Ultrasensitive multispecies spectroscopic breath analysis for real-time health monitoring and diagnostics

Qizhong Liang[a,b], Ya-Chu Chan[a,c], P. Bryan Changala[d], David J. Nesbitt[a,b,c], Jun Ye[a,b,1], and Jutta Toscano[a,b,1]

[a]JILA, National Institute of Standards and Technology and University of Colorado Boulder, Boulder, CO 80309; [b]Department of Physics, University of Colorado Boulder, Boulder, CO 80309; [c]Department of Chemistry, University of Colorado Boulder, Boulder, CO 80309; and [d]Center for Astrophysics, Harvard College Observatory and Smithsonian Astrophysical Observatory, Cambridge, MA 02138

Edited by Marsha I. Lester, University of Pennsylvania, Philadelphia, PA, and approved August 30, 2021 (received for review March 15, 2021)

Breath analysis enables rapid, noninvasive diagnostics, as well as long-term monitoring of human health, through the identification and quantification of exhaled biomarkers. Here, we demonstrate the remarkable capabilities of mid-infrared (mid-IR) cavity-enhanced direct-frequency comb spectroscopy (CE-DFCS) applied to breath analysis. We simultaneously detect and monitor as a function of time four breath biomarkers—CH_3OH, CH_4, H_2O, and HDO—as well as illustrate the feasibility of detecting at least six more (H_2CO, C_2H_6, OCS, C_2H_4, CS_2, and NH_3) without modifications to the experimental apparatus. We achieve ultrahigh detection sensitivity at the parts-per-trillion level. This is made possible by the combination of the broadband spectral coverage of a frequency comb, the high spectral resolution afforded by the individual comb teeth, and the sensitivity enhancement resulting from a high-finesse cavity. Exploiting recent advances in frequency comb, optical coating, and photodetector technologies, we can access a large variety of biomarkers with strong carbon–hydrogen-bond spectral signatures in the mid-IR.

breath analysis | spectroscopy | frequency combs

Breath analysis is an exceptionally promising and rapidly developing field of research, which examines the molecular composition of exhaled breath (1–6). The hundreds of different gases that are present in exhaled breath include inorganic compounds, as well as volatile organic compounds (VOCs), and can either result from internal metabolic activity (endogenous emissions) or external factors, such as food consumption or environmental exposure (exogenous emissions). Despite its distinctive advantages of being a rapid, noninvasive technique and its long history dating back to Hippocrates, breath analysis has not yet been as widely deployed for routine diagnostics and monitoring as other methods, such as blood-based analysis. This is partly due to the experimental challenges of dealing with extremely small amounts of gas-phase molecules—in the parts-per-million (ppm) to parts-per-billion (ppb) range for most VOCs—and partly due to the relative scarcity of large-scale clinical studies that can reliably correlate specific diseases with biomarkers present in breath. Nevertheless, through close collaborations between instrument developers, breath-analysis experts, and clinicians, the field of breath analysis is fast approaching its goal of enabling real-time, noninvasive early detection and long-term monitoring of temporary and permanent health conditions (1, 3). Several biomarkers present in breath have been associated with specific conditions—for instance, nitrogen monoxide with asthma, acetone with diabetes, and ammonia with renal failure (5)—and breath is increasingly being used to track diseases and infections, both bacterial and viral (7). Recently, three studies have demonstrated the use of breath analysis to discriminate between SARS-CoV-2–infected patients and patients affected by other conditions (including asthma, chronic obstructive pulmonary disease, bacterial pneumonia, and cardiac conditions) (8, 9) or influenza A-infected patients (10). The possibility of real-time testing for highly infectious diseases in a noninvasive manner, without the need for chemical reagents and complex laboratory facilities, is particularly appealing in view of the current global pandemic.

Technologies being explored and adopted for breath analysis include mass spectrometry, nanomaterial-based sensors, and laser spectroscopy. To date, the most widely used analytical technique in breath research is gas chromatography combined with mass spectrometry, which allows for the sensitive detection of hundreds of exhaled molecules, albeit with relatively long analysis times (tens of minutes) limited by the elution time of the various species. On the other hand, selected ion-flow-tube mass spectrometry and proton-transfer reaction mass spectrometry allow for real-time breath analysis at the expense of a reduced number of simultaneously detectable molecules (11). Sensor arrays offer an inexpensive and practical alternative for identifying the presence of a class of compounds based on their functional groups, but they generally do not permit identification of the specific molecules present in the samples (9, 12). Laser spectroscopy is intrinsically fast ($\ll$ second timescale), allowing breath-cycle-resolved (i.e., respiratory-phase-resolved) sampling of breath with high precision and absolute accuracy. Achieving high sensitivity requires both signal enhancement and noise reduction: The former is attained by using multipass cells or high-finesse cavities, while the latter is accomplished

Significance

Determining the identity and concentration of the molecules present in breath is a powerful tool to assess the overall health of a person, analogous to blood testing in clinical medicine, but in a faster and less invasive manner. The presence of a particular molecule (or combination of molecules) can indicate the presence of a certain health condition or infection, facilitating a diagnosis. Monitoring the concentration of the molecules of interest over time can help track the development (or recurrence) of a condition, as well as the effectiveness of the administered treatment. To make breath analysis more accessible and widely adopted, we explore a technique to perform breath analysis with high resolution, wide spectral coverage, high specificity, and high sensitivity.

