In situ Quantitative THz Gas Analysis

John C. Pearson, Brian J. Drouin, Ken B. Cooper, and Robert A. Stachnik

Abstract— **The rotational spectrum of gas phase molecules offers the potential to be an extremely powerful tool for analytical chemistry. Every polar molecule features a unique rotational spectrum with a band origin at DC. The resolution of phase locked microwave techniques can easily be 1 part in 10^{10} while the Doppler limited line widths are one part in 10^6. In the absence of significant optical depth the intensities of Doppler limited lines are linearly proportional to the total column. For *in situ* applications a microwave source illuminates the gas and power is measured as a function of frequency allowing the comparatively narrow lines to be detected even when there is significant gain or power structure. The illuminating source can easily have a brightness temperature greater than 10^{14}K so that extremely weak features can be detected. The result is that rotational techniques facilitate extremely sensitive chemical identification and have the potential for very accurate total abundance determination. The development and calibration of ambient temperature *in situ* rotational spectroscopic systems operating between 0.1 and 0.6 THz for planetary science is presented. Optimization of this type of system requires some rather different considerations than are typical for space based remote sensing instruments. The system considerations, the calibration and potential applications of *in situ* spectroscopy are discussed.**

Index Terms—**Gas detectors, Spectroscopy, Signal Detection, Heterodyning**

I. INTRODUCTION

THE idea of employing microwave spectroscopy as an analytical chemistry tool was first extensively discussed in 1948 by Gordy [1]. An entire chapter of the classic 1955 text Microwave Spectroscopy by Townes and Schawlow [2] was devoted to the subject. With the exception of a massive, late 1960's, commercial, system produced by Hewlett Packard, microwave spectroscopy has yet to be commercially exploited or used in any real sense for *in situ* analytical chemistry. However, microwave spectroscopy has proven to be a valuable remote chemical analysis tool in Limb Sounding and Radio Astronomy applications. The reasons for the microwave spectroscopic technique not achieving its *in situ* chemical analysis potential are primarily due to some rather significant technological challenges in generating and detecting submillimeter radiation that had to be overcome. The rapid improvements in digital electronics and wide bandwidth submillimeter sources have now eliminated the last of these original technical obstacles and building small powerful gas analysis systems are now feasible.

J.C. Pearson, B.J. Drouin, K.B. Cooper and R. Stachnik are with the Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109. Phone: 818-354-6822; fax 818-393-2430; e-mail: John.C.Pearson@jpl.nasa.gov

Design and optimization of an *in situ* gas detector requires a different approach than passive remote sensing where the sensitivity and performance is driven by minimizing the effective T_{sys}. First, the *in situ* implementation requires detection of a small change in power on a high power carrier signal caused by a molecular absorption in the sample gas. The relevant figure of merit is then the smallest change in power, rather than the smallest power that can be detected. Second, an *in situ* design must be able to do quantitative analysis in an absolute sense, extracting the total column of absorbing gas and determine with the help of a pressure and temperature measurement the fractional abundance of the observed molecule. In the planetary science application a key aspect is determination of the isotopic ratios of simple molecules for derivation of information about the evolution of the body in question relative to the initial solar nebula. From a spectroscopic point of view the isotopes are distinct molecules the so quantitative analysis of both determines the isotope ratio.

In this paper we discuss the three aspects of in situ rotational gas spectroscopy that are most critical to a system design: an understanding of the "target" signatures as a function of conditions, the fundamental limits of system performance as well as what is required to detect and quantify, and the practical system considerations that can prevent an instrument from reaching theoretical limits.

II. MOLECULAR ABSORPTION

All molecules with different atoms have polar bonds. As long as there is not accidentally high symmetry such as in CO_2 (OCO) or benzene (C_6H_6 in a ring), all molecules with dissimilar atoms have a permanent dipole moment and exhibit a rotational spectrum. Unlike infrared and electronic bands, the rotational band of all molecules features an origin at zero in frequency. Therefore, every polar molecule features transitions which can be accessed with millimeter and submillimeter techniques. The great promise of microwave spectroscopy is that a single detection system can in theory detect and quantify any molecule that is polar in sufficient abundance to be detectable.

