



OPEN

Compressive dual-comb spectroscopy

Akira Kawai¹, Takahiro Kageyama¹, Ryoichi Horisaki^{2,3} & Takuro Ideguchi^{1,3,4}✉

Broadband, high resolution, and rapid measurements of dual-comb spectroscopy (DCS) generate a large amount of data stream. We numerically demonstrate significant data compression of DCS spectra by using a compressive sensing technique. Our numerical simulation shows a compression rate of more than 100 with a 3% error in mole fraction estimation of mid-infrared (MIR) DCS of two molecular species in a broadband (~30 THz) and high resolution (~115 MHz) condition. We also numerically demonstrate a massively parallel MIR DCS spectrum of 10 different molecular species can be reconstructed with a compression rate of 10.5 with a transmittance error of 0.003 from the original spectrum.

In the last decade, intensive attention has been cast on dual-comb spectroscopy (DCS), which allows one to measure broadband and high-resolution spectra with superior frequency accuracy at a high data acquisition rate^{1–3}. DCS provides outstanding spectroscopic features, especially for multiplex gas-phase molecular sensing with its high spectral resolution of ~100 MHz spanning over ~10 s THz, enabling various applications such as precision metrology⁴, greenhouse gas sensing^{5,6}, combustion diagnosis⁷, etc. The broadband and high-resolution spectroscopy can generate a large data set, e.g., 1,000,000 spectral points for a single spectrum⁸. Now, if we imagine the DCS techniques are to be used for hyperspectral imaging of 1000 × 1000 pixels measured with a 16-bit analog-to-digital converter, it generates ~4 TB per single hyperspectral image. Taking such images would cause severe problems in data transportation and storage.

Compressive sensing (CS) is a signal processing technique that allows, by making use of sparsity of a signal, reconstruction of the signal from a significantly reduced number of data points than the full set of data points required from the Nyquist-Shannon sampling theorem⁹. If the signal is sparse on a certain basis, the sparsest solution can be found by algorithms with a sparsity constraint or regularization. Mathematical studies proved that, in some appropriate conditions, CS could reconstruct an exact signal even in the presence of measurement noise¹⁰. A variety of studies on CS have been reported especially in the field of optical imaging, where natural scenes such as landscapes or biological cells are well reconstructed from images with fewer pixels^{10–12}. Contrary to imaging, CS-based spectroscopy^{13–18}, especially for gas-phase molecular sensing, has not actively been investigated, although high-resolution broadband spectra of gaseous molecules are good candidates of CS because of their sparse nature originating from the narrow molecular lines spread in a broad spectral range.

In this study, we numerically demonstrate compressive dual-comb spectroscopy (C-DCS), in which a simple CS technique effectively compresses the data size of DCS. In our numerical demonstration, we show that quantitative estimation of mole fraction of molecules can be made with an error of 3% even when <1% of interferometric data points are used only. Also, we show a well-reconstructed complex spectrum of a mixture of 10 molecular species in the condition of using 10% of original data points.

Results

Concept of compressive dual-comb spectroscopy (C-DCS). The basic concept of C-DCS is described in Fig. 1. DCS is a type of Fourier-transform spectroscopy with two mutually coherent frequency combs that run at slightly detuned repetition rates^{1–3}. The spatially combined pulse trains are interferometrically detected with a single photodetector, and digitization of the signal generates an interferogram. Finally, Fourier-transforming of the interferogram shows a spectrum. Our simulation is conducted under the assumption that we have a complete original data set of a DCS interferogram and randomly resample the data points at a lower rate (reducing the data points) to reconstruct a CS spectrum. Note that the random sampling guarantees decorrelation of the measurement data and the sparse basis (Fourier basis, in this case), which is a requirement of efficient CS⁹. It also ensures the generality of CS that works for any spectral shapes (e.g., condensed or spread absorption

¹Department of Physics, The University of Tokyo, Tokyo, Japan. ²Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan. ³PRESTO, Japan Science and Technology Agency, Saitama, Japan. ⁴Institute for Photon Science and Technology, The University of Tokyo, Tokyo, Japan. ✉email: ideguchi@ipst.s.u-tokyo.ac.jp

