Appl. Phys. B 67, 379–390 (1998)

Applied Physics B
Lasers
and Optics

Gas sensing using terahertz time-domain spectroscopy

D.M. Mittleman[1], R.H. Jacobsen[2], R. Neelamani[1], R.G. Baraniuk[1], M.C. Nuss[3]

[1] Electrical and Computer Engineering Dept., MS-366, 6100 Main Street, Rice University, Houston, TX 77005, USA
(Fax: +1-713/524-5237, E-mail: daniel@rice.edu)
[2] Tele Danmark Udvikling, UPTA 11–190, Sletvej 30, 8310 Tranbjerg J, Denmark
[3] Bell Laboratories–Lucent Technologies, Room 4B-429, 101 Crawfords Corner Road, Holmdel, NJ 07733, USA

Received: 25 February 1998

Abstract. A method for detection and identification of polar gases and gas mixtures based on the technique of terahertz time-domain spectroscopy is presented. This relatively new technology promises to be the first portable far-infrared spectrometer, providing a means for real-time spectroscopic measurements over a broad bandwidth up to several THz. The measured time-domain waveforms can be efficiently parameterized using standard tools from signal processing, including procedures developed for speech recognition applications. These are generally more efficient than conventional methods based on Fourier analysis, and are easier to implement in a real-time sensing system. Preliminary results of real-time gas mixture analysis using a linear predictive coding algorithm are presented. A number of possible avenues for improved signal processing schemes are discussed. In particular, the utility of a wavelet-based signal analysis for tasks such as denoising is demonstrated.

Spectroscopic methods for the sensing and identification of gases have shown great promise, owing to their inherent non-invasive nature, relative simplicity, and high selectivity. The vast majority of the work in this area has relied on the “fingerprint” absorption in the mid-infrared ($\lambda \approx 2$–$20\,\mu$m), where molecular vibrations often provide a unique signature. Both incoherent (for example, Fourier-transform infrared, FTIR) [1] and laser-based coherent sources [2] have been commonly employed. These have demonstrated sensitive detection of many gases including greenhouse gases such as CO, CO_2, and CH_4, chemical etchants such as HCl and HF, and common smokestack pollutants such as SO_2 and N_2O. Despite these promising results, significant challenges remain, particularly in expanding the range of gases amenable to spectroscopic detection.

In contrast to mid-infrared gas sensing, the use of far-infrared or terahertz ($1\,\mathrm{THz} = 10^{12}\,\mathrm{Hz}$) radiation for sensing purposes is a field in relative infancy. In this frequency range, from $\lambda \approx 3$ mm to $50\,\mu$m (corresponding to frequencies between 0.1 and 6 THz), many polar molecules exhibit unique spectral signatures arising from transitions between rotational quantum levels. The use of these absorption signatures for detection or identification of gases is very much complementary to the more well-established mid-IR techniques, and will greatly expand the number of gas species that can be detected via laser-based methods. Yet, the development of THz sensing tools has been hindered in large part due to the lack of suitable radiation sources and detectors for use in the far-infrared.

Within the last decade, a number of new approaches to the generation and detection of THz radiation have been pursued with increasing interest. These techniques, based on frequency conversion using nonlinear optics [3–8], are often simpler, more reliable, and potentially much less expensive than the more traditional approaches such as molecular vapor lasers, free electron lasers, and synchrotrons. Although many of these nonlinear optical techniques were pioneered in the late 1960s or early 1970s [9], it is only relatively recently that it has become plausible to consider “real-world” applications based on these THz technologies. This excitement has been spurred in part by a number of important advances. First, a range of new fabrication techniques have been developed for nonlinear optical materials, such as the low-temperature growth of semiconductors and periodic poling. Second, there have been substantial improvements in the reliability, stability, size, and cost of the lasers required for these devices. The advances in the ultrafast pulsed laser systems are particularly notable, with the cumbersome argon-ion-pumped Kerr-lens-mode-locked Ti:sapphire lasers giving way to more compact and less costly all-solid-state systems [10] or to mode-locked fiber lasers [11].

