Application of Highly Sensitive Spectroscopic Techniques to the Study of Intramolecular Dynamics and Remote Sensing

by

Ilia A. Dubinsky

M.S. Physics. Moscow Institute of Physics and Technology, 1991
M.S. Chemistry. The Catholic University of America, 1994

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1998

Signature of Author_____
Department of Chemistry
December 19, 1997

Certified by_____
Jeffrey I. Steinfeld
Professor of Chemistry, Thesis Supervisor

Certified by_____
Robert W. Field
Professor of Chemistry, Thesis Supervisor

Accepted by_____
Dietmar Seyferth
Chairman, Department Committee on Graduate Students

Application of Highly Sensitive Spectroscopic Techniques to the Study of Intramolecular Dynamics and Remote Sensing

by

Ilia A. Dubinsky

M.S. Physics. Moscow Institute of Physics and Technology, 1991
M.S. Chemistry. The Catholic University of America, 1994

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1998

Signature of Author_____
Department of Chemistry
December 19, 1997

Certified by_____
Jeffrey I. Steinfeld
Professor of Chemistry, Thesis Supervisor

Certified by_____
Robert W. Field
Professor of Chemistry, Thesis Supervisor

Accepted by_____
Dietmar Seyferth
Chairman, Department Committee on Graduate Students

This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Professor Keith A.
Nelson_____
Chairperson

Professor Robert W. Field_____
Thesis Supervisor

Professor Jeffrey I. Steinfeld_____ _____
Thesis Supervisor

Professor Mario J. Molina_____

To my family

Application of Highly Sensitive Spectroscopic Techniques to the Study of Intramolecular Dynamics and Remote Sensing

by Ilia A. Dubinsky

Submitted to the Department of Chemistry on January 8, 1998
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemistry

ABSTRACT

A highly sensitive Pulse Amplified Frequency Modulation based spectroscopic technique for remote sensing applications (FMRS) is developed. A detailed consideration of noise sources in the experiment is followed by a sensitivity analysis for the technique and comparison with the existing absorption and fluorescence detection-based remote sensing techniques. A model, describing FMRS signal in the case when light falling on the detector is scattered by a large diffuse scatterer, is presented. Estimates of sensitivity and signal levels, expected in realistic remote sensing environments, indicate feasibility of implementation of the technique in the nearest future.

Isocyanogen (CNCN) was synthesized by flash pyrolysis of N-cyano-2,3-diphenylcyclopropeneimine. Modification of the precursor synthesis, pyrolysis process, and apparatus and purification procedures are described in detail. The results of the study of CNCN using a diffraction grating spectrometer are reported. Although the sensitivity of this method was insufficient for detection of isocyanogen, the appearance of cyanogen in the spectrum suggests isomerization of CNCN on the excited state potential energy surface as a result of excitation by the broadband UV source (hydrogen discharge). The spectroscopy of the $\tilde{A}\,^1A''$ state of isocyanogen is discussed and simulations of the expected spectra are reported.

A Cavity Ring Down setup, constructed for conducting spectroscopic studies deep in the ultraviolet, is described. Improvements of the experimental technique, allowing for measurements in spectral regions where high reflectance mirrors are not available, with sensitivities comparable to those achieved when very high reflectivity "super" mirrors are used, are reported. The nature and origin of the intensity noise (previously assumed to originate in energy redistribution among the transverse cavity modes) is explained. A simple and elegant way to eliminate this important source of noise is given. Theoretical and practical issues of sensitivity of CRD are discussed. The design of a specialized CRD cell, constructed for the isocyanogen experiment and containing no metal parts, is presented.

Thesis Supervisor: Jeffrey I. Steinfeld
Title: Professor of Chemistry

Thesis Supervisor: Robert W. Field
Title: Professor of Chemistry

Acknowledgments

Looking back at all the years I spent in graduate school, here at MIT, in the Catholic University, and back in Russia in Moscow Institute of Physics and Technology, I remember so many people who provided me with help, guidance and friendship, that merely listing their names here would make this thesis a hundred pages thicker. I am deeply grateful to all. There are certain people who must be acknowledged individually.

I am truly grateful to my advisors, Bob Field and Jeff Steinfeld, for having me as a graduate student, for teaching and guiding me through my time at MIT. I appreciate the freedom you gave me in choosing the topics of my research and understanding and sympathy you showed toward all my ideas and projects (scientific or not), however wild or stupid. The value of your support in these finishing months is impossible to overestimate.

I would like to thank Jon Bloch, who introduced me to the world of lasers, optics, and FM spectroscopy, and greatly influenced my research. A large part of this thesis grew out of our endless conversations.

I would like to thank Leah Ruslen, a wonderful friend. I have learned very much from her: about science and about life. Finishing this work would have been much more difficult without her hard work and input. I wish you the very best in your years at MIT and beyond, and look forward to your continuing friendship and collaboration.

James Janni deserves very special thanks. He and Leah kept me sane through many of my failures. James was always ready to help (to the point that he was taking apart and reconfiguring his setup to take an FTIR spectrum for me).

Brian Gilbert kept amazing me by the depth of his knowledge as much as by his optimism and other cheerleading qualities. Brian also introduced me to squash, and I still hope to beat him someday. Becky Janni and Dayna Gilbert will always be remembered for sharing warm moments with me.

Sergey Panov played a very unique role in my life. He is, more than anybody , responsible for my coming to the USA to a graduate school, getting a green card and meeting two very special people - late David Gutman and Sasha Kachanov. Sergey likes to do good things for no reason, and his help was instrumental in completing large parts of this thesis. I would also like to thank him and Nastja and Petja for wonderful time I had when visiting them.

David Gutman was the first American I met and I could not have had a better luck. He was a very good friend and teacher, a truly special person. He took deep interest in lives of many of us and will always be remembered and missed. In difficult times I am always looking up to you for strength.

Alexander (Sasha) Kachanov has always been an inspiration to me. I think it is fair to say that he is the most talented experimentalist with whom I've had the privilege of working. His incredible knowledge, energy and determination made CRD part possible. I am looking forward to further collaborations with him.

Mike Fickes introduced me to a field of synthetic chemistry and was enormously patient deciphering and teaching the synthesis of N-cyano-2,3-diphenyl-

cyclopropeneimine to me. Without him, isocyanogen part of this thesis would not have been done.

Steve Coy has always been a great person to talk with and to try ideas on, and to extract bits of knowledge from.

Dave Moss played a very important role in my understanding of FM remote sensing. His logic, knowledge, and clarity of thinking made me, finally, understand it.

Jon O'Brian was willing to give a helping hand even after he graduated and left MIT. I liked talking to Jon about science, life and life in science.

I was very fortunate to work with a very talented and driven undergraduate, Krzysztof Rybak. Beyond his natural gifts, he is a very motivated and hard working person. His help was very important in completing remote sensing project and working with him was always a pleasure. We found more common interests outside the work, and I only regret that our collaboration started so recently, is finishing so soon, and hope for more work together in the future.

Johnny, Ed and Bob helped a lot in putting together my numerous projects.

Pat Vaccaro helped us to start isocyanogen research by providing us with an outline of its synthesis.

Jan Hall and Ed Eyler taught me a lot during my short visits to their labs.

This acknowledgment will not be complete if I do not mention Peter Giunta, whose friendship and help I will always remember. Charlotte Russ was very helpful untangling our complicated dealings with the laser companies.

Kevin, Michelle, Stephanie, Matt Gardner, Steve Drucker, Jason, Matt Jacobson, Selen - many thanks to you all and best of luck!

I thank Bernd Abel for a wonderful time I had in Goettingen, for all the talks about science we had, and simply for friendship.

I thank Carter Kittrell for being a great host during my visit to Houston.

Above all, I thank my wife, Luda, and the rest of my family for love and never-ending support. You have always believed in me and without your understanding, encouragement, and enthusiasm this thesis could have never been finished. To you I dedicate this thesis.

Contents

1 Introduction

The main part of this dissertation is centered around the development and improvement of two highly sensitive spectroscopic detection methods, which are based on Pulse-amplified Frequency Modulation Spectroscopy (PFMS) and Cavity Ring-Down Spectroscopy (CRD).

The first method, Frequency Modulation facilitated Remote Sensing (FMRS), grew out of our previous work on the development of Pulse-amplified FM spectroscopy as a natural extension made to accommodate a useful and important application. Very often, working in the basic sciences environment, our group addresses esoteric and highly specialized questions. Few people around the world get excited by our findings, and few of our efforts bear practical fruit that can benefit the society in the immediate future. That is the reason why, when we understood that we are able to develop a tool that would provide the remote sensing community with a much needed versatile, absorption based spectroscopic technique, we jumped on the opportunity. Although, due to technical limitations of our laboratory, we can not make a viable "field" device, we carried out experiments and analysis that demonstrate the feasibility of the method and suggest a technological solution to the problem. The method is discussed in the first part (Section 2) of the thesis.

While working on this project, we have started the investigation of the spectroscopy of the first excited singlet state of isocyanogen (CNCN). This is the first step in our effort to use this molecule to study intramolecular dynamics of isomerization and dissociation. In many respects, CNCN is an ideal system for such studies. It is a stable molecule which has a stable isomer (cyanogen). According to *ab initio* calculations, the first excited state of isocyanogen would provide good Franck-Condon access to the barrier of the isomerization region on the ground electronic potential energy surface. In order to utilize the $\tilde{A}$ state as the intermediate state in a SEP-type experiment, spectroscopic characterization of the $\tilde{A}$ state will be necessary. Section 3 describes the

synthesis of isocyanogen and provides the latest results of the study of the spectroscopy of the excited state. Although we have successfully synthesized CNCN, our attempts to record its spectrum using a grating spectrometer failed due to a lack of sensitivity of the experimental setup used. This prompted us to build a Cavity Ring Down (CRD) spectrometer, which will be used in further studies.

Thus, the second spectroscopic method discussed in the Thesis, CRD Spectroscopy, has been employed in our efforts to detect the first excited singlet electronic state of isocyanogen (CNCN).

2 Frequency Modulation Enhanced Absorption Based Remote Sensing (FMRS).

2.1 Introduction.

In the following sections I will discuss the possibility of using the spectroscopic technique of Pulse Amplified Frequency Modulation Spectroscopy (FMS) in remote sensing applications requiring very high sensitivity detection of various atmospheric molecules. Recent advances in laser technology suggest that construction of a mobile, easy to operate remote sensing FM spectrometer might be both technically and economically feasible. Namely, we are seeking to improve techniques that are effective for distances of up to a mile and do not require (or rely on use of) a back-reflector or solid scatterer at the remote end of the optical path.

If scattered light is

(i) of sufficiently high intensity to be used with heterodyne detection methods,

(ii) retains the FM signal due to the absorption of the species under study, and

(iii) does not have any amplitude modulation as a result of the scattering process itself

it might provide a very useful technique for various remote sensing applications.

In this chapter I will address the following issues:

1. A short introduction to the current state of the field of remote sensing (RS) will help us to understand requirements faced by designers of a new RS technique. This review will stress the main features of these methods that demand improvement (Section 2.2).

2. An overview of FMS, including exact and approximate theories of FM signal, sensitivity and signal-to-noise analyses, are given in Sections 2.3 and 2.4. Comparison of sensitivities for direct absorption and fluorescence spectroscopies with that of FMS is presented. Results of this analysis show that use of FM spectroscopy is likely to provide necessary improvements to existing RS methods. The development of Pulsed Amplified FMS (PFM), the reasons for its development and the importance of PFM for FMRS are briefly discussed in Section 2.5.
3. Experiments using FMS for sensing of important molecules in ambient atmosphere are well known. The *caveat* is that in all the cases light has been directed onto the detector either by a mirror or by a solid point scatterer. Can FMS be used for FMRS when no solid back-scatterer is present? Following are the questions that will be considered in the next two sections and will help us to build a model of FMRS.
    a. Is the intensity of light that reaches the detector sufficient for heterodyne detection?
    b. Is the coherence of the RF modulation preserved in the scattering event?
    c. Even if coherence is preserved, does this imply that scattering from the *distributed* scatterer will result in interference, canceling the FM signal on the detector? If some signal will still be present, how will its magnitude compare to that of an FM signal in the light returned by a mirror with the reflectance equivalent to the scattering efficiency of the cloud under consideration.
    d. Sensitivity issues specific to the operation of an FM spectrometer at ambient pressures in the field conditions will be discussed. An exemplary setup will be introduced. Some of the issues are: use of dye versus diode lasers, fiber amplifiers, fiber EOMs, collection optics, and seeding of the photodetector with the source beam to improve sensitivity.

2.2 Spectroscopy in Remote Sensing and Environmental Science.

In today's fast changing , heavily industrialized world, the topic of sustainability of mankind's growth, and, in fact, its very existence, is closely connected to the understanding of the factors that control and cause environmental change. Development of new powerful experimental and theoretical tools is especially important to be able to more effectively address numerous problems posed by environmental science. A great number of global (such as stratospheric ozone destruction and global warming) and local (such as smog, carbon monoxide, tropospheric ozone, PAHs, etc.) atmospheric problems have major environmental consequences and, therefore, require serious study. Spectroscopic methods have proved to be invaluable in the context of atmospheric science[3,5,11,13]. The following short overview will provide a starting point for the discussion of our method and will focus on the spectroscopic techniques that give information about the presence and concentration of chemical species of interest.

LIDAR Methods.

- Raman scattering based techniques.