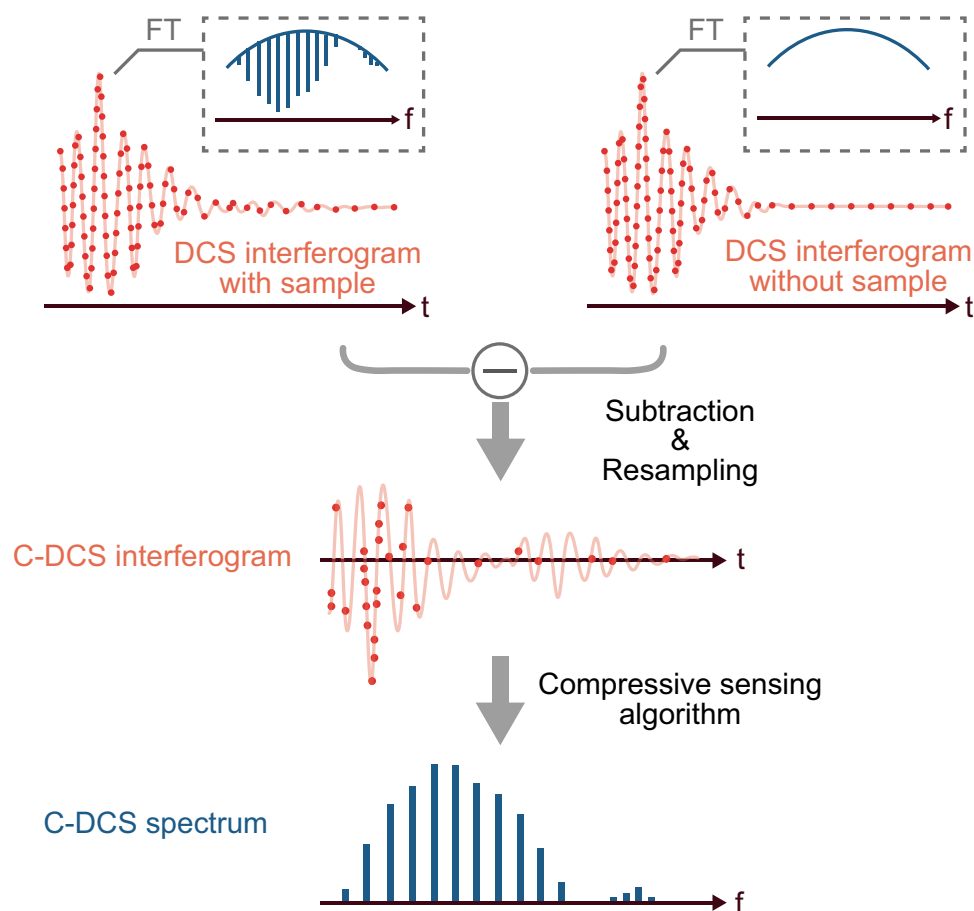


Figure 1. A conceptual illustration of compressive dual-comb spectroscopy (C-DCS).

lines) when the sparsity is at the same level. Suppose that the number of data points of the original interferogram is N and that of the resampled data points is M ($M < N$), the resampled data set can be represented as $\Omega = \{\omega_j\}_{j=1}^M \subset \{1, 2, \dots, N\}$. Note that we can arbitrarily set a probability mass function (PMF) for the random resampling. From the resampled data set, we can estimate a spectrum vector $x \in \mathbb{C}^N$ as an answer to the l_1 minimization problem⁹ described as Eq. (1),

$$\min \|\tilde{x}\|_1 \text{ subject to } \|A\tilde{x} - y\|_2 < \epsilon \quad (1)$$

where $A \in \mathbb{C}^{M \times N}$ ($M < N$) is a sensing matrix, $y \in \mathbb{R}^M$ a measurement vector, and $\epsilon \in \mathbb{R}$ a constraint value determined by noise of the system. In our case, $A = R\Phi$, where $R \in \{0, 1\}^{M \times N}$ is a subsampling operator which obeys $(Rx)_j = x_{\omega_j}$, and Φ a $N \times N$ discrete Fourier-transform operator. Considering a case of DCS operated with relatively low-chirped combs, signal intensities of the sampled points around the zero delay between the pulses have a larger magnitude (called “center-burst”) than the other points, including the signals showing molecular induction decays¹⁹. Therefore, it is effective to select a sloped PMF that samples more points around the center burst.

To fully utilize the sparse nature of the absorption spectrum and efficiently compress the data points, we suggest operating background subtraction of the interferogram. It can be implemented either by hardware instrumentation or post numerical processing. For the hardware instrumentation, a Michelson-type interferometer is added to make the π phase difference between the pulses from the two arms due to the reflection of the beam splitter, realizing the background subtraction due to the destructive interference on the detector²⁰. On the other hand, for the post numerical processing of background subtraction, a reference background interferogram can be obtained either by an additional measurement²¹ or a numerical baseline reconstruction⁷. Although in this proof-of-concept demonstration we operate the background subtraction for better reconstruction, we expect improved algorithms would make it possible to reconstruct spectra with a background.

