Johannes Gutenberg-Universität
Mainz

Fachbereich Physik, Mathematik und
Informatik

Doctoral thesis

# Towards rapid, sensitive and portable optical polarimetry

*Author:*
Jim C. Visschers

*Supervisor:*
prof. dr. Dmitry Budker

August 19, 2021

# Contents

*Gute Nacht Freunde*
*Es wird Zeit für mich zu gehen*
*Was ich noch zu sagen hätte*
*Dauert eine Zigarette*
*Und ein letztes Glas im Stehen*
-Reinhard Mey

# 1 Declaration of authorship

Herewith I,

**Jim C. Visschers**

declare that the present thesis, entitled:

"***Towards sensitive and portable optical polarimetry***"

and the works presented within it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the Johannes Gutenberg University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at the Johannes Gutenberg University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Mainz, **August 19, 2021**

Signature:

## 2 Published manuscripts within this thesis

The following papers have been published in peer-reviewed journals and, in combination with the article in the process of publication, comprise the body of this thesis:

**Visschers JC**, Tretiak O, Budker D, Bougas L. Continuous-wave cavity ring-down polarimetry. The Journal of chemical physics. 2020 Apr 30;152(16):164202.

**Visschers JC**, Wilson E, Conneely T, Mudrov A, Bougas L. Rapid parameter determination of discrete damped sinusoidal oscillations. Optics Express. 2021 Mar 1;29(5):6863-78.

In addition to the papers already published in peer review journals, the following paper has, at the time of writing, completed the peer-review process and is formally accepted for publication in Machine Learning: Science and Technology. The most recent publicly available version of the manuscript can be found as:

**Visschers JC**, Budker D, Bougas L. Rapid parameter estimation of discrete decaying signals using autoencoder networks. arXiv preprint arXiv:2103.08663. 2021 Mar 10.

The previously mentioned papers are directly incorporated into chapters 4, 5, 6 of this thesis.

Finally, the work outlined in chapter 7 is not, and will not be, part of a work intended for peer-reviewed publication. However, it is available on arXiv and the accompanying code and designs are freely available in a GitHub repository: `https://github.com/JCVis/GrowFacility`.

# 3 Introduction

## 3.1 Optical polarimetry

Optical polarimetry is one of the oldest tools in the field of analytical chemistry. The technique measures polarization dependant rotation (birefringence) and absorption (dichroism) within a medium. Over the last centuries, this technique has been essential in scientific discoveries and investigations. These include advances in identification of the structure of complex molecules [1, 2] and optical magnetometry [3]. In industrial applications (e.g. pharmaceutical, chemical and agricultural) optical polarimetry is routinely applied in process verification and quality control[4]. The birefringent and dichroic effects that optical polarimetry measures can arise from various sources. However, for the scope of this thesis, we will focus mainly on circular birefringent effects that arise from natural optical activity (e.g. chirality) and the Faraday effect. These effects are highly substance-specific and can be used to:

1. Identify substances within the medium;
2. Measure the concentration and purity of optically active substances within a medium;
3. Observe dynamic processes of optically active molecules within a medium.

An advantage of optical polarimetry compared to other analytical chemistry tools is the non-invasive nature of the technique; samples under investigation are simply illuminated by light and do not need to undergo any chemical reactions. Even the light probing the sample is, in most cases, off resonance to any atomic or molecular transitions. Typically, a polarimeter (Fig. 1) uses monochromatic light that is passed through a polarizing optic (polarizer) and then sent through the optically active medium rotating the polarization of the light. The light is then sent through another polarizing optic (analyser) after which the intensity is measured. For the case shown in Fig. 1 the analysing optic is a polarizing beamsplitter and both polarization states are measured and compared. The difference in angle between the polarizing and analysing optic corresponds to the total rotation from the medium under investigation. The angle of

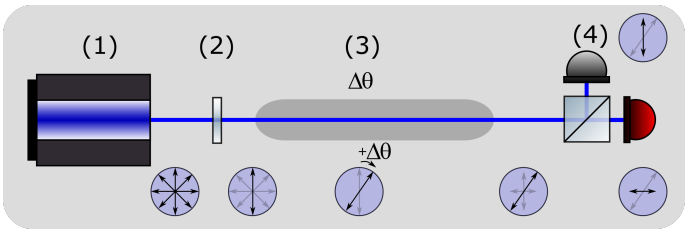


Figure 1: Schematic of a single-pass polarimeter consisting of (1) a light source emitting unpolarized light (in practice, lasers tend to emit polarized light), (2) polarizer, (3) optically active medium rotating the polarization $\Delta\phi$, (4) analysing optic and differential measurement of the reflected and transmitted polarizations. The blue circles represent the changing polarization orientation though the polarimeter.

rotation is given by:

$$\theta = \frac{\Delta n \, L \, \pi}{\lambda}, \tag{1}$$

where $\Delta n$ is the difference in refractive index between left- and right circularly polarized light, $L$ is the path length through the optically active medium and $\lambda$ is the wavelength of the light probing the medium. For the Faraday effect, where the rotation is not only dependant on the intrinsic property of the medium itself, but also on an external magnetic field, the optical rotation is given by:

$$\theta = \mathcal{V}(\lambda) \; \vec{B} \, L, \tag{2}$$

where $\mathcal{V}(\lambda)$ is the Verdet constant of the substance under investigation, $\vec{B}$ is the applied magnetic field along the direction of the light and $L$ is the interaction length. The Verdet constant is inversely proportional to the wavelength of the probing light ($\lambda$) just like the optical rotation in Eq. 1. A key difference between optical activity and the Faraday effect is that the resulting rotation in the former case is not dependant on the direction of the propagating light, whereas for the latter, the orientation of the magnetic field with respect to the propagation direction of the light determines the sign (direction) of the optical rotation. As a result, when light that passes through a medium affected by the Faraday effect, is reflected 180° and passes through the the medium again, the two rotations add up. In the case of optical activity, where the direction of the rotation is independant of the propagation direction of the light the

two subsequent rotations cancel each other out[1].

## 3.2 A brief history of optical polarimetry

### 3.2.1 Discovery of optical polarization and rotation

The orientation and possibility of rotation of linearly polarized light was discovered by the French physicist François Jean Dominique Arago in the early 1800s [5]. In the years after, optical rotation was observed in more and more substances such as crystalline structures [6], liquids [7], and organic compounds [8]. In 1845, Michael Faraday discovered the Faraday effect after extensive but, until then, unsuccessful attempts to find a relation between electricity and light. He placed a piece of glass in a magnetic field and shone a ray of linearly polarized light through and observed the subsequent rotation of its polarization [9].

### 3.2.2 Tartaric acid

As polarimeters gained wide use, a puzzling observation was made that was explained in 1849 by Louis Pasteur. Namely, tartaric acid derived from living organisms rotates the polarization plane of linearly polarized light, whereas, chemically synthesized tartaric acid does not affect the polarization of light whatsoever. This fact was puzzling because the chemical properties of both types of tartaric acid were – and are still – the same. Pasteur showed, through a thorough investigation of the salt-crystalline form of the acid (tartrate) that there were two distinct, asymmetric, forms of the tartrate present in the chemically synthesized tartaric acid. When separated, a solution of one form of the tartaric acid would rotate light clockwise, just like the organically derived tartaric acid, while a solution of the other form of the tartaric acid would rotate light counterclockwise. Pasteur came to the conclusion that tartaric acid must be an asymmetric molecule with two possible arrangements, much like left- and right hands and that biological tartaric acid consisted solely of one type [10]. This idea was far ahead of its time, as 9 years later, in 1859 chemist Archibald Couper proposed one of the first theories on molecular structure and chemical bonding [11].

[1] This fact will become very important when we consider multi-pass systems and cavity enhanced polarimetry schemes.

### 3.2.3 Wilhelmy and cane sugar

Later, in 1850, Ludwig Ferdinand Wilhelmy used polarimetry to investigate the chemical process (hydrolysis) where a water solution of cane sugar, consisting mainly of sucrose, is converted to an equal mixture of fructose (levulose) and glucose (dextrose) using an acid-catalytic reaction [12]. Cane sugar, rotates linear polarized light clockwise with a specific rotation of $\alpha_D^{20} = +66.47^\circ\,\mathrm{dm}^{-1}\,\mathrm{ml}\,\mathrm{g}^{-1}$, as their names suggest, fructose and glucose rotate, counter-clockwise and clockwise respectively. However, fructose has a specific rotation of $\alpha_D^{20} = -92^\circ\,\mathrm{dm}^{-1}\,\mathrm{ml}\,\mathrm{g}^{-1}$, whereas glucose only has a specific rotation of $\alpha_D^{20} = +52.7^\circ\,\mathrm{dm}^{-1}\,\mathrm{ml}\,\mathrm{g}^{-1}$. As the sucrose is converted to a mixture of fructose and glucose the overall optical activity in the solution goes from clockwise to counter-clockwise. Wilhelmy was able to interpret his results by writing down the differential equations describing the relevant chemical reactions. Furthermore, he studied the influence of different acids, acid concentrations, varying temperatures and concentrations of sugar, on the reaction and optical activities of the different sugars.

### 3.2.4 Industrial standard and possibilities for improvement

Although contemporary, industrial optical polarimetric instruments obtain a rotation angle precision of $<0.0001^o$, or approximately $1\,\mu$rad [13], the technique that these instruments employ does not differ fundamentally from the experimental technique conducted by Pasteur and Whilhelmy (described in Sec. 3.1). Instruments rely on single-pass measurements trough a measurement cell of 10 cm (which is the industry standard). These measurements are used to quantify the specific rotation of molecules, analyse enantiomeric ratios and asses the purity of solutions.

Development of more sensitive polarimetric methods will allow these measurements to be done (1) faster, (2) with minimal concentrations and (3) open up new fields of investigation previously unattainable hopefully driving down the development costs of drugs and other novel chemical compounds. Observing Eq.1 we see that $\Delta n$, the refractive index difference between left- and right circularly polarized light, is dependant on the concentration of the optically active substance under investigation. Reducing the wavelength $\lambda$ of the probing light increases the signal. However as the wavelength de-

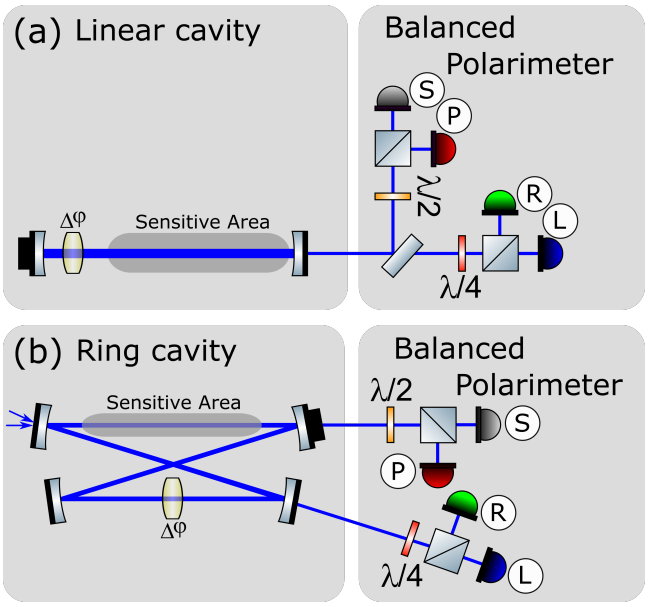


Figure 2: Two distinctly different cavity setup schemes for the use of cavity enhanced polarimetry. In both cases a balanced polarimeter is used to analyse the polarization state of the light leaking from the cavity. A half-waveplate ($\lambda/2$) is used to investigate the linearly polarized states (S and P) and a quarter-waveplate ($\lambda/4$) is used to decompose the leaking light into right- (R) and left (L) handed polarized light. (a) Illustrates a two-mirror linear cavity that can be used to study circular birefringent effects such as the Faraday effect. (b) Four-mirror ring cavity setup that allows the investigation of reciprocal circular birefringence, such as optical activity.

creases, losses from Raleigh scattering increase much faster ($\sim \lambda^{-4}$) resulting in the fact that minimization of the wavelength does not systematically enhance the sensitivity of an measurement. As the interaction length between the polarized light and the optically active medium increases, the polarization rotation increases. However, so does the size of the instrument and the required amount of optically active substance to fill the instrument.

### 3.3 Cavity enhanced polarimeters

State-of-the-art polarimeters use optical cavities to increase the interaction length between the polarized light and the optically active substance [14–28]. Using high quality mirrors, optical cavities can achieve the path-length enhancements in excess of $10^5$, enabling record sensitivities for absorption and birefringence measurements.

Importantly, instruments using optical cavities can easily be designed compactly, robustly and even in a portable package. Cavity enhanced spectroscopy is, by itself, a well-established research field [29, 30]. Implementing optical cavities has its drawbacks as well: either the majority of light is lost by coupling into the cavity or a non-trivial locking mechanism needs to be introduced to allow light from its source to enter into the optical cavity. However, there exist several cavity-based polarimetric designs for the measurement of linear and circular, birefringence and dichroism. Although there are fundamental symmetry considerations for which type of cavity one employs for the measurement of reciprocal effects such as natural optical activity or non-reciprocal birefringent effects such as the Faraday effect, for measurement of non-reciprocal birefringent effects, a simple (linear) two-mirror cavity can be used, as illustrated in Fig. 2a. However, for the measurement of reciprocal birefringent effects a cavity supporting a running wave needs to be employed, typically, a four-mirror cavity as illustrated in Fig. 2b. Alternatively, reciprocal birefringent effects can be measured using a two-mirror cavity when two intracavity waveplates are introduced [15]. A more thorough exposé of the different forms of cavity enhanced polarimetery and the novel technique we demonstrate is given in Ch. 4 of this thesis.

### 3.4 Fast optical polarimetry and analysis

In Ch. 4 of this thesis we describe a novel method of polarimetry, that we named continuous-wave cavity ring-down polarimetry. In this chapter, the theory behind continuous-wave cavity ring-down polarimetry is presented alongside a demonstration of a working instrument and measurements of the Faraday effect in solid $SiO_2$ and $CeF_3$, and gaseous butane. The work is accompanied with an analysis of the fundamental sensitivity limits for other cavity ring-down polarimetry methods. The analysis reveals that the technique we demonstrate is able to achieve competitive sensitivities using cavities of much lower finesse than other methods. Lower finesse cavities require mirrors with lower reflectively and are subsequently much easier to work with compared to higher finesse optical systems. In essence, we believe that continuous-wave cavity ring-down polarimetry is the most suitable polarimetry technique available today for the development of a portable instrument. Various, more

optimal modalities for the technique, which do not necessarily require an extensive locking mechanism, are discussed in Ch. 4.

### 3.4.1 Fast signal analysis; a comparison

Another main advantage of the continuous-wave cavity ring-down polarimetry technique that we demonstrate in Ch. 4 is that it is easily capable of reaching measurement repetition rates of 50 kHz. A 20 $\mu$s temporal resolution to investigate circular birefringent effects is a great improvement for the direct observation (and control) of fast physical and chemical processes. However, control and observation of fast process in real time does not only require a fast and sensitive measurement technique, but also an equally fast method of analysis. Without the latter, dynamic processes of interest can only be observed in post-processing, too late to control the environment and optimize the process under investigation. Luckily, our consortium collaborators in the European Commission Horizon 2020 project ULTRACHIRAL designed a fast analysis method based on the polynomial Prony method which they describe in their paper [31]. The assumption of the Prony method is that a measurement sample can be expressed by a linear combination of the preceding samples in the measurement and therefore can effectively be described by an auto-regressive model. From the roots of this auto-regressive model the characteristics (decay time and frequency) of the signal (decaying oscillations) can be determined. In Ch. 5 we investigate the performance of this novel Prony method and compare it in terms of precision, accuracy and analysis speed over an extensive range of signal characteristics to a least-squares fitting algorithm that we employ for our findings in Ch. 4 and a non-iterative fast Fourier transform method. From our investigation, we conclude that the Fourier transform method and not the Prony method delivers the optimal combination of precision, accuracy and speed. Low kHz analysis rates are achievable using the Fourier transform method and affordable computational resources, which is great, but not sufficiently fast to analyse single measurements of the continuous-wave cavity ring-down polarimetry technique which, in its current state, runs with a repetition rate of 50 kHz.

### 3.4.2 Fast analysis with Neural networks

After analysing the Prony, least-squares and Fourier transform methods we thought of the opportunity of using neural networks to analyse the signals from the continuous-wave cavity ring-down polarimetry instrument. Neural networks seem like an ideal candidate for a fast analysis method because of the non-iterative nature of the technique. Furthermore, neural networks can be small, meaning that storing the neural network on an device such as a field-programmable gate array would be possible, and more importantly, are fast to use. The time-consuming work of training the neural network can be done offline using simulated signals, after which the trained network can be used to rapidly analyse measurement data. Chapter 6 outlines the work we have done developing an analysis method based on neural network autoencoders. To our delight, the autoencoder neural network approach is able to achieve a precision and accuracy similar so the Fourier transform method that we determined as the most optimal analysis method in Ch. 5 and close to the fundamental estimation limit given by the Cramér-Rao lower bound. Moreover, we demonstrate that we are able to achieve analysis rates of 75 kHz with cost-effective computational facilities. These high analysis rates are sufficiently fast to analyse single-shot measurements from the continuous-wave cavity ring-down polarimetry instrument. Furthermore, the two model signals that we demonstrate the autoencoder analysis method on are signals encountered in many fields of research and applicable in fields such as cavity enhanced spectroscopy and nuclear magnetic resonance.

## 3.5 Plant-growing facility

Finally, we present a design and realization of a plant-growing facility for the study of (chiral) plant volatile organic compound emissions. This plant-growing facility will be used in future, long-term measurement campaigns attempting to connect plant emissions to external, stress-inducing conditions. The designs and measurement and control scripts for the plant-growing facility are accessible via `https://github.com/JCVis/GrowFacility`.

# 4 Continuous-wave cavity ring-down polarimetry

The following chapter is a direct adaptation of a peer-reviewed and published work. It can be found as:

**Visschers JC**, Tretiak O, Budker D, Bougas L. Continuous-wave cavity ring-down polarimetry. The Journal of chemical physics. 2020 Apr 30;152(16):164202.

The minimal changes made to this version of the manuscript compared to the published work limit themselves to formatting and style alterations. The method itself is not one of my own ideas; the design and building of the experimental setup, data-acquisition, analysis and interpretation have been done by myself with help from my co-authors.

## 4.1 Introduction

Polarimetry is one of the oldest research tools available to characterize the optical properties of a substance, and has been essential in scientific developments, such as, for identifying a molecule's structure [2, 32] in tests of fundamental symmetries of nature [33], in optical magnetometry [3]. Polarimetry is also used in industrial applications where it is routinely applied in quality and process control in the pharmaceutical, chemical, and agricultural industries [34]. Despite their extensive use, the sensitivity limits of commercially available optical spectro-polarimeters are at the $\sim 10\,\mu$rad levels, which constrain the applicability of polarimetry in a wide range of important research and industrial applications, e.g., in chemical analysis and drug design and development where the relatively poor polarimetric sensitivities result in poor analyte concentration detection limits at the micromolar levels [2, 35], or in trace-gas detection and analysis.

Different techniques have been developed to push optical polarimetry to its limits that, in principle, are fundamentally constrained by the photon shot-noise of the light source used for the measurements, which for a $\sim 1\,$mW of visible radiation are at the $\sim 10^{-8}\,\mathrm{rad}/\sqrt{\mathrm{Hz}}$ levels. However, there exist different sources of

noise (e.g. light-intensity noise and detector electronic noise) that are usually much larger than the photon shot noise and limit polarimetric sensitivities to the $10^{-7} - 10^{-6}\,\mathrm{rad}/\sqrt{\mathrm{Hz}}$ levels. Modulation techniques are typically employed to approach shot-noise limited polarimetric sensitivities, but even then these limits preclude broad application of polarimetry, especially in trace-gas detection and analysis. The most straightforward solution towards increased spectropolarimetric sensitivities is through the enhancement of the interaction path-length of the probing radiation in the substance under investigation. This can be achieved with the use of optical mirrors to create either multipass cells or optical cavities, where in both cases spectropolarimetric signals are enhanced by the average number of passes through the medium. Multipass cells [36, 37], such as White cells [38] or Herriott cells [39], are technically easy to construct and implement, and path-length enhancements as large as $\times 500$ have been demonstrated [40, 41]. However, multipass-based techniques are still limited by laser intensity fluctuations and the probing radiation travels along a different path for each pass, and, thus, relatively large substance volumes are typically required, making, for example, multipass cells hard to implement for measurements of liquid samples. Importantly, multipass techniques are suitable for absorption measurements and, in principle, can not be employed for the measurement of natural optical activity. On the contrary, using stable optical cavities one can achieve path-length enhancements of up to $10^5$ using state-of-the-art high quality mirrors, with effective path-lengths of up to several hundred km (compare this to the 10 cm interaction path-length of a single-pass commercial polarimeter), enabling record sensitivities for measurements of absorption and birefringence. Moreover, optical cavities can be easily made compact and allow for light-medium interactions in small volumes.

There exist several cavity-based polarimetric designs for the measurement of linear and circular, birefringence and dichroism. Although there are fundamental symmetry considerations for which type of cavity one employs for the measurement of reciprocal or non-reciprocal birefringent effects [14, 15, 17], in general, stable optical cavities consist of two to four optical mirrors, and all designs can be realized using (a) continuous-wave (CW) laser light or (b) short laser pulses. In the case of using CW laser sources, state-of-the-art polarimetric sensitivities of $10^{-13}\,\mathrm{rad}\sqrt{\mathrm{Hz}}$ have been

demonstrated [21]. There exist several approaches for performing CW cavity-based polarimetric measurements, including modulation-based ones (e.g. Ref. [25]), or techniques developed in the field of frequency metrology [22, 24]. However, these CW-laser-based cavity-enhanced polarimetric techniques require complicated and extremely stable optomechanical setups together with state-of-the-art electronics to achieve this kind of sensitivity levels, precluding their translation into portable instrumentation and, therefore, for applicability in field studies.

In the case of using short laser pulses, cavity-based polarimetric techniques build upon the inherent sensitivity of cavity ring-dow n spectroscopy (CRDS) [29, 30], via the insertion of polarization-selective components into the preparation, interaction, and analysis stage of a conventional CRDS apparatus. Particularly, in CRDS a laser pulse is stored in a stable optical cavity containing a sample, and the pulse decay, characterized by the "ring-down" time $\tau$, is monitored with the aid of a detector that measures the intensity of light transmitted through the mirrors. Sensitive detection of this decay time, which depends on the intracavity absorption losses, allows for highly sensitive absorption measurements that are inherently insensitive to the intensity noise originating from the light source. For the case of pulsed polarimetry we can distinguish up to date two distinct measurement approaches. The first is based on monitoring polarization-dependent changes in $\tau$ as the frequency of the laser source is scanned across a transition of the system under investigation that demonstrates dichroism (linear or circular). The system's dichroism (and its associated birefringence) will transfer a fraction of the intra-cavity optical power from one of its eigenpolarization states to its orthogonal one, which thus becomes observable through the losses measured by the ring-down decays on two orthogonal polarization-sensitive detection channels. Such measurements are effectively intensity-based and can be categorized under the general term of polarization-dependent cavity ring-down spectroscopy (PD-CRDS), a terminology introduced in the first demonstration by Engeln and co-workers [42, 43]. PD-CRDS schemes have achieved near-shot-noise limited sensitivities at $< 10^{-9}\,\mathrm{rad}\sqrt{\mathrm{Hz}}$ levels, and have been employed for measurements of linear and non-reciprocal circular birefringent effects, such as the measurement of residual or stress-induced linear birefringence of supermirrors [19, 20, 26, 44],

and the measurement of resonant Faraday optical rotation (otherwise known as FR-CRDS) as employed for the sensitive and selective detection of paramagnetic gaseous molecules and radicals (e.g. $O_2$ [45–48], $HO_2$ [48, 49]). However, PD-CRDS is specifically suitable for resonant effects and requires ultrahigh finesse optical cavities (as the sensitivity in the decay time is translated directly into polarimetric sensitivity), which effectively increases the cost of the apparatus and can limit its operational lifetime (degradation of the high quality optical mirrors results in reduced sensitivities), precluding this approach from being the most suitable one for portable instrumentations for field studies.

The second approach, alternative to PD-CRDS, relies on the use of an intracavity "bias" polarization anisotropy that introduces a large background polarization rotation and results in a rapid oscillating signal superimposed upon the ring-down signal as detected in a polarization-sensitive analysis stage. This polarization bias rotation offers two critical advantages. The first one is to provide an easily measured polarization beat frequency that is altered with the addition of an incremental anisotropy (i.e. an effect under investigation), transforming the polarimetric measurement into a frequency-based one. For this reason, we categorize these techniques under the general term of cavity ring-down polarimetry (CRDP). The second advantage is that the large bias polarization rotation can suppress intracavity anisotropies of opposite symmetry that might otherwise affect the desired measurement. This advantage becomes apparent in the case of chiral sensing, where CRDP approaches have been implemented for highly sensitive chiral polarimetry [15, 16, 18, 23]. Particularly, in the case of chiral sensing, any intracavity linear birefringence effect would otherwise inhibit the sensitive measurement of the expected weak chiroptical signals. The solution is to use a large intracavity circular birefringence, much larger than any residual linear birefringence present within the cavity, which introduces a large background polarization rotation and protects the measurement of the weak chiroptical signals. Despite the impressive successes of (chiral-sensitive) CRDP [23], the demonstrated sensitivities are still several orders of magnitude worse than their expected shot-noise limits.

Most of the PD-CRDS demonstrations have employed CW laser sources, while all CRDP demonstrations have employed thus far

only pulsed laser sources. The extension in using CW laser sources in CRDS was first shown by Romanini and co-workers, who demonstrated how CW-based CRDS allows for gains in spectral resolution, signal intensity, and data acquisition rates [50]. In CW-CRDS, the frequency of the CW laser is either locked into resonance with the CRD cavity or brought into resonance slowly (at rates sufficient slow to allow for maximum input coupling, constrained by the cavity lifetime), and a ring-down signal is observed after the laser beam is quickly interrupted (typically much faster than the cavity lifetime, i.e. $\tau$). This procedure guarantees a substantial buildup of the intracavity field leading to strong ring-down signals, as the signal intensity of the transmitted beam can reach close to the input intensity, enabling photon-shot-noise limited CRDS measurements. Several commercially available portable systems based on diode-laser CW-CRDS already exist [2], however none yet for highly sensitive polarimetry.

In this article, we present the technique of CW-CRDP, a methodology that combines the strengths of all the aforementioned techniques and alleviates their weaknesses, and which we consider to be the ideal modality for highly sensitive, cost-effective, polarimetry, particularly suitable for portable and robust spectro-polarimetric instrumentation. CW-CRDP extends the operational principles of pulsed-based CRDP by employing CW-laser sources, enabling time-resolved shot-noise limited spectropolarimetric sensitivities.

We start by presenting the general principles of operation of CW-CRDP and discuss how these can be generally adapted for measurements of birefringence and dichroism, both circular and linear. Our proposed methodology relies on the practice of using a bias intracavity polarization anisotropy as a means of protection of weak polarimetric signals under investigation, and while this approach has already been applied several times in cavity-based polarimetry, particularly in the case of studying chirality [15, 16, 18, 23], the fundamental measurement sensitivity benefits behind this methodology have not been explored in detail before, and we demonstrate here how CW-CRDP can reveal these.

In particular, in Sec. II we present the theory behind CW-CRD polarimetry, and its operational principles. In Sec. III, we present the details of a prototype setup that we employ for measurements of

[2] Los Gatos Research, http://www.lgrinc.com; Picarro, https://www.picarro.com

non-resonant Faraday optical rotation from solid and gaseous systems using CW-CRDP. In Sec. IV we present results for the Faraday effect in solid $SiO_2$ and $CeF_3$, and in gaseous butane, where we measure their Verdet constants to quantitatively validate our method. Finally, in Sec. V we discuss the fundamental sensitivity limits for any CW-CRD-based polarimetric scheme, which we employ to analyze the stability of our experimental results. Most importantly, we demonstrate how the introduction of an intracavity bias anisotropy, which enables frequency-based polarimetric measurements and is adaptable in all CRD-based polarimetric schemes, can effectively alleviate the necessity for high finesse optical cavities and, therefore, the need of ultrahigh-quality optics, which can be costly and not available at all optical spectral regions, a crucially important aspect for robust and versatile portable spectropolarimetric instrumentations.

### 4.2 Theory

We start by discussing the theory of CW-CRDP. Although several extended analyses of the theory behind CW-based cavity-enhanced polarimetric schemes exist in the literature [48, 51–53], we repeat here key theoretical aspects to clarify the necessary terminology and concepts required to understand the experimental principles of the CW-CRDP technique.

To examine the principles of operation of a CW-CRD polarimetric protocol we focus, for simplicity and without loss of generality, on linear optical cavities, consisting of two mirrors and, as a bias intracavity polarization anisotropy, a non-reciprocal circular birefringent effect (e.g. Faraday effect). However, we emphasize that the general theoretical approach we present here can be similarly applied for the case of reciprocal circular birefringence (e.g. for a four-mirror cavity for the study of chirality [53]) and linear birefringence (e.g. linear cavity for the study of Voigt effect, mirror-related birefringence) [26, 42], and, when required, we discuss extensions to different designs. Furthermore, we assume that the laser beam is mode-matched into the $TEM_{00}$ mode of the optical cavity and, therefore, we focus our analysis on the polarization properties of the cavity's longitudinal modes. We also neglect any changes in the spatial profile of the laser beam, possibly introduced by the intracavity element(s).

We present an eigenpolarization theory for the cavity-based polarimeter based on the Jones matrix calculus that allows us to describe the full optical system and to incorporate a CW laser source and the ability to switch it on and off rapidly. Here, the full optical system includes the polarization-control optics before the cavity, the optical cavity including the intracavity anisotropy, and the polarization-analysis (detection) stage. In the Jones matrix formalism, the effect of any optical element on the polarization state vector of the laser light is described as a linear operator expressed by a $2\times2$ matrix whose elements are in general complex. We denote each of these matrices by boldface letters $\mathbf{J}$. Most importantly, the direct incorporation of amplitude and phase information in the Jones matrices allows for the investigation of coherent phenomena.

#### 4.2.1 Jones matrices for polarization optics

The Jones matrix representation of a mirror reflection is given as:

$$\mathbf{J}_{\mathrm{M}_i}(R_i,\delta_i)=\sqrt{R_i}\begin{pmatrix} -e^{i\delta_i/2} & 0 \\ 0 & e^{-i\delta_i/2}\end{pmatrix}, \tag{3}$$

where the index $i$ denotes each cavity mirror (in our case $i$=$\{1,2\}$). Since we focus on two-mirror cavities, with normal angle-of-incidence reflections, we set the Fresnel amplitude reflection coefficients for the $s$ and $p$ polarizations to be equal in magnitude (an assumption expressed by the common factor $\sqrt{R_i}$). The differential $s$-$p$ phase shift $\delta_i\equiv\delta_p^i-\delta_s^i$, represents the linear birefringence obtained upon mirror reflection. For normal incidence these $s$-$p$ phase shifts can be of the order of $10^{-7}-10^{-5}$ rad at a specific design wavelength [20, 22].

In the presence of a longitudinal magnetic field a medium becomes circular birefringent, an effect otherwise know as the Faraday effect [54]. The Faraday optical rotation is expressed as: $\theta_F = V\cdot \mathrm{B}\cdot l$, where B is the magnetic field strength along the direction of light propagation, $l$ is the path-length in the medium, and $V$ is the Verdet constant of the medium. The Jones matrix representation for the Faraday effect is an SU(2) rotation matrix with argument $\theta_{\mathrm{F}}$:

$$\mathbf{J}_{\mathrm{F}}(\theta_{\mathrm{F}})=\begin{pmatrix}\cos\theta_{\mathrm{F}} & -\sin\theta_{\mathrm{F}} \\ \sin\theta_{\mathrm{F}} & \cos\theta_{\mathrm{F}}\end{pmatrix}. \tag{4}$$

The physical direction of the polarization rotation is defined by the magnetic field orientation. Due to the non-reciprocal nature of the

Faraday effect, when either the magnetic field or the direction of propagation of the light reverses, the sign of rotation reverses. This directional symmetry breaking, induced by the Faraday effect, has been essential for the implementation of crucial signal reversals in chiral cavity-based polarimetry [18, 23].

#### 4.2.2 Frequencies and polarizations of the cavity spectrum

The Jones matrix representation of a round-trip in the optical cavity is obtained by the ordered multiplication of Jones matrices representing the independent optical elements (see Fig. 3),

$$\mathbf{J}^{\text{cav}} = \mathbf{J}_{\text{M}_1}(R,\delta) \cdot \mathbf{J}_{\text{F}}(-\theta_{\text{F}}) \cdot \mathbf{J}_{\text{M}_2}(R,\delta) \cdot \mathbf{J}_{\text{F}}(\theta_{\text{F}}), \tag{5}$$

where, again for simplicity, we assume that mirrors $\text{M}_1$ and $\text{M}_2$ have the same characteristics.