Raman-based techniques have great potential for measuring concentrations of certain atmospheric gases by monitoring the corresponding wavelength-shifted returns. On the other hand, as is true for all Raman-based techniques, they suffer from relatively small cross-sections. A basic set-up for such an experiment would include a laser with a very high pulse energy and high repetition rate which, if possible, would operate at frequencies taking advantage of the fourth-power dependence of the cross-section on optical frequency. The detection system must include a spectrometer or a tunable acousto-optic or interference "notch" rejection filter matched to the Raman-shifted wavelengths.

Most of the Raman-scattering based instruments were built to monitor water vapor concentration as a result of the acknowledged importance of water as an atmospheric constituent and a dominant greenhouse gas. For atmospheric water measurements, the Raman signal generated by nitrogen is observed simultaneously with the Raman signal generated by water. Ratioing of the two signals gives a water vapor mixing ratio. Similarly, very important information about the aerosol scattering ratio can be obtained by ratioing the contributions from Rayleigh and Mie scattering. A variety of aerosol and cloud parameters can be derived from such measurements.

Most of the systems in use employ frequency tripled output (355nm) of a Nd:YAG laser at for excitation and a set of interference filters in the detection channel (Melfi et al.[1] and Vaughan et al.[2]). To date, all the water-vapor Raman measurements had to be conducted at night due to difficulties of discriminating Raman signals from the solar background. Various groups have examined possibilities of working during the daytime using Raman lidar systems operating either in the solar blind ultraviolet spectral region, or with a very narrow spectral bandpass and restricted angular field of view. Analysis of such a system, carried out by Goldsmith et al.[3], showed that collection of water profiles with 10% precision for altitudes up to 4km would require a sampling time of at least 10 minutes. The most recent Raman lidar system[3a] uses a flash-lamp pumped Nd-YAG laser (Continuum 9030) with a KDP-based third harmonic generation at the source. The system operates at 355nm, producing 5ns 400mJ pulses at a repetition rate of 30Hz. Dichroic beamsplitters, incorporated into the receiving telescope, separate the photons into three groups: water vapor return at 408nm, nitrogen return at 387nm and Rayleigh/aerosol scatter at 355nm. Each group is then detected by a photomultiplier. In this system, a very high power of the source laser, along with a narrow field of view (FOV~ 0.3mrad) and narrowband filters, allows daytime operation. Temporal resolution is nominally one minute, but shorter times are, in principle, possible. A wide FOV mode (2mrad), intended for night use, is employed when weaker Raman signals must be detected.

- Differential absorption (DIAL)

Differential absorption lidar (DIAL) is the most widely used lidar-based remote sensing technique directed at composition measurements in the troposphere and stratosphere. The technique has been extensively used for detection of various trace gases since the mid-1960's and was most successful in the ultraviolet and visible spectral regions with ozone, nitric oxide, nitrogen dioxide, sulfur dioxide and chlorine (Grant et al.[4] and references therein). The main difficulty in the use of this technique in the infrared spectral region (which is a potentially more useful region for detection of many species as most everything absorbs at wavelengths longer than 2.5 μm, spectra are sharp and species specific) was the absence of high power continuously tunable light sources. The most commonly used lasers were line tunable systems such as DF (operating between 3.5 and 4 μm) and CO_2 (covering the wavelengths between 9.1 and 11 μm) lasers. Modern infrared laser systems, based on OPOs and difference frequency mixing as well as on new powerful diode lasers, would permit detection of a large number of species of interest.

The basic principle of DIAL is that it uses radiation at two wavelengths - one, λ_{on}, which is tuned to the absorption feature of interest, and the other, λ_{off}, tuned off this feature. The attenuation of the backscattered light at these two wavelengths is compared and the absorption spectrum of the species of interest is extracted from the difference. The main limitations of the technique become apparent from the following considerations. In addition to absorption by the substance of interest X, the extinction coefficient $\alpha(\lambda)$ is affected by absorption by other species and by extinction by aerosols.

$$\alpha(\lambda) = N_X \sigma_X(\lambda) + N_o \sigma_o(\lambda) + \alpha_a(\lambda), \qquad [2.01]$$

where N is concentration of molecules, $\sigma(\lambda)$ is an absorption cross-section, and subscripts X, o, and a specify the molecule of interest, other molecules, and aerosols, respectively. If the difference between parameter values at λ_{off} and λ_{on} is Δ, then the concentration of the molecule of interest can be expressed as:

$$N_X(h) = \frac{1}{2\Delta\sigma_X}\left[\frac{d}{dh}\left\{\ln\left(\frac{n_s(\lambda_{on},h)}{n_s(\lambda_{off},h)}\right)\right\} + \frac{d}{dh}\left\{\ln\left(\frac{\beta(\lambda_{on},h)}{\beta(\lambda_{off},h)}\right)\right\}\right] - \\ - \frac{1}{\Delta\sigma_X}\left[N_o(h)\Delta\sigma_o - \Delta\alpha_a(h)\right] \qquad [2.02]$$

where n_s is the number of photoelectrons generated by the backscattered pulse, h is the height of the sample volume of interest, and β is a volume backscatter coefficient[5]. It is quite easy to see from this expression that, when aerosol loading is negligible, the second term of the expression on the right side of the equation is largely determined by the Rayleigh scattering and is a constant. The fourth term is small and can be neglected. The third term can be accounted for by adopting an atmosphere density profile. Then, the concentration $N_X(h)$ may be derived solely from the first term. Unfortunately, such an ideal situation is rarely , if ever observed.

In the presence of aerosols, the second and fourth terms become very significant. Detailed information about the composition, shape and size distribution of the aerosols is necessary to calculate (usually using Mie theory) the wavelength dependence of the scattering, and, therefore, to determine its contribution to the β and α_a coefficients. Hence, in such a situation it is extremely difficult to extract information about the concentration (or even the presence) of species of interest, especially when widely separated probe wavelengths are used[6]. It is interesting to note here that the Raman signal in the backscattered radiation is dependent only on molecular terms and has no dependence on the presence of aerosols. So, one possible solution to the aerosol "uncertainty" problem would be to continuously calibrate DIAL against a Raman-scattering signal. It is clear that such a solution, although feasible, would be quite complicated and expensive to implement.

Another important sensitivity limitation of this technique arises from the fact that DIAL is in essence a direct absorption (homodyne) measurement. Even if the uncertainty due to the presence of aerosols is small and can be neglected, it is difficult to detect small changes on a large fluctuating background (an estimate of absorption-based detection

sensitivity and a comparison to the sensitivities of fluorescence-based and FM-based techniques will be presented in Section 2.4). In other words, the remote sensing community would greatly benefit from a *zero-background* technique.

Although differential absorption lidar systems are based on the same measurement principles as DOAS (Differential Optical Absorption Spectroscopy - very long absorption path technique using broadband sources) systems and the spectral range is similar, DIAL devices have substantial advantages, most notably a monostatic configuration (i.e. such that the source and detector are in the same location) that permits measurements in any direction and spatial resolution. The main deterrent to DIAL's wider use until recently has been the high cost of the laser source and the complexity of the system. A description of two recently developed DIAL systems will be provided to illustrate the preceding discussion.

The first system is a NASA airborne DIAL for measurement of atmospheric water vapor[10]. The system is based on a solid-state Alexandrite laser as an "on-line" source and a frequency doubled YAG pumped dye laser as an "off-line" source. In a subsequent paper[11], a second Alexandrite laser was added to replace the dye laser as the "off-line" source. This new modified "double ring cavity injection-seeded" system has a number of advantages over the older one. One of the major ones is that the near-identity of the two lasers simplifies the data reduction procedure. Avalanche photodiodes have been adopted for detection. A number of *in-situ* measurements were performed with this system and the authors state that the H_2O mixing ratios can be determined with an accuracy of almost 10%.

The second work I will mention here is interesting for its unusual choice of the source. DIAL wavelengths are often generated using a tunable laser (such as a Nd:YAG pumped dye laser). Sunesson et al.[12] used a fixed frequency UV laser together with stimulated Raman-shifting. The main advantage of this scheme is that it is relatively maintenance-free. The system described in the paper used two Nd:YAG lasers and two Raman cells - with H_2 and with D_2 providing radiation at 266, 289 and 299nm. Ozone profiles have been obtained at altitudes from 600 meters up to approximately 5 km.

Uncertainties in ozone concentration determination were reported to be up to 50 per cent in hazy weather.

The last work I would like to mention here is the most recent DIAL system reported[13]. It was constructed by CISE (Milan, Italy) and features the most advanced, to date, laser system. The source consists of a doubled Nd:YAG pumped Ti:sapphire laser that produces 100mJ pulses. Its radiation is subsequently frequency mixed with that of the YAG laser to produce 10mJ pulses, tunable in the 220-300nm range. This DIAL system was used in industrial and urban pollution studies to monitor pollutants such as SO_2, NO, NO_2, O_3 and VOC's, and the data obtained were compared to the data obtained from conventional sensors. Unfortunately, no error analysis or performance information was given in the article.

The previous discussion suggests that DIAL is plagued by two inherent problems: aerosol dependent returns and homodyne detection. Frequency Modulation methods, therefore, look very promising!

- Fluorescence based techniques

Although in principle a zero background technique, and, therefore, intrinsically more sensitive than direct absorption, fluorescence detection has not seen nearly as prolific use as DIAL or Raman based schemes. The main reason lies in the restriction imposed by collisional quenching, which is very rapid even in the stratosphere. Because of this difficulty, the only substance that has been relatively extensively studied by the fluorescence method is the hydroxyl radical. It is well known that hydroxyl radical plays a major role in the chemistry of the troposphere[14]. It is a primary oxidizing agent and is responsible for destruction of all major pollutants as well as carbon dioxide and methane. Measurements of hydroxyl concentrations with sensitivity better than 10^{12} m^{-3} is desirable. The traditional method of fluorescence based detection[7] of OH is based on its excitation from the $X^2\Pi$ (v"=0) state to the $A^2\Sigma(v'=1)$ state by laser radiation at 282nm. As a result, some of the OH radicals in $v'=0$ and $v'=1$ emit radiation near 310nm. Some interference is caused by non-resonant fluorescence from species other than OH, such as aerosol particulates and sulfur dioxide molecules. Another significant difficulty

in this measurement is caused by the photolytic production of hydroxyl radical by the probe laser itself. Laser light at 282nm photolyses ozone and produces $O(^1D)$ atoms, whose reaction with water results in hydroxyl radical formation. This probe laser generated hydroxyl is subsequently excited by the probe laser and fluoresces during the same laser pulse. Two improvements have been made to the original technique. The first one[8] uses an infrared-UV double resonance scheme to excite the hydroxyl radical. The first laser pumps OH molecules into the $v''=1$ vibrational state, while the second laser, tuned to near 343nm, excites hydroxyl into the *A* state. Neither photon has sufficient energy to generate $O(^1D)$ atoms, so photolytic production of hydroxyl does not occur. A laboratory version of this experiment has been successful, but its "field" implementation awaits the IR laser capable of pumping a sufficient number of OH radicals into the $v''=1$ excited state.

The second method[9] that improves the basic experimental scheme is based on rarefaction and purification of the air sample prior to LIF detection and requires the presence of the sensing system in the location under study (which implies expensive airborne experimental setups). Finally, an important limitation of LIF based detection methods should be emphasized one more time: the species under study must have a very strong fluorescent signature. In addition, these methods are not useful even for most of such species due to fast collisional relaxation at atmospheric pressure.

References:

1. S.H. Melfi and D. Whitemen, *Bull. Am. Meteorol. Sci.* **66**, 1288 (1985).
2. G. Vaughan, D.P. Wareing, L.Thomas and V. Mitev, *Q.J.R. Meteorol. Soc.* **114**, 1471 (1988).
3. J.E.M. Goldsmith and R.A. Ferrare. In M.P. McCormick (ed.). *Abstracts of the Sixteenth International Laser Radar Conference*, MIT, Boston, Mass., USA, 1992, p.667.

3a. J.E.M. Goldsmith, *Laser Focus World*, May (1997) 241.

4. W.B. Grant, R.H. Kagana and W.A. McClenny, *J. Air Management Assoc.*, **42**, 18(1992).

5. L. Thomas. In R.J.H. Clark and R.E. Hester (ed.). *Spectroscopy in Environmental Science*, Wiley & Sons, 1995, p.30.

6. T.J. McGee, M. Gross, R. Ferrare, W. Heaps and U. Singh *Geophys. Res. Lett.*, **20**, 955 (1993).

7. E.L. Baardsen and R.W. Terhune, *Appl. Phys. Lett.* **21**, 209 (1972).

8. J.D. Bradshaw, M.D. Rodgers and D.D. Davis, *Appl. Opt.* **23**, 2134 (1984).

9. T.M. Hard, R.J. O'Brien, C.Y. Chan and A.A. Mehrabzadeh, *Environ. Sci. Technol.* **18**, 768 (1984).

10. N.H. Higdon et al., *Applied Optics*, **33** (27), (1994) 6422.

11. N.H. Higdon, T.H. Chyba, P. Ponsadrin, Sixth Topical Meeting on Optical Remote Sensing in the Atmosphere, Feb. 6-10, 1995, Salt Lake City, Utah.