Numerical condition of C-DCS simulation. To show the above-mentioned C-DCS concept, we demonstrate numerical simulations of trace-gas DCS in the MIR region. We simulate a mimic condition of a previously reported experiment²¹, where a broadband spectrum covering from 2006.7 to 3013.4 cm^{-1} (60.159–90.339 THz) is measured at a resolution of 0.0038 cm^{-1} (115 MHz) that consists of 262,144 spectral points Fourier-transformed by temporal data points of 524,286. We assume a broadband Gaussian-profile spectrum as comb

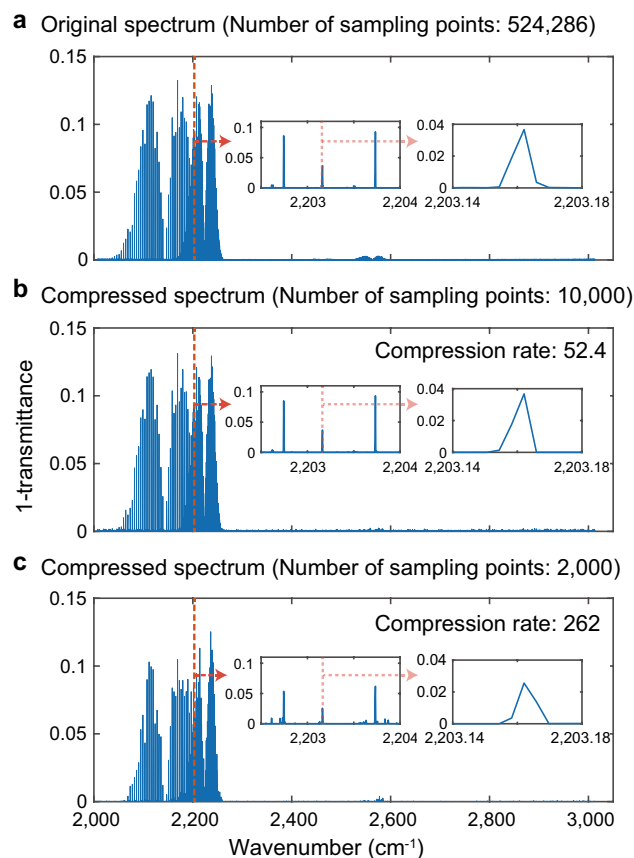


Figure 2. Simulated MIR DCS spectra of gaseous molecules. (a) An original spectrum Fourier-transformed by temporal data points of 524,286. (b, c) Compressed spectra with 10,000 and 2000 sampling points, respectively. The inset panels show zoom-in spectra around 2203 cm⁻¹.

sources. We first simulate an interferogram from the source spectrum with molecular absorptions with Doppler line profiles and create a background-free interferogram by baseline subtraction with a reference spectrum with no absorption lines. We numerically calculate the spectrum by referencing the HITRAN database and using its application programming interface HAPI²². Then, the interferogram is resampled with a sloped PMF, $C \min\{1, 1/|1 - N/2|\}$, where l is an index of sampling points of the interferogram in chronological order and C a normalization constant, which is found in the literature²³. For spectrum reconstruction, we use SPGL1 as an l_1 -minimization problem solver²⁴. We set an arbitrary signal-to-noise ratio (SNR) by assuming coherently averaged interferograms, which can be experimentally implemented with a variety of techniques^{8,25,26}. We note that the CS reconstruction algorithm, in general, works without having pre-knowledge of the molecular species.

Mole fraction estimation of two molecular species. To show how the data compression rate, which is defined as N/M , affects the quantitative capability of C-DCS, we demonstrate mole fraction estimation of two molecular species of trace gases. We numerically prepare mixed gases of N₂O (42 ppm) and CO (120 ppm) with a buffer gas at a pressure of 3 mbar filled in a 10-m-long multi-pass cell. We add a Gaussian measurement noise \mathbf{n} to the interferogram to make the estimation with different SNR conditions. We first set the SNR of the real part of FFT to be 1000. A constraint term ϵ in Eq. (1) is empirically set to an average of $\|\mathbf{n}/10\|$. The original spectrum converted from the full data points of 524,286, and compressive spectra with 10,000 and 2000 sampling points are shown in Fig. 2, where we show transmittance spectra of the sample. The compressive spectrum with 10,000 sampling points, which compression rate is 52.4, shows good agreement with the original one, while that with 2000 (compression rate of 262) shows clear distortions. We evaluate the mole fraction of N₂O molecules by spectrally fitting each absorption line and obeying Lambert–Beer law. The fitting is operated by a fixed-profile Gaussian function with a single free parameter of mole fraction. Here, the spectral points that satisfy $\log(1 - T) > 0.01$ (T : transmittance) are used for the evaluation. Figure 3a shows the ratio of the evaluated mole fraction of the compressed and original spectra for different compression rates under different SNR (1000, 500, 100) conditions. The result with SNR of 1000 shows that the compression rate of 105, which corresponds to the number of sampling points of 5000, leads to a 3% of deviation of mole fraction from that evaluated with the original spectrum. We can also see that lower measurement SNR degrades the evaluation results. Figure 3b shows root-mean-squared error (RMSE) of the spectral points that satisfy $\log(1 - T) > 0.01$ for each compression rate. We find that RMSE is proportional to (compression rate)^{0.93} by least-squares fitting.