Using Eq. 5 we can determine the allowed polarizations of the cavity modes (eigenpolarizations) along with their respective frequencies. Considering that the Jones matrices are unitary matrices of rank two, each matrix has two eigenvalues and two eigenvectors; the eigenvectors $\nu_{\pm}$ are orthogonal, in general complex, vectors and represent the eigenpolarizations of each cavity mode. The eigenvalues can be written in the general form $\lambda_{\pm} = e^{\pm i\alpha}$, and the phase of each eigenvalue, $\alpha$, is the round-trip optical phase shift obtained during light propagation, which, thus, yields the frequency splittings of the eigenmodes.

In the simple case of an isotropic cavity ($\theta_{\text{F}} = 0$ & $\delta = 0$), the two eigenmodes are degenerate and any polarization state can couple into the cavity (i.e. $\mathbf{J}^{\text{cav}}$ becomes proportional to the identity matrix; Fig. 3, upper panel). The introduction of any polarization anisotropy lifts this degeneracy. In the most general case, the spectrum of the cavity is represented by two non-degenerate modes of elliptical polarization, whose frequencies lie above and below the degenerate frequency of the isotropic case. However, when the intracavity anisotropy is a circular birefringence ($\theta_{\text{F}} \neq 0$ & $\delta = 0$), the two eigenpolarization modes are circular-polarization states, denoted hereafter as $R$ and $L$ modes, and, thus, the cavity spectrum is now represented by two modes split in frequency by $2\,\text{f}_{\theta_F} = 2\theta_{\text{F}} \cdot \text{FSR}/\pi$, where FSR$= (c/L_{rt})$ is the cavity's free spectral range, with $c$ the speed of light and $L_{rt}$ the round-trip cavity length (Fig. 3, lower panel). We note that this mode structure can

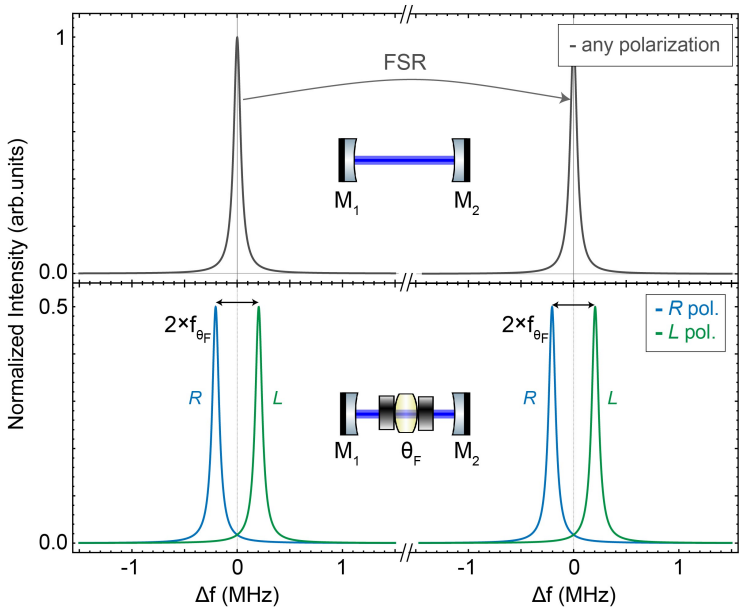


Figure 3: Cavity frequency polarization spectrum of a linear (two-mirror) optical cavity in the **(a)** absence and **(b)** in the presence of an intracavity polarization anisotropy. We choose here as an intracavity anisotropy a non-reciprocal Faraday effect, which splits the cavity spectrum into two (orthogonal) circularly polarized modes, $R$ and $L$, by $2\,\mathrm{f}_{\theta_F} = 2\,\theta_F \times (\pi/\mathrm{FSR})$ (for clarity we assume here a $\theta_F$ value much larger than the cavity linewidth).

be resolved in the case of a frequency splitting much larger than the cavity linewidth, which is the case when the polarization anisotropy (here a Faraday effect), obeys the relationship $\theta_F \gg \pi/\mathcal{F}$.

In the case of intracavity linear birefringence ($\theta_{\mathrm{F}} \neq 0$ & $\delta \neq 0$), which can also originate from thermal or mechanical stress in all intracavity optics apart from mirror-reflection-related phase shifts, the path-length-related enhancement of circular birefringence is inhibited through the transformation of the circularly polarized eigenmodes into linear ones. The equivalent time-domain explanation is that an incident linearly polarized light beam will oscillate between linear and circular polarization states, reducing the effective path-length enhancement and the sensitivity of the measurement. However, as long as the bias intracavity circular birefringence is much larger than linear birefringence, i.e. here $\theta_F \gg \delta$, then the effects of linear birefringence will be averaged out and the cavity modes will maintain the circular polarization character. The physics be-

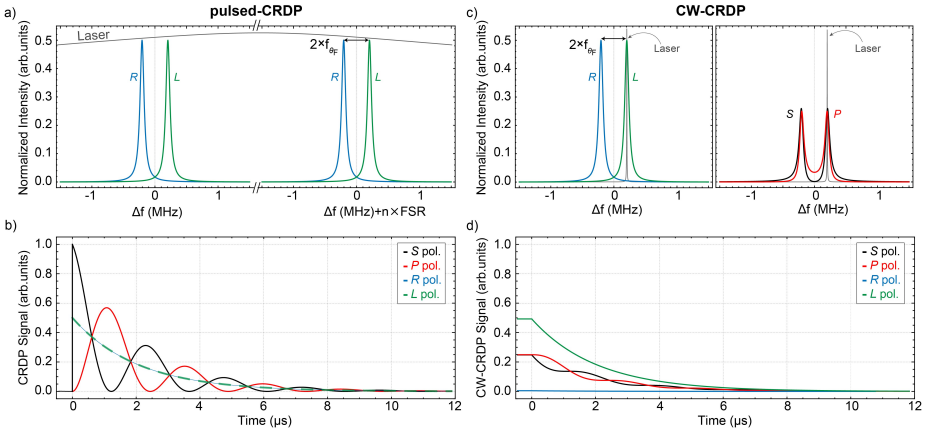


Figure 4: *Principles of CRDP measurements* - **(a)** & **(b)** pulsed-CRDP, **(c)** & **(d)** CW-CRDP. For the simulations we assume a two-mirror cavity with an intracavity circular birefringence, i.e. Faraday effect (we use cavity length of 0.6 m, cavity finesse of $\mathcal{F} \approx 3140$, and $\theta_F = 2.6$ mrad). Typically, the frequency linewidth of a pulsed laser is much larger than several ($n$) cavity FSRs, which allows for the direct coupling of an incident linearly polarized beam into the cavity [(a)]. The transmitted radiation is analyzed with a linear balanced polarimeter yielding a ring-down signal superimposed with a polarization frequency beating that has a modulation depth approaching unity, i.e. the CRDP trace (when the circular components are analyzed, pure exponentially decaying signals are observed demonstrating the equal coupling to both $R$ & $L$ modes) [(b)]. In contrast, in CW-CRDP the linewidth of the laser is typically much smaller than a cavity linewidth and the incident radiation couples into a single cavity mode. Depending on the detuning with respect to the $R$-$L$ cavity modes, different polarization states are allowed to couple into the cavity [(c)]. Once the incident (and frequency locked) radiation is switched off, at $t = 0$, the excited polarization mode evolves resulting in pure exponentially decaying signals when analyzed with a circularly sensitive balanced polarimeter, while when analyzed with a linearly sensitive balanced polarimeter, which appropriately mixes the circular polarization components of the transmitted radiation to produce two orthogonal, linearly polarized ($S$ and $P$) waves, it results in ring-down signals superimposed with a polarization beat frequency that now has a modulation depth and amplitude defined by the initially excited polarization state [(d)]. In (a) &(c) the gray dotted lines at $\Delta$f = 0 correspond to the two-fold degenerate axial mode of an isotropic cavity.

hind this process has been extensively discussed within the context of chiral-sensitive CRDP [16, 53, 55].

Finally, in the case of circular dichroism, the linewidths of the two cavity eigenpolarization modes become different, since the cavity finesse depends on the intracavity losses, which are in this case different for the two circular polarization states.

#### 4.2.3 Principles of a CRDP measurement: pulsed vs. CW

In traditional CRDP, laser pulses with linewidths much larger than a cavity FSR are used [Fig. 4 (a)]. For example, a gaussian-shaped Fourier-limited laser pulse of $\sim$ 1 ps in duration has a spectral bandwidth of $\sim$ 0.4 THz, which is much larger than the FSR of meter-long cavities. Typically, such a linearly polarized laser pulse is directly coupled into the cavity by coherently exciting multiple cavity (eigenpolarization) modes. The transmitted beam is also linearly polarized, and when analyzed using a linear-polarization detector, i.e. a balanced polarimeter that analyzes linear polarization components such as a Wollaston prism, the resulting ring-down signal is superimposed with the differential polarization beat frequency equal to the $R$-$L$ frequency mode-splitting [CRDP trace: Fig. 4 (b); note that if we analyze the circular components we see pure exponentially decaying ring-down signals]. If there are no depolarization mechanisms, such as dichroic absorption losses and/or any residual intracavity linear birefringence, the modulation depth of this polarization beat frequency approaches unity.

In the case of CW-CRDP, the measurement principle is different. Incident radiation from a CW laser source with a linewidth that is significantly smaller than that of a cavity, can couple into and build-up within that cavity by exciting only one particular cavity mode at a time [Fig. 4 (c)]. As shown in the previous section, for a linear cavity with a large intracavity circular birefringent anisotropy, the cavity eigenpolarization modes are frequency-separated circularly polarized modes ($R$ and $L$ modes), and the polarization state of the incident radiation will be projected onto these. Therefore, depending on the frequency detuning of the incident radiation with respect to the $R$-$L$ cavity eingepolarization modes, a different polarization state will be coupled into the cavity. For example, the frequency at which the $R$ and $L$ modes are equally excited [$\Delta$f $=$ 0 MHz; Fig. 4 (c)] corresponds to a mode of linear polarization, as this is an equal coherent superposition of right and left circularly polarized states. Similarly, the central frequencies of the individual $R$ and $L$ cavity modes correspond to elliptical polarization states (since these are an unequal superposition of the $R$ and $L$ modes), and the degree of ellipticity of these states depends on the detuning but also on the $R$-$L$ mode-splitting, i.e. on the value of the bias anisotropy (e.g. $\theta_F$), as the larger (smaller) the splitting, the larger the circular (lin-

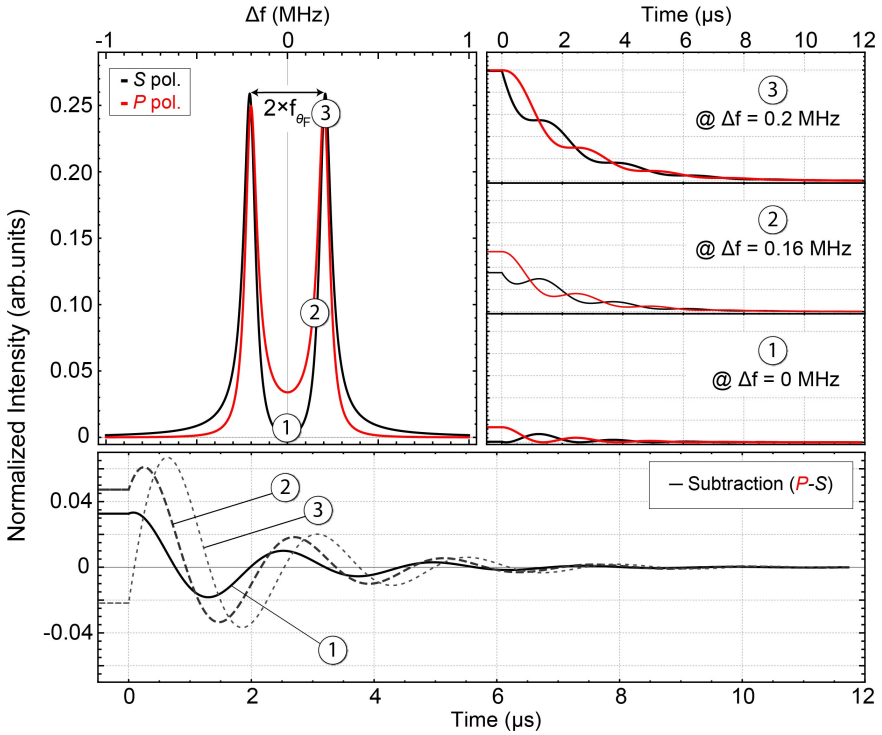


Figure 5: Depending on the detuning of the frequency of the incident radiation with respect to the $R$ and $L$ cavity eingepolarization modes, a different polarization state is coupled into the cavity, and once the incident (and frequency locked) radiation is switched off at $t = 0$, the excited polarization mode evolves resulting in a ring-down signal superimposed with a polarization beat frequency, as recorded by the $S$ and $P$ polarization channels of a linear balanced polarimeter. The amplitude, modulation depth, and phase of the resulting CW-CRDP traces depend on the frequency detuning (upper right panels). Obtaining the difference via a balanced detection scheme, yields an exponentially decaying sinusoidal signal where one can clearly observe the amplitude and phase differences between the resulting CW-CRDP traces (lower pannel). For the simulations we use the same parameters as for Fig. 3 and Fig. 4.

ear) character of the coupled, and consequently of the transmitted, polarization state.

As in pulsed-CRDP, in CW-CRDP one analyzes the polarization state of the transmitted radiation using a linearly-sensitive balanced polarimeter, which appropriately mixes the circular polarization components of the transmitted radiation to produce two orthogonal, linearly polarized ($S$ and $P$) waves incident upon separate photodetectors [Fig. 4 (c)]. However, in CW-CRDP, to initiate a ring-down event, the incident radiation which is coupled into and

built-up within the cavity, is abruptly switched off (at time-scales much faster than the cavity lifetime). Since the frequency detuning of the incident radiation defines the coupled polarization state (which can vary continuously from linear to circular), once the incident radiation is switched off, this coupled (intracavity) polarization state freely evolves, resulting in a transmitted ring-down signal superimposed with a polarization beat frequency [CW-CRDP trace: Fig. 4 (d)]. For the case of incident radiation tuned to the central frequency point ($\Delta \mathrm{f} = 0\,\mathrm{MHz}$), linearly polarized light couples to and is transmitted by the cavity, which results in a modulation depth that reaches unity, similarly to the case of pulsed-CRDP. However, in the CW case, the signal amplitude is reduced, and this reduction follows the value of the intracavity polarization anisotropy (i.e. the *R-L* mode-splitting). Similarly, the circularity of the coupled light increases as the incident radiation is detuned from this central frequency point, leading to a decrease in the modulation depth of the polarization beat frequency, though accompanied by an increase in amplitude. We demonstrate these cases in Fig. 5 where we present simulations of CW-CRDP traces for different possible frequency detunings of the incident radiation with respect to the *R-L* modes. Overall, an optimum detuning exists, for performing highly sensitive CW-CRDP measurements, that depends on the cavity linewidth (i.e. the cavity finesse), the strength of the intracavity anisotropy (e.g. $\theta_F$), and the frequency detuning of the incident radiation with respect to the cavity's eigenpolarization modes. In Sec. V we discuss how one should appropriately select the experimental parameters of a CW-CRDP technique when we analyze the fundamental sensitivity limits of CRDP methods.

In Fig. 5 we also show that subtracting the $P$ and $S$ signals generates a pure damped sinusoidal signal which clearly exhibits the change in signal amplitude as a function of the frequency detuning of the incident radiation with respect to the cavity's eigenpolarization mode-splitting. In a CW-CRDP protocol, a balanced detection scheme is advantageous as the signal subtraction suppresses common mode laser intensity noise measured by the two $S$- and $P$-wave photodetectors, resulting in improved signal-to-noise ratios. Furthermore, in addition to the amplitude change as a function of detuning, our simulations show an associated phase shift in the recorded beat frequency. This phase shift dependence on the frequency detuning

is critical for a CW-CRDP protocol because it indicates that the performance of a frequency-stabilization scheme, which controls the frequency of the incident radiation with respect to the cavity-mode structure, directly impacts the measurement sensitivity (in pulsed-CRDP this is not an issue as the incident radiation couples fully with multiple modes). We discuss this in more detail when we analyze the fundamental sensitivity limits of the CW-CRDP method.

Here, we wish to emphasize a critical difference between pulsed-based and CW-based CRDP. In the case of CW-CRDP the initial signal amplitude is defined by the cavity transmission on resonance, which in theory can reach unity, so detected optical powers of 1-100 mW can be achieved, whereas for pulsed-CRDP typical detected optical powers are at the $\mu$W levels at best [23, 50]. Therefore, CW-CRDP can result in signals with significantly higher signal-to-noise ratio than pulsed techniques, while attaining much higher repetition rates through the use of fast optical switches, leading to improved (ideally shot-noise limited) statistics for a given integration time.

Finally, although this discussion focuses on measurements of circular birefringence, the same measurement principles are applicable in studies of linear birefringence. Moreover, the case of linear (circular) dichroism is studied by measuring the ring-down times as recorded by a linearly- (circularly-)sensitive balanced polarimeter.

## 4.3 Experiment

### 4.3.1 CW-CRDP apparatus

In Fig. 6 we present a schematic diagram of the optical setup we use to demonstrate the experimental principles of the CW-CRDP technique. In this work we focus on studying the non-resonant Faraday effect from gaseous species. For this reason, we chose as our intracavity bias anisotropy a non-reciprocal Faraday optical rotation, i.e. $\theta_{\text{F}}$, which can be generated with the use of a material that has low enough absorption losses and large enough magneto-optic response, to enable sensitive CW-CRDP measurements. We describe in the next section the different options we consider in this work.

The ring-down cavity we use has a total length of 0.60 m and consists of two highly reflective concave mirrors with radii of curvature of 1 m (FiveNine Optics; specified reflectivity R$\sim$99.9% at 408 nm). The mirrors and intracavity optics are mounted on kine-

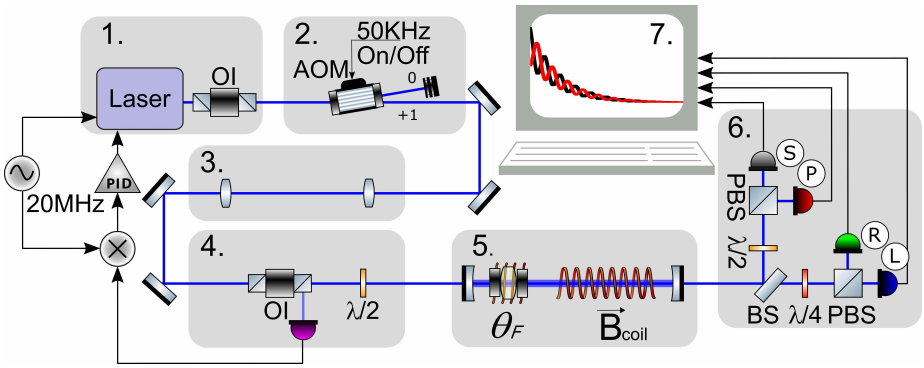


Figure 6: Schematic diagram of the CW-CRDP optical setup: 1) Laser operating in continuous-wave mode followed by an optical isolator (OI) that blocks unwanted back-reflections to the laser source; 2) a fast optical switch (acousto-optic modulator; AOM) is used to rapidly interrupt the laser beam and initiate a ring-down event; 3) beam-shaping optics to enable optimal coupling of the laser beam into the optical cavity; 4) polarization preparation stage and optoelectronics required to generate an error signal; 5) two-mirror optical cavity with an intracavity Faraday anisotropy, characterized by its rotational strength, $\theta_F$, and permanent magnets and coils for the generation of axial magnetic fields; 6) light transmitted through the cavity is split using a beam splitter (BS) and is sent to a linearly and a circularly sensitive balanced polarimeter for full polarization analysis; 7) analog-to-digital conversion electronics and data processing.

matic mounts. To be able to perform Faraday effect measurements of gaseous samples, we install an intracavity, 40 cm long, solenoid that generates uniform magnetic fields (with strengths of 40 G per ampere). The solenoid is driven by a DC power supply, and we switch the magnetic field using a relay circuit controlled with a digital-to-analog device (Labjack, U6). The whole cavity is housed within a custom stainless-steel enclosure, sealed with a Plexiglas lid. Anti-reflection (AR) coated flange viewports (Thorlabs, VPCH42-A) are mounted on the enclosure and allow for laser-light access into the cavity and for collecting the cavity reflection required for frequency locking. A rotary vane pump (Kurt J. Lesker, RV224) is connected to the enclosure and is used to pump down the system for measurements, and we monitor the pressure inside the enclosure using a vacuum gauge. We control the injection of gases into the cavity using a needle valve, which is connected directly to the enclosure.

The laser source is an external cavity diode laser with a center wavelength of $\lambda = 408$ nm (Toptica DL-PRO), and we use a set of lenses to achieve spatial mode-matching of the laser beam

to the fundamental transverse cavity mode. Our primary focus in this work is the study of non-resonant Faraday effects and, for this reason, we do not pre-stabilize the frequency of our laser to be resonant with a specific transition frequency. However, to establish a CW-CRDP protocol we lock the frequency of the laser to the cavity resonance using a Pound-Drever-Hall (PDH) scheme [56]. To prepare the polarization state of the input beam, and to collect the back-reflected one to a photodetector (Thorlabs, PDA10A-EC) for PDH-locking, we use an optical isolator that has side exit ports equipped with polarizing beam splitters (Thorlabs, IO-5-405-LP) followed by a half-wave plate. We generate the error signal using direct current modulation and a PDH module (Toptica, PDD 110). The error signal is then input to an analog PID controller (Toptica FALC 110) to provide a fast feedback to the laser-diode current, and to a second analog PID (SRS, SIM960) that creates a slower feedback that acts on the grating of the ECDL through a piezoactuator. The gain and bandwidths of the PID loops are different for low frequencies (<10 Hz) to prevent accumulation of DC-offsets in the error signals that would otherwise prevent a stable lock. To initiate a ring-down event, we use an acousto-optic modulator (AOM; Gooch and Housego 3200-125) driven with a homebuilt AOM-driver, whose RF output can be controlled though a high-bandwidth switch. We analyze the light transmitted through the cavity using two balanced polarimeters: a linearly and a circularly sensitive one, which consist of a half-wave plate (Thorlabs, WPH10ME-405) followed by a Wollaston prism (Thorlabs, WP10), and a quarter-wave plate (Thorlabs, WPQ10ME-405) followed by a Wollaston prism, respectively. All wave-plates and prism are placed within rotation mounts. In each balanced polarimeter, we use short lenses ($f = 50$ mm) to focus the emerging radiation on separate silicon amplified photodetectors (Thorlabs, PDA8A).

In Fig. 7 we show typical experimental CW-CRDP traces. With our optical setup and intracavity optics we can achieve ring-down times in the 0.7-1.5 $\mu$s range, and with our feedback system and AOM electronics, we are able to generate CW-CRDP traces at repetition rates as high as 50 kHz.

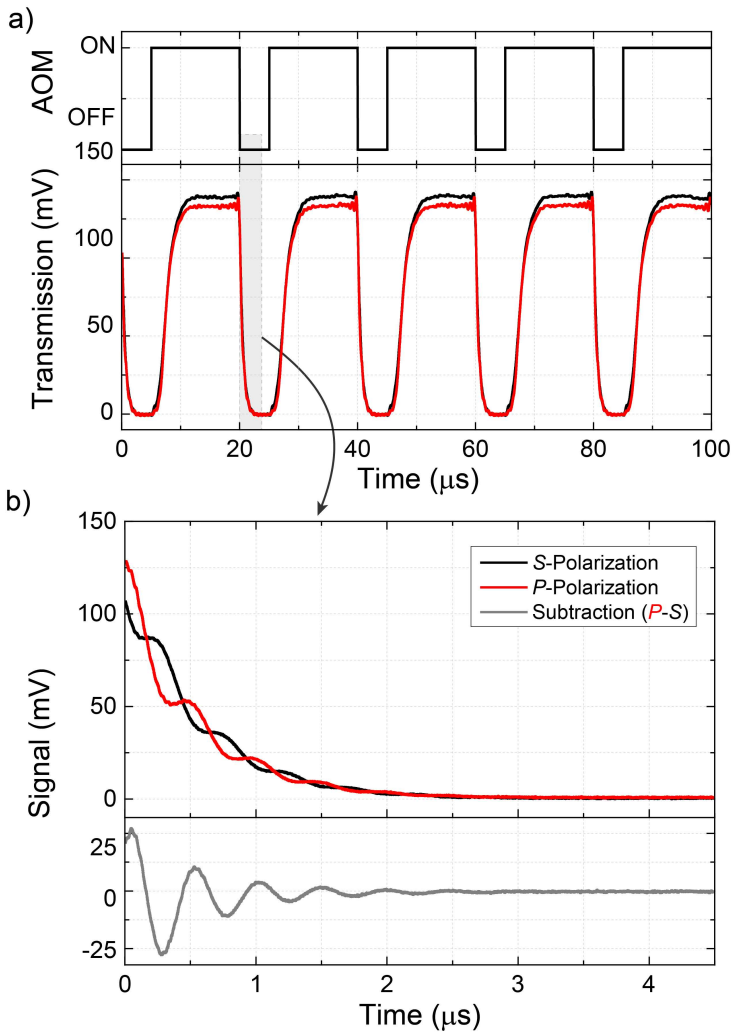


Figure 7: **(a)** To initiate a ring-down event, the incident radiation, which is frequency locked to the cavity resonance, is abruptly switched off (at time-scales much faster than the cavity lifetime) with the use of an acousto-optic modulator (AOM). **(b)** Typical CW-CRDP traces, showing polarization beat frequencies generated via the Faraday effect on a $SiO_2$ substrate, as recorded by the $S$ and $P$ channels of a linear balanced polarimeter (Fig. 6).

### 4.3.2 Data acquisition and signal analysis

We record and digitize the ring-down traces (photo-detector signals) using two different acqusition modules: an oscilloscope (LeCroy, Wavesurfer 510), which has a maximum acquisition rate of approx-

imately 200 Hz, an 8-bit resolution per channel, and permits on-board signal averaging; and a 14-bit digitizer (Teledyne, ADQ14DC-2X-PCIE, dual channel DC-coupled operation; sample rates of 2 GS/s per channel), which has a maximum acquisition rate of approximately 100 kHz (mainly limited by the data transfer rate), 14-bit resolution per channel, and permits on-board signal averaging. All collected data are transferred into a personal computer, and are processed by a custom Python program (SciPy [3]).

In this work we use time-domain analysis, based on nonlinear least-squares regression of the recorded time-traces, and all the data we present are the result of such an analysis. In particular, we record signals from both polarization channels (Fig. 6 & Fig. 7), and we fit the resulting CRDP-traces using the following model:

$$I(t) = A\, e^{-t/\tau}\, [\, \sin^2(2\pi f \cdot t + \phi) + B\, ] + C, \tag{6}$$

where $A$ is the amplitude of the trace, $\tau$ is the ring-down time, $f$ is the polarization beating frequency, $\phi$ represents a global phase offset, parameter $B$ takes into account any reduction in modulation depth, and $C$ is a global signal offset. With our acquisition systems and analysis process we can record and monitor in real-time the signal amplitude, ring-down time, frequency, and phase of both polarization channels independently. The subtraction of the signals, which yields a damped sinusoid, is performed through our computer-analysis software.

We wish to emphasize here that for CRDP-based portable spectropolarimetric instrumentations and, especially, experiments studying dynamics [57], online and fast, acquisition and analysis, are crucial. For the purposes of this work we focus on a post-processing time-domain analysis that allows us to thoroughly investigate the sensitivity limits of CW-CRDP experiments for total integration times ranging from a few ms to a few seconds. However, computational algorithms for time-domain analysis are typically slow (in our case we require 10-100 ms to fit a single trace) and might not be the appropriate methodology for rapid signal analysis (at time scales similar to a single ring-down event), particularly for spectropolarimetry where precision is the key criterion in choosing the appropriate analysis for CRDP techniques. A detailed investigation of the advantages and disadvantages of different signal analysis tech-

[3] https://www.scipy.org

niques, to the best of our knowledge, does not exist in the context of CRDP and we will address this in a follow-up work.

## 4.4 Results

As a proof-of-principle demonstration of the CW-CRDP technique we choose to study the non-resonant Faraday effect from a gaseous species, particularly butane. CW-CRDP relies on the implementation of a large intracavity bias anisotropy that assists in the measurement of the weak Faraday effect of butane (from studies in the literature we know that the expected Verdet constant of butane at 408 nm is approximately 30 nrad $G^{-1}$ $cm^{-1}$ $bar^{-1}$, at 408 nm). For this reason, we start by investigating different optical materials that can allow for the generation of a large enough intracavity bias circular birefringence while having as low as possible absorption. We are interested in measurements at 408 nm where typical magneto-optic crystals are expected to have significant absorption losses ($\gtrsim 1\,cm^{-1}$). We focus our efforts on two optical elements, a $SiO_2$ substrate and a $CeF_3$ crystal, to investigate their magneto-optic response and the possibility of employing them for gas-phase measurements.

### 4.4.1 $SiO_2$ Faraday effect measurement

From studies available in the literature we find that the absorption coefficient for $SiO_2$ is estimated to be $< 10^{-5}\,cm^{-1}$ [58], while its Verdet constant is 10 $\mu$rad $G^{-1}$ $cm^{-1}$ [59, 60], at 408 nm. Therefore, we expect that high quality $SiO_2$ substrates and AR-coatings can allow for ultra high-finesse cavities, and the Verdet constant is large enough to allow for a sufficiently large *R-L* mode-splitting for applied magnetic fields of approximately a few ∼kG (possible with the use of permanent magnets).

In Fig. 8 we present measurements of the (non-resonant) Faraday effect of a 6.35(1) mm thick, AR-coated $SiO_2$ substrate (FiveNines Optics; AR coated by FiveNine Optics with specified R< 0.01%), which is also the first demonstration of the CW-CRDP technique. To generate a large bias Faraday optical rotation we use permanent magnets (K&J Magnetics, Inc.) attached directly to the mount holder of the substrate. For measuring precisely the Verdet constant of the substrate we use an additional, homemade, solenoid

[with a length of 2.53(1) cm, and a diameter of 3.05(1) cm], which we place around the $SiO_2$ substrate. We calibrate the magnetic field of the solenoid using a Hall probe magnetometer (Hirst Magnetic Instruments GM05), and use a USB controlled digital-to-analog device (Labjack, U6) to drive the solenoid and scan the magnetic field it generates. The permanent magnets allows us to generate a large bias Faraday optical rotation, yielding a bias beat frequency of $f_0 = 2\,121\,123(3)$ Hz, and using the smaller solenoid we are able to induce frequency shifts around this central beat frequency [Fig. 8 (a)]. For each magnetic field strength we obtain the resulting frequency shift [$\Delta f = f(B) - f(0)$], which in turn yields the Faraday optical rotation for $SiO_2$ [Fig. 8 (b)] and which is directly proportional to the Verdet constant of the $SiO_2$ substrate. Importantly, we observe in our measurements frequency drifts that can be associated with mechanical drifts, but our subtraction procedure removes these and allows us to observe a linear dependence of the measured optical rotation on the applied magnetic field. With this procedure, we measure the Verdet constant to be $V^{SiO_2} = 10.1(3)\,\mu\text{rad}\,\text{G}^{-1}\,\text{cm}^{-1}$, at 408 nm, which is in accord with results available in the literature [Fig. 8 (c)]. The measurement error bars are derived through propagation of the errors ascribed to the measured polarization beat frequencies.

#### 4.4.2 $CeF_3$ Faraday effect measurement

Garnet single crystals, such as yttrium-iron-garnet (YIG) and terbium-gallium-garnet (TGG), are magneto-optic materials typically used in Faraday rotators, due to their transparency in the visible and NIR optical ranges and their high Verdet-constant values. However, all these materials exhibit significant losses at wavelengths $< 600$ nm. Recent studies have reported on alternative materials suitable for Faraday rotators at near-UV optical wavelengths, such as $CeF_3$ and $PrF_3$ [61, 62]. For $CeF_3$, in particular, the absorption coefficient at 408 nm is predicted to be $\lesssim 10^{-1}\,\text{cm}^{-1}$, and its Verdet constant is predicted to be approximately $390\,\mu\text{rad}\,\text{G}^{-1}\,\text{cm}^{-1}$ [62]. To verify these numbers and, therefore, investigate whether we can employ this crystal for inducing an intracavity bias anisotropy for CW-CRDP at 408 nm (and similar wavelengths), we use a 1.20(1) mm thick, AR-coated $CeF_3$ crystal (E-Crystal Co., Ltd.; AR-coated by E-Crystal Co., Ltd. with specified R$<$ 0.2%).