In this paper, we present a review of recent results in automated real-time gas sensing involving one of the more promising of these new THz techniques, known as terahertz time-domain spectroscopy, or THz-TDS. The THz-TDS system relies on the use of femtosecond laser pulses for the generation and detection of THz radiation, and has been a beneficiary of the rapid progress in this enabling technology. Using THz-TDS, one can obtain a rapid spectral measurement over a very large bandwidth in the terahertz range, where many gases exhibit “fingerprint” absorption spectra arising from rotational quantum transitions, analogous to the familiar vibrational fingerprint region. These rotational line spectra can be used for gas identification and gas mixture analysis. Broadly speaking, the advantages of this technique are straightforward: relatively easy access to a previously unexploited spectral range will expand the number of gas species amenable to remote detection. Many gases may be detected more efficiently or accurately by relying on their relatively simple rotational spectra, rather than on their more complex vibrational or rovibrational signatures. Other advantages include the great flexibility afforded by the extremely broad bandwidth accessible with THz-TDS, and the coherent detection, which permits far-infrared spectroscopic measurements of plasmas, flames, and other challenging samples.

Here, we describe the THz-TDS technique, highlighting those aspects which are uniquely advantageous for the purposes of gas sensing. We focus particularly on how the nature of the measured signals lends itself to rapid and highly selective data analysis. We review recent results involving the implementation of a real-time gas sensing system, with emphasis on the adaptation of powerful signal processing protocols for gas and gas mixture analysis. Because the ultimate sensitivity limits for a given gas will depend largely on the effectiveness of the signal processing, we have concentrated our recent efforts on further development of these algorithms. We provide a brief discussion of the future directions of this research, noting in particular the efficacy of wavelet analyses of the THz-TDS signals for signal processing challenges such as deconvolution and denoising.

1 Terahertz time-domain spectroscopy

The THz-TDS system described here has been reviewed extensively in the recent literature [12–14]. With THz-TDS, one overcomes many of the difficulties traditionally associated with research in the far-infrared, such as the need for cryogenics for cooling detectors. Because of this, a great deal of interest has been generated by the potential for novel far-infrared spectroscopic measurements [15–18], and by the recent demonstrations of imaging using THz waves [19–21]. The necessity for a source of fs pulses has hindered the development of this THz source for applications such as gas sensing, but the recent progress mentioned above has minimized many of these concerns. With the fs laser system no longer a substantial impediment to commercialization, the development of a compact portable coherent terahertz spectrometer is currently under way.

A typical THz-TDS gas sensing setup is shown in Fig. 1. Pioneered in the late 1980s by groups at Bell Laboratories [3] and IBM's T.J. Watson Research Center [23], this system relies on an ultrafast laser system, producing a train of optical (usually near-infrared) pulses, of $\approx$ 100-fs (10^{-13} s) duration. This pulse train is split into two arms, one for the THz emitter and one for the detector. The emitter is usually a direct-gap semiconductor (typically GaAs) with some simple antenna structure lithographically defined on its surface. This high-impedance dipole emitter is dc-biased, typically with a few tens of volts. When a fs pulse excites the semiconducting material in the antenna gap with an above-band-gap excitation, the photogenerated charge carriers are accelerated by the applied bias field. The rise time of this photocurrent can be nearly as fast as the fs optical pulse duration. This transient current radiates according to Maxwell's equations, $E(t) \propto \partial j(t)/\partial t$. Because $j(t)$ is approximately a step function, this radiated field is close to a single-cycle pulse, with extremely large fractional bandwidth. The radiated field can also contain contributions from effects other than charge transport, such as optical rectification [4], although in the experiments described here this contribution is negligible.

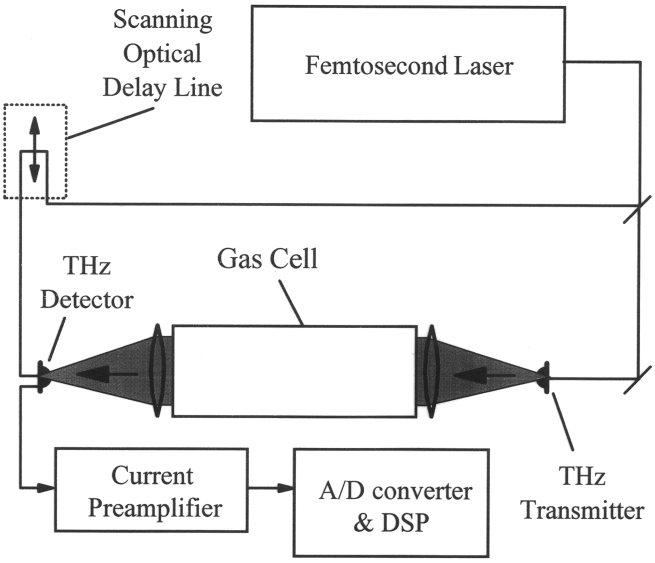

Fig. 1. Experimental setup for gas sensing measurements based on terahertz time-domain spectroscopy