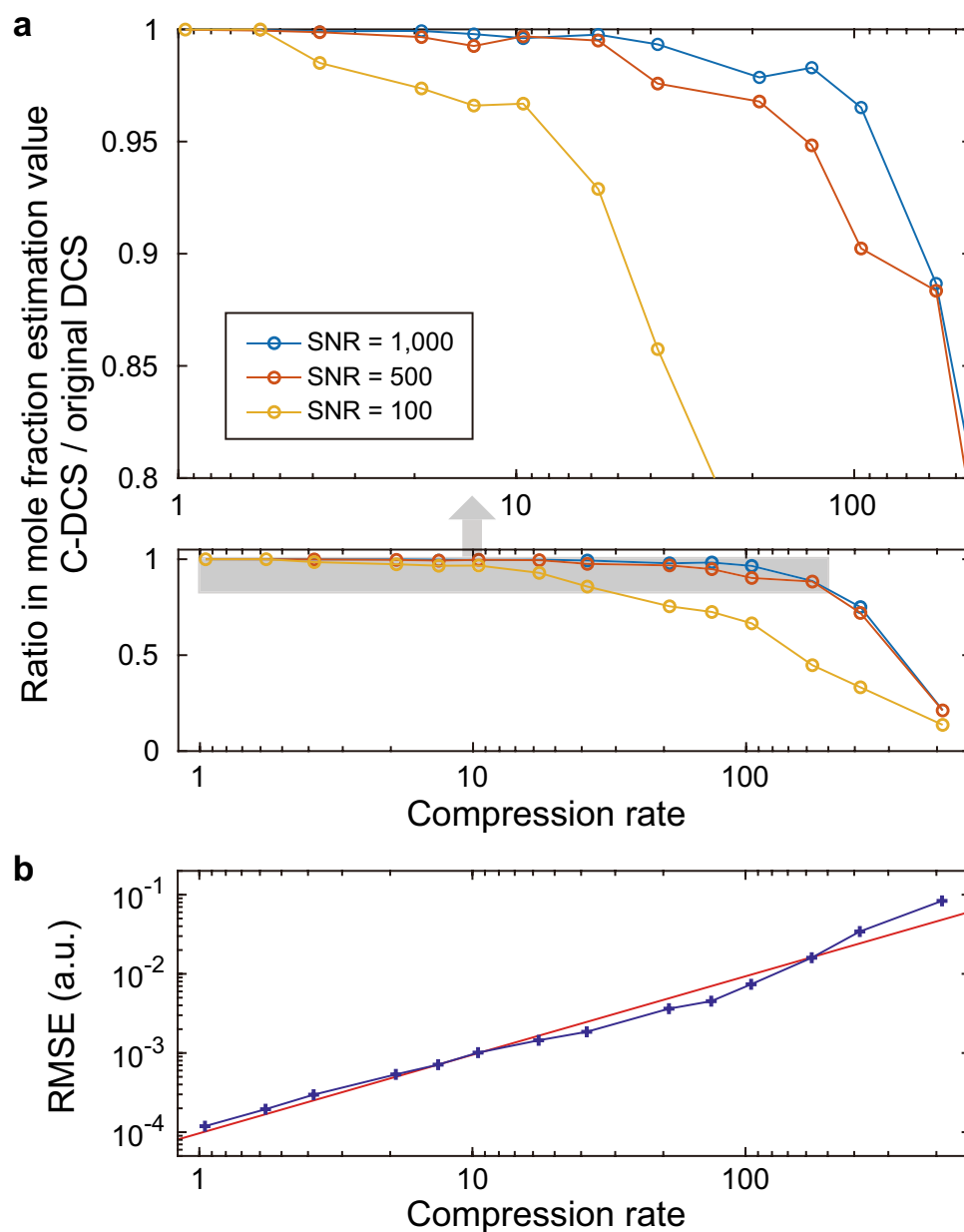


Figure 3. Mole fraction estimation from simulated spectra. (a) Ratio in mole fraction estimation value of C-DCS and original DCS with different compression rates. Blue, Red, and Yellow data show the results with different SNR. (b) Root-mean-squared error (RMSE) of the spectral points for each compression rate. The fitting (red line) shows that RMSE is proportional to (compression rate)^{0.93}.

