here, but will be discussed in the coherent averaging section of this document, is the uninteresting shape of the raw multiheterodyne spectrum. The cryocooler created a noisy and vibrating environment for the QCLs, which heavily destabilized their operation. As a result, the beat notes completely overlapped in the RF spectrum and no comb lines were identifiable there. To deal with such extreme cases we had to develop a completely new procedure for coherent averaging, which was further used in all our dual comb experiments. Currently, it can correct an overlapped beat note spectrum consisting of more than a hundred lines in almost real time, yet with a real-time implementation potential.

To demonstrate the spectroscopic usefulness of our system, we measured pure ammonia with a broadband absorption spectrum in a 10-cm long absorption cell equipped with TPX windows. The THz transmission spectrum at 306 Torr is plotted in Figure 3.26a together with a HITRAN fit, which is again in good agreement with the measurement. For illustrative purposes, Figure 3.26 (b) and (c) include the reference and sample RF beat note spectrum, respectively, used to generate the transmission spectrum in (a). Beat notes on the right are visibly attenuated. Of course, the slope-like shape of ammonia is not very rich, and it would be beneficial to test the system for absorbers more abundant in spectroscopic features. To do that, we measured poisonous nitrogen dioxide (NO_2) at 1.2 bar [92], as plotted in the bottom panel of Figure 3.26. For both measurements, we estimate the NEA to be below $1\times10^{-3}/\sqrt{\text{Hz}}$, mainly limited by the mechanical vibrations and optical feedback effects.

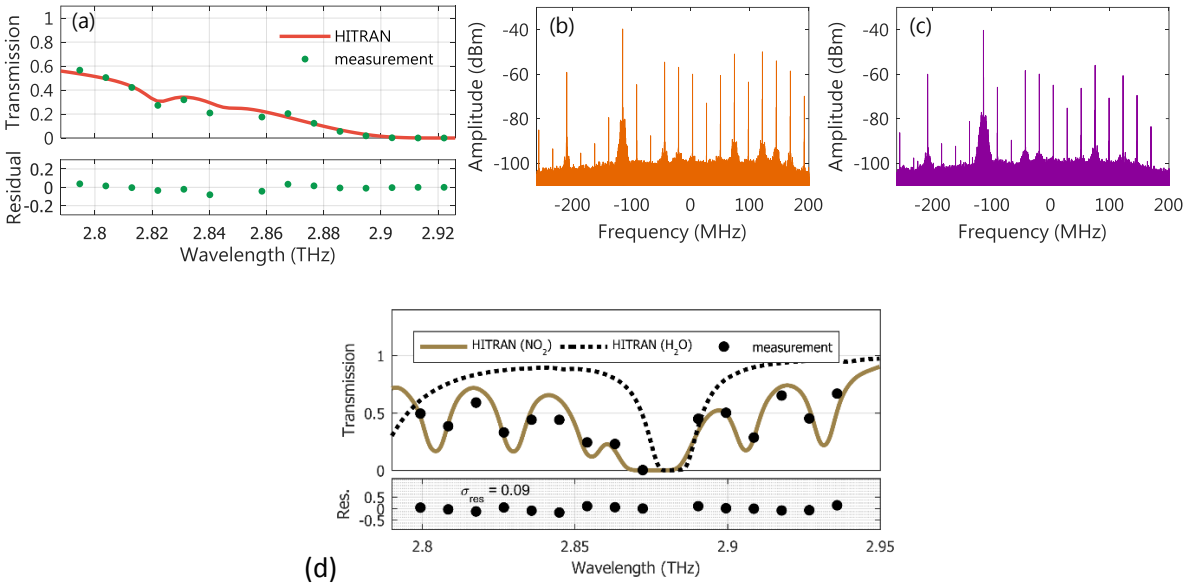

Figure 3.26. (a) THz transmission spectrum of ammonia at 306 Torr together with a HITRAN fit. (b) Reference beat note spectrum. (c) Sample beat note spectrum. Note the attenuation of the beat notes on the right. (d) Nitrogen dioxide at 1.2 bar. Results of (a)-(c) published in [91].

The high signal to noise ratios attainable with QCL-based DCS systems over micro- to millisecond time scales is very encouraging in terms of application of this technology in pharmaceutical industry. While the main limitation here is the narrow bandwidth (10% of THz-TDS systems), first research-grade THz QCLs can develop even a 600 GHz-wide comb spectrum [93]. What is also promising is that THz QCLs can even span an octave from 1.6 to 3.2 THz [94]. Unfortunately, dispersion management to obtain comb operation over such a wide frequency range will require more research in the coming years.

3.6. Dual comb spectroscopy with FP-QCLs at 7.8 μm

A careful reader probably noticed that our first spectroscopic experiments belonged to the family of multiheterodyne measurements, whereas starting with the terahertz system, we introduced the term "dual-comb" in the project name. What makes a multiheterodyne experiment a dual comb experiment?

The answer lies in the ability to demonstrate a strong narrow intermode beat note originating from electrical nonlinearities of the QCL, causing the optical modes to mix with one another. In a highly coherent QCL comb source all possible mixing products will form a clean microwave signal, extractable through a bias tee. Obviously, to obtain comb operation in a QCL, it is imperative to carefully design the lasing medium for the lowest dispersion possible, known in the literature as dispersion engineering [95]. In the long-wave infrared (LWIR), such carefully dispersion-engineered lasers were provided to us by one of the inventors of the QCL – prof. Jerome Faist – and his research group of ETH Zurich. His devices emitted broadband comb spectrum consisting of more than a hundred comb lines over extended current ranges with hundreds of milliwatts of optical power, however to identify which

regions were suitable for dual-comb spectroscopy, we had to develop a custom way to extract the intermode beat note, which is presented in the Appendix.

The home-made microwave probe served well for its purpose. The SNR of the beat note at ~9.8 GHz in general exceeded 30 dB with kHz linewidths, which was a signature of comb operation of the lasers. The optical spectra of lasers in the comb regime are shown in Figure 3.27 together with the experimental setup incorporating an FTIR to measure them. We can see the characteristic bimodal shape of the QCL spectrum with a visible gap at around ~1280 cm^{-1}. Whereas for the green spectrum this gap was not severe, in the orange spectrum case almost no modes developed over 10 cm^{-1}, which resulted in a degradation of the spectroscopic performance therein.

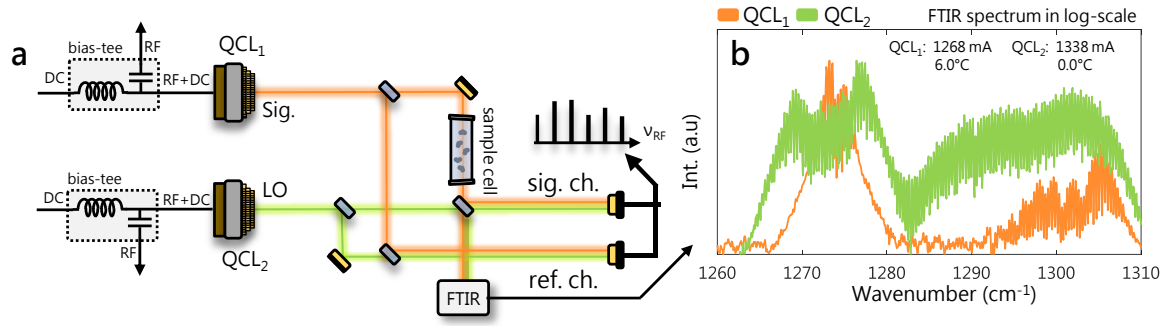

Figure 3.27. Dual comb spectroscopy with mid-IR QCLs. (a) Experimental setup. (b) Optical spectra of the combs spanning 50 cm^{-1} with a noticeable spectral gap around 1280 cm^{-1}. Results published in [96].

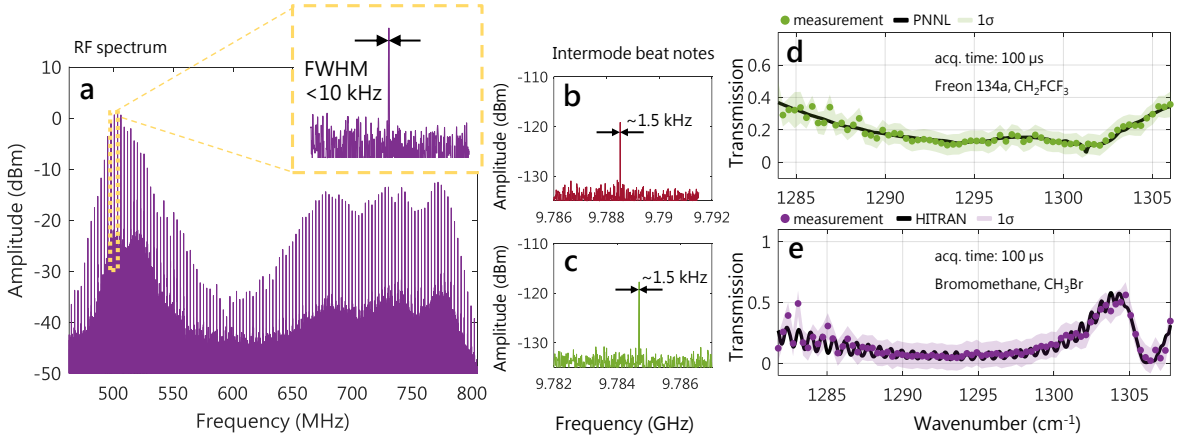

Figure 3.28. Radio frequency spectrum mapping 133 optical modes obtained with the help of computational phase and time correction. The linewidth of an individual beat note is limited by the acquisition time of 100 μs. (b) and (c) show the optically detected intermode beat notes with ~kHz linewidths and ~3 MHz of separation. (d) DCS spectrum of Freon 134a (CH_2FCF_3) measured with an acquisition time of 100 μs. (e) DCS spectrum of bromomethane (CH_3Br) at 160 Torr of pressure, measured with an acquisition time of 100 μs. Results published in [96].

What should be noted is that in this experiment we took full advantage of our novel signal processing framework. We have successfully mapped 133 optical modes to the RF domain in just 400 MHz of electrical bandwidth, which in our first mid-infrared MH systems was occupied by just a couple modes. This improvement has a lot of practical implications. Most importantly, it allows to use slower, hence cheaper photodetectors. Without using the procedure, only the envelope of the spectrum was visible due to a mutual overlap of the beat notes and no reliable DCS measurements were possible. Also the time of such a broadband acquisition was reduced to 100 μs, corresponding to a theoretical beat note linewidth limit of 10 kHz. Thanks to the equally-efficient correction procedure, it was achieved for all MH lines in the corrected spectrum, as plotted in Figure 3.28a together with the comb-indicating intermode beat notes in Figure 3.28b and c.

To demonstrate spectroscopy with these powerful sources, we measured broadband Freon 134a (CH_2FCF_3), and bromomethane with an interesting oscillatory shape of the optical spectrum. We obtained very good agreement with a PNNL (Pacific Northwest National Laboratory) database.

3.7. Dual comb spectroscopy with free running mode-locked ICLs with a saturable absorber

All the previously demonstrated sources for MHS relied on the generation of a comb-like spectrum in an FP chip with a nearly constant optical power, *i.e.* without optical pulses encountered in conventional mode-locked lasers. In 2016 a novel kind of semiconductor laser was developed as part of a collaborative effort between NRL and NASA Jet Propulsion Laboratory [97]. The researchers modified the structure of an interband cascade laser by monolithically-integrating a saturable absorber, hence creating world's first chip-scale electrically-pumped source of mode-locked mid-infrared light. We had access to these novel sources with a large potential for rapid broadband hydrocarbon detection, which we evaluated on a methane example. This demonstration was the most difficult of all in terms of signal processing: we showed that is possible to perform spectroscopy with free-running lasers, meaning that no frequency lock was used, and the position of the beat notes in the RF domain could freely drift between the sample and reference measurement. Although challenging, it proved that DCS systems in the mid-infrared can be considerably simplified, and any imperfections of the sources can be compensated through computational procedures. We will start by describing the mode locked sources from a fabrication point of view to understand how they are different from the previously used FP devices.

3.7.1. Mode-locked ICL sources

A photo of a wafer containing five cleaved but non-separated 4 mm long mode-locked ICL together with an SEM micrograph of the waveguide facet is shown in Figure 3.29. There are two major design changes compared to the previously shown bulk FP-ICLs. First is the use of a split-contact geometry. The three center chips have a biasable saturable absorber section (SA) of different lengths (for empirical determination of the optimal design), electrically isolated from the gain section. This allowed to control

the conditions of mode locking and extract the instantaneous repetition rate of the optical pulses, because the SA acted as a fast photodetector. Second change in the design is the reduction in the waveguide width down to ~3.5 μm, which promoted single transverse mode operation of the ICLs. As a result, we obtained a nearly Gaussian far-field beam pattern without any signatures of higher order transverse modes, which also translated into a more stable optical modes structure.

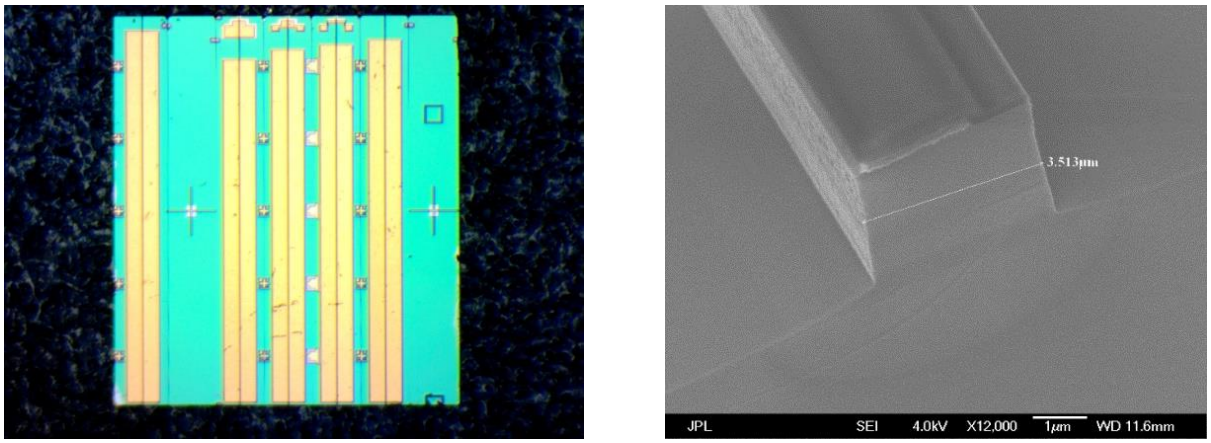

Figure 3.29. Left: microscope picture of an ICL comb wafer with five cleaved devices of different SA length. Right: SEM micrograph of the waveguide facet indicating its width of 3.5 μm. Note the different intensities of gray, which reveal the layered structure of the ICL. The SEM picture was taken thanks to the courtesy of Dr. Clifford Frez, and Dr. Mahmood Bagheri of NASA Jet Propulsion Laboratory.

3.7.2. Experimental setup

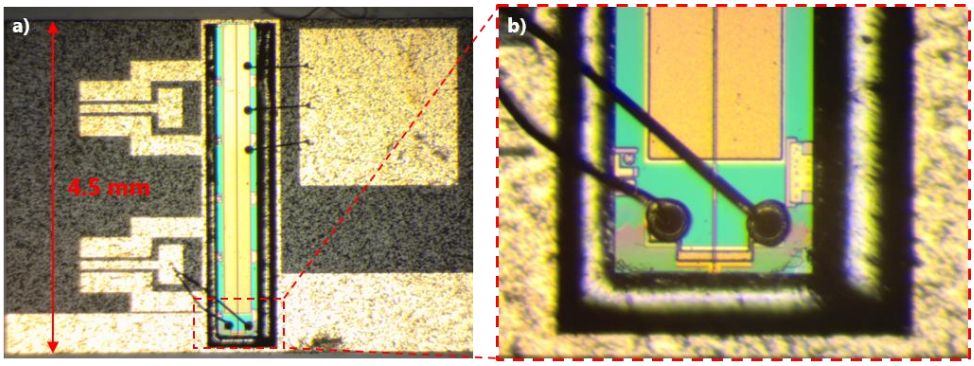

Figure 3.30. (a) ICL chip with a saturable absorber mounted on a beryllium oxide submount with gold electrodes deposited on its surface. The pads on the left are compatible with a GSG microwave probe to enable the extraction of the intermode beat note signal indicating comb operation. The length of the submount is 4.5 mm. (b) Zoom of the saturable absorber section with the 3.5 μm ridge waveguide visible in the center. This device was mounted by the author at JPL facilities in Pasadena, CA.

By the same token as in the LWIR QCL experiment, we developed a custom microwave probe station to characterize the comb operation of the mode-locked ICLs. The cleaved lasers were first soldered to a gold-plated beryllium oxide (BeO) submount utilizing a layout of the top conducting surface compatible

with a marketed 150 μm pitch 40 GHz ground-signal-ground (GSG) microwave probe (Cascade Microtech), as visible in a microscope photography in Figure 3.30.

Next, to simultaneously bias and monitor comb properties of the ICLs, we included a pair of such microwave probes in our DCS system, and mounted them on micromanipulator stages fixed to two separate thermoelectric cooling blocks for individual control of the sources' temperatures. The light was collected from the facets using a mid-infrared anti-reflection coated lens on an XYZ stage for precise collimation. A picture of our compact dual-comb ICL system is shown in Figure 3.31.

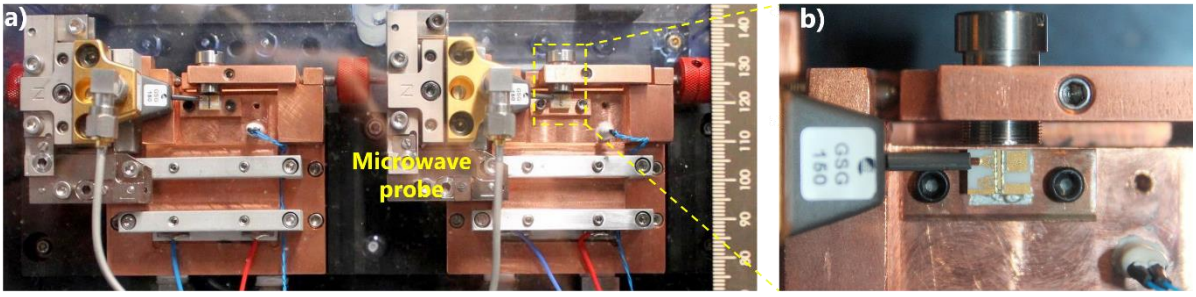

Figure 3.31. Microwave probe station for two simultaneously run ICL combs equipped with two TEC coolers. (a) Overview photography. (b) Zoom into one of the lasers with the BeO submount visible.

3.7.3. Comb operation of mode-locked ICLs

When biased with currents exceeding threshold 3-4 times, the ICLs developed broad optical spectra with kHz-wide intermode beat notes centered at ~9.7 GHz (Figure 3.32). Their relative strength of 30 dB is a clear signature of high mode-to-mode coherence, and it is somewhat unexpected due to the low amount of electrical power fed to the cavity of less than a watt. The optical spectra span approximately 33 cm^{-1} (1 THz), but if one includes the weakest mode structures slightly the noise floor, in can be as large as 46 cm^{-1}. This finds confirmation in the multiheterodyne spectrum in Figure 3.32d, where the dominant group of modes in the optical spectrum appears together with numerous low-intensity lines above 1 GHz, all coherent with each other, as detected using the phase correction algorithm.

