Time-domain Compressed Sensing

Kilian Scheffter, Jonathan Will, Claudius Riek, Herve Jousselin, Sebastien Coudreau, Nicolas Forget, Hanieh Fattahi

Time-domain Compressed Sensing

Kilian Scheffter[1,2], Jonathan Will[1,2], Claudius Riek[3], Herve Jousselin[4], Sébastien Coudreau[4], Nicolas Forget[4], Hanieh Fattahi[1,2]

1. Max Planck Institute for the Science of Light, Staudtstr. 2, 91058 Erlangen, Germany
2. Friedrich-Alexander University Erlangen-Nürnberg, Staudtrstr. 7, 91085 Erlangen, Germany
3. Zurich Instruments Germany, Mühldorfstraße 15, 81671 Munich, Germany
4. Fastlite, rue des Cistes 165, 06600 Antibes, France

Abstract

Ultrashort time-domain spectroscopy, particularly field-resolved spectroscopy, are established methods for identifying the constituents and internal dynamics of samples. However, these techniques are often encumbered by the Nyquist criterion, leading to prolonged data acquisition and processing times as well as sizable data volumes. To mitigate these issues, we have successfully implemented the first instance of time-domain compressed sensing, enabling us to pinpoint the primary absorption peaks of atmospheric water vapor in response to terahertz light transients that exceed the Nyquist limit. Our method demonstrates successful identification of water absorption peaks up to 2.5 THz, even for sampling rates where the Nyquist frequency is as low as 0.75 THz, with a mean squared error of 12×10^{-4}. Time-domain sparse sampling achieves considerable data compression while also expediting both the measurement and data processing time, representing a significant stride towards the realm of real-time spectroscopy.

Keywords: sparse sampling, field-resolved spectroscopy, time-domain spectroscopy, compressed sensing, real-time detection

Introduction

Detailed description of the constituent and internal dynamics of matter is mirrored in its transient response to an external field. Resolving and monitoring the encoded information in the electric field of the ultrashort excitation field or in the time-dependent changes of the optical properties of the sample, provides deep understanding and insights of matter [1, 2]. The availability of few-cycle pulses has not only advanced various spectroscopic methods such as pump-probe spectroscopy and Fourier transform spectroscopy, but has also led to the emergence of innovative techniques such as dual-comb spectroscopy and field-resolved spectroscopy [3–14].

Employing ultrashort pulses offers two distinct advantages for spectroscopic applications (Fig. 1). Firstly, their broad spectral bandwidth enables simultaneous data acquisition of the sample, eliminating the need for repeated measurements or laser tuning. With high-bandwidth acquisition, prior knowledge of the sample is not required, as all available information can be extracted from the measurement during post-processing. Secondly, their extreme temporal confinement allows for temporal gating of the sample's response from the excitation pulses. This response, which is enriched with comprehensive spectroscopic information, lasts from tens of femtoseconds to nanoseconds and is commonly probed by a shorter pulse at various time intervals. When combined with additional temporal or spatial dimensions, such as multi-dimensional coherent spectroscopy [15], four-dimensional imaging [16], and hyperspectral imaging [17], ultrashort spectroscopic techniques render quantitative, multivariate characterization of the sample under scrutiny and facilitate the identification of unknown constituents. However, real-time measurements are prevented due to the need to record the high bandwidth spectrum at each pixel image and time delay, which results in a prohibitively long acquisition time to attain sufficient signal-to-noise ratio. The measurement speed is limited to i) the required number of sample points dictated by the Nyquist-Shannon criteria, ii) the speed of spatio-temporal scanning, and iii) transportation and storage speed of the measured data. Although the use of short, high-bandwidth excitation pulses offers significant benefits, such as simultaneous data acquisition, the response of the sample is often a linear combination of K basis vectors leading to a K-sparse frequency domain. To minimize data redundancy and obtain the desired spectral content with the minimum amount of data necessary, it is necessary to apply algorithms to the acquisition process. According to the Nyquist-Shannon sampling theorem, a signal having a frequency f needs to be sampled at a minimum sample rate f_{samp} of twice the frequency of the signal ($f_{\text{samp}} \geq 2 \cdot f$). This criterion sets a minimum requirement for the amount of measured data necessary for a successful sampling. Compressed Sensing can circumvent this fundamental barrier by exploiting sparsity and incoherence of the signal, enabling the recovery of the original signal from fewer samples or measurements while preserving its quality [18–23]. The prior knowledge of a signal's sparsity enables the formulation of an optimization problem that allows for the reconstruction of the signal using a reduced number of sampling points below the Nyquist-Shannon criteria. The sparsity of the measured spectroscopic information of various materials has been discussed by analytical treatment on over-sampled measurements [24–28] or analytical compression prior to storage or transmission [29, 30]. However, the requirements of incoherence and random sampling have

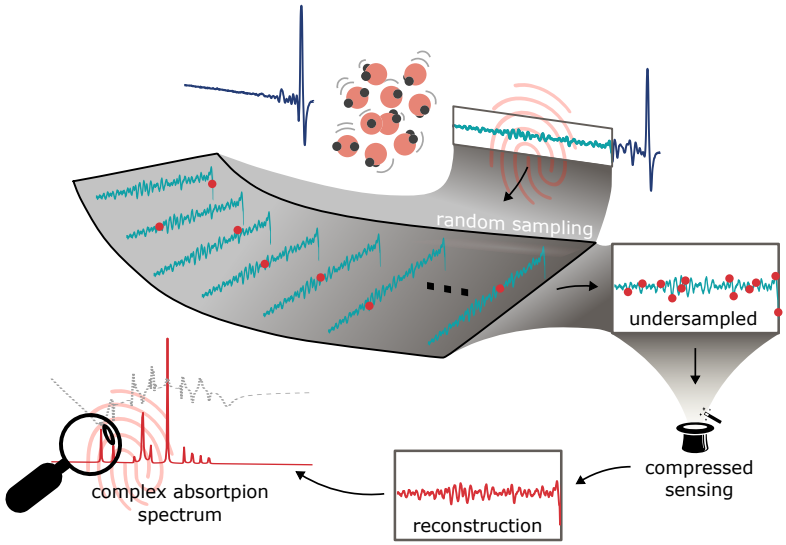


Figure 1: Visual summary of time-domain compressed sensing.

made real-time, time-domain sparse sampling challenging. In this work we demonstrate real-time, field-resolved compressed sensing of water vapour molecules beyond the Nyquist criteria. The reconstruction of the absorption spectrum of water vapor sets clear boundaries on the required sparsity of a sample for the effectiveness of compressed sensing, as the absorption spectrum includes both high cross-section peaks, and low amplitude adjacent peaks. Our approach is enabled by randomly sampling, rapidly scanning delay line, enabling sensitive, real-time sample analysis.

Results

Experimental concept

Fig. 2a shows the experimental setup for terahertz (THz) field-resolved compressed sensing. 2.1 mJ, 54 fs pulses of a Ti:Sapphire amplifier operating at 1 kHz are used to generate THz transients via optical rectification in a 1 mm-thick ZnTe crystal (see Fig. 2b and c). The THz pulses propagated through a box filled and sealed with water vapour molecules at 50% concentration. To characterize the electric field of the THz pulses, an electro-optic sampling (EOS) setup incorporating a 0.1 mm-thick ZnTe was developed. 40 µW of the amplifier's output power is used to probe the THz pulses at the EOS stage, where the polarization changes of the probe pulses due to interaction with THz electric field strength is detected in an ellipsometer incorporating balanced photo-diodes. Afterwards, the detected signal is fed into a lock-in amplifier. A mechanical chopper is used to modulate the repetition frequency of the THz pulses at 500 Hz. The modulated signal in combination with a boxcar filter is used to eliminate systematic drifts of the measurement. Moreover, the box-car filter of the lock-in amplifier ensures the high bandwidth detection of the signal during the random scanning. For real-time, random scanning of the probe pulses over the water vapour's molecular response, an acousto-optical delay line with kHz scanning rates was integrated into the probe's beam path [31, 32]. The acousto-optic delay line allows for arbitrary relative time delay between the THz pulse and probe pulses and shot-to-shot random scanning of the electric field. Data acquisition is performed by the lock-in amplifier triggered by the radio frequency control signals from the acousto-optic delay line. Data acquisition trigger (green) and synchronization paths (red) are indicated in Fig. 2a by dashed lines. The time-delay module has the refresh time of 2 ms at an arbitrary temporal position within its scanning range. In order to capture the complete molecular response encoded in the THz field, which lasts for tens of picoseconds, a mechanical delay-line was added to the acousto-optic delay line, extending the scanning range of 6400 fs for a single acousto-optical delay line to > 40 ps. Alternatively, the scanning range can be extended by coupling multiple acousto-optic modulators.