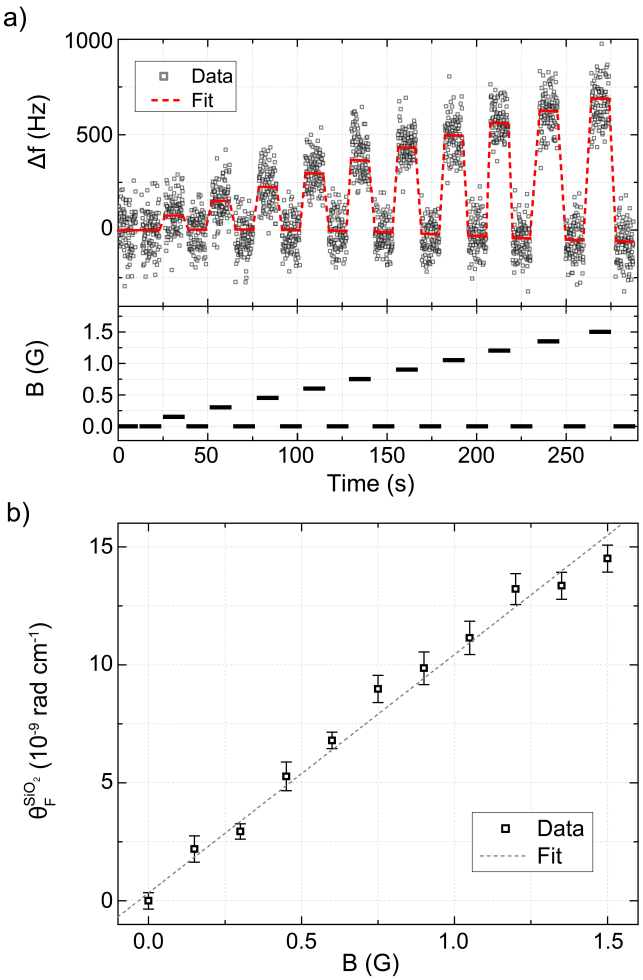


Figure 8: **(a)** Measurements of frequency shifts [$\Delta f = f(B) - f_0$] with respect to the bias beat frequency induced by a set of permanent magnets placed around a 6.35 mm thick $SiO_2$ substrate [$f_0 = 2\,121\,123(3)$ Hz], as a function of an externally applied magnetic field (B). **(b)** Measurements of Faraday optical rotation per unit length for a $SiO_2$ substrate, $\theta_F^{SiO_2}$, at 408 nm, as a function of applied magnetic field. The dashed line is the result of a linear least-squares regression analysis that yields the Verdet constant for $SiO_2$ at 408 nm: $V^{SiO_2} = 10.1(3)\,\mu\text{rad}\,\text{G}^{-1}\,\text{cm}^{-1}$ (one standard deviation uncertainties).

In Fig. 9 we present measurements of the (non-resonant) Faraday effect of crystalline $CeF_3$ as a function of magnetic field. Using the crystal, we observe ring-down times of approximately $\tau_{CeF_3} \approx 200$ ns (compared to the attainable ring-down time of $\tau \approx 1.5\,\mu$s

for the empty cavity), which suggests overall losses of $\approx 0.14\,\mathrm{cm}^{-1}$ at 408 nm. Although, these losses are the combined result of scattering losses from the AR-coatings and of material absorption, we attribute these to absorption. To measure precisely the material's Verdet constant we follow a similar measurement approach as described above. A set of permanent magnets allows us to generate a large bias Faraday optical rotation frequency, yielding a bias beat frequency of $f_0 = 7\,151\,229(229)\,\mathrm{Hz}$, and using the smaller solenoid we are able to induce frequency shifts around this central bias beating frequency. We measure the Verdet constant to be $V^{CeF_3} = 462(16)\,\mu\mathrm{rad\,G^{-1}\,cm^{-1}}$, at 408 nm. Our findings are in relative agreement with the results presented in Ref. [62].

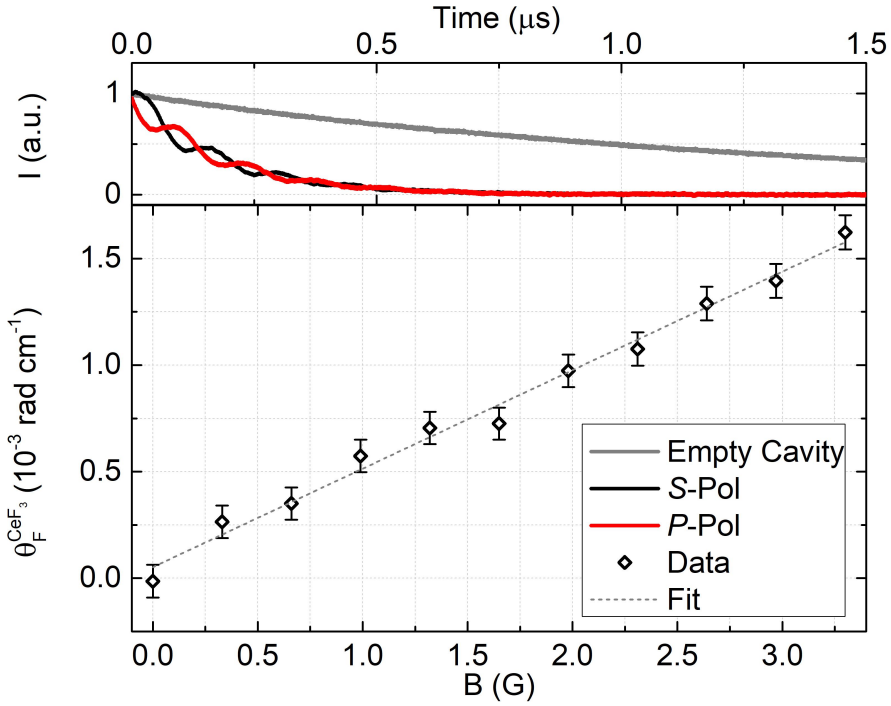


Figure 9: **Upper panel -** CW-CRDP traces for an empty cavity vs. one with an intracavity 1.20 mm thick AR-coated $CeF_3$ crystal. The bias polarization beat frequency, $f_0 = 7\,151\,229(229)$ Hz, is the result of large Faraday optical rotation generated using a set of permanent magnets. **Lower panel -** Measurements of Faraday optical rotation per unit length for the $CeF_3$ crystal, $\theta_F^{CeF_3}$, at 408 nm, as a function of applied magnetic field. The dashed line is the result of a linear least-squares regression analysis that yields the Verdet constant for $CeF_3$ at 408 nm: $V^{CeF_3} = 462(16)\,\mu\mathrm{rad\,G^{-1}\,cm^{-1}}$ (one standard deviation uncertainties).

### 4.4.3 Gaseous butane Faraday effect measurement

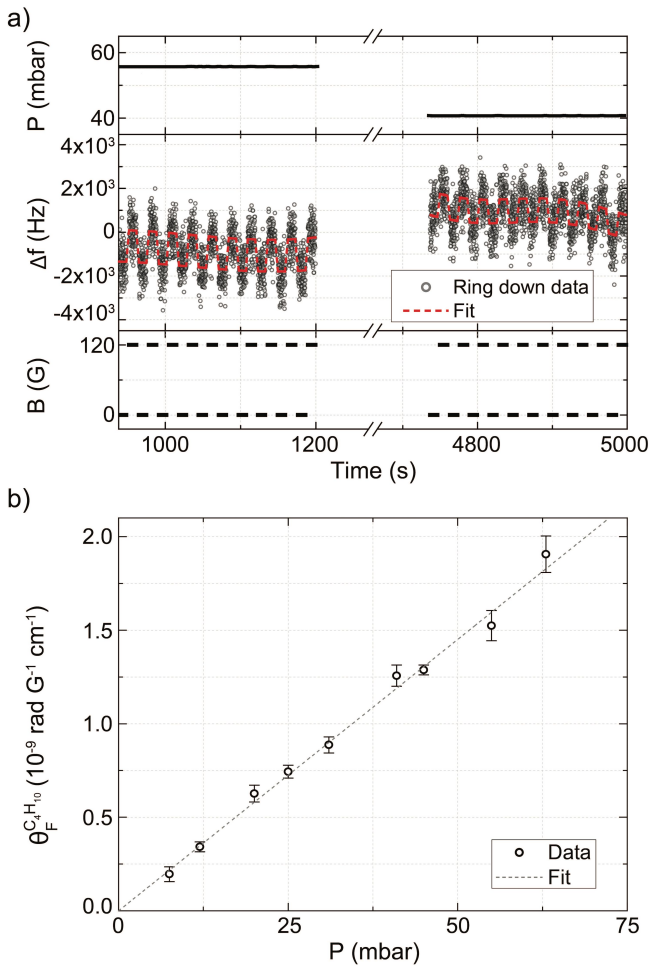


Figure 10: (a) Measurements of frequency shifts [$\Delta f = f(B) - f_0$] with respect to the bias beat frequency induced by the Faraday rotation from a $SiO_2$ substrate [$f_0 = 2\,858\,570(20)$ Hz], due to the non-resonant Faraday effect of gaseous butane at 408 nm, for two different gas pressures [55.6(1) mbar and 40.7(1) mbar] and for an applied magnetic field of 120 G. (b) Measurements of Faraday optical rotation per unit length for gaseous butane, $\theta_F^{C_4H_{10}}$, at 408 nm, as a function of pressure for a magnetic field of 120 G. The dashed line is the result of a linear least-squares regression analysis that yields the Verdet constant for butane at 408 nm: $V^{C_4H_{10}} = 29.0(7)\,\mathrm{nrad\,G^{-1}\,cm^{-1}\,bar^{-1}}$ (one standard deviation uncertainties). (c) Comparison of the measured Verdet constant for butane with results presented in the literature (Ingersoll et al. 1954: Ref. [63]; Ingersoll et al. 1956: Ref. [64]).

For the measurement of the Faraday effect in gaseous butane

we induce an intracavity bias anisotropy with the $SiO_2$ substrate (Fig. 8). We use a similar measurement procedure as described before and employ a homemade solenoid (40 cm) for the generation of adjustable magnetic fields (Fig. 6). We use an external power supply to control the current of the solenoid and a relay circuit to controllably switch the magnetic field on and off.

In Fig. 10 we present measurements of the (non-resonant) Faraday effect of butane ($C_4H_10$) as a function of pressure. For each pressure, we perform approximately 10 magnetic field cycles and obtain the frequency difference with respect to the bias Faraday beat frequency generated using a $SiO_2$ substrate [$\Delta f = f(B) - f_0$, with $f_0 = 2\,858\,570(20)$ Hz]. This process ensures that we remove any contribution to the signal originating from the permanent magnets used to create this large bias Faraday anisotropy. We verify the linear dependence between the observed Faraday optical rotation (per unit length) and butane's vapor pressure, and we measure its Verdet constant to be $V^{C_4H_{10}} = 29.0(7)\,\mathrm{nrad\,G^{-1}\,cm^{-1}\,bar^{-1}}$, at 408 nm and 21 °C [Fig. 10 (b)]. Our measurements are in accord with results available in the literature [Fig. 10 (c)].

## 4.5 Sensitivity

### 4.5.1 Cramér-Rao Lower Bound: fundamental sensitivity limit in CRDP

The noise in any CRDP-based polarimetric measurement will contain contributions from both technical and intrinsic (fundamental) sources. For typical experimental setups, similar to the one we use for the measurements, technical noise sources are predominantly of mechanical and acoustic nature with non-white power noise spectral densities, and their associated frequency/phase drifts will effectively be the major limiting factor influencing high-sensitivity measurements. However, we wish to consider here the fundamental sensitivity of any CRDP scheme that is directly related to the measurement sensitivity of the polarization beat frequency. While we generally focus on time-domain analysis of the CRDP signals, nothing precludes us from using alternative analysis schemes, such as frequency-domain approaches. Notwithstanding, independent of which analysis methodology we choose, the fundamental limit for the statistical uncertainty of determining the beat frequency from

a given CRDP trace will be given by the Cramér-Rao lower bound (CRLB) [65], which sets the lower limit on the variance $\sigma_{\mathrm{f}}^2$ of any frequency estimator. The CRLB condition for the frequency extracted from a discrete damped sinusoid - the CW-CRDP signal in our case as obtained from the balanced detection stage - is given by [65–68],

$$\sigma_{\mathrm{f}}^2 = \frac{6}{(2\pi)^2\,\mathrm{SNR}^2\,\mathrm{f_{BW}}\,\mathrm{T_{meas}^3}}\chi\left(\tau/\mathrm{T_{meas}}\right), \tag{7}$$

where SNR is the measured signal-to-noise ratio, which is defined as the ratio between the signal amplitude and the standard deviation of the noise (effectively the electronic noise of the acquisition system); $\mathrm{f_{BW}}$ is the sampling-rate-limited bandwidth of the measurement; $\mathrm{T_{meas}}$ is the measurement time window; and $\chi(r)$ is a corrective factor that takes into account the signal decay that is given by,

$$\chi(r) = \frac{e^{2/r}-1}{3r^3\,\cosh(2/r) - 3r\,(r^2+2)}. \tag{8}$$

The factor $\chi\left(\tau/\mathrm{T_{meas}}\right)$ in Eq. 7 is a compensation factor that penalizes measurement of the tails of the exponential decay when the signal has effectively died out. Importantly, Eq. 7 assumes that the period of the oscillation is much shorter than the decay constant of the envelope and that a sufficient number of oscillations occurs in the signal, without limiting the estimation accuracy. Equation 7 sets the frequency detection limit and, therefore, the CRLB-related optical rotation sensitivity limit is equal to,

$$\sigma_\theta = \frac{\pi}{\mathrm{FSR}}\sigma_{\mathrm{f}}. \tag{9}$$

We emphasize here that Eq. 7 dictates that any noise sources affecting the experimental measurements are contributing to the fundamental CRLB limit through their effect on the SNR of the CRDP signal.

To test the validity of Eq. 7 as the appropriate estimator of the fundamental sensitivity for a frequency-based CRDP measurement, we use an arbitrary waveform generator (AWG; Tektronix AWG7122C, 10-bit resolution) to simulate CRDP signals and determine the sensitivity in measuring their beat frequency. In particular, we simulate CRDP signals with a constant beat frequency of 2 MHz and variable signal-to-noise ratios (SNR) and ring-down times. We use the 14-bit digitizer (Teledyne, ADQ14DC-2X-PCIE) to record (over a fixed

time window of approximately $\sim 7\tau$) and integrate AWG-generated CRDP traces, which are analyzed using a time-domain approach to obtain their amplitude, frequency and respective SNR. Although we vary the SNR of the simulated signals by the introduction of white gaussian noise through the AWG, we use the time domain analysis for the estimation of the SNR value we use in Eq. 7.

In Fig. 11 we show two special cases of the simulated CRDP traces together with the results of our analysis. We confirm that the frequency estimation errors coincide with the CRLB limit (Eq. 7) for a wide range of ring-down decay times and SNRs. Most importantly, we simulate a CRDP trace with $\tau = 15\,\mu$s, and confirm that the CRLB limit for a CRDP frequency measurement of such a trace with an SNR=3000 to be at the sub-Hz level, corresponding to sub-ppm fractional uncertainties (i.e. a $\sim$ 100 mHz uncertainty on a measurement of a 2 MHz beating frequency). We also verify that for CRDP signals with short ring-down times, e.g. $0.5\,\mu$s, where one observes only a limited number of oscillations within the whole acquisition window of the ring-down signal, the results start to deviate from the CRLB limit.

#### 4.5.2 CW-CRDP experimental sensitivity limits

We proceed by analysing the sensitivity and stability of the experimental apparatus using an Allan-variance methodology, which allows us to study the temporal characteristics of the frequency measurements. In Fig. 12 we present an Allan plot of the CW-CRDP measurements of the polarization beat frequency induced by the Faraday effect in the $SiO_2$ substrate under the influence of a magnetic field from permanent magnets. We use two separate acquisition systems of different resolution and sampling rates to acquire and integrate CW-CRDP signals for over $\sim$1000 s, which enables us to examine the stability of our apparatus and the validity of the CRLB limit for the different acquisition conditions. Using time-domain analysis (Sec. III B), we estimate the uncertainty in our frequency measurements over a specific integration time. We observe that the system exhibits a white-noise like behaviour with a slope of approximately $35\,\mathrm{Hz}/\sqrt{\mathrm{Hz}}$, which corresponds to polarimetric sensitivities of approximately $448 \times 10^{-9}\,\mathrm{rad}/\sqrt{\mathrm{Hz}}$ per round-trip (the cavity's finesse is $\approx$1200), and this behaviour holds for integration

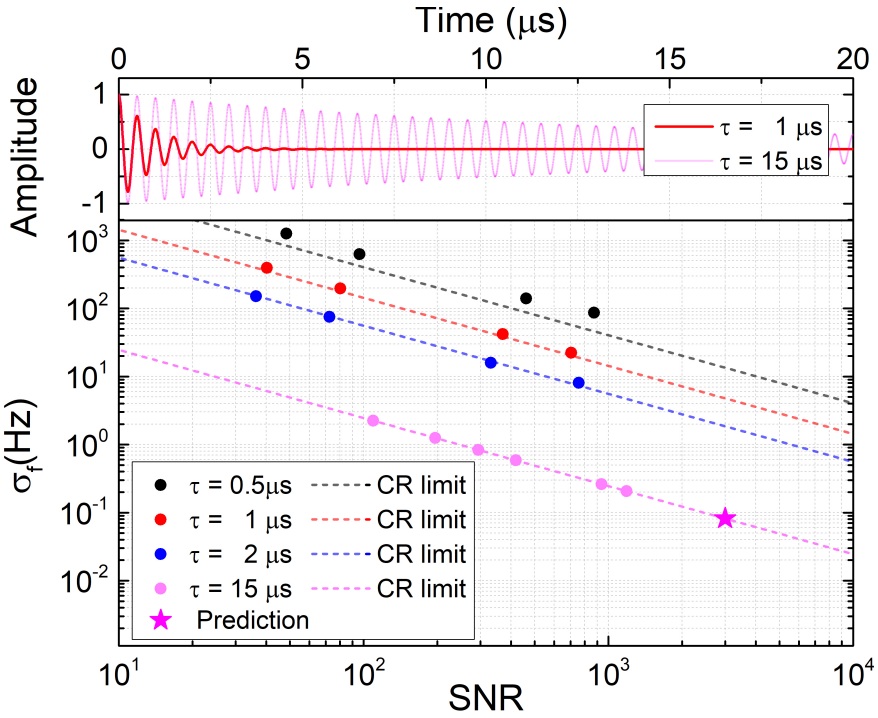


Figure 11: **Upper panel -** Simulated CRDP traces generated using an arbitrary waveform generator (8-bit resolution) for different ring-down times and a fixed beat frequency of 2 MHz, which are recorded using a 14-bit digitizer. **Lower Panel -** Frequency estimation errors obtained using a time-domain analysis of the simulated CRDP traces. The dashed lines are the estimation limits set by the Cramér-Rao lower bound limit (Eq. 7) for the specified decay times ($\tau$) and signal-to-noise ratios (SNR) of the simulated signals.

times of up to ∼30 s, where drifts associated with our frequency stabilization system do not allow for further sensitivity improvements. Most importantly, we observe that our experimental results are in agreement with Eq. 7 (we achieve a SNR≈ 700 within 1 s of integration time for repetition rates of 50 kHz). As a comparison we note that pulsed-CRDP techniques have achieved polarimetric sensitivities at the $\sim 10^{-6}$ rad levels for several minutes of integration time using optical cavities of similar finesse as the one we use here [16, 23], clearly demonstrating the benefits of employing CW-laser sources. In addition, using our experimental conditions, we predict the shot-noise limit for our CRD-based polarimetric measurements to be $\sim 10 \times 10^{-9}\,\mathrm{rad}/\sqrt{\mathrm{Hz}}$ (per round-trip).

To demonstrate the merits of CW-CRDP, in Fig. 12 we include for comparison the reported polarimetric sensitivities of two recent state-of-the-art PD-CRDS-related works. In particular, Westberg and Wysocki [69] using an optical cavity with $\mathcal{F} \approx 50000$ demonstrated Faraday optical rotation sensitivities of $\sim 1.3\times10^{-9}\text{rad}/\sqrt{\text{Hz}}$ (per cavity round-trip, for 1 s of integration time), at 762 nm. Similarly, Gianella et al. [48] using an optical cavity with $\mathcal{F} \approx 175000$, demonstrated almost identical Faraday optical rotation sensitivities of $\sim 1.24 \times 10^{-9}\,\text{rad}/\sqrt{\text{Hz}}$ (per cavity round-trip, for 1 s of integration time), at 1506 nm. Using our analysis on the fundamental frequency sensitivity limits of CRDP (see Fig. 11), we predict that a CW-CRDP experiment realized using a 0.60 m long two-mirror optical cavity with $\mathcal{F} \approx 11000$, and which yields signals with an SNR of 3000 within 1 s of integration time (we assume here only white noise), one can achieve Faraday optical rotation sensitivities of $\sim 1 \times 10^{-9}\,\text{rad}/\sqrt{\text{Hz}}$. Therefore, one can achieve similar state-of-the-art sensitivities with a CW-CRDP polarimetric approach using cavities of significantly lower finesse, opening the possibility for highly sensitive polarimetry to spectral regimes where high quality optics might not be available and/or material losses can be higher. Finally, it is important to emphasize that for our prediction, where we consider a reasonably attainable SNR for 1 s of integration time, we do not consider the case of correlated noise cancellations as a result of the balanced detection scheme [48, 69] or possible SNR improvements that arise from the application of rapid signal reversals [18, 23, 46].

#### 4.5.3 Alternative modalities for CW-CRDP

In the theory section (Sec. II) we discuss the principles of CW-CRDP for the case of a CW laser source with a linewidth that is significantly smaller than that of the optical cavity. For this case we show how the incident radiation can couple into and build-up within that cavity by exciting only one particular cavity polarization eigenmode at a time, and how this in turn defines the amplitude, phase, and modulation depth of the resulting CW-CRDP signals. From Fig. 5 we see that these parameters can change significantly as a function of laser detuning with respect to the $R$-$L$ mode splitting, and for this reason a robust frequency stabilization scheme is required to ensure minimal frequency deviations that result in signal phase fluctuations, which

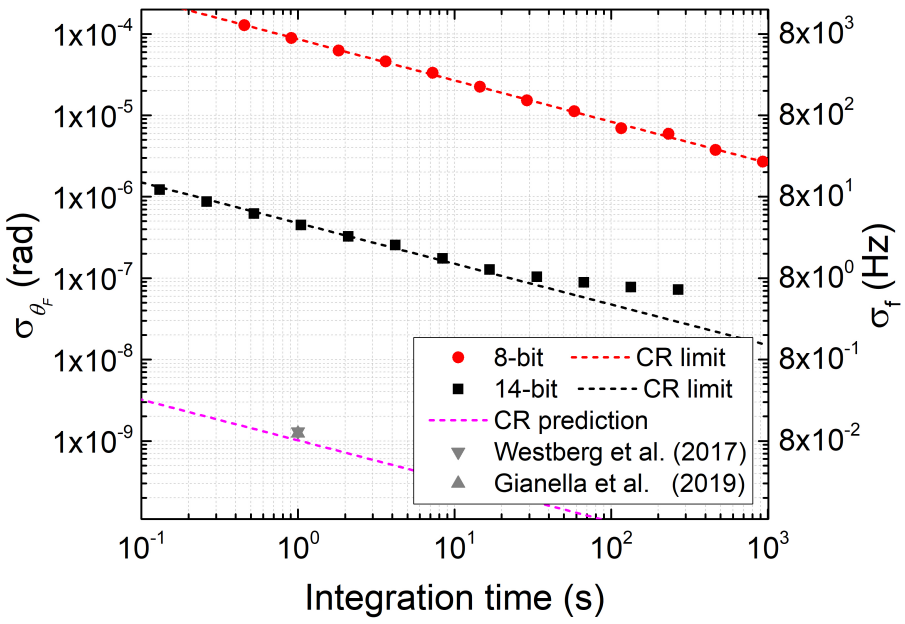


Figure 12: Allan variance of CW-CRDP polarization beat frequency measurements, $\sigma_f^2$, of the non-resonant Faraday effect of an $SiO_2$ substrate, using two different acquisition modalities (of 8-bit vs. 14-bit resolution). At short integration times white noise dominates and the variance decreases proportionally as the inverse of the square-root of the integration time. Mechanical drifts influence the stability of the measurements beyond an optimum integration time of $\approx 30$ s. The dashed lines (black, red, magenta) are the Cramér-Rao limits for our experimental parameters and for a predicted set of experimental parameters.

in turn hinder the sensitivity of a CW-CRDP measurement. A possible solution towards highly sensitive CW-CRDP instrumentation that does not require complicated electronics, is to use a single-frequency laser with a linewidth that is significantly broader than that of the cavity mode-splitting. In this case, the incident radiation excites both modes coherently and the frequency-stabilization scheme can be replaced by an intensity threshold comparator-circuit that switches off the laser radiation once a transmission threshold is recorded, similarly to what has been done in CW-CRDS [50]. For example, for optical cavities as the one we present here and similar polarization mode-splittings of the order of 1-2 MHz, the laser light

linewidth should be at least an order of magnitude larger (i.e. 10-20 MHz). Most importantly, using this approach one is not anymore limited by the finesse of the cavity and the desired mode-splitting can be significantly larger than the cavity linewidth, contrary to the case of a CW laser source with a narrow linewidth for which one must generate a mode-splitting of the order of the cavity linewidth for optimal mode coupling. Free-running diode and distributed-feedback lasers typically have linewidths of <5 MHz, while the design approach for commercially available external cavity diode lasers is to reduce the diode-laser linewidth as much as possible (with current linewidths at the sub-100 kHz range). A direct method to achieve the desired broad linewidth is to broaden the linewidth of the laser source in a controllable fashion by, e.g., modulating the laser current using white noise with high bandwidth. This method is, in principle, applicable with both diode and distributed-feedback lasers (however, diode-based lasers might face the problem of mode fragmentation due to the noise modulation). There exist several additional modulation approaches that are suitable for CW-CRDP, such as the case of modulating the laser current at the frequency of the mode-splitting, and we will explore these in future works. However, we consider the approach of broadening the linewidth of the laser source to be larger than the mode-splitting, as the most suitable one for portable spectropolarimetric instrumentations.

An alternative option is to implement an optical-feedback technique, whereby the circulating wave within the cavity is allowed to return to the laser source to injection-seed it, inducing, thus, a direct frequency-locking without the need for the generation of an error signal and the associated feedback electronics. This is the principle behind optical feedback cavity-enhanced absorption spectroscopy (OF-CEAS) [70], which has already been employed in CRD-based spectroscopic [71] and polarimetric [25] measurements. In this case, the linewidth and central frequency of the laser source are effectively matched and (phase-)stabilised to that of the cavity, respectively, allowing for optimal injection and the reduction of phase-fluctuations related to frequency drifts of either the laser or the cavity.

### 4.6 Conclusions

In this paper, we present the general principles of CW-CRDP, a CRD-based polarimetric scheme that employs CW laser sources and benefits from the use of large intracavity polarization anisotropies for highly sensitive polarimetric measurements. Contrary to PD-CRDS polarimetric approaches that are effectively intensity-based measurement schemes (since these monitor changes in the ring-down decay time), CRDP schemes are frequency-based measurements where sensitive detection of shifts in their induced polarization beat frequency enables highly sensitive polarimetric measurements. CW-CRDP builds upon the measurement methodologies that have been developed within the context of chiral pulsed-CRDP [15, 18, 23], and we show how the use of CW laser sources allows for significant gains in signal intensity and data acquisition rates, enabling shot-noise limited measurements that are not easily attainable using pulsed laser sources. Furthermore, we discuss in depth the fundamental limits of CRDP protocols, and demonstrate how CW-CRDP instruments can perform highly sensitive measurements even with the use of optical cavities with modest finesse. We confirm the principles and merits of CW-CRDP by measuring non-resonant Faraday optical rotations from solid and gaseous samples using a prototype design.

In conclusion, CW-CRDP is a powerful and versatile technique suitable for the measurement of reciprocal and non-reciprocal birefringence and dichroism, linear and circular. We believe that CW-CRDP is the ideal modality for portable spectropolarimetric instrumentations as it allows for time-resolved, highly sensitive, and cost-effective operation at a broad spectral region. Future work will focus on improvements in system design and sensitivity, and on the possibility of coupling CW-CRDP with chromatographic techniques, towards the development of a portable instrument suitable for breath analysis and monitoring in clinical settings [72, 73], and for trace gas analysis and monitoring of paramagnetic species in field settings [46].

# 5 Rapid parameter determination of discrete damped sinusoidal oscillations

The following chapter is a direct adaptation of a peer-reviewed and published work. It can be found as:

**Visschers JC**, Wilson E, Conneely T, Mudrov A, Bougas L. Rapid parameter determination of discrete damped sinusoidal oscillations. Optics Express. 2021 Mar 1;29(5):6863-78.

The minimal changes made to this version of the manuscript compared to the published work limit themselves to formatting and style alterations. The conceptualization and implementation of the Prony method described are not mine and are attributed to E. Wilson, T. Conneely and A. Mudrov. The work comparing the three analysis methods and the interpretations of the results presented in this work is mine.

## 5.1 Introduction

Precise and rapid signal-parameter estimation is important for both fundamental and applied research, and becomes particularly crucial when observing and controlling fast processes in real time (e.g., chemical reactions), and in the development of portable instrumentation where fast, real-time, data streaming and inspection is essential.

Several different research fields rely on the precise and accurate extraction of the time constants and frequencies of damped sinusoidal signals. Prominent examples include: nuclear magnetic resonance (NMR) [74], where information on the structure and the spin environment of a target molecule is extracted from precise determination of the frequency and decay constant of a damped sinusoidal signal; free-induction-decay (FID) optical magnetometry [66, 68, 75–78], where the magnetometric sensitivities depend on the precision of the measurement of the oscillating frequency; and pulsed/continuous-wave cavity ring-down polarimetry (CRDP) [15, 16, 18, 23, 26–28, 79] and ellipsometry (CRDE) [57, 80–82], where polarization-dependent absorption and refraction/reflection through/by an optical medium is extracted with high sensitivity through the pre-

cise measurement of the signal-decay time and its polarization beat frequency.

A distinction among the aforementioned examples can be made according to their respective decay constants and oscillating frequencies. In routine NMR, typical decay times are in the $10^{-2}-10\,\mathrm{s}$ range, while frequencies are in the $10-800\,\mathrm{MHz}$ range; especially, portable NMR instruments operate in the $10-30\,\mathrm{MHz}$ frequency range [83–86]. In FID optical magnetometry, typical decay times are in the $10^{-2}-1\,\mathrm{s}$ range, while frequencies are within the $10^{2}-10^{5}\,\mathrm{Hz}$ range (see for instance [75, 77]). In CRDP/CRDE demonstrations, however, decay times are typically in the $10^{-7}-10^{-5}\,\mathrm{s}$ range, while polarization beat frequencies are in the $1-100\,\mathrm{MHz}$ range.

For all these applications, significant data processing is typically required to determine the signal parameters, and, in general, experimental sensitivity is improved by averaging over many measurement runs. As such, when developing portable instruments one needs to appropriately select the instrument's sampling and acquisition rates, but also carefully consider the computational cost, i.e. the calculation time, to analyze each acquired signal. For applications where the relevant time-scales are relatively long (10 ms - 1 s), such as NMR or FID magnetometry, there are several options that can provide precise results sufficiently fast (with respect to "single-events"), such as, e.g., frequency counters (see [87, 88] and references therein). However, in applications where the relevant time-scales are much shorter than a few ms, as in the case of CRDP/CRDE, acquisition and computational speeds ultimately define, respectively, the measurement and analysis repetition rates.