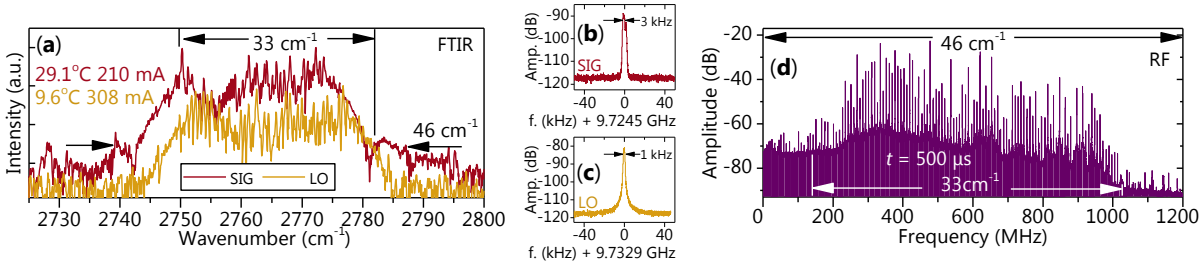

Figure 3.32. Optical spectra of the two closely matched ICL combs (vertically offset by 10 dB for clarity) and their respective kHz-wide intermode beat notes in (b) and (c) (RBW=100 Hz, VBW=1 Hz). (d) Coherently-averaged MH spectrum containing beat notes spanning 46 cm^{-1} of optical bandwidth, of which 33 cm^{-1} is suitable for spectroscopy on the 500 μs timescales. Results published in [98].

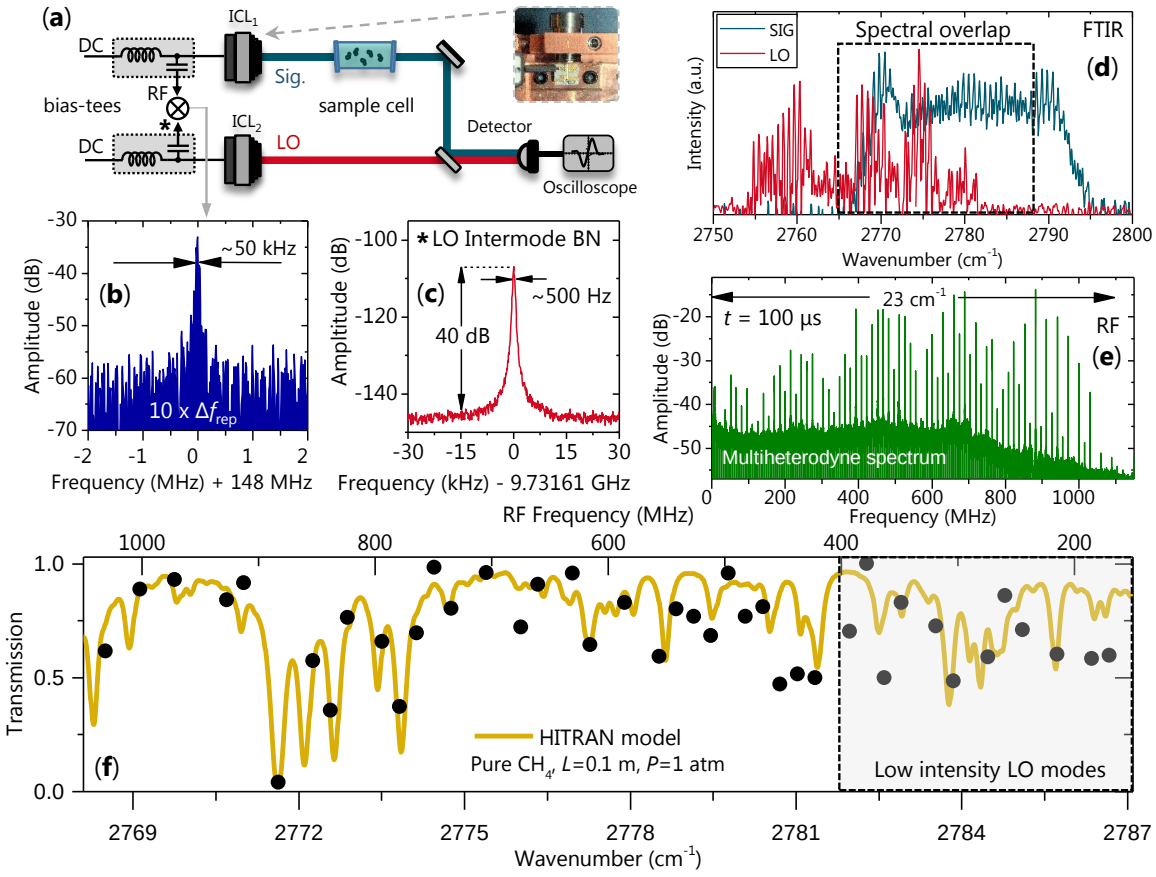

Figure 3.33. Simplified experimental setup for the DCS using mode-locked ICLs. The intermode beat notes are extracted using microwave probes (photo) and bias-tees. (b) 10th harmonic of the repetition rate difference, which shows that the fluctuations of the repetition rate between the combs (FWHM) are ~5 kHz over 100 µs. (c) Intermode beat note of the local oscillator comb with FWHM ~500 Hz. (d) Log-scale optical spectra of the two ICL combs. (e) Coherently averaged reverse-mapped MH spectrum acquired within 100 µs. (f) Spectroscopic measurement of pure methane at atmospheric pressure in a 10-cm absorption cell using beat notes with SNR>10 dB. Results published in [98].

3.7.4. Broadband and high-resolution spectroscopy of methane

To test the feasibility of performing spectroscopy with free-running ICL combs using a single-detector (asymmetric) configuration, we measured methane at atmospheric pressure in a 10-cm absorption cell, as presented in Figure 3.33. Unfortunately, the spectroscopic potential of the ICLs pair shown before is awaiting to be revealed, thus we will show results obtained with a pair with comparably good comb properties, yet with spectral match not as close as before. Panels (a) – (d) show the experimental setup together with numerous diagnostic signals to characterize the sources and their spectral properties. A coherently averaged multiheterodyne spectrum (e) down-converted ~23 cm^{-1} of optical into 1 GHz electrical bandwidth, which was predominantly limited by the difference in repetition rates of ~14.8 MHz. The reverse-mapped RF spectrum includes beat notes with significantly varying intensities, even between neighbors, exceeding in extreme cases 30 dB caused by the uneven intensities of the

optical modes, as visible in the FTIR spectra. From a spectroscopic perspective, it translates into a larger variance in transmission in regions, where LO modes are barely distinguishable from the noise floor [see panel (f)], however the overall agreement with a HITRAN fit for such a spectrally rich absorber is encouraging. Especially the region between 2768 cm^{-1} and 2878 cm^{-1} shows a large promise, and proves the necessity of equalizing the mode structure and ensuring good spectral overlap between the combs, which was not fulfilled at higher frequencies. Beat notes resulting from mixing between strong signal comb, and weak LO or vice versa simply require longer averaging times to obtain the same uncertainty.

This experiment concludes the validation of our computational procedure for using phase-noisy combs in molecular dual-comb spectroscopy. We proved that scenarios generally incompatible with DCS can be resolved thanks to signal processing procedures like adaptive resampling and frequency tracking. We demonstrated that in principle, there is no need for additional optical and electronic components in the system, proposing instead to free-run the lasers and burden the signal acquisition side to correct for any drifts. Consequently, a DCS spectrometer needs just two comb sources and a photodetector, however even the latter can be mitigated due to the recent demonstration of self-detection capabilities of THz QCLs [99].

3.8. Coherent averaging of multiheterodyne spectra

As it was presented in the previous sections, progress in modern quantum electronics enabled the development of truly miniaturized high power multimode sources of coherent mid-infrared and terahertz laser radiation with wide spectral coverage, narrow linewidths below a MHz, and comb-like properties. Some of the sources, such as ICLs offer additionally high suitability for battery operation for use in portable spectrometers in the mid-IR. To take full advantage of their broadband operation, the dual-comb technique enables to turn two lasers into a real-time multi-parallel optical spectrum analyzer with an unprecedented scale of miniaturization, and no moving parts.

Unfortunately, the price for this convenience is high: optical modes must be relatively stable in amplitude, and frequency, as well as show stable comb or comb-like operation over a wide range of temperatures and currents. Additionally, not only is it necessary to find one perfect device. In fact, one needs to find the perfect matched pair. Such requirements are stringent and can effectively exclude a majority of devices in a batch of fabrication from use in DCS. While it is imperative to satisfy the close spectral and free spectral range match criterion, we know that frequency-and amplitude- unstable devices should not be disqualified from using in DCS. Amplitude fluctuations can be averaged, if we ensure constant RF frequencies of the MH beat notes.

Previously we showed that one way to tackle the sources instability problem is to utilize the frequency- and phase locked loop stabilization, which additionally allow for ramped (swept) acquisitions bypassing the limitations of discrete sampling of the optical spectrum. Unfortunately, that solution to deal with frequency instabilities has some disadvantages. The servo loop has a finite bandwidth and allows for

locking only the offset frequencies difference. For the most demanding acquisitions, where the number of modes exceeds a hundred, that scheme was not sufficient, since the repetition rate fluctuations caused a mode number dependent frequency shift. As a result, beat notes lying far away from the locking point mutually overlapped. To overcome this issue, we have developed an alternative computational solution, equivalent to a digital phase locked loop (DPLL) [100], which was intensively employed in most of our DCS experiments to coherently average noisy MH signals. It enables to use noisy (but still coherent) lasers in the dual-comb arrangement in extreme conditions, *i.e.* when the down-converted RF beat notes overlap and the spectrum looks like a featureless blob. In contrast to the analog PLL and FLL that account only for offset frequency instabilities, the DPLL corrects for both offset and repetition rate jitter. With the computational technique, it is possible to perform DCS event with free running lasers, as proven in the ICL comb example.

To further motivate the necessity of using coherent averaging techniques, one should mention that a freshly fabricated laser requires some time to stabilize its optical mode structure, which in practical considerations takes at least a hundred hours per device. The semiconductor structure burns in, inducing changes in its structure such as surface passivation [101], which inevitably affect the characteristics of the device. In some cases, a laser that showed decent spectral and comb properties, suddenly changes its operation into more chaotic. Even after the physico-chemical properties of the semiconductor structure stabilize, dynamic effects such as optical feedback in the spectroscopic system can cause unstable operation too. The latter was very severe in the terahertz QCL setup, where vibrations from the cryocooler caused extreme instabilities in the MH spectrum. Consequently, it did not have any identifiable discrete components, which was successfully corrected with the help of our algorithm, as illustrated in Figure 3.34.

So far, we focused our attention on phase noise, but amplitude noise is the actual reason for coherent averaging. Intensity of modes in semiconductor lasers fluctuates together with multiheterodyne beat notes, which can be suppressed by performing measurements relying on their mean value evaluated over extended time scales. As long as the character of fluctuations is white, longer averaging will reduce the uncertainty of the mean estimate, thus yielding shorter error bars. Phase correction ensures, that averaging will take place within narrow frequency bands, thereby making it more efficient compared to signals with power distributed over a wide frequency range.

The procedures described here are universal to any dual-comb spectroscopy system, and were successfully applied to improve spectroscopic results in systems utilizing quantum cascade lasers (QCLs) operating in the mid-infrared and terahertz range, as well as interband cascade lasers (ICLs) in the mid-infrared. We will start their description by analyzing effects of acquisition time on the shape of multiheterodyne spectrum.

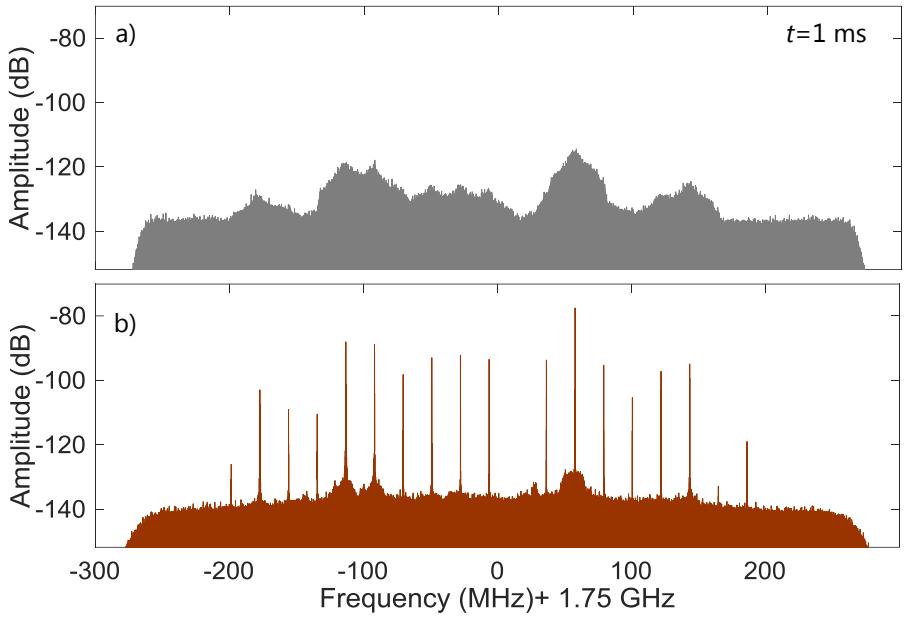

Figure 3.34. Demonstration of the coherent averaging technique in application to the terahertz QCLs centered at ~2.8 THz operated at cryogenic temperatures (~35 K) in a pulsed tube cryocooler (PTC), which created severe mechanical vibrations with a non-negligible effect on the multiheterodyne spectrum (a). Thanks to the algorithm, the discrete nature of the MH lines can be recovered (b). The gaps in the corrected spectrum correspond to modes absorbed by water vapor. Results published in [102].

3.8.1. Effect of acquisition time

Multiheterodyne beating in DCS produces a time-domain electrical signal on a photodetector. Conventionally, we would like to analyze it in the spectral domain to retrieve information about the optical spectrum mapped at different radio frequencies. A straightforward way is to employ the Fourier Transform to find the frequency representation of the signal, but there is one risk associated with this operation if the signal is non-stationary.

Changes in frequency of a narrowband signal throughout the acquisition time cause the Fourier spectrum to contain power at all swept frequencies. In a monocomponent, frequency-noisy sine wave case, the amplitude of different frequency bins will be related to the amount of time the signal "spends" at a given frequency. Next, if we have a comb-like structure that sweeps the RF spectrum, as previously shown in Figure 3.34, the spectrum will look even worse since the discrete MH lines mutually overlap, making the resulting spectrum look featureless. Clearly, long-term Fourier Transform is not suitable for frequency-unstable signals. A natural solution would be to limit the observation time, and assume local stationarity of the signal before the Fourier spectrum is computed, which is known as a short time Fourier transform (STFT)

$$X(\tau,t)=\int_{-\infty}^{\infty}x(t)w_f(t-\tau)\mathrm{e}^{-\mathrm{j}\omega t}\mathrm{d}t. \qquad (3.15)$$

The STFT calculates the Fourier transform of the input signal multiplied by a window function w_f which passes it only for a short period of time. This tool enables to deal with non-stationary (locally stationary) signals, however at the expense of a lowered frequency resolution limited by the length of the window. This is highly undesired in DCS because at short timescales the weakest beat notes will not be distinguishable from the noise floor, hence excluding them from use in spectroscopy and limiting the coverage. Still, the STFT is a useful tool for diagnosis of instabilities in a MH system. Let us look at the spectrogram of multiheterodyne beat notes observed in a vibrating terahertz QCL setup, which is a plot of STFT as a function of τ (Figure 3.35).

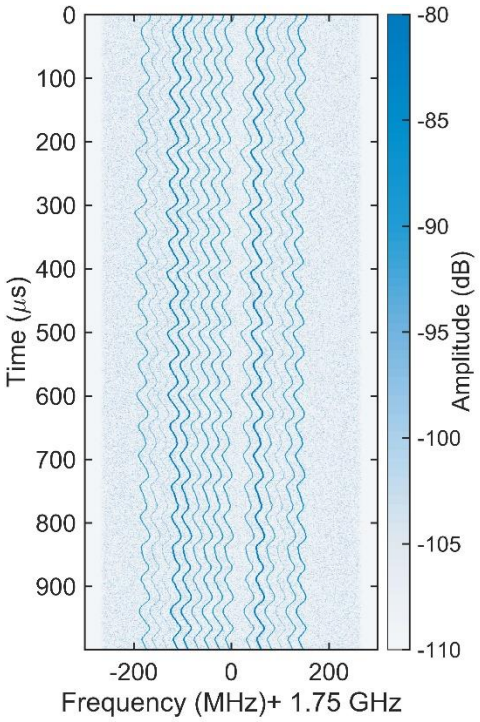

Figure 3.35. Spectrogram of a multiheterodyne signal in the DCS THz experiment. Results published in [102].

The discrete component structure of the MH spectrum changes its position in a periodic-like fashion. Since the displacement is comparable to the spacing between the lines, it is obvious that longer timescales will yield a featureless shape. This is additionally confirmed in Figure 3.36, which plots the MH spectrum at different acquisition times. In the spectrum acquired over 1 μs one can clearly see the dominant MH lines but MHz spectral resolution makes the weakest beat notes almost non-distinguishable from the large-variance noise floor. When the acquisition is increased to 10 μs, the beat notes are more pronounced, however they take irregular, asymmetric shapes. It is not obvious how to extract a representative quantity characterizing the power of such beat notes. Finally, when the

acquisition is extended to 100 μs, the spectrum looks like composed of almost overlapping rectangles with a width spanning tens of MHz. Clearly, we need to counteract this undesired effect.

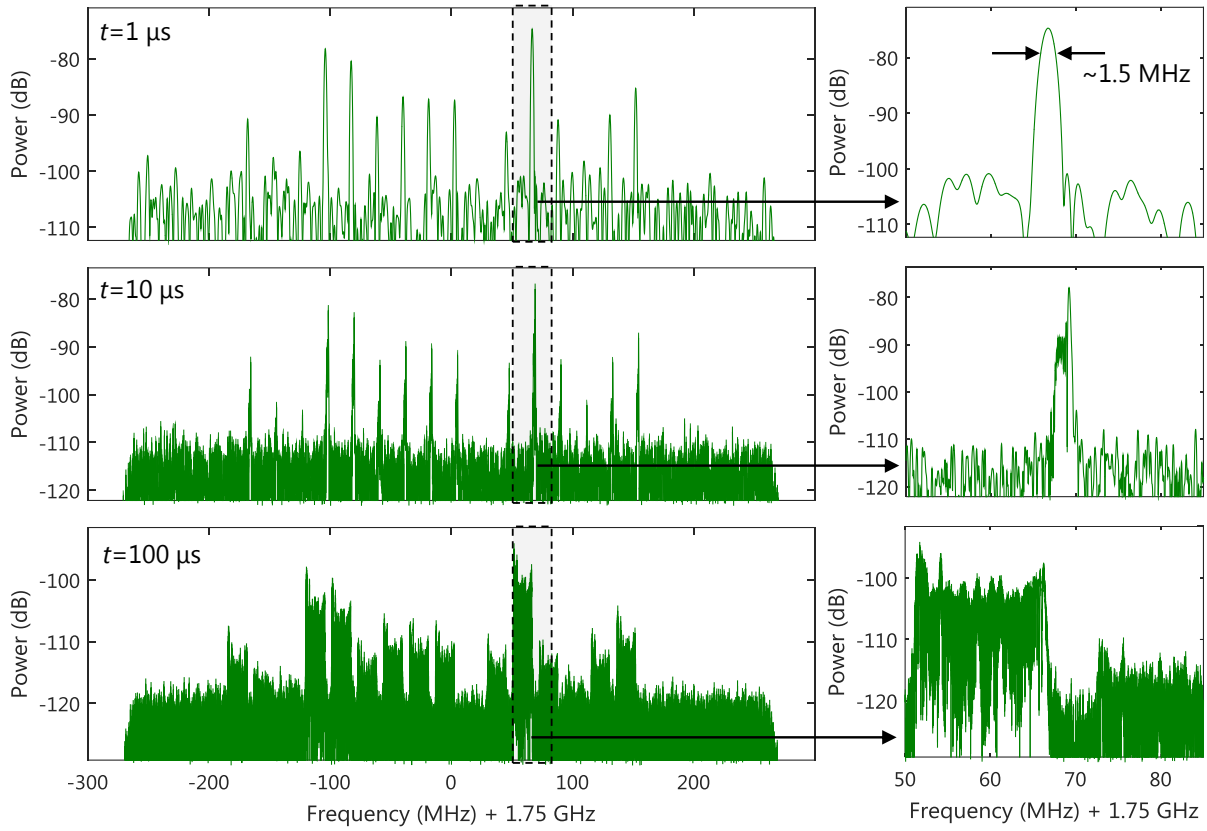

Figure 3.36. Short time Fourier Transform of a multiheterodyne signal rich in phase noise. A shorter timescale of 1 μs yields beat notes with linewidths almost limited by the acquisition time but allows for detection of fewer lines, and imposes a requirement of large separation between the lines. With an increasing acquisition time, we can see that the beat notes do not take Lorentz-like shapes. Instead, they turn into broad and uneven rectangles. At 1 ms, shown previously in Figure 3.34, the spectrum turns into a featureless blob. Results published in [102].