When the radiation emerges from the rear surface of the emitter, it is collected, first with a substrate lens, then using conventional far-infrared optics. The resulting THz beam can be collimated and focused just as any laser beam, except that the long-wavelength radiation diffracts more readily than optical beams [23]. Even so, a well-collimated THz beam of $\approx$ 1-cm radius can propagate for several meters before diffraction becomes significant. This large beam diameter precludes the use of most commercially available multi-pass cells, such as those used in many mid-infrared gas sensing systems to increase optical path lengths. One could probably design a multi-pass arrangement appropriate for the THz beam, but for present purposes a single-pass cell of length $\approx$ 30 cm has been used (Fig. 1).

The detection of these THz transients is accomplished using the second arm of the fs beam. The traditional scheme relies on photoconductive sampling in an antenna similar to that used in the emission process. In this case, the THz pulse provides the bias field, which drives the photogenerated carriers in the antenna circuit. This photocurrent is sampled by varying the delay of the fs gate beam with respect to the THz beam, using an optical delay line. This measurement relies on the fact that the photogenerated carriers have a high mobility (because the induced current depends on the carrier mobility) and also a fast carrier lifetime (because a sampling measurement is only useful if the sampling window is fast compared to the transients being measured). These requirements place rather stringent demands on the semiconductor materials used in the detector. Currently, low-temperature-grown (LT) GaAs and radiation-damaged silicon-on-sapphire are the two most common choices. These materials have been used for spectroscopic measurements spanning bandwidths from below 0.1 THz to beyond 5 THz.

A typical THz pulse measured in this fashion is shown in Fig. 2. Here, curve (a) shows the measured photocurrent as a function of the delay between the two beams. This photocurrent is essentially proportional to the terahertz electric field, $E_{\mathrm{THz}}(t)$. On the bottom (curve (b)), the magnitude of the Fourier transform of $E_{\mathrm{THz}}(t)$ is shown. In this data, the antenna used for detection provided useful bandwidth from 100 GHz to ≈ 2.5 THz. The lower limit is determined by diffraction losses in the optical system, as well as the acceptance bandwidth of the receiver antenna. The upper limit is largely determined by the free-carrier lifetime in the semiconductor, which limits the minimum temporal width of the sampling gate. In principle, the bandwidth is also limited by the duration of the fs pulses used to gate the antenna, but the free-carrier lifetime is, in general, longer than the pulse duration and is therefore the more important limiting factor [24].

The speed at which these waveforms can be measured depends on several factors, including the speed of the A/D conversion, the required signal-to-noise ratio of the measurement, and the speed of the mechanical optical delay line. Using a scanning optical delay line (a retroreflector mounted on a galvanometric motor), these THz waveforms can be measured in only a few tens of milliseconds with signal-to-noise in excess of 10^3. The scanning delay line can be driven by a triangle wave, providing a linear variation of optical delay with time, at a rate of 10–100 Hz. This effectively down-converts the THz waveforms into the kHz range, where they can be processed with slow electronics. This fast data acquisition permits rapid analysis of the full spectral bandwidth spanned by the THz pulse, and thus real-time sensing and identification of gases with absorption signatures within the measured bandwidth. Further, because the measured waveforms are in the kHz range, they are well suited for processing with available digital signal processing algorithms. These algorithms are often more efficient than the more familiar frequency-domain analogs, and can provide a high degree of selectivity in gas identification. An example of an implementation of such a processing scheme is described below.