The critical element of spectroscopic detection is the line strength with emission and absorption being completely correlated through the Einstein A and B coefficients [3]. The integrated absorption intensity of any molecular gas transition at any wavelength is calculated from the following equation [4]:

$$I(T) = \int_0^\infty \alpha_v \, dv = \left(\frac{8\pi^3}{3hc}\right) v S \mu^2 \left[e^{-E''/kT} - e^{-E'/kT}\right] / Q. \quad (1)$$

Here α_ν is the frequency dependent absorption coefficient, h is Planck's constant, c is the speed of light, ν is the frequency, E'' and E' are the upper and lower state energies, respectively, k is Boltzmann's constant, T is the temperature, S is the geometrical projection of the transition moment, μ is the transition dipole moment and Q is the partition function. In rotational spectroscopy the transition dipole is the permanent electric dipole moment and the energy exponent is a few percent unless the temperature is very low. The transition line width of the absorption is determined by summing in frequency space the spontaneous emission line width (inverse of Einstein A coefficient), the Doppler line width and the collisional line width. Spontaneous emission line widths are proportional to the transition frequency cubed and are only a few kHz for the strongest submillimeter transitions [4]. The Doppler width is given by Equation (2), below, and is typically less than 1 MHz for frequencies less than 700 GHz. The pressure broadening width (typically 5-15 MHz/Torr) is linear in pressure, independent of wavelength, and is a function of quantum state and broadening gas. Pressure broadening can be ignored by operating only in the Doppler limit i.e. at very low pressure (<10 mTorr) gas. The Doppler half width at half maximum in MHz is [4]:

$$\Delta\nu_d = 1.17221 \times 10^{-6} \times \nu_{ab} \sqrt{\left(\frac{T}{300K}\right)\left(\frac{28}{m}\right)} \ , \qquad (2)$$

where ν_{ab} is transition frequency in MHz, T is the gas temperature and m is the mass in atomic units. Doppler limited transitions in the submillimeter are only a few 100 kHz wide, but increase linearly with frequency. In the Doppler limit, the line width is on the order of one part in 10^6 of the frequency. This is effectively a very high Q which makes it comparatively easy to avoid interference on the same scale from either electronic or optical effects. The other implication of the narrow line width is that Doppler limited spectroscopy in the millimeter and submillimeter with a modest fractional bandwidth will cover about a million line widths. Since pressure broadening in an unknown gas mixer is an intractable problem [2] staying in the Doppler limit is an obvious operational requirement for a practical system. The peak absorption [4] is given by:

$$\alpha = I(T) f / \Delta\nu_d \times 1.51149 cm^{-1} \ . \qquad (3)$$

Here $\Delta\nu$ is the Doppler half width at half maximum of a Doppler limited 10 mTorr pressure and f is the fraction of molecules of a given species. In the Doppler limit the peak line strength is linear with column density. Fractional concentration can be directly converted to a column using the ideal gas law where there are 3.9×10^{14} molecules/cm^2 in a 1 cm long column at a pressure of 10 mTorr at 300 K. There are several public rotational line catalogs http://spec.jpl.nasa.gov and http://www.astro.uni-koeln.de/site/vorhersagen/, which tabulate line strengths, *I(T)*, and provide some instruction and software (JPL) to do so for other molecules.

The power absorbed per unit length *l* of gas is given by:

$$P_x = P_0 e^{-\alpha d} \ . \qquad (4)$$

Where P_0 is the power applied to the absorbing gas and l is the path length of propagation in the medium. For the line to be detectable P_x must be larger than the receiver ΔP_{min} in the measurement period.

Molecules feature a number of transitions depending on their structure. In just about all cases there is a line every 2B or B+C where these are the inverse moments of inertia along the axes of rotation. In symmetric top and asymmetric top molecules there will be additional series due to projection (K quantum number) on the molecular axis. As a result, a system with a modest fractional bandwidth can detect at least several lines of all but the lightest molecules. In heavy molecules thousands of lines will likely be present in a wideband spectrum. The relative intensity and the line width are determined by Equations (1) and (2) adding to the confidence in any detection.

The last molecular consideration is picking a frequency band which is ultimately a compromise between light and heavy target molecules with light molecules like water being much stronger at high frequency while heavy molecules are better suited for lower frequencies. A very good compromise at room temperature is the 300 GHz range where the heavy molecules will be near the Boltzmann peak and the light ones, which are generally easier to detect, are usually still present. For planetary applications lower temperature suggest lower frequency, however the light molecules like HCl and ammonia are often the primary science targets and these require submillimeter detection.

III. Spectroscopic Detection

The ultimate figure of merit for an absorption spectrometer is the smallest change of power that can be detected. In general it is easy to detect power and compare the power level to the noise floor of the detector to determine a carrier-to-noise floor ratio (CNR). For small signals the detector noise dominates, but in the large signal limit, noise related to the source always plays a significant role. A previous paper discusses the fundamental detection limits and compares in situ to active remote (standoff) detection [5].