Author contributions: Q.L., Y.-C.C., P.B.C., D.J.N., J.Y., and J.T. designed research; Q.L., Y.-C.C., and J.T. performed research; Q.L., Y.-C.C., P.B.C., D.J.N., J.Y., and J.T. contributed new reagents/analytic tools; Q.L., Y.-C.C., and J.T. analyzed data; and Q.L., Y.-C.C., P.B.C., D.J.N., J.Y., and J.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jutta.toscano@jila.colorado.edu or ye@jila.colorado.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2105063118/-/DCSupplemental.

Published October 1, 2021.

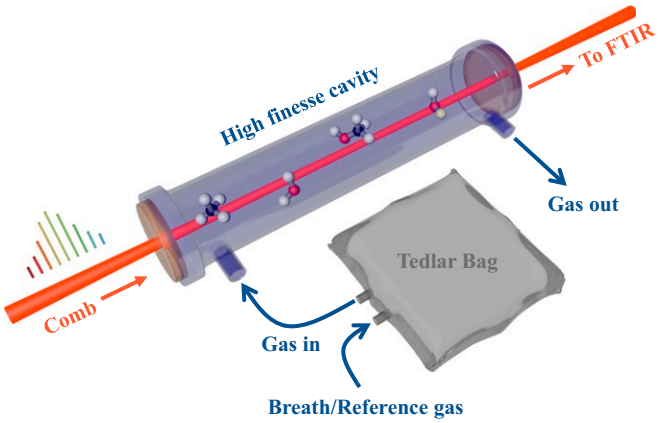

Fig. 1. Schematic representation of the experimental apparatus. Breath—or, alternatively, a reference gas—is collected into a Tedlar bag and subsequently loaded into a high-finesse cavity comprising a pair of high-reflectivity mirrors. A mid-IR frequency comb is resonantly coupled into the cavity and interacts with the molecules present inside the cavity during several thousand round trips. The cavity-transmitted comb light is analyzed by a Fourier-transform IR spectrometer to determine the identity and concentration of the molecular species present in the gas sample.

through intensity or frequency-modulation techniques. Among others, tunable diode laser absorption spectroscopy, cavity ring-down spectroscopy, cavity-enhanced absorption spectroscopy, and photoacoustic spectroscopy have all successfully been employed in breath analysis, but are typically limited in tunability and therefore in the number of detectable analytes (1). Cavity-enhanced direct-frequency comb spectroscopy (CE-DFCS) offers substantially enhanced capabilities for the simultaneous detection of multiple species due to the combination of high spectral resolution, wide spectral coverage, and high sensitivity (13–18). An early study from 2008 demonstrated this by detecting carbon monoxide, carbon dioxide, methane, ammonia, and water in breath samples by CE-DFCS (19). This previous work measured vibrational (mainly first) overtone transitions in the near-infrared (near-IR) region of the spectrum, from 1.5 μm to 1.7 μm.

Here, we report a 2-orders-of-magnitude improvement in the detection sensitivity for multiple species relevant to breath analysis by using CE-DFCS in the mid-infrared (mid-IR) molecular fingerprint region (3.4–3.6 μm). We gain access to fundamental vibrational transitions, as well as employ higher-finesse mid-IR cavity mirrors (15, 20), compared to previous work in this spectral region (21). Exploiting recent advances in frequency comb, high-reflectivity optical coating, and photodetector technologies, we can detect a large variety of biomarkers simultaneously, sensitively, and unambiguously, providing exciting prospects to connect breath to a range of biological functions and diseases.

Results

A frequency comb is a coherent light source consisting of hundreds of thousands of discrete, equidistant, and narrow frequency modes, often referred to as comb teeth. Despite the large number of comb teeth spanning a broad wavelength range, the exact frequency of each individual tooth can be precisely controlled through frequency stabilization (22–25). The