3.8.2. General scheme for fast coherent averaging

As we introduced at the beginning of this chapter, MH signals suffer from two superimposed undesired effects. First is the varying period of the beating signal (interferogram), stemming from the non-stationary instantaneous repetition rate difference $\Delta f_{\mathrm{rep}}(t)$. The second is related to the time-varying difference in offset frequencies $\Delta f_0(t)$, which equally shifts all MH lines. We proposed to correct an unstable MH signal in a two-step procedure. We equalize the duration of consecutive interferograms first, to apply a global offset phase correction as a final step. This algorithm is summarized graphically in Figure 3.37.

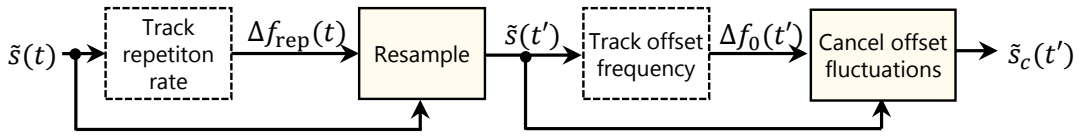

Figure 3.37. General scheme for fast coherent averaging of MH spectra. The tilde operator denotes complex signal.

This scheme remained unchanged throughout all revisions. What changed is the way how the correction signals were extracted. We will start with its most straightforward implementation, which will be extended by more sophisticated variants in further steps.

3.8.3. Fast computational adaptive sampling with frequency tracking

If multiheterodyne lines do not overlap, one can implement phase correction in an intuitive way using RF filters, adaptive resamplers, and phase shifters. An obvious fact is that wo neighboring beat notes in a MH spectrum are spaced by $\Delta f_{\mathrm{rep}}(t)$. When mixed together, they yield the instantaneous repetition rate, which can be used to resample non-uniformly sampled data, or serve as an adaptive clock for an acquisition card. This is also known as timing correction. Its principle is equivalent to playing an old magnetic tape that locally stretched or compressed with a variable rate to compensate for aging. The instantaneous "playback" rate of the interferogram can be expressed as

$$r(t) = \frac{\langle \Delta f_{\mathrm{rep}} \rangle}{\Delta f_{\mathrm{rep}}(t)}. \qquad (3.16)$$

where $\langle \cdot \rangle$ is the expected value operator. To equalize the duration of interferogram frames, one needs to resample (linearly interpolate) data on a new, non-uniform grid analogous to a non-linear time axis obtained by cumulative integration of r

$$t'(t) = \int_0^t \frac{\langle \Delta f_{\mathrm{rep}} \rangle}{\Delta f_{\mathrm{rep}}(\tau)} \mathrm{d}\tau. \qquad (3.17)$$

For discretely sampled data, the above turns into a sum, as in [77]. The result of this operation is a suppression of spectral "breathing" – all MH lines preserve mutual distances over time. Next step after resampling is to correct offset fluctuations using the instantaneous offset frequency $\Delta f_0(t)$ or phase $\Delta \varphi_0(t)$ retrieved from an arbitrary beat note, however stronger ones with a sufficient SNR are preferred. While it is possible to retrieve $\Delta f_0(t)$ using a frequency tracker, we will present it in further sections due to its increased complexity. Instead, we will discuss the most conventional way as first, which is used in various demodulation techniques and utilizes the complex envelope, with two almost equivalent implementations. One approach is to mix a selected RF beat note in quadrature with a fixed-frequency local oscillator (carrier), centered at the expected value of its frequency (jitter-free) f_0. We will refer to this case as baseband complex signal

$$\tilde{s}_{\mathrm{BB}}(t') = s(t') \exp(-\mathrm{j}2\pi \langle f_0 \rangle t'). \qquad (3.18)$$

This down-converts a passband signal to baseband, yielding two components: X and Y, which can be represented in polar coordinates

$$\tilde{s}_{\mathrm{BB}}(t') = R(t')\exp\big(\mathrm{j}\Theta(t')\big). \tag{3.19}$$

The instantaneous phase of $\tilde{s}_{\mathrm{BB}}(t')$ used for offset correction is simply

$$\Theta(t') = \Delta\varphi_0(t') = \tan^{-1}\left(\frac{Y(t')}{X(t')}\right), \tag{3.20}$$

and the amplitude (envelope)

$$R(t') = \sqrt{X(t')^2 + Y(t')^2} \tag{3.21}$$

used for instance to characterize the amplitude stability of a beat note. A schematic of a typical IQ demodulator is shown in Figure 3.38. It can be implemented using either hardware mixers and filters, or computational procedures with real-time FPGA implementation compatibility.

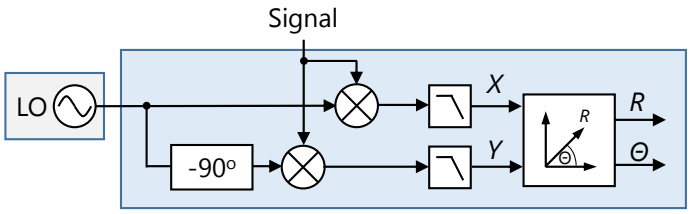

Figure 3.38. Schematic of an IQ demodulator.

An alternative formulation of the problem is to convert a real signal from the photodetector after resampling $s(t')$ into analytic form using the Hilbert transform

$$\tilde{s}(t') = s(t') + \mathrm{j}\mathcal{H}[s(t')], \tag{3.22}$$

where the Hilbert transform $\mathcal{H}[s(t')]$ shifts the phase of the negative frequency components by +90° ($\pi/2$ radians), and the positive by -90°. When multiplied by the imaginary unit j, an additional shift of +90° is introduced, which restores the positive frequencies and inverts the phase of the negative. After combining with the original signal $s(t')$, the negative frequencies cancel out due to their out-of-phase relationship, hence leaving a complex signal $\tilde{s}(t')$ with positive frequencies only. The asymmetry of the frequency spectrum is possible only for complex signals. If one drops the imaginary part, we can easily recover the original signal

$$\mathrm{Re}\{\tilde{s}(t')\} = s(t'). \tag{3.23}$$

The Hilbert transform is essentially a filter with impulse response

$$h(t) = \frac{1}{\pi t}. \tag{3.24}$$

Given a signal $x(t)$, we compute its Hilbert transform through a convolution with $h(t)$

$$\mathcal{H}[x(t)] = \hat{x}(t) = \frac{1}{\pi}\,\mathrm{p.v.}\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}\,\mathrm{d}\tau, \tag{3.25}$$

where p.v. is the Cauchy principal value. Having obtained the analytic signal representation (possible to implement in real time thanks to the multiply-accumulate architecture of the Hilbert filter), we can band-pass an arbitrary beat note centered at f_B, and retrieve its instantaneous phase φ_B in the same way as in IQ demodulation. The difference, however, will lie in the non-zero frequency shift equal to the expected value of the beat note's frequency $\langle f_\mathrm{B}\rangle$, which we can easily compensate numerically. We will refer to this case as passband complex signal.

To obtain the instantaneous offset frequency fluctuations in the baseband case, one calculates the derivative of the phase jitter $\Delta\varphi_0$

$$\Delta f_0(t') = \frac{1}{2\pi}\cdot\frac{\mathrm{d}(\Delta\varphi_0)}{\mathrm{d}t'}(t'). \tag{3.26}$$

For smoother estimates, it is also recommended to smoothen the frequency estimate using a Savitzky-Golay or similar filter. Since we want to avoid additional frequency shifts, we compensate for the expected frequency of the filtered beat note in the passband case, hence we obtain

$$\Delta f_0(t') = \left(\frac{1}{2\pi}\cdot\frac{\mathrm{d}\varphi_\mathrm{B}}{\mathrm{d}t'}(t')\right) - \langle f_\mathrm{B}\rangle. \tag{3.27}$$

Consequently, to convert φ_B to $\Delta\varphi_0$, we have to remove the linear phase term arising from non-zero value of $\langle f_\mathrm{B}\rangle$

$$\Delta\varphi_0(t') = \Delta\varphi_0(t') - 2\pi\langle f_\mathrm{B}\rangle t. \tag{3.28}$$

The final step in the coherent averaging algorithm is to correct for global offset fluctuations through phase-shifting in counterphase.

$$\tilde{s}_c(t') = \exp\big(-\mathrm{j}\Delta\varphi_0(t')\big)\,\tilde{s}(t'). \tag{3.29}$$

If one retrieves the instantaneous passband frequency as in frequency trackers, the formula takes form

$$\tilde{s}_c(t') = \exp\left(-\mathrm{j}2\pi\int_0^t f_\mathrm{B}(\tau) - \langle f_\mathrm{B}\rangle\mathrm{d}\tau\right)\cdot\tilde{s}(t'). \tag{3.30}$$

Again, the shifting operation can be realized either through fast point-wise operations or using an analog phase shifter or mixer. For convenience, the essential procedures of the correction algorithm are summarized in table form together with an example corrected MH spectrum obtained from two ICLs in Figure 3.39. More details regarding the procedure can be found in Ref. [77].

Simplified algorithm for phase correction of MH signals

1: Filter two neighboring strong beat notes, mix them to obtain $\Delta f_\mathrm{rep}(t)$.

2: Acquire data at a rate equal to a multiple of $\Delta f_\mathrm{rep}(t)$, or resample on a non-linear grid $t'(t)$.

3: Filter any strong beat note after correction, IQ-demodulate to obtain $\Delta\varphi_0(t')$.

4: Shift in counterphase to compensate for $\Delta\varphi_0(t')$.

End of algorithm

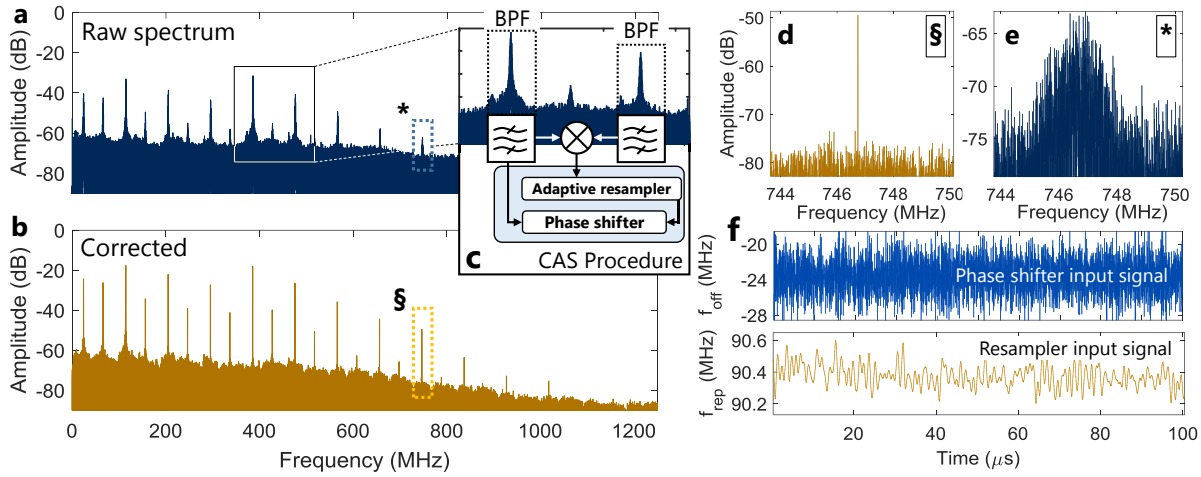

Figure 3.39. General scheme for fast coherent averaging of MH spectra. (a). Raw spectrum. (b) Phase-corrected spectrum. (c) Schematic of the computational adaptive sampling (CAS) procedure. (d) Beat note after correction. (e) Beat note before correction. (f) Computationally-retrieved correction signals. Results published in [103].

3.8.4. Extension to overlapping beat notes and harmonic resampling

The straightforward two-beat-note coherent averaging scheme has one significant drawback: it cannot deal with spectrally overlapping beat notes. Also by using neighboring beat notes to retrieve the correction signals, we observed a gradual increase in phase noise at beat notes located far from the correction pair. To mitigate both limitations, we have developed a simple yet powerful technique, which is a slightly modified version of the previous algorithm. A phase-noisy multiheterodyne signal with overlapping beat notes is first resampled using harmonics of the instantaneous repetition frequency retrieved from all (even overlapping) beat notes to make sure that only offset fluctuations remain in the signal. Next, by tracking the frequency of any beat note we can globally correct for the common offset phase noise. The only requirement is a complex representation of the signal, which can be obtained either through broadband IQ demodulation or the Hilbert transform.

To describe how this procedure works mathematically, we will start with the resampling step first, which employs Digital Difference Frequency Generation (DDFG). Consider a radio-frequency N-component complex multiheterodyne signal in form

$$\tilde{s}(t) = \sum_{n=1}^{N} \cos\big((\omega_0 + n\omega_r)t\big) + \mathrm{i}\sum_{n=1}^{N} \sin\big((\omega_0 + n\omega_r)t\big) = \sum_{n=1}^{N} \exp(\,\mathrm{i}(\omega_0 + n\omega_r)t), \qquad (3.31)$$

where ω_0 is the angular offset frequency, and ω_r is the angular repetition frequency. For simplicity, we assume equal amplitudes of the N frequency components. Our goal is to retrieve N harmonics of ω_r, while eliminating any frequency components containing ω_0. This problem is almost equivalent to difference frequency generation (DFG) – a process well known from the domain of nonlinear optics [104]. In the optical domain, the second order nonlinear susceptibility $\chi^{(2)}$ of a medium is responsible for multiple frequency mixing phenomena, such as sum frequency (SFG), second harmonic (SHG), and difference frequency generation (DFG), all dependent upon squared electric field. In principle, this scheme can be easily applied here to obtain the difference frequencies just by squaring the input signal. However, for a complex input we will just double and sum the signal's frequencies without creating any difference components. Using the well-known formula for the square of the sum of N numbers, we can write:

$$\left(\sum_{n=1}^{N} a_n\right)^2 = \sum_{n=1}^{N} a_n^2 + 2\sum_{m=1}^{N}\sum_{l=1}^{m-1} a_l a_m. \qquad (3.32)$$

If we plug the input signal model to the above equation, we arrive at

$$[\tilde{s}(t)]^2 = \underbrace{\sum_{n=1}^{N} \exp(\mathrm{i}(2\omega_0 + 2n\omega_r)t)}_{\text{SHG}} + 2\underbrace{\sum_{m=1}^{N}\sum_{l=1}^{m-1} \exp(\mathrm{i}(2\omega_0 + (l+m)\omega_r)t)}_{\text{SFG}}. \qquad (3.33)$$

The result is complex and spectrally asymmetric but there is no DFG term. Only the second harmonic, and sum frequency components appear in the equation.

Let us analyze now, what happens if we process the real and imaginary components separately. If we square the real part of $\tilde{s}(t)$, we obtain the same components as yielded by $\chi^{(2)}$ in nonlinear optics:

$$(\mathrm{Re}\{\tilde{s}(t)\})^2 = \underbrace{\frac{N}{2}}_{\text{DC}} + \underbrace{\frac{1}{2}\sum_{n=1}^{N} \cos(2\omega_0 t + 2n\omega_r t)}_{\text{SHG}} + \underbrace{\sum_{m=1}^{N}\sum_{l=1}^{m-1} \cos(2\omega_0 t + (l+m)\omega_r t)}_{\text{SFG}} + \underbrace{\sum_{m=1}^{N}\sum_{l=1}^{m-1} \cos\big((m-l)\omega_r t\big)}_{\text{DFG}}. \qquad (3.34)$$

In the above formula, we employed the product-to-sum formula for the cosine, namely $2\cos x \cos y = \cos(x+y) + \cos(x-y)$, where $x = \omega_0$, and $y = n\omega_r$. Since the complex character of the signal is lost, the frequency spectrum becomes symmetric with a significant DC term. The latter in nonlinear optics is known as optical rectification, and here the same process takes place: by squaring a signal, we rectify it thus producing DC. More importantly, the DFG term appears now in the equation,

yielding harmonics of ω_r. Unfortunately, they will often be overlaid by spectrally-broad low frequency SHG and SFG components containing ω_0. To make it even worse, higher frequency SFG and SHG products can spectrally alias if they appear at frequencies greater than half of the sampling frequency, thereby spectrally overlaying the DFG products as well. Both effects will introduce a cross-talk between the DFG products used to retrieve correction signals, and the unwanted SHG and SFG by-products. Clearly, a nonlinear operation that yields DFG alone without any second harmonic or sum products including ω_0 is of utmost importance.

If square the imaginary part of $\tilde{s}(t)$ now, the result is nearly the same, however the sign of the components containing ω_0 flips:

$$(\mathrm{Im}\{\tilde{s}(t)\})^2 = \underbrace{\frac{N}{2}}_{\mathrm{DC}} - \underbrace{\frac{1}{2}\sum_{n=1}^{N}\cos(2\omega_0 t + 2n\omega_r t)}_{\mathrm{SHG}} - \underbrace{\sum_{m=1}^{N}\sum_{l=1}^{m-1}\cos(2\omega_0 t + (l+m)\omega_r t)}_{\mathrm{SFG}} + \underbrace{\sum_{m=1}^{N}\sum_{l=1}^{m-1}\cos\big((m-l)\omega_r t\big)}_{\mathrm{DFG}}. \quad (3.35)$$

Note that the minus sign appears exactly where it is needed. If we combine the two squared parts of the complex signal, any influence on ω_0 disappears, therefore we arrive at the final formula for multiple harmonics of the repetition rate of a multiheterodyne signal, referred here to as Digital Difference Frequency Generation (DDFG):

$$(\mathrm{Re}\{\tilde{s}(t)\})^2 + (\mathrm{Im}\{\tilde{s}(t)\})^2 = \underbrace{N}_{\mathrm{DC}} + 2\underbrace{\sum_{m=1}^{N}\sum_{l=1}^{m-1}\cos\big((m-l)\omega_r t\big)}_{\mathrm{DFG}}. \quad (3.36)$$

Why is this formula useful? Because it gives easy access to all harmonics of the repetition rate and allows for direct use of fast harmonic frequency trackers [105]–[107]. Alternatively, seeing that fluctuations of the repetition rate have a cumulative effect, we can pick a high-order repetition rate harmonic beat note, filter it using less computationally demanding low-order bandpass filters, next retrieve its instantaneous frequency through IQ demodulation, and finally scale the estimated frequency by the harmonic number. In contrast to ultra-sharp narrowband filters used for retrieving the instantaneous repetition frequency of low-order beat notes close to DC, as only such are usable in signals contaminated by SHG and SFG products, the proposed high-order harmonic approach introduces less pronounced edge effects and ringing artifacts, which translates into more efficient correction of a noisy multiheterodyne signal. A comparison of the different nonlinear operations on a complex MH signal is shown in Figure 3.40.