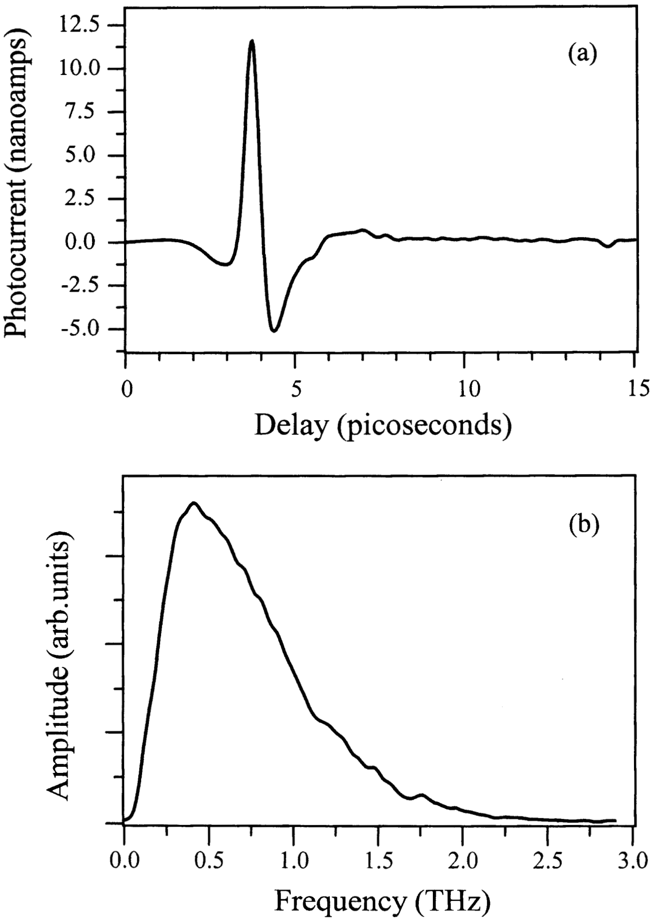

Fig. 2a,b. THz waveform measured using photoconductive sampling. **a** Average photocurrent (proportional to THz electric field) as a function of optical delay. **b** amplitude of Fourier transform of the signal in **a**, showing the spectral content of the radiation

Recently, a second detection scheme has been gaining in popularity. This newer technique is known as free-space electro-optic sampling [25]. Here, the THz beam co-propagates with a synchronized fs probe pulse in an electro-optically active material. The electric field of the THz pulse induces a polarization rotation of the fs probe beam, via the Pockel's effect. This rotation can be measured as a function of the delay between the THz pulse and the optical pulse, giving a direct measurement of the THz electric field. In contrast with the photoconductive technique outlined above, the electro-optic nonlinearity is nonresonant, and thus responds essentially instantaneously to the applied field. As a result, this technique can provide much more detection bandwidth, limited by the duration of the fs probe pulse. Using ZnTe as a detector, bandwidths as large as 37 THz have been reported [26]. Presently, the photoconductive sampling technique is more favorable for gas sensing applications, because of the ease of implementing real-time measurements. A fast scanning delay line can be used to acquire the THz waveforms with the electro-optic sampling technique, but the hardware requirements are somewhat more cumbersome [27]. Also, the use of fiber-optic-coupled detection antennas is substantially easier with the photoconductive technique, resulting in a system which is much less sensitive to alignment and external perturbation. In situations where waveform acquisition times of a few minutes are tolerable, the broader bandwidth available from electro-optic sampling will be quite valuable.

It should be emphasized that both of these detection techniques are coherent; they measure the THz electric field $E_{\mathrm{THz}}(t)$, not the intensity $|E_{\mathrm{THz}}(t)|^2$. As a result, the phase information is preserved, and one may determine both the real and imaginary parts of a sample's dielectric function simultaneously, without resorting to Kramers–Kronig analysis. This phase-sensitive detection has a further advantage in the near-complete rejection of incoherent radiation. Thus, these detectors generally operate at room temperature and in ambient light. This obviates the need for the cumbersome cryogenic apparatus typically required in far-infrared systems. Additionally, it permits spectroscopic and imaging studies of hot samples such as flames and plasmas, whose large thermal signatures would ordinarily swamp an incoherent detector [28]. An example of this is shown in Fig. 3, which is an image of a $\approx$ 1-cm gas flame [14]. The waveforms collected at each pixel of this image display no measurable variation in spectral amplitude. Instead, small changes in phase are detected, corresponding to variations in the transit time of the THz pulse through different regions of the flame. This