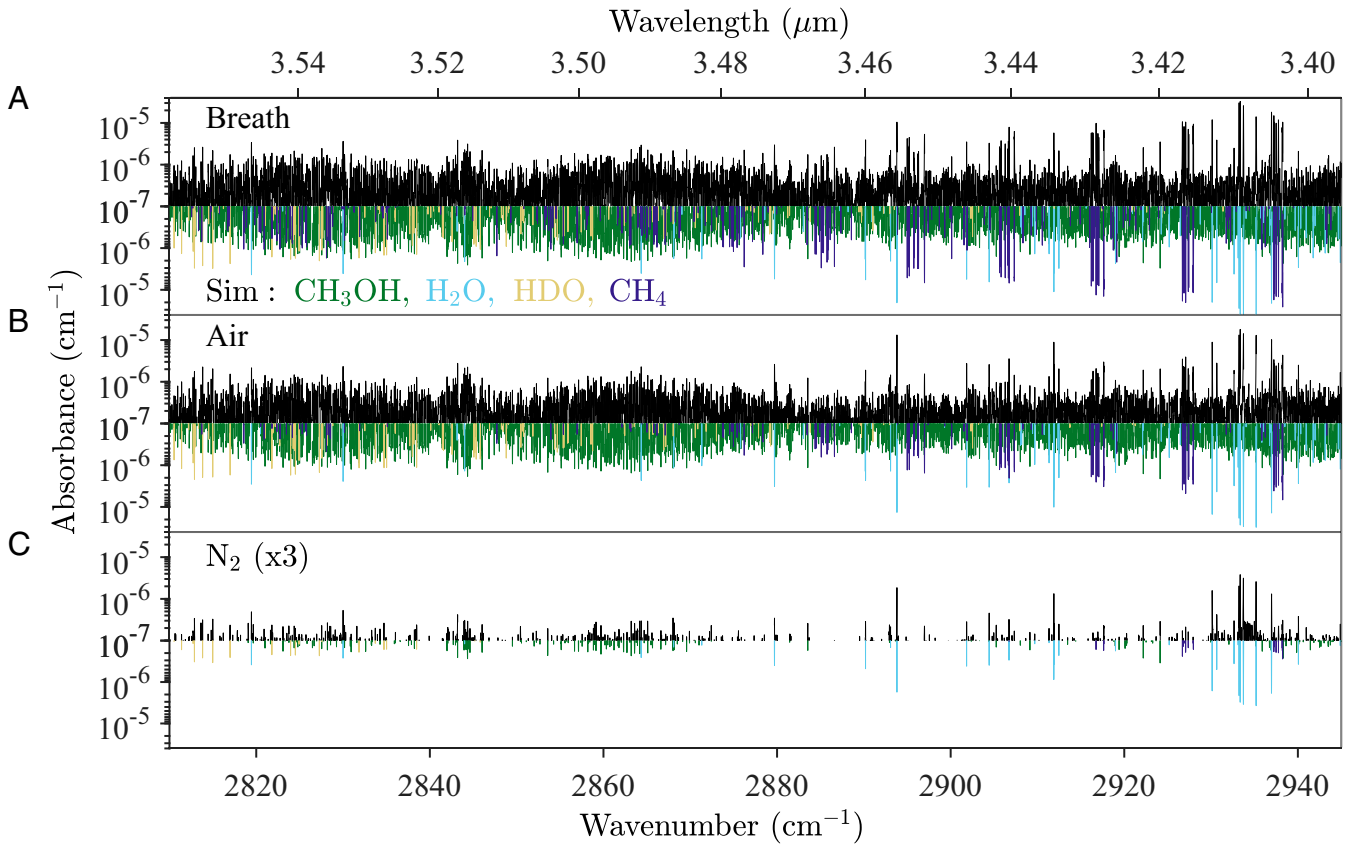

Fig. 2. Experimental and simulated absorbance spectra of breath (*A*), laboratory-environment air (*B*), and nitrogen gas (*C*). The experimental data are depicted in black, while the simulated data are shown inverted with the color corresponding to the absorbing species: methanol (green), water (light blue), partially deuterated water (gold), and methane (dark blue). Note that the molecules detected in the nitrogen gas spectrum (which has been magnified by a factor of 3) are remaining impurities following the cleaning protocol performed between measurements (see *Results*).

simultaneous highly resolved and broadband nature of frequency combs make them ideal light sources for spectroscopy. Additionally, the high degree of control over the exact frequency of each comb tooth enables efficient coupling to a high-finesse enhancement cavity by matching the frequencies of the comb teeth with those of the resonant cavity modes. The combination of frequency combs and high-finesse cavities adds high sensitivity to the previously mentioned benefits of high resolution and broadband spectral coverage. To exploit all three of these features for breath-analysis applications, we employ a high-power mid-IR frequency comb (26) in combination with an ultrahigh-finesse mid-IR enhancement cavity and near-shot-noise-limited detection to achieve exceptionally high sensitivity for a range of molecules present in breath.

The breath sample or a reference gas is introduced directly into the enhancement cavity, where it absorbs part of the comb light; the absorbed frequency components are subsequently detected by using a Fourier-transform spectrometer (Fig. 1). Typical absorption spectra obtained in this way are shown in Fig. 2. For the breath measurements (Fig. 2*A*), the volunteer (one of the coauthors; age: 25, body weight: 55 kg) is asked to inhale to total lung capacity before blowing exclusively the second half of the exhalation into the reusable Tedlar bag through a mouthpiece (end-tidal breath sampling; inhalation time: 2 s, breath-holding time: 1 s, exhalation time: 10 s). Immediately after collection, the sample contained in the Tedlar bag is loaded into the stainless steel cavity for analysis via a Teflon tubing gas manifold. Measurements of the background laboratory air (Fig. 2*B*) are taken for comparison and constitute a baseline for the breath measurements. The nitrogen gas measurements (Fig. 2*C*), on the other hand, are taken before and after each breath or air measurement, following the cleaning of the Tedlar bag and of the enhancement cavity by flowing nitrogen gas through the system for 5 min (with a flow of $\approx$ 30 standard liters per minute). These serve to check the cleanliness of the system and control against cross-contamination from one sample to the next.

Within the examined spectral window (2,810 cm^{-1} to 2,945 cm^{-1}), we observe absorption features from four molecular species: methanol (CH_3 *s*-stretch, ν_3, with Q-branch at 2,844 cm^{-1} and CH_3 *d*-stretch, ν_2 and ν_9), methane (P-branch of the ν_3 band), water (first overtone of the bend, $2\nu_2$), and partially deuterated water (first overtone of the bend, $2\nu_2$). We fit the experimental absorption spectra to those calculated from the High-Resolution Transmission Molecular Absorption (HITRAN) database (27, 28) to extract the concentration of each species. These simulated spectra are shown in Fig. 2 (see *SI Appendix* for more details). The multiline, global fitting procedure we adopt allows us to fit all species at once across the entire spectral window, enabling a precise determination of the molecular concentrations, despite the presence of overlapping spectral features. We find the measured concentrations of methanol, methane, and water in the volunteer's breath sample

PHYSICS

Table 1. Molecules absorbing in the spectral region probed in this work that are relevant for breath analysis