It is instructive to consider here a general data acquisition system,

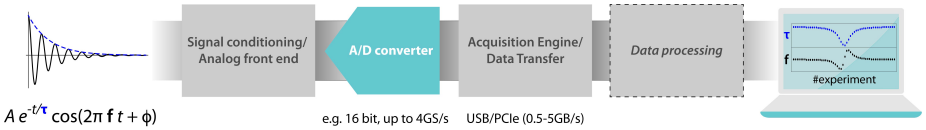


Figure 13: *Data acquisition block diagram* - a damped sinusoidal, physical, signal acquired by an experimental apparatus is digitized using an analog-to-digital (A/D) converter and then transferred to a data processing unit, which estimates the signal parameters and displays them on screen for real-time monitoring.

as depicted in Fig. 13: such a system collects the physical signal, in our case a damped sinusoidal signal, through an analog front-end

that typically performs some signal conditioning (e.g. analog filtering, signal amplification). The acquired signal is subsequently digitized using an analog-to-digital converter and transferred to a data processing unit for parameter analysis. A particular case example is the field-programmable gate array (FPGA), whose development has lead to the emergence of stand alone multi-channel (e.g., four) high-precision (e.g., 16-bit) data acquisition systems with high sampling (e.g., 2 GS/s) and triggering rates (e.g., 1 MHz), which have nowadays become commercially available at cost-effective rates. In such systems, data are (typically) transferred via USB or PCIe interfaces and, as such, data transfer rates as high as 5 GB/s are feasible (e.g. Teledyne SP Devices, ADQ series [4]). Using such systems, therefore, it is possible to perform (sub-)microsecond-resolved CRDP/CRDE measurements [57]. However, the principal limiting factors towards an online, real-time, processing system that needs to be operable in a *running* fashion with no dead-times, is the data-processing module of the overall data acquisition system (Fig. 13) and its limited memory storage capacity. Considering that in CRDP/CRDE experiments demonstrated decay time constants are in the $10^{-7} - 10^{-5}$ s range, it is important to identify appropriate computational approaches that can be implemented in data acquisition systems, such as FPGA-based digitizers, to allow for online, real-time, signal analysis at such fast time scales.

Time- and frequency-based computational methodologies for rapid parameter estimation have been developed within the context of CRD spectroscopy, and these have been evaluated and compared in terms of their speed and precision [89–91]. Most notably, Fourier transform methods have been implemented on FPGAs for fast analysis of exponentially decaying signals [92], with demonstrated analysis rates as high as 4.4 kHz [93]. However, while several works discuss the performance of various time- and frequency-domain analysis algorithms for damped sinusoidal signals [94, 95], a direct comparison between their attainable precision and computational speed is currently missing.

In this work we compare three specific analysis methods of discretely sampled damped sinusoidal signals in terms of their speed, and attainable accuracy and precision. These methods are: (a) a time-domain least-squares analysis based on a Levenberg-Marquardt

[4] https://www.spdevices.com

algorithm [96]; (b) a frequency-domain analysis based on a fast Fourier algorithm [90, 91, 97, 98] in combination with a quadratic interpolation of the frequency components of the resulting Fourier transform [81]; and (c) a time-domain analysis based on the Prony method [31]. We evaluate their efficacy in terms of the signal's parameters, and discuss how each of these affect the sensitivity limits for each computational methodology. Finally, we present an experimental, proof-of-principle, demonstration of the capabilities of such methods for the online analysis of CRDP signals.

## 5.2 Theory

### 5.2.1 Damped sinusoidal signals

A damped sinusoidal signal can be characterized in terms of a model function as:

$$y(t) = A \cdot e^{-t/\tau} \cdot \cos\left(2\pi \cdot \mathrm{f} \cdot t + \phi\right) + y_0, \tag{10}$$

where $t$ is the (discretely sampled) independent (time) variable of the signal, $A$ is the amplitude of oscillation, $\tau$ is the characteristic decay time, f and $\phi$ the frequency and phase of the oscillation, respectively, and $y_0$ is a global signal offset. Under realistic experimental conditions, all the signal parameters will be time-dependent, and the power spectral density of the signal will be proportional to their respective noise contributions. Here, for simplicity, and to clarify the main results of our findings, we assume that $\tau$ and f are constant parameters and restrict the investigation of noise contributions to the global offset parameter, i.e. $y_0(t)$, which we assume to be normally distributed [$\langle y_0(t)\rangle = 0$, $\langle y_0^2(t)\rangle = \sigma_{y_0}^2$].

In Fig 14 we show an example of a discretely sampled damped sinusoidal signal. For the analysis of such a signal we consider four key parameters that affect the expected precision and accuracy: a) the number of signal oscillations per typical decay time, $\mathrm{f} \times \tau$; b) the number of samples per typical decay time, $n \times \tau^{-1}$ (i.e. the sampling rate); c) the number of decay times measured in a measurement time window $T_\mathrm{m}$, $T_\mathrm{m} \times \tau^{-1}$; and d) the signal-to-noise ratio, defined as $\mathrm{SNR} = A \times \sigma_{y_0}^{-1}$.

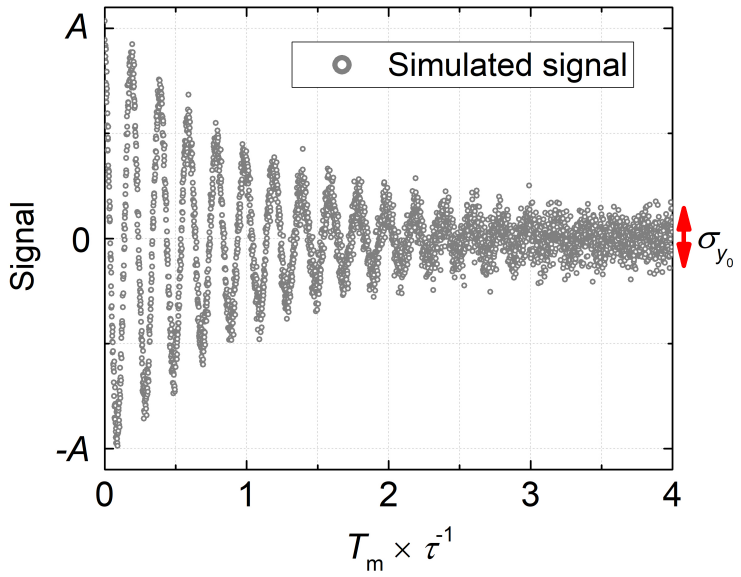


Figure 14: Example of a discretely sampled damped sinusoid as described by Eq. 10, for $f \times \tau = 5$, $n \times \tau^{-1} = 200/\tau$, and $SNR = 2^4$.

#### 5.2.2 Cramér-Rao lower bound

The fundamental limit for the statistical uncertainty of determining the oscillating frequency of a damped sinusoidal signal (Eq. 10) is described by the Cramér-Rao lower bound (CRLB) [65, 66], which sets the lower limit on the variance $\sigma_f^2$ of any frequency estimator. The CRLB condition for the frequency extracted from a discrete damped sinusoid is given by [65–68]:

$$\sigma_f^2 = \frac{6}{(2\pi)^2 \, SNR^2 \, f_{BW} \, T_m^3} \, \chi\left(\tau/T_m\right), \tag{11}$$

where SNR is the signal-to-noise ratio of the signal; $f_{BW}$ is the sampling-rate-limited bandwidth of the measurement; $T_m$ is the measurement time window; and $\chi\left(\tau/T_m\right)$ is a correction factor that takes into account the signal decay, which is given by:

$$\chi(r) = \frac{e^{2/r} - 1}{3r^3 \cosh\left(2/r\right) - 3r\left(r^2 + 2\right)}. \tag{12}$$

The factor $\chi\left(\tau/T_m\right)$ serves as a compensation factor in Eq. 11 that penalizes measurement of the tails of the exponential decay when the signal has effectively died out. Equation 11 remains valid under

the condition that the period of the oscillation is much shorter than the decay time of the signal and that a sufficient number of oscillations occurs in it. Moreover, Eq. 11 dictates that any noise sources affecting the signal detection are contributing to the fundamental CRLB limit through their effect on the SNR of signal.

In Ref. [79], the authors demonstrate that the CRLB limit is the appropriate estimator of the fundamental sensitivity of frequency-based measurements within the context of CRDP, as the frequency measurements are directly translated into polarimetric results. However, one needs to carefully investigate whether different signal processing techniques can approach the CRLB, and if yes, under what conditions this is possible. Moreover, considering our motivation is the development of a portable CRDP instrumentation operating with similar principles as recent demonstrations of it [57, 79], we focus on investigating and comparing different signal processing approaches in terms of their speed and attainable accuracy and precision for damped sinusoidal signals with decay times in the range of $1 - 10\,\mu s$ and frequencies in the range of $1 - 10\,\mathrm{MHz}$.

#### 5.2.3 Signal analysis

##### 5.2.3.1 Least-squares estimation of nonlinear parameters

For time-domain analysis we focus on an optimized least squares curve fitting approach based on the Levenberg-Marquardt algorithm (LMA) [96]. The algorithm minimizes the sum of the squared residuals,

$$S = \sum_{i=1}^{n} \left[ y_i - f\left(t_i, \beta\right) \right]^2, \tag{13}$$

where $y_i$ is the $i^{th}$ sample of the discretized recorded signal $y(t)$ (Eq. 10), $t_i$ is the $i^{th}$ time sample, and $f(x, \beta)$ is the non-linear fit function (Eq. 10) with $\beta$ representing the guess fitting parameters for $\{A, \tau, \mathrm{f}, \phi, \mathrm{y_0}\}$. The LMA algorithm iteratively finds the optimal guess parameters $\beta$ describing the recorded signal $y$.

The LMA is, in itself, an efficient algorithm, but it relies heavily on the initial guess parameters of the iterative process. However, we wish to identify the precision and speed limitations of computational implementations of a least-squares algorithm and, hence, we assume for our computational investigations that the initial conditions are

well-defined and known in advance (with our experimental investigation we examine the dependence of the LMA algorithm on the initial guess parameters under realistic conditions; see Sec. 5.4.3). Furthermore, the time required for the convergence of a fit using LMA is highly dependent on the platform used. In this work we choose to work with a CPU-based code for the implementation of the LMA that employs a Python optimized package (SciPy) based on the MINPACK library [99].

#### 5.2.3.2 Fast Fourier transform

For frequency-domain analysis we use a fast Fourier transform (FFT) algorithm, as introduced by Cooley and Tukey [97], to calculate the discrete Fourier transform (DFT) of the signal,

$$\mathcal{F}(k) = \sum_{n=0}^{N-1} y_n \, e^{-\frac{2\pi i}{N} n \, k}, \quad k = 0, \ldots, N-1, \tag{14}$$

where $y_n$ is the $n^{\text{th}}$ sample of the discretized time-domain signal $y(t)$ (Eq. 10). We note here that the simplest and most common implementations of the FFT algorithm introduced by Cooley and Tuckey assume that $N$ is a power of two.

The Fourier transform of a monochromatic damped sinusoidal signal corresponds to a single Fourier (frequency) component with a spectral width inversely proportional to the signal's decay time. Our aim is to estimate accurately and precisely the central value of this component, rapidly. One approach is to perform a least-squares curve fitting on the resulting FFT spectrum to obtain the central value of the frequency component and its width. However, the accuracy and precision of such process depends strongly on the curve fit-model selected and its initial guess parameters, but, importantly, the speed of such an approach would be at least equal to the overall time required to perform both the FFT and the least-squares fitting. Furthermore, in order for such an approach to be as precise as the direct time-domain analysis approach using, e.g., LMA, one typically employs additional data manipulation techniques (e.g. zero-padding, apodization).

Here, we focus on algebraic approaches for the rapid extraction of the central value of the Fourier (frequency) component from the FFT spectrum. One such approach is to determine the center value of

this component by considering the three closest neighbouring points to the maximum frequency value (peak): $(k_i, \mathcal{F}(k_i)) \equiv (k_i, b_i)$, $(i = 1, 2, 3)$, and use a quadratic estimator to find $\mathrm{f}_{\max}$ as:

$$\mathrm{f}_{\max} = \frac{k_1^2 b_1 (b_3 - b_2) + k_2^2 b_2 (b_1 - b_3) + k_3^2 b_3 (b_2 - b_1)}{2 [k_1 b_1 (b_3 - b_2) + k_2 b_2 (b_1 - b_3) + k_3 b_3 (b_2 - b_1)]}. \quad (15)$$

It is important to emphasize that the selection of the neighbouring points is crucial for the accuracy (not the speed) of the frequency estimation using such an approach. For high sampling rates, for instance, one can choose - symmetrically, or even asymmetrically - points further away from the closest neighbouring points to the peak, and preferably points lying near to the half-maximum of the Fourier component [this can be easily pre-set in the algorithm if the decay time and the sampling rate are (approximately) known in advance]. Such an algebraic approach on analysing FFT spectra has already been successfully implemented for rapid frequency estimation in CRDP-based experiments (see [81]). By choosing such an approach, we ensure that the computational speed remains as close as possible to the speed required to employ a FFT algorithm.

There exist several CPU-based codes available for FFT analysis, but for an appropriate speed comparison between the alternative signal processing methodologies presented in this work, we use a DFT algorithm directly from a Python-based scientific environment (here NumPy; see also discussions in the Appendix).

#### 5.2.3.3 Prony method

The Prony method, is a time domain approach originally designed for processing discrete time signals that are superpositions of damped sinusoids. The Prony method is closely related to the Matrix Pencil method (both estimate the signal as a sum of complex exponentials) [100, 101], the latter being used in NMR analysis [102, 103]. However, Prony analysis takes a polynomial approach in parameter (frequencies and damping factors) estimation whereas Matrix Pencil Method locates the signal parameters by finding the eigenvalues to a matrix pencil.

The application of the Prony method follows in three steps: (a) an autoregressive model is built employing discrete measurements; (b) the roots of the characteristic polynomial for the corresponding

finite difference equation are statistically estimated; and (c) estimates of the parameters of the signal are derived from the roots.

For the special case of one damped sinusoid, a discrete time sampling of such a signal gives rise to an autoregressive model of order 3 where the measurement $y_k$ at time $k$ is expressed through 3 preceding measurements in a linear way:

$$y_{k+3} + \alpha_2 y_{k+2} + \alpha_1 y_{k+1} + \alpha_0 y_k = 0, \tag{16}$$

where $k$ varies from 0 to $n+2$, and $n+5$ is the number of measurements (sampling points). The coefficients $\alpha_i$ are determined by any of the linear systems

$$\begin{bmatrix} y_{k+2} & y_{k+1} & y_k \\ y_{k+3} & y_{k+2} & y_{k+1} \\ y_{k+4} & y_{k+3} & y_{k+2} \end{bmatrix} \cdot \begin{bmatrix} \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{bmatrix} = - \begin{bmatrix} y_{k+3} \\ y_{k+4} \\ y_{k+5} \end{bmatrix}, \tag{17}$$

with $k = 0, \ldots, n$.

In the presence of noise, the $3 \times 3$-matrix in the left-hand side and the 3-vector in the right-hand side are random, so the coefficients $\alpha_i$ can be found, e.g., by the least square method minimising the loss function

$$\sum_{k=0}^{n} (y_{k+3} + \alpha_2 y_{k+2} + \alpha_1 y_{k+1} + \alpha_0 y_k)^2.$$

These constitute the characteristic polynomial equation

$$q(z) = z^3 + \alpha_2 z^2 + \alpha_1 z + \alpha_0, \tag{18}$$

whose roots $u_0, u_\pm$ incorporate the parameters of the signal. In the special case under study, $u_0 = e^{-\frac{\Delta}{\tau}}$, $u_\pm = e^{-\frac{\Delta}{\tau}} e^{\pm i 2\pi \mathrm{f} \Delta}$, where $\Delta$ is the sampling time interval. Then the frequency and decay constant of the signal are found as

$$\tau = -\ln(u_0)/\Delta\,, \tag{19}$$

$$\mathrm{f} = \mathrm{Im}[\ln(u_+)/\Delta]. \tag{20}$$

The roots can be calculated, e.g., by the Cardano formulas, or by employing $\alpha_0 = -u_0^3$ (in the limit of no noise). The root $u_+$ can be distinguished from $u_-$ as the one with positive imaginary part if $u_+ + u_- > 0$, and negative otherwise.

A practical realization of this scheme has to take into account the role of the sampling rate $n \times \tau^{-1} = \frac{1}{\Delta}$. Even in the absence of noise,

there exist singular values at $n \times \tau^{-1} = \frac{\mathrm{f}}{\pi N}$ (with positive integer $N$) for which the matrix in Eq. 17 becomes degenerate (degeneracy occurs for half-integer $N$), and the sampling rate has to be chosen to be different from such singular values. Another thing is that the coefficients $\alpha_i$ depend on f via $\cos(\mathrm{f}\Delta)$. Keeping the sampling rate above $\mathrm{f}_{max}\Delta$ confines $\mathrm{f}\Delta$ in $[0, \pi]$ and determines f from Eq. 20 uniquely.

Furthermore, in the presence of noise, the accuracy still depends on $n \times \tau^{-1}$ even if it exceeds $\frac{\mathrm{f}}{\pi}$. So, if the sampling rate is too high, the sampled points are too close to each other (note that their number is fixed), and small variations of the signal (a smooth function) from point to point are distorted by random jumps which deteriorate estimation. Therefore the frequency should be bounded from below, say with $\mathrm{f}_{min}$. In practice, for a reasonable SNR the dependence of the result on the sampling rate is weak in a wide range of $n \times \tau^{-1}$ values, and this observation can be used for estimation.

## 5.3 Methods

### 5.3.1 Signal simulation

To compare the three methods of analysis on their respective precision and accuracy in estimating the central frequency of damped sinusoidal signals (Fig. 14), we generate and analyze sets of 500 such signals on a homemade Python CPU-code on a Windows 10 workstation [CPU: AMD Ryzen 7 2700, RAM: 16.0 GB 1330 MHz DDR4]. All simulated signals have the following non-changing parameter values: $A = 1$, $\langle y_0 \rangle = 0$, and $\phi = 0$. We also choose the following baseline values for the key parameters of each simulated signal: $\mathrm{f} \times \tau = 5$, $n \times \tau^{-1} = 1000/\tau$, $T_\mathrm{m} \times \tau^{-1} = 5$, and $\mathrm{SNR} = 2^{12}$. We choose here a high baseline value for the SNR to clearly examine whether the computational approaches can reach the fundamental CRLB limit as a function of the other key signal parameters. We proceed by varying each key parameter over several orders of magnitude while keeping the other parameters at their baseline value, to explore the dependence of the precision and accuracy of each computational approach on these parameters.

#### 5.3.2 Precision and accuracy

As a way to quantify the precision of each computational approach we use the standard deviation from the distribution of frequency values obtained through the analysis of the 500 simulated signals, i.e., $\sigma_{\rm f}$, to estimate the fractional uncertainty $\sigma_{\rm f}/{\rm f}$ (i.e. smaller fractional uncertainty corresponds to higher precision). Similarly, we define the accuracy of a method as:

$$\text{accuracy} = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \frac{|{\rm f}_{i,\rm est} - {\rm f}_{\rm act}|}{{\rm f}_{\rm act}}, \tag{21}$$

where ${\rm f}_{i,\rm est}$ is the frequency estimated by the analysis method for a single signal, ${\rm f}_{\rm act}$ is the actual (input) frequency of the simulated signal (again here, $\mathcal{N} = 500$). An analysis method is predicted to have no bias as long as the accuracy of its frequency estimation falls within the precision of the estimation.

#### 5.3.3 Computation time

We determine the speed of each computation method by estimating the time required to analyze a single signal using the internal timing functions of the Python software (e.g., function *timeit*).

### 5.4 Results

#### 5.4.1 Precision and accuracy

In Fig. 15 we present results on the precision and accuracy achieved when analyzing discrete simulated damped sinusoidal signals using the least-squares, FFT, and Prony computation analysis methods, as a function of varying signal conditions. In particular:

***Frequency -*** In Fig. 15 (a) we show a comparison of the attainable precision and accuracy between the three different approaches as a function of the frequency of oscillation. For all frequencies, the least-squares approach results in optimal accuracies compared to the other approaches, with the FFT approach being consistently less accurate (this is largely related to the peak-finding algebraic methodology we employ here). The Prony method becomes particularly inaccurate for low frequencies, which is related with the computational formulation of the Prony method that doesn't allows

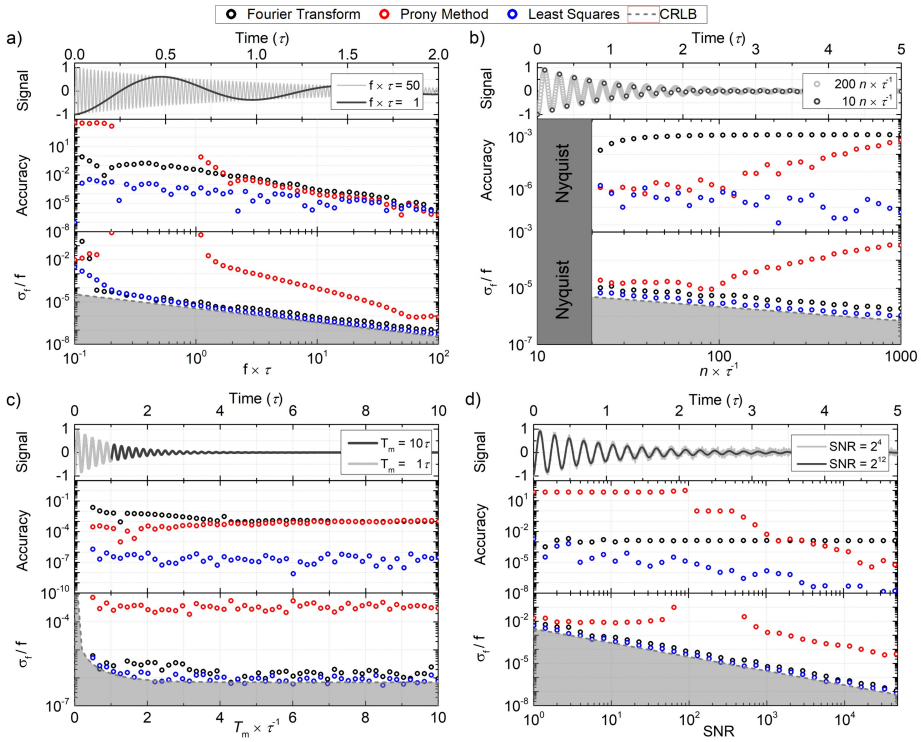


Figure 15: Results of attainable fractional uncertainty, $\sigma_{\mathrm{f}}/\mathrm{f}$, and accuracy using the FFT (black points), Prony (red points), and least-squares (blue points) analysis methods as a function of varying signal conditions: **(a)** signal frequency, $\mathrm{f}\times\tau$; **(b)** sampling rate, $n \times \tau^{-1}$; **(c)** measurement time-window, $T_m \times \tau^{-1}$; and **(d)** signal-to-noise-ratio (SNR). In combination with the performance of each method, the fundamental frequency estimation limit given by the Cramér Rao lower bound (CRLB) is also shown (dashed gray line). The least-squares method yields results close to the CRLB limit in virtually all situations, with the FFT method yielding similar results, while the Prony yields poor results under most selected conditions.

us to investigate a large parameter space without approaching singular points in the analysis. In terms of precision, both the least squares and FFT methods approach closely the CRLB, the latter being approximately a factor of three less precise than the former, while for both methodologies the precision is not influenced by the frequency value (the least squares method deviates from the CRLB at frequencies $\mathrm{f} \times \tau < 0.2$, as expected, since the observed time window does not contain a full period of oscillation). The Prony method yields results with poor accuracy for $\mathrm{f} \times \tau < 1$ and does not reach the expected precision limits for the whole simulated fre-

quency range.

***Sampling rate -*** In Fig. 15 (b) we show a similar comparison as a function of the sampling rate, i.e., as a function of the number of sample points per decay time $n \times \tau^{-1}$. We note again here that for these estimations we choose a constant frequency of $\mathrm{f} \times \tau^{-1} = 5$. As such, the Nyquist criterion limits the lowest sampling rate for a sensible frequency estimate to $n \times \tau^{-1} = 20$. In terms of accuracy and precision the least squares method yields optimal results, while the FFT method yields optimal precision but relatively poor accuracy (at the $10^{-3}$ level), both related to the peak-finding algorithm (these can be improved by performing additional signal manipulation, such as zero padding, but this will significantly affect the computational speed). The Prony method provides accurate and precise results for low sampling rates, but these deteriorate for high sampling rates (see Sec. 5.2.3.3). We also observe that the least squares method is limited by the CRLB over the entire range of sampling range we investigate, while the FFT method remains consistently less precise.

***Measurement window -*** In Fig. 15 (c) we present results as a function of the measurement window, i.e. $T_{\mathrm{m}} \times \tau^{-1}$. The least-squares method yields optimal accuracy ($\sim 10^{-7}$) and precision ($\sim 10^{-5}$) results for $T_{\mathrm{m}} \times \tau^{-1} > 2$, however, for short measurement windows the precision becomes poor [as predicted by the CRLB limit, Eq. 11]. The FFT method reaches its optimum accuracy ($\sim 10^{-4}$) and precision for $T_{\mathrm{m}} \times \tau^{-1} \approx 4$. Importantly, we observe that both the least-squares and FFT methods reach the CLRB limit for $T_{\mathrm{m}} \times \tau^{-1} > 4$, with an optimal measurement window for precise signal analysis using both methods to be $\sim 5\tau$. Similar conclusions have already been reported in [91], suggesting that $\sim 5\tau$ can be considered the optimum repetition rate for, e.g., CRDP/CRDE experiments, as compared to the longer acquisition windows ($T_{\mathrm{m}} \times \tau^{-1} > 5$) typically required in traditional CRD spectroscopy [104]. The Prony method reaches similar accuracies as the FFT method but its precision is two orders of magnitude larger than the predicted CRLB limit.

***SNR -*** The final key parameter we vary is the signal's SNR, with the results seen in Fig. 15 (d). The least-squares analysis yields results close to the CRLB limit, while the precision attained using FFT analysis method is approximately a factor of two ($\times 2$) higher. Notwithstanding, we see that for an optimum measurement time-

window of $5\tau$ [Fig. 15 (c)] and a SNR $\approx 2^{12}$ both the least-squares and FFT methods yield precisions at the $\sim 10^{-6}$ levels. In contrast, the Prony method does not provide reliable frequency estimates for signals with SNR $< 500$, while, for higher SNRs, the attainable precision is two orders of magnitude above the CRLB limit.

#### 5.4.2 Speed

In Fig. 16 we present results on the dependence of the fitting (computation) time for each method on the number of data points in a single damped sinusoidal signal, which we also compare with the attainable precision for each case. For these simulations, we use the results presented in Fig. 15 to choose optimum values for the signal's key parameters: $\mathrm{f} \times \tau = 5$, SNR $= 2^{12}$ and $T_m \times \tau^{-1} = 5$ (such values are also realistically attainable in experiments; see discussions in [79]).

Overall we observe a non-linear increase in the calculation time as the number of samples is increased, with the least-squares and Prony algorithms being more than an order of magnitude slower than the FFT+peak-finding algorithms. In addition, while the least-squares method reaches the CLRB limit, the FFT method yields fractional uncertainties approximately a factor of two larger than the predicted CRLB limit, with the Prony method practically never reaching optimal precision levels. Most importantly, we observe that for a discrete signal with $\sim 10^3$ sample points, using the FFT+peak-finding algorithm one can achieve ppm sensitivities ($10^{-6}$ fractional uncertainties) for computational times of $\sim 200\,\mu\mathrm{s}$. Under the same conditions, the least-squares algorithm requires approximately $\sim 5\,\mathrm{ms}$ to reach similar fractional uncertainties. In addition, we observe that under similar conditions the Prony method results in poor sensitivities ($10^{-4}$ fractional uncertainties) requiring long ($> \mathrm{ms}$) computational times.

#### 5.4.3 Experiment

As a proof-of-principle demonstration for the capabilities of our analysis methodology we use a CRD-based polarimetric instrument we have developed in our laboratory for the attainment of experimental CRDP, i.e. damped sinusoidal, signals (see for details [79]).

Briefly, the ring-down cavity of the instrument has a total length

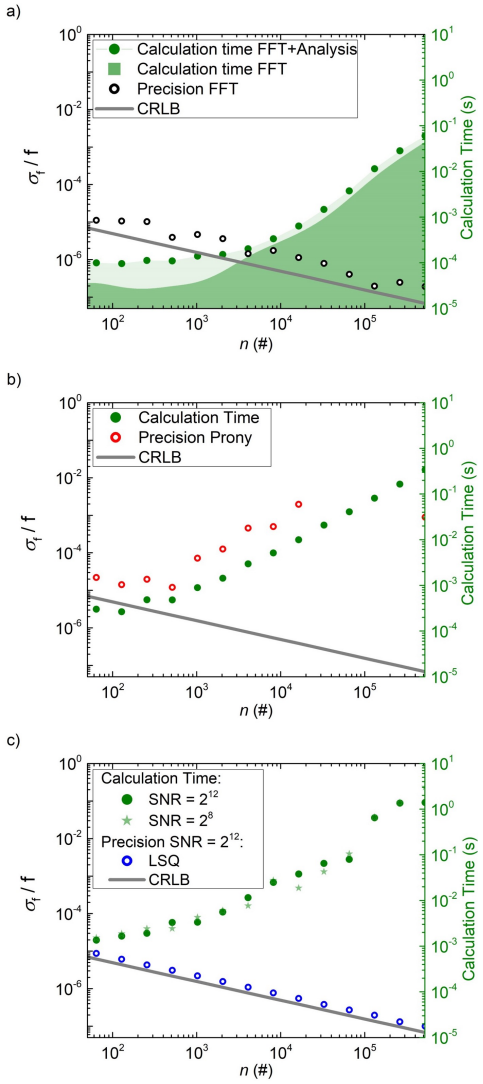


Figure 16: Dependence of precision and calculation time (green points) on the number of samples for the: **(a)** FFT (black points), **(b)** Prony (red points), and **(c)** least-squares (blue points) evaluation methods. In all cases, we compare the obtained results with the expected Cramér-Rao lower bound limit (CRLB; solid gray line) for increasing sample size $[n\,(\#)]$ of simulated signals with $\mathrm{f} \times \tau = 5$, $\mathrm{SNR} = 2^{12}$ and $T_m \times \tau^{-1} = 5$. Using the FFT algorithm methodology, for a signal with $10^3$ data points we obtain fractional uncertainties at the sub-$10^{-5}$ levels within a computational time of $\sim 200\,\mu$s, while for the same conditions using a least-squares approach we obtain similar precisions but for a computational time of $\sim 5$ ms. In overall, the Prony method results in poor sensitivities and with $>$ ms computational times.

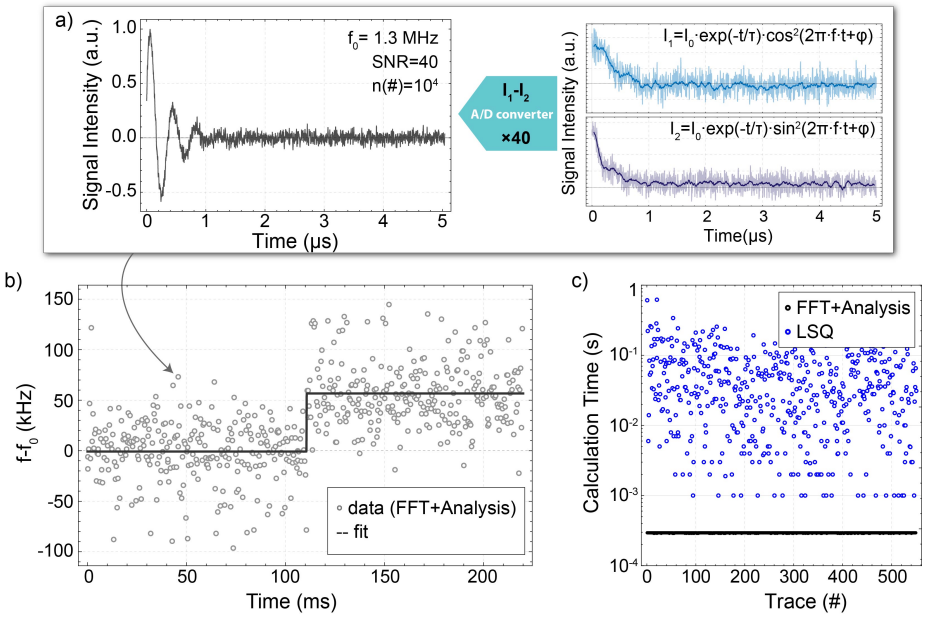


Figure 17: *Rapid analysis of experimental signals:* **(a)** CRDP experimental signals showing polarization beat frequencies generated via the Faraday effect on a $SiO_2$ substrate, as recorded by two orthogonal channels of a linear balanced polarimeter (right side). By subtracting these and averaging 40 consecutive traces, at a repetition rate of 100 kHz, we acquire a CRDP trace (i.e. damped sinusoidal oscillation) with a SNR=40 within 400 µs. **(b)** Direct observation of frequency shifts with respect to the (bias) Faraday polarization beat frequency [$f_0 = 1.3$ MHz], as a function of an externally applied, rapidly pulsed, magnetic field: each frequency (measurement) point is the result of an online FFT analysis of a CRDP trace, as the one shown in (a), and requires $\sim 300$ µs of calculation time [(c)]. The (black) line is the result of a nonlinear least-squares regression analysis used to fit a sigmoid function to the data, demonstrating that we can resolve sub-µs dynamics in a running-fashion. Note that the CRLB estimation for such signals (with $10^4$ data points, SNR=40, and $\tau \approx 0.33$ µs) is $\sim 1.3$ kHz (Eq. 11), and the observed scattering is associated with experimental noise. **(c)** Computational time required to perform the FFT+peak-finding analysis algorithm is $\sim 300$ µs per trace and remains constant during the analysis of the complete set of traces, while a least-squares approach requires computational times ranging from $\sim 1$ ms $- 1$ s (since least-square fitting is highly dependent on the initial fit parameters).

of 0.60 m and consists of two concave mirrors with radii of curvature of 1 m and specified reflectivity R∼99.9% at 408 nm (Five-Nine Optics). We use a single-frequency CW laser source (Toptica DL-PRO; $\lambda = 408$ nm) that we rapidly pulse to initiate ring-down events [with the use of an acousto-optic modulator (AOM; Gooch and Housego 3200-125)]. In our optical setup we can gener-

ate CRDP signals with ring-down times in the 0.3-1.5 µs range (depending on the usage of intracavity optics), at repetition rates as high as 100 kHz, that we record and digitize using a 14-bit digitizer (Teledyne, ADQ14DC-2X-PCIE, dual channel DC-coupled operation; sample rates of 2 GS/s per channel), which has a maximum acquisition rate of 100 kHz (mainly limited by the data transfer rate), 14-bit resolution per channel, and permits on-board channel subtraction and signal averaging.