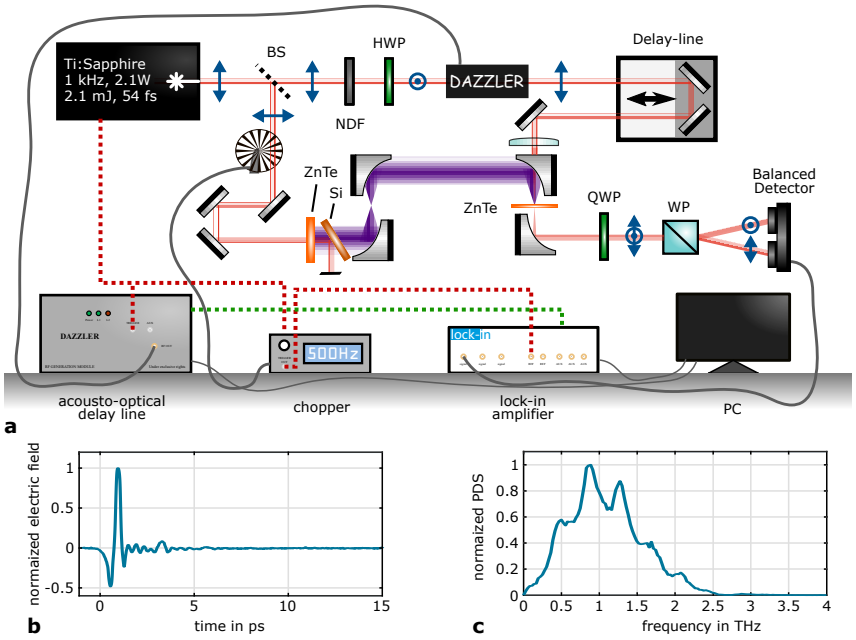


Figure 2: a) Details of the experimental setup. b) Field-resolved measurement of the THz pulses via electro-optic sampling. c) Spectrum of the THz pulses retrieved from the time-domain measurement. BS: beam splitter, HWP: half wave plate, NDF: neutral density filter, QWP: quarter wave plate, Si-plate: silicon-plate, ZnTe: Zinc Telluride, WP: Wollaston prism

Reconstruction strategy

Two categories of compressed sensing algorithms were investigated for the reconstruction of the electric field of the randomly sampled molecular response of water: convex compressed sensing and greedy algorithm (see Fig. 3). Basis Pursuit Denoising (BPD) and Lasso algorithms [33, 34] from the family of convex compressed sensing algorithms were used for reconstruction, due to their noise robustness and low reconstruction error for signals with moderate sparsity [19, 35]. When comparing the mean squared error of the reconstruction for both algorithms, it was noted that the Lasso algorithm has a unique global minimum at a specific threshold value in the optimization problem, while the BPD algorithm demonstrates a negligible mean squared error across a range of threshold parameters (τ and σ in supplementary information). BPD is also recognized for its ability to function as a filter by eliminating noise from oversampled signals [36]. As finding a reference for arbitrary measurements to optimize the reconstruction threshold value is not always feasible, BPD was selected for further analysis. From the second category, the Stagewise Orthogonal Matching Pursuit (StOMP) greedy algorithm was chosen and developed owing to its fast computation and robustness to noise [19]. To address the issue of significant amplitude reconstruction error in StOMP, a Nonlinear Least Square (NLS) algorithm has been incorporated into the StOMP reconstruction. Additionally, StOMP necessitates a distinct reconstruction threshold value at every iteration, which is inefficient and impractical. Consequently, an approach based on the Interquartile Range method [37] was developed to calculate the threshold parameter automatically (see supplementary information for more detail).

Real-time field-resolved compressed sensing

As a proof of principle experiment for time domain compressed sensing, we chose to resolve the spectroscopic information of atmospheric water vapor in response the THz excitation pulses centered at 1 THz. Among the most abundant molecules in the atmosphere, only water possesses a permanent dipole in this spectral range [38]. As a result, the spectral coverage of THz excitation pulses serves as a filter, isolating the study to only water vapor and its isotopes [39]. The ambient air's absorption spectrum in this spectral range is characterized by a high density of absorption peaks [40], which makes it an ideal platform for assessing the efficacy of compressed sensing in reconstructing moderately sparse spectra while also highlighting its limitations in reconstructing prominent absorption peaks and adjunct frequencies. Moreover, such realization holds promise for real-time gas detection in open-air environments [41, 42]. For a comparison between conventional sampling and Compressed sensing, two categories of measurements with different number of sampling points were performed on the water vapor response:

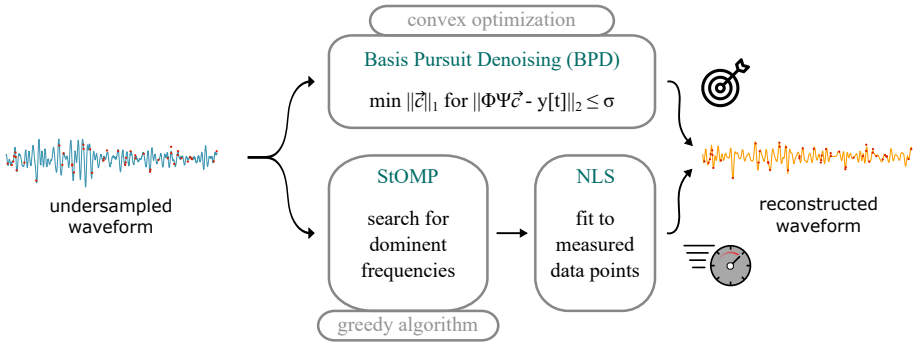


Figure 3: The reconstruction algorithms used for the data analysis of the sparse measurements: The BPD and StOMP algorithms. While StOMP has the benefit of fast data processing, it requires an input reconstruction threshold value. To overcome this inefficiency IQR method was used to search for the dominant frequencies while NLS algorithm is used to optimize the amplitude of the reconstructed traces.

i) linear scans with sampling points at the time intervals of $t_{n+1} = t_n + \Delta t$, and ii) randomly distributed sampling points. Each measurement was repeated for ten times. Hereby, the location of the randomly located sample points of the second category varies for every measurement. To generate a reference, ten oversampled, linearly scanned measurements with $N = 2034$ sample points are averaged. This reference trace is shown by the blue curve in Fig. 4a. For the spectroscopic analysis, only the sparse molecular response in the decaying tail of the THz light transient, ranging from 5 to 35 ps was considered. The upper limit of this range was determined by the internal reflection of the THz light transient in the sampling crystal, occurring at 36 ps. The sampling points were reduced from 2023 to 46 data points in eighteen discrete steps and the reconstruction was performed by using both BPD and StOMP algorithms.

Fig. 4a displays a measurement at an extreme limit, with the minimum random number of sample points at $N = 46$, indicated by red dots, along with the reconstructed fields obtained using the BPD algorithm (in orange) and the StOMP algorithm (in green). The averaged Fourier transformations of ten reconstructions, as well as the Fourier transformation of a measurement with uniformly distributed $N = 46$ sampling point are shown in Fig. 4b. The uniformly sampled waveform with $N = 46$ has temporal steps of $\Delta t = 666$ fs, corresponding to a Nyquist frequency of 0.75 THz. The resulting spectrum clearly displays aliasing and incorporate fake frequencies at below 0.5 THz. However, the spectrum obtained from the compressed sensing reconstructed waveform successfully retrieves water absorption peaks beyond the Nyquist limit, up to 2.2 THz. To achieve a similar outcome using conventional sampling, the number of sampling points would need to be increased by at least three times.

To quantitatively evaluate the performance of compressed sensing, we calculated the mean squared error between the oversampled reference and the reconstructed waveforms via BPD and StOMP, respectively. Fig. 4c shows the average and standard deviation of the mean squared error as a function of N for each algorithms, while the purple curve represents the corresponding Nyquist frequency. We determined the value of the measurement noise by averaging the mean squared error of the conventionally sampled waveforms with respect to the oversampled reference for each N. As this value exhibits only slight variations across different N, it is indicated by a horizontal dotted line (see supplementary information for more details). Not only does the BPD algorithm outperform the StOMP method, but for larger values of N, the waveforms reconstructed using BPD show a mean squared error that is lower than the measurement noise. This is due to the tendency of BPD to converge to a solution with maximum sparsity, which acts as a filter rejecting signal amplitudes other than absorption frequencies. For lower values of N, the mean squared error increases non-linearly, reaching 12×10^{-4} at $N = 46$. The low standard deviation between the ten different measurements for each value of N demonstrates the stability of the reconstructions. Fig. 4d and e summarize the performance of the BPD and StOMP reconstructions for measurements at various N, where the averaged Fourier transformation of the ten compressed sensing reconstructed waveforms at different N are presented. The amplitude at each N is normalized to one, and the area bordered by the purple line denotes the recovered frequencies beyond the Nyquist limit. The spectral region ranging from 1 THz to 1.5 THz is crucial for this investigation, due to the presence of a large number of absorption lines with high cross-section and the high spectral density of the THz excitation pulses. Remarkably, the primary absorption peaks in this region are exceptionally well-recovered beyond the Nyquist limit with excellent quality. When dealing with lower values of N, the reconstruction of absorption lines with lower amplitude and the adjunct frequencies are more prone to noise, even for those below the Nyquist limit. A comparison of panel d) and e) reveals that while the less complex StOMP algorithm has the potential of a fast reconstruction speed compared to BPD, it results in noisier reconstructions of