Robustness evaluation of C-DCS reconstruction. We quantitatively evaluate the robustness (deviation from the ground truth) of the CS reconstruction in terms of the peak transmittance, center frequency, and linewidth of an absorption line by simulating spectra with 100 different patterns of random sampling for each condition. We analyze a single absorption line of N₂O at 2238.36 cm⁻¹ in the spectra calculated in the same condition as that shown in Fig. 2 with the SNR of 1000. Here we change the sample lengths (76, 11.4, 1.67, 0.76, 0.15 m) so that we can see how the absorption peak transmittance affects the CS reconstruction quality. Figure 4a shows the peak transmittance of the absorption line as a function of the compression rate. The standard deviations of the 100 spectra calculated with the different random samplings are illustrated as weak color bands around the mean values. Figure 4b shows the mean value divided by the standard deviation of the data points shown in Fig. 4a. It clearly shows the reconstruction degrades at higher compression rates, and the absorption lines with higher absorption intensities (longer sample lengths) are reconstructed more robustly. Figure 4c shows the deviation in the center wavenumber of the absorption line from the ground truth. The mean values are mostly within the spectral resolution (0.0038 cm⁻¹), showing its high robustness in the CS reconstruction. Figure 4d shows the relative linewidth to the ground truth. Although the absorption lines with higher peak

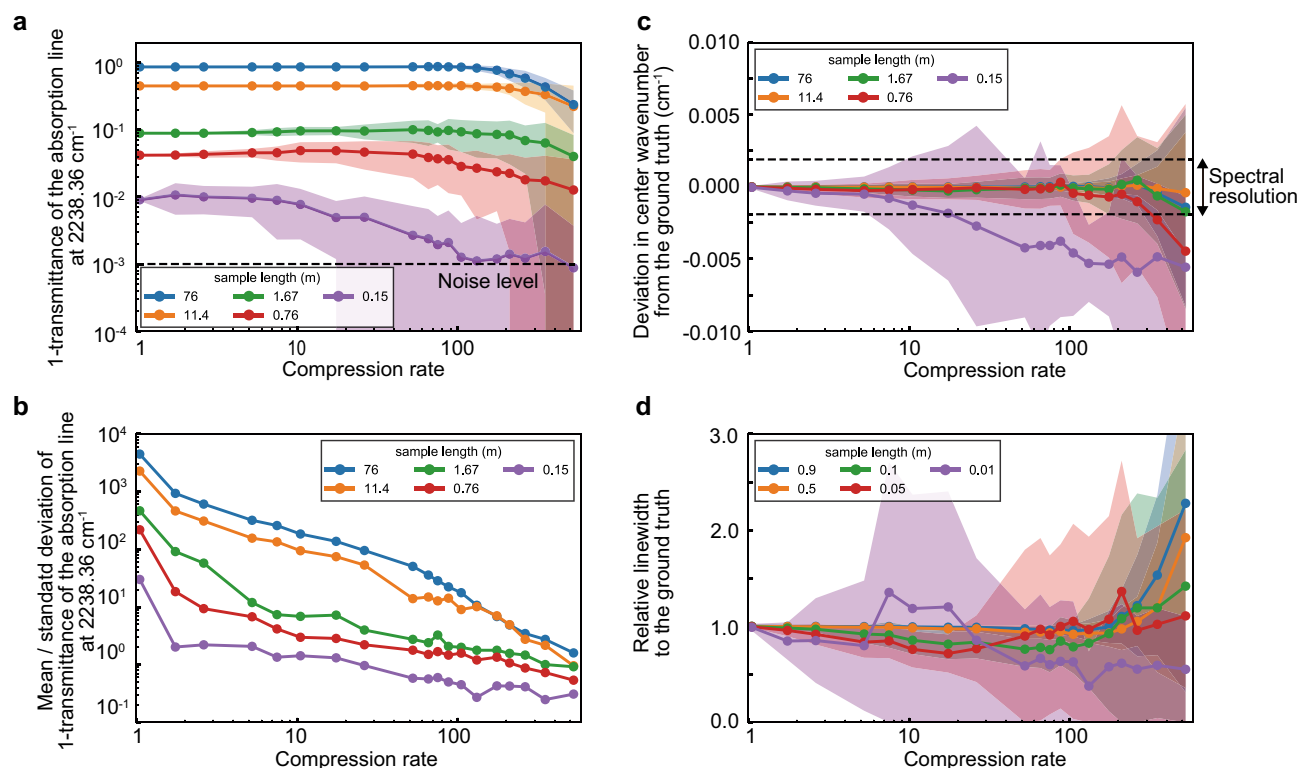


Figure 4. Robustness evaluation of CS reconstruction as a function of the compression rate. **(a)** peak transmittance. **(b)** mean/standard deviation of peak transmittance. **(c)** deviation in center wavenumber from the ground truth. **(d)** relative linewidth to the ground truth, of the absorption line at 2238.36 cm^{-1} . The plots with different colors show the calculation results with different sample lengths (blue: 76 m, yellow: 11.4 m, green: 1.67 m, red: 0.76 m, purple: 0.15 m). The dots and bands represent the mean values and standard deviations of the 100 spectra calculated with the different random samplings.

intensities are well reconstructed and no resolution degradation is observed up to the compression rate of ~ 100 , the ones with lower peak intensities largely deviate from the ground truth at the higher compression rates.