For our demonstration, we use the (non-resonant) Faraday effect of a 6.35(1) mm thick, AR-coated $SiO_2$ substrate (FiveNines Optics; AR coated by FiveNine Optics with specified R$<$ 0.01%). In particular, using permanent magnets directly attached to the substrate, we can generate large enough Faraday optical rotations $\theta_F$ [17, 18, 79] that result in CRDP signals with polarization beat frequencies in the range of 1-3 MHz [the beat frequency is proportional to the induced Faraday rotation as: $\mathrm{f} = \theta_\mathrm{F} \cdot \mathrm{FSR}/\pi$, where FSR$= (c/L_{rt})$ is the cavity's free spectral range, with $c$ the speed of light and $L_{rt}$ the round-trip cavity length].

To demonstrate our ability to monitor and analyze CRDP experimental signals in a running fashion at $\sim$kHz rates using the FFT+peak-finding analysis algorithm [in accord with the results shown in Fig. 16 (a)], we proceed as follows: we initiate ring-down events at a rate of 100 kHz and continuously record the polarimetric ring-down signals (i.e., the photo-detector signals) for $\approx$ 220 ms, while channel subtraction allows for the generation of damped sinusoidal signals (CRDP traces) and on-board signal averaging enables the average of 40 consecutive traces; each trace, therefore, requires an integration time of 400 µs. In Fig. 17 (a) we show such an experimentally acquired CRDP trace with a Faraday-rotation-related polarization beat frequency of $\mathrm{f}_0 = 1.3$ MHz, a ring-down time of $\tau = 0.324(2)$ µs, and an SNR$\simeq$ 40. Note that, given the digitizer's sampling rate, the CRLB limit in estimating the signal's beat frequency is 1.3 kHz, i.e. $\sigma_\mathrm{f}/\mathrm{f} = 10^{-3}$ (Eq. 11; Fig. 15). Furthermore, using an additional, homemade, solenoid [with a length of 2.53(1) cm, and a diameter of 3.05(1) cm] placed around the $SiO_2$ substrate, we induce a rapid frequency shift on the recorded CRDP signal by applying a (rapidly pulsed) external magnetic field [using a USB controlled metal-oxide-semiconductor field-effect transistor-based switching circuit resulting in switch-on times of $< 10$ µs].

In Fig. 17 (b) & (c) we show our experimental results. We see that using an online FFT+peak-finding approach we can analyze CRDP traces at a constant rate of ∼3.3 kHz, which is comparable to the acquisition rate of the individual CRDP traces, and we can sensitively follow parameter changes - in our case, frequency changes - at µs time scales [evident from the analysis of our results using a sigmoid function that yields a $\approx 10\,\mu s$ rise-time; Fig. 17 (b)]. We emphasize here that the frequency fluctuations present in the recorded signals [Fig. 17 (b)] are the result of experimental noise sources (see [79]). As a comparison, in Fig. 17 (c) we also demonstrate that a least-squares approach would require computational times ranging from $\sim 1\,ms$ to $1\,s$ to analyze the same CRDP traces and, hence, we would be unable to analyze such a stream of traces online in a running fashion (this wide range of calculation times is related to the dependence of the least-square fitting algorithm to the initial guess fit parameters).

### 5.5 Discussion & conclusion

In this work, we consider different time- and frequency-domain-based computational algorithms for the rapid estimation of the signal parameters of damped sinusoidal signals. We analyze their accuracy and precision in terms of key signal parameters and estimate the computational time required to obtain these using standard computational platforms (e.g. a desktop computer) and software (e.g. Python). Overall, we see that a time-domain-based least-squares algorithm reaches the expected fundamental estimator limits in terms of precision and accuracy, and requires ms-long computational times to obtain fractional uncertainties at the ppm levels (for signals with high SNR), while a Fourier-based algorithm can achieve similar sensitivities at, at least, an order of magnitude faster computational times, even when one employees *standard* computational platforms and algorithms. We also consider in comparison to the least-squares and FFT analysis methods an alternative computational approach based on the Prony method, recently proposed for rapid analysis of damped sinusoidal signals; we observe that an implementation of the Prony method for our parameter-range of interest, fails to provide comparable accuracies and precisions to the least-squares and FFT methods, particularly within similar computational time-scales. We then validate our results using an experimental CRDP setup and a

FRGA-based acquisition system, to demonstrate the online recording and analysis of damped sinusoidal signals in a running fashion at $\sim$kHz rates using an FFT+peak-finding alogrithm, and, in particular, we demonstrate the ability to observe signal changes at time scales as fast as 10 µs.

Overall, based on our results, we recommend adopting FFT analysis approaches for rapid parameter analysis of damped sinusoidal signals within the context of CRDP/CRDE (and similar) techniques, which offers an optimum combination of speed and performance.

As a concluding remark we note that the exact computational speeds for the presented analysis methods depend significantly on the computational platform and software used. In Ref. [31], the authors demonstrate more than an order of magnitude improvements in computational times for the FFT and Prony methods by implementing alternative computational packages/softwares. As such, we expect that an optimized FFT algorithm may very well outperform our results by more than an order of magnitude, suggesting, therefore, that even with the use of standard data-acquisition systems, highly precise microsecond-resolved signal parameter estimation is possible. This becomes particularly important for our application of interest, CRDP: using the results presented in [79] we can estimate that in CRDP experiments with typical decay times constants of a few µs and oscillating frequencies of >1MHz (typically corresponding to polarimetric signals of a few rad; see [18, 23, 79]), and which can yield (acquired) signals with SNR $> 2^{12}$, we see that it is possible to perform online, in a running-fashion, microsecond-resolved experiments with µrad polarimetric sensitivities (i.e. sub-nrad/$\sqrt{\text{Hz}}$ sensitivities). Such a possibility is of paramount importance for emerging applications in chiral sensing and analysis, in surface catalysis, and indicates that real-time monitoring of gas/liquid flows in GC/HPLC and of surface (chiral) dynamics is nowadays feasible. Finally, we anticipate that an overall improvement in computational speeds could also result from implementing a network theory approach for rapid (spectroscopic/spectropolarimetric) signal analysis; for intance, network theory has been recently used for the classification and search of spectral features of various molecules [105, 106]. However, rapid parameter extraction of time-sampled signals could prove to be nontrivial with a network approach, and we will investigate such a possibility in future works.

## A. Speed for different implementations

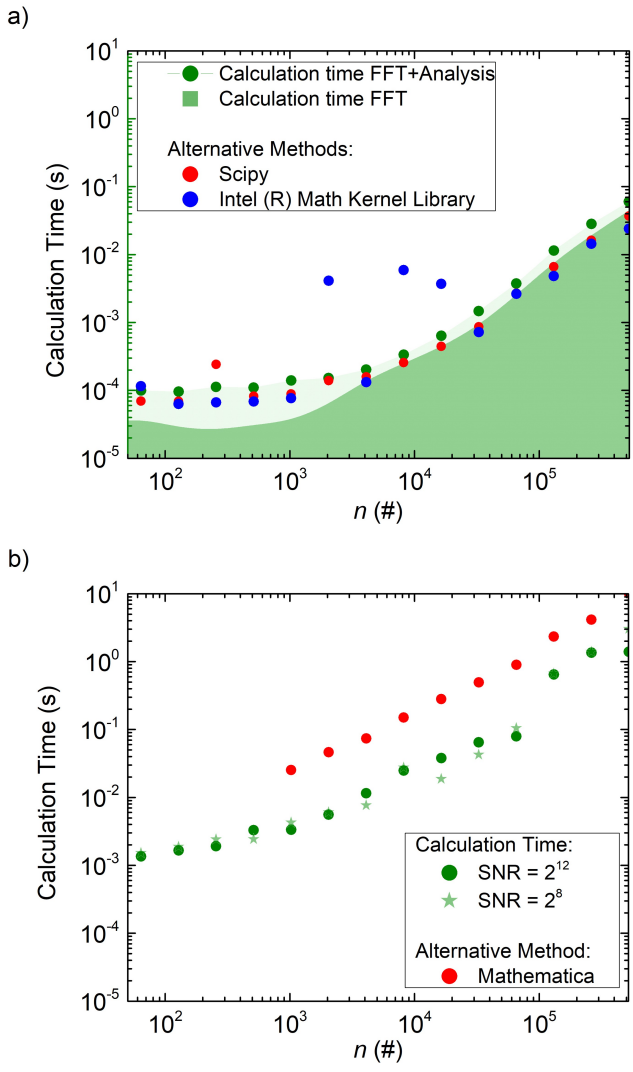


Figure 18: Dependence of calculation time on the number of samples $[n\,(\#)]$ for the: **(a)** FFT and **(b)** least-squares evaluation methods, using different libraries and/or computing systems. In all cases, we analyze simulated signals with $\mathrm{f}\times\tau = 5$, $\mathrm{SNR} = 2^{12}$ and $T_m \times \tau^{-1} = 5$ (Fig. 16).

In Fig. 18 we demonstrate how different libraries for Python and

computing systems can affect the attainable computational speeds, for instance, of the FFT and LS analysis methods we use in our work.

For the case of the FFT analysis (Fig. 18 a), we compare the calculation time required for the NumPy library to analyze a given signal (which is used for the FFT-based analysis of this work) with that required by SciPy and by Intel (R) Math Kernel Library (MKL). We observe that different libraries can improve the computational speeds of the FFT algorithm required for the analysis of a single trace, but only by a (maximal) factor of $\sim 3$. Notwithstanding, in the case of the Intel MKL, which provides the fastest calculation times, we observe that the analysis becomes less stable with varying number of samples.

For the case of the least-squares analysis (Fig. 18 b), we provide a comparison with a high-end computing system (Wolfram Mathematica). We observe that Mathematica takes at least an order of magnitude more time to analyze the same signal compared to a Python-based approach, which is undesirable for our applications of interest (as the Python-based computing systems already reach the fundamental CRLB limit faster).

It is important to emphasize that a crucial factor for all the above discussions/computations is the computational platform (hardware) used. The authors in [31], perform their Prony analysis on a GPU-based computer workstation and are able to observe large differences in calculation analysis times when using different Python implementations. In contrast, we employ a CPU-based computer workstation and a FPGA card, and when we use, for instance, the Numba compiler in our implementation of the Prony method we are unable to achieve (improved) speeds that match the results shown in [31]. However, since our focus is the development of a portable instrumentation, our aim is to identify an optimum platform - in terms of both the cost and performance - for signal acquisition and analysis. As such, while a GPU-based workstation seems to be the optimum solution in terms of speed performance (we expect this to be the case not only for the Prony method), its (typically large) cost and versatility remains - to date - a bottleneck, while CPU/FPGA-based analysis systems offer optimum performance, for our applications of interest, and are cost effective.

# 6 Rapid parameter estimation of discrete decaying signals using autoencoder networks

The following chapter is a direct adaptation of a manuscript that, at the time of writing, has completed the peer-review process and has been formally accepted for publication in Machine Learning: Science and Technology. The latest publicly available version of the manuscript can be found as:

The most recent available version can be found as:

**Visschers JC**, Budker D, Bougas L. Rapid parameter estimation of discrete decaying signals using autoencoder networks. arXiv preprint arXiv:2103.08663. 2021 Mar 10.

The minimal changes made to this version of the manuscript compared to the manuscript submitted for publication limit themselves to formatting and style alterations. The ideas and achievements outlined in this chapter, from conceptualization of the idea to submission of the manuscript has been mine, with the support of my co-authors.

## 6.1 Introduction

Machine learning (ML) is becoming a widespread method for the generation, and analysis of big data. ML uses networks of interconnected neurons (i.e., neural networks) that, much like real brains, recognize patterns in data structures [107, 108]. Generally, these neural networks are first trained in situations where the desired action/output of the network is known and subsequently used in similar, real-world situations. Research fields where a precise mathematical description of the problem is difficult -if not impossible- stand much to gain from implementing ML solutions. Areas such as image processing [109, 110] or text generation and analysis [111] have seen innovations that would not have been possible without the implementation of ML.

In physical sciences, ML-based techniques are becoming more and more widespread (see Ref. [112] for an extended review). Besides unlocking completely new innovations, ML techniques have been utilized in situations where finding solutions to physical problems

requires a lot of computing effort using conventional methods. For example, in fluid simulations where normally one is required to solve the Navier–Stokes equations, ML can be used to accurately predict the evolution of a fluid simulation [113], while reducing the amount of computation time of these complex simulations significantly. Furthermore, neural networks have been used to simulate light scattering by multi-layer nanoparticles and the subsequent design of these nanoparticles using backpropagation [114]. In spectroscopy, machine learning is used to accurately classify physical objects based on noisy/complex spectroscopic data [115–117]. Another field where ML techniques find their way is in experiments with extremely high inference rates and low latency constraints, such as in high-performance detector triggers [118] (e.g. the ATLAS experiment at the LHC at CERN). In these experiments, field-programmable gate array (FPGA) implementations of ML techniques are used to create complex hardware triggers for events with sub-microsecond lifetimes.

However, ML techniques are notoriously opaque, meaning that, although they deliver desired results, our understanding of how the results are obtained and how patterns in data are detected by the underlying neural network is very limited [119]. Often, little or no information on the physical system that is solved/simulated can be gained by studying the neural network itself, which is a significant drawback of these techniques. Furthermore, basic questions regarding the minimal amount of training data, or the optimal size of the underlying neural network required to obtain good precision for a pre-defined problem remain unanswered [112].

Signal-parameter estimation is an integral part of both fundamental and applied research, with precision, accuracy and speed being crucial for the real-time observation and control of physical and chemical dynamic processes. A multitude of research fields rely on determination of time constants and frequencies of decaying signals, examples of which include: nuclear magnetic resonance (NMR) [74] where molecular structures are analysed from precise determination of the frequency and decay time of the Larmor precession of nuclear spins in a magnetic field; free-induction-decay (FID) optical magnetometry [66, 68, 120–122] where measurement sensitivities depend on the precision of the measurement of the precession of magnetized spins about a magnetic field; and cavity ring-down spectroscopy

(CRDS) [29, 30, 50, 123, 124] that relies on the detection of variations of the photon lifetime in optical cavities to identify trace gasses, measure absorption cross-sections or observe chemical reactions in real time. Other cavity-enhanced methods, such as cavity ring-down polarimetry (CRDP) [15–18, 23, 26, 79] and ellipsometry (CRDE) [81, 125, 126], measure birefringence and/or dichorism of an optical medium through the precise estimation of the signal-decay time and the polarization beat oscillations superimposed on such decaying signals.

All the aforementioned methods require some form of data processing to determine their relevant signal parameters. Optimizing sensitivity and minimizing computational costs is usually done by averaging multiple, consecutive measurements, sacrificing temporal resolution. Therefore, it is imperative that one considers the sampling and acquisition rates of an instrument and balances that against the computational costs (calculation time) to analyse the acquired data, the time scales of relevant dynamics to be observed, and the requirement to observe those dynamics in real time. In cavity-enhanced spectroscopy, CRDP/CRDE techniques in particular, the decay times of interest are typically in the $10^{-7} - 10^{-5}$ s range, a few orders of magnitude smaller than the typical decay times of NMR experiments ($10^{-2} - 10^{1}$ s) or FID optical magnetometry techniques ($10^{-3} - 10^{1}$ s). As such, analysis methods that are sufficiently fast for the real-time analysis of single measurements of NMR or FID magnetometry experiments do not have the capability to offer real-time analysis in situations where the relevant time scales are much smaller than 1 ms. Different time- and frequency-based computational methods for rapid parameter estimation have been demonstrated, and evaluated, for cavity ring-down (CRD) spectroscopy methods [91, 127, 128]. Notably, Fourier-transform implementations on FPGAs have demonstrated data analysis rates as high as 4.4 kHz [93]. Several works discuss time- and frequency-domain analysis algorithms for damped sinusoidal signals [94, 95] and a recently published comparative study includes three analysis methods of discretely sampled damped sinusoidal signals in terms of their speed, and attainable accuracy and precision [129].

In this work we demonstrate a ML based approach to extract the relevant signal parameters from experimentally relevant signals with well-defined functional form. More specifically, we use dense

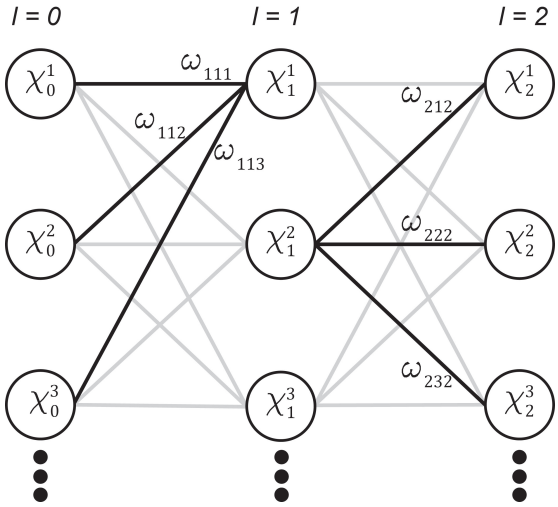


Figure 19: Example of a dense neural network with $L = 2$ layers. The first layer, $l = 0$ is the input layer, $l = 1$ a hidden layer and $l = 2$ is the output layer. $\chi_y^x$ specifies neuron $x$ in layer $y$. Each neuron contains a bias and layer specific activation function (not illustrated). $\omega_{y,x,z}$ denotes the weight (strength) between neuron $x$ in layer $y$ and neuron $z$ in layer $x - 1$. For illustrative purposes, some connections have been highlighted and are shown with their corresponding weight. The output of the neuron depends on its inputs, weights and bias put through the activation function (Eq. 22).

autoencoder networks to encode, extract parameters from, and subsequently reconstruct two types of discretely sampled, decaying signals: (1) exponentially decaying signals and (2) decaying oscillations. We evaluate the autoencoder network on its precision and accuracy in parameter extraction and compare its performance to the fundamental estimation limits of such signals given by the Cramér-Rao lower bound (CRLB). We show that the dense autoencoder network is able to reach analysis rates of 75 kHz of 1000 sample signals with cost-effective computational facilities. This makes the ML method ideal for implementations where computational capabilities come at a premium, such as fast, portable cavity-enhanced sensing instruments.

## 6.2 Theory

### 6.2.1 Dense autoencoder neural networks

A neural network is an ordered group of neurons that, much like the neurons in brains, communicate by signaling to each other. Each

neuron has a number of incoming and outgoing connections from and to other neurons. Each connection has a strength (weight) and can be stimulating or inhibiting the response of the receiving neuron. To calculate the activity (value) of a neuron, one compares the sum of its weighted inputs to a neuron specific reference (bias). The difference is passed through an activation function which produces the output of the neuron [107, 108]. Initially, the weights and biases within a neural network start off randomly. As the neural network is trained, these weights and biases are altered to optimize the performance of the network with respect to the task it is asked to perform. The manner in which neurons are ordered and connected influence the capabilities of the overall network and a variety of different network configurations with different purposes have been demonstrated [130]. A dense neural network is a network where the neurons are arranged in sequential layers and all the layers are fully connected. This means that every single neuron in one layer is connected to all neurons in the next layer. There are no neural connections within a layer or connections spanning multiple layers. An example of a dense neural network is shown in Fig. 19. For such a neural network, one can calculate its output as:

$$\begin{aligned} &\text{for } l : 0 \longrightarrow L \\ &\vec{\chi}_{l+1} = \mathcal{F}_l \left( \boldsymbol{\omega}_{l+1} \cdot \vec{\chi}_l - \vec{b}_l \right), \end{aligned} \tag{22}$$

where $L$ is the number of layers in the neural network (excluding the input layer), $\vec{\chi}_l$ is the output of the network at layer $l$, $\boldsymbol{\omega}_l$ is the weights matrix, and $\vec{b}_l$ is the bias vector for the corresponding neurons of layer $l$. Finally, $\mathcal{F}_l$ is the activation function of the network layer that acts piece-wise on each neuron in layer $l$. The input layer ($l = 0$) has no bias or activation function.

Autoencoders are symmetric, hourglass-shaped neural networks [131–135] (Fig. 20). This means that the number of neurons in the middle layer of the network is small compared to the number of neurons in the input and output layer of the neural network. The goal of the autoencoder network is to recreate its input signal. In the first half, the autoencoder network learns to encode the input data into a lower dimensional representation. In the second half, the autoencoder network learns to subsequently decode that lower dimensional representation and reconstruct the original signal with

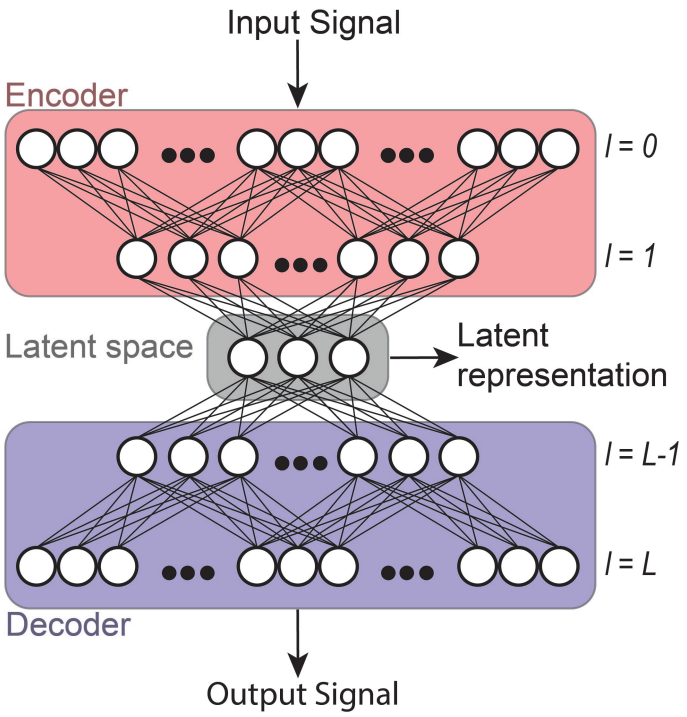


Figure 20: Example of an autoencoder neural network with $L$ layers. Signals are fed to the encoder, which compresses the data into a limited number of parameters after which the decoder reconstructs the original data from the encoded parameters whilst minimizing the losses. The layer with the smallest number of neurons is called the latent space. From the latent space the encoded, or latent, representation of the input signal can be extracted.

minimal losses. The middle layer of the network, in which the maximal compression of the signal takes place, is called the latent space. The latent representation of the original signal can be extracted from the latent space. Autoencoder networks are typically used in data compression [134] and de-noising [135].

The choice of dimensionality of the latent space is "*problematic*" [132]. In order to achieve maximum data compression, the number of neurons in the latent space of the autoencoder network should be as small as possible. However one would also like to be able to reconstruct the data with minimal error. If the inherent dimensionality of the data is not known *a priori* (as is typical) [132, 134, 135] but important, sparse, or variational autoencoder networks with Bayesian regularization methods should be used [136]. Furthermore, inference from the latent space representation is nontrivial because the way a network "learns" to encode a signal in training is not necessarily unique. Another way of structuring the latent space of an autoencoder network is by including the Kullback-Leibler (KL) divergence during training. However, KL divergence inevitably introduces a trade-off between the ability of the autoencoder network to reconstruct the original signal and a structured latent space [137].

In this work, we examine how an autoencoder network consisting solely of dense layers can be employed for direct extraction of the parameters of a signal with well-defined functional form from its latent signal-representation, as this becomes relevant in real-time signal analysis. To achieve such a task, we must make sure that (1) the dimensionality of the latent space matches the dimensionality of the input signal, and (2) that the autoencoder network encodes the input signal in a specific way where the latent representation of the signal matches the independent parameters of the signal.

We are able to satisfy the first requirement because the number of parameters in our signal's functional form are known, which allows us to match the number of neurons in the latent space of the autoencoder network to the number of parameters of the function (Secs. 6.2.2 & 6.3.2). Thereby we achieve efficient encoding without redundant neurons in the latent space, or redundant values in the latent representation of the signal. We are able to satisfy the second requirement by training the autoencoder network in a specific, three stage training method (Sec. 6.3.4).

### 6.2.2 Model signals

We use autoencoder networks to estimate the parameters of two types of experimentally relevant model signals: (i) exponentially decaying signals and (ii) decaying oscillations.

#### 6.2.2.1 Exponentially decaying signals

We start by investigating purely exponentially decaying signals. Such signals are encountered in a wide range of spectroscopic techniques, such as cavity ring down spectroscopy and cavity-enhanced sensing methods [29, 123, 138] and time-resolved fluorescence spectroscopy [139, 140]. An exponentially decaying signal can be characterized in terms of a model function as:

$$y(t) = A_0 \cdot e^{-t/\tau} + y_0(t), \tag{23}$$

where $A_0$ is the initial amplitude and $\tau$ is the decay constant of the signal, $y_0(t)$ is the signal offset, and $t$ is the independent (time) variable that is discreetly sampled. Here, for simplicity and without loss of generality, we assume that $A_0 = 1$ and $\langle y_0(t) \rangle = 0$. Furthermore we restrict the investigation of noise contributions to the global offset parameter, i.e. $y_0(t)$. We model the noise to be normally distributed, i.e., $\langle y_0(t) \rangle = 0, \langle y_0^2(t) \rangle = \sigma_{y_0}^2$ and define the signal-to-noise ratio as: $\mathrm{SNR} = \mathrm{A}_0/\sigma_{\mathrm{y}_0}^2 = \sigma_{\mathrm{y}_0}^{-2}$. Under realistic experimental conditions, signal-amplitude fluctuations can be incorporated into the SNR through $A_0$. This way, the dimensionality of the exponentially decaying signal is reduced to a single parameter that in most cases of interest carries the valuable information: the decay constant $\tau$.

#### 6.2.2.2 Decaying oscillations

The second type of model signals we investigate are decaying oscillations, as these become relevant in a wide range of experimental techniques, particularly within nuclear magnetic resonance (NMR) [74], optical magnetometry [66, 68, 120–122] and CRDP [15–18, 23, 26, 79, 141] and CRDE [81, 125, 126, 142]. An exponentially decaying oscillation can be characterized in terms of a model function as:

$$y(t) = A_0 \cdot e^{-t/\tau} \cdot \cos\left(2\pi \cdot \mathrm{f} \cdot t + \phi\right) + y_0(t), \tag{24}$$

were t is the discretely sampled independent variable of the signal (time), $A_0$ is the oscillation's initial amplitude, $\tau$ is the decay constant of the signal's envelope, f is the frequency of the oscillation, and $\phi$ its phase. Finally $y_0(t)$ is the global signal offset. Again, we assume the signals amplitude to be normalized ($A_0 = 1$) and no global signal offset ($\langle y_0(t) \rangle = 0$), and SNR is defined as $\mathrm{SNR} = \sigma_{\mathrm{y_0}}^{-2}$ identical to the exponentially decaying signals. The experimentally relevant parameters $\tau$, f and $\phi$, are the free parameters we extract from the latent representation. We recognize that, similar to the case of pure exponential decays, the amplitude parameter $A_0$ could potentially be an experimentally relevant parameter for inspection purposes, and could be considered as a free parameter of the model signal.

#### 6.2.2.3 Cramér-Rao lower bound

The fundamental limits for the statistical uncertainties of determining the decaying time-constant and oscillation frequency of pure and oscillating decaying signals are described by the Cramér-Rao lower bound (CRLB). For the case of a decaying (oscillating) signal (Eq. 23 & Eq. 24) the CRLB [65, 66] sets the lower limit on the variance of both the decay constant estimator $\sigma_\tau^2$ and the frequency estimator $\sigma_{\mathrm{f}}^2$. In general, the CRLB limit can be defined for any parameter extracted by an unbiased estimator. Here however, we focus on the two most important parameters for the experimental techniques of interest, e.g., [15–18, 23, 26, 29, 30, 50, 66, 68, 74, 79, 81, 120–123, 125, 126, 138–142].

The relation between the variance limit of the decay-time estimator and the lower limit of the frequency estimator is given by: $\sigma_\tau^2 = 2\pi\,\sigma_{\mathrm{f}}^2$. The CRLB for the frequency estimator $\sigma_{\mathrm{f}}^2$ is given by [65, 66]:

$$\sigma_{\mathrm{f}}^2 = \frac{6}{(2\pi)^2\,\mathrm{SNR}^2\,\mathrm{f_{BW}}\,T_{\mathrm{m}}^3}\,\xi\left(\tau/T_{\mathrm{m}}\right), \tag{25}$$

where SNR is the signal-to-noise ratio of the decaying oscillating signal; $\mathrm{f_{BW}}$ is the sampling-rate-limited bandwidth of the measurement; $T_{\mathrm{m}}$ is the measurement time window and, $\xi\left(\tau/T_{\mathrm{m}}\right)$ is a correction factor that takes into account the signal decay, which is given by

[65, 66]:

$$\xi(r) = \frac{\exp(2/r) - 1}{3r^3 \cosh(2/r) - 3r(r^2 + 2)}. \tag{26}$$

The factor $\xi(\tau/T_{\mathrm{m}})$ serves as a compensation factor in Eq. 25 that penalizes measurement of the tails of the exponential decay when the signal has effectively died out. The validity of Eq. 25 hinges on the assumption that the period of the oscillation is shorter than the typical decay time ($\tau$) of the signal envelope, and that a sufficient number of oscillations occur in the measurement time window. Moreover, Eq. 25 dictates that any noise sources affecting the signal detection are contributing to the fundamental CRLB limit through their effect on the SNR of signal.

In Ref. [79], it was demonstrated that the CRLB limit is the appropriate estimator of the fundamental sensitivity of frequency-based measurements within the context of cavity-enhanced spectropolarimetric techniques [15–18, 23, 26, 79, 81, 125, 126, 141, 142].

## 6.3 Methods

### 6.3.1 Simulated signals and data-sets

We use simulated signals generated using the model functions presented in Eqs. 23 & 24 to train and demonstrate the parameter-extraction capabilities of dense autoencoder networks. For our simulated signals - both pure exponentially decaying and decaying oscillating ones - we assume a total signal duration of $5\,\mu$s and an effective bandwidth-limited sampling rate of 200 MHz, resulting in discretely sampled signals with a length of 1000 samples. Such signals are frequently encountered in CRDP experiments [79]. We generate training data consisting out of multiple signals with varying experimentally relevant parameters. From the training data the autoencoder network learns the relationship between the shape of the signal and the values of the relevant parameters. For exponentially decaying signals (Eq. 23), the only parameter of interest is the decay constant $\tau$, which we vary between simulated signals as follows:

$$\tau = \left|\mathrm{norm}\left(\mu_\tau = 1\,\mu s, \quad \zeta_\tau = 0.5\,\mu s\right)\right|, \tag{27}$$

where $\mu_\tau$ and $\zeta_\tau$ are the average and standard deviation of a normal distribution, from which we take the absolute value as a negative

decay constant would lead to exponentially increasing signal. For the decaying oscillations (Eq. 24) we vary the decay constant using the same process, but also vary the other two experimentally relevant parameters, the frequency and phase of the oscillation, f and $\phi$ respectively, using:

$$\mathrm{f} = \mathrm{norm}\,(\mu_{\mathrm{f}} = 3\,\mathrm{MHz}, \quad \zeta_{\mathrm{f}} = 0.1\,\mathrm{MHz}) \tag{28}$$

$$\phi = \mathrm{norm}\,(\mu_{\phi} = 0, \quad \zeta_{\phi} = 0.1) \tag{29}$$

For the exponentially decaying signals a training data set consists of 200 signals, while for the decaying oscillations, a training data set consists of 1000 signals. These signals have a SNR $= 2^{20}$, to allow the autoencoder network to learn the features of each model signal and the signals variance under changing underlying parameters of interest without being hindered by noise artefacts.
We would like to mention here that, in real-world applications of this analysis method, the parameter range on which the autoencoder network is trained is of crucial importance to the functionality of the method. The training parameter range should correspond to the expected experimental parameter range. Therefore, prior knowledge about the parameters and the expected range of these parameters is required for the autoencoder to analyse experimental signals. However, in a real-world experiment, or commercially available instrument, baseline signal parameters will be well defined and within well specified ranges making the analysis method viable.