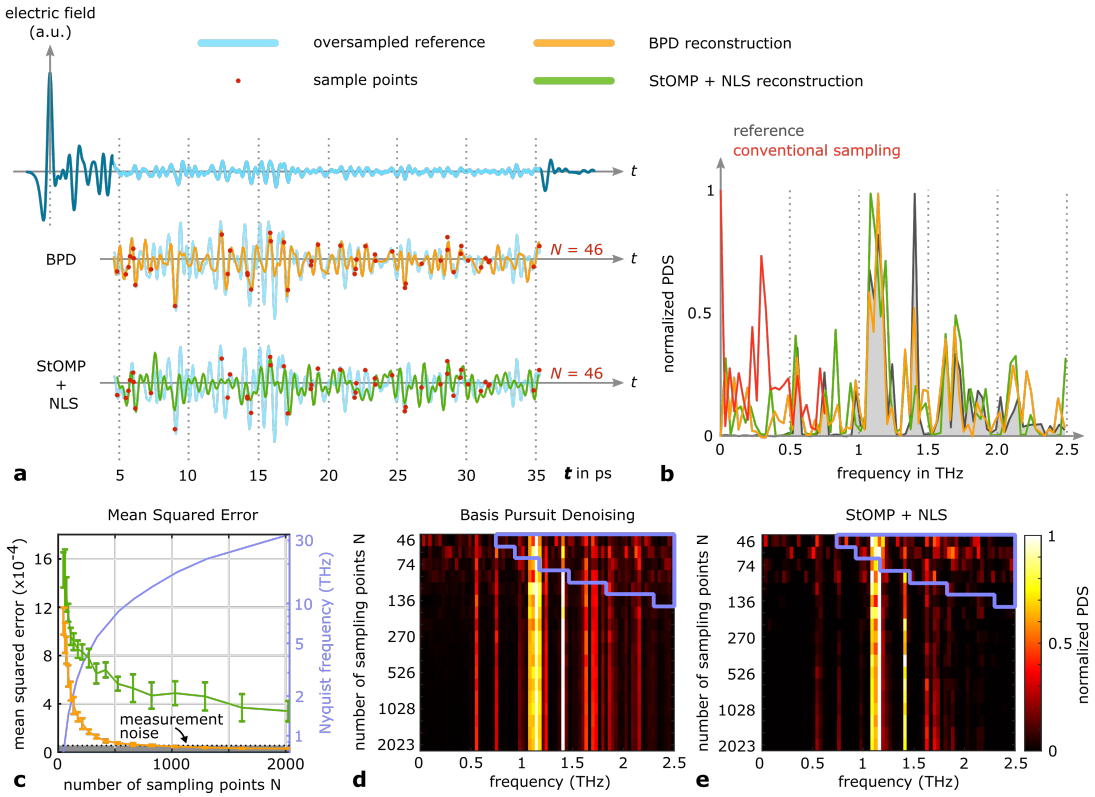


Figure 4: Real-time field-resolved compressed sensing. a) Field-resolved measurement of atmospheric water vapor response to THz excitation pulses. The light blue curve shows an oversampled measurement, while the red dots illustrate the random spars detection at $N = 46$ sampling points. The reconstructed waveform via BPD and StOMP algorithm are shown in orange and green respectively. b) Fourier transform counterpart of the sparse measurements at $N = 46$ sampling points. The red spectrum illustrate the direct Fourier transformation, while the orange and green spectra show the averaged Fourier transformation of the reconstructed waveforms via BPD and StOMP, respectively. c) Mean Squared Error of the reconstructed waveforms via BPD and StOMP at various N. d) Averaged reconstructed spectra by BPD at various N. e) Averaged reconstructed spectra by StOMP at various N. The reconstructed spectral components beyond the Nyquist limit are denoted by the bordered purple area.

both the low amplitude frequencies below the Nyquist limit and the absorption lines beyond the Nyquist limit.

Discussion

Ultrashort time-domain spectroscopy and in particular field-resolved spectroscopy have been gold standards for accurately identifying the constituents and dynamics of matter without the need for labels. Despite numerous attempts to increase measurement speed in ultrafast spectroscopy [32, 43, 44], real-time measurements remain challenging due to a prolonged acquisition time, significant data volume, and processing time. For the first time, this study showcases the use of Compressed Sensing with rapidly sampled field resolved spectroscopy, providing a solution to overcome these limitations. In particular, when short excitation laser pulses interacting with matter, the signal carrying information on this interaction are temporally separated from the main pulse turning the characterization problem to a compressed sensing problem. Crucial for compressed sensing is the possibility for random sampling which has been achieved by employing an acousto-optic delay line and a box-car filter for broadband data acquisition with a high dynamic range. The relatively high density of absorption peaks of ambient air water in terahertz spectral range provides an ideal platform for assessing the efficacy of compressed sensing in reconstructing moderately sparse spectra while also highlighting its limitations. We report on resolving the absorption frequencies three times higher than the Nyquist limit. The reconstruction success rate for the most prominent absorption peak is 100 % showing the potential of successfully reconstructing the absorption for even lower number of sampling points. However, the success rate decreases below 90 % for adjacent frequencies, emphasizing the importance of sparsity of the signal for full reconstruction (see supplementary information). Additionally, we demonstrate that

compressed sensing below the Nyquist criteria can suppress measurement noise, making it valuable not only for speeding up measurement time but also for denoising sensitive measurements.

Determining when a sufficient number of sample points have been acquired for measuring an unknown spectrum is a major challenge in compressed sensing. However, due to the ability of random sampling, the number of sampling points increases uniformly across the entire sampling range, allowing for the observation of the convergence of the reconstructed waveform or spectrum. With our compressed sensing algorithm possessing processing times below 3 ms, real time analysis of the convergence is possible. As the refresh time of the acousto-optic delay line is at 30 kHz, individual random shot to shot sampling can be performed for laser pulses below this repetition rates. For higher repetition rates, partial random scanning can be performed. Here each launched acoustic wave packet inside the acousto optical delay line delays an incoming pulse train with equal spaced delays [32], while the relative delay between different scans are randomized showing promise to introduce compressed sensing.

Real-time ultrafast spectroscopy is of crucial importance in various fields. Our innovative technique can greatly accelerate data acquisition in ultrafast spectroscopy, particularly in higher-dimensional analyses, by data volume minimization, signal acquisition time reduction, and a contraction in the required number of measurements in each dimension. These advances alleviate the requirements for specialized measurement instruments, offering a range of benefits that extend well beyond traditional spectroscopy applications. For instance simplifying the handling of fragile specimens, enabling real-time environmental monitoring of short-lived pollutants, and real-time, open-air diagnostics of toxic and hazardous gases.

Methods

The laser delivers 54 fs pulses centered at 810 nm with 2.1 mJ energy at 1 kHz repetition rate. In the setup, the diameter of the beam is first reduced from 4.1 mm to 2.8 mm at $1/e^2$ via a Gallilean beam expander with the focal length of $f = -100$ mm and $f = 150$ mm. The beam is then split into two paths via a 90:10 unpolarized beamsplitter. The reflected beam containing 90 % of energy is used for THz-generation and is chopped at a frequency of 500 Hz via a mechanical chopper. Therefore, every second pulse is blocked by the chopper. This modulation frequency in combination with a lock-in amplifier is used to reduce the systematic noise in the measurement. Afterwards, the laser pulses are sent through a THz-crystal to generate THz-pulses via optical rectification. After the THz generation, a 1.6 mm thick Silicon plate is used to filter the 800 nm pump-beam, while transmitting the THz-beam. The entire THz-beam path is placed inside a purge-box, filled with nitrogen to control the humidity level. The THz-pulses are focused on the detection crystal with a pierced parabola with the focal length of $f = 50.3$ mm, while spatially overlapped with the probe beam.

The 10 % transmitted 800 nm beam is used to probe the electric field of the THz-pulses. An acousto optical delay-line (Dazzler) is implemented in the probe beam path for fast and random delay scanning. In front of the Dazzler, the polarization of the probe is flipped via a half-waveplate to provide the required input polarization. The diffracted output beam from the Dazzler has an orthogonal polarization relative to the input beam. Eventually the probe beam is focused through a pierced parabola to the detection crystal with a plano-convex lens ($f = 130$ mm). In the detection crystal, the THz-beam co-propagates with the probe beam with altered polarization depending on the instantaneous THz field strength via the Pockels-effect. After collimation of the probe beam with an off-axis parabola ($f = 100$ mm), its polarization status is analyzed with an ellipsometric detection scheme consisting of a quarter-waveplate, a Wollaston prism, and a balanced photodetector. The signal of the balanced photodetector is fed into a lock-in amplifier using a boxcar filter to eliminate systematic drifts. Data acquisition is performed by the lock-in amplifier triggered by control signals leaving the Dazzler radio frequency module. For more details we refer the readers to the supplementary information.