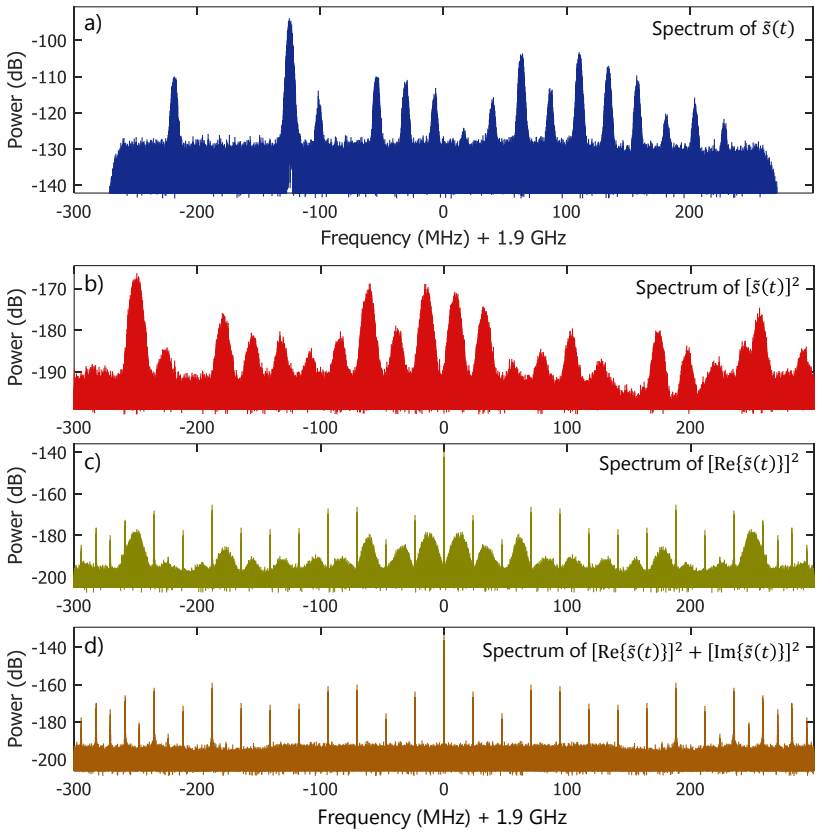

Figure 3.40. Spectral comparison of the different nonlinear operations applied to the acquired complex multiheterodyne signal in (a). Panel (b) shows the result of squaring the complex signal, which yields SHG and SFG albeit without any DFG products. Panel (c) shows the spectrum of the squared real part, which creates a DC component together with SHG, SFG, and DFG. Note the spectral symmetry, and overlap of the components. (d) Spurious-free DFG signal obtained by adding the independently squared real and complex part of the signal. Slight aliasing is also visible for the highest order repetition rate harmonics. Results published in [102].

After filtering one of the repetition rate harmonics obtained through DDFG (or tracking its frequency using a tracking algorithm), we resample the MH signal by the same token as in the first revision of the correction procedure. Another thing the two revisions have in common, is that after resampling we can choose any region of the frequency spectrum to track the RF offset frequency. While it is possible to track all of them, as in the Multiple Frequency Tracker case, it always comes at the expense of increased computational complexity, therefore narrowed regions are preferred. As long as the repetition rate correction is efficient, no difference should be observed between a broadband multi-line and narrowband near-single-line tracking. A flowchart illustrating how signals are retrieved in the overlapped beat note case is plotted in Figure 3.41, whereas the algorithm is summarized in table form.

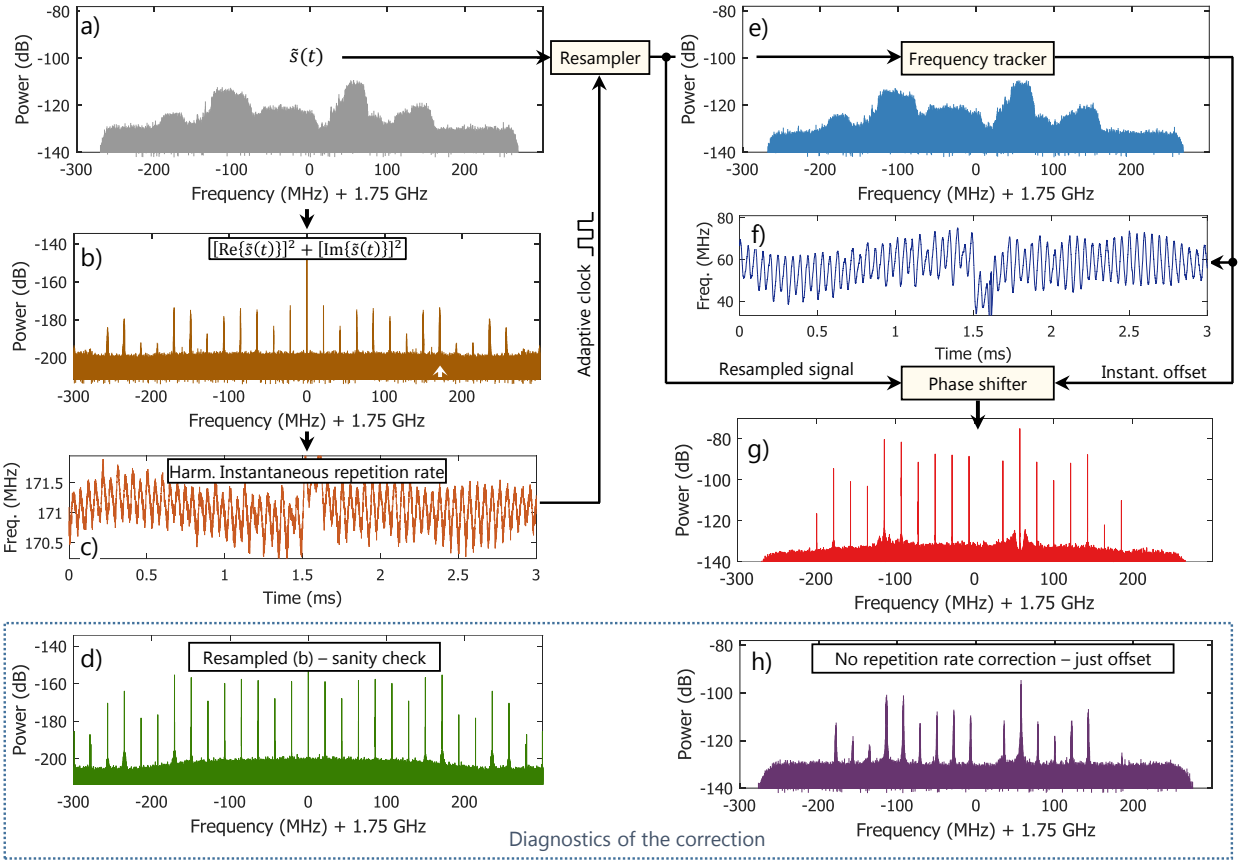

Figure 3.41. Signal flowchart in the extended phase correction algorithm. (a) Raw signal (b) DDFG signal (c) Instantaneous frequency of the 8th harmonic marked with white arrow. (d) Resampled DDFG signal as a validation tool. (e) Resampled original signal. (f) Instantaneous frequency of one of the tracked beat notes. (g) Corrected signal. (h) Signal corrected for offset only. Results published in [107].

Algorithm for phase correction of MH signals with overlapped beat notes

1: Obtain a complex representation of the signal, i.e. using the Hilbert transform or IQ demodulation.

2: Retrieve the harmonics of the repetition rate through DDFG.

3: Track the frequency (phase) the highest order harmonic possible with a sufficient SNR.

4: Resample the MH signal using the instantaneous repetition rate.

5: Track the instantaneous offset frequency using one of the frequency trackers.

6: Correct the phase of the resampled signal.

End of algorithm

One of the crucial elements in the correction algorithm is the frequency tracker. We have successfully implemented three different procedures, whose derivation will be presented now.

3.8.5. Frequency tracker using the Kalman filter

The first algorithm enabling frequency tracking is the celebrated Kalman Filter. To derive the necessary formulas, we consider a non-stationary real sinusoid with a time-varying amplitude $A(t)$, phase $\varphi(t)$, and frequency $\nu(t)$. At each time instant, it can be expressed as

$$s(t) = A(t)\sin(\varphi(t)) = A(t)\sin\left(2\pi\int_0^t \nu(\tau)\mathrm{d}\tau + \varphi_0\right). \quad (3.37)$$

Our goal is to estimate the instantaneous parameters A, ν, and φ provided data points from a noisy measurement. Unfortunately, sinusoid is a non-linear function and special tools must be used to solve this non-linear estimation problem, which in general is not an easy task. One of the well-suited methods is the extended Kalman filter (EKF). In contrast to the regular (non-extended) version, the EKF can deal with the nonlinearity, however the solution is no longer optimal due to the approximation of the nonlinear function around a working point with Taylor Series. Despite this imperfection, we will show that the EKF has practical relevance as a sufficiently good frequency tracker for correcting a noisy multi-heterodyne signal.

The data in a real-world spectroscopic experiment are discretely sampled with an interval Δt, therefore we need to estimate the state vector $\mathbf{x}$ at each step k with the arrival of a new sample. For clarity, we will assume that $\mathbf{x}_k = \mathbf{x}(t_k)$, and that the estimate of the true state vector $\mathbf{x}$ is denoted by $\hat{\mathbf{x}}$. Also note that superscript in $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k-1|k-1}$, means k-th, and $(k-1)$-th iteration, given $k-1$ samples, respectively.

The state vector, we will try to estimate, consists of three time-varying parameters: amplitude, phase, and frequency:

$$\mathbf{x}_k = \begin{bmatrix} A_k \\ \varphi_k \\ \nu_k \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \quad (3.38)$$

A change of the state vector in the next time instant can be modeled as an operation on the nonlinear process (state transition) function plus an external perturbation of the parameters such as a change in amplitude or frequency

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}, \quad (3.39)$$

where $\mathbf{w}_{k-1}$ is the process Gaussian noise vector assumed to have zero mean and covariance $\mathbf{Q}_k$, and $\mathbf{f}(\mathbf{x}_k)$ is the nonlinear process function

$$\mathbf{f}(\mathbf{x}_k) = \begin{bmatrix} A_k \\ \varphi_{k-1} + 2\pi\nu_k\Delta t \\ \nu_k \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 + 2\pi x_3\Delta t \\ x_3 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}. \quad (3.40)$$

Unfortunately, in the real world we do not have access to the true state vector $\mathbf{x}$, but instead we measure samples corrupted by instrumental noise $\mathbf{v}_k$ (also assumed to be zero-mean Gaussian with covariance $\mathbf{R}_k$). The samples are governed by the measurement equation

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k, \tag{3.41}$$

where $\mathbf{v}_k$ is the measurement noise vector, and the observation model function $\mathbf{h}(\mathbf{x}_k)$ is a vector function of the amplitude and phase (frequency is the derivative of the phase and is a by-product of the latter):

$$\mathbf{h}(\mathbf{x}_k) = A_k \cdot \sin(\varphi_k) = [x_1 \cdot \sin(x_2)]_{\mathbf{x}_k}. \tag{3.42}$$

Given the set of the above model equations, we will try to estimate the time-varying $\mathbf{x}_k$. In contrast to the non-extended (linear) Kalman Filter, we cannot plug **f** and **h** directly to obtain the estimates. Instead, we need to perform linearization of the observation and process functions about the working point. To do that, we will calculate the Jacobian, which is a matrix of partial derivatives of a function. Since we have only one observable variable in **h** (it is a scalar function), the Jacobian **H** becomes a row vector

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k|k-1}} = \left[\frac{\partial h}{\partial x_1} \quad \frac{\partial h}{\partial x_2} \quad \frac{\partial h}{\partial x_3}\right]_{\hat{\mathbf{x}}_{k|k-1}} = [\sin(x_2) \quad x_1 \cos(x_2) \quad 0]_{\hat{\mathbf{x}}_{k|k-1}}. \tag{3.43}$$

The process **f**, on the contrary, is a vector function of three variables (amplitude, phase and frequency), therefore the Jacobian **F** becomes a 3x3 matrix

$$\mathbf{F}_{k-1} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{k-1|k-1}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_3} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_3}{\partial x_1} & \cdots & \frac{\partial f_3}{\partial x_3} \end{bmatrix}_{\hat{\mathbf{x}}_{\mathrm{k-1|k-1}}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2\pi\Delta t \\ 0 & 0 & 1 \end{bmatrix}_{\hat{\mathbf{x}}_{k-1|k-1}}. \tag{3.44}$$

Equipped with this model, we can perform the estimation of the noisy sinusoid. However, to enable fast convergence of the Kalman Filter, it is reasonable to incorporate some prior knowledge about the signal. For instance, a short time Fourier Transform can be used to provide an initial guess of the amplitude, phase, and frequency. Alternatively, one can guess the values. While the exact numerical value is not critical, the order of magnitude is. We set the initial guess of the state vector in

$$\hat{\mathbf{x}}_0 = \begin{bmatrix} A_0 \\ \varphi_0 \\ \nu_0 \end{bmatrix}. \tag{3.45}$$

What is of even larger importance than the initial guess, is how much the process variables, and measured samples are expected to change, which is reflected in the noise covariance matrices. First, we need to roughly approximate the variance of the additive noise the measurement $\mathbf{v}_k$, contained in the instrumental noise covariance $\mathbf{R}_k$ matrix, which can be obtained by measuring the noise floor profile within an arbitrary frequency range in the system. Recall that the noise spectral density of white Gaussian noise assumed in the model is flat and frequency independent (factor of two comes from the contribution of positive and negative frequencies)

$$S_{\mathrm{vv}}(\nu) = S_{\mathrm{vv}} = \frac{N_0}{2}. \tag{3.46}$$

A good estimate N_0, can be obtained from averaged periodograms of a detector signal with a blocked light input, where N_0 is taken as the mean across a frequency interval. Since S_{vv} is in units of V^2/Hz, in the discrete case we divide it by the sampling interval Δt to preserve energy, as in Parseval's integral

$$\sigma_{\mathrm{v}}^2 = \frac{S_{\mathrm{vv}}}{\Delta t} \quad \left[\frac{\mathrm{V}^2}{\mathrm{Hz}} \cdot \frac{1}{\mathrm{s}} = \frac{\mathrm{V}^2}{\mathrm{Hz}} \cdot \mathrm{Hz} = \mathrm{V}^2\right]. \tag{3.47}$$

Luckily, for the single sine wave case, $\mathbf{R}_k$ simplifies to only one numerical quantity: the noise variance σ_{v}^2, which additionally is assumed to not depend on time, therefore $\mathbf{R}_k = \mathbf{R} = \sigma_{\mathrm{v}}^2$.

Secondly, the covariance matrix $\mathbf{Q}_k$ of the process noise vector $\mathbf{w}_k$ needs to be approximated to fully describe the model. We have three variables that change over time and the degree of their mutual dependencies – covariance σ_{xy} – must be evaluated. For the continuous case, the covariance matrix is in form

$$\mathbf{Q}_{\mathrm{c}} = \begin{bmatrix} \sigma_A^2 & \sigma_{A\varphi} & \sigma_{A\nu} \\ \sigma_{\varphi A} & \sigma_{\varphi}^2 & \sigma_{\varphi\nu} \\ \sigma_{\nu A} & \sigma_{\nu\varphi} & \sigma_{\nu}^2 \end{bmatrix}. \tag{3.48}$$

In our model, we expect only the amplitude and frequency to jitter. The time-varying phase can be seen as a secondary effect of the varying frequency. Therefore, the covariance matrix takes form

$$\mathbf{Q}_{\mathrm{c}} = \begin{bmatrix} \sigma_A^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{\nu}^2 \end{bmatrix}. \tag{3.49}$$

Unfortunately, the discrete time steps require to discretize the process covariance matrix $\mathbf{Q}_{\mathrm{c}}$. To do that, we need to take several steps. Initially, we convert our continuous time process model into state space form, and we express frequency through phase derivative $\nu = \dot{\varphi}$, where the over-dot denotes a time derivative.

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{A} \\ \dot{\varphi} \\ \dot{\nu} \end{bmatrix} = \mathbf{F}_{\mathrm{d}}\mathbf{x} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} A \\ \varphi \\ \nu \end{bmatrix}. \tag{3.50}$$

Next, we will calculate the fundamental matrix $\mathbf{\Phi}$ using Taylor expansion of the exponential function. Since $\mathbf{F}_{\mathrm{d}}^{\mathbf{2}} = \mathbf{0}$, it will be limited to the first term

$$\mathbf{\Phi} = \mathrm{e}^{\mathbf{F}_{\mathrm{d}}t} = \mathbf{I} + \mathbf{F}_{\mathrm{d}}t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.51}$$

By substituting Δt for t in $\mathbf{\Phi}$, the discrete fundamental matrix is given by

$$\boldsymbol{\Phi}_k = \boldsymbol{\Phi}(\Delta t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.52)$$

Finally, $\boldsymbol{\Phi_k}$ is plugged into the covariance discretization integral to obtain the discrete process noise matrix:

$$\mathbf{Q}_k = \int_0^{\Delta t} \boldsymbol{\Phi}_k \mathbf{Q}_c \boldsymbol{\Phi}_k^{\mathrm{T}} = \int_0^{\Delta t} \begin{bmatrix} \sigma_A^2 & 0 & 0 \\ 0 & \sigma_v^2 \Delta t^2 & \sigma_v^2 \Delta t \\ 0 & \sigma_v^2 \Delta t & \sigma_v^2 \end{bmatrix} = \begin{bmatrix} \sigma_A^2 \Delta t & 0 & 0 \\ 0 & \frac{\Delta t^3 \sigma_v^2}{3} & \frac{\Delta t^2 \sigma_v^2}{2} \\ 0 & \frac{\Delta t^2 \sigma_v^2}{2} & \sigma_v^2 \Delta t \end{bmatrix}. \quad (3.53)$$

where the time index *k* can be skipped due to the time-invariance, thus $\mathbf{Q}_k = \mathbf{Q}$. The author has found the most critical to set the diagonal entries. The filter provided a reasonably good estimate of the state vector even without the off-diagonals, but obviously a better model will likely lead to faster convergence.

Having derived all the necessary variables in the non-stationary sinusoid tracking problem, we can summarize the algorithm for its tracking.

Algorithm for the estimation of a noisy sinusoid using the Extended Kalman Filter

Initialize (only once):

1: Guess the initial value of state vector $\mathbf{x}_{0|0}$ i.e. supported by short term Fourier Transform. In not possible, set to the expected value of **x**.

2: Estimate the discrete noise variance σ_v^2 in **R** from the PSD of the noise floor of the spectrum.

3: Estimate the noise of amplitude σ_A^2 and frequency σ_v^2 in **Q**, for instance by IQ-demodulating around the center frequency of the beat note to track. For a typical multiheterodyne system, they remain relatively reproducible and the crucial part is to estimate the order of magnitude rather than absolute numerical value.

4: Set the covariance (uncertainty) matrix of the initial state estimate $\mathbf{P_{0|0}}$ to either low values on the diagonal if the initial guess of $\mathbf{x}_{0|0}$ is very accurate, or $\mathbf{P_{0|0}} = \mathbf{Q}$ otherwise.

Predict:

5: Predict the state vector estimate in k^{th} iteration given *k*-1 samples

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{x}}_{k-1|k-1}). \quad (3.54)$$

6: Predict the covariance (uncertainty) of the state estimate

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^{\mathrm{T}} + \mathbf{Q} \quad (3.55)$$

Update:

7: Calculate the measurement residual – the difference between the observation model and new sample

$$\tilde{\mathbf{y}}_k = \boldsymbol{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}). \quad (3.56)$$

8: Obtain the covariance (uncertainty) of the measurement residual

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \mathbf{R} = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \sigma_{\mathrm{V}}^2. \quad (3.57)$$

9: Calculate the near-optimal Extended Kalman Filter gain

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} \mathbf{S}_k^{-1}. \quad (3.58)$$

10: Update the state vector estimate using the EKF gain and residual

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{y}}_k. \quad (3.59)$$

11: Update the covariance estimate (uncertainty) of the state vector estimate

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}. \quad (3.60)$$

12: Return to point (5) and loop while new samples in $\boldsymbol{z}_k$ are available.