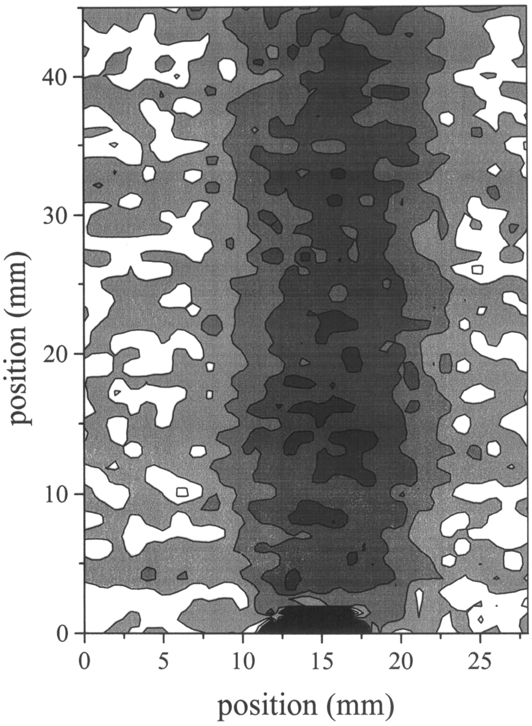

Fig. 3. THz image of a small gas flame, in which each contour corresponds to a shift in the transit time of the THz pulse of 5 fs. The pulse travels faster through the center of the flame because of the lower density (and thus lower index) of heated air

variation is due to the change in refractive index of the air which results from heating, and is displayed as gray scale contours. Adjacent contours correspond to shifts in the arrival time of only 5 fs, corresponding to changes in the phase of ≈ 0.015 radians. The ability to perform measurements such as this opens up many new possibilities in gas sensing, including the monitoring of gases in ovens or of plasmas, for example.

2 Gas sensing with terahertz time-domain spectroscopy

Although far-infrared spectroscopy of gases has been a vibrant area of research for many years [29], the use of active devices in the THz range for gas sensing and identification is only just beginning to be explored [30, 31]. The more mature technology based on mid-IR spectral signatures has demonstrated sensitive detection of many gases at the level of a few parts per billion, based on vibrational fingerprint spectroscopy [2]. However, larger molecules such as polar organic solvents and chlorofluorocarbons tend to exhibit extremely complex and congested line shapes in this spectral range. As a result, distinguishing an individual molecular species based purely on mid-infrared spectroscopy can be quite challenging. In the far-infrared regime, the line shapes tend to simplify considerably, since one is left with pure rotational spectra. In many cases, these spectra can be characterized by a small number of rotational energy parameters. Thus, characterization of more complex molecules is frequently easier with pure rotational spectra than it is when vibrational levels are also excited. A device that spans the rotational spectral region constitutes a powerful complement to established sensing technologies, expanding the number of gases amenable to spectroscopic identification. Here, the gas sensing methods based on THz-TDS are described in some detail, with particular focus on the important signal processing issues. The examples presented here are chosen, not because they represent insurmountable challenges with more well established technologies, but because they serve to illustrate the capabilities and limitations of the time-domain spectrometer.

In recent years, THz-TDS has been used as a tool for spectroscopic studies of a number of gases, including water vapor [32], methyl chloride [33], and nitrous oxide [34]. As with most polar molecules, these gases have characteristic fingerprint absorption spectra within the bandwidth of the far-infrared radiation. These spectra, which originate from transitions between rotational quantum levels, can be theoretically modeled quite accurately [35]. The narrow rotational absorption lines attenuate selected frequency components within the bandwidth of the THz beam as it propagates through the sample under study. The resulting waveform, measured in the time domain, exhibits small oscillatory features following the main pulse. Examples of this are depicted in Figs. 4 and 5, which show the waveforms that result when the THz pulse is passed through water vapor and acetonitrile (methyl cyanide, CH_3CN) vapor, respectively. In the case of water (Fig. 4) the irregularly spaced absorption lines produce a complex beat pattern following the initial pulse. These absorption lines are observed in the Fourier transform of the THz waveform as