Acknowledgment

The authors thank Philip J. Russell, Francesco Tani, Martin Butryn, Stefan Malzer and Heidi Potts for their support.

Author Contributions

H.F. envisioned the experiment. K.S. performed the measurements. H.J, S.C., N.F., K.S., H.F. designed the random scanning. C.R. contributed in devising the strategy for data acquisition. K.S., J.W., H.F. performed the data reconstruction and analysis. All authors reviewed and contributed to the manuscript text.

Funding

This work was supported by the research funding from Max Planck Society. K.S. is part of the Max Planck School of Photonics supported by the German Federal Ministry of Education and Research (BMBF), the Max Planck Society and the Fraunhofer Society.

Disclosures

The authors declare no conflicts of interest.

References

1. Wang, Z.-G. & Xia, H.-R. *Molecular and laser spectroscopy* (Springer Science & Business Media, 2012).
2. Coutaz, J.-L., Garet, F. & Wallace, V. *Principles of Terahertz time-domain spectroscopy* (CRC Press, 2018).
3. Zewail, A. H. Femtochemistry: Atomic-Scale Dynamics of the Chemical Bond. *The Journal of Physical Chemistry A* **104,** 5660–5694. ISSN: 1089-5639. https://doi.org/10.1021/jp001460h (June 2000).
4. Keilmann, F., Gohle, C. & Holzwarth, R. Time-domain mid-infrared frequency-comb spectrometer. *Opt. Lett.* **29,** 1542–1544. https://opg.optica.org/ol/abstract.cfm?URI=ol-29-13-1542 (July 2004).
5. Schiller, S. Spectrometry with frequency combs. *Opt. Lett.* **27,** 766–768. https://opg.optica.org/ol/abstract.cfm?URI=ol-27-9-766 (May 2002).
6. Picqué, N. & Hänsch, T. W. Frequency comb spectroscopy. *Nature Photonics* **13,** 146–157. ISSN: 1749-4893. https://doi.org/10.1038/s41566-018-0347-5 (Mar. 2019).
7. Lanin, A. A., Voronin, A. A., Fedotov, A. B. & Zheltikov, A. M. Time-domain spectroscopy in the mid-infrared. *Scientific Reports* **4,** 6670. ISSN: 2045-2322. https://doi.org/10.1038/srep06670 (Oct. 2014).
8. Wu, Q. & Zhang, X.-C. Free-space electro-optic sampling of terahertz beams. *Applied Physics Letters* **67,** 3523–3525. eprint: https://doi.org/10.1063/1.114909. https://doi.org/10.1063/1.114909 (1995).
9. Timmers, H. *et al.* Molecular fingerprinting with bright, broadband infrared frequency combs. *Optica* **5,** 727–732 (2018).
10. Laubereau, A. & Kaiser, W. Vibrational dynamics of liquids and solids investigated by picosecond light pulses. *Rev. Mod. Phys.* **50,** 607–665. https://link.aps.org/doi/10.1103/RevModPhys.50.607 (3 July 1978).
11. Movasaghi, Z., Rehman, S. & ur Rehman, D. I. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews* **43,** 134–179. eprint: https://doi.org/10.1080/05704920701829043. https://doi.org/10.1080/05704920701829043 (2008).
12. Pupeza, I. *et al.* Field-resolved infrared spectroscopy of biological systems. *Nature* **577,** 52–59. ISSN: 1476-4687. https://doi.org/10.1038/s41586-019-1850-7 (Jan. 2020).
13. Kowligy, A. S. *et al.* Infrared electric field sampled frequency comb spectroscopy. *Science Advances* **5,** eaaw8794. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.aaw8794. https://www.science.org/doi/abs/10.1126/sciadv.aaw8794 (2019).
14. Herbst, A. *et al.* Recent advances in petahertz electric field sampling. *Journal of Physics B: Atomic, Molecular and Optical Physics* **55,** 172001. https://dx.doi.org/10.1088/1361-6455/ac8032 (Aug. 2022).
15. Smallwood, C. L. & Cundiff, S. T. Multidimensional Coherent Spectroscopy of Semiconductors. *Laser & Photonics Reviews* **12,** 1800171. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/lpor.201800171. https://onlinelibrary.wiley.com/doi/abs/10.1002/lpor.201800171 (2018).
16. Cocker, T. L., Peller, D., Yu, P., Repp, J. & Huber, R. Tracking the ultrafast motion of a single molecule by femtosecond orbital imaging. *Nature* **539,** 263–267. ISSN: 1476-4687. https://doi.org/10.1038/nature19816 (Nov. 2016).

17. Vicentini, E., Wang, Z., Van Gasse, K., Hänsch, T. W. & Picqué, N. Dual-comb hyperspectral digital holography. *Nature Photonics* **15,** 890–894. ISSN: 1749-4893. https://doi.org/10.1038/s41566-021-00892-x (Dec. 2021).

18. Edgar, M. P., Gibson, G. M. & Padgett, M. J. Principles and prospects for single-pixel imaging. *Nature Photonics* **13,** 13–20. ISSN: 1749-4893. https://doi.org/10.1038/s41566-018-0300-7 (Jan. 2019).

19. Rani, M., Dhok, S. B. & Deshmukh, R. B. A Systematic Review of Compressive Sensing: Concepts, Implementations and Applications. *IEEE Access* **6,** 4875–4894 (2018).

20. Candes, E. J. & Wakin, M. B. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine* **25,** 21–30 (2008).

21. Candes, E., Romberg, J. & Tao, T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* **52,** 489–509 (2006).

22. Donoho, D. Compressed sensing. *IEEE Transactions on Information Theory* **52,** 1289–1306 (2006).

23. Candes, E. J. & Tao, T. Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory* **52,** 5406–5425 (2006).

24. Ostic, R. & Ménard, J.-M. Speeding Up Ultrafast Spectroscopy. *Physics* **14,** 23. https://www.physics.aps.org/articles/v14/23 (2021).

25. Takizawa, S., Hiramatsu, K. & Goda, K. Compressed time-domain coherent Raman spectroscopy with real-time random sampling. *Vibrational Spectroscopy* **107,** 103042. ISSN: 0924-2031. https://www.sciencedirect.com/science/article/pii/S0924203119303121 (2020).

26. Adhikari, S. *et al.* Accelerating Ultrafast Spectroscopy with Compressive Sensing. *Phys. Rev. Appl.* **15,** 024032. https://link.aps.org/doi/10.1103/PhysRevApplied.15.024032 (2 Feb. 2021).

27. Katz, O., Levitt, J. & Silberberg, Y. Compressive Fourier Transform Spectroscopy. *Optics InfoBase Conference Papers* (June 2010).

28. Kaestner, B. *et al.* Compressed sensing FTIR nano-spectroscopy and nano-imaging. *Optics Express* **26,** 18115 (July 2018).

29. Kawai, A., Kageyama, T., Horisaki, R. & Ideguchi, T. Compressive dual-comb spectroscopy. *Scientific Reports* **11,** 13494. ISSN: 2045-2322. https://doi.org/10.1038/s41598-021-93005-1 (June 2021).

30. Zhu, N. *et al.* Midinfrared Compressed Fourier-Transform Spectroscopy with an Optical Frequency Comb. *Phys. Rev. Appl.* **18,** 024025. https://link.aps.org/doi/10.1103/PhysRevApplied.18.024025 (2 Aug. 2022).

31. D. Kaplan & P. Tournois. Theory and performance of the acousto optic programmable dispersive filter used for femtosecond laser pulse shaping. *J. Phys. IV France* **12,** 69–75. https://doi.org/10.1051/jp4:20020098 (2002).

32. Schubert, O. *et al.* Rapid-scan acousto-optical delay line with 34 kHz scan rate and 15 as precision. *Optics letters* **38,** 2907–2910 (2013).

33. Van den Berg, E. & Friedlander, M. P. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing* **31,** 890–912. http://link.aip.org/link/?SCE/31/890 (2008).

34. Van den Berg, E. & Friedlander, M. P. *SPGL1: A solver for large-scale sparse reconstruction* https://friedlander.io/spgl1. Dec. 2019.

35. Arjoune, Y., Kaabouch, N., El Ghazi, H. & Tamtaoui, A. *Compressive sensing: Performance comparison of sparse recovery algorithms* in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)* (2017), 1–7.

36. Chen, S. S., Donoho, D. L. & Saunders, M. A. Atomic Decomposition by Basis Pursuit. *SIAM Review* **43,** 129–159. eprint: https://doi.org/10.1137/S003614450037906X. https://doi.org/10.1137/S003614450037906X (2001).

37. Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P. & Meester, L. E. *A Modern Introduction to Probability and Statistics: Understanding why and how* (Springer, 2005).

38. Cox, A. N. *Allen's astrophysical quantities* (Springer, 2015).

39. Van Exter, M., Fattinger, C. & Grischkowsky, D. Terahertz time-domain spectroscopy of water vapor. *Opt. Lett.* **14,** 1128–1130. https://opg.optica.org/ol/abstract.cfm?URI=ol-14-20-1128 (Oct. 1989).