The subject of sensitivity limitations for absorption spectrometers where large signals are involved was first discussed by Townes and Geschwind [6] who pointed out on the basis of amplitude modulation that the theoretical minimum detectable change in power of a large signal is proportional to the geometric mean of the noise and the signal powers:

$$\Delta P_{min} \propto \sqrt{kT\Delta\nu P_{rx}} \ , \qquad (5)$$

where k is Boltzmann's constant, T is the receiver's physical noise temperature in the absence of signal power, $\Delta\nu$ is the detection bandwidth and P_{rx} is the received signal power. Historically, spectroscopic systems have been operated in the small signal power regime, so that the minimum detectable absorbed power is determined by non-source noise terms not

related to the source, such as the NEP of the detector. However, the improvements in sources at THz frequencies have resulted in the source noise being an essential consideration.

Practical detection systems require choices be made in the sort of detector employed. Historically the emphasis has been on achieving thermal background limited detection, however, this is not necessarily the best optimization for a large signal detection system.

Equation (5) states that the smallest detectable signal is at best proportional to the square root of the received power. The sensitivity limit can be rigorously derived from the Rician probability distribution of signal detection as has been done for radar detection by Levanon [7]. The result also agrees with radar cross section measurement theory discussed by Dybdal [8] and proportionality factor is of order unity:

$$\Delta P_{min} = \sqrt{2kT\Delta v P_{rx}} \quad . \qquad (6)$$

An intuitive justification is that any power detector squares the sum of a signal and a noise voltage, $P = (Vs+Vn)^2$, so that for large signals, the cross-term $2Vs \cdot Vn \propto \sqrt{Ps \cdot Pn}$ dominates the fluctuations of the measured power. In heterodyne detection, the system temperature, T_{sys}, includes quantum noise, intrinsic mixer noise (conversion loss), as well as the thermal background noise. As a result, T_{sys} is the relevant quantity in Equations (5-6) for all heterodyne detection of changes in transmitted power.

The minimum detectable power absorption of Equation (6) must also apply to the signals received by a direct detector. However, analysis of direct detectors is complicated by the fact that a direct detector's NEP almost always quantifies the dominant noise that is additive after the square law action, rather than multiplicative prior to squaring as in Equation (6). In the large signal limit, the internal NEP must eventually become irrelevant in the absence of saturation. Direct detectors also have quantum efficiency, γ, that is less than unity. The performance of direct detector systems have been discussed at length by many authors, however we will adopt the formalism derived by Brown [9]. The SNR at the detector output includes the background modulation term, the background seen by the detector, the shot noise, and the intrinsic NEP.

$$SNR^2 = \frac{P_{rx}^2}{\Delta P_{min}^2} = \frac{P_{rx}^2}{2P_{rx}kT\Delta v + 2(kT)^2\Delta v\Delta f + 2P_{rx}hv_c\Delta v + NEP^2\Delta v} \qquad (7)$$

Here Δf is the input bandwidth and Δv is the output bandwidth. In the limit $hv << kT$, the cross-term noise will always dominate the shot noise. The SNR with respect to the detector input can be derived by multiplying Equation 7 by the quantum efficiency.

It is instructive to compare direct detectors using Equation 7 with heterodyne detectors in an ideal case where $T_{sys} \sim T$ (no mixer conversion loss or shot noise) and unity quantum efficiency. Regardless of input bandwidth and NEP both detection schemes ultimately converge to the same value: $SNR = (CNR/2)^{1/2}$, where CNR is the carrier to thermal noise at

the detector input.

The figure of merit for a heterodyne detector is the transmit temperature at which the IF starts to compress relative to T_{sys}. In a direct detector it is generally better to be able to accept higher input power levels as long as Equation (6) is not dominated by the NEP term, then a better detector is advantageous. Detector selection ultimately requires an understanding of the available source power and actual detector properties such as saturation, NEP, and load curves.

An interesting consequence of the sensitivity being proportional to the square root of input power and detection (output) bandwidth is that the absolute absorption can be almost trivially determined. The real CNR is proportional to the number of counts between the unilluminated detector and the illuminated detector in the absence of gas and the SNR can be obtained by taking the appropriate square root. The square root of the number of counts of the absorption line counts relative to the CNR is the absolute fractional absorption of the line which can be used in equation 3 to derive the fractional abundance of the gas. A challenge is to assure that the detector is very close to linear and not dominated by some intrinsic noise (NEP). The signal-to-noise of a heterodyne detection system used to measure a change in power on a carrier is always proportional to the square root of input power as long as the IF is not compressed. Direct detectors have a signal to noise proportional to the square root of the change in counts, but the zero (noise floor) or a strong absorption will always sample the NEP limited regime making the calibration significantly more difficult.