#### 6.3.2 Dense autoencoder networks

The autoencoder networks that we use to reconstruct and extract the parameters from decaying signals, have an input layer and output layer of 1000 neurons, equal to the number of samples in a signal. However the number of neurons in the latent space of the autoencoder networks are different between autoencoder networks that analyse exponentially decaying signals and decaying oscillations. In the case of the exponentially decaying signals, we use a network with only a single neuron in the latent space, corresponding with the single parameter of interest that we wish to extract from the signal, i.e., $\tau$. In the case of the decaying oscillations, we need three neurons in the latent space of the network in order to be able to extract the three interesting parameters of that signal,

i.e., $\tau, \mathrm{f}, \phi$. Furthermore, the signal complexity of a decaying oscillation with three parameters is higher than a exponentially decaying signal with only one parameter of interest. For this reason, to analyse decaying oscillations we increase the number of neurons in the layers of the encoder and decoder part of the autoencoder network (Fig. 20). The network we use to analyse the exponentially decaying signal has 5 total layers with 1000 - 50 - 1 - 50 - 1000 neurons per layer respectively. For the decaying oscillations, we use a network with 7 layers and 1000 - 50 - 10 - 3 - 10 - 50 - 1000 neurons per layer respectively. We choose to use a hyperbolic tangent as activation function for all relevant layers. This choice allows the neural network to express values between -1 and 1, which matches the maximal amplitude of the input and output signals. We create the aforementioned neural networks using the Tensorflow and Keras [143] libraries in a homemade Python script.

#### 6.3.3 Signal reconstruction and parameter extraction

Autoencoder networks are used to reconstruct the original input data. However, our focus here is to use autoencoder networks for the rapid extraction of the signal parameters from exponentially decaying signals and decaying oscillations from the latent space of the autoencoder network. To do this we map the signal parameters of interest to a number between -1 and 1 that the autoencoder network can express in the latent space:

$$x_{\mathrm{lat}} = \frac{x - \mu_x}{3 \times \zeta_x}, \tag{30}$$

where $x_{\mathrm{lat}}$ is the latent representation of the value of parameter $x$ ($\tau$, f, and $\phi$). $\mu_x$ and $\zeta_x$ are the mean and standard deviation of the parameter variation defined in Eqs. 27-29. We use Eq. 30 to create a desired latent representation of signal parameters during training, and to convert the latent representation back into the signal parameter when using the autoencoder network to analyse signals. For a trained neural network, a signal only needs to be passed through the encoding part of the network, up to the latent space, in order to find its latent representation. To investigate the analysis speed of the neural networks, we only take into account the time it takes to encode the original signal into the latent representation.

#### 6.3.4 Training protocol

The autoencoder networks are trained following a three-stage training scheme, using a stochastic gradient descent (SGD) algorithm [143] and use a mean squared error loss function given by

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{31}$$

where $n$ are the total number of samples in a signal, $y$ is the desired signal or representation and $\hat{y}$ is the predicted outcome by the autoencoder network. Even though the sizes of the autoencoder networks are different given the two different model signals, the training method for the autoencoder networks is the same:

1. The complete autoencoder network is trained to recreate simulated input signals for 100 epochs.

2. The encoder part of the autoencoder network, up to the latent space (Fig. 20), is trained to generate the desired latent-signal representation from simulated input signals.

3. Finally, the decoder part of the autoencoder network is trained with the desired latent-signal representation as input and the desired signals as output for 100 epochs.

We repeat these steps 10 times after which a new training data set is generated to avoid over-fitting a single data set. In total, each network is trained on 10 different data sets. By implementing three distinct training steps we are able to train the neural network to not only recreate the original signal, but also to encode the signal in a specific way that allows extraction of the signal parameters from the latent space.

Figure 21 compares the prediction error (loss) of two autoencoder networks during training. We measure the loss of each neuron in the output layer and the latent space using the mean squared error (MSE) [143]. The smaller the loss of the autoencoder network, the better the network is in recreating the desired output. The first network (red and black dots) is trained using our three-stage training method. The second network (dark-green line) is only trained on replicating the input data (first step). Both networks encode and reconstruct exponentially decaying signals described by Eq. 23. We

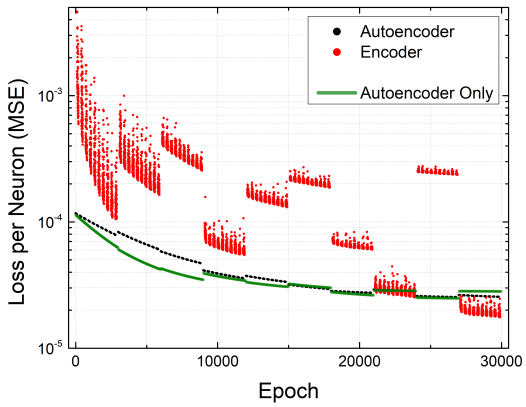


Figure 21: Average mean squared error (MSE) loss per neuron as a function of training epochs for two networks. One network is trained using the three-stage training method outlined in Sec. 6.3.4 (dots) while the second network is trained in a typical fashion for an autoencoder network consisting of the first step only (dark-green line). The two networks have the same shape and size, are trained for an equal number of epochs using an equal amount of training data. The black and red dots correspond to the losses of the first network at training step 1 and 2 respectively (see text). The losses at training step 3 are omitted from the figure because they overlap with the losses of the first training step. The breaks in the losses of the encoder network training (red dots) when a new data set is generated are attributed to the small number of neurons in the latent space of the autoencoder network and the specifics of the small validation data set.

observe that both training methods converge to similar losses per neuron at the end of training for the complete autoencoder networks (black dots versus dark-green line), meaning that both networks, at the end of training, are able to reproduce the input data with a similar level of precision and accuracy. However, in the autoencoder network that is trained using our three-stage training method, we see a decrease in losses from the encoder part of the autoencoder network (red dots), indicating that by using the three-stage training method we can achieve a desired latent signal representation without sacrifice of the capability of the autoencoder network to reconstruct the original signal.

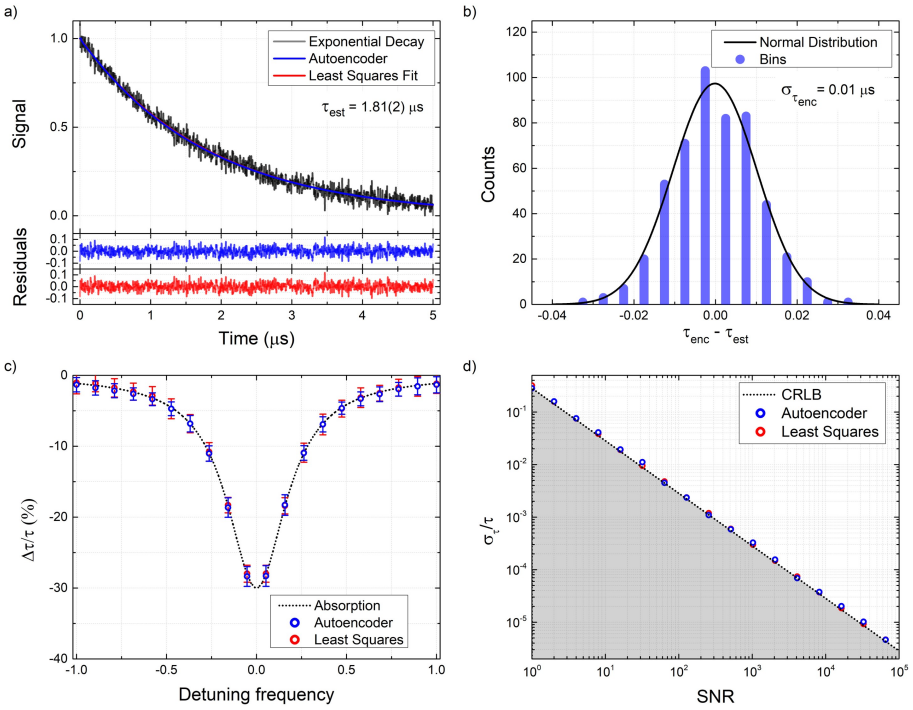


Figure 22: Parameter extraction from an exponentially decaying signal using an autoencoder network. **(a)** Example of an exponentially decaying signal with 1000 samples and SNR= $2^5$ (black line) reconstructed by a trained autoencoder network (blue line) and analysed using a least squares fitting method (red line). The residuals for each method are shown in the bottom insets of the figure with their corresponding colour. The decay constant estimated by the least squares method is $\tau_{\text{est}} = 1.81(2)\,\mu$s. **(b)** Histogram of 500 decay-constant estimates extracted from the latent space of the trained autoencoder network. Analysed signals have decay constants and SNR equal to the signal shown in (a). The value of the decay constant estimated by the least squares method, $\tau_{\text{est}}$, is subtracted from the decay constant extracted from the latent space of the autoencoder network, $\tau_{\text{enc}}$. The full width at half maximum (FWHM) of the fitted Gaussian distribution coincides with the uncertainty of the least squares method while the center of the distribution is not significantly different from 0. **(c)** Accurate and precise signal parameter extraction using an autoencoder network is possible in the case of (relative) changes in the decay constant through a simulated absorption feature (dotted line) as one would expect in, for instance, cavity ring-down absorption spectroscopy when a laser is detuned from the resonance frequency of the absorption feature. The feature is followed by both the trained autoencoder network (blue dots) and a least squares algorithm (red dots). The error bars represent the standard deviation of 100 parameter estimations from the latent space of the trained autoencoder network and the least squares algorithm respectively. **(d)** Precision of a least squares algorithm and a trained autoencoder network on estimating the decay constant from the latent space over multiple orders of magnitude in SNR. The Cramér-Rao lower bound (CRLB) giving the fundamental estimation limit on the decay constant is also shown (dotted line) and the area below the fundamental estimation limit is greyed out.

## 6.4 Results

### 6.4.1 Exponentially decaying signals

Figure 22 shows the results for a trained autoencoder network reconstructing and extracting parameters from exponentially decaying signals. The ability of the autoencoder network to extract the signal parameters from the latent space is compared to the performance of an ordinary least squares algorithm [79]. In Fig. 22(a) we show an example of a generated exponential decaying signal with SNR $= 2^5$ and a randomly selected time constant ($\tau = 1.81\,\mu$s). We use a trained autoencoder network (Sec. 6.3.4) to reconstruct this generated signal. In addition, we use a least-squares algorithm to analyse the generated signal, from which we extract the expected decay constant $\tau = 1.81(2)\,\mu$s, with a precision in accord to the CRLB limit (see related discussion in Ref. [129]). In Fig. 22(b) we show a histogram of 500 estimates of the decay constant extracted from the latent space of the autoencoder network. By subtracting the least squares decay constant estimate ($\tau_{est}$) and fitting the histogram to a Gaussian distribution we show that there is no difference in accuracy between the two analysis methods as the center of the distribution is not significant from zero. The width of the distribution of the decay constant estimates coincides with the uncertainty of the least squares estimation method indicating that also the precision of the two methods is equal. In Fig. 22(c) we demonstrate that the trained autoencoder network is able to, accurately and without loss in precision, follow signal changes. In particular, we simulate a spectral absorption feature having a characteristic dispersive (Lorentzian) profile. In this case, the decay constant changes by 30% and both the autoencoder network and the least squares algorithm are able to accurately detect such a change. . This situation is representative of a spectroscopy experiment investigating, for instance, absorption from gaseous species using CRDS [124, 144]. In Fig. 22(d) we show that the precision of the neural network is able to match the precision of the least squares algorithm and that both are limited by the CRLB over several orders of magnitude in SNR. It is crucial to emphasize that the network is able to accurately and precisely extract the correct signal parameters over a wide range of SNR values, despite the fact that we do not vary the SNR of the input training data during the training of the network.

### 6.4.2 Decaying oscillations

In Fig. 23 we present the results of a trained autoencoder network reconstructing and extracting parameters from decaying oscillations. In Fig. 23(a) we show an example of a decaying oscillating signal with SNR = $2^5$, $\tau = 1.28\,\mu$s, $\phi = -0.243$, and f= 2.972 MHz. We fit the signal to Eq. 24 using a least squares algorithm and use a trained autoencoder (Sec. 6.3.4) to reconstruct the signal. Figure 23(b) shows the histograms of 500 parameter estimates of $\tau$ , $\phi$ and f, extracted from the latent space of the autoencoder network. We extract the latent parameters $\tau_{\text{lat}}, \phi_{\text{lat}}, \text{f}_{\text{lat}}$ from the latent space of the autoencoder network and use Eq. 30 to reconstruct the estimate of the signal parameters found by the autoencoder $\tau_{\text{enc}}, \phi_{\text{enc}}, \text{f}_{\text{enc}}$. Following the same procedure described for the case of pure exponentially decaying signals, for each parameter we subtract the parameter's estimated value obtained through the least squares method and fit the histograms to a Gaussian distribution. Each distribution has a center that is not significant from zero and a width equal to the uncertainty of the fit parameters found by the least squares fit indicating that there is no difference in accuracy or precision between the two analysis methods. In Fig. 23(c) we demonstrate that the trained autoencoder network is able to, accurately and without loss in precision, follow signal changes. In particular, in Fig. 23(c) we simulate a spectral feature having a characteristic change in both its absorption and dispersion (e.g., Cotton effect). This results in signals whose frequency and decay constant change in a correlated way. Finally, in Fig. 23(d) we show that the precision of the parameter estimates from the latent space of the trained autoencoder network follows the precision of the least squares method over five orders of magnitude in SNR. Moreover, the precision of the parameters extracted from the latent space of the autoencoder network approaches the fundamental estimation limit as given by the CRLB.

### 6.4.3 Complexity vs. calculation time

Analysis speed is crucial for the real-time investigation and control of fast processes. Faster methods of analysis allow for large data stream to be analysed quicker, driving down computational costs.

For our particular cases of interest, we recently showed (Ref. [129]) that non-iterative FFT methods require $\sim 300\,\mu$s of calcula-

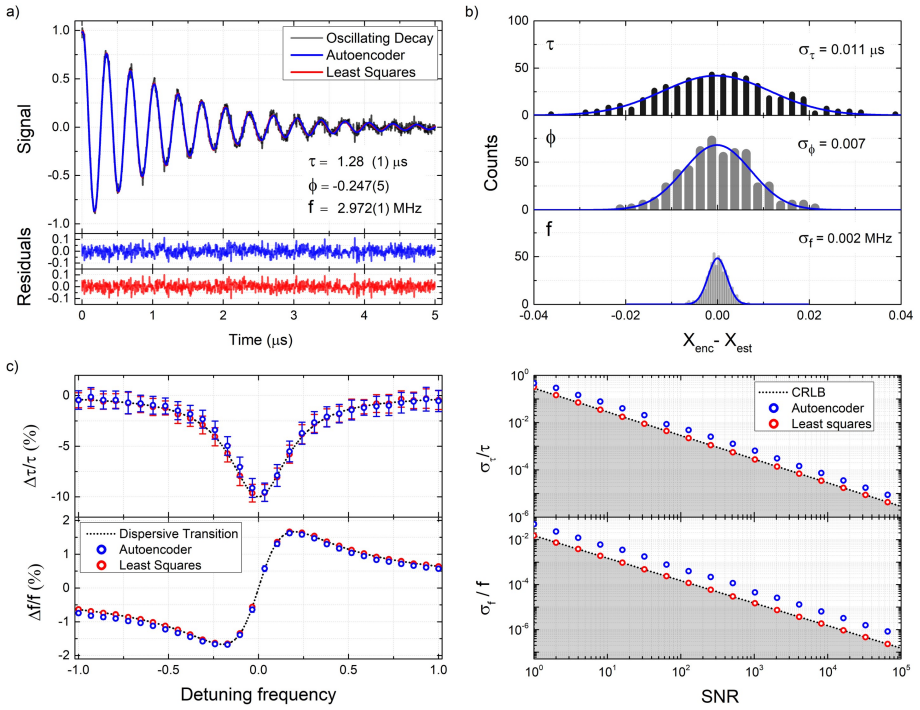


Figure 23: Parameter extraction from decaying oscillating signals using an autoencoder network. **(a)** Example of a decaying oscillation (black line) with SNR$= 2^5$ reconstructed by a trained autoencoder network (blue line) and analysed using a least squares fit (red line). The residuals for each method are shown in the bottom insets with their corresponding colour. The signal parameters found by the least squares fit are shown in the figure with their corresponding uncertainties. **(b)** Histograms of 500 signal parameter sets (top: $\tau$, middle: $\phi$ and bottom: f) extracted from the latent space of the autoencoder network. The analysed signals have the same parameters and SNR as the example signal shown in (a). For each of the signals parameters, we subtract the parameter estimate of the least squares fit $X_{est}$ from the parameter estimate extracted latent space of the autoencoder network $X_{enc}$. The width of the fitted Gaussian distributions coincide with the uncertainty of the least squares method for each individual parameter of the signal and the centers of the distributions are not significantly different form 0. **(c)** Accurate and precise signal parameter extraction using an autoencoder network is possible in the case of (relative) changes in the decay constant through a simulated disperive absorption feature, or cotton effect, (dotted line) as one would expect in a cavity enhanced polarimetry experiment when a laser is detuned from the resonance frequency of the absorption feature. The feature is followed by both the trained autoencoder network (blue dots) and the least squares algorithm (red dots).The error bars represent the standard deviation of 100 parameter estimations from the latent space of the trained autoencoder network and the least squares algorithm respectively. **(d)** Precision of the trained autoencoder network (blue dots) and the least squares algorithm (red dots) in estimating both the frequency (f) and decay constant ($\tau$) over several orders of magnitude in SNR. The fundamental estimation limits of these parameters, given by the Cramér-Rao lower bound (CRLB) is also shown (dotted line) and the area below the fundamental estimation limit is greyed out.

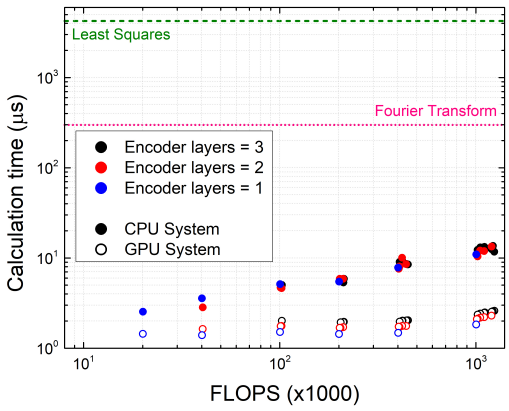


Figure 24: Calculation time required by the encoder part of the autoencoder network to extract the latent parameters of a single damped oscillating signal as a function of floating point operations (FLOPS) required to calculate the network using a CPU- (solid dots) and a GPU- (open dots) based system. The analysed signals consist of 1000 samples. The networks are varied in size, and thereby the required number of FLOPS needed to calculate latent signal representation. We vary the size of the network by varying the number of layers in the encoder network and the number of neurons within each layer. The lengths of these layers range from 10 to 500 neurons, where the smallest network investigated has a single layer with 10 neurons, and the largest network has 3 layers with 500, 200 and 100 neurons between the input layer and the latent space respectively. The dashed and dotted horizontal lines represent the calculation times required to analyse the same signal using a least squares algorithm and non-iterative Fast Fourier transform method, respectively [129].

tion time whereas a least-squares fitting algorithms requires $> 1\,\mathrm{ms}$ of calculation time under identical signal conditions as we present in this work. Similarly, other works have shown that FPGA- based systems running FFT algorithms are able to determine the decay constant of exponentially decaying signals with analysis rates of 4.4 kHz [93]. Here, we show that trained autoencoder networks are able to analyse both exponentially decaying signals and decaying oscillations significantly faster than any previously reported method. We compare the calculation speed of the encoder part of trained autoencoder networks of different sizes on different systems: a central processing unit (CPU-) and a graphics-processing unit (GPU-) based system. The CPU-based system is based around an Intel(R) Xeon(R) W-2123, 3.6 GHz processing unit with access to 16 GB of random-access memory (RAM) with a frequency of 1330 MHz, while the GPU system is based on a NVIDIA Quadro p5000 graphics board.

In Fig. 24 we present the calculation time required by the encoder part of an autoencoder network to analyse a single decaying oscillation signal for varying sizes of networks. For our estimation of the overall CPU calculation time (Fig. 24) we count the matrix multiplications, the adding/subtracting of the network biases and the application of the activation function. Other prepossessing operations for the CPU-based system, such as data collection, take between 3.6 - 4.9 $\mu$s per signal and are not included. For our estimation of the overall GPU calculation time we include the aforementioned operations (such as matrix multiplications, the adding/subtracting of the network biases, and the application of the activation function) and the transfer time of the results of the encoder onto the RAM of the computer. We do not include, however, the time it takes for the original signal to be transferred from the RAM onto the GPU memory ($\sim 5\,\mu$s) or any other pre/post-processing operations conducted by the CPU of the system ($\sim 6\,\mu$s). For the GPU-based system there is a clear difference in calculation time between neural networks with different numbers of layers in the encoder network, something we do not observe for the same neural networks on the CPU-based system. The results we present in Figs. 22 - 23 are obtained with autoencoder networks that require the calculation of ∼100,000 floating point operations (FLOPS) to reach the latent space representation of the signal. Using the trained autoencoder networks we are able

to achieve analysis rates upwards of 75 kHz using the GPU-based system. This includes the data transfer time, the pre/post processing operations conducted by the CPU, and the actual calculation time required of the autoencoder network itself. In future work, optimization of the data transfer rate between the data-acquisition system and data-analysis system and minimization of preprocessing operations, down to a combined $\sim 3\,\mu$s, the method of analysis we propose in this paper could be shown to be capable of reaching analysis rates of >200 kHz, in which we include the $<2\,\mu$s calculation time required by the GPU system.

### 6.5 Outlook and conclusion

In summary we have shown that we can accurately and precisely extract the signal parameters of decaying signals using simple autoencoder networks. We demonstrate that our approach is orders of magnitude faster than conventional algorithmic methods (e.g., least-squares or FFT), regardless of CPU- or GPU implementation of the neural network. We demonstrate analysis rates upwards of 75 kHz for signals with 1000 samples, and illustrate that analysis rates of >200 kHz are feasible with further optimization of data transfer speed between a data-acquisition and data-analysis device, which would allow for real time signal analysis rates of >200 kHz using state-of-the-art GSa/s sampling rates. Such capabilities can enable the real-time signal analysis of, for instance, CRDP signals at > 200 KHz rates using state-of-the-art acquisition modules that have GSa/s sampling rates.

Concluding, we wish to note that the methodology of signal parameter extraction directly from the latent space of dense autoencoder networks could be applicable to other signal types that currently use fitting models for parameter extraction. Presently, neural networks are used for the classification of spectroscopic data [115–117], however, our approach can be directly implemented to analyse spectroscopic data for quantitative rather than qualitative results. Signals of higher complexity, such as signals with additional decay constants or frequency components, will require larger networks and larger training data sets for the network. If the number of extra parameters is known, a similar technique to what we demonstrate can be employed. However, if the number of additional parameters is

unknown, a regularization method should be used to adequately choose the number of latent space parameters of the network to analyse the signal.

## A. Influence of realistic noise sources on autoencoder reconstructions and parameter estimations.

Studying additional noise sources and their influence on the autoencoder networks signal reconstructions and parameter estimations can be highly informative, considering that real-world experiments encounter real-world noise. We consider frequency- and amplitude noise as two additional noise sources and, in this section, study how these noise sources affect the signal reconstruction and parameter estimation differently compared to the offset noise discussed in the main text of this work. For this investigation we limit ourselves to decaying oscillations only.

Decaying oscillations with frequency noise are given by:

$$y(t) = A_0 \cdot e^{-t/\tau} \cdot \cos\left(2\pi \cdot \mathrm{f}\left(t\right) \cdot t + \phi\right) + y_0(t), \tag{32}$$

where the frequency is given by

$$\mathrm{f}(t) = \mathrm{f}_0 + \mathrm{norm}\left(\mu, \sigma\right) \tag{33}$$

and $\mathrm{norm}\left(\mu, \sigma\right)$ is a normally distributed random process with mean $\mu$ and standard deviation $\sigma$. We compare autoencoder reconstructions and latent space frequency estimates of decaying oscillations under three distinct cases of frequency noise where:

1. $\mu = 0\,\mathrm{kHz}; \quad \sigma = 0\,\mathrm{kHz},$
2. $\mu = 0\,\mathrm{kHz}; \quad \sigma = 30\,\mathrm{kHz},$
3. $\mu = 30\,\mathrm{kHz}; \quad \sigma = 30\,\mathrm{kHz}.$

The first case gives a baseline estimate without frequency noise, the second case adds frequency noise to the signals, the third and final case we investigate has both frequency noise and a frequency offset of equal magnitude. In each case we investigate the additional frequency noise (or lack thereof) on a baseline offset noise corresponding to a signal-to-noise ratio of $2^5$, an equivalent noise level shown in Fig. 23.

Figure 25(a)-(c) shows the autoencoder reconstructions and least-squares fits for the three cases of frequency noise. It can clearly be observed from the residuals of the autoencoder network reconstructions that, in the presence of frequency noise, the autoencoder is unable to fully reconstruct the signal (Fig. 25(b) & (c)). Figure 25(d)

shows histograms of 500 frequency estimates of the cases of frequency noise and area-equalized Gaussian fits of said histograms. The estimates and uncertainties from the least-squares fits are also shown. The autoencoder network frequency estimate distributions in the cases with additional frequency noise are $\sim 20\%$ wider compared to the case without frequency noise. Both cases where frequency noise is present pick up a bias compared to the real frequency of the decaying oscillating signal. Remarkably, the frequency estimates obtained form the least-squares fitting technique also pick up a slight, but significant bias in the presence of frequency noise. The width of the frequency estimates from the least-squares method also widens by $\sim 8\%$. The autoencoder network is able to differentiate between cases 2 and 3 where a frequency offset of the same magnitude as the frequency noise is applied.

Considering amplitude noise in decaying oscillating signals, we take the a similar and express signals containing amplitude noise as:

$$y(t) = A(t) \cdot e^{-t/\tau} \cdot \cos\left(2\pi \cdot \mathrm{f} \cdot t + \phi\right) + y_0(t), \tag{34}$$

where the amplitude of the signal is given by

$$A(t) = A_0 + \mathrm{norm}\left(\mu, \sigma\right). \tag{35}$$

Again we study three cases where:

1. $\mu = 0\,; \quad \sigma = 0,$
2. $\mu = 0\,; \quad \sigma = 0.03,$
3. $\mu = 0.03\,; \sigma = 0.03.$

Figure 26(a)-(c) shows the autoencoder reconstructions and least square fits for the three cases of amplitude noise. Interestingly, cases 2 and 3 show progressively more remaining signal in the residuals of the autoencoder reconstructions, In contrast, the least-squares method is able to reconstruct the signal effectively in all cases. Figure 26(d) shows area-equalized Gaussian fits from 500 parameter estimations of the respective amplitude noise cases and the estimates and uncertainties corresponding to the least-squares method. The least-squares method has no problem handling the amplitude noise cases and none of the parameter estimates differ significantly between the different cases of amplitude noise. It can be seen from

Fig. 26(d) that the autoencoder network attempts to accommodate the amplitude noise with an increased decay time (top) and slightly retarded phase (middle). The frequency-estimate distributions for the cases with amplitude noise are slightly narrower ($\sim 8\%$) compared to the frequency-estimate distribution of the case without amplitude noise (bottom).

Overall, the addition of extra, realistic noise sources such as frequency and amplitude noise increases the estimation uncertainty of, and can introduce biases in, the parameter estimates of the autoencoder networks. However, the parameter estimates of the autoencoder networks do not completely deteriorate from the addition of these noise sources. The results shown in Figs. 22 & 23 of the main text were chosen as the most general representation of a wide range of possible experiments ranging in SNR over five orders of magnitude (for offset noise). We would also like to note that the autoencoder network used to reconstruct and predict the parameters from signals affected by frequency- and amplitude noise are the same networks used to acquire the results presented in Fig. 23 in the main body of our work and have not been re-trained to accommodate these types of noise specifically.

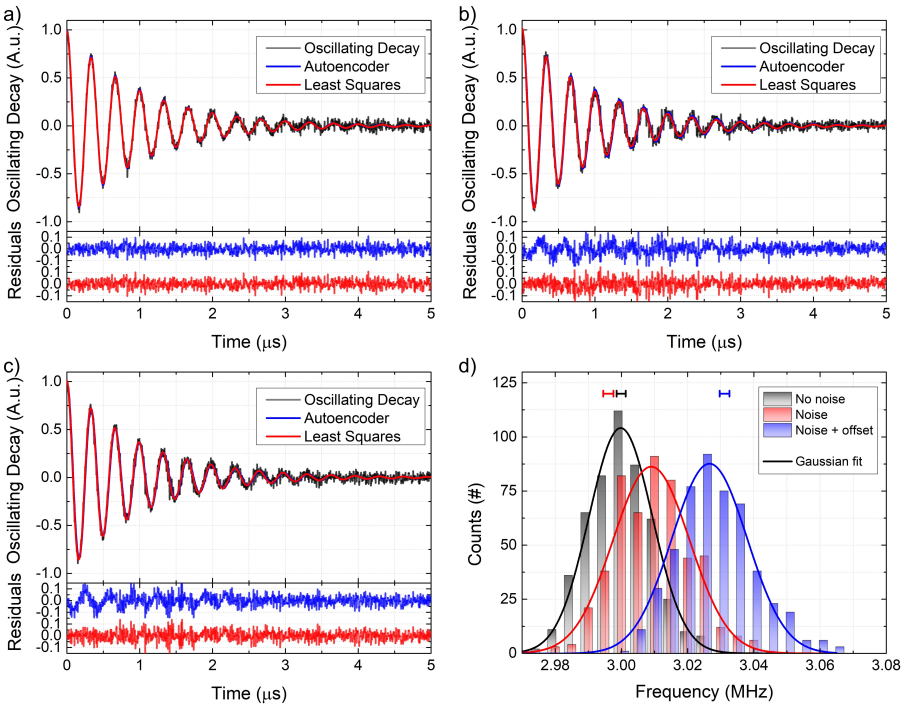


Figure 25: Autoencoder reconstructions and least squares fits of decaying oscillating signals with baseline SNR $= 2^5$ for the cases of **(a)** no additional frequency noise, **(b)** additional frequency noise with $\sigma = 30\,\text{kHz}$, and **(c)** additional frequency noise and frequency offset of $\mu = 30\,\text{kHz}$, $\sigma = 30\,\text{kHz}$. The residuals for the autoencoder reconstruction and the least-square fit are shown in the bottom insets of each figure with their corresponding colour. **(d)** Histograms and area-equalized Gaussian fits of 500 frequency estimates of the three cases of frequency noise (black = no frequency noise, red = additional frequency noise, blue = additional frequency noise and frequency offset). The least-squares estimates are shown along with their respective uncertainty above the histograms.