12. J.A. Sunesson, A. Apituley, D. Swart *Applied Optics*, **33** (30),(1994) 7045.

13. L. Garifo *Laser Focus World*, April (1997) 95.

14. Wayne, R.P., "Chemistry of Atmospheres", 2nd ed. Oxford University Press, 1988.

2.3 Frequency Modulation Spectroscopy.

Section 2.2 gave a brief overview of experimental techniques currently used for optical remote sensing. A new, universally applicable optical spectroscopic technique that is capable of higher sensitivity and higher signal to noise ratio is necessary to address a large number of atmospheric sensing problems and to improve the quality of experimental data obtained in field experiments. In this section we will introduce the method of Frequency Modulation Spectroscopy (FMS), rigorously derive the FM signal expression for all possible values of experimental parameters, and look at the simplified version of the theory applicable to small modulation indices.

A discussion of signal to noise and sensitivity issues and the comparison of FM spectroscopy with direct absorption and fluorescence based techniques follows in Section 2.4. We hope to show why we expect significant sensitivity improvements when FMS instead of direct laser absorption spectroscopy (LAS) or Fluorescence Detection (FD) is used. This section will also prepare the foundation for building a model describing FM-Remote Sensing (FMRS) and addressing the sensitivity issues of this technique, which will follow in subsequent sections.

Frequency modulation spectroscopy was born out of the need for a sensitive absorption based spectroscopic technique to study atomic and molecular physics and chemistry. Absorption spectroscopy is, perhaps, the most generally applicable and versatile spectroscopic method. Its primary disadvantage lies in the fact that the measurement relies on the detection of a change of energy of a light beam that passes through a sample. In most interesting cases, absorption signals are very small, due either to minute amounts of detected substance or tiny transition strengths or both. At the same time, the sensitivity of the direct absorption technique is limited by random noise in the laser intensity that causes strong background fluctuations. Ordinary direct absorption spectroscopy does not distinguish between the changes in intensity of a laser beam caused by molecular absorption and those caused by laser noise or environmental interference. On the other hand, absorption spectroscopy has several major advantages. It is a linear technique - the magnitude of the absorption signal linearly and directly relates to the

amount of the substance. This relation is very simple and requires knowledge of just one parameter, the absorption cross section. In contrast, some other techniques require knowledge of more than one parameter (e.g. for fluorescence detection one needs to know absorption cross section and fluorescence quantum yield), while in other techniques (such as four-wave mixing) the signal may depend non-linearly on the laser power, absorption cross section and other possible experimental parameters. Another important feature of absorption spectroscopy is that it is largely non-destructive and does not depend on the stability of the species under investigation (in contrast to such techniques as fluorescence, which cannot be applied to species which undergo rapid dissociation or non-radiative deactivation).

FM spectroscopy was not the first attempt to extend the sensitivity limits of traditional absorption spectroscopy, but one of the most successful ones. It was first introduced in 1980 by G. Bjorklund[1] to solve problems inherent to direct laser absorption spectroscopy (LAS), while retaining the positive features of LAS. Modulation techniques are often used to extract weak signals from a noisy background. Examples of using sample modulation as well as source intensity modulation are well known, but are not always practicable. It is, in general, quite difficult to modulate the amount of sample at a high enough frequency to escape outside of the laser noise spectrum (which, for a typical cw single longitudinal mode laser, is on the order of magnitude of several MHz, while for typical pulsed lasers can be up to a few GHz). Intensity modulation of a light source has the disadvantage of modulating both signal and noise simultaneously.

Frequency modulation (or *phase modulation*, which is a different name for the same phenomenon) provides a significant, up to 10 orders of magnitude[2,3,4], sensitivity improvement over traditional LAS. However, often FMS fails to reach its ultimate sensitivity limit because of its own very specific limitations, the source and character of which will become clear from the discussion in Section 2.4.

2.3.1 A complete non-approximate theory of FM signal .

In this sub-section, a complete, approximation-free treatment of FM spectroscopy will be presented. A simple, approximate treatment will follow in order to create a more intuitive physical picture of the technique.

The expression relating the photocurrent on the detector to the electric field of the laser beam is

$$I(t) = Q_{\text{det}} \frac{1}{T} \int_0^T E_R(t) E_R(t) dt, \qquad [2.1]$$

where $I(t)$ is the photocurrent, $E_R(t)$ is the electric field at the detector, T is the period of optical oscillation, and Q_{det} is the quantum efficiency of the detector. The electric field is presented as a scalar because we can treat light as being linearly polarized without sacrificing generality. Subscript *"R"* stands for *real*. $E_R(t)$ is the real electric field of the laser beam. This is emphasized because, although electric field is an observable (and, therefore, real) quantity, it is often desirable to treat it as a complex vector, $E(t)$, to simplify calculations. In this case, $E_R(t)$ is the real part of $E(t)$:

$$E_R(t) = \text{Re}[E(t)] = \frac{1}{2}[E(t) + E(t)^*], \qquad [2.2]$$

where $E(t)^*$ is the complex conjugate of $E(t)$. To avoid confusion we will always specify which form of electric field expression, real or complex, we are using. The averaging over the optical period introduced in expression [2.1] is necessary because no conventional optical detector is fast enough to respond at optical frequencies[5].

A simplified outline of a typical FM experimental setup is presented in Figure 2.1.

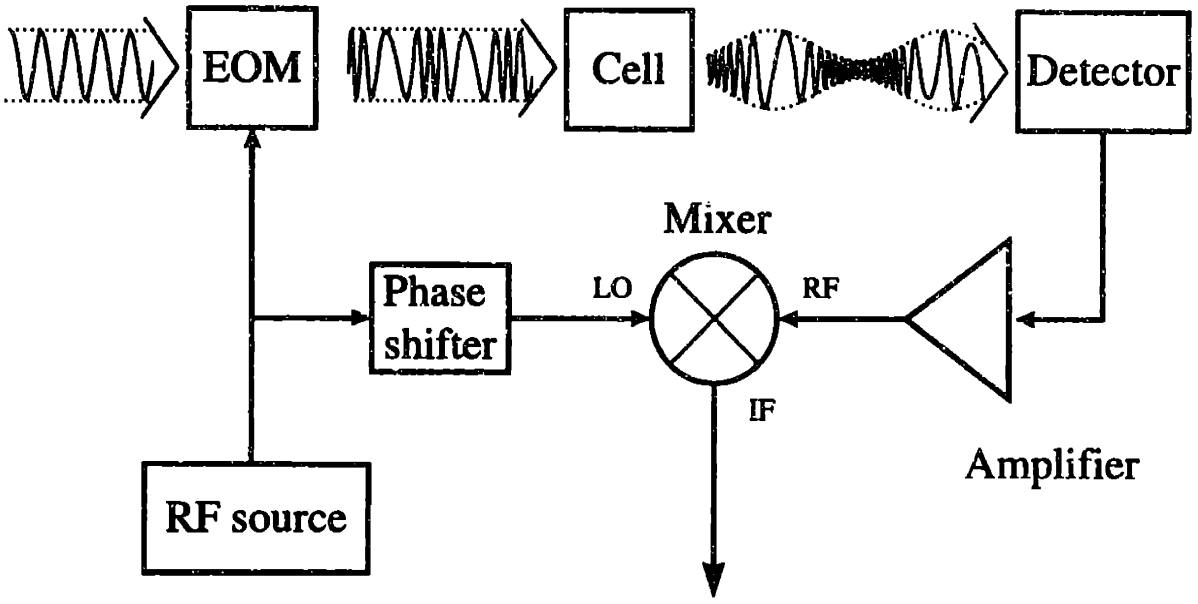


Figure 2.1 Simplified experimental setup for a typical FMS experiment. **LO** stands for *local oscillator*, **IF** is a standard notation for low frequency output.

Light from a single longitudinal mode dye laser is frequency (phase) modulated by an RF-field driven electro-optical modulator (EOM)[6]. Light exiting the EOM is directed through a sample cell and onto a photodetector. The output of the photodetector is conditioned (filtered using bandpass filter centered at the RF modulation frequency, then amplified) and sent into an RF mixer, which is a device that extracts the component of the photocurrent that oscillates at the frequency of the RF modulation. This photocurrent component is the FM signal.

The mathematical theory of FM has been developed earlier in a number of papers, so the treatment presented here is not completely new[1,11,19]. On the other hand, in most of the previous treatments, authors tailored their discussions to the particular FM experiments involving further refinements of the FM method[2,17,18]. Our treatment is unique in that it will be both very general and will, for the first time, introduce dependence of the frequency modulated optical field on the spatial position. This last feature will later allow us to develop a model for FM-enhanced remote sensing.

The optical field of laser light entering the EOM is

$$E_c(t) = E_c e^{i(\Omega t+\varphi)}, \qquad [2.3]$$

where $E_c(t)$ and E_c are, respectively, complex field and complex field amplitude of the *carrier* (hence subscript "c"), i.e. the field at the frequency of the source laser. The optical frequency of the carrier is denoted Ω and the phase is φ. We can set φ to be zero without loss of generality. Ideally, the electric field after the modulator is purely phase modulated. In other words, an ideal EOM does not introduce any amplitude modulation. Then, the field, upon exiting EOM, is

$$E_{FM}(t) = E_c e^{i(\Omega t + M\sin(\omega t+\phi))}, \qquad [2.4]$$

where $M\sin(\omega t+\phi)$ is the time dependent phase shift introduced by the EOM, M is the modulation index, ω is the modulation frequency and ϕ is RF phase (we can assume ϕ to be zero without loss of generality). The modulation index is the peak phase retardation, which is determined by the non-linear response of the crystal as well as the magnitude of the applied RF field[6]. The modulation index should not be confused with the modulation depth, which is the ratio of sideband to carrier powers. Rewriting the expression for the modulated field

$$E_{FM}(t) = E_c e^{i\Omega t} e^{iM\sin(\omega t)} \qquad [2.5]$$

we notice that $e^{iM\sin(\omega t)}$ is a periodic function and, therefore, can be represented as a Fourier series:

$$e^{iM\sin(\omega t)} = \sum_{n=-\infty}^{\infty} c_n e^{in\omega t}. \qquad [2.6]$$

The Fourier coefficients, c_n, are found using the standard formula:

$$c_n = \frac{1}{2\pi}\int_{-\omega t}^{\omega t} e^{iM\sin(\omega t)} e^{-in\omega t}\,dx = J_n(M), \qquad [2.7]$$

where $J_n(M)$ is an n-th order Bessel function of the first kind, of argument M. Expression [2.7] is a generating integral for Bessel functions[8].

Combining equations [2.5]-[2.7] and introducing the field dependence on the spatial coordinate, we can rewrite the frequency modulated field as

$$E_{FM}(t) = E_c \sum_{n=-\infty}^{+\infty} J_n(M) e^{i(\Omega t + n\omega t - k_n x)}, \qquad [2.8]$$

where $k_n = \frac{2\pi}{\lambda_n} = \frac{2\pi(\Omega + n\omega)}{c}$ is the wave number for each sideband.

Light passing through the sample interacts with it. Attenuation of the intensity of light at frequency ω_n is expressed by δ_n, where δ_n is related to the pathlength through the sample, L, and the frequency dependent intensity (as opposed to amplitude) absorption coefficient α_n as

$$\delta_n = \frac{\alpha_n L}{2}. \qquad [2.9]$$

The notation is chosen to simplify the expression

$$\omega_n = \Omega + n\omega. \qquad [2.10]$$

Analogously, we can introduce a phase shift of a frequency component ϕ_n ,which is related to the frequency dependent index of refraction n_n as

$$\phi_n = \frac{n_n L \omega_n}{c}. \qquad [2.11]$$

Then the overall effect of the light field interaction with the sample for each frequency component will be expressed in terms of the function T_n:

$$T_n = e^{-\delta_n} e^{-i\phi_n}. \qquad [2.12]$$

Using equation [2.12] we can express the effect of the sample on the light electric field as

$$E_{FM}^{T}(t) = E_c \sum_{n=-\infty}^{+\infty} T_n J_n(M) e^{i(\Omega t + n\omega t - k_n x)}. \qquad [2.13]$$

Using equation [2.1] we can calculate the photocurrent on the detector. It is easy to notice that each component of the sum in equation [2.13] can be represented as a product of two terms, a rapidly oscillating (optical frequency) term and a slowly oscillating term (containing all the constants) at multiples of the RF frequency:

$$E_{FM}^{T}(t) = \left[E_c \sum_{-\infty}^{+\infty} T_n J_n(M) e^{i(n\omega t - k_n x)} \right] \cdot \left[e^{i\Omega t} \right]. \qquad [2.14]$$

Denoting the slowly varying part as

$$\xi = E_c \sum_{-\infty}^{+\infty} T_n J_n(M) e^{i(n\omega x - k_n x)}, \qquad [2.15]$$

we can calculate the photocurrent as

$$\begin{aligned} I(t) &= Q_{\det} \frac{1}{T} \int_0^T E_R(t) E_R(t) dt \\ &= \frac{Q_{\det}}{4T} \int_0^T (E(t) + E(t)^*)(E(t) + E(t)^*) dt = \\ &= \frac{Q_{\det}}{4T} \int_0^T \left(E^2(t) + E(t)^{*2} + 2E(t)E(t)^*\right) dt. \end{aligned} \qquad [2.16]$$

The integral over the first two components is zero because

$$\int_0^T e^{2i\Omega t} dt = \int_0^T e^{-2i\Omega t} dt = 0 \quad \text{for } T = \frac{2\pi}{\Omega}. \qquad [2.17]$$

In the cross-term, the rapidly oscillating terms cancel each other while the slowly oscillating terms can be pulled out of the integral, because on the time scale of the optical period, the slowly varying terms can be considered constant. Thus, the total photocurrent caused by the frequency modulated electric field impinging on the detector is given by the equation:

$$\begin{aligned} I_{FM}(t) &= \frac{Q_{\det}}{2} E(t) E(t)^* \\ &= \frac{Q_{\det} E_c^2}{2} \sum_{l=-\infty}^{+\infty} T_l J_l(M) e^{i(l\omega x - k_l x)} \sum_{m=-\infty}^{+\infty} T_m^* J_m(M) e^{-i(m\omega x - k_m x)}. \end{aligned} \qquad [2.18]$$

A double balanced mixer extracts the current component oscillating at ω (ω is the modulation frequency) from the photocurrent represented by equation [2.18]. Therefore we need to take into account only terms with $|l-m|=1$:

$$I_{FM}^{\omega}(t) = \frac{Q_{\det} E_c^2}{2} \sum_{m=-\infty}^{+\infty} J_m J_{m+1} \left[T_m T_{m+1}^* e^{-i(\omega x - k_{RF} x)} + T_m^* T_{m+1} e^{i(\omega x - k_{RF} x)} \right], \qquad [2.19]$$

where $I_{FM}^{\omega}(t,x)$ is a photocurrent component oscillating at the modulation frequency, and

$$k_{RF} = \frac{2\pi\omega}{c} \qquad [2.20]$$

is a wave number corresponding to the RF modulation. This expression for the current is

a sum of RF waves, whose phases depend on position along the propagation path. This is a very important result that will be discussed in the following sections.