It is useful to consider the sensitivity of a system in a 1 Hz bandwidth with a mixer that has a 10 dB conversion loss (T_{sys}=3000 K) and starts to compress with a signal level of 100 μW of RF (10 μW at the IF). In this case ΔP_{min}=2.88x10^{-12} Watts and the signal-to-noise is 3.5x10^7. After taking into account the square root of the input power level, it is hard to conceive of a system with a signal-to-noise better than 1x10^8 or, equivalently, a minimum detectable absorption of 1 part in 100 million. So for existing technology this can be considered an upper limit on the possible performance of the microwave portion of the system.

IV. DETECTION REQUIREMENTS

A. Calibration of Absorption

An analytical chemistry system must be able to determine the absolute abundance. The test requirements associated with determining the absolute absorption in a direct detector make it highly advantageous for *in situ* applications to choose heterodyne detection as does the general requirement for ambient temperature detection in planetary applications. In principle most of the issues subsequently discussed apply to both kinds of detection, however for simplicity heterodyne detection will be considered hereafter. In the heterodyne system measurement of an absolute concentration corresponds to a calibration test where the IF compression is carefully measured as a function of RF frequency at the IF frequency. Calibration requires a "dark" or "cold" input level for the

mixer which is just the pumped mixer looking at the ambient environment. This can be accomplished by simply turning off the transmitted signal, but it must be done periodically in frequency since T_{sys} will be a function of frequency. The "hot" level is the illuminated mixer with no gas in the cell. It should be noted that this can usually be well determined even with gas in the cell because the spectrum is rarely completely line confused.

The difference between the hot and cold is required for calibration but only a relative scale in arbitrary count units is needed. This is due to the fact that measurement of the absorption is effectively a % absorbed result and is normalized to such in the data reduction. The % absorbed is an absolute number that can be traced directly to the fractional abundance of the molecule in question. The accuracy of the abundance measurement can be determined by the noise (square root of hot counts) relative to the depth of the line plus the uncertainty in the depth measurement. In this case of a line absorbing 10% of the power would be dominated by the accuracy in determining the line depth while a signal-to-noise of 5 line would be dominated by the noise.

The other calibration considerations are that the path length of the absorbing column is known and that the temperature and pressure can be measured. Errors in these quantities correspond directly to errors in the determined abundance as can be seen from Equations (1-3).

B. Dynamic Range of Electronics

A general requirement for a heterodyne spectrometer is that its dynamic range be limited by the first mixer. The corollary to this requirement is the noise be preserved after the mixer IF amplifier combination and that no subsequent component significantly adds to the noise. Both are standard for heterodyne systems but the need to preserve a maximum dynamic range of 160 dB corresponding to the CNR must be considered carefully and is far from trivial.

Equation (6) states that the signal-to-noise is proportional to the square root of the input power. In the limit of a "dark" calibration pulse, or an optically thick absorption, the power at the detector is just its noise. Since counting from this "zero" is necessary for calibration the detection scheme must have enough dynamic range to count zero and full power on the same scale. In our limiting case of a signal-to-noise of 1×10^8 this corresponds to 160 dB, which is a significant problem for the ultimate detector in the system. Unfortunately, only HEMT amplifiers generally have this sort of dynamic range, while diodes, digitizers, and other sorts of analog detectors do not achieve the required performance. However, lock-in amplification theory provides the basis of a solution. The 160 dB corresponds to the final filtered 1 Hz bandwidth. The electric field sampled in a much larger bandwidth will have a correspondingly lower signal-to-noise. The noise will increase 10 dB with every decade of bandwidth digitized so a digitizer with an 80 dB spur free dynamic range would have to digitize a 100 MHz bandwidth to have the digitizer noise not be the limiting factor. Sampling faster does not help since the digitizer noise also increases with the reduced time. The

currently available state-of-the art 16 and 24 bit digitizers are close to this level of performance and improvements are constantly being made so that they will not likely be a future limitation. The signal can then be extracted from the noise with digital lock-in amplification techniques that narrow the bandwidth to the desired level by having the integration time define the detection bandwidth. Once sufficient bandwidth to have the signal-to-noise of the digitizer not be the limiting factor is digitized, then digital lock-in techniques can provide the bandwidth limitation and accumulate the necessary bits for the required accuracy. 160 dB corresponds to 54 bits so the digital lock-in should be a 64 bit implementation. Digitization and lock-in processing must be done with both the in phase and quadrature components due to the unknown and changing phase. The digitizer linearity is a calibration error and any fluctuations will appear as a system gain change. Any digitizer instability will be noise.