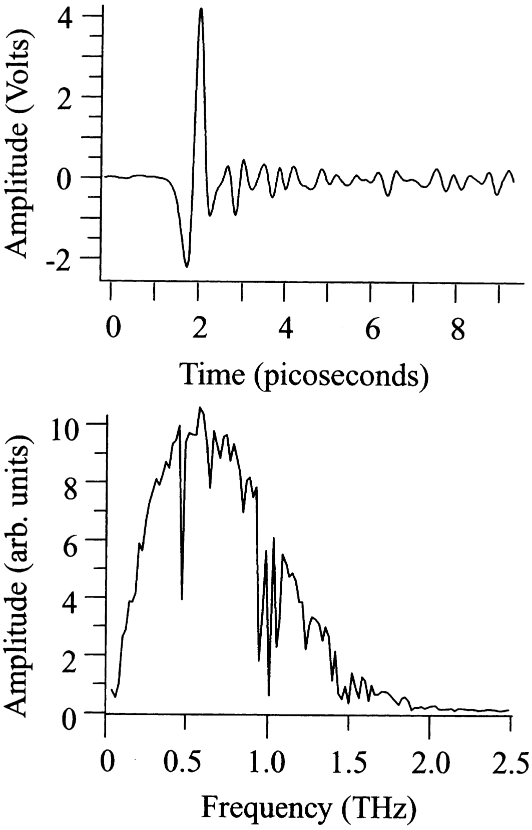

Fig. 4. THz waveform modified by propagation through humid air. The oscillations following the initial pulse result from the effects of water vapor. The H_2O absorption lines attenuate selected narrow frequency intervals within the spectrum of the THz pulse, as seen in the Fourier transform, at right

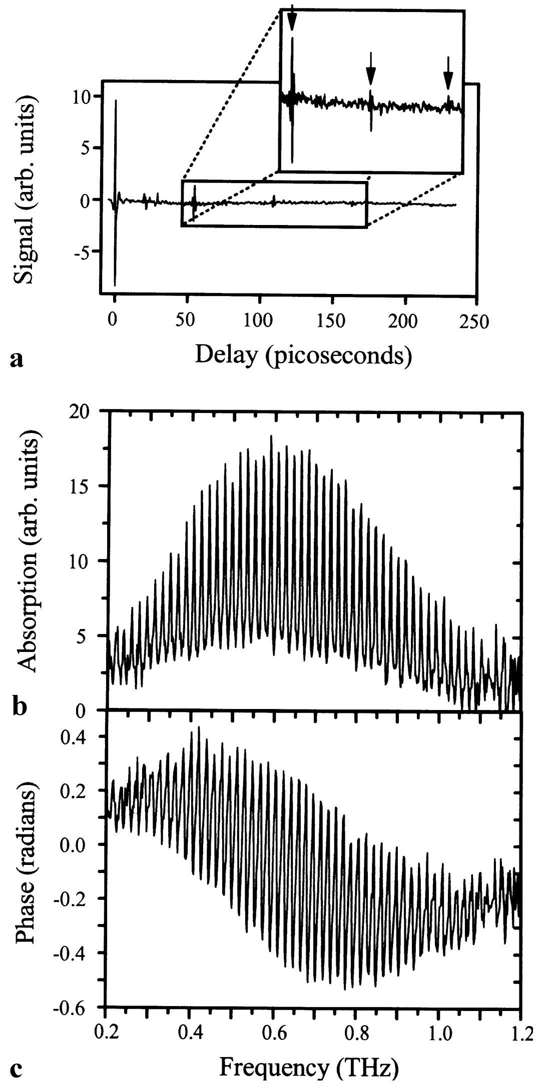

Fig. 5. a THz waveform modified by propagation through acetonitrile vapor, and the corresponding absorption (**b**) and dispersion (**c**) spectra. *Arrows* in **a** indicate the positions of the periodic echoes which result from the coherent excitation of the rotational manifold

sharp dips in the spectrum. In the case of the symmetric rotor acetonitrile, the regularly spaced rotational manifold generates a series of recurrences or echoes of the initial pulse [36], at temporal delays that are multiples of the inverse of the rotational line spacing (Fig. 5). In some cases, more than 25 such echoes can be observed, extending for more than 1 ns after the initial transient [33]. The absorption coefficient and dispersive phase can be independently derived from this time-domain data by Fourier transform, and are shown in Fig. 5b,c.

It should be noted that the linewidths of these rotational transitions are generally far narrower than the spectral resolution of the THz-TDS measurements, which is determined by the range of the temporal scan and is typically 1–10 GHz. A method has been reported for high-resolution (better than 1 MHz) gas phase spectroscopy using demodulation of a fs THz source, but the advantage of coherent detection is lost [37]. The coherent detection methods described above, although not well suited for detailed lineshape measurements due to the lack of sub-MHz resolution, are more than sufficient for distinguishing between different gas species in most cases, and are therefore adequate for many practical gas sensing applications.