The mixer has two inputs - signal input (RF) and local oscillator (LO). It works by extracting the component of signal input which is oscillating at the same frequency and in-phase with the local oscillator (analogous to what a lock-in amplifier does with a modulated signal). The output of the mixer is a product of LO and signal (RF) inputs:

$$V(t) = RI_{FM}(t)\zeta \cos(\omega + \phi_{RF}). \qquad [2.21]$$

Therefore, the mixer output contains frequency components at DC and at twice the modulation frequency, 2ω. Either component can be used for further signal processing, but the high frequency component is most often removed by low-pass filtering and only the DC component is used. The constant ζ is inserted into the equation to show that the output signal will be somewhat reduced by losses in the mixer.

Now, we can write the FM signal in the output of the mixer as a sum of all terms in equation [2.18] with $|l-m| = 1$:

$$V_{FM-SIG}(t) = R\zeta\, Q_{det} \frac{E_c^2}{4} \sum_{m=-\infty}^{+\infty} J_m J_{m+1} \left[T_m T_{m+1}^* e^{-i(\phi_{RF} + k_{RF} x)} + T_m^* T_{m+1} e^{i(\phi_{RF} + k_{RF} x)} \right]. \qquad [2.22]$$

Remembering that the sidebands with $+n$ and $-n$ are related by

$$J_{-n}(M) = (-1)^n J_n(M), \qquad [2.23]$$

and renaming $R\zeta\, Q_{det} \frac{E_c^2}{4}$ as Ξ and $\phi_{RF} + k_{RF} x$ as ϕ we can rewrite [2.22] as

$$\begin{aligned} V_{FM-SIG}(t) &= \Xi \sum_{m=0}^{\infty} J_m J_{m+1} \left[\left(T_m T_{m+1}^* - T_{-m}^* T_{-m-1} \right) e^{-i\phi} + \left(T_{m+1} T_m^* - T_{-m-1}^* T_{-m} \right) e^{i\phi} \right] \\ &= \Xi \sum_{m=0}^{\infty} J_m J_{m+1} \left[\begin{array}{l} e^{-\delta_m - \delta_{m+1}} \left(e^{-i(\phi_m - \phi_{m+1} + \phi)} + e^{i(\phi_m - \phi_{m+1} + \phi)} \right) - \\ - e^{-\delta_{-m} - \delta_{-(m+1)}} \left(e^{i(\phi_{-m} - \phi_{-(m+1)} - \phi)} + e^{-i(\phi_{-m} - \phi_{-(m+1)} - \phi)} \right) \end{array} \right] \\ &= \Xi \sum_{m=0}^{\infty} J_m J_{m+1} \left[e^{-\delta_m - \delta_{m+1}} \cos(\phi_m - \phi_{m+1} + \phi) - e^{-\delta_{-m} - \delta_{-(m+1)}} \cos\left(\phi_{-(m+1)} - \phi_{-m} + \phi \right) \right]. \end{aligned} \qquad [2.24]$$

This expression describes the FM signal dependence in the most general way for any modulation index and transition strength. The signal on the output of the mixer will have two components which can be simplified by presentation in the form of quadratures: in-

phase and 90^{O} out-of-phase relative to ϕ. The implications of this quadrature feature for detection schemes will be discussed later. These two components are:

$$V_{FM-SIG} = V_{FM-SIG}^{Inphase}\cos(\phi) + V_{FM-SIG}^{Quad}\sin(\phi), \qquad [2.25]$$

where

$$V_{FM-SIG}^{Inphase} = \sum_{m=-\infty}^{+\infty} J_m J_{m+1}\left[e^{-\delta_m-\delta_{m+1}}\cos(\phi_m-\phi_{m+1}) - e^{-\delta_{-m}-\delta_{-(m+1)}}\cos(\phi_{-(m+1)}-\phi_{-m})\right],$$
$$V_{FM-SIG}^{Quad} = \sum_{m=-\infty}^{+\infty} J_m J_{m+1}\left[e^{-\delta_{-m}-\delta_{-(m+1)}}\sin(\phi_{-(m+1)}-\phi_{-m}) - e^{-\delta_m-\delta_{m+1}}\sin(\phi_m-\phi_{m+1})\right]. \qquad [2.26]$$

If no sample is present, the terms within the brackets cancel and the entire expression is equal to zero, as expected. These expressions are useful when very accurate simulations of FM signal are necessary, although such a task would require use of a computer.

On the other hand, an understanding of the physical nature of processes leading to noise cancellation in FMS is obscured by the large number of parameters and terms involved in expression [2.24]. This is the main reason we are going to develop a simplified model of FMS in the following section. As we shall see later, the simplified, approximate theory, valid only for small modulation indices M, is a special case of this treatment. All further discussion, although illustrated by the examples using the simplified model, will be applicable to the general case of FMS as well.

A few important questions should be addressed before we proceed to development of a simplified theory. The first one illustrates energy conservation in the frequency modulation. Namely, if only frequency (or phase) is modulated, then no amplitude modulation should have been introduced and the light energy before and after the modulation should be equal. Indeed, there is a theorem for Fourier series (Parseval's Theorem) that states that the total average intensity of a periodic signal is equal to the sum of the squares of its Fourier coefficients. For FM spectroscopy, the Fourier coefficients are Bessel functions of the first kind. A Bessel function series relation

$$\sum_{n=-\infty}^{+\infty} J_n(M) = 1 \qquad [2.27]$$

says that, regardless of the value of modulation index (including M=0), the average energy remains the same.

Another important note concerns the transition from the strict to a simplified treatment. The approximate form of FMS equations are applicable when the M-parameter size justifies considering only a carrier and its first order sidebands, neglecting all higher order sidebands. From expression [2.22] we can see that the retained accuracy, after such an approximation is made, depends mostly on the size of the $J_k(M)J_{k+1}(M)$ terms. Figure 2.2 shows the sizes of these terms for some values of k and M. Clearly, this product for larger values of k will have an even smaller magnitude and can therefore be neglected.

Equation [2.24] can be optimized for the largest FM signal. It has been found in a number of previous studies that the largest FM signal is obtained when the modulation frequency is on the order of the linewidth of the absorption feature under study[10,11,19]. For such a value of the modulation frequency the optimal modulation index is around M=1.4[10].

Figure 2.2 shows the dependence of the magnitude of the first several terms in FM signal expression. It is quite obvious that, for small modulation index M, the first order sidebands dominate and all other orders can be neglected (see next section for more discussion).

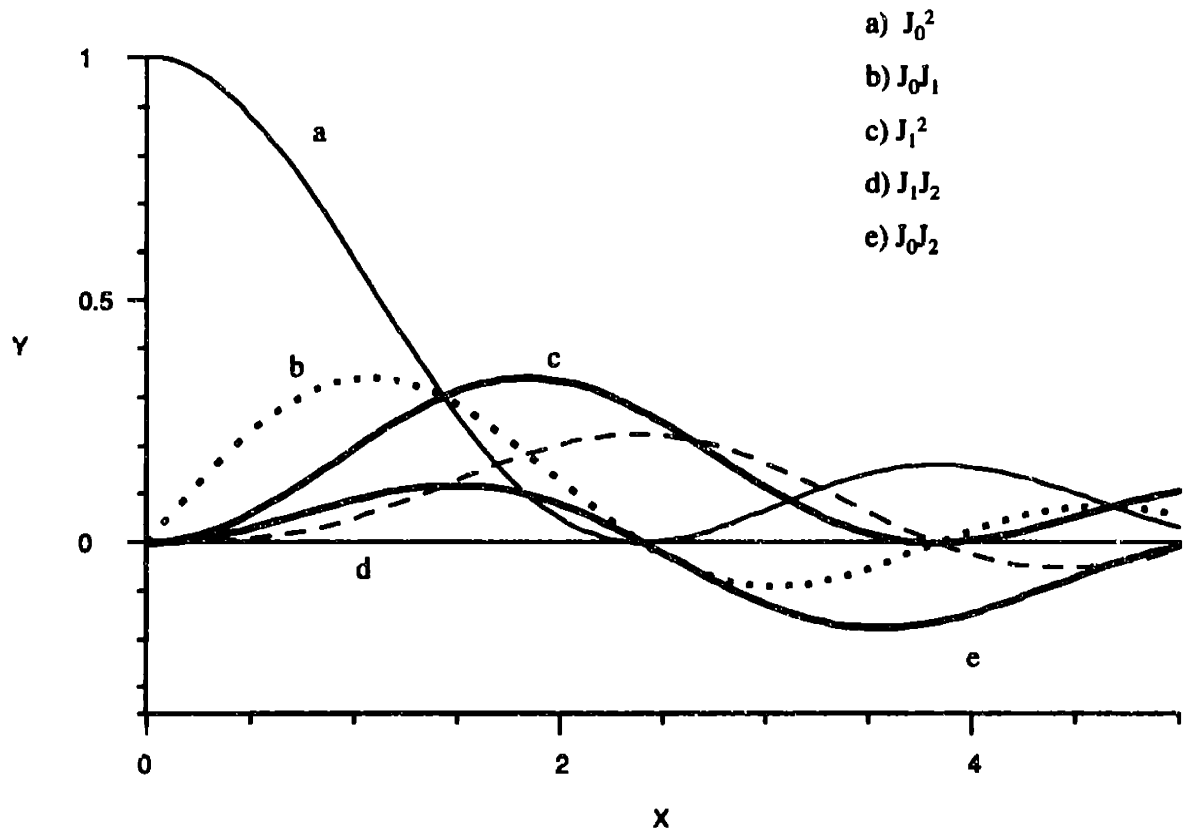


Figure 2.2. Products of Bessel functions of the first kind.(see text)

A decrease in modulation frequency decreases the signal magnitude. This relationship between modulation frequency, linewidth of the spectral feature and sensitivity is very important for the assessment of parameters of the model for the remote sensing experiment. Because of significant pressure broadening at ambient conditions, the choice of modulation frequency becomes crucial to the design of an optimally sensitive spectrometer.

2.3.2 Approximate Theoretical Treatment of FMS signal.

Let us again introduce a frequency modulated complex electric field of the laser:

$$E_{FM}(t) = E_c e^{i(\Omega t + M\sin(\omega t + \phi))}. \qquad [2.4]$$

For small values of the modulation index, M, Bessel function of the first kind can be approximated as:

$$J_n(M << 1) = (-1)^n \frac{M^n}{2^n n!}. \qquad [2.28]$$

For small M we shall assume that only the first order sidebands are important, all others are too weak to have significant influence on the signal and can be omitted from the discussion. Thus, $J_0=1$ and $J_{\pm 1} = \mp\frac{M}{2}$, while other terms vanish. In such a case, frequency modulated light has a spectrum with a strong carrier at frequency Ω and two weak sidebands at frequencies $\Omega \pm \omega$. Additionally, to simplify the transition to the discussion of FMS-based remote sensing, we will introduce the dependence of the frequency modulated optical field on the spatial position. The overall sample effect (recall formulae [2.9]-[2.12]) on each frequency component, in terms of the transmission function T_n, is:

$$T_n = e^{-\delta_n} e^{-i\phi_n}. \qquad [2.12]$$

Then, the transmitted field is:

$$E_{FM}(t,x) = E_c\left[T_{-1}\frac{M}{2}e^{i(\Omega t - \omega t - k_{-1}x)} + T_0 e^{i(\Omega t - k_0 x)} - T_1\frac{M}{2}e^{i(\Omega t + \omega t - k_1 x)}\right], \qquad [2.29]$$

where $k_n = \frac{2\pi}{\lambda_n} = \frac{2\pi(\Omega + n\omega)}{c}$ is the wavenumber for each sideband.