40. Cui, H. *et al.* Vibration–rotation absorption spectrum of water vapor molecular in frequency selector at 0.5–2.5THz range. *Optik* **126,** 3533–3537. ISSN: 0030-4026. https://www.sciencedirect.com/science/article/pii/S003040261500827X (2015).

41. Sitnikov, D. S., Romashevskiy, S. A., Pronkin, A. A. & Ilina, I. V. Open-path gas detection using terahertz time-domain spectroscopy. *Journal of Physics: Conference Series* **1147,** 012061. https://dx.doi.org/10.1088/1742-6596/1147/1/012061 (Jan. 2019).

42. Liu, H.-B., Zhong, H., Karpowicz, N., Chen, Y. & Zhang, X.-C. Terahertz Spectroscopy and Imaging for Defense and Security Applications. *Proceedings of the IEEE* **95,** 1514–1527 (2007).

43. Weigel, A. *et al.* Ultra-rapid electro-optic sampling of octave-spanning mid-infrared waveforms. en. *Opt Express* **29,** 20747–20764 (June 2021).

44. Mohler, K. J. *et al.* Dual-comb coherent Raman spectroscopy with lasers of 1-GHz pulse repetition frequency. *Opt. Lett.* **42,** 318–321. https://opg.optica.org/ol/abstract.cfm?URI=ol-42-2-318 (Jan. 2017).

Supplementary Information

Kilian Scheffter[1,2], Jonathan Will[1,2], Claudius Riek[3], Herve Jousselin[4], Sébastien Coudreau[4], Nicolas Forget[4], Hanieh Fattahi[1,2]

1. Max Planck Institute for the Science of Light, Staudtstr. 2, 91058 Erlangen, Germany
2. Friedrich-Alexander University Erlangen-Nürnberg, Staudtrstr. 7, 91085 Erlangen, Germany
3. Zurich Instruments Germany, Mühldorfstraße 15, 81671 Munich, Germany
4. Fastlite, rue des Cistes 165, 06600 Antibes, France

1 Experimental setup

1.1 Lock-in detection

To minimize measurement noise, the experiment uses lock-in detection to measure the signal from the balanced photodetector. Ideally, the balanced photodetector output yields 0 V in the absence of the THz-pulse. However, drifts can occur due to alignment imperfections or changes caused by thermal effects or by scanning the delay with the Dazzler or mechanical delay stage. To remove these drifts, the THz-pulses are chopped at half the laser repetition rate. Consequently, the balanced photodetector measures periodically one probe pulse with the THz-field and one without it. Since drifts are present in both pulses, it results in a 1 kHz signal. However, the information about the THz-field strength is present in only one probe pulse yielding a signal at 500 Hz. By demodulating the signal of the balanced photodetector with a reference signal at 500 Hz in a lock-in amplifier, the resulting signal is free from systematic noise and possesses only information about the THz-field.

1.2 Synchronization

Various trigger signals are present in the experimental setup to synchronize the different devices and to ensure accurate data acquisition. Fig. 1 provides a layout of the different trigger signals. The front-end generates a trigger signal at a frequency of 1 kHz (red) corresponding to the outgoing optical pulse train. This signal triggers both the chopper and the Dazzler. To block every other optical pulse, the chopper uses the first subharmonic of this trigger signal with a frequency of 500 Hz. This chopping frequency is further used as the demodulation frequency of the lock-in amplifier (green). The Dazzler utilizes the trigger of the Legend to trigger an RF-pulse that drives an acoustic wave in the crystal-unit. A control signal is then sent to the electronic control-board via the "GATE"-output S1 of the Dazzler (blue, at 1 kHz) for every launched RF-pulse, even if the Dazzler is running but not controlled by the computer.

When a measurement is started by the computer, the information about the acoustic pulses launched within the experiment is send via a USB connection to the Dazzler. The "PATTERN-GATE" output of the Dazzler becomes high before the first acoustic wave of the sequence is launched, indicating that the Dazzler is running a user-defined sequence controlled by the computer. This output signal remains high until the sequence is over. To ensure that the lock-in signal is sampled only when the user-defined sequence is played and when a pulse passes through the chopper, an electronic control board is used. The board consists of an AND gate and a D-flip-flop. The AND gate combines the "GATE" and "PATTERN-GATE" outputs of the Dazzler resulting in a 1 kHz trigger signal whenever a user-defined sequence is played from the computer. If no user-defined sequence is playing, the AND gate outputs a low signal. In the presence of a sequence, the frequency of the 1 kHz trigger signal after the AND gate is divided by two using a D-flip-flop to sample only the lock-in amplifier signal of pulses which are passing through the chopper. The signal leaving the electronic board is triggering the data acquisition of the lock-in amplifier.

Due to the lack of synchronization between the chopper and the Dazzler in the setup described above, the Dazzler is unable to determine if a pulse is blocked or transmitted by the chopper during the start of a sequence. As a result, in every second measurement in average, the electronic control board produces an incorrect trigger signal sampling the lock-in amplifier when a pulse is blocked by the chopper. In this case, the measurement must be discarded.

1.3 Rapid field measurements

To minimize measurement time, the Dazzler rapidly scans in a shot-to-shot manner, producing the same time delay for a pair of two consecutive probe pulses. Following one pair of probe pulses, a next pair of probe pulses experiences a new delay. As the beam generating the THz-pulse is chopped at half the repetition frequency, each pair of probe pulses coincides with one generated THz-pulse and one blocked pulse. Therefore, for every time-delay

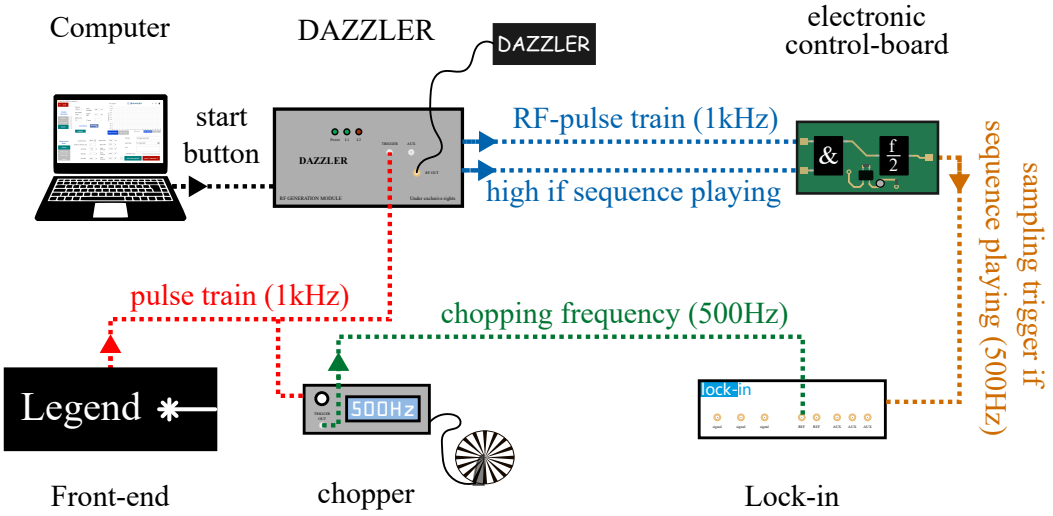


Figure 1: Path of trigger signals required for synchronization and data acquisition

within a scan, one probe pulse probes the THz-pulse while the other one serves as a reference. By measuring the probe pulses via a lock-in amplifier, systematic noise present in both probe pulses of a pair is eliminated, leaving only the THz-signal.

For large step sizes or random scanning, the measured signal amplitude can jump from zero to the peak value of the THz-pulse. Hence, the probe pulses have to resolve this and it must be captured with the detection electronics. The UHFLI lock-in amplifier from Zurich Instruments can follow such a jump when using the boxcar averager functionality while reducing noise due to the signals low duty-cycle. Here, the temporal moving average (boxcar) ensures a pure signal of the last data point without any residual contribution by prior data due to the exponential filter function used in traditional lock-in amplifiers.

2 Reconstruction strategy

2.1 Mathematical Description of Compressed Sensing

Compressed Sensing is a technique where an analog signal $s(t)$ is sampled with a low number of M sampling points resulting in the measured $M \times 1$ discrete signal $y[t]$. As it is not feasible to mathematically describe sampling an analog signal $s(t)$, an oversampled $N \times 1$ digital signal $s[t]$ with $N >> M$ sampling points represents the original signal. Sampling this signal can then be described with the $M \times N$ sensing matrix Φ

$$y[t] = \Phi s[t]. \tag{1}$$

As an example, the sensing matrix Φ can be the spike basis possessing the columns $\phi_i[t] = \delta[t-i]$ in the sensing matrix [1].