	Achieved (2,810 to 2,945 cm^{-1})				Potential (2,100 to 3,600 cm^{-1})	
Molecule	Max. absorption cross-section (cm^2 $molecule^{-1}$)	Single-elem. detection limit	Multi-elem. detection limit	Spectral element utilization (%)	Max. absorption cross-section (cm^2 $molecule^{-1}$)	Clinical relevance (3, 6, 12)
H_2CO	4×10^{-18} (at 2,815 cm^{-1})	467 ppt	126 ppt	98.30	4×10^{-18} (at 2,815 cm^{-1})	Lung cancer
C_2H_6	1×10^{-18} (at 2,907 cm^{-1})	2.4 ppb	378 ppt	99.08	1×10^{-17} (at 2,983 cm^{-1})	Asthma; Chronic obstructive pulmonary disease; Inflammatory bowel disease
CH_4*	4×10^{-18} (at 2,938 cm^{-1})	2.3 ppb	644 ppt	60.98	1×10^{-17} (at 3,058 cm^{-1})	Intestinal problems; Bacteria and colonic fermentation
CH_3OH*	3×10^{-19} (at 2,924 cm^{-1})	4.1 ppb	722 ppt	99.33	5×10^{-19} (at 2,982 cm^{-1})	Hangover
OCS	8×10^{-19} (at 2,926 cm^{-1})	5.0 ppb	900 ppt	50.11	8×10^{-19} (at 2,926 cm^{-1})	Liver fetor
HDO*	8×10^{-19} (at 2,815 cm^{-1})	4.8 ppb	1.4 ppb	44.60	2×10^{-18} (at 2,720 cm^{-1})	Measurement of total body water weight (through measurement of HDO/H_2O concentration ratio)
C_2H_4	9×10^{-20} (at 2,943 cm^{-1})	72 ppb	19 ppb	11.58	6×10^{-19} (at 3,085 cm^{-1})	Lipid peroxidation
CS_2	2×10^{-20} (at 2,840 cm^{-1})	65 ppb	17 ppb	15.04	2×10^{-18} (at 2,178 cm^{-1})	Halitosis; Schizophrenia
NH_3	1×10^{-20} (at 2,893 cm^{-1})	201 ppb	103 ppb	2.22	8×10^{-19} (at 3,336 cm^{-1})	Asthma; Halitosis; Chronic renal failure/uremia
H_2O*	3×10^{-21} (at 2,935 cm^{-1})	1.5 ppm	813 ppb	0.52	2×10^{-18} (at 3,589 cm^{-1})	Measurement of total body water weight (through measurement of HDO/H_2O concentration ratio)

The maximum absorption cross-section—with the corresponding wavenumber shown in parentheses underneath—indicates the strongest absorption peak in the probed spectral range. The single-element detection limit is determined at the single frequency with highest signal-to-noise ratio within the probed spectral range. It represents the detection limit by a single frequency comb tooth. Using the transition lines contained in the entire spectrum—instead of the absorption at a single frequency—yields the multi-element detection limit, which is (depending on the molecule) between two and seven times lower than the single-element detection limit. The spectral element utilization column reports the percentage of the entire spectrum that contains absorbing features for each molecule, using an arbitrary threshold cross-section of 1×10^{-22} $cm^2 \cdot molecule^{-1}$. With the spectral tunability of the mid-IR frequency comb covering the range 2,100 to 3,600 cm^{-1}, the reported detection limits could be further improved by probing the molecules closer to their maximum absorption features, shown under the "Potential" column, which takes into account a wider spectral region (2,100 to 3,600 cm^{-1}) compared to the one probed in this work (2,810 to 2,945 cm^{-1}). The average simulated Lorentzian and Gaussian linewidth components are 0.31(8) and 0.21(6) GHz, respectively. (See *SI Appendix* for more details.)

*Molecules detected in this work.

(CH_3OH = 12.39(2) ppm, CH_4 = 8.52(9) ppm, H_2O = 17.6(3) ‰) to be higher than the respective background concentrations in air (CH_3OH = 8.64(5) ppm, CH_4 = 1.95(3) ppm, H_2O = 11.5(2) ‰). In contrast, we measure similar concentrations of partially deuterated water in breath (HDO = 4.2(1) ppm) and air samples (HDO = 3.17(8) ppm). The concentrations of the four species in the nitrogen gas sample are 1 or 2 orders of magnitude lower than those measured in the breath and air samples (CH_3OH = 0.410(4) ppm, CH_4 = 34(2) ppb, H_2O = 0.432(3) ‰, HDO = 152(7) ppb), demonstrating the effectiveness of the aforementioned cleaning protocol.

Although the breath samples analyzed in this work only contain the four species discussed above, several other molecules absorb within the probed spectral region and could therefore also be detected if they were present in the analyzed samples (Table 1). From the noise floor of our spectra, we estimate the absorption sensitivity per spectral element of our experiment to be $5(1)\times10^{-11}$ $cm^{-1}\cdot Hz^{-1/2}$ (see *SI Appendix* for more details). Such exceptionally high sensitivity corresponds to especially low minimum detectable concentrations, which are reported for several species of interest in Table 1. In particular, detection sensitivity at the parts-per-trillion (ppt) level is achievable in our apparatus for five molecules of clinical relevance to breath analysis: H_2CO, C_2H_6, CH_4, CH_3OH, and OCS. A ppb-level sensitivity is attainable for the other five molecules: HDO, C_2H_4, CS_2, NH_3, and H_2O. Given the high concentration of water vapor present in breath, its slightly higher detection limit (813 ppb) is not detrimental, but rather advantageous, since it mitigates potential saturation issues without affecting the ability to detect its presence in breath samples. Compared to a single-frequency spectroscopy experiment (single-element detection limit in Table 1), the simultaneous detection of a broad range of frequencies demonstrated in this study affords a multi-element detection limit that is between two and seven times lower, depending on the molecule. In essence, CE-DFCS not only enables the detection of several molecules simultaneously, but it also lowers each of their minimum detectable concentrations.