Massively parallel C-DCS of 10 trace gas species. Finally, to show the compression capability of C-DCS for denser molecular lines, we demonstrate massively parallel spectroscopy of 10 trace gas species, which resembles the previously reported experiment²¹. We assume a 76-m-long multi-pass gas cell filled with nitrous oxide ($^{14}\text{N}_2^{16}\text{O}$) at 42 ppm, nitric oxide ($^{14}\text{N}^{16}\text{O}$) at 420 ppm, carbon monoxide ($^{12}\text{C}^{16}\text{O}$) at 120 ppm, carbonyl sulfide ($^{16}\text{O}^{12}\text{C}^{32}\text{S}$) at 26 ppm, methane ($^{12}\text{CH}_4$) at 1,500 ppm, ethane ($^{12}\text{C}_2\text{H}_6$) at 490 ppm, ethylene ($^{12}\text{C}_2\text{H}_4$) at 540 ppm, acetylene ($^{12}\text{C}_2\text{H}_2$) at 6,600 ppm, carbon dioxide ($^{12}\text{CO}_2$, $^{13}\text{CO}_2$), at 280 ppm, water vapor (H_2^{16}O) at 2,100 ppm and a buffer gas. The most abundant isotope of each element except for CO_2 is included in this simulation. The pressure is set to 3 mbar. To relax regularization and let small absorption peaks remain, we increase the constraint term ϵ to the average of $\|n\|$. Figure 5 shows the C-DCS spectra with a compression rate of 10.5. We observe that vibrational absorption lines are well reconstructed compared to the original ones with the error (standard deviation) in transmittance of less than 0.003.

Discussion

The concept of C-DCS can apply to other Fourier-transform spectroscopy (FTS), including Michelson-type FTS, FT-Raman spectroscopy²⁷, and FT-CARS spectroscopy^{28,29}, etc. It can also be used for a spectrum measured in the spectral domain by calculating an interferogram by Fourier-transforming the spectrum. We point out that the CS is, in particular, useful for DCS because it generates a large amount of data with large spectral bandwidth and high spectral resolution, which ensures the most significant sparsity hence the efficient data compression. Also, the high scan rate of DCS generates a high-speed data stream that would cause foreseeable problems of data transportation and storage.

The performance of C-DCS can be improved by using other PMFs and/or reconstruction algorithms specifically developed for the use of CS imaging. The compression of C-DCS would become more valuable when we use it for higher dimensional measurements such as hyperspectral imaging³⁰ or multi-dimensional DCS³¹. Lastly, we note that C-DCS is possibly used for speeding up DCS measurement by implementing the compressive sampling in hardware. For that purpose, for example, we can arbitrarily sweep the difference in repetition rate during measurement of an interferogram, allowing a non-uniform temporal waveform sampling.