End of algorithm

Without major changes, the complex sinusoid (cisoid) case can be resolved too. The model of the signal takes form

$$\tilde{s}(t) = A(t) \mathrm{e}^{\mathrm{i}\left(2\pi \int_0^t \nu(\tau) \mathrm{d}\tau + \varphi_0\right)}. \quad (3.61)$$

Rather than real, at each time instant we get a new complex sample

$$\mathbf{h}^{(\mathbf{c})}(\mathbf{x}_k) = A_k \cdot \mathrm{e}^{\mathrm{i}\varphi_k} = \left[x_1 \mathrm{e}^{\mathrm{i}x_2}\right]_{\mathbf{x}_k}. \quad (3.62)$$

The internal process governing the state vector and its Jacobian remain unchanged (we observe the same variables: amplitude, phase and frequency), but the Jacobian of the observation function needs to be replaced to account for the complex exponential:

$$\mathbf{H}_{\mathrm{k}}^{(\mathbf{c})} = \left.\frac{\partial \mathbf{h}^{(\mathrm{c})}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{\mathrm{k|k-1}}} = \left[\mathrm{e}^{\mathrm{i}x_2} \quad \mathrm{i}x_1 \mathrm{e}^{\mathrm{i}x_2} \quad 0\right]_{\hat{\mathbf{x}}_{\mathrm{k|k-1}}}. \quad (3.63)$$

These are almost the only changes in the model to enable the tracking of the cisoid with the EFK. The newly derived $\mathbf{H}_{\mathrm{k}}^{(\mathbf{c})}$ must obviously replace $\mathbf{H}_{\mathrm{k}}$, and the instrumental noise covariance $\mathbf{R}_k$ matrix must include complex white Gaussian noise with a variance $\boldsymbol{\Gamma}$, rather than real σ_{V}^2. The process covariance matrix $\mathbf{Q}$ on the contrary will remain real, since the amplitude, phase, and frequencies remain real even in the complex case.

While it is possible to reconstruct all multiheterodyne lines using this approach by re-formulating the initial problem in terms of the sum of all multiheterodyne lines [76], the computational complexity can be cumbersome for instances with tens of hundreds lines. Indeed, there are implementations of the Fast Kalman Filter [108], but we will propose several other equally efficient yet less computationally demanding approaches.

3.8.6. Multifrequency tracker (MFT)

At first glance, the intrinsic complexity of the Extented Kalman Filter can be overwhelming. It requires to make many assumptions regarding the underlying process and involves a careful characterization of noise variances to ensure convergence of the algorithm. While it models the problem very accurately, such knowledge is not always obtainable. Additionally, from a practical standpoint it takes much effort to fine-tune the EKF to get near-optimal results due to the large number of parameters.

An alternative approach, which seems to be quite robust and applicable for phase correction of noisy multiheterodyne systems, is to assume that over shorter timescales (on the order of the sampling period) frequency changes in a linear fashion and amplitude varies slowly. P. Tichavsky *et al.* [109] proposed an algorithm to track multiple frequencies corrupted by white Gaussian noise with a near zero-tracking delay (lower bias), which cannot be easily obtained with the Extended Kalman Filter. Also in the case of crossing frequencies it is analytically proven to perform better.

In the algorithm, at each time instant we model the multiheterodyne signal as a sum of *p* cisoids (complex sinusoids):

$$y_n = \sum_{k=1}^{p} A_{kn} + v_n, \quad n = 0,1,2 \dots , \tag{3.64}$$

where A_{kn} denotes the *k*-th component at *n*-th time instant, and v_n is the noise. Notably this model assumes no phase noise correlation between the lines, which can be potentially improved in the future. Let us remind that after the repetition rate correction, there is no need to track all frequency components in a MHS signal. It is sufficient to track only one component (for instance filtered using a bandpass filter whose center frequency follows the frequency estimate) or just a few, if beat notes mutually overlap, assuming a fixed frequency bandpass filter is used to isolate a group of such overlapping MH lines.

Each component A_{kn} can be represented in analytic form with a magnitude and phase

$$A_{kn} = |A_{kn}|\mathrm{e}^{\mathrm{i}\varphi_{kn}}, \quad k = 1,2, \dots p. \tag{3.65}$$

and stored in the signal decomposition vector

$$\boldsymbol{A}_n = \left(A_{1n}, \dots, A_{pn}\right)^{\mathrm{T}}. \tag{3.66}$$

Having defined the signal components A_{kn}, we can express the instantaneous angular frequency

$$\omega_{kn} \triangleq \mathrm{Arg}\left[\frac{A_{kn}}{A_{k,n-1}}\right], \tag{3.67}$$

and the frequency rate-of-change

$$\alpha_{k,n-1} \triangleq \mathrm{Arg}\left[\frac{A_{kn}A_{k,n-2}}{A_{k,n-1}^2}\right]. \tag{3.68}$$

The index of $n-1$ reflects the fact that for $\alpha_{k,n}$ we would need to have access to a future sample $A_{k,n+1}$, which does not exist. Both the angular frequency and its rate-of-change can be summarized in the frequency prediction equation:

$$\widetilde{\omega}_{kn} = \widehat{\omega}_{kn} + \hat{\alpha}_{k,n-1}, \tag{3.69}$$

where the hat symbol denotes an estimate, and tilde the prediction. For convenience, all variables are normalized to the sampling frequency, so that no time interval is included in the formulas. Now, if we define a vector of ones:

$$\boldsymbol{J} = (1 \dots 1)^{\mathrm{T}} \qquad (p \times 1), \tag{3.70}$$

a diagonal matrix with angular frequencies in the complex exponential:

$$\mathbf{F}_n = \mathrm{diag}\left(\mathrm{e}^{\mathrm{i}\omega_{1n}}, \dots, \mathrm{e}^{\mathrm{i}\omega_{pn}}\right) \qquad (p \times p), \tag{3.71}$$

and recall the inversion property of the diagonal matrix

$$\mathbf{F}_n^{-1} = \mathrm{diag}\left(\left(\mathrm{e}^{\mathrm{i}\omega_{1n}}\right)^{-1}, \dots, \left(\mathrm{e}^{\mathrm{i}\omega_{pn}}\right)^{-1}\right) = \mathrm{diag}\left(\mathrm{e}^{-\mathrm{i}\omega_{1n}}, \dots, \mathrm{e}^{-\mathrm{i}\omega_{pn}}\right), \tag{3.72}$$

we can describe the measured signal at any time instant as a sum of all past frequency shifts multiplied by the signal decomposition vector and the noise:

$$y_{n-m} = \boldsymbol{J}^{\mathrm{T}}\mathbf{F}_{n-(m-1)}^{-1}\mathbf{F}_{n-(m-2)}^{-1} \dots \mathbf{F}_n^{-1}\boldsymbol{A}_n + v_{n-m}. \tag{3.73}$$

The above equation is subject to quadratic optimization with a forgetting factor λ, which yields the recursion formulas. For convenience, we will summarize the algorithm in the same form as for the Kalman Filter. Note that in the algorithm formulas we will use hats above variables to denote their estimates, since we do not have access to their true values.

Algorithm for the estimation of a noisy sinusoid using the Multiple Frequency Tracker

Initialize (only once):

1: Guess the initial value of the p normalized angular frequencies ω_{k0}, set them on the diagonal of $\widehat{\mathbf{F}}_0$.

2: Set $\widehat{\boldsymbol{A}}_0 = \widehat{\boldsymbol{A}}_{-1} = \widehat{\boldsymbol{A}}_{-2}$, and $\boldsymbol{\alpha}_0 = \left[\alpha_{1n} \dots \alpha_{pn}\right]^{\mathrm{T}}$ to a vector of zeros $\mathbf{0}^{(p\times 1)}$.

3: Choose the desired limit of the forgetting factors for the amplitude λ_∞, angular frequency ρ_∞, and frequency rate of change μ_∞ together with the convergence factors for these variables: λ_c, ρ_c and μ_c, all from the interval (0,1). Set the initial value of each design variable using a general formula: $\lambda_0 = (1-\lambda_c)\lambda_\infty$. These factors suppress initial transients occurring in the initial phase of tracking.

4: Set $\mathbf{G_0} = (1-\lambda_\infty)^{-1}\mathbf{I}_{p\times p}$, where $\mathbf{I}_{p\times p}$ is a diagonal matrix.

5: Iterate the steps below while forcing $\hat{\alpha}_{kn} = 0$, until rough convergence is reached or over *m* burn-in samples. Once reached, include the estimation of $\hat{\alpha}_{kn}$.

Estimate:

6: Estimate the decomposition vector in the *n*-th iteration given *n*-1 samples

$$\widehat{\boldsymbol{A}}_n = \hat{\mathbf{F}}_n \widehat{\boldsymbol{A}}_{n-1} + \mathbf{G}_n^{-1}\boldsymbol{J}\left[y_n - \boldsymbol{J}^{\mathrm{T}}\hat{\mathbf{F}}_n \widehat{\boldsymbol{A}}_{n-1}\right], \tag{3.74}$$

where the auxiliary $p \times p$ matrix is given by

$$\mathbf{G}_{\boldsymbol{n}} = \boldsymbol{J}\boldsymbol{J}^{\mathrm{T}} + \lambda \hat{\mathbf{F}}_n \mathbf{G}_{n-1} \hat{\mathbf{F}}_n^{\mathrm{H}}, \tag{3.75}$$

with

$$\hat{\mathbf{F}}_{\mathrm{n}} = \mathrm{diag}\left(\mathrm{e}^{\mathrm{i}(\widehat{\omega}_{1,n-1} + \widehat{\alpha}_{1,n-1})}, \ldots, \mathrm{e}^{\mathrm{i}(\widehat{\omega}_{p,n-1} + \widehat{\alpha}_{p,n-1})}\right). \tag{3.76}$$

7: Estimate the frequency rate-of-change

$$\hat{\alpha}_{kn} = \hat{\alpha}_{k,n-1} + (1-\mu)\,\mathrm{Arg}\left[\frac{\hat{A}_{kn}\hat{A}_{k,n-2}}{\hat{A}_{kn}^2 \mathrm{e}^{\mathrm{i}\hat{\alpha}_{k,n-1}}}\right]. \tag{3.77}$$

8: Estimate the angular frequencies

$$\widehat{\omega}_{kn} = \widehat{\omega}_{k,n-1} + \hat{\alpha}_{kn} + (1-\rho)\,\mathrm{Arg}\left[\frac{\hat{A}_{kn}}{\hat{A}_{k,n-1}\mathrm{e}^{\mathrm{i}(\widehat{\omega}_{k,n-1} + \widehat{\alpha}_{kn})}}\right], \qquad k = 1 \ldots p. \tag{3.78}$$

9: Update the forgetting factors λ, ρ and μ in the *n*-th iteration using the convergence speed constants λ_c, and the desired limits λ_∞in a general formula below. Substitute the remaining design variables: ρ and μ respectively.

$$\lambda_n = \lambda_c \lambda_{n-1} + (1-\lambda_c)\lambda_\infty \tag{3.79}$$

10: Return to point (6) and loop while new samples in y are available.

End of algorithm

It should be noted that the design variables can be used to fine-tune the algorithm for a specific character of fluctuations. In the ICL-based experiments, the limit of the forgetting factor λ_∞ of 0.95 and convergence parameters λ_c=0.2 yielded the least oscillatory frequency estimate and the cleanest noise pedestal, whereas in the QCL experiments, slightly higher values closer to 1 were preferred.

3.8.7. Fast Frequency Tracker (Fast FT)

A common feature of the Kalman filter and the Multifrequency Tracker is that they require matrix inversion, known as one of the most computationally-inefficient operations, hence should be avoided whenever possible. A remedy for this issue was proposed in Ref. [110] by R. Aarts, who proved that the recursion

$$\hat{r}_k = \hat{r}_{k-1} + x_{k-1}\gamma[x_k + x_{k-2} - 2x_{k-1}\hat{r}_{k-1}] \tag{3.80}$$

can be used to track the instantaneous frequency f_0 of a real sampled signal in $x(k)$, given by

$$r_k = \cos(2\pi f_0(k) T_s), \tag{3.81}$$

where $T_s = 1/f_s$ is the sampling period. Since the tracking formula $\hat{r}_k$ uses only multiplications and additions, it can undoubtedly be implemented in a real-time platform. The convergence parameter γ is

a small positive number that controls the convergence speed of the algorithm, and enables tracking slightly non-stationary signals with time-varying amplitude A_0, and obviously f_0. The time constant of the tracking behavior is related to γ through

$$\tau_d = \frac{T_s}{\gamma A_0^2}, \quad (3.82)$$

which can be approximated a priori to compensate for the tracking delay. From a practical standpoint, it may be beneficial to scale the signal using automatic gain control (AGC), or by a constant factor to keep τ_d approximately invariant on signal amplitude. Since the algorithm models the signal r_k, to obtain the ordinary instantaneous frequency, we need to calculate its arccosine followed by scaling by the ratio of sampling frequency and 2π:

$$\hat{f}_0(k) = \cos^{-1}(\hat{r}_k) \cdot \frac{f_s}{2\pi}. \quad (3.83)$$

To further reduce the computational complexity and tracking time, it is possible to tabulate arccosine for processing units without efficient floating-point units. The final step of tracking is to smoothen the frequency estimate using for instance a moving average, or a Savitzky-Golay (SG) filter. As before, for convenience the algorithm is summarized in table form.

Algorithm for the estimation of a noisy sinusoid using the Fast Frequency Tracker
Initialize (only once):
1: Choose the convergence parameter γ. A good starting point for ICL/QCL-based MHS signals sampled at 4 GS/s was $\gamma = 0.3$.
2: Iterate the recursive formula $\hat{r}_k$ over *n* burn-in (initial) samples. n=100 is a good starting point.
Estimate:
3: Iterate the recursive formula $\hat{r}_k$, compute the arccosine of the estimate.
4: Smoothen the estimate $\hat{f}_0(k)$ using a moving average or a S-G filter.
5: Return to point (3) and loop while new samples in x are available.
End of algorithm

This algorithm was preferred as a frequency tracker in our most recent DCS experiments due to its high speed, and tracking capabilities comparable to the other two algorithms. Nevertheless, one should note its general incompatibility with multifrequency signals, where the MFT or EKF is preferred.

3.9. Amplitude noise modulation effects on the frequency spectrum

In addition to multiplicative phase noise in MHS, multiplicative amplitude noise also cannot be neglected, since it considerably changes the shape of the spectrum. Let us assume that the intensity of a

beat note at a constant angular (carrier) frequency ω_c varies in a random fashion and is corrupted by zero-mean white Gaussian noise with variance σ_w^2. The noisy signal is given by

$$x(t) = A(t)\cos(\omega_c t) = \big(A_{\mathrm{DC}} + w(t)\big)\cos(\omega_{\mathrm{c}} t). \tag{3.84}$$

The above formula essentially describes amplitude modulation (AM). In particular, the effect of $w(t)$ is well-known for the case of a modulating sine or cosine with frequency ω_m, which produces sidebands at $\omega_c \pm \omega_m$. Consequently, we cannot assume that random fluctuations of $A(t)$ will affect only one frequency ω_c, but we expect it to corrupt the entire spectrum. By calculating the Fourier Transform of $x(t)$, we can formulate the problem in the frequency domain

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = \frac{A_{\mathrm{DC}}}{2}\,\delta(\omega \pm \omega_{\mathrm{c}}) + \int_{-\infty}^{\infty} w(t)\cos(\omega_{\mathrm{c}} t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t. \tag{3.85}$$

The first term is just a Dirac Delta at positive and negative carrier frequency, but the second part cannot be directly evaluated, because $w(t)$ is random and the integral does not converge.

A satisfactory solution for our needs would be to derive the power spectral density of the random process $x(t)$, unfortunately at the expense of phase information, which is lost in PSD. To do that, we will employ the Wiener-Khintchin relationship between PSD and the autocorrelation function R_{xx}. If x(t) is a wide-sense stationary process, meaning that is has a time-independent mean, its autocorrelation depends only on time difference τ, and has a finite variance, the PSD is given by

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau = \mathcal{F}\{R_{xx}(\tau)\}, \tag{3.86}$$

where the autocorrelation function is the expected value of the product of $x(t)$ and $x(t+\tau)$

$$R_{xx}(\tau) = \mathrm{E}\{x(t)x(t+\tau)\}. \tag{3.87}$$

Let us start with the mean of the process

$$m_x(t) = \mathrm{E}\big\{\big(A_{\mathrm{DC}} + w(t)\big)\cos(\omega_{\mathrm{c}} t)\big\} = A_{\mathrm{DC}}\cos(\omega_{\mathrm{c}} t), \tag{3.88}$$

which unfortunately depends on time, in contrast to what is required for a WSS process. Also, the autocorrelation depends on t :

$$\begin{aligned} R_{xx}(t,\tau) &= \frac{1}{2}\big(A_{\mathrm{DC}}^2 + \mathrm{E}\{(w(t)w(t+\tau)\}\big)[\cos(\omega_{\mathrm{c}}\tau) + \cos(2\omega_{\mathrm{c}} t + \omega_{\mathrm{c}}\tau)] \\ &= \frac{A_{\mathrm{DC}}^2 + \sigma^2\delta(\tau)}{2}[\cos(\omega_{\mathrm{c}}\tau) + \cos(2\omega_{\mathrm{c}} t + \omega_{\mathrm{c}}\tau)]. \end{aligned} \tag{3.89}$$

The above was derived by utilizing the property that for a white Gaussian noise process $w(t)$, the autocorrelation function is a Dirac delta function (because $\mathrm{E}\{w(t)w(t+\tau)\} = 0$ for $\tau \neq 0$) scaled by the variance σ_w^2

$$R_{ww}(\tau) = \sigma^2\delta(\tau) \tag{3.90}$$

with the spectral density constant over frequency

$$S_{ww}(\omega) = \mathcal{F}\{R_{ww}(\tau)\} = \sigma^2. \tag{3.91}$$

In the expression for $R_{xx}(t,\tau)$ and $m_x(t)$ we see, that the noise-modulated cosine process is not WSS, however both the mean and autocorrelation are periodic. Consequently, we can consider a special class of stationarity of periodic functions called *cyclostationarity*, which will enable us to derive the *cyclic spectrum* rather than power spectral density for a *wide sense cyclostationary process*. We can consider the following cyclic expression:

$$R_{xx}(t,\tau) = \frac{A_{\mathrm{DC}}^2 + \sigma^2\delta(\tau)}{2}[\cos(\omega_{\mathrm{c}}\tau) + \cos(2\omega_{\mathrm{c}}t + \omega_{\mathrm{c}}\tau)], \tag{3.92}$$

where the periodicity T_0 with respect to t is equal to $2\pi/\omega_{\mathrm{c}} = 1/f_{\mathrm{c}}$ and is chosen so that both $m_x(t)$ and $R_{xx}(t,\tau)$ are periodic (the choice of $T_0 = \pi/\omega_{\mathrm{c}}$, which works for $R_{xx}(t,\tau)$, does not satisfy the requirement for $m_x(t)$). Also the mean value is cyclic, which means that the process is *wide sense cyclostationary*, therefore we can plug the periodic autocorrelation to the cyclic autocorrelation given in general by

$$R_x^{\alpha=\frac{n}{T_0}}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} R_x(t,\tau)\mathrm{e}^{-\frac{\mathrm{j}2\pi nt}{T_0}}\mathrm{d}t, \tag{3.93}$$

which we will evaluate at frequencies $\alpha = n/T_0$ known as cyclic frequencies. Consequently, we get

$$R_x^{\alpha=\frac{n}{T_0}}(\tau) = \frac{\left(A_{\mathrm{DC}}^2 + \sigma^2\delta(\tau)\right)\omega_{\mathrm{c}}}{2\cdot 2\pi}\int_{-\frac{\pi}{\omega_{\mathrm{c}}}}^{\frac{\pi}{\omega_{\mathrm{c}}}} (\cos(\omega_{\mathrm{c}}\tau) + \cos(2\omega_{\mathrm{c}}t + \omega_{\mathrm{c}}\tau))\mathrm{e}^{-\mathrm{j}nt\omega_{\mathrm{c}}}\,\mathrm{d}t. \tag{3.94}$$

The first element in the integral is the Fourier Transform of a rectangular window function, and is essentially a scaled *sinc* function

$$\begin{aligned}\int_{-\frac{\pi}{\omega_{\mathrm{c}}}}^{\frac{\pi}{\omega_{\mathrm{c}}}} \cos(\omega_{\mathrm{c}}\tau)\, e^{-\mathrm{j}nt\omega_{\mathrm{c}}}\mathrm{d}t &= \frac{\mathrm{j}\cos(\omega_{\mathrm{c}}\tau)}{n\omega_{\mathrm{c}}}\mathrm{e}^{-\mathrm{j}nt\omega_{\mathrm{c}}}\Big|_{t=-\frac{\pi}{\omega_{\mathrm{c}}}}^{t=\frac{\pi}{\omega_{\mathrm{c}}}} = \\ &= \frac{-2\cos(\omega_{\mathrm{c}}\tau)}{2\mathrm{j}n\omega_{\mathrm{c}}}\left(\mathrm{e}^{-\mathrm{j}n\pi} - \mathrm{e}^{\mathrm{j}n\pi}\right) = \frac{2\cos(\omega_{\mathrm{c}}\tau)\sin(n\pi)}{n\omega_{\mathrm{c}}}.\end{aligned} \tag{3.95}$$