To demonstrate the feasibility of applying CE-DFCS to real-time breath monitoring, we tracked the concentration of the molecules present in the breath of the same volunteer before and after the consumption of ripe fruit (specifically, 10 ripe baby bananas, ≈ 0.5 kg). For the duration of the experiment (7 h), as well as for a period of 12 h before the beginning of the experiment, the volunteer did not consume any other food or drink. To achieve better time resolution, here we focus on a narrower spectral range (2,853.0 cm^{-1} to 2,864.5 cm^{-1}) (Fig. 3*A*) while maintaining the ability to track all four molecules observed in Fig. 2. A breath spectrum is recorded approximately every 15 min and an air spectrum approximately every 1.5 h, for a total of 23 breath and 5 air spectra collected. Between each measurement, the Tedlar bag and enhancement cavity are cleaned following the procedure outlined above, and then a nitrogen gas spectrum is collected to check for remaining impurities (for a total of 28 N_2 spectra). Throughout the experiment, the breath fractional concentrations of water, partially deuterated water, and methane fluctuate around average values of 14(4) ‰, 4.3(4) ppm, and 2.4(5) ppm, respectively (Fig. 3*B*). However, the breath fractional concentration of methanol increases from 534(64) ppb to 929(16) ppb after the consumption of fruit, leveling off after ∼2.5 h (Fig. 3 *B* and *C*). A similar trend is observed on a second day of measurements following the same experimental protocol (Fig. 3*C*). The initial concentration and the observed increase in methanol following fruit consumption are consistent with

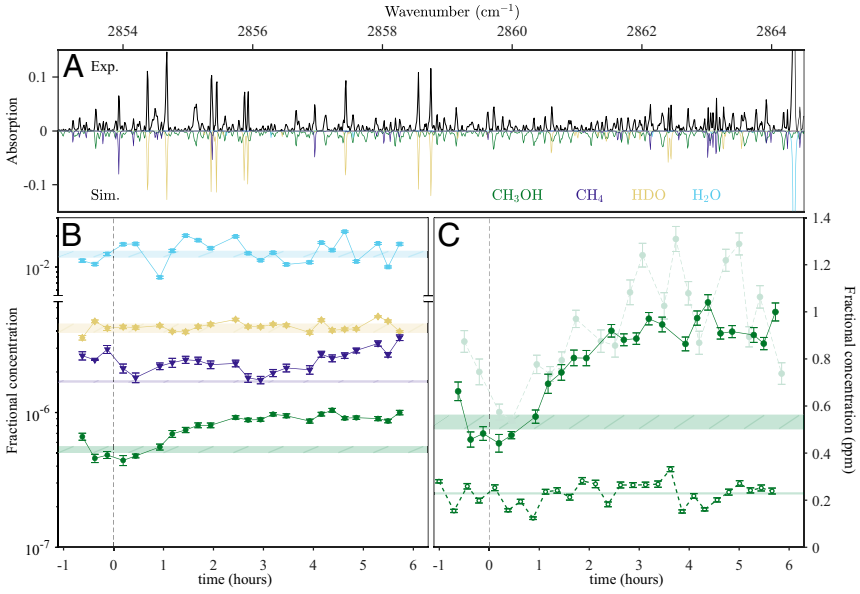

Fig. 3. Measurement of concentrations over time. (*A*) Sample spectrum of breath in the spectral region used for this measurement. The experimental (Exp.) data are depicted in black, while the simulated (Sim.) data are shown inverted with the color corresponding to the absorbing species: methanol (green), water (light blue), partially deuterated water (gold), and methane (dark blue). (*B*) Fractional concentration of the four absorbing species as a function of time. The data points correspond to the concentrations obtained from the breath samples, with the error bars showing the 1σ uncertainty from the fit to the experimental spectra. The shaded hatched regions represent the mean and SE of the concentrations measured for the laboratory air samples (background measurements). Note the large dynamic range of measured fractional concentrations in the order of 10^5; the linearity of the instrument response upon dilution has been verified. (*C*) Methanol concentration as a function of time for breath (filled circles) and nitrogen gas (open circles). The shaded hatched (nonhatched) regions represent the mean and SE of the laboratory air (nitrogen gas) measurements. Note that the breath data points (filled circles) and shaded hatched region are the same as in *B*, but plotted on a linear scale. For reference, a second set of data for methanol concentration in breath collected on a different day following the same protocol is also shown (pale filled circles); the mean and SE of the background methanol fractional concentration for the second data set ($5.6(6)\times10^{-7}$) is comparable to the one shown. The vertical dashed line in *B* and *C* denotes the time of fruit consumption.

previous work, with the latter resulting from endogenous production of methanol (29, 30).