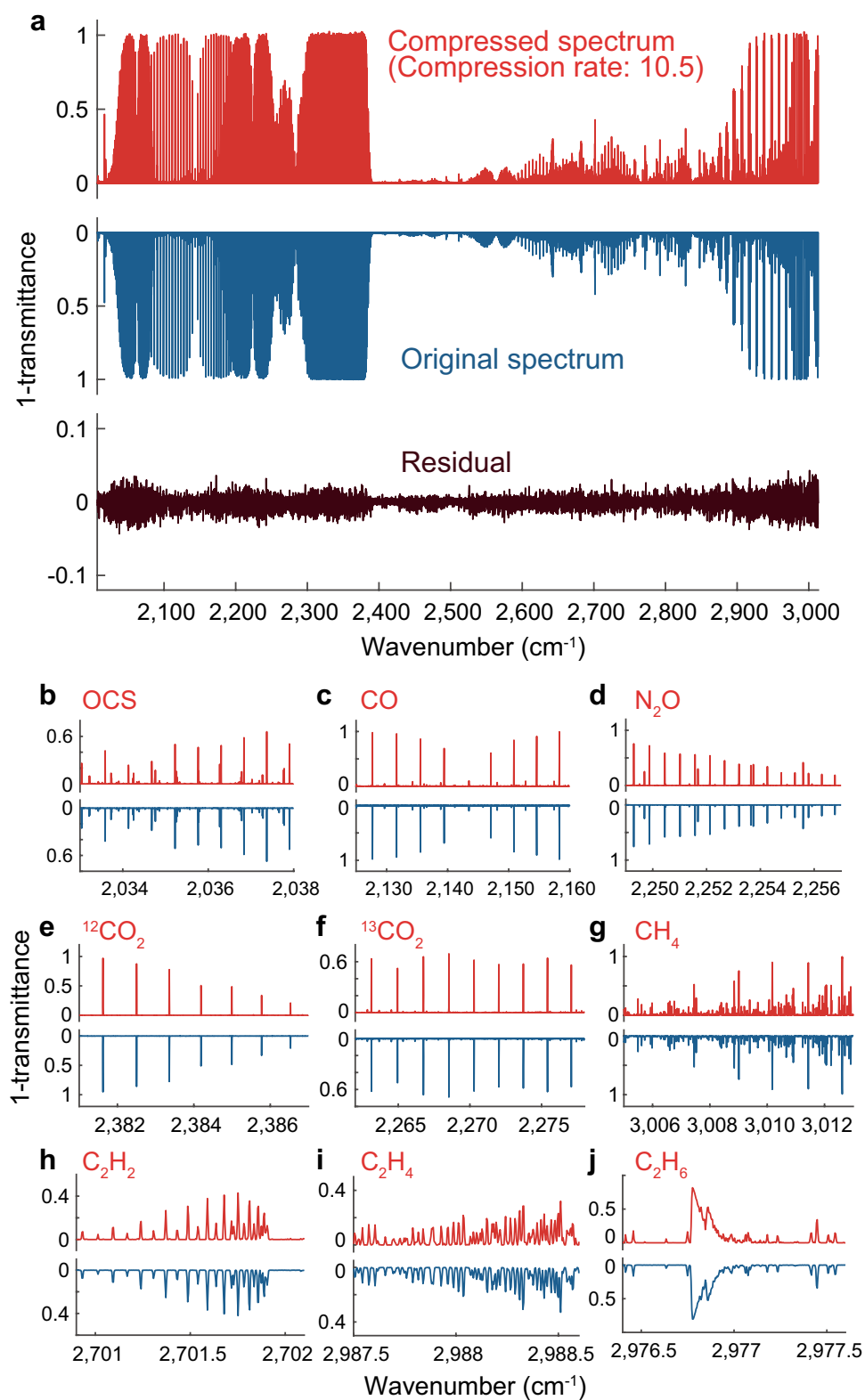


Figure 5. Massively parallel spectra of 10 trace gas species. **(a)** Comparison between a compressed spectrum and an original spectrum. The compressed spectrum is reconstructed from 50,000 points (compression rate of 10.5). **(b–j)** Zoom-in spectra of characteristic absorption lines of different molecular species.

Conclusion

We proposed compressive dual-comb spectroscopy (C-DCS) to address the data size problem of high-speed, broadband, and high-resolution dual-comb spectroscopy. Our numerical demonstration of C-DCS of the two molecular species (N_2O and CO) showed the reduction of the required data points by more than a factor of two under the allowable mole fraction estimation error of 3%. To investigate the robustness of C-DCS, we evaluated how the compression rate affects the transmittance, center wavenumber, and linewidth of the reconstructed absorption lines. Finally, we demonstrated the massively parallel sensing of 10 molecular species and showed the compression rate over 10. Our numerical studies show that the C-DCS is promising for solving the foreseeable data size problem in future dual-comb spectrometer deployment.

Data availability

The data provided in the manuscript are available from T.I. upon request.