C. Lock-in Detection Approach

Microwave spectroscopy has traditionally utilized a modulation scheme to minimize the impact of 1/f noise in the detector and to minimize the impact of standing waves. Modulation is always less than unit efficiency so there is a sensitivity penalty when using it. As long as the mixer is not operated in a homodyne mode, 1/f will not be a problem in the conversion. The overall receiver system gain is subject to 1/f noise that will manifest itself as structure on the power spectrum (noise) and a calibration error (shift of zero). Phase or frequency modulation allows the phase to be detected, which is independent of the receiver gain. While this is advantageous, the phase in a heterodyne system is proportional to the path difference between the LO and signal path, and will in general depend on frequency as well as (to a lesser extent) the absorbing gas pressure. As a result, traditional lock-in amplifier type detection with a fixed phase shift will only work over a limited range of conditions and frequencies. The only solution is to capture a complete description of the electric field at the detector by detecting both in phase, I, and quadrature, Q. Early attempts at build heterodyne absorption spectrometers without IQ detection had major problems with phases and could only be used in narrow frequency bands. So detecting in phase and quadrature must be considered a detection requirement.

Modulation represents a trade-off between standing waves and system stability with the ultimate detection limit and calibration accuracy. Modulation may be advantageous in the final implementation, however the final minimum detectable concentration will be larger and there will be a calibration penalty due to the line width changing the modulation efficiency that will be particularly severe when lines are blended. In an ideal system modulation would not be necessary, however achieving this requires a number of aspects of the system to be carefully designed and tuned up. These are discussed in Section VI.

D. Other Considerations

The effective filtering in the detection scheme makes this

system sensitive to phase noise. All power that falls out of the detection band due to phase noise is effectively lost. For a high quality local oscillator and transmit signal, spectral purity only becomes a consideration at longer integration times where the bandwidth is effectively extremely narrow. In this case a phase reference (difference between LO and transmit) can be multiplied by the same factor and then be used in a second down conversion to cancel all phase noise within a bandwidth defined by path difference in two signal paths.

A variety of other standard microwave techniques such as the use of a correlation receiver can be used to suppress various sources of noise from other aspects of the system such as the local oscillator if necessary. However for planetary applications simplicity is preferred since mass, power and volume are extremely expensive.

V. GENERAL REQUIREMENTS

A. Source and Reference

A general requirement is that the frequency is sampled on a fine enough grid so that each absorption feature is covered by at least 10 points with 20 or more being better whenever it is desirable to determine the absorption line width. This corresponds to a step size of 10 to 50 kHz at the transmit frequency. Since the signal-to-noise is proportional to the square root of integration time per point, there is a direct trade-off between speed and sensitivity and ultimately accuracy. A "discovery mode" would sample more coarsely in frequency with shorter integration times while careful measurement of isotope ratios will need to be done with a finer frequency step. The strict square root in the integration time means that there is no advantage to stepping slowly relative to averaging more samples as long as the system overhead is negligible. Unfortunately synthesizers suitable for significant fractional bandwidths with the ~1kHz step size required often take a few 10's of milliseconds to switch frequencies. Much faster synthesizers can be constructed at the cost of additional power. Ultimately, digital electronics will enable direct synthesis of any arbitrary waveform with good phase noise and low power. Currently this approach requires more DC power than a slow synthesizer however it should eventually become an extremely attractive option.

The fundamental synthesizer must cover the same fractional bandwidth as the resultant system and have a step size small enough to cover a molecular line with at least 20 points. Software control can degrade the resolution as desired. Faster switching is always desirable since this affords flexibility in the data collection, however this can be traded with system stability. A very stable system will follow the square root of integration time independent of the data collecting strategy, whereas in an unstable system averaging many fast frequency scans is significantly better.

Molecular line frequencies are constants of nature and cannot change. As a result, they define a frequency to the accuracy of their measurement. A trivial peak picking algorithm will determine line center to better than 10% of the line width. A more sophisticated algorithm, such as fitting the

base line structure and line to a dispersion and a Gaussian line shape, will return the center frequency to better than 1% of the line width. Historically, microwave spectroscopy has utilized a highly accurate reference frequency to derive the absolute frequency of the signal absorbed by the molecules. However, the reference frequency can be calibrated to extremely high precision by making measurements of lines as long as it is stable in time. It is possible to implement a spectrometer without a locked reference or even without an electronic reference and still achieve frequency accuracy of better than 1 part in 10^7 [10]. However when unlocked, calibration becomes significantly more numerically intensive and signal averaging cannot be done without first calibrating and then re-gridding prior to averaging. For planetary application a frequency accuracy of about 1 part in 10^7 is sufficient and well in range of capability of temperature controlled crystal oscillators. Higher accuracy is better but only if the reference quality is free in terms of other system constraints.