Discussion

The ultrahigh molecular detection sensitivity achieved in this study stems from three factors: 1) the exceptionally high absorption sensitivity per spectral element of $5(1) \times 10^{-11}\text{cm}^{-1} \cdot \text{Hz}^{-1/2}$; 2) the large absorption cross-sections available in the mid-IR spectroscopic region (up to 4 orders of magnitude larger than those in the near-IR); and 3) the simultaneous measurement of multiple spectral frequencies enabled by the wide spectral coverage of a frequency comb. These factors are responsible for the 2-orders-of-magnitude improvement over the sensitivity per spectral element recorded by the previous CE-DFCS study of breath, which probed the near-IR spectral region instead (19). Compared to previous direct-frequency comb spectroscopy work in the mid-IR region, our detection limits for methane and methanol are 3 and 5 times lower, respectively, and our predicted detection limit for formaldehyde is 18 times lower (31). In this case, the critical factor for such remarkable improvement comes from using a high-finesse enhancement cavity instead of a multipass cell. Overall, our sensitivity is comparable to the highest sensitivity per spectral element achieved in the mid-IR region by CE-DFCS to date [$6.9 \times 10^{-11}\text{cm}^{-1} \cdot \text{Hz}^{-1/2}$ (32)]. It is worth noting that, while the goal of this work is to enable the simultaneous detection of numerous different species, it is possible to achieve even lower detection limits than those reported here by probing the spectral region with the strongest absorption for each individual molecule (Table 1). This could be easily attained with the current setup, upon replacement of the high-reflectivity cavity mirrors. In general, the examined spectral region can be selected depending on the specific clinical needs.

Detecting multiple species simultaneously is particularly desirable in the exploratory phase of breath research, as well as with the view to monitor health through breath analysis as it is currently done through blood analysis. One of the main drawbacks of using optical methods for breath analysis is the limited number of molecular species that can be detected compared to mass-spectrometric techniques. Indeed, all but one of the optical-based breath tests approved by the US Food and Drug Administration detect a single species (1). Here, we push this limit and demonstrate that CE-DFCS can potentially detect tens of species simultaneously, provided that they exhibit sharp absorption features (*Materials and Methods*). The latter requirement makes this technique most suitable for detecting light biomarkers, which—depending on the exact ionization technique employed—can be challenging to sensitively and reliably detect by mass spectrometry. This is due to the fragmentation of large molecules, with fragments potentially appearing at the same mass-to-charge ratio as smaller molecules in mass spectra, making it difficult to disentangle the contribution to the signal from light biomarkers and that from fragments of heavier biomarkers. Additionally, IR absorption spectroscopy is both isomer- and isotopologue-specific, whereas mass spectrometry is not typically able to distinguish between isomers. In these respects, frequency comb spectroscopy and mass spectrometry are complementary techniques for breath analysis. Finally, it is worth mentioning that it is possible to use independent laser sources to track different molecules in breath simultaneously, as it has been elegantly demonstrated for the real-time monitoring of O_2, CO_2, and H_2O to track oxygen consumption during an aortic aneurysm repair operation (33) and of NO and CO_2 to record accurate NO expirograms (34). Using a single broadband laser source to detect all molecules, such as is shown in this work with a frequency comb, allows more accurate determination of the relative concentrations of the probed molecules and may therefore be preferable for some applications.

Conclusion

By operating in the mid-IR molecular fingerprint region, we have demonstrated the unique and powerful capabilities of mid-IR CE-DFCS for breath analysis, including:

- Ultrahigh detection sensitivity at the ppt and ppb levels;
- Simultaneous detection of four breath biomarkers, with the potential to detect at least six more;
- Real-time, simultaneous monitoring of biomarkers to track the physiological response to stimuli (food).

This is made possible by the combination of the broadband spectral coverage of a frequency comb, the high spectral resolution afforded by the individual comb teeth, and the sensitivity enhancement resulting from the high-finesse cavity. Probing the mid-IR spectral region is beneficial due to the presence of strong fundamental vibrational transitions, while probing a large spectral range allows us to detect different molecules simultaneously, as well as to improve our detection limit. The simultaneous detection of different species leads to faster measurements, but also to more reliable determination of relative and absolute concentrations. In conclusion, we have shown that this technique offers unique advantages and opportunities for the detection of light biomarkers in breath, and it is poised to facilitate real-time, noninvasive monitoring of breath for clinical studies, as well as for early detection and long-term monitoring of temporary and permanent health conditions.

Materials and Methods

This work is a prelude for a planned breath analysis study for SARS-CoV-2 screening, which is a medical study for which an institutional review board (IRB) approval has been sought and obtained (protocol no. 21-0088). The work presented here, on the other hand, constitutes a validation study with the purpose of characterizing the experimental apparatus using breath samples volunteered by one of the coauthors, and, as such, it is exempt from IRB approval according to the University of Colorado Boulder IRB.

Optical Frequency Comb. The mid-IR frequency comb used in this study is generated from a singly-resonant optical parametric oscillator (OPO) synchronously pumped by a femtosecond Yb:fiber laser at a repetition rate of 136 MHz. The center frequency of the mid-IR comb is tunable from 2,100 cm^{-1} to 3,600 cm^{-1} by adjusting the OPO crystal position inside the oscillator cavity. In this work, we have fixed the center frequency to be near 2,880 cm^{-1} with the intent of detecting a large number of molecular species of clinical relevance, instead of maximizing the detection sensitivity for any particular species to its ultimate limit. We typically use 50 to 100 mW of mid-IR optical power. More details on the light source can be found in ref. 26.