Indeed, a great deal of information can be gained directly from the time-domain waveform, without the need for a Fourier transform. For example, in Fig. 5a, the observed delay Δt between the periodic recurrences in the time-domain waveform (indicated by arrows) is roughly 54 ps. Since the rotational line spacing is given, to a good approximation by $\Delta\nu = 2B(J+1)$, it follows that $B = (2\Delta t)^{-1} \approx 9.2$ GHz for acetonitrile, in good agreement with the literature value [38]. Values for the rotational constant derived in this fashion can be used to provide a rapid identification of the absorbing species. Given the nature of the THz-TDS signals, such time-domain methods for characterizing and identifying molecular species are more natural than frequency-domain techniques, and often more efficient as well.

The selectivity of this technique is demonstrated in Fig. 6, which shows the temporal waveforms obtained by passing the THz beam through a sample of acetonitrile (CH_3CN, curve a) and its fully deuterated analog (CD_3CN, curve b). Only the first ≈ 70 ps of each waveform is shown, so only the first recurrence is visible in each case. The difference between the two gases is quite dramatic, as the temporal delay of the first echo shifts from $\Delta t \approx 54$ ps in CH_3CN to 64 ps in CD_3CN. This reflects the decrease in the rotational constant of the molecule from ≈ 9.2 GHz to ≈ 7.8 GHz, resulting from the larger moment of inertia of the deuterated molecule [38]. As mentioned above (see Fig. 3), extremely small shifts in Δt (less than 10 fs) can be detected using THz-TDS. This is therefore a sensitive method for distinguishing among several closely related gas species.

Additionally, the use of THz spectroscopy for gas identification is greatly facilitated by the simplicity of the spectral signatures of these gases. Because of the periodicity of the rotational manifold, the THz spectra (and consequently the measured time-domain waveforms) can be parameterized by a single number. Although some information is neglected (for example, the envelope of the rotational manifold), that which remains (for example, the value of the rotational constant B) is often sufficient for gas identification. This is in contrast

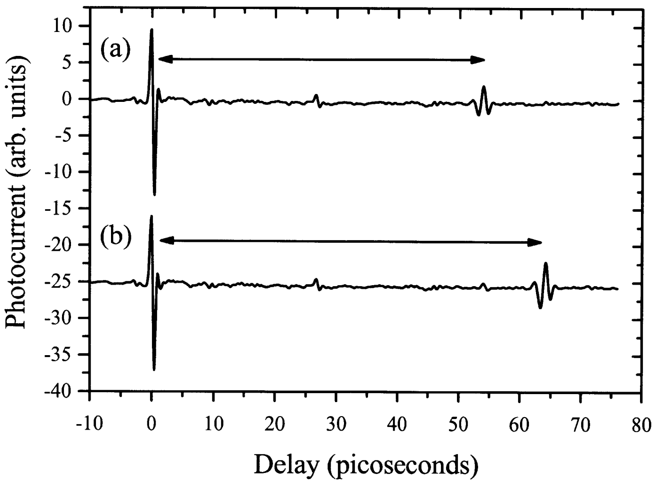

Fig. 6. Time-domain waveforms showing the effects of propagation through small amounts of gaseous CH_3CN (a) and CD_3CN (b). The shift in the delay of the first echo (indicated by *arrows*) results from the larger moment of inertia of the deuterated species. Waveforms offset for clarity

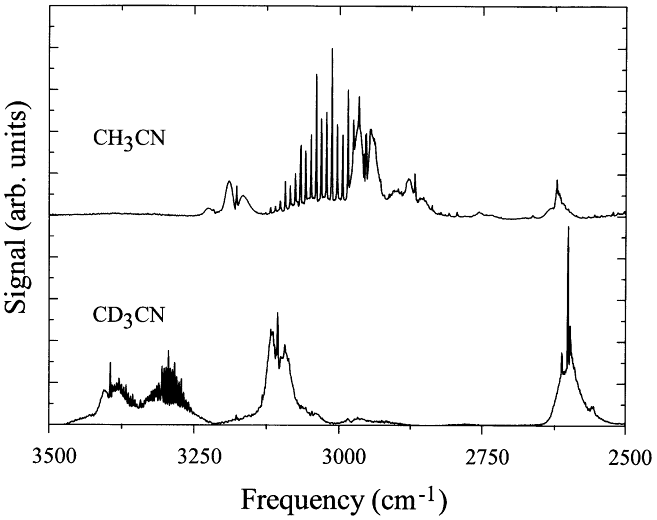

Fig. 7. Mid-infrared absorption spectra of CH_3CN and CD_3CN, measured with a commercial FTIR spectrometer. Spectra vertically offset for clarity