Compressed Sensing requires the signal $s[t]$ to be sparse in some orthonormal basis. The original signal $s[t]$ can be expressed in this basis with a $N \times N$ basis matrix Ψ and a $N \times 1$ amplitude vector $\vec{c}$

$$s[t] = \Psi \vec{c}. \tag{2}$$

One example for an orthonormal basis is the Fourier-basis with the discrete-Fourier-transform (DFT) matrix as the basis matrix Ψ.

The original signal can be regained by solving the equation

$$y[t] = \Phi \Psi \vec{c} \tag{3}$$

for the vector $\vec{c}$. Since this system is undetermined, the unique solution can be found by solving for the minimum l_0 norm of $\vec{c}$. However, to obtain the minimum l_0 norm, all possible solutions need to be tried out making the problem not feasible to compute.

Hence, in Compressed Sensing one usually solves the optimization problem [1, 2]

$$\min ||\vec{c}||_1 \text{ subject to } y[t] = \Phi\Psi\vec{c}, \tag{4}$$

where $||\vec{c}||_1 = \sum_i |c_i|$ is the l_1 norm. The solution of equation 3 has the minimum l_1 norm of $\vec{c}$.

The necessary condition to regain $s[t]$ from the measured signal $y[t]$ is that the reconstruction matrix $\Theta = \Phi\Psi$ needs to obey the restricted isometry property (RIP). As this condition is hard to check, the incoherence between the matrices Φ and Ψ can be used as an indication for a stable and faithful reconstruction. The coherence between the two matrices is defined as [1, 2]

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{i \leq M, j \leq N} |\langle \phi_i, \psi_j \rangle|, \tag{5}$$

measuring the largest correlation between any of the elements in the two matrices.

The coherence values that can yield values in the range from 1 to $\sqrt{N}$, with 1 being the value of maximum incoherence. A higher degree of incoherence, which translates to lower coherence values, increases the likelihood of accurately reconstructing the signal $s[t]$. In certain instances, the incoherence value is calculable. For instance, the spike basis possesses a maximum incoherence of $\mu = 1$ with the Fourier-basis, while the noiselet basis has a coherence value of $\mu = \sqrt{2}$ with the wavelet basis. For a fixed basis with matrix Ψ of unknown coherence to any sensing matrix Φ, the sensing matrix can be chosen to be a random, orthonormal matrix. This results in a high probability of having a low coherence value with the fixed matrix Ψ [1, 2].

2.2 Convex Compressed Sensing Reconstruction

Equation 4 can be understood as a convex optimization problem which can be solved by linear programming algorithms. Since the problem as stated in equation 4 is not considering noise, the so called basis pursuit algorithm can only reconstruct a stable, faithful solution for noise-free measurements. For noisy measurements, it is not optimal to solve equation 3 exactly. Therefore, a relaxed problem needs to be formulated [2]. Two convex approaches capable of dealing with noisy measurements are investigated for compressed sensing. The first algorithm is called basis pursuit denoising optimizing the problem [3]

$$\min ||\vec{c}||_1 \text{ subject to } \frac{1}{2} ||\Phi\Psi\vec{c} - y[t]||_2^2 \leq \sigma, \tag{6}$$

where σ is a parameter chosen according to the magnitude of noise in the measurement. The second algorithm is solving the so called Lasso problem given as [3]

$$\min \frac{1}{2} ||\Phi\Psi\vec{c} - y[t]||_2^2 \text{ subject to } ||\vec{c}||_1 \leq \tau, \tag{7}$$

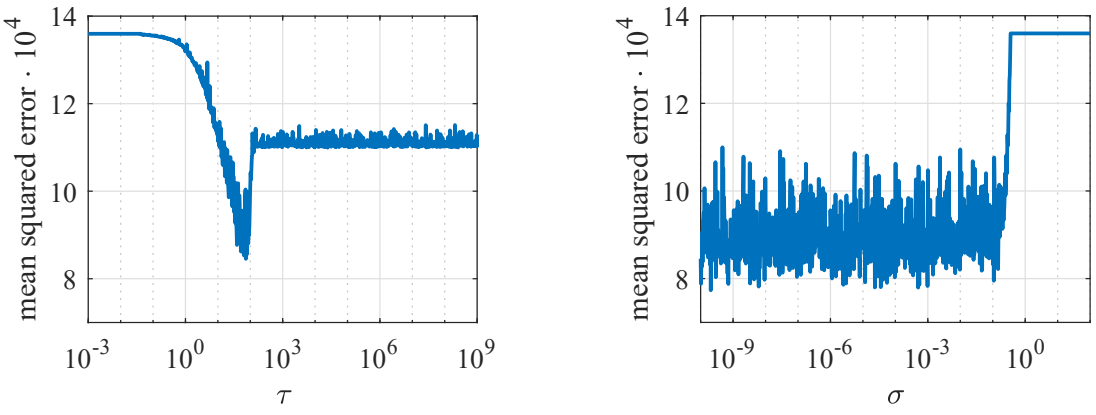
where the parameter τ is the regularization parameter of the problem.

Open source Lasso and basis-pursuit denoise algorithms are investigated [3, 4]. For the Lasso as well as the basis-pursuit problem, the value of τ, respectively σ needs to be determined empirically. For that, one randomly sampled trace with $N = 88$ sampling points of the UHLFI measurement set is used for the determination of τ and σ. This explicit measurement is taken, since the number of sampling points correspond to the measurement with the linear sampling scheme with $\Delta t = 341\,\text{fs}$. This step size yields a Nyquist-frequency of $f_{\text{Nyquist}} \approx 1.5\,\text{THz}$ positioned in the middle of the absorption spectrum of water vapor. Therefore, this measurement correlates to the border in which compressed sensing could outperform the Fourier-transformation of a traditional linear sampling scheme.

The data is fed to the Lasso as well as the basis-pursuit denoise algorithm for different values of τ and σ, respectively [5]. For the respective reconstructed traces, the mean squared error to a oversampled reference trace is calculated. The results are depicted in Fig. 2. Both algorithms exhibit a similar minimal value of the mean squared error. However, the Lasso algorithm possesses a distinct global minimum at $\tau \approx 70$, whereas the basis-pursuit denoise algorithm shows a constant minimum value for $\sigma < 10^{-1}$. Since it is not always possible to find a reference for arbitrary measurements to optimize the value of τ, the basis pursuit algorithm fits better as a compressed sensing algorithm for accelerating time domain spectroscopy. Consequently, solely the basis pursuit algorithm with a fixed value of $\sigma = 10^{-6}$ is investigated.

2.3 StOMP

A reconstruction program consisting of the three algorithms StOMP, TRF and iFFT was developed for evaluating the measured data [The code will be publicly available upon acceptance for publication]. The code aims for



(a) Lasso

(b) Basis-pursuit denoise

Figure 2: Mean squared error of reconstructed traces from a randomly sampled measurement with $N = 88$ sampling points with respect to an oversampled reference trace. The figure compares the mean squared error for the reconstruction performed by the Lasso (a) and the basis-pursuit denoise (b) algorithm for different values of τ, respectively σ.

fast data processing, minimum human intervention and control characteristics of result. Stagewise Orthogonal Matching Pursuit is a greedy algorithm developed in 2012 which is popular due to the low level of computational complexity resulting in fast data processing [6].

The developed implementation of the iterative algorithm StOMP is shown in Algorithm 1. In each iteration s, the different entries of the sparse domain $\vec{c}$ for a residual signal r_s are weighted via the matched filter $\kappa_s = \Theta^{\mathrm{T}} r_s$ (In case of the first iteration: $r_0 = y$). The result κ_s scales the presence of the different entries of $\vec{c}$ in the signal r_s. The most prominent entries are selected via a threshold τ_s yielding a set of outliers $J_s = \{j : |\kappa_s(j)| > \tau_s\}$. The threshold $\tau_s = t_s \cdot \sigma_s$ is constructed by a threshold parameter t_s and the formal noise level of the residual r_s given by

$$\sigma_s = \frac{||r_s||_2}{\sqrt{\text{dimension}(r_s)}}. \tag{8}$$

The entries in this set are merged with all the previous found outliers given in the set I_{s-1} updating the total set of detected outliers $I_s = I_{s-1} \cup J_s$. The projection of the measured data on the detected outliers I_s yields the result for this iteration

$$\vec{c}_s = \left(\Theta_{I_s}^{\mathrm{T}} \Theta_{I_s}\right)^{-1} \Theta_{I_s}^{\mathrm{T}} y. \tag{9}$$

Here, Θ_{I_s} indicates the $M \times |I_s|$ matrix which is constructed of Θ by choosing only the columns according to the index set I_s. With that, the result $\vec{c}_s$ possesses nonzero values for the dominant entries in I_s weighted by the projection of the measured signal y. If a certain exit condition is fulfilled, c_s yields the final result of the algorithm. If the exit condition is not yet fulfilled, the residual for the next iteration is updated by $r_{s+1} = y - \Theta\vec{c}_s$. Consequently, the residual of the next iteration r_{s+1} contains the remaining signal contributions which have not been yet detected.