In a similar fashion, the second element can be evaluated

$$\int_{-\frac{\pi}{\omega_{\mathrm{c}}}}^{\frac{\pi}{\omega_{\mathrm{c}}}} \cos(2\omega_{\mathrm{c}}t + \omega_{\mathrm{c}}\tau)\mathrm{e}^{-\mathrm{j}nt\omega_{\mathrm{c}}}\,\mathrm{d}t = \frac{2\sin(n\pi)\left(n\cos(\omega_{\mathrm{c}}\tau) + 2\mathrm{j}\sin(\omega_{\mathrm{c}}\tau)\right)}{(n^2-4)\omega_{\mathrm{c}}}. \tag{3.96}$$

Plugging the solved integrals into the cyclic autocorrelation, we arrive at

$$R_x^{\alpha=\frac{n}{T_0}}(\tau) = \frac{2\left(\mathrm{A}_{\mathrm{DC}}^2 + \sigma^2\delta(\tau)\right)\omega_{\mathrm{c}}\sin(n\pi)}{4\pi}\left[\frac{\cos(\omega_{\mathrm{c}}\tau)}{n\omega_{\mathrm{c}}} + \frac{n\cos(\omega_{\mathrm{c}}\tau) + 2\mathrm{j}\sin(\omega_{\mathrm{c}}\tau)}{(n^2-4)\omega_{\mathrm{c}}}\right] = \\ = \frac{\left(A_{\mathrm{DC}}^2 + \sigma^2\delta(\tau)\right)\sin(n\pi)\left((n^2-2)\cos(\omega_{\mathrm{c}}\tau) + \mathrm{j}n\sin(\omega_{\mathrm{c}}\tau)\right)}{\pi n(n^2-4)}. \quad (3.97)$$

Now, we can calculate the cyclic spectrum known as the spectral correlation density function at cyclic frequency α using the cyclic Wiener relation

$$S_x^{\alpha}(\omega) = \int_{-\infty}^{\infty} R_x^{\alpha}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau, \quad (3.98)$$

which for the noise-modulated cosine is

$$S_x^{\alpha=\frac{n}{T_0}}(\omega) = \frac{\sin(n\pi)}{n\pi(n^2-4)}\int_{-\infty}^{\infty}\left(A_{\mathrm{DC}}^2 + \sigma^2\delta(\tau)\right)\left((n^2-2)\cos(\omega_{\mathrm{c}}\tau) + \right. \\ \left. + \mathrm{j}n\sin(\omega_{\mathrm{c}}\tau)\right)e^{-\mathrm{j}\omega\tau}\mathrm{d}\tau. \quad (3.99)$$

In fact, we have four integrals to solve, which are respectively:

$$\int_{-\infty}^{\infty} A_{\mathrm{DC}}^2(n^2-2)\cos(\omega_{\mathrm{c}}\tau)\,\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau = \frac{(n^2-2)A_{\mathrm{DC}}^2}{2}\delta(\omega \pm \omega_{\mathrm{c}}) \quad (3.100)$$

$$\int_{-\infty}^{\infty} \sigma^2\delta(\tau)(n^2-2)\cos(\omega_{\mathrm{c}}\tau)\,\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau = (n^2-2)\sigma^2 \quad (3.101)$$

$$\int_{-\infty}^{\infty} A_{\mathrm{DC}}^2\left(\mathrm{j}\,n\sin(\omega_{\mathrm{c}}\tau)\right)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau = \frac{nA_{\mathrm{DC}}^2}{2}\delta(\omega \pm \omega_{\mathrm{c}}) \quad (3.102)$$

$$\int_{-\infty}^{\infty} \sigma^2\delta(\tau)\left(\mathrm{j}\,n\sin(\omega_{\mathrm{c}}\tau)\right)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau = 0 \quad (3.103)$$

which plugged into the expression for the spectral correlation density function yield

$$S_x^{\alpha=\frac{n}{T_0}}(\omega) = \frac{\sin(n\pi)}{n\pi(n^2-4)}\left(\frac{(n^2+n-2)A_{\mathrm{DC}}^2\delta(\omega \pm \omega_{\mathrm{c}})}{2} + (n^2-2)\sigma^2\right) = \\ = \frac{\sin(n\pi)}{n\pi}\left(\frac{(n-1)}{(n-2)}\frac{A_{\mathrm{DC}}^2\delta(\omega \pm \omega_{\mathrm{c}})}{2} + \frac{(n^2-2)}{(n^2-4)}\sigma^2\right). \quad (3.104)$$

For $n = 0$, if we assume that $\mathrm{sinc}(0) = 1$, or calculate the limiting case for $n \to 0$, the average power spectral density can be expressed as

$$S_x^{\alpha=0}(\omega) = \frac{A_{\mathrm{DC}}^2}{4}\delta(\omega \pm \omega_{\mathrm{c}}) + \frac{\sigma^2}{2}. \quad (3.105)$$

The obtained result is somewhat expected because the modulating signal decomposed spectrally into flat-profile noise, and a constant amplitude cosine at positive and negative frequencies. We predicted

that the noise would create uniform "sidebands" at all possible frequencies. Note also that the average power spectral density does not reflect instantaneous changes of the carrier, but characterizes its average behavior, as the name suggests.

3.9.1. Numerical experiment of noise-modulated cosine

To evaluate the effect of noisy amplitude modulation on the carrier to noise ratio, we will perform a numerical experiment with different acquisition times and signal-to-noise ratios. Since it will be conducted in the discrete time domain, we will need to adjust our model accordingly.

First, the discrete noise variance $\sigma_{w\mathrm{D}}^2$ in the simulation must be scaled by the sampling frequency (bandwidth) of the experiment to be consistent with Parseval's theorem. We will consider a single-sided power spectral density; therefore, we multiply S_{ww} by 2:

$$\sigma_{w\mathrm{D}}^2 = 2S_{ww} \cdot f_{\mathrm{S}} = \sigma^2 \cdot f_{\mathrm{S}}. \tag{3.106}$$

Also the Dirac delta distribution in the discrete time and finite acquisition case will not be accurately reproduced because it is not physically realizable. A finite acquisition translates into an uncertainty in frequency. Consequently, rather than being infinitely narrow with infinite amplitude, the discrete spectrum will mimic δ function in an area under curve sense: the height of the DFT peak will increase proportionally to an increasing acquisition time, while decreasing its 3-dB width. For the simplest case of a rectangular window, we can approximate the peak height of the cosine's PSD estimate using the gain-bandwidth product:

$$\hat{S}_m^{\mathrm{d}}(f_k = \pm f_{\mathrm{c}}) \approx \frac{A^2}{4\Delta f}, \tag{3.107}$$

where A denotes the amplitude, and Δf is the DFT bin width proportional to $1/t_{\mathrm{acq}}$. For a single-sided PSD, this value will be doubled.

In our numerical experiment, whose results are plotted in Figure 3.42, a cosine wave at $f_{\mathrm{c}} = \omega_{\mathrm{c}}/2\pi = 100$ Hz with a mean amplitude $A = 1$ is modulated by white Gaussian noise with zero mean and unitary variance σ_{ND}^2. The sampling frequency f_{s} is 10 times larger than the carrier (1 kHz). The experiment is carried out in two scenarios: in the first, the acquisition time is varied in factors of 10 ranging between 0.1 s and 100 s while keeping the amplitude and variance constant, whereas in the second, the acquisition time is constant and the variance (signal-to-noise ratio) is changed. What do we expect to see? Let us start with the noise floor.

At a constant sampling frequency $f_{\mathrm{s}} = 1$ kHz, we expect the noise spectral density to be independent on the acquisition time

$$\hat{S}_w^{\mathrm{d}}(f) = 10\log_{10}\left(\frac{\sigma_{w\mathrm{D}}^2}{f_{\mathrm{s}}}\right) = 10\log_{10}\left(\sigma_{w\mathrm{D}}^2\right) - 30\ \left[\frac{\mathrm{dB}}{\mathrm{Hz}}\right]. \tag{3.108}$$

However, when the power spectrum is measured, as in a typical spectrum analyzer, by increasing the acquisition time, we can reduce the resolution bandwidth (RBW), and hence the Displayed Average Noise Level (DANL). In other words, at a constant noise level, longer acquisitions lower the noise floor, thus enabling to resolve components with a lower signal-to-noise ratio.

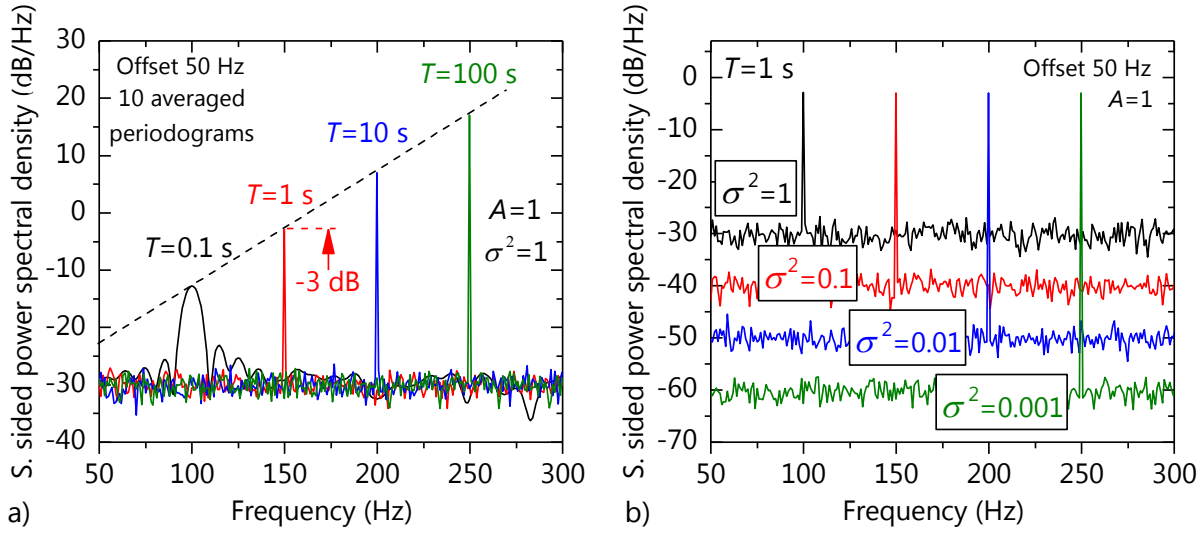

Figure 3.42. Single sided power spectral density of an amplitude modulated cosine by a white Gaussian process with unitary mean in a discrete- and finite-time experiment. (a) The acquisition time was changed while keeping the variance unitary. (b) The variance was changed in steps by a factor of 10 while keeping the acquisition time fixed.

Next thing we expect to see is that at a constant acquisition time, the spectral signal-to-noise ratio changes proportionally to $(A/\sigma)^2$. Figure 3.42 proves our expectations and validates the previously derived formulas. In principle, one can extrapolate this experiment to estimate the minimum detectable signal (spectral sensitivity) within a given t_{acq}. We clearly see the expected linear trend. However, in a real spectroscopic system one has to deal with a superposition of many additional noise sources in contrast to the simplified single-noise-source model that we derived. A thorough review of numerous contributors is given in Ref. [111]. One factor that deserves a more careful analysis is that in our model we assumed a white character of amplitude fluctuations, which is almost never true. In a bandwidth-limited system, white noise samples are in fact correlated. Rather than being a Dirac delta, the autocorrelation of a WSS white noise signal turns into a *sinc* function scaled by the bandwidth B:

$$R_W(\tau) = BN_0\,\mathrm{sinc}(2B\tau), \qquad (3.109)$$

where in the sampled case we replace τ with k/f_s. If our noise was purely white, we could sample the signal at zeros of the *sinc* function occurring at $\tau_i = i/2B$ to get uncorrelated samples, but a real-world beat note shows colored noise characteristics, which further modify the autocorrelation function. For

instance, in the OPLL system we observed that the Allan deviation curve considerably oscillated. This will inevitably modulate the carrier amplitude. Simply speaking, modulation of a cosine with a colored noise can cause the carrier to fluctuate, so that the estimation of the mean value without long-term averaging would be very inaccurate. Nevertheless, our simplified model is useful to understand the effects of averaging on amplitude estimation over shorter timescales, where white noise characteristics dominate.

3.10. Uncertainty analysis

So far, we analyzed, how phase and amplitude noise affects the measured RF spectrum. Now we will discuss how it translates into uncertainties in transmission spectrum – a format preferred for spectroscopic analysis and quantification.

3.10.1. FFT-related uncertainty in amplitude and power estimation

As it was shown in Chapter 2, amplitude estimation using the Discrete Fourier Transform can be subject to many errors. One of them is related to the shape of the used window. In DCS, even a slightest change in the position of the beat notes has serious implications on the accuracy of amplitude retrieval due the problem of scalloping loss discussed earlier, analyzed here again in linear scale. Recall that after phase correction a dual-comb spectrum consists of purely discrete, narrowband frequency components. A likely to occur sub-bin-wide shift of beat notes' frequencies can cause the peak amplitude to be modulated by the magnitude frequency response of the used window. If the peak frequency misses the center of the DFT bin by half of the bin width, the underestimation of amplitude can be as large as ~36%. The severity of this undesired effect for different windows is depicted in Figure 3.43.

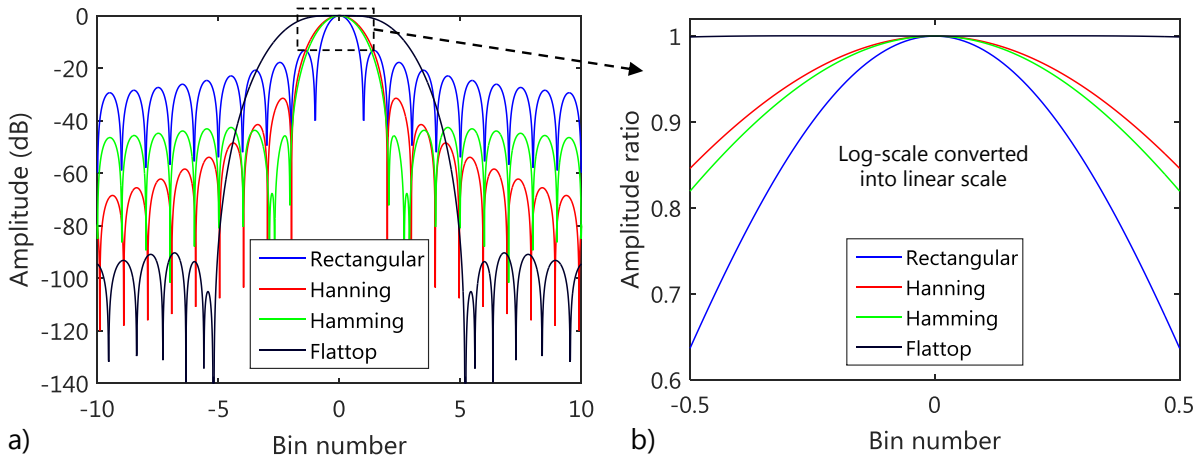

Figure 3.43. Magnitude response and scalloping loss for the most common DFT windows. (a) Spectral leakage and frequency resolution. (b) Severity of the scalloping loss plotted in linear scale. A fixed frequency sinusoidal signal can fall anywhere within the bin, hence suffer from a large underestimation of its amplitude.

Amplitude underestimation is the tip of the iceberg. In DCS the measured quantity is RF power, which squares the amplitude, thereby making the errors even higher. If residual shifts in the absolute position of the RF spectrum have a random-like character, we can characterize the problem in statistical terms: the mean, standard deviation, and worst-case error, as summarized in Table 3.1, and Table 3.2.

Table 3.1. Relative errors in amplitude and power estimation (normalized) due to a sub-bin-wide shift in position of a single-frequency sine wave for different spectral windows. The most extreme cases occur whenever the frequency lies exactly between two FFT bins.

Window name	Max amplitude error (%)	Max power error (%)	Mean amplitude error (%)	Mean power error (%)
Rectangular	36.34	59.47	12.97	23.00
Hamming	18.05	32.84	6.34	11.96
Hanning	15.40	28.43	5.37	10.23
Flat-top	0.10	0.21	0.0015	0.003

Table 3.2. Standard deviation of amplitude and power estimation (normalized) due to a sub-bin wide shift in position of a single-frequency sine wave for different spectral windows.

Window name	Standard deviation of beat note amplitude σ_A	Standard deviation of beat note power σ_P
Rectangular	0.1128	0.1897
Hamming	0.0557	0.1026
Hanning	0.0475	0.0885
Flat-top	0.0003	0.0006

3.10.2. Uncertainty in transmission

The underestimation of amplitude and power has even more severe implications when optical transmission is calculated by rationing the power of an absorber spectrum to a reference spectrum. Because any uncertainties arising from scalloping losses add in quadrature, the transmission spectrum can be tremendously erroneous. We will illustrate this problem on an example of two consecutive measurements of an empty absorption cell in an asymmetric DCS configuration, expected to yield a perfectly flat line. A detailed analysis and derivation of the formulas, including a balanced asymmetric arrangement can be found in Ref. [83]. The propagation of uncertainty states that the relative error of transmission is

$$\frac{\sigma_T}{T} \approx \sqrt{\left(2\frac{\sigma_{\mathrm{GS}}}{V_{\mathrm{GS}}}\right)^2 + \left(2\frac{\sigma_{\mathrm{ZS}}}{V_{\mathrm{ZS}}}\right)^2}, \qquad (3.110)$$

where $\left(\frac{\sigma_{\mathrm{GS}}}{V_{\mathrm{GS}}}\right)^2$ denotes the relative error of a beat note's power in the absorber spectrum, and $\left(\frac{\sigma_{\mathrm{ZS}}}{V_{\mathrm{ZS}}}\right)^2$ is the relative error in the reference spectrum. Since we measure an empty cell twice, the above formula simplifies to

$$\frac{\sigma_T}{T} \approx \sqrt{8\sigma_A^2} \approx 2\sqrt{2}|\sigma_A| \approx 2.82\sigma_A. \qquad (3.111)$$

Now, if we plug the standard deviation of amplitude σ_A, we can estimate the severity of scalloping loss on the transmission, which is summarized in Table 3.3. It is evident, that in the rectangular window case the uncertainty exceeds 30%, whereas the choice of flat-top window almost eliminates this effect.

Table 3.3. Uncertainty of the transmission of an empty cell assuming a change in position of the multiheterodyne beat notes by less than one DFT bin (typically kHz to tens of kHz).

Window name	**Error in transmission σ_T/T (%)**
Rectangular	31.81
Hamming	15.71
Hanning	13.40
Flat-top	0.08

Unfortunately, the price we pay for this convenience is high – flat-top windows have considerably lower resolution capabilities, hence are not very well suited for weak beat notes. A better solution that efficiently suppresses scalloping loss is to zero-pad the signal windowed first using for instance a Hanning window, followed by computing the FFT. As a result, even the narrowest beat notes, whose linewidth is limited by the acquisition time, will always occupy more than one FFT bin, hence slight frequency shifts will not severely affect the amplitude estimate. This approach obviously involves additional computational complexity, but it can be minimized using clever techniques of local spectral interpolation [112], taking advantage of the sparse structure of a dual-comb RF spectrum.

Another technique used in combination with zero-padding is the Welch method of overlapped periodograms, as introduced in the previous chapter. Since it is a low-variance periodogram estimator, it naturally provides more accurate values of spectral amplitude, and should be used whenever possible.

Chapter 4

Conclusions and outlook

In summary, in this thesis we studied the signal processing aspects of spectroscopy with frequency combs. We considered two distinct classes of experiments: a single and dual comb arrangement.