The other major constraint on the synthesizer driving both the LO and the transmitter is that it does not blind the receiver with noise at the IF frequency. Generally this is not a problem as long as signal levels are maintained with significant margin above the noise. The dynamic range of a 10 dBm signal relative to thermal noise is 184 dBc/Hz. A 10 dB noise figure at the first step of amplification would reduce this to 174 dBc/Hz, which can in principle be preserved until the multiplication where the signal-to-noise degrades at least as fast as the conversion loss and in the case of other signals as $20 \log(N)$ where N is the harmonic number. For the mixer to limit the system dynamic range, the carrier-to-noise ratio in the local oscillator and transmit signals must be better than in the mixer.

Several practical approaches can be applied. First is to saturate all amplifiers to effectively compress the amplifier noise. Second is to maintain large signal levels, never getting below a few dBm. Practically this means making sure the noise figure of the entire amplification chain is determined by the first amplifier.

B. Gas Handling

Gas handling is by definition separate from the problem of detection aside from the requirement that the sample must have at least the minimum absorption to be detected. The strength of the absorption is linearly proportional to the path length in the absorption cell. As a result longer sample cells allow for detection of lower concentrations. Unfortunately there is always a standing wave between the transmitter and the receiver doing the down conversion. This standing wave causes several significant problems. Most importantly it causes some fraction of the signal power to see a longer path length resulting in a direct error in the amplitude. Second it has a frequency dependent phase which results in an apparent distortion in power (a reason to modulate). If the path length is such that a standing wave period becomes close to the line width the system sensitivity is enormously degraded and no easy signal processing can retrieve the lost information. This results in a rule of thumb that the path length should at the

longest result in a standing wave period 10% of the line width. The worst case is at high frequency for a light molecule in a hot system. For example, NH_3 has a mass of 17 and a transition frequency of 572 GHz at 370 K, and it would have a full width at half maximum of 2 MHz. Applying the factor of 10 the system could tolerate a standing wave period of 20 MHz corresponding to a maximum tolerable path length of 15 meters. Longer is possible at lower frequencies and temperatures where the line widths are narrower.

Several gas collection strategies exist. The simplest is just to leak in the required amount of gas and then pump it out once the spectrum has been collected. If sensitivity is a problem a sorbate can be used to collect specific classes of molecules and the target molecules desorbed during filling. This makes it more challenging to determine absolute concentrations but it identifies the trace constituents. There are several considerations for the gas handling. Pumping is expensive in terms of power, and the smallest low power pumps have a finite lifetime. Thus, pumping should only be done when needed, requiring the absorption cell to be very vacuum tight. Another consideration is a strategy to keep previous fills of the cell from contaminating the current fill. Bake out heaters are effective in decontamination as is pure purge gas and pumping time.

C. Resources

The planetary application requires low power and very compact and light weight electronics as well as small volumes. These constraints are lifted in many other analytical chemistry applications. Limited data rates suggest that the ability to signal average and transmit only the resulting data is desirable. The other constraints push the design to use a phase locked oscillator and to focus on obtaining sufficient gain stability to not require rapid frequency switching.

Another obvious power saving consideration is to turn off the RF and detection electronics while gas handling operations are taking place. This will have the effect of averaging the power over the operation and will limit the maximum power. Realistically a complete *in situ* system could be implemented in less than 10 kg and would require less than 20 Watts. The details of the accommodation and other mission driven constraints could potentially have major implications to these values making the overall system architecture the single most important feature in saving mass and power.

D. Mixer and IF Amplifier

The best mixer is the one with the largest range between T system and T saturation, generally leading to a higher LO power implementation in a balanced configuration. Here the LO power requirement is increased but so is the compression point of the IF, while the noise temperature will be largely unaffected. Fundamental mixers generally have higher power levels at the right frequency so the IF compression sets in later as well as featuring slightly better noise temperatures.

The IF amplifier needs to comply with the usual receiver requirement of not dominating the system noise, but in this case the IF amplifier must also not compress on a macroscopic signal that can approach -15 dBm. As a result, linearity rather than ultimate noise figure is the more important selection consideration.