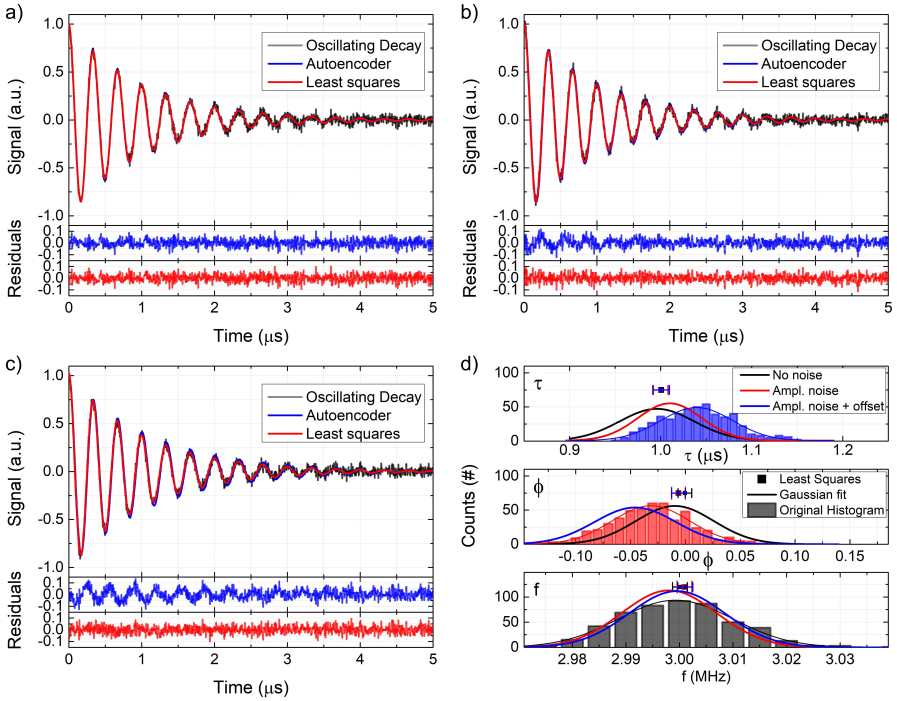


Figure 26: Autoencoder reconstructions and least squares fits of decaying oscillating signals with baseline SNR $= 2^5$ for the cases of **(a)** no additional amplitude noise, **(b)** additional amplitude noise with $\sigma = 0.03$, and **(c)** additional amplitude noise and amplitude offset where $\mu = 0.03$ and $\sigma = 0.03$. The residuals for the autoencoder reconstructions are given in the bottom insets of each figure with their corresponding colour. **(d)** Area-equalized Gaussian fits of histograms from 500 parameter estimates for the three noise cases shown in (a,b,c). For illustrative purposes only one original histogram is shown in each subplot. The least-squares estimates and corresponding uncertainties are given in each subplot above the distributions.

# 7 Growth box design and fabrication for observation of emissions from living plants

The following chapter has not been and will not be submitted for peer-review and/or publication. However, accompanying designs, software and scripts are available at `https://github.com/JCVis/GrowFacility`. The work described in this chapter has been done by myself. Design of both the physical structure and the data structure was done in close collaboration with L. Bougas to define the requirements and desires from the plant-growing facility and align its characteristics and capabilities to an existing experimental setup.

## 7.1 Introduction

In this work we present the design, realization and characterization measurements of a plant-growing facility. The motivation behind building the plant-growing facility are planned measurements of time-resolved chiral volatile organic compound (VOC) emission measurements from plants within the lab. The growth facility is required to provide an optimal, observable, controllable and importantly, repeatable growing environment for the plants of interest in various growth stages.

The growth facility needs to fulfill certain requirements in order to be used effectively for the previously mentioned experiments and to be able to facilitate many experimental variations. For instance, individual plants should be in their individual chambers to minimize unintentional cross talk between plants and cross contamination between plants, however the ability to study plant communication through the emissions of VOC's or soil should be an option. A high degree of automation, such as automated watering and an automated day-night cycle, is crucial to ensure optimal and repeatable growth conditions within a lab environment and minimize the need for outside intervention during long-running experiments or multi-week growth stages of certain plant species.

To accurately, and sufficiently monitor the growing conditions within each individual growing compartment within the plant-growing facility we need to observe some basic environmental parameters. These include: (1) Air temperature, (2) Air humidity, (3) Air Pressure, (4) Light levels, and, (5) Soil humidity levels. These basic mea-

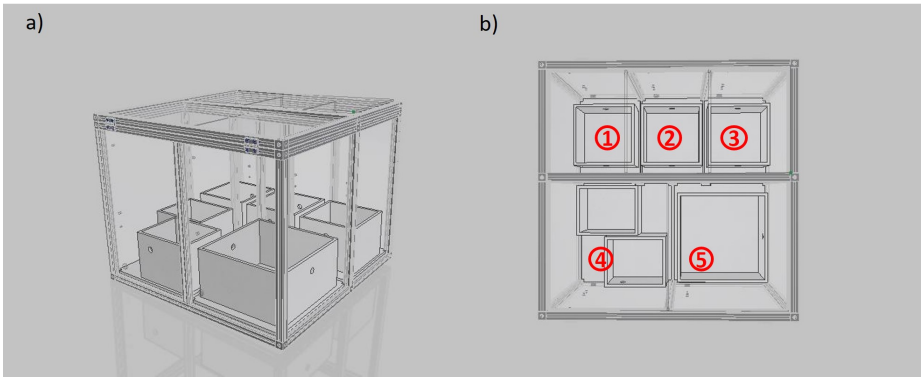


Figure 27: Design schematics of the plant-growing facility. **a)** Perspective view of the plant-growing facility. **b)** Top view of the plant-growing facility. The different growth compartments are labeled 1 through 5. Compartments 1,2, and 3 are individual compartments allowing to house a single plants each. Compartment 4 is able to accommodate two plants with separated soil containers. Compartment 5 features a single, large soil container that can house two plants at the same time.

surements are needed to ensure optimal growth conditions within a compartment of the growth-chamber. By also measuring some gasses present within the growth chamber we hope to gain insight into plant VOC emissions corresponding to the environment. The desired gas concentration measurements include: (1) total volatile organic compounds (tVOC), (2) carbon dioxide, (3) nitrogen dioxide, (4) ethyl alcohols, (5) ozone, and finally (6) oxygen. Importantly, from the plant emissions we also hope to gain insight on additional plant conditions that are imposed onto the plants under observation.

In the following sections we will describe the physical design and layout of the plant-growing facility followed by a description of the data handling and measurement scheme as well as a description of the home-made climate sensors. Finally some characterization measurements are presented which are important for the eventual use of plant-growing facility.

## 7.2 plant-growing facility

We designed a plant-growing facility consisting of five separated growth chambers, a side- and top-view of the design shown in Fig. 27. Three of the five compartments house individual plants and allow for controlled studies of individual plants. The fourth chamber

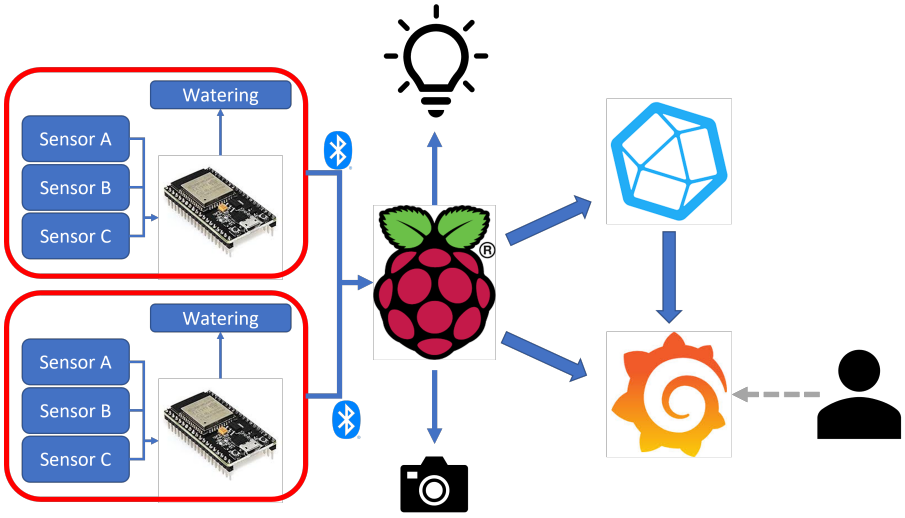


Figure 28: Visualisation of the experimental control and data flow for the plant-growing facility. The red boxes represent individual compartments of the plant-growing facility. Each functional growth chamber has an ESP-32 based sensor board which measures the climate parameters within each box and controls the watering of the soil. Each sensor board advertises the various sensor values in a Bluetooth Low Energy (BLE) service. The BLE service is read-out periodically by a Raspberry-Pi 4 which saves the data into an InfluxDB database. The Raspberry-Pi 4 also controls the light cycle of the plant-growing facility and has the capabilities of controlling a camera to periodically image the growing plants. Finally the Raspberry-Pi 4 hosts a Grafana server that visualizes the data from the ESP-32 sensor boards and is remotely accessible to anyone within the institutes network.

houses two plants but with individual soil containers and allows for communication between two plants through VOC emissions. The fifth and final chamber in the plant-growing facility is also able to house two plants sharing the air and soil allowing for both soil and VOC communication between plants. The footprint of the complete plant-growing facility is 75 cm x 75 cm with a height of 60 cm. The soil containers have a height of 17cm. The footprint of the soil containers in the individual growth compartments (1, 2, and 3) have a size of 20 cm x 20 cm. The soil containers in the fourth compartment have a footprint of 20 cm x 16 cm, and finally, the soil box in the combined compartment has a footprint of 30 cm x 30 cm.

The frame of the plant-growing facility is made from Thorlabs optical rails while the windows and soil containers are made in-house and are made from perspex and PVC respectively. We are able to access the growth compartments in the box through two top folding doors. The first of which gives access to the three individual compartments and the second one opens up the two double compartments. Electrical access is provided through a 2 cm hole in one of the outside window panels of each individual growth chamber at the heights of the top of the soil compartments. Two smaller holes provide access for a sampling tube for sampling of the plant emissions at the top and base of the plant respectively. A final hole at the height of the top of the soil container provides access for a watering tube.

Light is provided from high-efficiency, plant-growth optimized LED lighting (NEONICA Growy). The lights consume 13 W/m. The ceiling of each growth compartment is lined with 20 cm strips of LED lighting totalling 1 m. In growth chamber 3 (Fig 27b) we put an additional 0.6 m of LED lighting controlled separately form the regular lighting which gives us the ability to induce light stress in the plant growing in that compartment. To maximize the amount of light within the growth chamber we coat the insides of each compartment with aluminium foil. This subsequently limits the light leakage form the plant-growing facility [5].

[5] Much to the delight of the other occupants in the lab.

## 7.3 Measurements, Data logging and handling

Structured experimental controls and a clear data collection scheme are crucial for the ease-of-use of the growth box. In an effort to retain oversight we take a top down approach: we use a Raspberry-Pi 4 to collect, store and visualize all the relevant data from the different growth compartments. Furthermore, it is in control of all -facility wide- processes (such as the light cycle). Each individual growth compartment is fitted with a home-made climate sensor that observes the relevant environmental parameters within said growth compartment and reports them to the Raspberry-Pi 4, the climate sensors also regulate processes specific to a single growth-box (e.g. watering of the plants). A visualisation of the controls and data flow is given in Fig. 28.

### 7.3.1 Raspberry-Pi 4

A Raspberry-Pi 4 platform is used as the main data center for the plant-growing facility. It is ideal for this specific project because it is a lightweight and affordable computer platform that supports Bluetooth Low Energy (BLE) and accommodates remote access via SSH. The Raspberry-Pi 4 is used to collect and store the sensor data from the individual growth chambers in a database. It controls the light cycle of the plant-growing facility through a USB connected relay circuit. The four relays allow for four independent light cycles within the plant-growing facility. Currently the growth facility is set-up on a single light cycle and a separate relay is used to control additional lights in compartment 3 (see Fig. 27) which enables us to induce light stress in that specific compartment. We make extensive use of the Linux job scheduler Cron to perform periodic tasks, such as turning on and off the lights in the plant-growing facility, monitoring the climate parameters and pushing/pulling new versions of the GitHub repository.

### 7.3.2 Climate sensor

For the backbone of our home-made climate sensors we use an ESP-32 Development-board platform with integrated Bluetooth and Wi-Fi capabilities. It is cheaper compared to Arduino micro-controllers with similar features and is more suitable for permanent configurations as it has a smaller footprint. We designed a motherboard con-

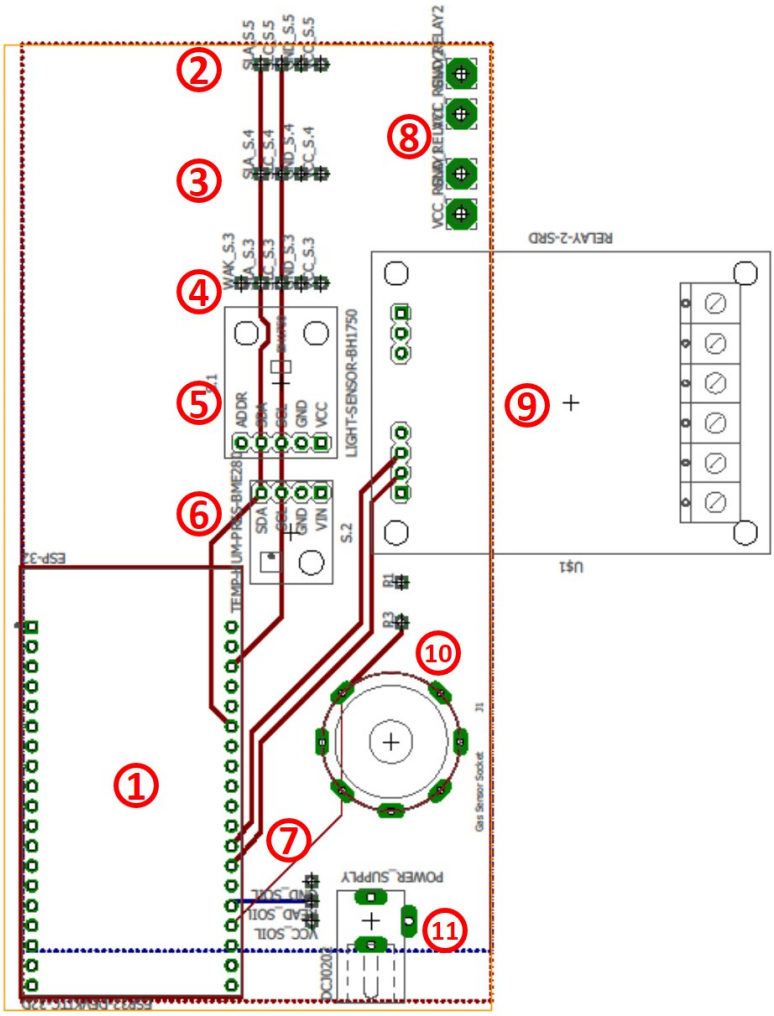


Figure 29: Design of the climate sensor motherboard. The front and back of the motherboard have a +5V and GND plate respectively. The motherboard features the following components: (1) ESP-32 Development board, (2)-(6) $I^2C$ connections for multiple sensor bake-out boards, (7) analog readout for a soil humidity sensor, (8) two power supplies for auxiliary components. (9) Two relay circuit bakeout board to toggle auxiliary components, (10) analog readout for ozone sensor, (11) power supply socket.

necting the various climate sensors to the ESP-32 micro-controller itself avoiding electronic breadboards and jumper cables, which quickly become a mess. Figure 29 shows the schematic of the climate sensor motherboard which facilitates up-to five $I^2C$ connections to the ESP-32 micro-controller alongside analog connections to ozone and soil humidity sensors. The motherboard also provides sockets and connections for two relay switches, one of which is used to water the plants when the soil humidity readout is below a threshold value. The ESP-32 micro-controller runs a script that reads out each climate sensor available and presents the sensor values as characteristics in a Bluetooth Low Energy (BLE) server. A readout device (client) can connect to the BLE server and subsequently read-out the various climate parameters. In our case we use the previously mentioned Raspberry-Pi 4 as a readout device, but for debugging purposes a phone using an appropriate application (e.g. nRF connect) can also be used to observe the climate parameters. The BLE server-client protocol is identical to how many wireless sports equipment works, such as, for example, heart rate chest band (BLE server) and a sports watch (BLE client).

#### 7.3.3 Sensors

The climate sensors are equiped with various individual sensors to monitor the climate within each growth container. Table 7.3.3 gives an overview of the different sensors that we fit on the climate sensor motherboard to monitor the climate within each growth container.

## 7.4 Data Collection

#### 7.4.1 Methods of Acquisition

The plant-growing facility supports various methods of data collection through different channels and at different rates of acquisition. The following methods can be used for climate data collection and logging:

1. The Linux job-scheduler Cron periodically runs a homemade python script on the Raspberry-Pi 4 that reads out all the relevant sensors on each ESP-32 sensor board. The python script subsequently uploads the sensor readouts into the database where the data is stored. The maximum rate at which the

| **Measurement** | **Sensor** | **Range** | **Type** |
|---|---|---|---|
| Pressure<br>Temperature<br>Humidity | BME280 | 300-1100 hPa<br>-40-85 $^{o}$C<br>0-100% rel. hum | I2C |
| eCO2<br>VOC | CCS811 | 400-32768 ppm<br>0 - 29206 ppb | I2C |
| Light | BH175 | 0-65535 lx | I2C |
| CO<br>$NO_2$<br>$C_2H_5OH$<br>VOC | Grove-Multichannel Gas Sensor V2 | ∼5-5000 ppm<br>∼0-2 ppm<br>∼1-500 ppm<br>∼1-500 ppm | I2C |
| Ozone | MQ131 | 10-1000 ppb<br>The other one | Analog |
| Soil Humidity | Capacitive Sensor | 0-4096 | Analog |

Table 1: Overview of the different sensors on the climate sensor motherboard. We give the measurement, type of the sensor, and the connection why which it is connected to the ESP-32 micro-controller.

Cron job scheduler is able to run the readout script is once per minute.

2. Manually running a streaming script on the Raspberry-Pi 4 to continually read out the relevant sensors on the climate sensor board. Identical to the first method the streaming script subsequently uploads the received sensor values into the database for storage and remote visualisation. The streaming script can be configured to (I) readout a single climate sensor or (II) readout all available climate sensors. This data acquisition mode results in a measurement every 2 seconds, or a measurement rate of 0.5 Hz.

3. The final method of data acquisition is to connect a single sensor board to a computer and establish a serial connection. The ESP-32 boards are programmed to print the sensor values at the refresh-rate of the micro-controller firmware ($\sim$ 2 Hz). This method was originally intended for debugging the firmware of the ESP-32, however it can also be highly advantageous to see fast, dynamic changes within the growing facility. The downside of this acquisition mode it that the sensor measurements are not stored automatically in the database. Data taken through this acquisition mode needs an external program if it needs to be saved for later use.

The Cron based data acquisition method is ideal for continuous observation of the overall climate within the plant-growing facility and runs continually. The two other acquisition methods are ideal for fast observation of dynamic processes in circumstances when external measurement devices are sampling within the growth chamber and these data steams have to be initiated manually.

#### 7.4.2 Database and data visualization

We use InfluxDB as the database structure to store and handle the data from the plant-growing facility. InfluxDB is an open-source database structure optimised for handling time series data. It allows high throughput data logging and native data analysis. The structure requires data to be written as JSON objects which can be added to the database. The JSON objects contain the measurement time, location within the plant-growing facility, and the respective measurement values. The database can be queried to selectively return data for subsequent analysis. To continually visualize the data stream from the growing compartments the ESP-32 sensors we use a Grafana dashboard. Grafana is an open source analytic and interactive visualization web application. It provides charts, graphs, and alerts for the growth chamber. The Grafana service is hosted on the Raspberry-Pi 4 and accessible for viewing to everyone on the institutes network accessible via IP address and port 10.106.9.83:3000.

### 7.5 Characterisation measurements

Characterization measurements are crucial for the operation of the plant-growing facility and for the understanding of future plant emission studies. To explain deviations from baseline conditions, those baseline conditions should be known and well understood. In Fig. 30 we show two distinct characterization measurements. Figure 30(a) shows response of the climate sensor to exposure to two distinctly different conditions, human respiration and gaseous butane. We are able to distinguish between these two exposures by identifying temperature and humidity changes in the case of exposure to human respiration and the subsequent absence absence of those changes in the exposure to gaseous butane whereas both respiration and butane register on the tVOC sensor. Figure 30(b) shows the measured $CO_2$ and tVOC concentrations in growth cham-

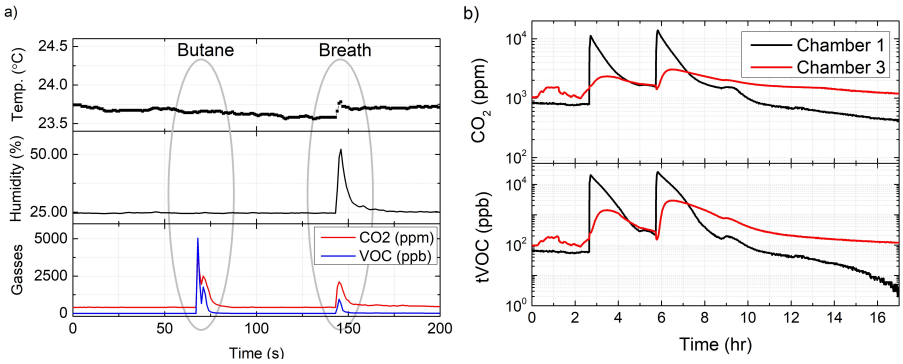


Figure 30: Characterization measurements of the ESP32 sensor-board and plant-growing facility as a whole. **(a)**. Consecutive exposures of climate sensor to butane and respiration. The butane is only registered with an increase of the measured $CO_2$ and tVOC concentrations by the CSS811 sensor, whereas the exhaled breath is also measured to have a higher temperature (top) and relative humidity (bottom). **(b)**. Measured $CO_2$ (top) and tVOC (bottom) concentrations in growth chambers 1 and 3 over a period of 17 hours. At t = 2.7 Hr and t = 5.75 Hr a petri-dish with a ∼ml of methanol is introduced into growth chamber 1. The $CO_2$ and tVOC concentrations are monitored as the methanol evaporates and diffuses through the entire growing facility. The measured $CO_2$ and tVOC concentrations in chamber 1 spike up immediately and slowly, over the course of a couple of hours, decay to their ambient levels. In chamber 3, the $CO_2$ and tVOC concentrations rise, albeit more slowly and not as drastically as in chamber 1. The highest level of recorded concentrations measured in chamber 3 are an order of magnitude lower than the maximum concentrations measured in chamber 1. Moreover, the time delay between the highest measured $CO_2$ and tVOC concentrations is approximately 45 minutes.

bers 1 and 3 over a period of 17 hours as an amount of methanol is introduced twice (at t = 2.7 hr and t = 5.75 hr) into chamber 1 and subsequently evaporates. The evaporation is registered on by the $CO_2$ and tVOC sensor on the climate sensors. The better the insulation between chambers 1 and 3, the slower the rise of $CO_2$ and tVOC concentrations will be in chamber 3. We observe that their concentrations do rise, albeit more slowly as in chamber 1. Furthermore, the the highest level of concentrations measured in chamber 3 are an order of magnitude lower than the maximum concentrations measured in chamber 1. Most importantly, the time delay between the highest measured $CO_2$ and tVOC concentrations is approximately 45 minutes. This means that any plant-to-plant signalling though VOC emissions will have a significant time delay and will not disturb control measurements taken at the same time as plant affliction measurements. The final characerization measurement, shown in Fig. 31, shows 10 day-night cycles of chambers 1 and 3 in the plant-growing facility. The parameters given include temperature, relative humidity, light level, $CO_2$ concentrations and tCOV concentrations. For all parameters except the latter two, the baseline/ambient lab levels are also recorded. Presence of people in the lab can be tracked by the lab lights sporadically switching on/off. From the 10-day observation a clear relation between the temperature in the growth compartments and the on/off status of the light can be seen, the growing lights are also responsible for lowering the relative humidity during day cycles. A high degree of temperature isolation is achieved through the reflective aluminium foil which lines the inside of each growth compartment and no temperature increase in growth chamber 1 can be observed while extra growing lights in box 3 are turned on to simulate"*high noon*" conditions.

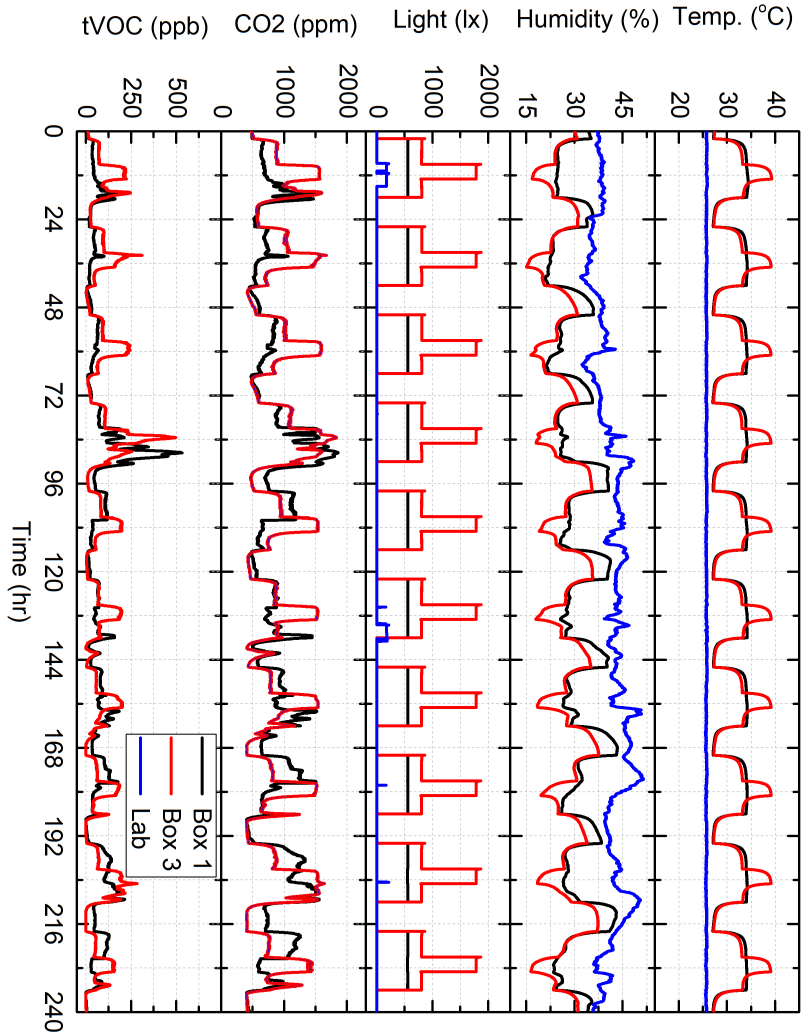


Figure 31: Observation of the climate parameters in growth boxes 1 (black line) and 3 (red line) and the ambient lab environment (Blue line) over a continuous period of 10 days (240 hours), starting t = 0 at 30-07-2021 00:00 until 09-08-2021 00:00. For the duration of this measurement, the plant-growing facility remained closed and the chambers being observed are empty. A clear correlation between the temperature and humidity within each box and the on/off status of the growing lights can be identified. Measured $CO_2$ and tVOC concentrations in the growth cambers are also highly correlated to the applied light cycle. Additional growing lights are turned on in box 3 for 4 hours each cycle to simulate high noon.

## 7.6 Outlook & Conclusion

In summary, this work shows a design, realisation and characterization of a cost-effective plant-growing facility with integrated monitoring capabilities. We show the data acquisition, handling and visualisation scheme of the facility and demonstrate a couple of characterizing measurements that are important for its application. Importantly, all the software associated to the growth-facility, such as the ESP-32 firmware and the data acquisition scripts are freely available on the GitHub development platform (`https://github.com/JCVis/weatherstation`) and will be updated as improvements are made to the growth facility and as inspiration for other designs. Future work will include more characterizations of the growing facility containing plants, the subsequent growing of plants from seeds and various plant VOC emission studies are planned with this plant-growing facility.

# 8 Conclusion and outlook

In summary, this thesis presents the general principles of continuous-wave cavity ring-down polarimetry, a cavity ring-down-based polarimetric scheme that employs continuous wave laser sources and benefits from the use of large intracavity polarization anisotropies for highly sensitive polarimetric measurements. Contrary to polarization-dependant cavity ring-down spectroscopy approaches that are effectively intensity-based measurement schemes (since these monitor changes in the ring-down decay time), cavity ring down polarimetry schemes are frequency-based measurements where sensitive detection of shifts in their induced polarization beat frequency enables highly sensitive polarimetric measurements. We demonstrate that the use of continuous wave laser sources facilitate gains in signal intensity and data acquisition rates, enabling shot-noise limited measurements. The fundamental limits of continuous-wave cavity ring-down polarimetry are considered, outlining the capability of this method to achieve high sensitivity using optical cavities with modest finesse. We demonstrate the merits of continuous-wave cavity ring-down polarimetry by measuring non-resonant Faraday optical rotation in gaseous and solid samples using our prototype design.

This thesis also considers different computational algorithms based on the time- and frequency-domain for the rapid estimation of signal parameters from damped sinusoidal signals, as encountered in continuous-wave cavity ring-down polarimetry experiments. We analyse the accuracy and precision of a least-squares, fast Fourier transform and Prony analysis method in terms of key signal parameters and estimate the computational time required to obtain these using standard computational platforms (e.g. a desktop computer) and software (e.g. Python). We observe that the time domain-based least squares method performs to the expected fundamental estimator limits given by the Cramér-Rao lower bound. However, the least-squares analysis method requires $\sim$ms calculation time whereas non-iterative Fourier-based algorithm achieves similar sensitivities with computation times an order of magnitude lower compared to the least-squares method. The third computational approach we consider for the analysis of damped sinusoidal signals is based on the Prony method. However, we observe that the im-

plementation fails to provide comparable accuracy and precision to the least-squares and Fourier methods. To validate the finding we demonstrate the online recording and analysis of a continuous-wave cavity ring-down polarimetry and show that we can observe signal changes at time scales of 10 $\mu$s using the Fourier based method, which based on the results of chapter 5, we recommend over the Prony or least-squares method for fast and accurate analysis.

This thesis also presents a novel method of parameter estimation using dense autoencoder networks. We demonstrate the parameter estimation method on exponentially decaying signals and decaying oscillations, the latter we observe in continuous-wave cavity ring-down polarimetry. We demonstrate that the precision and accuracy of the autoencoder networks approaches the fundamental estimation limit as given by the Cramér-Rao lower bound for both model signals. Furthermore we show that the novel approach is orders of magnitude faster than conventional algorithmic methods (e.g., least-squares or FFT), regardless of CPU- or GPU implementation. Analysis rates upwards of 75 kHz for signals with 1000 samples are demonstrated, and analysis rates of >200 kHz are feasible with optimization of data transfer speed between a data-acquisition and data-analysis device, which would allow for real time signal analysis rates of >200 kHz using state-of-the-art GSa/s sampling rates. The demonstrated analysis rate easily accommodates the measurement repetition rate of the current continuous-wave cavity ring-down polarimetry instrument and is therefore ideal for an on-line analysis implementation.

Finally this thesis presents the design and implementation of a plant-growing facility from which plant emissions can be detected. The detection could possibly be done using an altered continuous-wave cavity ring-down polarimetry instrument using the demonstrated autoencoder network analysis method. The design outlines a climate monitoring and control system with a fully automated day-night cycle and integrated watering capabilities. The growth facility offers optimal control and inspection, with minimal effort for an optimal growing climate.

Concluding, continuous-wave cavity ring-down polarimetry is a

powerful and versatile technique suitable for the measurement of reciprocal and non-reciprocal birefringence and dichroism, linear and circular. We believe that continuous-wave cavity ring-down polarimetry is the ideal modality for portable a spectropolarimetric instrument as it allows for a time-resolved, highly sensitive, and cost-effective operation at a broad spectral region. The technique is delivered with a fast and lightweight analysis method based on neural network autoencoders capable of analyzing single measurements from the continuous-wave cavity ring-down polarimetry instrument in an on-line and running fashion. Furthermore, a cost-effective plant-growing facility is presented from which the volatile organic compound emissions can be measured with a modified continuous-wave cavity ring-down polarimetry instrument.

Future work on the continuous-wave cavity ring-down polarimetry technique will focus on improvements in system design and sensitivity, and on the possibility of coupling continuous-wave cavity ring-down polarimetry with chromatographic techniques, toward the development of a portable instrument suitable for breath analysis and monitoring in clinical settings and for trace-gas analysis and monitoring of paramagnetic species in field settings.
The development of the autoencoder network based analysis method is not only advantageous for the rapid analysis of cavity ring-down polarimetry experiments. The method can also be employed for the direct analysis of cavity ring-down spectroscopy measurements, as we demonstrate the use of autoencoder networks on exponentially decaying signals which are usually encountered in intensity based cavity ring-down spectroscopy measurements. One area where future work can focus on is expanding the autoencoder analysis method to a wider variety of signals, thus opening up the possibility of rapid analysis in more experimental techniques and fields of research. Another area which would improve the applicability of the analysis technique in an experimental setting is integration of the trained autoencoder networks onto a field programmable gate array for a combined data acquisition and analysis system. This will reduce the data-transfer time, which is the greatest bottleneck of the analysis method, and will allow analysis rates, and therefore real-time observation of dynamic processes, upwards of 200 kHz of optical spectro-polarimetric measurements.

# 9 Acknowledgements

First and foremost I would like to thank my supervisor, Dmitry Budker. Dima, thank you for giving me the opportunity to do my doctorate in your group. Over the past years you have, despite your own ever-busy schedule, always been available for advice, insight and discussions. The time you put in helping me with all of the manuscripts has been most helpful.

Second, I would like to thank the other members of my committee for critically addressing my thesis and the scientific contributions I have made over the past years.