Cavity-Enhanced Spectroscopy. The frequencies of the mid-IR comb and the enhancement cavity are fully stabilized and referenced to an ultralow-phase–noise quartz oscillator, which is slaved to a cesium clock (see *SI Appendix* for details on the servo loops implementation). The cavity free spectral range is maintained at twice the comb repetition rate, which results in the filtering of every other cavity incident comb line and doubling of the repetition rate of the transmitted comb light to 272 MHz. The cavity finesse ranges between 6,000 and 8,000 in the probed spectral region (see *SI Appendix* for more details), and the spectral coverage of the high-reflectivity mirrors coating extends from 2,800 cm^{-1} to 3,300 cm^{-1}. The cavity-transmitted comb light is detected by a home-built Fourier-transform IR (FTIR) spectrometer with an autobalancing HgCdZnTe photodetector, referenced to a passively stable 1,064-nm continuous-wave laser. By employing the properties of the known instrument function (a sinc function) and setting the spectral resolution to half the comb spacing (136 MHz), the intensity of each individual comb line can be extracted. The resulting comb-mode-resolved spectra have an effective spectral resolution limited by the linewidth of the individual comb teeth ($\approx$ 50 kHz), rather than by the resolution of the spectrometer (4.5×10^{-3} cm^{-1}, which is our data-sampling frequency interval) (35). While the spectral bandwidth of the incident comb light covers more than 100 cm^{-1}, we limit only a narrow spectral portion of $\approx$ 10 cm^{-1} to be simultaneously coupled into the enhancement cavity, so as to significantly reduce the intensity noise (see *SI Appendix* for details). Spectra collected at different central wavelengths are concatenated

PHYSICS

to form a total spectrum with broader spectral coverage. All experimental spectra are collected at 50 Torr (67 mbar) and 20 °C. Note that the decrease in temperature of the exhaled gas during the breath collection (from body to room temperature) leads to the formation of water droplets inside the Tedlar bag, which could potentially dissolve some of the methanol present. Upon testing, we find that this does not significantly contribute to a systematic underestimation of the measured methanol concentration. Additionally, although Tedlar bags and Teflon tubing have both been chosen to minimize adsorption of polar molecules, these may adsorb to the walls of the stainless steel chamber, which could lead to an underestimation of the absolute methanol and water concentration.

Data Processing. Optical etalon fringes that interfere with the molecular absorption features are removed by notch filtering the experimental data in the postprocessing. The zero-absorption baseline is determined by fitting a cubic polynomial to match the upper envelope of the cavity-transmitted light spectrum. We note that some molecules (such as methanol) exhibit continuous absorption bands with spectral widths comparable to the spectral bandwidth of the cavity-transmitted comb light. In our experiment, these broad molecular absorption components are indistinguishable from the baseline of the light source and are removed during the baseline subtraction process. This results in an alteration of the molecular lineshape, with the extracted absorbance spectrum containing only the sharp molecular absorption features. The experimental absorbance spectrum is fitted by using the Beer–Lambert law with the varying effective interaction length, L_{eff}, considered in ref. 32. In the weak absorption limit, L_{eff} converges to $2FL/\pi$, where L is the cavity length and F is the wavelength-dependent cavity finesse (see *SI Appendix* for details on the cavity finesse). The HITRAN database (27) and HAPI code (28) are used to obtain the molecular absorption cross-sections and generate a simulated spectrum. Starting from an initial guess of the concentrations for the molecules known to be present in the sample, we calculate a simulated absorption spectrum ($A_{sim} = 1 - e^{-\alpha 2FL/\pi}$, where α is the total absorption coefficient) containing both the broad and sharp molecular absorption features. The broad absorption feature ($A_{broad} = A_{sim} - A_{sharp}$) is then removed by using the same baseline extraction method that is used for processing the experimental spectrum, and the resulting spectrum normalized ($A_{norm} = A_{sharp}/(1 - A_{broad})$). The normalized simulated absorbance spectrum is fitted to the normalized experimental absorbance spectrum by using a nonlinear least-squares method. Absolute molecular concentrations are calibrated by using a certified mixture of 1.04(5) ppm methane diluted in nitrogen from Matheson Tri-Gas (the company name is mentioned for technical communication only, and it does not represent an endorsement of the company). The detection sensitivity analysis is discussed in *SI Appendix*.

Data Availability. All data have been deposited in CU Scholar and can be accessed at https://doi.org/10.25810/HDXE-2M79 (36).

ACKNOWLEDGMENTS. We thank S. Diddams, T. Brown, M. Thorpe, D. Lesko, B. Bjork, and T. Bui for helpful discussions. This work is supported by Air Force Office of Scientific Research Grant 9FA9550-19-1-0148; NSF Grants CHE-1665 271 and PHY-1734 006; and the National Institute of Standards and Technology. J.T. was supported by the Lindemann Trust in the form of a Postdoctoral Fellowship.