VI. PRACTICAL CONSIDERATIONS

A. Stability

Several important practical considerations must be made in optimizing the overall system. These include the choice of modulation, the averaging strategy and minimization of standing waves. The optimal choices depend on the RF system gain stability and the stability of the optical alignment. Dealing with the RF system gain stability is more challenging because gain changes can originate in local oscillator power level fluctuations, the mixer conversion efficiency, the transmitted power, the IF amplification gain or drifts in the digitizer. Like other radiometers temperature fluctuations are the major contributor to gain instability. If the temperature cannot be regulated or is expected to be highly variable, faster frequency sweeping and more regular transmit off pulses will be required for calibration. Sweeping the frequency fast enough will always make the impact of drifts on the scale of a line negligible even for highly unstable operational conditions. As a result, a very high quality system can be built even for conditions unsuitable for a radiometer, but this places a significant demand on the synthesizer performance. Any sweep rate that covers a line width at a frequency beyond the 1/f knee of the gain stability will achieve the desired end. Covering a line width at a frequency within the 1/f knee will degrade the performance. Allen times at representative operating conditions can be used to determine the 1/f knee.

B. LO and Transmit Hardware

A practical consideration is to utilize identical local oscillator and transmit hardware. This has a distinct cost and development complexity advantage, but it requires something close to -20 dB of the transmit signal not make it to the mixer to avoid saturation. The need for the loss significantly relaxes a number of optical throughput requirements and probably necessitates including additional optical loss in the system. The simplest implementation, unthinkable in other types of receivers, is to put an attenuator in front of the mixer signal port. The ideal implementation is to distribute the loss to minimize the standing waves in the system.

The challenge for the LO and transmit hardware is to have a relatively constant power over frequency. This will minimize the need to actively control bias parameters during measurements and will greatly simplify the control software. An obvious compromise can be made in accepting some LO power starvation at the mixer, since the system performance is only a weak function of T_{sys} as long as the IF compression is not dramatically reduced.

C. Optical Alignment

A second consideration is to assure that temperature does not change the alignment significantly and that the standing waves are at a small level. Several elements are highly desirable, first is to have separate RF and LO ports on the

mixer so that the LO can be injected through a waveguide eliminating LO alignment as a potential problem. The second consideration is to fabricate the entire optical system out of the same material, including the feed horns, sample cell and mirrors. If these are all hard mounted to a mechanical structure of the same material then alignment will be well preserved over temperature and there will be no dramatic changes in standing waves over temperature.

D. Optics

The excess transmit power affords many luxuries in the optical design. As such the critical aspect of the design must be a focused on minimization of standing waves over the entire frequency rather than optical throughput. Optical throughput should be traded for minimal standing waves in all cases. This means liberal use of aperture stops, not using all the mirror surfaces and introducing loss where advantageous. Unlike receiver signal paths which must use reflective optics, lenses in transmission may be used with no penalty as long as the surfaces do not cause standing waves. Since the RF system covers a significant fractional bandwidth it is highly desirable to design the optical system to be frequency independent. This requires designing the optics to take into account the constant radius of phase curvature in the feed horns rather than the usual beam properties which are frequency dependent. The bandwidth makes it difficult to use quarter wave plates, but the polarized nature of the source and detector make it possible to use the polarization as a means of standing wave control. The mechanical design should allow for optimization of standing waves before it is locked in place.

VII. PLANETARY SYSTEM

The general design philosophy is to do as much of the signal processing as possible at frequencies below 20 GHz where high quality commercial components are available. The feasibility of this approach has been demonstrated in THz astronomy by the Heterodyne Instrument for Far Infrared (HIFI) [11] and in THz laboratory spectroscopy [12].

The basic building block for a planetary system would be to have two frequency channels implemented in a cross polarized configuration and launched into a common cell with a polarizing grid. The detection end would be accomplished in an identical way. The high frequency channel would cover the 530-630 GHz range while the low frequency channel would cover 225-330 GHz. The high frequency channel would be turned off when the common source is out of its frequency range. Software control would allow one or both channels to be operated simultaneously. A reference channel comprising the difference frequency multiplied by the correct harmonic number would be generated in common with the high frequency channel simply being the low frequency channel multiplied by two.

A synthesizer in the 12.5 to 18.4 GHz range with a 1 kHz step and a <1 ms switching time would drive the LO. The transmit signal would be derived by mixing the LO signal with a fixed oscillator used for the down conversion as well. It may prove advantageous to use the digital clock frequency for this purpose. The fundamental mixer would down convert to a MHz base band at approximately ¼ of the digitizer frequency, i.e. 32 MHz for a 16 bit 130 M sample per second digitizer collecting both the real and imaginary part over its Nyquist limited bandwidth.