Received: 24 March 2021; Accepted: 18 June 2021

Published online: 29 June 2021

References

- Coddington, I., Newbury, N. R. & Swann, W. C. Dual-comb spectroscopy. *Optica* **3**, 414–426 (2016).
- Ideguchi, T. Dual-comb spectroscopy. *Opt. Photon. News* **28**, 32–39 (2017).
- Picqué, N. & Hänsch, T. W. Frequency comb spectroscopy. *Nat. Photon.* **13**, 146–157 (2019).
- Baumann, E. *et al.* Spectroscopy of the methane ν_3 band with an accurate midinfrared coherent dual-comb spectrometer. *Phys. Rev. A* **84**, 062513 (2011).
- Rieker, G. B. *et al.* Frequency-comb-based remote sensing of greenhouse gases over kilometer air paths. *Optica* **1**, 290–298 (2014).
- Coburn, S. *et al.* Regional trace-gas source attribution using a field-deployed dual frequency comb spectrometer. *Optica* **5**, 320–327 (2018).
- Schroeder, P. J. *et al.* Dual frequency comb laser absorption spectroscopy in a 16 MW gas turbine exhaust. *Proc. Combust. Inst.* **36**, 4565–4573 (2017).
- Okubo, S. *et al.* Ultra-broadband dual-comb spectroscopy across 1.0–1.9 μm . *Appl. Phys. Express* **8**, 082402 (2015).
- Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
- Edgar, M. P., Gibson, G. M. & Padgett, M. J. Principles and prospects for single-pixel imaging. *Nat. Photon.* **13**, 13–20 (2019).
- Sun, M. J. *et al.* Single-pixel three-dimensional imaging with time-based depth resolution. *Nat. Commun.* **7**, 1–6 (2016).
- Ota, S. *et al.* Ghost cytometry. *Science* **360**, 1246–1251 (2018).
- Katz, O., Levitt, J. M. & Silberberg, Y. Compressive Fourier transform spectroscopy. in *Frontiers in Optics FTuE3* (Optical Society of America, 2010).
- Sanders, J. N. *et al.* Compressed sensing for multidimensional spectroscopy experiments. *J. Phys. Chem. Lett.* **3**, 2697–2702 (2012).
- Podmore, H. *et al.* Demonstration of a compressive-sensing Fourier-transform on-chip spectrometer. *Opt. Lett.* **42**, 1440–1443 (2017).
- Kästner, B. *et al.* Compressed sensing FTIR nano-spectroscopy and nano-imaging. *Opt. Express* **26**, 18115–18124 (2018).
- Takizawa, S., Hiramatsu, K. & Goda, K. Compressed time-domain coherent Raman spectroscopy with real-time random sampling. *Vib. Spectrosc.* **107**, 103042 (2020).
- Mehta, N., Chen, J., Zhang, Z. & Liu, Z. Compressive multi-heterodyne optical spectroscopy. *Opt. Express* **20**, 28363–28372 (2012).
- Coddington, I., Swann, W. C. & Newbury, N. R. Time-domain spectroscopy of molecular free-induction decay in the infrared. *Opt. Lett.* **35**, 1395–1397 (2010).
- Tomberg, T., Muraviev, A., Ru, Q. & Vodopyanov, K. L. Background-free broadband absorption spectroscopy based on interferometric suppression with a sign-inverted waveform. *Optica* **6**, 147–151 (2019).
- Muraviev, A. V., Smolski, V. O., Loparo, Z. E. & Vodopyanov, K. L. Massively parallel sensing of trace molecules and their isotopologues with broadband subharmonic mid-infrared frequency combs. *Nat. Photon.* **12**, 209–214 (2018).
- Kochanov, R. V. *et al.* HITRAN Application Programming Interface (HAPI): A comprehensive approach to working with spectroscopic data. *J. Quant. Spectrosc. Radiat. Transf.* **177**, 15–30 (2016).
- Moshtaghpour, A., Bioucas-Dias, J. M. & Jacques, L. Compressive single-pixel Fourier transform imaging using structured illumination. in *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 7810–7814 (IEEE, 2019).
- Van Den Berg, E. & Friedlander, M. P. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* **31**, 890–912 (2009).
- Chen, Z., Yan, M., Hänsch, T. W. & Picqué, N. A phase-stable dual-comb interferometer. *Nat. Commun.* **9**, 1–7 (2018).
- Sterczewski, L. A., Westberg, J. & Wysocki, G. Computational coherent averaging for free-running dual-comb spectroscopy. *Opt. Express* **27**, 23875–23893 (2019).
- Griffiths, P. R. & De Haseth, J. A. *Fourier Transform Infrared Spectrometry* (Wiley, 2007).
- Ideguchi, T. *et al.* Coherent Raman spectro-imaging with laser frequency combs. *Nature* **502**, 355–358 (2013).
- Hashimoto, K., Omachi, J. & Ideguchi, T. Ultra-broadband rapid-scan Fourier-transform CARS spectroscopy with sub-10-fs optical pulses. *Opt. Express* **26**, 14307–14314 (2018).
- Martín-Mateos, P., Khan, F. U. & Bonilla-Manrique, O. E. Direct hyperspectral dual-comb imaging. *Optica* **7**, 199–202 (2020).
- Lomsadze, B., Smith, B. C. & Cundiff, S. T. Tri-comb spectroscopy. *Nat. Photon.* **12**, 676–680 (2018).

Acknowledgements

R.H. and T.I. thank Japan Science and Technology Agency for their support. This work was financially supported by JST PRESTO (JPMJPR17G2), and JSPS KAKENHI (20H00125).

Author contributions

T.I. conceived the concept. A.K. designed the numerical experiments. A.K. and T.K. performed the numerical experiments. R.H. contributed to the design of the numerical experiments and data interpretation. T.I. supervised the entire work. A.K. and T.I. wrote the manuscript with contributions from all the other authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021