1. B. Henderson *et al*., Laser spectroscopy for breath analysis: Towards clinical implementation. *Appl. Phys. B* **124**, 161 (2018).
2. C. Lourenço, C. Turner, Breath analysis in disease diagnosis: Methodological considerations and applications. *Metabolites* **4**, 465–498 (2014).
3. T. H. Risby, S. F. Solga, Current status of clinical breath analysis. *Appl. Phys. B* **85**, 421–426 (2006).
4. A. S. Modak, Barriers to overcome for transition of breath tests from research to routine clinical practice. *J. Breath Res.* **5**, 030202 (2011).
5. S. Das, M. Pal, Non-invasive monitoring of human health by exhaled breath analysis: A comprehensive review. *J. Electrochem. Soc.* **167**, 037562 (2020).
6. C. Wang, P. Sahay, Breath analysis using laser spectroscopic techniques: Breath biomarkers, spectral fingerprints, and detection limits. *Sensors* **9**, 8230–8262 (2009).
7. R. T. Jones *et al*., Could bio-detection dogs be used to limit the spread of COVID-19 by travellers? *J. Travel Med.* **27**, taaa131 (2020).
8. D. M. Ruszkiewicz *et al*., Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry—A feasibility study. *EClinicalMedicine* **29**, 100609 (2020).
9. B. Shan *et al*., Multiplexed nanomaterial-based sensor array for detection of COVID-19 in exhaled breath. *ACS Nano* **14**, 12125–12132 (2020).
10. C. Steppert, I. Steppert, W. Sterlacci, T. Bollinger, Rapid detection of SARS-CoV-2 infection by multicapillary column coupled ion mobility spectrometry (MCC-IMS) of breath. A proof of concept study. *J. Breath Res*, **15**, 027105 (2021).
11. D. Smith, P. Španěl, J. Herbig, J. Beauchamp, Mass spectrometry for real-time quantitative breath analysis. *J. Breath Res.* **8**, 027101 (2014).
12. Y. Y. Broza, H. Haick, Nanomaterial-based sensors for detection of disease by volatile organic compounds. *Nanomedicine* **8**, 785–806 (2013).
13. F. Adler, M. J. Thorpe, K. C. Cossel, J. Ye, Cavity-enhanced direct frequency comb spectroscopy: Technology and applications. *Annu. Rev. Anal. Chem.* **3**, 175–205 (2010).
14. P. Maslowski, K. C. Cossel, A. Foltynowicz, J. Ye, "Cavity-enhanced direct frequency comb spectroscopy" in *Cavity-Enhanced Spectroscopy and Sensing*, G. Gagliardi, H.-P. Loock, Eds. (Springer Berlin Heidelberg, Berlin, Germany, 2014), pp. 271–321.
15. P. B. Changala, B. Spaun, D. Patterson, J. M. Doyle, J. Ye, "Sensitivity and resolution in frequency comb spectroscopy of buffer gas cooled polyatomic molecules" in *Exploring the World with the Laser: Dedicated to Theodor Hänsch on his 75th Birthday*, D. Meschede, T. Udem, T. Esslinger, Eds. (Springer International Publishing, Cham, Switzerland, 2018), pp. 647–664.
16. M. L. Weichman *et al*., Broadband molecular spectroscopy with optical frequency combs. *J. Mol. Spectrosc.* **355**, 66–78 (2019).
17. M. J. Thorpe, J. Ye, Cavity-enhanced direct frequency comb spectroscopy. *Appl. Phys. B* **91**, 397–414 (2008).
18. A. Foltynowicz *et al*., Optical frequency comb spectroscopy. *Faraday Discuss.* **150**, 23–31, 113–160 (2011).
19. M. J. Thorpe, D. Balslev-Clausen, M. S. Kirchner, J. Ye, Cavity-enhanced optical frequency comb spectroscopy: Application to human breath analysis. *Opt. Express* **16**, 2387–2397 (2008).
20. G. D. Cole *et al*., High-performance near- and mid-infrared crystalline coatings. *Optica* **3**, 647–656 (2016).
21. A. Foltynowicz, P. Maslowski, A. J. Fleisher, B. J. Bjork, J. Ye, Cavity-enhanced optical frequency comb spectroscopy in the mid-infrared application to trace detection of hydrogen peroxide. *Appl. Phys. B* **110**, 163–175 (2013).
22. J. Ye, S. T. Cundiff, eds. *Femtosecond Optical Frequency Comb: Principle, Operation and Applications* (Springer US, New York, 2004).
23. A. Schliesser, N. Picqué, T. W. Hänsch, Mid-infrared frequency combs. *Nat. Photonics* **6**, 440–449 (2012).
24. T. Fortier, E. Baumann, 20 years of developments in optical frequency comb technology and applications. *Commun. Phys.* **2**, 153 (2019).
25. S. A. Diddams, K. Vahala, T. Udem, Optical frequency combs: Coherently uniting the electromagnetic spectrum. *Science* **369**, eaay3676 (2020).
26. F. Adler *et al*., Phase-stabilized, 1.5 W frequency comb at 2.8-4.8 microm. *Opt. Lett.* **34**, 1330–1332 (2009).
27. I. E. Gordon *et al*., The HITRAN2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **203**, 3–69 (2017).
28. R. V. Kochanov *et al*., HITRAN Application Programming Interface (HAPI): A comprehensive approach to working with spectroscopic data. *J. Quant. Spectrosc. Radiat. Transf.* **177**, 15–30 (2016).
29. A. Lagg, J. Taucher, A. Hansel, W. Lindinger, Applications of proton transfer reactions to gas analysis. *Int. J. Mass Spectrom. Ion Process.* **134**, 55–66 (1994).
30. J. Taucher, A. Lagg, A. Hansel, W. Vogel, W. Lindinger, Methanol in human breath. *Alcohol. Clin. Exp. Res.* **19**, 1147–1150 (1995).
31. F. Adler *et al*., Mid-infrared Fourier transform spectroscopy with a broadband frequency comb. *Opt. Express* **18**, 21861–21872 (2010).
32. A. Foltynowicz, T. Ban, P. Masłowski, F. Adler, J. Ye, Quantum-noise-limited optical frequency comb spectroscopy. *Phys. Rev. Lett.* **107**, 233002 (2011).
33. L. Ciaffoni *et al*., In-airway molecular flow sensing: A new technology for continuous, noninvasive monitoring of oxygen consumption in critical care. *Sci. Adv.* **2**, e1600560 (2016).
34. L. S. Petralia *et al*., Accurate real-time F_ENO expirograms using complementary optical sensors. *J. Breath Res.* **14**, 047102 (2020).
35. P. Maslowski *et al*., Surpassing the path-limited resolution of Fourier-transform spectrometry with frequency combs. *Phys. Rev. A* **93**, 021802 (2016).
36. Q. Liang *et al*., Dataset for Ultra-sensitive multi-species spectroscopic breath analysis for real-time health monitoring and diagnostics." CU Scholar. https://doi.org/10.25810/HDXE-2M79. Deposited 19 August 2021.