From this point the signal would be digitally lock-in amplified increasing the number of significant bits to the desired level. One result for each sample point would be stored at the output. The scan strategy would be to integrate for 10 ms per point and signal average to the desired level. The gas cell could be a 5 meter long folded path avoiding beam overlap in a welded aluminum box with one wedged window for the beam entrance and exit. The total volume of the cell would be about 2 liters. Pumping would be accomplished with redundant micro turbo pumps similar to those designed for the Sample Analysis at Mars instrument suite on the Mars Science Lander. A molecular drag pump could serve as the backing pump.

One electronics board would provide the conditioned power, a second board would control all the bias lines, valves and pumps while the third would analyze the data and log the results as well as communicate with the rest of the instrument system.

An expected sensitivity calculation can be done assuming a mixer IF compression level of 5 μW at 300 GHz and 2 μW at 600 GHz and a conversion loss of 10 dB at 300 GHz and 13 dB at 600 GHz. Assuming a factor of two loss for standing wave, and instability and a 90% integration efficiency the signal-to-noise ratio at 300 GHz would be 1.1×10^6 in the 10 ms integration and 7×10^5 at 600 GHz. Assuming a 1 kHz step at the synthesizer, the step size at 300 GHz would be 18 kHz and 36 kHz at 600 GHz. The 300 GHz channel would cover 5.9 GHz per hour while the 600 GHz system would double this. In 10 ms at 300 K, HCN at 265.8 GHz could be detected at 4.5 parts per billion of the 10 mTorr sample. NH_3 at 572 GHz could be detected under the same conditions at 51 parts per billion. The detection sensitivity will improve at the square root of integration time. All of this is prior to any attempt to concentrate gas constituents or signal average.

VIII. CONCLUSIONS

Sophisticated and extremely sensitive gas analysis systems can now be constructed and calibrated thanks to a number of THz developments originally for heterodyne local oscillator development. Application of submillimeter technology promises to make a new generation of analytical chemistry instruments possible for planetary exploration as well as for applications in terrestrial chemical detection and identification. The chemical specificity is unparalleled due to the resolution of the rotational spectroscopic technique. Additionally, many more orders of magnitude in sensitivity can be obtained by using concentration techniques, potentially making this a valuable tool in civil defense and environmental regulation enforcement.

REFERENCES

[1] W. Gordy, "Microwave Spectroscopy," Reviews of Modern Physics 20, 1948, pp. 668-717.

[2] C.H Townes and A.L. Schawlow, *Microwave Spectroscopy*, New York, McGraw-Hill, 1955.

[3] L.I. Schiff, *Quantum Mechanics3rd Ed.*, New York, McGraw-Hill, 1949.

[4] W. Gordy and R.L. Cook, *Microwave Molecular Spectroscopy*, New York, John Wiley and Sons, 1984.

[5] J.C. Pearson, K. Cooper, and B.J. Drouin, "Spectroscopic Detection, Fundamental Limits and System Considerations," IEEE MTTs Proceedings of IRMMW-THz 2008, 2008, paper M3B2.

[6] C.H. Townes and S. Geschwind "Limiting Sensitivity of a Microwave Spectrometer" J. Appl. Phys. 19, 1948, pp. 795-796

[7] A. Levanon, *Radar Principles*. Toronto: Wiley-Interscience, 1988.

[8] R.B. Dybdal, "Radar cross section measurements," Proc. IEEE 75 (4), 1987, pp. 498-516.

[9] E.R. Brown, "Fundamentals of Terrestrial Millimeter-Wave and THz Remote Sensing," in Terahertz Sensing Technology Volume II: Emerging Scientific Applications and Novel Device Concepts, D. L. Woolard, W.R. Loerop, and M. S. Shur, Eds, Singapore: World Sci., 2003.

[10] D.T. Petkie, T.M. Goyette, R.P.A. Bettens, S.P. Belov, S. Albert, P. Helminger and F.C. De Lucia, "A Fast Scan Submillimeter Technique," Rev. Sci. Instrument. 68, 1997, pp1675-1683.

[11] J. Pearson, R. Guesten, T. Klein, and N. D. Whyborn, "Local Oscillator System for the Heterodyne Instrument for FIRST (HIFI)," *Proceedings of SPIE* **4013**, 2000, pp. 264-274.

[12] B.J. Drouin, F.W. Maiwald, and J.C. Pearson, "Application of Cascaded Frequency Multiplication to Microwave Spectroscopy," Rev. Sci. Instrum. 76, 2005, 093113.