Recalling the derivation of the photocurrent expression given by formulae [2.16]-[2.18], the photodetector current caused by frequency modulated light is

$$I_{FM}(t) = \frac{Q_{det}}{2} E_{FM} E_{FM}^* =$$

$$= \frac{Q_{det} E_c^2}{2} \left[\begin{array}{l} -T_{-1}T_0^* \frac{M}{2} e^{-i\omega t - i(k_0 - k_{-1})x} - T_{-1}^* T_0 \frac{M}{2} e^{i\omega t + i(k_0 - k_{-1})x} \\ + T_0 T_1^* \frac{M}{2} e^{-i\omega t - i(k_1 - k_0)x} + T_0^* T_1 \frac{M}{2} e^{i\omega t + i(k_1 - k_0)x} + |T_0|^2 \\ + \frac{M^2}{4} \left(|T_{-1}|^2 + |T_1|^2 \right) + \{(2\omega)term\} \end{array} \right] \qquad [2.30]$$

Differences in wavenumber values for the carrier and sidebands in the expression above are equal to the wavenumber k, which corresponds to the modulating RF field:

$$k_0 - k_{-1} = k_1 - k_0 = \frac{2\pi\omega}{c} = k. \qquad [2.31]$$

The term oscillating at twice the modulation frequency can be omitted because the double balanced mixer will allow us to chose only the term oscillating at the modulation frequency. Analogously, the DC term can be omitted from consideration. Terms containing $\frac{M^2}{4}$ will also be neglected because of their size.

Calculating values of products of transmission functions:

$$\begin{aligned} |T_0|^2 &= e^{-2\delta_0}; \\ T_{-1}T_0^* &= e^{-2\delta_0} e^{\delta_0 - \delta_{-1}} e^{i(\phi_0 - \phi_{-1})}; \\ T_{-1}^* T_0 &= e^{-2\delta_0} e^{\delta_0 - \delta_{-1}} e^{-i(\phi_0 - \phi_{-1})}; \\ T_1 T_0^* &= e^{-2\delta_0} e^{\delta_0 - \delta_1} e^{i(\phi_0 - \phi_{-1})}; \\ T_1^* T_0 &= e^{-2\delta_0} e^{\delta_0 - \delta_1} e^{-i(\phi_0 - \phi_{-1})}; \end{aligned} \qquad [2.32]$$

and using [2.28], the photodetector current can be written as:

$$I_{FM}(t,x) = \frac{Q_{det} E_c^2}{2} e^{-2\delta_0} \left[\begin{array}{l} 1 - e^{(\delta_0 - \delta_{-1})} \frac{M}{2} \left(e^{i(\phi_0 - \phi_{-1})} e^{-i(\omega t - kx)} + e^{-i(\phi_0 - \phi_{-1})} e^{i(\omega t + kx)} \right) \\ + e^{(\delta_0 - \delta_1)} \frac{M}{2} \left(e^{-i(\phi_0 - \phi_1)} e^{-i(\omega t - kx)} + e^{i(\phi_0 - \phi_{-1})} e^{i(\omega t + kx)} \right) \end{array} \right] \qquad [2.33]$$

Assuming that the absorption and dispersion introduced by the sample are weak (in fact all we need is a small difference between either absorption or dispersion experienced by the carrier and sidebands) we can expand the exponentials in the expression [2.33]:

$$I_{FM}(t,x)=\frac{Q_{det}E_c^2}{2}e^{-2\delta_0}\begin{bmatrix}1+\frac{M}{2}(\delta_{-1}-\delta_1)(e^{i\omega t+ikx}+e^{-i\omega t-ikx})\\+\frac{M}{2i}(\phi_{-1}-\phi_0+\phi_1-\phi_0)(e^{i\omega t+ikx}-e^{-i\omega t-ikx})\end{bmatrix}= \quad [2.34]$$

$$=\frac{Q_{det}E_c^2}{2}e^{-2\delta_0}\cdot\begin{bmatrix}[1+M(\delta_{-1}-\delta_1)\cos(\omega t-kx)\\+M(\phi_{-1}+\phi_1-2\phi_0)\sin(\omega t-kx)]\end{bmatrix}.$$

Once again we notice that the functional form of the photocurrent is the same as for the RF wave. This is a very important result that we will use in the development of the FMRS theory. One can imagine a simplified picture of the propagation of frequency modulated light, where the current from the detector is connected to the amplitude modulation intensity envelope, resulting from the imbalance of the sidebands. Another nice feature of the simplified treatment is that the nature of the FM signal is very clearly illustrated by equation [2.34]. Figure 2.2, which shows the values of products of the several largest Bessel function coefficients, justifies our use of the simplified theory, because for all practical values of the modulation index, the first-order sidebands will have the strongest impact on the FM signal.

References:

1. Bjorklund, G.C. "Frequency-modulation spectroscopy: A new method for measuring weak absorptions and dispersions", *Opt. Lett*. **5**, 15-17 (1980).
2. Ye, J. Ma, L.-S., Hall, J., "Ultrasensitive detections in atomic and molecuar physics - demonstration in molecular overtone spectroscopy, manuscript in preparation.
3. Gehrtz, M., Bjorklund, G., Whittaker, E. "Quantum-limited laser frequency modulation spectroscopy", *J. Opt. Soc. B*, **2**, 1510 (1985).
4. Whittaker, E., Gehrtz, M., Bjorklund, G., " Residual amplitude modulaiton in laser electro-optic modulation", *J. Opt. Soc. B*, **8**, 1320 (1985).

5. Rieke, G.H., *Detection of Light: From the Ultraviolet to the Submillimeter.*, Cambridge University Press, 1996.
6. Yariv, A. *Optical Electronics*, 1991, Saunders College Publishing.
7. Abramowitz, M., Stegun, I. ed. by, *Handbook of Mathematical Functions*, NBS Applied Series, 1964.
8. Lipson, S., Lipson, H., Tanhauser, D., *Optical Physics*, 3rd edn., Cambridge University Press.
9. Hsu, H. P., *Schaum's outline of theory and problems of analog and digital communications*, New York, McGraw-Hill, 1933.
10. Bjorklund, G.C., Levenson, M.D., Lenth, W., Ortiz, C., "FM spectroscopy: Theory of Lineshapes and Signal to-Noise Analysis", *Appl. Phys. B*, **32**, 145 (1983).
11. Silver, J.A., Bomse, D.S., Stanton, A.C., "Frequency-modulaiton absorption spectroscopy for trace species detection: theoretical and experimental comparison among methods" in Optical Methods for Ultrasensitive Detection and Analysis: Techniques and Applications, *Proc. SPIE*, **1435**, 64 (1991).
12. Gower, J., *Optical Communication Systems*, Englewood Cliffs, N.J., Prentice-Hall, 1984.
13. Keiser, G., *Optical Fiber Communications*, New York, McGraw-Hill, 1983.
14. Levenson, M.D., Moerner, W.E., Horne, D.E., "FM Spectroscopy detection of stimulated Raman gain", *Opt. Lett.*, **8**, 108 (1983).
15. Wong, N.C., Hall, J.L., "Servo control of amplitude modulation in frequency-modulation spectroscopy: demonstration of shot-noise limited detection.", *J. Opt. Soc. Am. B*, **9**, 1527 (1985).
16. Whittaker, E.A. et al., "ND_4 Schuler band absorption observed by laser FM spectroscopy in a photochemical reaction", *J. Chem. Phys.*, **80**, 961 (1984)
17. Whittaker, E.,A., et al., "Laser FM spectroscopy with photochemical modulation: a sensitive high resolution technique for chemical intermediates." *Appl. Phys. B*, **35**, 105 (1984).

18. McCarthy, M.C., Bloch, J.C., Field R.W., "Frequency-modulation enhanced magnetic rotation spectroscopy: a sensitive and selective absorption scheme for paramagnetic molecules", *J. Chem. Phys.*, **100**, 6331 (1994).

19. Levenson, M.D., Kano, S.S., *Introduction to Nonlinear Spectroscopy*, Boston: Academic Press, 1988.

20. Bloch, J.C., "Extending Frequency Modulation Spectroscopy: Sensitive and selective High Resolution Spectroscopy in the Visible and Ultraviolet", *Ph.D. Thesis*, M.I.T. 1996.

21. Avetisov, V.G., Kuranen, P., "Two-tone frequency-modulation spectroscopy for quantitative measurements of gaseous species: theoretical, numerical, and experimental investigation of line shapes", *Appl. Opt.*, **35**(24), 4705 (1996).

2.4 Sensitivity and Signal-to-Noise of the Frequency Modulation Spectroscopy.

The discussion in this section will be focused on developing an understanding of noise and background sources in traditional FM spectroscopy. We will explain how the background level depends on the signal and how the sensitivity of the technique depends on the background. Results of this discussion will be applied to analysis of sensitivities of three major laser based spectroscopic techniques - ordinary Laser Absorption Spectroscopy (LAS), Frequency Modulation Spectroscopy (FMS) and Fluorescence Spectroscopy (FS). Additional sources of noise due to the pulse amplification process will be addressed in the final section. Discussion of sensitivity issues has been presented in a number of papers published to date[6,7]. In most previous publications the authors specifically addressed questions of achieving the *limiting* sensitivity of the FM spectroscopy. Our goal is to determine the noise sources for all signal levels and to include background in the noise analysis. Several sources of noise must be addressed: laser intensity fluctuations and noise in the modulation and demodulation circuitry, the detector and the amplifiers. All the noise components will be analyzed based on the electrical signals they generate.

Whenever a measurement of an FM signal is made, the photocurrent from the detector has two components to it - the "theoretical" signal itself and noise:

$$i_{Total} = i_{FM} + i_{Noise}(t). \qquad [2.35]$$

The signal current i_{FM} is constant because we assume that the laser is locked at a certain wavelength. Thus, the time dependence of the total current comes solely from the noise. In the previous section we introduced a relationship between the photocurrent and the electrical field on the detector [2.1]. Detector efficiency Q_{det} can be expressed as:

$$Q_{\text{det}} = \frac{e\eta}{\hbar\omega} G_{\text{det}} = G_{\text{det}}\rho, \qquad [2.36]$$

where e is electron charge, η is quantum efficiency of the detector (probability of generation of an electron by a photon falling on the detector), G_{det} is the gain of the detector (in current units) and ρ is detector responsivity in amps/(watt of optical power).

Quantum efficiency and detector responsivity are frequency dependent, but change slowly enough with laser frequency to be considered the same for all the components of the frequency modulated light. At this point we consider a situation where there is no stable, constant parasitic background in the FM signal. The contribution to the total current that is due to the time dependent noise is a signed quantity, which adds or subtracts to the signal and has a zero time averaged mean:

$$\begin{aligned} i_{FM} &= \lim_{T\to\infty}\int_0^T i_{Total}(t)dt = \overline{i_{Total(t)}}, \\ \overline{i_{Noise}} &= \lim_{T\to\infty}\int_0^T i_{Noise}(t)dt = 0. \end{aligned} \qquad [2.37]$$

Then, the total noise power is given by:

$$\overline{i_{Noise}^2}R = \overline{\left(i_{Total}(t) - i_{FM}\right)^2}R. \qquad [2.38]$$

Signal-to-Noise Ratio (SNR or S/N) is often used to quantify the level of noise contamination of the signal for a given signal level. It can be expressed in units of power or units of current. Which units are used has to be made clear because the expressions differ:

$$SNR_P = \frac{i_{FM}^2}{i_{Noise}^2} = \left(SNR_I\right)^2. \qquad [2.39]$$

It is very important to note that the noise magnitude depends on the signal level. Thus, the usual practice of estimating the noise as the background signal and then calculating the SNR using the FM signal value when the FM signal is quite substantial, can lead to errors. Sensitivity can be calculated using formulas [2.24]-2.26] or [2.34] for the smallest detectable FM signal, which, in term, can be inferred from the smallest SNR that is tolerable for the given experiment.

2.4.1 Definition of Noise Sources in FMS experiment.

In our analysis of noise sources we will treat *shot noise* first. For simplicity we will assume use of cw laser and that there are no other sources of noise present. Although this is clearly an ideal, it sets a benchmark on the expected noise levels against which we will judge other noise sources. In a classical physics picture, shot noise arises from unavoidable fluctuations in rates of arrival of photons onto the detector, or the resulting fluctuations in the number of photoelectrons. The shot noise of the photons giving the optical noise power can be obtained, or, equivalently, the shot noise of all the photoelectrons generated by the laser beam can be calculated. The process of signal detection can be looked at from the point of view of counting these photoelectrons[5]. If an average number of N electrons is counted, the counting error will be (according to the uncertainty of Poisson distribution) equal to $\sqrt{N}$.

The frequency spectrum of shot noise is flat (hence its other name - "white" noise). This means that shot noise power will depend on the detection bandwidth . Thus we can write the shot noise power due to a cw laser of total average power of P_0 as:

$$\overline{i_{SN}^2} = 2eG_{det}^{2+x}\sqrt{\overline{i_{Total}^2}}\,\Delta f = 2eG_{det}^{2+x}P_0\rho\,\Delta f \qquad [2.40]$$

where Δf is detection bandwidth in Hz and x is an additional exponential factor necessary to account for additional shot noise found in the detectors with gain.

Because shot noise power is proportional to the square root of the optical power, when the optical power increases, the fraction of the total current corresponding to shot noise decreases. Taking as an example[5] 1mW of cw laser at 600nm detected with a photodiode having ρ = 0.5A/W, Δf = 1 Hz and G_{det}=1, the total current will be 500μA, while the shot noise current is 1.4×10^{-11}A. It means that if shot noise were the limiting noise source, 1mW of optical power of the laser would provide six orders of magnitude of dynamic range from the noise floor to the total current. This is clearly not the case in the traditional LAS, where laser intensity fluctuations limit the sensitivity of the technique.

Now that we have determined the shot noise contribution to the total noise, we will turn to the noise sources specific to the detection system. At this point we are assuming that no other noise is present.