Third, my gratitude goes out to Lykourgos Bougas, with whom I have had the honor to collaborate throughout my time in Mainz. Lykourgos, working with you on what have mainly been your ideas regarding chiral polarimetry over these last few years has been an inspiring experience, at times completely overwhelming and educational through out. Thank you for your infinite patience, never ending motivation and limitless insight.

To all my other colleagues, with whom I have had the pleasure of competing in triathlons, cycling on the road, cooking food, playing board games, coming up with side projects, and discussing the meaning of life: Thank you for making my time here not only educational, but also enjoyable.

Oriane, I have to thank my doctorate for you; without it we would not have met. At the same time, I also have to thank you for helping me to keep my eye on the ball and for motivating me in the final stages of my doctorate. I am excited to see what the future has in store for us.

Lastly, I would like to thank my parents, who, even though the last year has been tough, have always and will continue to provide me a secure home base from which I get to explore the world and create my own path within it. Pap en mam, thank you for all your support in my adventures.

# References


[1] E. L. Eliel and S. H. Wilen. *Stereochemistry of Organic Compounds*. Wiley, 1994.

[2] Sharon M Kelly, Thomas J Jess, and Nicholas C Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1751(2):119–139, 2005.

[3] Dmitry Budker and Dereck Jackson K. *Optical Magnetometry*. Cambridge University Press, 2013.

[4] Kenneth W Busch and Marianna A Busch. *Chiral analysis*. Elsevier, 2011.

[5] F Arago. Mémoire sur une modification particuliere qu'eprouvent les rayons lumineux dans leur passage à travers certains corps diaphanes et sur plusieurs autres nouveaux phénomenes d'optique. *Moniteur*, 73:282–284, 1811.

[6] John Frederick William Herschel. *On the Rotation Impressed by Plates of Rock Crystal on the Planes of Polarization of the Rays of Light, as Connected with Certain Peculiarities in Its Crystallization. From the Transactions of the Cambridge Philosophical Society*. J. Smith, 1820.

[7] Jean-Baptiste Biot. Phénomènes de polarisation successive, observés dans des fluides homogènes. *Bull. Soc. Philomath*, 190:1815, 1815.

[8] Jean-Baptiste Biot. *Sur les rotations que certaines substances impriment aus axer de polarisation des rayons lumineux*. 1819.

[9] Ole Knudsen. The faraday effect and physical theory, 1845–1873. *Archive for history of exact sciences*, 15(3):235–281, 1976.

[10] Forris Jewett Moore. *A history of chemistry*. McGraw-Hill, 1918.

[11] Archibald Couper.

[12] Ludwig Ferdinand Wilhelmy. *Ueber das Gesetz, nach welchem die Einwirkung der Säuren auf den Rohrzucker stattfindet*. Number 29. W. Engelmann, 1891.

[13] SCHMIDT + HAENSCH GmbH & Co. Polartronic H532, 2021.

[14] Laurence D. Barron. *Molecular Light Scattering and Optical Activity*. Cambridge University Press, 2 edition, 2004.

[15] Thomas Müller, Kenneth B Wiberg, and Patrick H Vaccaro. Cavity ring-down polarimetry (CRDP): a new scheme for probing circular birefringence and circular dichroism in the gas phase. *The Journal of Physical Chemistry A*, 104(25):5959–5968, 2000.

[16] Thomas Müller, Kenneth B. Wiberg, Patrick H. Vaccaro, James R. Cheeseman, and Michael J. Frisch. Cavity ring-down polarimetry (CRDP): theoretical and experimental characterization. *J. Opt. Soc. Am. B*, 19(1):125–141, Jan 2002.

[17] L. Bougas, G. E. Katsoprinakis, W. von Klitzing, J. Sapirstein, and T. P. Rakitzis. Cavity-enhanced parity-nonconserving optical rotation in metastable Xe and Hg. *Phys. Rev. Lett.*, 108:210801, May 2012.

[18] Lykourgos Bougas, Dimitris Sofikitis, Georgios E Katsoprinakis, Alexandros K Spiliotis, Paraskevas Tzallas, Benoit Loppinet, and T Peter Rakitzis. Chiral cavity ring down polarimetry: Chirality and magnetometry measurements using signal reversals. *The Journal of Chemical Physics*, 143(10):09B603_1, 2015.

[19] Adam J. Fleisher, David A. Long, Qingnan Liu, and Joseph T. Hodges. Precision interferometric measurements of mirror birefringence in high-finesse optical resonators. *Phys. Rev. A*, 93:013833, Jan 2016.

[20] Haifeng Huang and Kevin K. Lehmann. Effects of linear birefringence and polarization-dependent loss of supermirrors in cavity ring-down spectroscopy. *Appl. Opt.*, 47(21):3817–3827, Jul 2008.

[21] Mathieu Durand, Jérôme Morville, and Daniele Romanini. Shot-noise-limited measurement of sub–parts-per-trillion birefringence phase shift in a high-finesse cavity. *Phys. Rev. A*, 82:031803, Sep 2010.

[22] John L. Hall, Jun Ye, and Long-Sheng Ma. Measurement of mirror birefringence at the sub-ppm level: Proposed application to a test of qed. *Phys. Rev. A*, 62:013815, Jun 2000.

[23] Dimitris Sofikitis, Lykourgos Bougas, Georgios E Katsoprinakis, Alexandros K Spiliotis, Benoit Loppinet, and T Peter Rakitzis. Evanescent-wave and ambient chiral sensing by signal-reversing cavity ringdown polarimetry. *Nature*, 514(7520):76–79, 2014.

[24] Gilles Bailly, Raphaël Thon, and Cécile Robilliard. Highly sensitive frequency metrology for optical anisotropy measurements. *Review of Scientific Instruments*, 81(3):033105, 2010.

[25] Michele Gianella, Tomas H. P. Pinto, Xia Wu, and Grant A. D. Ritchie. Intracavity faraday modulation spectroscopy (infamos): A tool for radical detection. *The Journal of Chemical Physics*, 147(5):054201, 2017.

[26] Patrick Dupré. Birefringence-induced frequency beating in high-finesse cavities by continuous-wave cavity ring-down spectroscopy. *Physical Review A*, 92(5):053817, 2015.

[27] A.K. Spiliotis, M. Xygkis, E. Klironomou, E. Kardamaki, G.K. Boulogiannis, G.E. Katsoprinakis, D. Sofikitis, and T.P. Rakitzis. Optical activity of lysozyme in solution at 532 nm via signal-reversing cavity ring-down polarimetry. *Chem. Phys. Lett.*, 747:137345, 2020.

[28] A K Spiliotis, M Xygkis, E Klironomou, E Kardamaki, G K Boulogiannis, G E Katsoprinakis, D Sofikitis, and T P Rakitzis. Gas-phase optical activity measurements using a compact cavity ringdown polarimeter. *Laser Phys.*, 30(7):075602, jun 2020.

[29] Giel Berden and Richard Engeln. *Cavity ring-down spectroscopy: techniques and applications*. John Wiley & Sons, 2009.

[30] Martyn D. Wheeler, Stuart M. Newman, Andrew J. Orr-Ewing, and Michael N. R. Ashfold. Cavity ring-down spectroscopy. *J. Chem. Soc., Faraday Trans.*, 94:337–351, 1998.

[31] Emma Wilson, Thomas M Conneely, Andrey Mudrov, and Ivan Tyukin. Implementation of the prony method for signal deconvolution. *IFAC-PapersOnLine*, 52(29):269–273, 2019.

[32] E. L. Eliel and S. Wilen. *Stereochemistry of Organic Compounds*. Wiley, 2008.

[33] E.N. Fortson and L.L. Lewis. Atomic parity nonconservation experiments. *Physics Reports*, 113(5):289 – 344, 1984.

[34] Kenneth W. Busch and Marianna A. Busch. *Chiral Analysis*. Elsevier, 2006.

[35] Mark D. Argentine, Paul K. Owens, and Bernard A. Olsen. Strategies for the investigation and control of process-related impurities in drug substances. *Advanced Drug Delivery Reviews*, 59(1):12 – 28, 2007. Pharmaceutical Impurities: Analytical, Toxicological and Regulatory Perspectives.

[36] S. Li, P. Vachaspati, D. Sheng, N. Dural, and M. V. Romalis. Optical rotation in excess of 100 rad generated by rb vapor in a multipass cell. *Phys. Rev. A*, 84:061403, Dec 2011.

[37] Béla Tuzson, Markus Mangold, Herbert Looser, Albert Manninen, and Lukas Emmenegger. Compact multipass optical cell for laser spectroscopy. *Opt. Lett.*, 38(3):257–259, Feb 2013.

[38] John U. White. Long optical paths of large aperture. *J. Opt. Soc. Am.*, 32(5):285–288, May 1942.

[39] D. Herriott, H. Kogelnik, and R. Kompfner. Off-axis paths in spherical mirror interferometers. *Appl. Opt.*, 3(4):523–526, Apr 1964.

[40] D Das and A C Wilson. Very long optical path-length from a compact multi-pass cell. *Applied Physics B*, 103(3):749–754, 2011.

[41] Karol Krzempek, Mohammad Jahjah, Rafał Lewicki, Przemysław Stefański, Stephen So, David Thomazy, and Frank K Tittel. CW DFB RT diode laser-based sensor for trace-gas detection of ethane using a novel compact multipass gas absorption cell. *Applied Physics B*, 112(4):461–465, 2013.

[42] Richard Engeln, Giel Berden, Esther van den Berg, and Gerard Meijer. Polarization dependent cavity ring down spectroscopy. *The Journal of Chemical Physics*, 107(12):4458–4467, 1997.

[43] Richard Engeln, Giel Berden, Rudy Peeters, and Gerard Meijer. Cavity enhanced absorption and cavity enhanced magnetic rotation spectroscopy. *Review of Scientific Instruments*, 69(11):3763–3769, 1998.

[44] Jae Yong Lee, Hai-Woong Lee, Jae Wan Kim, Yong Shim Yoo, and Jae Won Hahn. Measurement of ultralow supermirror birefringence by use of the polarimetric differential cavity ringdown technique. *Appl. Opt.*, 39(12):1941–1945, Apr 2000.

[45] Giel Berden, Richard Engeln, Peter C. M. Christianen, Jan C. Maan, and Gerard Meijer. Cavity-ring-down spectroscopy on the oxygen a band in magnetic fields up to 20 t. *Phys. Rev. A*, 58:3114–3123, Oct 1998.

[46] Jakob Hayden, Jonas Westberg, Charles Link Patrick, Bernhard Lendl, and Gerard Wysocki. Frequency-locked cavity ring-down faraday rotation spectroscopy. *Opt. Lett.*, 43(20):5046–5049, Oct 2018.

[47] Charles Link Patrick, Jonas Westberg, and Gerard Wysocki. Cavity Attenuated Phase Shift Faraday Rotation Spectroscopy. *Analytical Chemistry*, 91(3):1696–1700, feb 2019.

[48] Michele Gianella, Sioned A. Press, Katherine M. Manfred, Helen C. Norman, Meez Islam, and Grant A. D. Ritchie. Sensitive detection of ho2 radicals produced in an atmospheric pressure plasma using faraday rotation cavity ring-down spectroscopy. *The Journal of Chemical Physics*, 151(12):124202, 2019.

[49] M Gianella, S Reuter, S A Press, A Schmidt-Bleker, J H van Helden, and G A D Ritchie. HO2 reaction kinetics in an atmospheric pressure plasma jet determined by cavity ring-down spectroscopy. *Plasma Sources Science and Technology*, 27(9):095013, sep 2018.

[50] Daniele Romanini, AA Kachanov, N Sadeghi, and F Stoeckel. CW cavity ring down spectroscopy. *Chemical Physics Letters*, 264(3-4):316–322, 1997.

[51] A. C. Nilsson, E. K. Gustafson, and R. L. Byer. Eigenpolarization theory of monolithic nonplanar ring oscillators. *IEEE Journal of Quantum Electronics*, 25(4):767–790, April 1989.

[52] M Vallet, F Bretenaker, A Le Floch, R Le Naour, and M Oger. The malus fabryâperot interferometer. *Optics Communications*, 168(5):423 – 443, 1999.

[53] L. Bougas, G. E. Katsoprinakis, W. von Klitzing, and T. P. Rakitzis. Fundamentals of cavity-enhanced polarimetry for parity-nonconserving optical rotation measurements: Application to xe, hg, and i. *Phys. Rev. A*, 89:052127, May 2014.

[54] D. Budker, W. Gawlik, D. F. Kimball, S. M. Rochester, V. V. Yashchuk, and A. Weis. Resonant nonlinear magneto-optical effects in atoms. *Rev. Mod. Phys.*, 74:1153–1201, Nov 2002.

[55] Dimitris Sofikitis, George E. Katsoprinakis, Alexandros K. Spiliotis, and T. Peter Rakitzis. Chapter 16 - cavity-based chiral polarimetry. In Prasad L. Polavarapu, editor, *Chiral Analysis (Second Edition)*, pages 649 – 678. Elsevier, second edition edition, 2018.

[56] Eric D. Black. An introduction to pound-drever-hall laser frequency stabilization. *American Journal of Physics*, 69(1):79–87, 2001.

[57] D. Sofikitis, A. K. Spiliotis, K. Stamataki, G. E. Katsoprinakis, L. Bougas, P. C. Samartzis, B. Loppinet, T. P. Rakitzis, M. Surligas, and S. Papadakis. Microsecond-resolved SDR-based cavity ring-down ellipsometry. *Appl. Opt.*, 54(18):5861–5865, Jun 2015.

[58] Rei Kitamura, Laurent Pilon, and Miroslaw Jonasz. Optical constants of silica glass from extreme ultraviolet to far infrared at near room temperature. *Appl. Opt.*, 46(33):8118–8133, Nov 2007.

[59] S Ramaseshan. Determination of the magneto-optic anomaly of some glasses. *Proceedings Mathematical Sciences*, 24(5):426–432, 1946.

[60] S Ramaseshan and V Sivaramakrishnan. *Faraday effect in diamagnetic crystals*, volume 1. Interscience, New York, 1958.

[61] Kiyoshi Shimamura, EncarnaciÃ³n G VÄ±Ìllora, Satoshi Nakakita, Martin Nikl, and Noboru Ichinose. Growth and scintillation characteristics of cef3, prf3 and ndf3 single crystals. *Journal of Crystal Growth*, 264(1):208 – 215, 2004.

[62] P. Molina, V. Vasyliev, E. G. Víllora, and K. Shimamura. Cef3 and prf3 as uv-visible faraday rotators. *Opt. Express*, 19(12):11786–11791, Jun 2011.

[63] LR Ingersoll and DH Liebenberg. The faraday effect in gases and vapors. i. *JOSA*, 44(7):566–571, 1954.

[64] L. R. Ingersoll and D. H. Liebenberg. Faraday effect in gases and vapors. ii∗. *J. Opt. Soc. Am.*, 46(7):538–542, Jul 1956.

[65] Ying-Xian Yao and S. M. Pandit. Cramér-rao lower bounds for a damped sinusoidal process. *IEEE Transactions on Signal Processing*, 43(4):878–885, April 1995.

[66] C. Gemmel, W. Heil, S. Karpuk, K. Lenz, Ch. Ludwig, Yu. Sobolev, K. Tullney, M. Burghoff, W. Kilian, S. Knappe-Grüneberg, W. Müller, A. Schnabel, F. Seifert, L. Trahms, and St. Baeßler. Ultra-sensitive magnetometry based on free precession of nuclear spins. *Eur. Phys. J. D*, 57(3):303–320, Apr 2010.

[67] Hans-Christian Koch, Georg Bison, Zoran D. Grujić, Werner Heil, Malgorzata Kasprzak, Paul Knowles, Andreas Kraft, Anatoly Pazgalev, Allard Schnabel, Jens Voigt, and Antoine Weis. Design and performance of an absolute 3He/Cs magnetometer. *Eur. Phys. J. D*, 69(8):202, Aug 2015.

[68] D. Hunter, S. Piccolomo, J. D. Pritchard, N. L. Brockie, T. E. Dyer, and E. Riis. Free-induction-decay magnetometer based on a microfabricated Cs vapor cell. *Phys. Rev. Applied*, 10:014002, Jul 2018.

[69] Jonas Westberg and Gerard Wysocki. Cavity ring-down faraday rotation spectroscopy for oxygen detection. *Applied Physics B*, 123(5):168, May 2017.

[70] J Morville, S Kassi, M Chenevier, and D Romanini. Fast, low-noise, mode-by-mode, cavity-enhanced absorption spectroscopy by diode-laser self-locking. *Applied Physics B*, 80(8):1027–1038, 2005.

[71] Johannes Burkart, Daniele Romanini, and Samir Kassi. Optical feedback frequency stabilized cavity ring-down spectroscopy. *Opt. Lett.*, 39(16):4695–4698, Aug 2014.

[72] Yin Wang, Michal Nikodem, Eric Zhang, Frank Cikach, Jarrod Barnes, Suzy Comhair, Raed A Dweik, Christina Kao, and Gerard Wysocki. Shot-noise Limited Faraday Rotation Spectroscopy for Detection of Nitric Oxide Isotopes in Breath, Urine, and Blood. *Scientific Reports*, 5:9096, mar 2015.

[73] Ben Henderson, Amir Khodabakhsh, Markus Metsälä, Irène Ventrillard, Florian M. Schmidt, Daniele Romanini, Grant A. D. Ritchie, Sacco te Lintel Hekkert, Raphaël Briot, Terence Risby, Nandor Marczin, Frans J. M. Harren, and Simona M. Cristescu. Laser spectroscopy for breath analysis: towards clinical implementation. *Applied Physics B*, 124(8):161, Jul 2018.

[74] Harald Günther. *NMR spectroscopy: basic principles, concepts and applications in chemistry*. John Wiley & Sons, 2013.

[75] I. M. Savukov and M. V. Romalis. NMR detection with an atomic magnetometer. *Phys. Rev. Lett.*, 94:123001, Mar 2005.

[76] Anna Nikiel, Peter Blümler, Werner Heil, Manfred Hehn, Sergej Karpuk, Andreas Maul, Ernst Otten, Laura M Schreiber, and Maxim Terekhov. Ultrasensitive 3He magnetometer for measurements of high magnetic fields. *Eur. Phys. J. D*, 68(11):330, 2014.

[77] Zoran D Grujić, Peter A Koss, Georg Bison, and Antoine Weis. A sensitive and accurate atomic magnetometer based on free spin precession. *Eur. Phys. J. D*, 69(5):135, 2015.

[78] Dominic Hunter, Ricardo Jiménez-Martínez, Juan Herbsommer, Srinath Ramaswamy, Wen Li, and Erling Riis. Waveform reconstruction with a Cs based free-induction-decay magnetometer. *Opt. Express*, 26(23):30523–30531, Nov 2018.

[79] Jim C Visschers, Oleg Tretiak, Dmitry Budker, and Lykourgos Bougas. Continuous-wave cavity ring-down polarimetry. *The Journal of Chemical Physics*, 152(16):164202, 2020.

[80] Vassilis Papadakis, Michael A. Everest, Katerina Stamataki, Stelios Tzortzakis, Benoit Loppinet, and T. Peter Rakitzis. Development of cavity ring-down ellipsometry with spectral and submicrosecond time resolution. In Michael T. Postek, editor, *Instrumentation, Metrology, and Standards for Nanomanufacturing, Optics, and Semiconductors V*, volume 8105, pages 104 – 112. International Society for Optics and Photonics, SPIE, 2011.

[81] Katerina Stamataki, Vassilis Papadakis, Michael A Everest, Stelios Tzortzakis, Benoit Loppinet, and T Peter Rakitzis. Monitoring adsorption and sedimentation using evanescent-wave cavity ringdown ellipsometry. *Applied Optics*, 52(5):1086–1093, 2013.

[82] Dimitris Sofikitis, Katerina Stamataki, Michael A. Everest, Vassilis Papadakis, Jean-Louis Stehle, Benoit Loppinet, and T. Peter Rakitzis. Sensitivity enhancement for evanescent-wave sensing using cavity-ring-down ellipsometry. *Opt. Lett.*, 38(8):1224–1226, Apr 2013.

[83] Hakho Lee, Eric Sun, Donhee Ham, and Ralph Weissleder. Chip-NMR biosensor for detection and molecular analysis of cells. *Nature Med.*, 14(8):869–874, 2008.

[84] Juan Perlo, Vasiliki Demas, Federico Casanova, Carlos A. Meriles, Jeffrey Reimer, Alexander Pines, and Bernhard Blümich. High-resolution NMR spectroscopy with a portable single-sided sensor. *Science*, 308(5726):1279–1279, 2005.

[85] K. Lei, H. Heidari, P. Mak, M. Law, F. Maloberti, and R. P. Martins. A handheld high-sensitivity micro-NMR CMOS platform with B-field stabilization for multi-type biological/chemical assays. *IEEE Journal of Solid-State Circuits*, 52(1):284–297, 2017.

[86] Ka-Meng Lei, Dongwan Ha, Yi-Qiao Song, Robert M Westervelt, Rui Martins, Pui-In Mak, and Donhee Ham. Portable NMR with parallelism. *Anal. Chem.*, 92(2):2112–2120, jan 2020.

[87] R Prigl, U Haeberlen, K Jungmann, G zu Putlitz, and P von Walter. A high precision magnetometer based on pulsed NMR. *Nucl. Instrum. Methods. Phys. Res. B*, 374(1):118 – 126, 1996.

[88] H. Dong, H. Liu, J. Ge, Z. Yuan, and Z. Zhao. A high-precision frequency measurement algorithm for FID signal of proton magnetometer. *IEEE T. Instrum. Meas.*, 65(4):898–904, 2016.

[89] Daniel Halmer, Golo von Basum, Peter Hering, and Manfred Mürtz. Fast exponential fitting algorithm for real-time instrumental use. *Rev. Sci. Instrum.*, 75(6):2187–2191, 2004.

[90] M Mazurenka, R Wada, A J L Shillings, T J A Butler, J M Beames, and A J Orr-Ewing. Fast Fourier transform analysis in cavity ring-down spectroscopy: application to an optical detector for atmospheric $NO_2$. *Appl. Phys. B*, 81(1):135–141, 2005.

[91] Michael A Everest and Dean B Atkinson. Discrete sums for the rapid determination of exponential decay constants. *Review of Scientific Instruments*, 79(2):023108, 2008.

[92] T. G. Spence, M. E. Calzada, H. M. Gardner, E. Leefe, H. B. Fontenot, L. Gilevicius, R. W. Hartsock, T. K. Boyson, and C. C. Harb. Real-time FPGA data collection of pulsed-laser cavity ringdown signals. *Opt. Express*, 20(8):8804–8814, Apr 2012.

[93] G Bostrom, D Atkinson, and A Rice. The discrete fourier transform algorithm for determining decay constants—implementation using a field programmable gate array. *Review of Scientific Instruments*, 86(4):043106, 2015.

[94] Elias Aboutanios. Estimating the parameters of sinusoids and decaying sinusoids in noise. *IEEE Instrumentation & Measurement Magazine*, 14(2):8–14, 2011.

[95] Elias Aboutanios. Estimation of the frequency and decay factor of a decaying exponential in noise. *IEEE Transactions on Signal Processing*, 58(2):501–509, 2009.

[96] Jorge J Moré. The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[97] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[98] T. K. Boyson, T. G. Spence, M. E. Calzada, and C. C. Harb. Frequency domain analysis for laser-locked cavity ring-down spectroscopy. *Opt. Express*, 19(9):8092–8101, Apr 2011.

[99] J. J. More, B. S. Garbow, and K. E. Hillstrom. User guide for MINPACK-1. [In FORTRAN]. 8 1980.

[100] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. Acoust.*, 38(5):814–824, 1990.

[101] F. Sarrazin, A. Sharaiha, P. Pouliguen, J. Chauveau, S. Collardey, and P. Potier. Comparison between matrix pencil and Prony methods applied on noisy antenna responses. In *2011 Loughborough Antennas Propagation Conference*, pages 1–4, 2011.

[102] Yung-Ya Lin, Paul Hodgkinson, Matthias Ernst, and Alexander Pines. A novel detection estimation scheme for noisy NMR signals: Applications to delayed acquisition data. *J. Magn. Reson.*, 128(1):30 – 41, 1997.

[103] S.N. Fricke, J.D. Seymour, M.D. Battistel, D.I. Freedberg, C.D. Eads, and M.P. Augustine. Data processing in NMR relaxometry using the matrix pencil. *J. Magn. Reson.*, 313:106704, 2020.

[104] Haifeng Huang and Kevin K Lehmann. Sensitivity limits of continuous wave cavity ring-down spectroscopy. *J. Phys. Chem. A*, 117(50):13399–13411, dec 2013.

[105] Daniel P. Zaleski and Kirill Prozument. Automated assignment of rotational spectra using artificial neural networks. *J. Chem. Phys.*, 149(10):104–106, 2018.

[106] Roland Tóbiás, Tibor Furtenbacher, Irén Simkó, Attila G Császár, Meissa L Diouf, Frank M J Cozijn, Joey M A Staa, Edcel J Salumbides, and Wim Ubachs. Spectroscopic-network-assisted precision spectroscopy and its application to water. *Nature Comm.*, 11(1):1708, 2020.

[107] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.

[108] John A Hertz. *Introduction to the theory of neural computation*. CRC Press, 2018.

[109] Chen Gao, Ayush Saraf, Jia-Bin Huang, and Johannes Kopf. Flow-edge guided video completion. In *European Conference on Computer Vision*, pages 713–729. Springer, 2020.

[110] Ariel Shamir, Niloy J Mitra, Nobuyuki Umetani, and Yuki Koyama. Intelligent tools for creative graphics. In *ACM SIGGRAPH 2020 Courses*, pages 1–11. 2020.

[111] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

[112] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

[113] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning*, pages 8459–8468. PMLR, 2020.

[114] John Peurifoy, Yichen Shen, Li Jing, Yi Yang, Fidel Cano-Renteria, Brendan G DeLacy, John D Joannopoulos, Max Tegmark, and Marin Soljačić. Nanophotonic particle simulation and inverse design using artificial neural networks. *Science advances*, 4(6):eaar4206, 2018.

[115] Jinmei Wang, Xiangyu Liao, Peichao Zheng, Shuwen Xue, and Rui Peng. Classification of chinese herbal medicine by laser-induced breakdown spectroscopy with principal component analysis and artificial neural network. *Analytical letters*, 51(4):575–586, 2018.

[116] FG Del Moral, A Guillén, LG Del Moral, F O'valle, L Martínez, and RG Del Moral. Duroc and iberian pork neural network classification by visible and near infrared reflectance spectroscopy. *Journal of Food Engineering*, 90(4):540–547, 2009.

[117] Monika Gniadecka, Peter Alshede Philipsen, Sonja Wessel, Robert Gniadecki, Hans Christian Wulf, Sigurdur Sigurdsson, Ole Faurskov Nielsen, Daniel Højgaard Christensen, Jana Hercogova, Kristian Rossen, et al. Melanoma diagnosis by raman spectroscopy and neural networks: structure alterations in proteins and lipids in intact cancer tissue. *Journal of investigative dermatology*, 122(2):443–449, 2004.

[118] N Nottbeck, C Schmitt, and V Büscher. Implementation of high-performance, sub-microsecond deep neural networks on FPGAs for trigger applications. *Journal of Instrumentation*, 14(09):P09014, 2019.

[119] Jenna Burrell. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1):2053951715622512, 2016.

[120] IM Savukov and MV Romalis. NMR detection with an atomic magnetometer. *Physical review letters*, 94(12):123001, 2005.

[121] Anna Nikiel, Peter Blümler, Werner Heil, Manfred Hehn, Sergej Karpuk, Andreas Maul, Ernst Otten, Laura M Schreiber, and Maxim Terekhov. Ultrasensitive $^3$He magnetometer for measurements of high magnetic fields. *The European Physical Journal D*, 68(11):1–12, 2014.

[122] Zoran D Grujić, Peter A Koss, Georg Bison, and Antoine Weis. A sensitive and accurate atomic magnetometer based on free spin precession. *The European Physical Journal D*, 69(5):135, 2015.

[123] Giel Berden, Rudy Peeters, and Gerard Meijer. Cavity ring-down spectroscopy: Experimental schemes and applications. *International reviews in physical chemistry*, 19(4):565–607, 2000.

[124] Zhiyan Li, Renzhi Hu, Pinhua Xie, Hao Chen, Xiaoyan Liu, Shuaixi Liang, Dan Wang, Fengyang Wang, Yihui Wang, Chuan Lin, et al. Simultaneous measurement of NO and $NO_2$ by a dual-channel cavity ring-down spectroscopy technique. *Atmospheric Measurement Techniques*, 12(6):3223–3236, 2019.

[125] Dimitris Sofikitis, Katerina Stamataki, Michael A Everest, Vassilis Papadakis, Jean-Louis Stehle, Benoit Loppinet, and T Peter Rakitzis. Sensitivity enhancement for evanescent-wave sensing using cavity-ring-down ellipsometry. *Optics Letters*, 38(8):1224–1226, 2013.

[126] D Sofikitis, AK Spiliotis, K Stamataki, GE Katsoprinakis, L Bougas, PC Samartzis, B Loppinet, TP Rakitzis, M Surligas, and S Papadakis. Microsecond-resolved sdr-based cavity ring down ellipsometry. *Applied Optics*, 54(18):5861–5865, 2015.

[127] Daniel Halmer, Golo von Basum, Peter Hering, and Manfred Mürtz. Fast exponential fitting algorithm for real-time instrumental use. *Review of scientific instruments*, 75(6):2187–2191, 2004.

[128] M Mazurenka, R Wada, AJL Shillings, TJA Butler, JM Beames, and AJ Orr-Ewing. Fast fourier transform analysis in cavity ring-down spectroscopy: application to an optical detector for atmospheric $NO_2$. *Applied Physics B*, 81(1):135–141, 2005.

[129] Jim C. Visschers, Emma Wilson, Thomas Conneely, Andrey Mudrov, and Lykourgos Bougas. Rapid parameter determination of discrete damped sinusoidal oscillations. *Opt. Express*, 29(5):6863–6878, Mar 2021.

[130] F. van Veen & S. Leijnen. The neural network zoo, 2019.

[131] Hervé Bourlard and Yves Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59(4-5):291–294, 1988.

[132] David DeMers and Garrison W Cottrell. Non-linear dimensionality reduction. In *Advances in neural information processing systems*, pages 580–587, 1993.

[133] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[134] Lucas Theis, Wenzhe Shi, Andrew Cunningham, and Ferenc Huszár. Lossy image compression with compressive autoencoders. *arXiv preprint arXiv:1703.00395*, 2017.

[135] Xugang Lu, Yu Tsao, Shigeki Matsuda, and Chiori Hori. Speech enhancement based on deep denoising autoencoder. In *Interspeech*, volume 2013, pages 436–440, 2013.

[136] Andrew Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011.

[137] Andrea Asperti and Matteo Trentin. Balancing reconstruction error and kullback-leibler divergence in variational autoencoders. *IEEE Access*, 8:199440–199448, 2020.

[138] Gianluca Gagliardi and Hans-Peter Loock. *Cavity-enhanced spectroscopy and sensing*, volume 179. Springer, 2014.

[139] Noël Boens, Wenwu Qin, Nikola Basarić, Johan Hofkens, Marcel Ameloot, Jacques Pouget, Jean-Pierre Lefevre, Bernard Valeur, Enrico Gratton, Martin VandeVen, et al. Fluorescence lifetime standards for time and frequency domain fluorescence spectroscopy. *Analytical chemistry*, 79(5):2137–2149, 2007.

[140] R Cundall. *Time-resolved fluorescence spectroscopy in biochemistry and biology*, volume 69. Springer Science & Business Media, 2013.

[141] AK Spiliotis, M Xygkis, E Klironomou, E Kardamaki, GK Boulogiannis, GE Katsoprinakis, D Sofikitis, and TP Rakitzis. Optical activity of lysozyme in solution at 532 nm via signal-reversing cavity ring-down polarimetry. *Chemical Physics Letters*, page 137345, 2020.

[142] Vassilis Papadakis, Michael A Everest, Katerina Stamataki, Stelios Tzortzakis, Benoit Loppinet, and T Peter Rakitzis. Development of cavity ring-down ellipsometry with spectral and submicrosecond time resolution. In *Instrumentation, Metrology, and Standards for Nanomanufacturing, Optics, and Semiconductors V*, volume 8105, page 81050L. International Society for Optics and Photonics, 2011.

[143] François Chollet et al. Keras. `https://keras.io`, 2015.

[144] Edward H Wahl, Sze M Tan, Sergei Koulikov, Boris Kharlamov, Christopher R Rella, Eric R Crosson, Dave Biswell, and Barbara A Paldus. Ultra-sensitive ethylene post-harvest monitor based on cavity ring-down spectroscopy. *Optics express*, 14(4):1673–1684, 2006.