The value of the threshold parameter t_s to achieve an optimal reconstruction depends on the input signal y and is varying significantly between different measurements. Hence, we developed an automatic thresholding approach based on the Inter-Quartile Range (IQR) method for outlier detection [7]. The IQR indicates the midspread of a dataset and is defined as the difference of the third and first quartile $IQR = Q_3 - Q_1$. The first and third quartile are hereby defined as the mean of the lower and upper half of a statistical distribution, respectively. In the developed algorithm, t_s is calculated by the formula

$$t_s = \frac{Q_3 + \gamma \cdot IQR}{\sigma_s}, \tag{10}$$

where γ is a set filter value dictating the sensitivity to reconstruct less dominant entries of $\vec{c}$ in the measured signal.

The discussed reconstructions in this work were computed with a filter value of $\gamma = 1.8$. The developed algorithm applies the threshold only to entries $\vec{c}_s$ within in a set of given boundary conditions. In the case of the

presented measurements, only entries corresponding to frequencies within the spectrum of the excitation pulse $f \leq 2.5\,\mathrm{THz}$ are considered. To improve the robustness of the algorithm, a range of sparsity for the reconstructed signal is defined. In case the set of detected outliers I_s contains a number of entries below the minimum allowed sparsity, the algorithm fills this set by the most prominent entries of $\vec{c_s}$ according to the matched filter until the set possesses the minimum allowed sparsity. In case the set I_s is above the maximum allowed sparsity, the algorithm stops and returns the solution of the current iteration c_s.

As StOMP is indifferent towards amplitude optimization, a Nonlinear Least Squares (NLS) solver was included in the reconstruction algorithm. The "Nonlinear Least Squares" solver is a wrapper for scipy.optimize.curve_fit which is a wrapper for scipy.optimize.least_squares, the used algorithm is "Trust Region Reflective". It has been chosen because it is well suited for large sparse optimization problems. We choose to use scipy.optimize.curve_fit instead of scipy.optimize.least_squares because its user-friendliness compared to other optimization methods in python. Due to the fact that NLS only optimizes the result of StOMP and does not "'add' any frequency to the optimized result, only a few iterations are needed. NLS tries to minimize the error between the reconstruction and the measurement in time domain. Therefore, it is capable of changing/readjusting all nonzero values of the reconstruction in frequency domain.

Algorithm 1 Our Implementation of StOMP ("IQR"-Mode)

Require: compressed measurement: y, reconstruction matrix: Θ, maximum stages: max_{stages}, filter value: γ, boundary condition: b, minimum sparsity ratio: $min_{sparsity}$, maximum sparsity ratio: $max_{sparsity}$

for $s = 1, s \leq max_{stages} : i+ = 1$ **do**
- $\kappa_s = \Theta^T r_s$ ▷ Matched Filter
- **if** $s == 1$ or $J_{s-1} == \{\}$ **then**
  - $t_s = \frac{\text{TRESHOLD}(\kappa_s, b, \gamma)}{\sigma_s}$ for $\kappa_s \in \{b\}$
- **else**
  - $t_s = t_{s-1}$
- **end if**
- $J_s = \{j : |\kappa_s(j)| > t_s \sigma_s\}$ ▷ Hard Thresholding
- **if** SPARSITY RATIO(κ_s, I_{s-1}, J_s) $<$ $min_{sparsity}$ **then**
  - **while** SPARSITY RATIO(κ_s, I_{s-1}, J_s) $<$ $min_{sparsity}$ **do** ▷ If sparsity to low increase it until it is ok
    - $J_s \cup \{j : |\kappa_s(j)| \geq |\kappa_s(n)| \forall n\}$
  - **end while**
- **end if**
- $I_s = I_{s-1} \cup J_s$ ▷ Updating Index Set
- $(c_s)_{I_s} = (\Theta_{I_s}^T \Theta_{I_s})^{-1} \Theta_{I_s}^T y$ ▷ Projection
- $r_{s+1} = y - \Theta_{I_s} c_s$ ▷ Updating residual
- **if** SPARSITY RATIO(κ_s, I_{s-1}, J_s) $\geq$ $max_{sparsity}$ **then** ▷ If sparsity ok or to hight stop
  - **end for**
- **end if**

end for

return c_s

Algorithm 2 sparsity ratio

function SPARSITY RATIO(κ_s, I_{s-1}, J_s)
- **return** $\frac{|I_{s-1}|+|J_s|}{dim(\kappa_s)}$

end function

3 Reconstruction details

3.1 Basis and measurement matrix

Compressed sensing algorithms require a measurement matrix Φ as well as a basis matrix Ψ. In this work, the discrete-Fourier-transformation (DFT) is used as the basis matrix Ψ. Hence, the entries $\vec{c}$ correspond to the frequency spectrum of the measured signal.

The size of the basis and the corresponding matrix Ψ is determined by the length of the time vector for the reconstructed signal. The time vector and the correlating measurement matrix can be chosen arbitrarily. In the

Algorithm 3 Threshold t_{IQR} Calculation

function THRESHOLD(κ_s, b, γ)
  $C = \{|\ \kappa_s(j)\ |: j \in b\}$ ▷ absolut values with respect to index boundaries
  $Q_1 = P_{25}^{C}$ ▷ Percentil 0 to 25%
  $Q_3 = P_{75}^{C}$ ▷ Percentil 75 to 100%
  $IQR = Q_3 - Q_1$ ▷ Interquartil range calculation
  $t_{\mathrm{IQR}} = Q_3 + \gamma \cdot IQR$ ▷ threshold calculation
  return t_{IQR}
end function

present work, the minimum and maximum of the time vector is bound to the temporal window of the molecular response. This allows to compare the Fourier-transformation of the reconstruction to the one of the linear sampled data laying within this window with equal spectral resolution. The step size of the time vector for the reconstruction is upper bounded by the Nyquist criteria of the signal. By decreasing the step size, the quality of the reconstruction improves since differences between time points of the reconstruction and the random time points of the measurement are decreasing. However, decreasing the step size is increasing the size of the matrices and hence the computation time. In this work, we use a moderate step size of $\Delta t = 75\,\mathrm{fs}$.

In the case of the basis pursuit denoise algorithm, a boundary matrix Σ able to set boundary conditions is added to the basis matrix $\Psi_{\mathrm{BPD}} = \Psi\Sigma$. Σ allows to define a frequency range for non-zero values while keeping the DFT matrix large. We selected a frequency range from 0 up to 2.5 THz corresponding to the spectrum of the excitation pulse. A large DFT matrix is favorable in BPD, since this corresponds to a fine time step of the reconstructed time signal. For finer time steps, the error calculation $||\Phi\Psi\vec{c} - y[t]||$ is more accurate, as the time points of the reconstructions are closer to the random time points of the measured signal. The boundary matrix transports the vector $\vec{c}$ with entries laying in the defined spectral region into the columns of a large fft matrix corresponding to the respective frequencies of $\vec{c}$.

3.2 BPD reconstruction of measurements with different lock-in amplifier

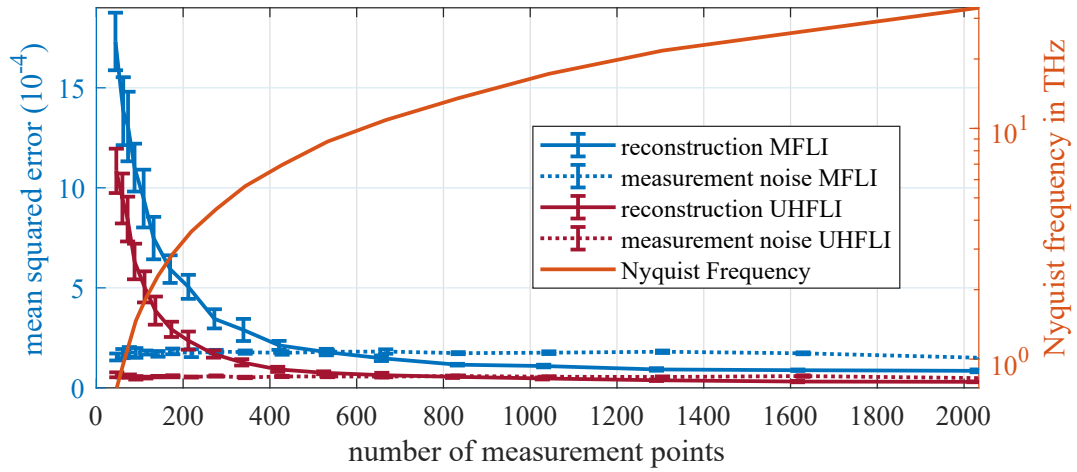


Figure 3: Mean and indicated standard deviation of the mean squared error of reconstructed traces for three different sets of measurements. The mean squared error is calculated with respect to an average of ten oversampled reference traces of the respective measurement set. In the figure, the reference shows the mean squared error of the linear sampled data with respective number of sampling points to the same reference trace.