with many examples in sensing based on mid-infrared vibrational spectroscopy, where even the spectra of relatively simple molecules such as acetonitrile are quite complicated [39] and require detailed analysis for band assignment. Figure 7 illustrates this point, showing the mid-infrared spectra of the two gases of Fig. 6, obtained using a commercial FTIR spectrometer. These two spectra, although quite different from one another, are both rather complex, as a result of the interplay between different vibrational and rovibrational modes of the molecules. Evidently, distinguishing between these two species is far simpler and more amenable to automated analysis with the THz time-domain signals. One can imagine that a mixture of these two gas species would produce a very complicated superposition of the two spectra in Fig. 7, whereas the resulting THz waveform would simply contain two echoes, rather than one. The nature of these time-domain signals affords a great deal of flexibility in a gas sensing system, both with regard to the number and types of gases to be detected.

3 Linear predictive coding

The utility of techniques such as those outlined above will depend on the accuracy with which spectroscopic information can be extracted from the time-domain waveform, relative to the differences between the signatures of different gases. This will depend on the effectiveness of the signal processing algorithms used for data analysis. Indeed, the ultimate sensitivity limits of this technique will likely be determined by the degree to which one can extract the signature of a particular gas from a waveform containing a combination of complex time-varying signals. Because the necessary algorithms are under current development, it is premature to estimate sensitivity limits based on the data obtained to date. However, a discussion of the most recent results serves to illustrate the potential of this technology. These results, which rely on a well-known time-domain waveform analysis technique known as linear predictive coding (LPC), are described here.

The value of linear prediction in descriptions of time-domain digital signals which exhibit narrow spectral features is recognized in the field of automated speech recognition [40]. The analysis of a digitally sampled THz waveform $E(t)$ begins with the assumption that the action of the gas or gas mixture can be modeled as a time-invariant linear filter, of the form

$$E(t) = \sum_{k=0}^{M} a_k E(t - k\Delta) \,. \tag{1}$$

Here, Δ is the sampling interval. The coefficients a_k are the parameters of the linear prediction, and thus contain the information on the spectral content of the waveform. This can be seen by taking the discrete-time Fourier-transform of (1), to find the frequency response of the transfer function corresponding to this time-domain filter:

$$H(f) = \left(1 - \sum_{k=0}^{M} a_k \exp\left[-\mathrm{i}\,2\pi k f \Delta\right]\right)^{-1} \tag{2}$$

Equation (2) represents an infinite impulse response (IIR) filter; the coefficients a_k determine the poles of the filter function. These coefficients are determined directly from the signal waveforms using a correlation analysis. Here, one minimizes the squared sum of the difference between the measured waveform and that generated by the filtering process. This analysis yields a system of linear equations that relate the LPC coefficients to values of the autocorrelations of the sampled waveform. These equations can be solved using fast matrix inversion techniques such as the Levinson algorithm [40], providing a fast and simple means for generating an M-dimensional vector $\boldsymbol{a} = \{a_1, a_2, \ldots a_M\}$ that can be used to parameterize any measured waveform. It should be noted that, because these coefficients depend only on the autocorrelation of the time-domain waveform, all information involving the relative phases of the oscillatory signals is lost. This could pose a substantial problem in certain cases, as discussed below.

There are two important advantages of using LPC for the parameterization of THz waveforms for gas recognition. First, an expansion in a series of poles provides a parsimonious representation of a power spectrum with sharp spectral features. LPC methods are particularly well suited for extracting the frequencies and amplitudes of a collection of slowly decaying sinusoids from a superposed waveform. This is in contrast to a Fourier expansion, which can require a large number of coefficients to represent narrow spectral lines. Second, the number of coefficients used to represent each waveform, M, is the same for all gases, regardless of the number of absorption resonances. This greatly facilitates the comparison of different waveforms as it permits simple geometric representations. An equivalent treatment using Fourier analysis would necessitate some form of template matching, and would be more cumbersome to implement. On the other hand, the optimal choice of the order parameter M can be problematic, depending on the selection of gases to be sensed. This difficulty is discussed further below.

The data compression inherent in the LPC analysis can be illustrated by comparing the spectral estimation of the filter function of a given sample, using both an LPC analysis