If the goal is to create a shot noise limited detection system, all other noise sources have to be eliminated, or their noise powers have to be reduced to a level below the shot noise power, since the total noise power in the experiment is the sum of all the individual contributing noise powers. (This, of course, is true only if the noise sources are uncorrelated, which we assume to be the case here.)

Let us separate the detection system into several "blocks" to facilitate the analysis. Figure 2.3 shows three such blocks: receiver (photodetector and preamplifier), demodulator (double balanced mixer) and recorder (lock-in amplifier for cw experiment, boxcar or digitizer for pulsed experiment - the device that sets the bandwidth of the detection system). Each "block" will be considered separately.

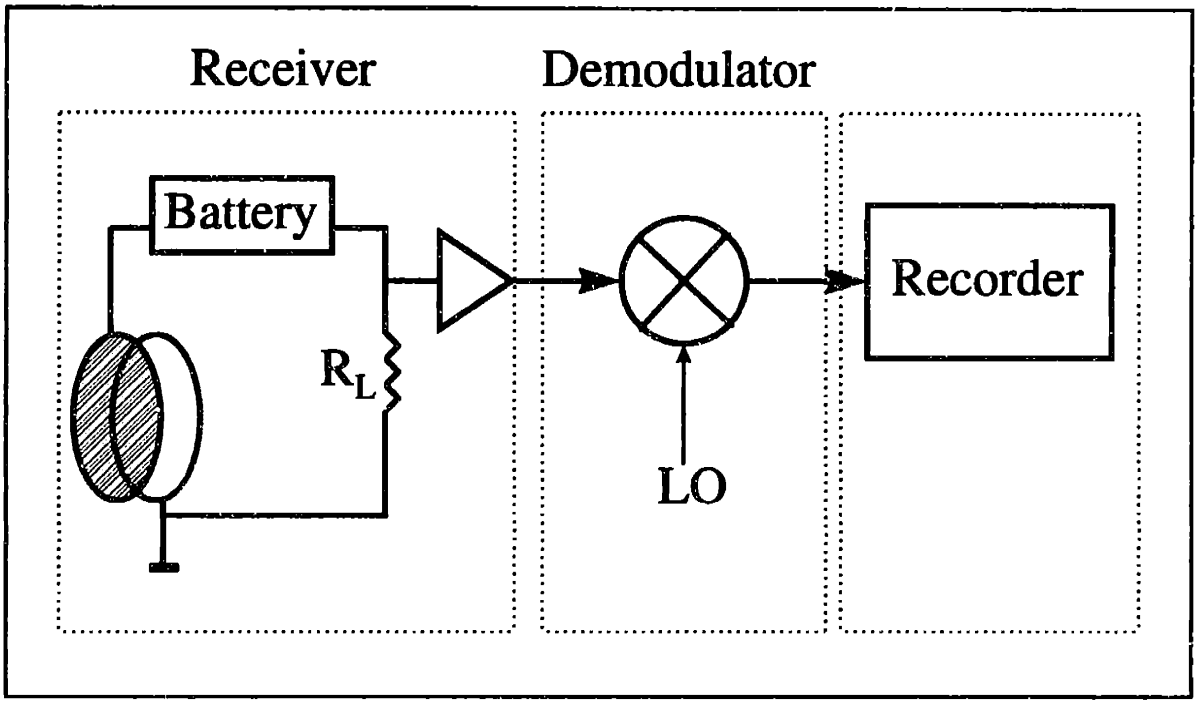


Figure 2.3 FM detection system. Simplified representation.

There have been many discussions of the available optical photoreceivers[4,5,8,9]. Because our discussion is directed at laying the ground for development of the FM enhanced remote sensing technique, we will only touch upon the most promising detectors for this type of application. It is quite clear that a detector with internal gain would be desirable for field measurements, especially if the source of the signal is a backscatter from a diffuse object. Until recently, such a detector was not available, or at least was not available cheaply; PMTs were relatively narrowband and avalanche photodiodes had gain of only 10-50 times. A solution to this problem is the Microchannel Plate Photomultiplier Tube (MCP-PMT). MCPs (for example Hamamatsu's R3809U series) are relatively fast, with typical rise time of 150 picoseconds, sensitive over a broad spectral range (various models covering from 115nm to 1200nm) and have a current amplification on the order of 10^5. Moderately priced (under $1000), they are good candidates for the FMRS detectors.

The standard arrangement is that the current output of a fast photodetector, be it photodiode, biplanar photoelement or MCP, is connected across a 50 Ohm impedance to a fast RF preamplifier. The choice of RF preamplifier is determined by the application. If FM spectroscopy is conducted in the cw arrangement with a very narrowband laser, any standard broadband RF amplifier can be used, as they are commercially available, inexpensive, and easy to operate. If, on the other hand, a pulsed experiment is conducted, one should be careful to use an amplifier that will not distort the pulse shape, as that will result in parasitic frequency generation that may dominate the noise. A good choice in such a case is to use traveling wave amplifiers, which are designed for amplification of short pulses.

Even when there is no light on the photodetector, it still produces some current which is called dark current, i_d. This dark current is due to the thermal generation of electrons. Lowering the operation temperature will lower the dark current. There is a shot noise that is associated with the dark current itself and it must be added to the shot noise due to the light. Shot noise due to both light and dark current is the only significant noise source in the photodetector.

The next source of noise is the motion of thermal electrons in the load resistor on the output of the detector. This statistical motion of electrons generates so called Thermal (also known as Johnson or Nyquist) noise, i_{Th}. Thermal noise, due to its statistical nature, is a white noise, analogous to the shot noise, and can be written as[4,5]:

$$\overline{i_{Th}^2} = \frac{4kT\Delta f}{R_L}, \qquad [2.41]$$

where k is Boltzmann constant and T is temperature.

The next element after the load resistor is the preamplifier. Every amplifier, in addition to providing gain of G_a, adds noise N_a. This can be illustrated by the equation:

$$\frac{(SNR)_{out}}{(SNR)_{in}} = \frac{\dfrac{G_a N_a}{G_a N_{in} + N_a}}{\dfrac{S_{in}}{N_{in}}} = \frac{G_a N_{in}}{G_a N_{in} + N_a} < 1. \qquad [2.42]$$

The reason this device is kept in the detection circuit notwithstanding the fact that it actually decreases the SNR is that some sort of amplification (either external or internal to the recorder) is necessary to record a signal. The reason amplification of the high frequency signal is preferred to amplification after demodulation is that the RF amplifiers are actually lower noise than low frequency amplifiers, and much less expensive as well. A traditional way to describe and quantify an amplifier's noise is to use a parameter known as *effective noise temperature* T_e. Then, the equation for the thermal noise can be modified by substitution of effective noise temperature instead of ambient temperature T:

$$\overline{i_{Th}^2} = \frac{4kT_e\Delta f}{R_L}, \qquad [2.41a]$$

and where effective noise temperature is given by:

$$T_e = T + T_a, \qquad [2.43]$$

where T_a is amplifier noise temperature. Two other characteristics of noise performance of an amplifier are noise factor, F (unitless), and noise figure, NF (in dB). All three of them are related as:

$$NF = 10\log F; \quad F = 1 + \frac{T_a}{290K}. \qquad [2.44]$$

The next element in the detection system is the mixer. From the previous discussion we can recall that only half of the output signal from the mixer is used - the low frequency component. Thus part of the mixer's action is attenuation. On the other hand, the photo-signal is multiplied by the local oscillator wave, which can yield effective amplification. This is why it is advantageous to use the highest level mixer for demodulation of weak signals. Cheap, commercially available mixers can work with as much as 23dBm of local oscillator power. As any other active device, mixer adds noise. If before the mixer SNR is expressed as

$$SNR = \frac{i_{nRF}}{i_{RF}}, \qquad [2.44a]$$

then after the mixer SNR will be

$$SNR = \frac{i_{nRF} + i_{nLO}}{i_{RF}}, \qquad [2.44b]$$

because in the mixer, multiplication of two currents - RF and LO is taking place:

$$(i_{RF} + i_{nRF}) \cdot (i_{LO} + i_{nLO}) = i_{RF} i_{LO} + i_{nRF} i_{LO} + i_{nLO} i_{RF} + i_{nRF} i_{nLO}. \qquad [2.44c]$$

Assuming, for simplification, equality of RF and LO terms, and neglecting of the last term due to its small value, one will obtain expression [2.44b].

The most important role of the recorder, the generic device we introduced to represent lock-in amplifier or transient digitizer, is that it determines the bandwidth of the measurement. Of course, in a pulsed experiment, bandwidth can be determined by the pulse length, but the final result will be the same. The reason the noise added by the recorder is not considered is that, in a properly designed experiment, noise due to the first amplifier is always larger than any added by the following amplifiers, simply because amplified input noise should be greater than the noise of the device itself.

We can summarize the noise contributions of the detection system by writing it as

$$\overline{i_N^2} = \overline{i_{SN}^2} + \overline{i_{amp}^2} + \overline{i_{Th}^2} = 2e\Delta f\left(\sqrt{\overline{i_d^2}} + \rho\, G_{det}^{2+x} P_0\right) + \frac{4kT_e\Delta f}{R_L}. \qquad [2.45]$$

It is quite clear that if one's objective is to design a shot noise limited detection system, the last term in the equation[2.45] should be made small compared to the first one. The reason I am disregarding the dark current is that it is usually very small (~1nA for MCP-

PMT). It means that, for the majority of cases, dark current noise is small compared to the shot noise, except for situations when the intensity of the detected light is very low. This situation should be avoided by all means.

The most important contributions to the noise are laser noise (shot noise and laser intensity fluctuations) and preamplifier noise. To indicate a level of amplifier noise we can take two amplifiers - with noise figure *NF*=1dB and with *NF*=10dB. For the low noise model T_a=75K, while for the noisier one T_a=2610K. It is quite clear that the amplifier with lower noise figure is preferable as a first amplifier in the detection circuit. It is, in principle, possible to build (or to buy) an amplifier with noise figure less than unity, but such an amplifier has to be cooled to achieve specified noise level.

The next, and arguably most important, source of noise is *laser noise*. Shot noise has already been considered and will be excluded from laser noise. The part of optical noise we are going to study is laser intensity fluctuations. Until now we avoided talking about it, because it would be advantageous not only to characterize it, but to discuss how FM spectroscopy deals with it, because FM spectroscopy was developed in an attempt to create an absorption based spectroscopic technique which is immune to laser intensity fluctuations. There is little doubt that the total laser noise $\overline{i_L^2}$ is much larger than the shot noise. The FMS can increase the SNR because it is designed to detect only a small part of the noise. A few attempts have been made to describe the dependence of this part on internal laser noise, signal level and imperfections of phase modulating circuitry[2,3]. The authors tried to determine the source of additional noise (which was, in fact, laser noise "leaking" in) and find ways to minimize it. Various extensions of the FM technique were introduced to eradicate this laser noise leakage[1,2,10-14] . Following is an attempt to describe how laser noise leaks into the FM detection and how it depends on various experimental parameters.

FMS is a zero background technique. It is not very obvious why it is, because if a laser before the modulation has time dependent amplitude and phase that each contain noise, then in the process of modulation this noise is being copied to every resulting frequency component. Phase modulation does not remove any noise. In fact, just as the

average power of light was not changed by the modulation (recall [2.27]), the average noise power is not changed by the modulation either. The zero background results from the perfect cancellation of the corresponding sideband-carrier beats in the absence of sample. Although each of the frequency components has excess laser noise associated with it, this noise is canceled during the detection process if there is no noise present at the modulation frequency (or, in other words, if the modulation frequency is larger than the laser noise bandwidth).

On the other hand, it is now clear that when the FM signal is present, laser noise cancellation is no longer perfect! Therefore, whenever an FM signal is present, it is accompanied by noise due to detection of uncanceled laser noise.

Let us introduce a laser noise characteristic of the laser RMS fractional power fluctuation, N_L. Then, the laser power to noise power ratio (in electrical power units) will be

$$LNR = \frac{1}{N_L^2}. \qquad [2.47]$$

Although we consider electrical power units (which is done for convenience of comparison with other noise sources) *LNR* does not depend on the detector characteristics, and depends only on laser.

Let us assume that we can use simplified FM theory, that the absorption feature is very narrow and interferes with only one sideband, that there are no additional losses, and that the signal due to dispersion can be neglected. For such a case, the FM current expression can be given (from [2.34]):

$$i_{FM} = \frac{Q_{det} M P_0 \delta}{2}, \qquad [2.48]$$

where δ is a fractional absorption experienced by one of the sidebands. This means that the fraction M of the total laser noise will leak through. Therefore, the laser noise that is present in the FM experiment due to detection of the FM signal can be expressed as:

$$\overline{i_{FM-LN}^2} = \frac{i_{FM}^2}{LNR} = \frac{\left[N_L \rho P_0 G_{det} \delta M\right]^2}{4}. \qquad [2.49]$$

This means that whenever an FM signal is detected there will be some associated noise. It also shows that if a parasitic FM signal is present, it will also introduce some laser noise.