The quality of reconstructed traces via BPD for two sets of measurements performed with two different lock-in amplifiers (UHLFI and MFLI from Zurich Instruments) are compared. The lock-in amplifier UHFLI is using the boxcar averager option to reduce noise due to the low duty cycle of the signal of the balanced photo diodes. The lock-in amplifier MFLI acquires data without a boxcar averager option. To speed up settling of the MFLI's low-pass filter to avoid unwanted contribution due to the exponential filter, the time constant is lowered. To attenuate signal contributions of DC offsets, non-linearities and signal components situated at twice the demodulation

frequency, a sinc-filter is applied. The combination of the low time constant with the sinc-filter results in higher measurement noise compared to the detection with the boxcar filter. Consequently, we are able to compare the quality of reconstruction of measurements with two different noise levels. For a quantitative comparison, the mean squared error of the reconstructed traces for the random sampled measurements is calculated with respect to the average of the ten linear sampled traces with the lowest step size for the respective set of measurements. The mean as well as the standard deviation for the mean squared error of the ten repetitions for different number of sampling points are shown in Fig. 3. As a measure for the measurement noise, the mean squared error of the linear sampled measurements is depicted for the different number of sampling points. The MFLI measurement exhibits measurement noise which is by a factor of three higher compared to the UHFLI measurement. For measurements possessing a large number of sampling points, the BPD reconstruction is below the noise level for both set of measurements. However, the reconstructions of the MFLI measurement show a bigger noise rejection with respect to the noise level compared to the UHFLI measurement. This shows the potential of compressed sensing in improving the signal quality of noisy measurements.

3.3 BPD reconstruction quality

The reconstruction quality of linearly sampled and randomly sampled measurement are compared in Fig. 4 and 5. The probability to reconstruct absorption peaks via BPD for a randomly sampled waveform with $N = 46$ is shown in Fig 6.

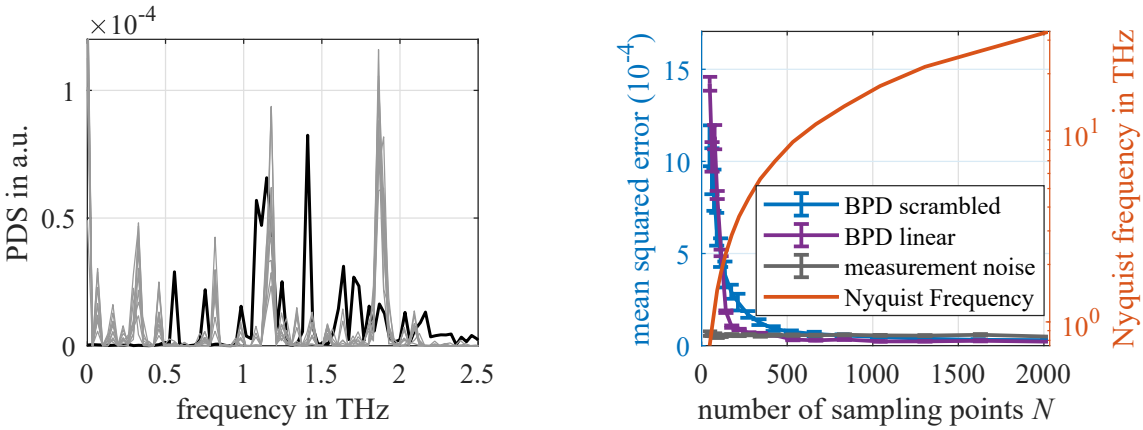


Figure 4: a) Power density spectrum (PDS) of the reconstruction of ten linear sampled measurements with $N = 47$ (gray) and the PDS of an oversampled reference trace (black). b) Mean squared error of the BPD reconstruction of linear (purple) and scrambled (blue) measurements with respect to an oversampled reference for different number of sampling points. The measurement noise (grey) is given by the mean squared error of the raw linear sampled measurements with respect to the oversampled reference. The Nyquist frequency for the corresponding number of sampling points is given in orange.

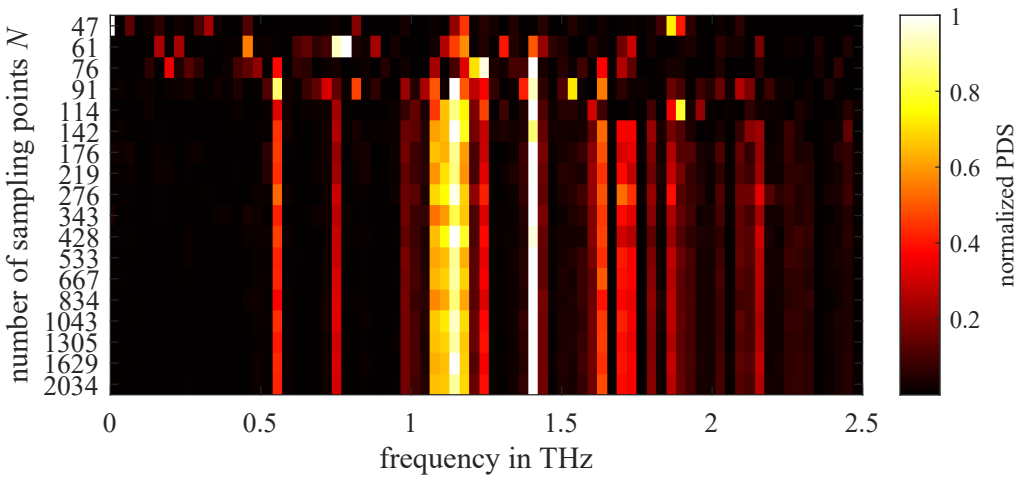
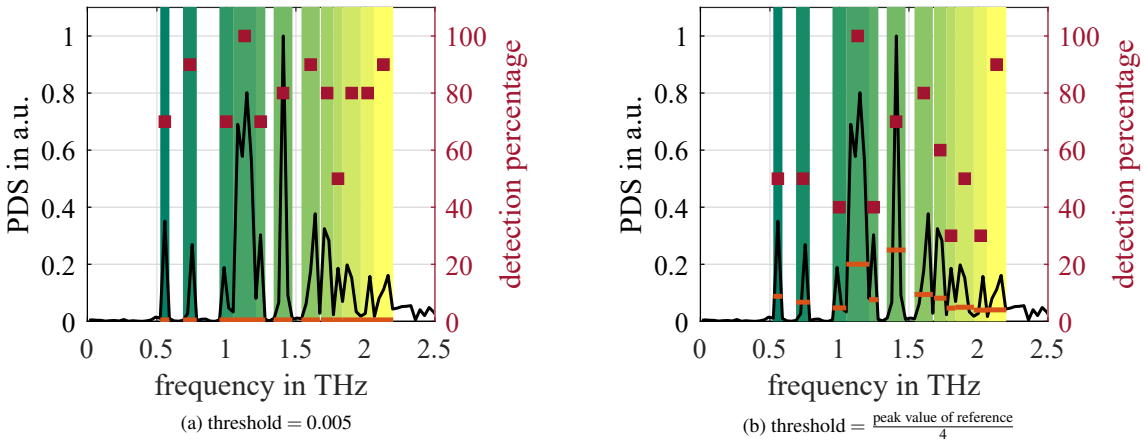


Figure 5: Averaged Power density spectra (PDS) for the BPD reconstruction of ten linear sampled measurements for different number of sampling points.



(a) threshold = 0.005

(b) threshold = $\frac{\text{peak value of reference}}{4}$

Figure 6: Probability to reconstruct absorption peaks via BPD for a randomly sampled waveform with $N = 46$ sampling points. An absorption peak is declared as detected if the reconstruction possesses a value greater than the given threshold (orange) in the area of the respective absorption. The threshold is given as the noise level of the reference (black) threshold = 0.005 in a) and as a quarter of the respective absorption peak value of the reference in b).

References

1. Candes, E. J. & Wakin, M. B. An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine* **25,** 21–30 (2008).
2. Rani, M., Dhok, S. B. & Deshmukh, R. B. A Systematic Review of Compressive Sensing: Concepts, Implementations and Applications. *IEEE Access* **6,** 4875–4894 (2018).
3. Van den Berg, E. & Friedlander, M. P. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing* **31,** 890–912. `http://link.aip.org/link/?SCE/31/890` (2008).
4. Van den Berg, E. & Friedlander, M. P. *SPGL1: A solver for large-scale sparse reconstruction* https://friedlander.io/spgl1. Dec. 2019.
5. Scheffter, K. *Accelerating Terahertz Field-Resolved Spectroscopy* MA thesis (Friedrich-Alexander-University Erlangen-Nürnberg, Max-Planck-Institute for the Science of Light, Germany, 2022).

6. Donoho, D. L., Tsaig, Y., Drori, I. & Starck, J.-L. Sparse Solution of Underdetermined Systems of Linear Equations by Stagewise Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory* **58,** 1094–1121 (2012).

7. Vinutha, H. P., Poornima, B. & Sagar, B. M. *Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset* in *Information and Decision Sciences* (eds Satapathy, S. C., Tavares, J. M. R., Bhateja, V. & Mohanty, J. R.) (Springer Singapore, Singapore, 2018), 511–518. ISBN: 978-981-10-7563-6.