In the single frequency comb case, we analyzed a terahertz time domain spectrometer operating in homodyne detection mode. We showed how phase noise caused by slight variations in the position of a moving mirror in the optical delay line precludes long term averaging by narrowing the bandwidth of the terahertz spectrum, and additionally gives rise to spectral artifacts. The Allan deviation characterization proves that thanks to cross-correlation based signal alignment one can coherently average multiple terahertz pulses to mitigate these effects, suppress the noise, improve the dynamic range, and lower the uncertainty of the measurement. We determined the optimal averaging time in the system, which is limited by a linear drift occurring after 80 seconds. These findings are supported by the example of solid state α-D-glucose dehydrate, as well as water vapor. Additionally, we showed that without accounting for the random character of the phase at low frequencies of the terahertz signal, the characterization of the refractive index may be severely erroneous, which was mitigated by linear extrapolation of the phase. Finally, we demonstrate the relevance of advanced signal processing in THz-TDS for characterization of the reaction kinetics of pharmaceutical ingredients.

In the dual comb case we developed novel analog and digital phase and frequency synchronization techniques for coherent averaging of optical beating signals to account for broad or even mutually overlapping RF lines caused by substantial amounts of phase noise corrupting two unsynchronized highly miniaturized frequency combs in the mid-infrared and terahertz regime. Under such conditions, unambiguous mapping of the optical spectrum to the RF domain is hindered, even though the mode-to-mode phase relationship within a comb is maintained. The developed tools were used in the first demonstration of multiheterodyne spectroscopy with phase locked QCLs in the LWIR, as well as highly miniaturized, low-power ICL comb sources in the MIR suitable for battery-operated dual comb spectrometers. The near real-time phase correction technique enabled to surpass some of the limitations of the hardware loop, thus narrowing the line width of the RF beat notes down to the acquisition time limit and enabling to break the barrier of 100 beat notes observed simultaneously with possibilities for further improvements. Furthermore, this thesis presents the first experimental demonstration of molecular dual comb spectroscopy with chip-scale THz QCL combs. The phase correction procedure efficiently suppressed increased amounts of frequency jitter caused by a noisy environment of the

cryostat housing the THz combs. In the final step, the feasibility of dual comb spectroscopy with completely free-running mode-locked ICL combs is presented. All the demonstrations are based on HITRAN database fits for evaluation purposes, proving the validity of the developed signal processing tools.

In both cases of vastly different frequency comb platforms, coherent averaging proved to be critical for the successful retrieval of spectroscopic data from raw time-domain measurements. Ordinary point-wise averaging either introduced severe spectral artifacts in the terahertz time-domain setup, or hindered the mapping of individual comb teeth to the radio frequency domain in the dual-comb arrangement. These issues were mitigated thanks to the application of the coherent averaging procedures, which suppressed the phase noise, increased the signal-to-noise ratio, and lowered the uncertainty in both amplitude and phase. Consequently, both transmission and dispersion spectra of the spectroscopically probed medium could be correctly characterized.

Due to the large interest of the scientific community in miniaturization of sensing platforms, and rapid progress in improving the performance of broadband low-dispersion QCL and ICL combs, it is expected that future chip-scale dual-comb spectrometers will cover bandwidths spanning tens or even hundreds of terahertz with much greater absolute frequency stability. Also, thanks to the detection capabilities of the sources, a dual comb spectrometer may consists of just two lasers without a separate photodetector.

In terms of signal processing, it is highly desired to implement the proposed phase correction procedures in a real-time FPGA platform for coherent averaging over timescales not limited by data storage capabilities. Additional work may be also required to take full advantage of the sparse structure of the comb for efficient data compression.

Chapter 5

Appendix

This chapter reviews some of the experimental details crucial for reproducibility of results. We focus on the dual-comb setup, where we encountered challenges in the design of radio frequency filters for comb stabilization, microwave probes for intermode beat note extraction, and the choice of the proper operating point in the superconducting bolometer for measurements in the terahertz range.

5.1. RF Filter design for the Frequency Locked Loop

The RF filter design for the frequency locked loop was based on two contradictory requirements: a wide range of linear frequency discrimination and a steep slope. While the actual design of such a low pass filter may not be of large interest for the reader, it is worth mentioning that we could not find a readily available solution on the market. Most low-pass filters, even when cascaded, did not provide a sufficiently steep slope and did not enable the ease of the fine tuning of the characteristics obtainable with wire-wound air core inductors. Therefore, we assembled a filter on a copper laminate using surface mount capacitors and home-made inductors enclosed in a metal shield with SMA ports, as shown in Figure 5.1. After several design iterations, we have found the circuit shown in the figure to be meet the requirements. The discrimination slope centered at 40 MHz (1/4 of the FSR difference) changed in a nearly linear fashion by 2.3 dB/MHz.

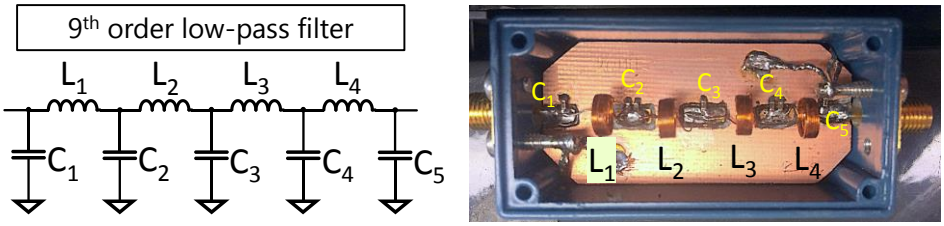

Figure 5.1. Custom RF low-pass filter acting as a frequency discriminator. (a) electrical schematic (b) implementation.

A careful reader will immediately realize that the PID control loop requires the error signal to be a linear function of frequency deviation. Since the filter's slope is linear only in the log-scale, it in needs to be linearized before further processing. This is realized by an integrated RF gain detector (Analog Devices, AD8302), which acts as a log-amplifier that compares the decimal logarithm of the ratio between its two inputs denoted by *A* and *B* multiplied by a slope coefficient V_{slp}

$$V_{\text{out}} = V_{\text{off}} + V_{\text{slp}} \cdot \log\left(\frac{V_A}{V_B}\right). \quad (5.1)$$

Effectively, the circuit combined with the filter provides nearly linear frequency to amplitude conversion over more than 10 MHz, yielding approximately 70 mV per each MHz of deviation with the advantage of being fast: the bandwidth of the log amplifier is 30 MHz.

5.2. Detection of THz radiation using NbN superconducting hot electron bolometer

The terahertz radiation in the dual comb setup was detected using a custom-designed quasioptical superconducting hot electron bolometer equipped with a helium bath cryostat (RS-LH2L-0.3-4T-0.05/1/10000, Scontel Russia), optimized for use in the 0.3-4 THz range. Its control unit allowed to stabilize either current or voltage of the superconductor, and supply sensitive high frequency cryogenic amplifiers.

The sensing element of the detector is a ~7 nm thick bridge of a superconducting niobium nitrate (NbN) film deposited on an insulating substrate kept at 4 K in a liquid helium bath cryostat. It acts as a hot electron bolometer: the incident radiation heats the electrons in the superconductor, which have a low, almost negligible thermal capacity, in contrast to that of the crystal lattice molecules [113], [114]. In superconducting materials, the electrical resistance is strongly dependent on the electron temperature, and it can be easily measured by providing an external voltage or current bias. The small size of the

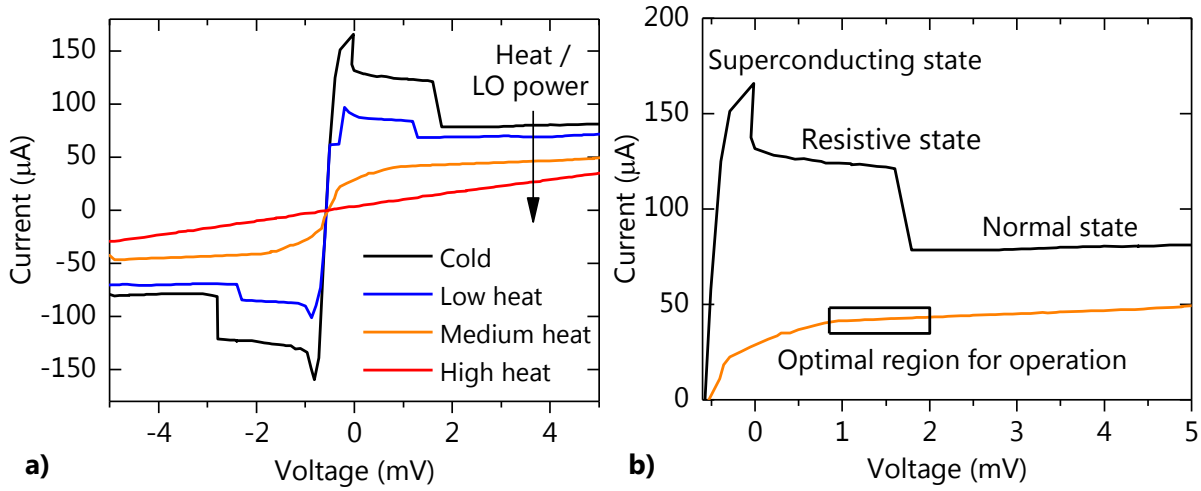

Figure 5.2. I-V characteristics of the liquid helium cooled NbN superconductor at different temperatures (equivalent to different pumping conditions). (a) Full family of curves. At low temperatures (top curves), several different operation regimes can be identified. At higher temperatures, superconductivity is suppressed and the sensing element turns into an ordinary resistor (red curve). (b) Zoom of the operation regimes of the cold and medium heat curves.

sensing element combined with the low thermal constants of the sensing mechanism allows to respond to changes in optical power within ns, yielding GHz bandwidths.

Figure 5.2 shows the I-V characteristics of the NbN HEB mixer, which strongly depend on temperature and take antisymmetric shapes varying between a highly nonlinear superconductor (black) and an ordinary linear resistor (red). The higher the temperature, the more superconductivity is suppressed. Of course, to obtain the optimal sensitivity and linearity of detection, we need to find the optimal operating point with a suitable power of the local oscillator (LO). Let us analyze the different regimes in the sensing element starting from near-zero bias. In principle, all curves should be antisymmetric around zero, however external circuitry can introduce a slight voltage offset, as in this example.

At low temperatures without LO pumping (Figure 5.2a), the I-V characteristics look like a single-cycle square wave with large overshoots. Initially, the NbN is in a superconducting state with a steep increase in current vs. voltage, however it tends to roll off at higher bias voltages, and eventually enters a negative resistance regime, where the current decreases with an increasing voltage. This is an erratic curve representing time-averaged fast (~MHz to kHz) oscillatory transitions between the superconducting state and the resistive state [115]. Interestingly, the amplitude of these oscillations can be as large as several millivolts, which complicates an initial alignment of the THz system. For instance, when a THz source is weakly coupled to the bolometer, it induces itself microvolts on the detector, being completely buried below these parasitic millivolt-level oscillations. This region should be also avoided because a periodic modulation of the operating point can create sidebands around a detected beat note. Nevertheless, from the network stability theory we know, that it is possible to stabilize a negative resistance system by using an external real-valued resistor in series [116], as in a typical He-Ne laser. Finally, at the highest bias voltages, the device reaches a normal state.

The same regimes can be identified in the medium heat curve, where the bolometer is pumped with an LO power equivalent to increasing its temperature with a built-in heater. The transitions between the regimes are now smooth, hence this curve shape will be favored for operation in DCS. The last, red straight-line curve stands for an ordinary resistor, which is a signature of too high operating temperature of the mixer. All these effects have been studied in more details in references [116], [117].

In conclusion, from the experimental point of view it is imperative to avoid operating the HEB in unpumped (cold) conditions promoting parasitic oscillations. On the other hand, we must ensure that the optical pumping without an absorber is not too strong, which would suppress superconductivity too much. Otherwise, the power of the detected signal would increase with decreasing power of the source, which is a clear indication of too much heat in the sensing element. In spectroscopic measurements, it translates to transmission values greater than 100%, which are obvious artifacts. Regarding the choice of the bias voltage, after ensuring the smooth sigmoidal shape of the I-V curve, one needs to find a

possibly linear region between the superconducting state, and normal state. For reference, the optimal region is surrounded by a rectangle, as in Figure 5.2b.

5.3. Extraction of intermode beat note in mid-infrared QCL combs

An idea that we implemented in our laboratory to extract RF signals from the mid-infrared QCL combs mounted on an NS-submount is shown schematically in Figure 5.3. We proposed to cut a short piece of a semi-rigid coaxial cable, and expose its center conductor over a millimeter. Next, the coax should be pulled through a small hole matched to the outer diameter of the shield, followed by soldering to a small piece of an FR4-based printed circuit board laminate. The center pin needs to be terminated with a possibly short spring-loaded contact, whereas the ground pins (use at least two) need to be slightly longer to match the difference in height between the metal submount, and the elevated gold-plated pad on an isolating substrate. This pad is wire-bonded to the chip, and this is where the center pin will contact. Obviously, the distance between the ground and center pins is critical, otherwise severe microwave reflections will appear. It should be noted, however, that this arrangement is far from optimality in terms of RF performance.

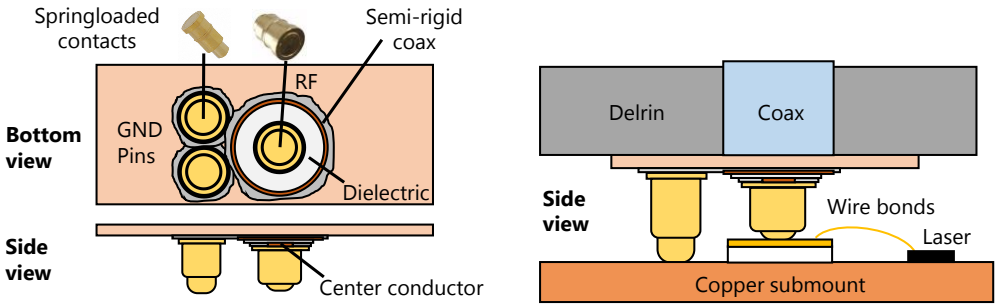

Figure 5.3. Drawing of a low-cost microwave probe for extracting the intermode beat note from a QCL.

This compact configuration enabled to house the laser in a small-volume leak-proof housing, which we pumped down to low pressure due to its operation far below the dew point. Its physical implementation is shown in Figure 5.4, where the QCL is marked with a yellow arrow.

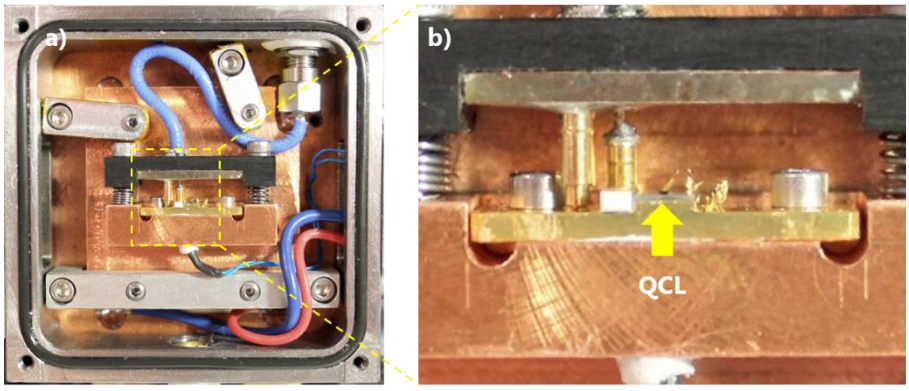

Figure 5.4. Physical implementation of the low-cost microwave probe compatible with the NS-submount QCL. (a) Overview of the laser housing with the probe. (b) Zoom into the probe.

Bibliography

[1] D. W. van der Weide and F. Keilmann, "Coherent periodically pulsed radiation spectrometer," US5748309A, 05-May-1998.

[2] S. A. Diddams *et al.*, "Direct Link between Microwave and Optical Frequencies with a 300 THz Femtosecond Laser Comb," *Phys. Rev. Lett.*, vol. 84, no. 22, pp. 5102–5105, May 2000.

[3] L. F. Tiemeijer, P. I. Kuindersma, P. J. A. Thijs, and G. L. J. Rikken, "Passive FM locking in InGaAsP semiconductor lasers," *IEEE J. Quantum Electron.*, vol. 25, no. 6, pp. 1385–1392, Jun. 1989.

[4] A. Hugi, "Single-mode and comb operation of broadband quantum cascade lasers," 2013.

[5] J. B. Khurgin, Y. Dikmelik, A. Hugi, and J. Faist, "Coherent frequency combs produced by self frequency modulation in quantum cascade lasers," *Appl. Phys. Lett.*, vol. 104, no. 8, p. 081118, 2014.

[6] C. Rullière, Ed., *Femtosecond laser pulses: principles and experiments*, 2nd ed. New York: Springer, 2005.

[7] S. L. Dexheimer, *Terahertz Spectroscopy: Principles and Applications*. CRC Press, 2007.

[8] M. Currie *et al.*, "Performance Enhancement of a GaAs Detector with a Vertical Field and an Embedded Thin Low-Temperature Grown Layer," *Sensors*, vol. 13, no. 2, pp. 2475–2483, Feb. 2013.

[9] K.-E. Peiponen, J. A. Zeitler, and M. Kuwata-Gonokami, Eds., *Terahertz spectroscopy and imaging*. Berlin ; New York: Springer, 2013.

[10] M. Tani, S. Matsuura, K. Sakai, and S. Nakashima, "Emission characteristics of photoconductive antennas based on low-temperature-grown GaAs and semi-insulating GaAs," *Appl Opt*, vol. 36, no. 30, pp. 7853–7859, Oct. 1997.

[11] S.-H. Yang and M. Jarrahi, "Frequency-tunable continuous-wave terahertz sources based on GaAs plasmonic photomixers," *Appl. Phys. Lett.*, vol. 107, no. 13, p. 131111, Sep. 2015.

[12] M. Naftaly, *Terahertz Metrology*. Artech House, 2015.

[13] L. Zhang, H. Zhong, C. Deng, C. Zhang, and Y. Zhao, "Polarization sensitive terahertz time-domain spectroscopy for birefringent materials," *Appl. Phys. Lett.*, vol. 94, no. 21, p. 211106, 2009.

[14] P. A. Elzinga, F. E. Lytle, Y. Jian, G. B. King, and N. M. Laurendeau, "Pump/Probe Spectroscopy by Asynchronous Optical Sampling," *Appl. Spectrosc.*, vol. 41, no. 1, pp. 2–4, Jan. 1987.

[15] T. Yasui, E. Saneyoshi, and T. Araki, "Asynchronous optical sampling terahertz time-domain spectroscopy for ultrahigh spectral resolution and rapid data acquisition," *Appl. Phys. Lett.*, vol. 87, no. 6, p. 061101, Aug. 2005.

[16] L. A. Sterczewski, M. P. Grzelczak, and E. F. Plinski, "Terahertz antenna electronic chopper," *Rev. Sci. Instrum.*, vol. 87, no. 1, p. 014702, Jan. 2016.

[17] L. A. Sterczewski, M. P. Grzelczak, K. Nowak, and E. F. Plinski, "Cast terahertz lenses made of caramelized sucrose," *Opt. Eng.*, vol. 55, no. 9, p. 090505, Sep. 2016.

[18] S. M. Teo, B. K. Ofori-Okai, C. A. Werley, and K. A. Nelson, "Invited Article: Single-shot THz detection techniques optimized for multidimensional THz spectroscopy," *Rev. Sci. Instrum.*, vol. 86, no. 5, p. 051301, May 2015.

[19] A. Rehn, D. Jahn, J. C. Balzer, and M. Koch, "Periodic sampling errors in terahertz time-domain measurements," *Opt. Express*, vol. 25, no. 6, p. 6712, Mar. 2017.

[20] D. W. Allan, "Statistics of atomic frequency standards," *Proc. IEEE*, vol. 54, no. 2, pp. 221–230, 1966.