These parasitic FM signals can occur even in the absence of sample. They can be generated anywhere along the beam path and two major sources have been identified in previous work[2,11,15]. Therefore we should distinguish between two different unwanted additions to the FM signal caused by the interference with the sample. First is noise and the second is the FM signal caused by anything other than the sample. We will call the latter background. This fact is emphasized here because there is a confusion in the literature, where a distinction between the two is often not made. Moreover, a term Residual Amplitude Modulation (RAM) was used for both the FM signal background and the noise associated with this background. Background signal can be stable, not changing on the timescale of the measurement, and then can be removed by simple subtraction, or it can be time dependent and thus add to the noise of the FMS. A number of authors discussed the possible sources of RAM[2,14] (we use the term in the meaning "background signal"). The most likely source of the frequency dependent RAM lies in multipassing of the EOM (accidental etalons in the EOM)[3], while the origin of frequency independent RAM is not yet clear. Usual range of RAM magnitudes is 10^{-2}-10^{-6} of the laser field[14]. Additionally, background FM signals can be caused by accidental etalons created by optical surfaces the beam passes on the way to the detector.

Whatever the sources of the background signal are, their effect can be described using the same formalism as the FM signal caused by the sample. Thus, following [2.48] we can write

$$i_B = \frac{Q_{det} M P_0 \Delta}{2}, \qquad [2.50]$$

where Δ is the attenuation factor for one sideband due to anything other than sample. Analogously to [2.49], this current will correspond to noise current caused by laser fluctuations expressed as

$$\overline{i_{BFM}^2} = \frac{\left(\frac{\Delta}{\delta} i_{FM}\right)^2}{LNR} = \frac{\left[N_L \rho P_0 G_{det} \Delta M\right]^2}{4}. \qquad [2.51]$$

It is interesting to note that this noise can be decreased by either decreasing the background signal, or decreasing laser noise, or both.

2.4.2 Sensitivity and Signal to Noise Ratio of FM Spectroscopy.

We are now ready to perform signal to noise analysis for the FMS and to find an expression for the smallest detectable signal. Putting together all the contributing noise currents we can write a complete expression for signal to noise as:

$$SNR = \frac{\overline{i_{FM}^2}}{\overline{i_{Noise}^2}} = \frac{\overline{i_{FM}^2}}{\overline{i_{SN}^2} + \overline{i_{amp}^2} + \overline{i_{Th}^2} + \overline{i_{BFM}^2} + \overline{i_{FM-LN}^2}} \qquad [2.52]$$

Assuming that the detection system was properly designed and amplifier noise, thermal noise and dark current noise are dominated by the shot noise from laser light, we will arrive at a simplified result:

$$SNR = \frac{\overline{i_{FM}^2}}{\overline{i_{SN}^2} + \overline{i_{BFM}^2} + \overline{i_{FM-L}^2}} = \frac{M^2\delta^2}{\xi + \frac{M^2\left(\delta^2 + \Delta^2\right)}{LNR}}, \qquad [2.53]$$

where ξ is defined as

$$\xi = \frac{8eG_{det}^{x}(\Delta f)}{\rho\ P_0} \qquad [2.54]$$

and has a physical meaning of the shot noise power to the laser power ratio.

This expression will allow us to determine the sensitivity of FMS - i.e. the smallest detectable signal. To do that we have to assume some acceptable value of SNR, say unity. Then, the smallest detectable absorption δ_{min} is:

$$\delta_{min} = \sqrt{\frac{\xi + M^2\Delta^2 LNR^{-1}}{M^2\left(1 - LNR^{-1}\right)}} = \sqrt{\frac{\frac{8eG_{det}^{x}(\Delta f)}{\rho\ P_0} + M^2\Delta^2 N_L^2}{M^2\left(1 - N_L^{\ 2}\right)}}. \qquad [2.55]$$

For most cases, the laser fluctuations are much smaller than unity (N_L<1), so the sensitivity is influenced by either shot noise or the product of laser noise, background and modulation index. If one attempts to work at the shot noise limit, this means that the background level of laser noise should be decreased to below shot noise level. In this limit (assuming still that laser fluctuations are much smaller than 100%) sensitivity is

largely independent of laser noise. Thus, the expression for the smallest detectable signal in the shot noise limit is:

$$\delta_{\min}^{shotnoise} \approx 2\sqrt{\frac{2eG_{det}^{I}(\Delta f)}{M^2 \rho P_0}}. \qquad [2.56]$$

Similar expressions have been obtained earlier by a number of authors[6,16].

An interesting observation can be made on the basis of equation [2.55]. For the shot noise to be larger than laser noise the bandwidth should be large. On the other hand, the larger the bandwidth (providing that the shot noise limit still applies), the worse the sensitivity! Thus, ideally, the bandwidth would be just large enough for the shot noise to dominate over background-laser noise product. Too large a bandwidth would provide shot noise limited detection, but poor sensitivity at the same time. Of course, in the case of a remote sensing experiment one is limited in the choice of the bandwidth by the necessity to use short pulse (i.e. large bandwidth) lasers.

To estimate signal to noise ratio we have to remember that this ratio is defined as ratio the of typical signal to the noise at this signal level. For FMS it means that the laser noise leaking through the imbalance of the beat notes is going to dominate shot noise and set the noise floor. We will call this case *background limited signal to noise*. From [2.53] :

$$SNR^B = \frac{LNR}{1+\frac{\Delta^2}{\delta^2}}. \qquad [2.57]$$

This result points out a very interesting feature of FMS. In the background limited case, when the detected signal is not vanishingly small, the SNR is never going to be better than laser noise (LNR). Even more so, if background value is relatively large, the SNR becomes even much worse than the LNR.

2.4.3 Comparison of LAS, FMS and Fluorescence Detected Spectroscopy.

For simplicity we shall divide this analysis into two parts - first comparing FMS with ordinary direct laser absorption spectroscopy (LAS), and then comparing FMS with fluorescence detection (FD).

When comparing two absorption based techniques we should remember that the magnitude of signal produced by the same absorption feature will differ from one technique to the other. In the case of LAS we can write usual Lambert-Beer's Law:

$$I = I_0 e^{-\delta} \qquad [2.58]$$

The signal will be proportional (for simplicity we shall omit the coefficient of proportionality, which is detector response for both techniques, by making it equal to unity) to the difference between I and I_0.

For FMS, signal magnitude depends on several factors: lineshape of the absorption feature, ratio of modulation frequency to the linewidth and bandwidth of the laser. A very good analysis of this can be found elsewhere[6]. The magnitude of the FM signal may de decreased relative to the LAS signal due to partial absorption of the carrier by the sample, or, even by the simultaneous absorption of the "competing" sidebands. To take all this into account we will introduce an attenuation factor $\chi \in [0;1]$ of FM signal reduction relative to LAS.

It is generally accepted that the dominant noise source for LAS is always laser noise. Recalling [2.47], we can use LNR to characterize SNR for a background limited LAS case:

$$SNR^{LAS} = (1 - e^{-\delta}) LNR \approx \delta LNR, \qquad [2.59]$$

where the second part of the equation is valid for small absorptions. If we assume that SNR=1 is acceptable, then the smallest absorption detected by LAS is

$$\delta = -\ln(1 - LNR^{-1}) \quad \text{- in the general case}$$

or [2.60]

$\delta = LNR^{-1}$ - for small absorptions.

Taking the worst case of background limited FMS (equation [2.57]), we have for the smallest detectable absorption:

$$\delta_{\min}^{FM,B} = \frac{\Delta}{\chi\sqrt{LNR-1}}. \qquad [2.62]$$

It is clear that FMS has much better sensitivity than LAS unless attenuation factor, χ, is very small (see Fig. 2.4).

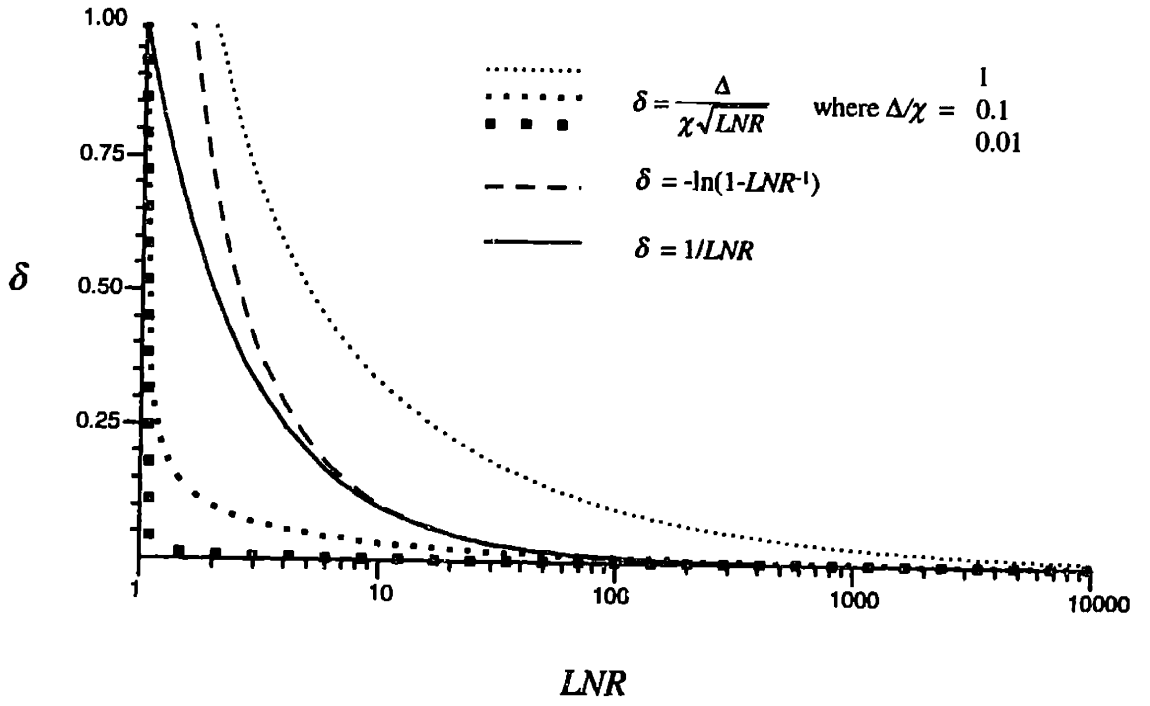


Figure 2.4 Comparison of background limited sensitivities of FMS and LAS.

Most often, FMS will be preferable to LAS even for extremely large FM background values.

Comparing FMS with FD is not as straightforward as both techniques can have shot noise limited as well as background limited sensitivity. Considering weak signals for simplicity we can write, for shot noise limited FD:

$$SNR_{SN}^{FD} = \frac{\delta\phi}{\xi}, \qquad [2.63]$$

where δ is the absorption coefficient defined earlier, ϕ is a fluorescence yield (which is a function of both fluorescence quantum efficiency and of the specific apparatus used for detection) and ξ is defined by equation [2.54]. For FMS in the shot noise limit (from [2.53]):

$$SNR_{SN}^{FM} = \frac{M^2\chi^2\delta^2}{\xi} \qquad [2.64]$$

if the laser noise is not taken into account at all. If both techniques are shot noise limited, then FMS is more sensitive than FD if

$$M\chi > \frac{\phi}{\sqrt{\xi}} \qquad [2.65]$$

Due to the fact that $(\xi)^{-0.5}$ is very large (for a 1mW cw laser at 600nm it is around 10^7) in the shot noise limit FMS will be better than FD only when the fluorescence yield is low and the fluorescence signal is poorly detected (which of course are both true for fluorescence detected remote sensing).

In the background limited case, SNR for FD is

$$SNR_{BL}^{FD} = \frac{LNR}{1 + \frac{\Delta_f^2}{\delta^2\phi^2}}, \qquad [2.66]$$

where Δ_f^2 is the factor describing fluorescence background, analogously to expression for FMS derived earlier (see [2.57]). FD remote sensing has an additional important source of background, so strong that most experiments cannot be run during the daytime because of solar background flux. The FMS background limited result is given by [2.62]. Compared to the smallest detected FD signal given by

$$\delta_{\min}^{FD-BL} = \frac{\Delta_f}{\phi\sqrt{LNR-1}}, \qquad [2.67]$$

FMS will have an advantage in remote sensing applications due to significantly smaller background (less interference with ambient light).

References:

1. Ye, J. Ma, L.-S., Hall, J., "Ultrasensitive detections in atomic and molecuar physics - demonstration in molecular overtone spectroscopy", manuscript in preparation.
2. Gehrtz, M., Bjorklund, G., Whittaker, E. "Quantum-limited laser frequency modulaiton spectroscopy", *J. Opt. Soc. B*, **2**, 1510 (1985).
3. Whittaker, E., Gehrtz, M., Bjorklund, G., " Residual amplitude modulaiton in laser electro-optic modulation", *J. Opt. Soc. B*, **8**, 1320 (1985).
4. Rieke, G.H., *Detection of Light: From the Ultraviolet to the Submillimeter.*, Cambridge University Press, 1996.
5. Yariv, A. *Optical Electronics*, 1991, Saunders College Publishing.
6. Bjorklund, G.C., Levenson, M.D., Lenth, W., Ortiz, C., "FM spectroscopy: Theory of Lineshapes and Signal to-Noise Analysis", *Appl. Phys. B*, **32**, 145 (1983).
7. Silver, J.A., Bomse, D.S., Stanton, A.C., "Frequency-modulaiton absorption spectroscopy for trace species detection: theoretical and experimental comparison among methods" in Optical Methods for Ultrasensitive Detection and Analysis: Techniques and Applications, *Proc. Soc. Photo-Opt. Instrum.*, **1435**, 64 (1991).
8. Gower, J., Optical Communication Systems, Englewood Cliffs, N.J., Prentice-Hall, 1984.
9. Keiser, G., Optical Fiber Communications, New York, McGraw-Hill, 1983.
10. Levenson, M.D., Moerner, W.E., Horne, D.E., "FM Spectroscopy detection of stimulated Raman gain", *Opt. Lett.*, **8**, 108 (1983).
11. Wong, N.C., Hall, J.L., "Servo control of amplitude modulation in frequency-modulation spectroscopy: demonstration of shot-noise limited detection.", *J. Opt. Soc. Am. B*, **9**, 1527 (1985).
12. Whittaker, E.A. et al., "ND_4 Schuler band absorption observed by laser FM spectroscopy in a photochemical reaction", *J. Chem. Phys.*, **80**, 961 (1984)
13. Whittaker, E.,A., et al., "Laser FM spectroscopy with photochemical modulation: a sensitive high resolution technique for chemical intermediates." *Appl. Phys. B*, **35**, 105 (1984).