[21] P. O. Werle, R. Mücke, and F. Slemr, "The limits of signal averaging in atmospheric trace-gas monitoring by tunable diode-laser absorption spectroscopy (TDLAS)," *Appl. Phys. B*, vol. 57, no. 2, pp. 131–139, 1993.

[22] P. P. Bohn, "The relationship between phase stability and frequency stability and a method of converting between them," 1971.

[23] J. Westberg, *Faraday modulation spectroscopy theoretical description and experimental realization for detection of nitric oxide*. Umeå: Department of Physics, Umeå University, 2013.

[24] Z. Bouguila, B. Fischer, A. Moukadem, and A. Dieterlen, “Time-frequency representation of terahertz time domain spectroscopy signals based on Stockwell transform,” 2014, pp. 1–2.
[25] Y.-C. Kim, K.-H. Jin, J.-C. Ye, J.-W. Ahn, and D.-S. Yee, “Wavelet Power Spectrum Estimation for High-resolution Terahertz Time-domain Spectroscopy,” *J. Opt. Soc. Korea*, vol. 15, no. 1, pp. 103–108, Mar. 2011.
[26] “FFTW Home Page.” [Online]. Available: http://www.fftw.org/. [Accessed: 02-Jan-2018].
[27] H. Gudbjartsson and S. Patz, “The Rician Distribution of Noisy MRI Data,” *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.*, vol. 34, no. 6, pp. 910–914, Dec. 1995.
[28] M. Naftaly and R. Dudley, “Methodologies for determining the dynamic ranges and signal-to-noise ratios of terahertz time-domain spectrometers,” *Opt. Lett.*, vol. 34, no. 8, p. 1213, Apr. 2009.
[29] H.-B. Liu and X.-C. Zhang, “Dehydration kinetics of D-glucose monohydrate studied using THz time-domain spectroscopy,” *Chem. Phys. Lett.*, vol. 429, no. 1–3, pp. 229–233, Sep. 2006.
[30] W. H. Press, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.
[31] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.
[32] R. G. Lyons, *Understanding Digital Signal Processing*. Pearson Education, 2010.
[33] H. Namkung, J. Kim, H. Chung, and M. A. Arnold, “Impact of Pellet Thickness on Quantitative Terahertz Spectroscopy of Solid Samples in a Polyethylene Matrix,” *Anal. Chem.*, vol. 85, no. 7, pp. 3674–3681, Apr. 2013.
[34] A. Heidari, M. Neshat, D. Saeedkia, and S. Safavi-Naeini, “Signal recovery in pulsed terahertz integrated circuits,” *Prog. Electromagn. Res.*, vol. 107, pp. 269–292, 2010.
[35] M. van Exter, C. Fattinger, and D. Grischkowsky, “Terahertz time-domain spectroscopy of water vapor,” *Opt. Lett.*, vol. 14, no. 20, p. 1128, Oct. 1989.
[36] D. M. Slocum, E. J. Slingerland, R. H. Giles, and T. M. Goyette, “Atmospheric absorption of terahertz radiation and water vapor continuum effects,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 127, pp. 49–63, Sep. 2013.
[37] Y.-S. Jin, G.-J. Kim, and S.-G. Jeon, “Terahertz dielectric properties of polymers,” *J. Korean Phys. Soc.*, vol. 49, no. 2, pp. 513–517, 2006.
[38] S. C. Gad, *Drug safety evaluation*. New York: J. Wiley, 2002.
[39] Z. Csakurda-Harmathy and I. K. Thege, “Transformation of chloramphenicol palmitate from therapeutically inactive polymorph A to active polymorph B,” *J. Therm. Anal.*, vol. 50, no. 5–6, pp. 867–871, Dec. 1997.
[40] “Guidance for Industry ANDAs: Pharmaceutical Solid Polymorphism Chemistry, Manufacturing, and Controls Information.” U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER).
[41] D. Giron, M. Draghi, C. Goldbronn, S. Pfeffer, and P. Piechon, “Study of the polymorphic behaviour of some local anesthetic drugs,” *J. Therm. Anal.*, vol. 49, no. 2, pp. 913–927, Aug. 1997.
[42] J. A. Zeitler *et al.*, “Temperature dependent terahertz pulsed spectroscopy of carbamazepine,” *Thermochim. Acta*, vol. 436, no. 1–2, pp. 71–77, Oct. 2005.
[43] M. Haaser, K. C. Gordon, C. J. Strachan, and T. Rades, “Terahertz pulsed imaging as an advanced characterisation tool for film coatings—A review,” *Int. J. Pharm.*, vol. 457, no. 2, pp. 510–520, Dec. 2013.
[44] C. J. Strachan *et al.*, “Using terahertz pulsed spectroscopy to quantify pharmaceutical polymorphism and crystallinity,” *J. Pharm. Sci.*, vol. 94, no. 4, pp. 837–846, Apr. 2005.
[45] J. A. Zeitler *et al.*, “Characterization of temperature-induced phase transitions in five polymorphic forms of sulfathiazole by terahertz pulsed spectroscopy and differential scanning calorimetry,” *J. Pharm. Sci.*, vol. 95, no. 11, pp. 2486–2498, Nov. 2006.
[46] J. Zeitler *et al.*, “Drug hydrate systems and dehydration processes studied by terahertz pulsed spectroscopy,” *Int. J. Pharm.*, vol. 334, no. 1–2, pp. 78–84, Apr. 2007.

[47] D. F. Pensenstadler, G. L. Carlson, W. G. Fateley, and F. F. Bentley, “The use of adamantane as a pelleting material for far infrared spectroscopy of solids,” *Spectrochim. Acta Part Mol. Spectrosc.*, vol. 28, no. 1, pp. 183–185, 1972.
[48] J. E. Bertie and S. M. Jacobs, “Far infrared absorption and rotational vibrations of the guest molecules in structure I clathrate hydrates between 4.3 and 100 K,” *Can. J. Chem.*, vol. 55, no. 10, pp. 1777–1785, 1977.
[49] H. Yoshinaga and R. A. Oetjen, “Sampling Technique for Transmission Measurements in the Far Infrared Region*,” *JOSA*, vol. 45, no. 12, pp. 1085–1085, Dec. 1955.
[50] O. Pirali, V. Boudon, J. Oomens, and M. Vervloet, “Rotationally resolved infrared spectroscopy of adamantane,” *J. Chem. Phys.*, vol. 136, no. 2, p. 024310, Jan. 2012.
[51] J. A. Zeitler, P. F. Taday, K. C. Gordon, M. Pepper, and T. Rades, “Solid-State Transition Mechanism in Carbamazepine Polymorphs by Time-Resolved Terahertz Spectroscopy,” *ChemPhysChem*, vol. 8, no. 13, pp. 1924–1927, Sep. 2007.
[52] “TPX Polymethylpentene (PMP) Transparent Polymer Properties.” [Online]. Available: https://www.mitsuichemicals.com/tpx_phy.htm. [Accessed: 30-Nov-2017].
[53] P. D. Cunningham *et al.*, “Broadband terahertz characterization of the refractive index and absorption of some important polymeric and organic electro-optic materials,” *J. Appl. Phys.*, vol. 109, no. 4, pp. 043505–043505, 2011.
[54] “Advanced COC Polymers | TOPAS.” [Online]. Available: https://topas.com/sites/default/files/Products_E_170406.pdf. [Accessed: 30-Nov-2017].
[55] R. Rungsawang, Y. Ueno, I. Tomita, and K. Ajito, “Terahertz notch filter using intermolecular hydrogen bonds in a sucrose crystal,” *Opt. Express*, vol. 14, no. 12, pp. 5765–5772, 2006.
[56] J. Kröll, J. Darmo, and K. Unterrainer, “Terahertz optical activity of sucrose single-crystals,” *Vib. Spectrosc.*, vol. 43, no. 2, pp. 324–329, Mar. 2007.
[57] R. Singh, D. K. George, and A. Markelz, “Anisotropy and birefringence in molecular crystals,” in *2011 International Conference on Infrared, Millimeter, and Terahertz Waves*, 2011, pp. 1–2.
[58] G. Acbas, E. Snell, and A. G. Markelz, “Orientation sensitive Terahertz resonances observed in protein crystals,” in *2012 37th International Conference on Infrared, Millimeter, and Terahertz Waves*, 2012, pp. 1–3.
[59] P. U. Jepsen and S. J. Clark, “Precise ab-initio prediction of terahertz vibrational modes in crystalline systems,” *Chem. Phys. Lett.*, vol. 442, no. 4–6, pp. 275–280, Jul. 2007.
[60] G. M. Day, J. A. Zeitler, W. Jones, T. Rades, and P. F. Taday, “Understanding the influence of polymorphism on phonon spectra: lattice dynamics calculations and terahertz spectroscopy of carbamazepine,” *J. Phys. Chem. B*, vol. 110, no. 1, pp. 447–456, 2006.
[61] K. Nowak *et al.*, “Selected aspects of terahertz spectroscopy in pharmaceutical sciences,” *Acta Pol. Pharm.*, vol. 72, no. 5, p. 851—866, 2015.
[62] L. A. Sterczewski, M. P. Grzelczak, and E. F. Plinski, “Heating system of pellet samples integrated with terahertz spectrometer,” *Rev. Sci. Instrum.*, vol. 87, no. 1, p. 013106, Jan. 2016.
[63] L. A. Sterczewski, M. P. Grzelczak, K. Nowak, B. Szlachetko, and E. F. Plinski, “Bayesian separation algorithm of THz spectral sources applied to D-glucose monohydrate dehydration kinetics,” *Chem. Phys. Lett.*, vol. 644, pp. 45–50, Jan. 2016.
[64] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, “Bayesian separation of spectral sources under non-negativity and full additivity constraints,” *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, 2009.
[65] M. A. Ponschke and J. E. House, “Kinetic studies on the loss of water from α-d-glucose monohydrate,” *Carbohydr. Res.*, vol. 346, no. 14, pp. 2285–2289, Jul. 2011.
[66] G. Villares, A. Hugi, S. Blaser, and J. Faist, “Dual-comb spectroscopy based on quantum-cascade-laser frequency combs,” *Nat. Commun.*, vol. 5, p. 5192, Oct. 2014.
[67] P. Jouy *et al.*, “Dual comb operation of $\lambda \sim 8.2\ \mu$m quantum cascade laser frequency comb with 1 W optical power,” *Appl. Phys. Lett.*, vol. 111, no. 14, p. 141102, Oct. 2017.
[68] Y. Bidaux, F. Kapsalidis, P. Jouy, M. Beck, and J. Faist, “Coupled-waveguides for dispersion compensation in semiconductor lasers,” *ArXiv171109116 Phys.*, Nov. 2017.
[69] H. Hundertmark, *Erbium fiber lasers for a frequency comb at 1560 nm*. Cuvillier Verlag, 2006.

[70] "Encyclopedia of Laser Physics and Technology - quantum cascade lasers, QCL, mid-infrared, terahertz, semiconductor lasers." [Online]. Available: https://www.rp-photonics.com/quantum_cascade_lasers.html. [Accessed: 08-Apr-2018].
[71] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, "Quantum Cascade Laser," *Science*, vol. 264, no. 5158, pp. 553–556, Apr. 1994.
[72] A. Hugi, G. Villares, S. Blaser, H. C. Liu, and J. Faist, "Mid-infrared frequency comb based on a quantum cascade laser," *Nature*, vol. 492, no. 7428, pp. 229–233, Dec. 2012.
[73] E. Baumann, F. R. Giorgetta, W. C. Swann, A. M. Zolot, I. Coddington, and N. R. Newbury, "Spectroscopy of the methane ν 3 band with an accurate midinfrared coherent dual-comb spectrometer," *Phys. Rev. A*, vol. 84, no. 6, p. 062513, 2011.
[74] S. Schiller, "Spectrometry with frequency combs," *Opt. Lett.*, vol. 27, no. 9, pp. 766–768, 2002.
[75] I. Coddington, N. Newbury, and W. Swann, "Dual-comb spectroscopy," *Optica*, vol. 3, no. 4, p. 414, Apr. 2016.
[76] D. Burghoff, Y. Yang, and Q. Hu, "Computational multiheterodyne spectroscopy," *Sci. Adv.*, vol. 2, no. 11, Nov. 2016.
[77] L. A. Sterczewski *et al.*, "Multiheterodyne spectroscopy using interband cascade lasers," *Opt. Eng.*, vol. 75, no. 1, 2017.
[78] K. Krzempek, G. Dudzik, A. Hudzikowski, A. Gluszek, and K. Abramski, "Highly-efficient fully-fiberized mid-infrared differential frequency generation source and its application to laser spectroscopy," *Opto-Electron. Rev.*, vol. 25, no. 4, pp. 269–273, 2017.
[79] Y.-D. Hsieh *et al.*, "Spectrally interleaved, comb-mode-resolved spectroscopy using swept dual terahertz combs," *Sci. Rep.*, vol. 4, Jan. 2014.
[80] M. R. St-Jean *et al.*, "Injection locking of mid-infrared quantum cascade laser at 14 GHz, by direct microwave modulation," *Laser Photonics Rev.*, vol. 8, no. 3, pp. 443–449, May 2014.
[81] J. Faist, *Quantum Cascade Lasers*. OUP Oxford, 2013.
[82] L. A. Sterczewski, J. Westberg, and G. Wysocki, "Molecular dispersion spectroscopy based on Fabry–Perot quantum cascade lasers," *Opt. Lett.*, vol. 42, no. 2, p. 243, Jan. 2017.
[83] J. Westberg, L. A. Sterczewski, and G. Wysocki, "Mid-infrared multiheterodyne spectroscopy with phase-locked quantum cascade lasers," *Appl. Phys. Lett.*, vol. 110, no. 14, p. 141108, Apr. 2017.
[84] A. Hangauer, J. Westberg, E. Zhang, and G. Wysocki, "Wavelength modulated multiheterodyne spectroscopy using Fabry-Pérot quantum cascade lasers," *Opt. Express*, vol. 24, no. 22, pp. 25298–25307, Oct. 2016.
[85] R. Q. Yang, "Infrared laser based on intersubband transitions in quantum wells," *Superlattices Microstruct.*, vol. 17, no. 1, pp. 77–83, 1995.
[86] I. Vurgaftman *et al.*, "Rebalancing of internally generated carriers for mid-infrared interband cascade lasers with very low power consumption," *Nat. Commun.*, vol. 2, p. 585, 2011.
[87] C. R. Webster *et al.*, "Mars methane detection and variability at Gale crater," *Science*, vol. 347, no. 6220, pp. 415–417, 2015.
[88] D. Burghoff *et al.*, "Terahertz laser frequency combs," *Nat Photon*, vol. 8, no. 6, pp. 462–467, Jun. 2014.
[89] R. Köhler *et al.*, "Terahertz semiconductor-heterostructure laser," *Nature*, vol. 417, no. 6885, pp. 156–159, May 2002.
[90] Y. Yang, D. Burghoff, D. J. Hayton, J.-R. Gao, J. L. Reno, and Q. Hu, "Terahertz multiheterodyne spectroscopy using laser frequency combs," *Optica*, vol. 3, no. 5, p. 499, May 2016.
[91] J. Westberg *et al.*, "Terahertz dual comb spectroscopy using quantum cascade laser frequency combs," in *CLEO: 2018*, 2018.
[92] J. Westberg *et al.*, "Terahertz multiheterodyne spectroscopy of molecular samples with quantum cascade laser frequency combs," in *2017 14th International Conference on Intersubband Transitions in Quantum Wells (ITQW)*, Singapore, 2017.
[93] Y. Yang, D. Burghoff, J. Reno, and Q. Hu, "Achieving comb formation over the entire lasing range of quantum cascade lasers," *Opt. Lett.*, vol. 42, no. 19, p. 3888, Oct. 2017.
[94] M. Rösch, G. Scalari, M. Beck, and J. Faist, "Octave-spanning semiconductor laser," *Nat. Photonics*, vol. 9, no. 1, p. 42, Jan. 2015.

[95] G. Villares *et al.*, “Dispersion engineering of quantum cascade laser frequency combs,” *Optica*, vol. 3, no. 3, p. 252, Mar. 2016.
[96] J. Westberg *et al.*, “Quantum cascade laser-based dual-comb spectroscopy in the mid-infrared,” in *CLEO: 2018*.
[97] M. Bagheri *et al.*, “Passively mode-locked interband cascade optical frequency combs,” *Sci. Rep.*, vol. 8, no. 1, p. 3322, Feb. 2018.
[98] L. A. Sterczewski *et al.*, “Dual-comb spectroscopy with passively mode-locked interband cascade laser frequency combs,” in *CLEO: 2018*, 2018.
[99] M. Rösch, G. Scalari, G. Villares, L. Bosco, M. Beck, and J. Faist, “On-chip, self-detected terahertz dual-comb source,” *Appl. Phys. Lett.*, vol. 108, no. 17, p. 171104, Apr. 2016.
[100] S. R. Al-Araji, Z. M. Hussain, and M. A. Al-Qutayri, *Digital Phase Lock Loops*. Springer, 2006.
[101] “Semiconductor surface passivation,” *Mater. Sci. Rep.*, vol. 3, no. 3–4, pp. 139–216, Jan. 1988.
[102] L. A. Sterczewski, J. Westberg, and G. Wysocki, “Fast coherent averaging for multiheterodyne spectroscopy,” *Manuscr.*, 2018.
[103] L. A. Sterczewski, L. A. Sterczewski, J. Westberg, L. Patrick, and G. Wysocki, “Computational Adaptive Sampling for Multiheterodyne Spectroscopy,” in *Conference on Lasers and Electro-Optics (2017), paper JW2A.80*, 2017, p. JW2A.80.
[104] C. Li, “Polarization Theory of Nonlinear Medium,” in *Nonlinear Optics*, Singapore: Springer Singapore, 2017, pp. 23–50.
[105] P. Tichavsky and P. Handel, “Recursive estimation of linearly or harmonically modulated frequencies of multiple cisoids in noise,” in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997, vol. 3, pp. 1925–1928 vol.3.
[106] J. Van Zaen, “Efficient schemes for adaptive frequency tracking and their relevance for EEG and ECG,” 2012.
[107] L. Tan and J. Jiang, “Novel adaptive IIR filter for frequency estimation and tracking [DSP Tips Tricks],” *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 186–189, Nov. 2009.
[108] M. Piovoso and P. A. Laplante, “Kalman Filter Recipes for Real-time Image Processing,” *Real-Time Imaging*, vol. 9, no. 6, pp. 433–439, Dec. 2003.
[109] P. Tichavsky and P. Handel, “Recursive estimation of frequencies and frequency rates of multiple cisoids in noise,” *Signal Process.*, vol. 58, no. 2, pp. 117–129, Apr. 1997.
[110] R. M. Aarts, “Low-complexity tracking and estimation of frequency and amplitude of sinusoids,” *Digit. Signal Process.*, vol. 14, no. 4, pp. 372–378, Jul. 2004.
[111] N. R. Newbury, I. Coddington, and W. Swann, “Sensitivity of coherent dual-comb spectroscopy,” *Opt. Express*, vol. 18, no. 8, pp. 7929–7945, 2010.
[112] R. Lyons, “Reducing FFT Scalloping Loss Errors Without Multiplication [DSP Tips and Tricks],” *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 112–116, Mar. 2011.
[113] M. O. Reese, “Superconducting hot electron bolometers for terahertz sensing,” Citeseer, 2006.
[114] A. Shurakov, Y. Lobanov, and G. Goltsman, “Superconducting hot-electron bolometer: from the discovery of hot-electron phenomena to practical applications,” *Supercond. Sci. Technol.*, vol. 29, no. 2, p. 023001, Feb. 2016.
[115] T. M. Klapwijk and A. V. Semenov, *Engineering Physics of Superconducting Hot-Electron Bolometer Mixers*, vol. PP. 2017.
[116] Y. Zhuang and K. S. Yngvesson, “Negative resistance effects in NbN HEB devices,” in *Proc. 12th Int. Symp. Space Terahertz Technol.*, 2001.
[117] Y. Zhuang, D. Gu, and S. Yngvesson, “Bistability in NbN HEB mixer devices,” in *Proc. 14th Int. Symp. Space Terahertz Technol.*, 2003, pp. 289–298.