14. McCarthy, M.C., Bloch, J.C., Field R.W., "Frequency-modulation enhanced magnetic rotation spectroscopy: a sensitive and selective absorption scheme for paramagnetic molecules", *J. Chem. Phys.*, **100**, 6331 (1994).
15. Levenson, M.D., Kano, S.S., *Introduction to nonlinear spectroscopy*, Boston: Academic Press, 1988.
16. Bloch, J.C., "Extending Frequency Modulation Spectroscopy: Sensitive and selective High Resolution Spectroscopy in the Visible and Ultraviolet", *Ph.D. Thesis*, M.I.T. 1996.

2.5 Pulse Amplified Frequency Modulation Spectroscopy.

Employing pulsed, frequency modulated laser radiation is crucial to the success of the application of FM spectroscopy to remote sensing applications. This requirement is caused by the very low efficiency of light backscattering by aerosols and clouds. In addition, in all shot noise limited experiments, in FM studies of samples with very high optical density (that would absorb all the light from the cw laser in a very short distance), and in studies of processes taking place in highly scattering media, detection sensitivity would improve with an increase of laser power, so that pulsed operation of an FM spectrometer would be desirable. In a series of experiments conducted in our lab in collaboration with Prof. E. Eyler (Univ. of Connecticut) we have developed and successfully demonstrated a pulsed FM spectrometer[4,5,6].

Prior to our work, the whole approach to producing pulsed frequency modulated radiation was totally different. The first attempt was made by Gallagher and co-workers[1] who in 1981 produced pulsed FM radiation by directly modulating the output of an etalon-narrowed pulsed dye laser. The sensitivity of the spectrometer was sufficient to detect 2% absorption after averaging 10 shots. The original work employed modulation frequencies of 1-2GHz. Subsequent attempts to improve the sensitivity[2] employed even higher (4GHz) modulation frequency, but the SNR remained poor, thus limiting the applicability of the technique to relatively strong absorptions. The work by Tran and co-workers[2] made another very important step in the development of FM spectroscopy - it was the first work in which frequency modulated light was subsequently frequency mixed to obtain radiation in the UV. The final work by the same group[3] determined that in order to improve the sensitivity they had to increase the modulation frequency to 12GHz, and to narrow the bandwidth of the laser. They stated that the primary sensitivity parameter is the ratio of the modulation frequency to the laser bandwidth, and that FM spectroscopy employing large modulation frequencies is best suited for observation of broad spectral features. Both of these conclusions are easy to understand from our discussion in Section 2.3.

It is quite apparent that the major obstacle these researchers encountered was that they used lasers that had relatively broad (>1GHz) bandwidth. If the frequency of modulation is not high enough, the spectral wings of the sidebands and the carrier will overlap, thereby making the perfect amplitude and phase balance required for full FM beat cancellation impossible.

The solution to the problem we found was to create narrow (~50MHz) transform limited frequency modulated pulses by using a system in which a chain of amplifiers is seeded by the frequency modulated radiation of a single longitudinal mode cw laser. Another significant improvement came from understanding that the output of the amplifier is a convolution of an input (seeding) frequency modulated cw light field and of the pump laser field. In other words, the bandwidth of laser radiation after the amplifier will be that of the pump laser. Therefore, the pump laser used is a relatively long pulse (15-17nsec) , injection seeded Nd:YAG laser that produces essentially transform limited pulses. A thorough description of the experimental setup and discussion of this particular implementation of the technique is given elsewhere[4,5,6]. An important improvement to the setup reported in the references above would be to find a way avoid one of the major sources of noise in the pulse amplification - amplified spontaneous emission (ASE). This can be accomplished by replacing the Bethune[7] amplifier cells by transversely pumped cells in which the amplified laser beam experiences total internal reflection[8] from the cell window (see Figure 2.5).

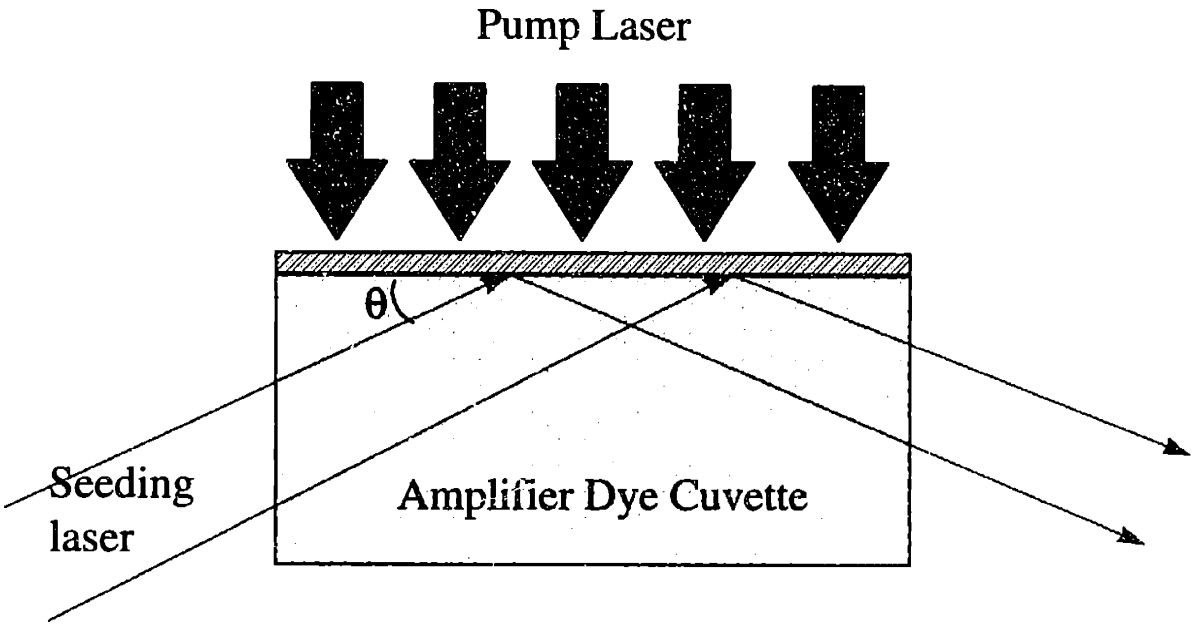


Figure 2.5 ASE-free amplifier cell. Angle θ is angle of incidence of totally internally reflected beam.

The reason this arrangement provides virtually ASE-free amplification is that while the amplified seeding beam exits the cell at an angle $\pi/2 - \theta$ to the front window, the ASE propagates in the direction perpendicular to the pump laser propagation direction - i.e. parallel to the window. It is, therefore, easy to separate them. Multistage amplification systems are capable of producing significant powers and broad spectral range through the use of frequency mixing.

One of the major drawbacks of the use of any cw dye laser based system in the field conditions would be its high cost and very high complexity of operation and maintenance. Solid state laser based systems are likely to be more suitable for remote sensing applications, although no such system (except for a Ti:Sapphire system which is analogous to the dye system described) is currently available.

References

1. Gallagher, T.F.,Kachru, R., Gounand, F., Bjorklund, G.C., Lenth,W., "Frequency modulation spectroscopy with pulsed laser", *J. Opt. Soc. Am.*, **7**, 28, (1981).
2. Tran, N.H., Kachru, R., Gallagher, T.F., Watjen, J.P., Bjorklund, G.C., "Pulsed frequency modulation spectroscopy at 330nm", *J. Opt. Soc. Am.*, **8**, 157, (1983).
3. Tran, N.H., Kachru, R., Pillet, P., van Linden van den Heuvell, H.B., Gallagher, T.F., Watjen, J.P., "Frequency-modulaiton spectroscopy with a pulsed dye laser: experimental investigations of sensitivity and useful features", *Appl. Opt.*, 23, 1354, (1984).
4. Eyler, E.E., Gangopadhyay, S., Melikechi, N., Bloch, J., Field, R. W. "FM spectroscopy with transform-limited nanosecond pulses", *Opt. Lett.*, **21**, 225, 1996.
5. Nicolov, A., Cheng, C., Eyler E., Dubinsky, I., Field, R.W., "Cavity enhanced FM spectroscopy with pulsed lasers", *OSA Ann. Meeting*, 1997.
6. Bloch, J.C., *Ph.D. Thesis*, MIT, (1996).
7. Bethune D.S., "Dye cell design for high-power low-divergence eximer-pumped dye lasers", *Appl. Opt.*, **20**, 1897, (1981).
8. Kachanov, A., private communication.

2.6 Experimental Study of Scattering Properties of Frequency Modulated Light.

When attempting to develop a remote sensing technique based upon Frequency Modulation spectroscopy which is applicable when measurements require collection of light backscattered by large diffuse objects, one must understand how modulated light, scattered from different points in space, will interfere on the detector. When there is no single point scattering object, a superposition of scatterers that are distributed over a large volume (cloud, plume of smoke, etc.) must be taken into account.

The simplest model would be to introduce two point scatterers and to observe the dependence of the FM signal on their relative positions. Consequently, it will be simple to extend the result of this experiment to an arbitrarily large number of scatterers.

2.6.1 Experimental Apparatus.

The following figure (Fig. 2.6) is a general schematic of the experiment we undertook with the goal of understanding the character of the aforementioned interference effect and its influence on the FM signal recovery.

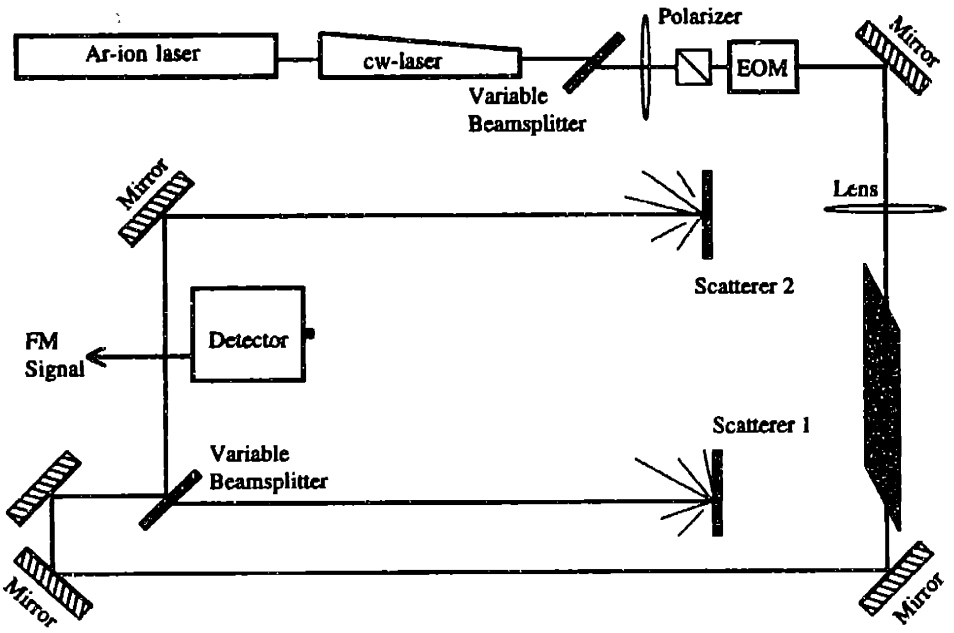


Figure 2.6. Schematic of the experimental setup for demonstration of the coherence of the FM signal.

An argon ion laser (Coherent Innova-100) was used to pump an actively stabilized, single longitudinal mode (>1MHz bandwidth) cw dye laser (Coherent, Inc., Model 899-29), lasing around 596nm. The output of the dye laser, after first passing through a beamsplitter, a variable attenuator and a linear polarizer, passed through an electro-optic modulator (EOM) (New Focus, Inc., Model 4421, standing wave, tunable 500-1500MHz). The linearly polarized output of the polarizer was precisely aligned along the principal optical axis of the EOM crystal, to minimize residual amplitude modulation (RAM). The EOM was driven at ω_{rf} of 860MHz, imposing sidebands at frequencies $\Omega \pm n\omega_{rf}$, where Ω is the laser carrier frequency and n is the order of sideband. The amplification of the RF signal applied to EOM crystal was chosen such that the intensity of each of the first order sidebands was on the order of 20% of the carrier intensity (modulation index M=0.8), and the intensities of sidebands of higher orders were negligible. The depth of modulation and the optical mode were monitored with optical spectrum analyzer (Tropel, Model 240). The variable attenuator was used to prevent photorefractive damage of the EOM from happening by keeping the power of the light

entering the crystal below 60 mW. After the EOM, the beam, collimated by a telescope, was sent through a 30 cm cell containing iodine at room temperature. As the laser frequency is scanned preferential attenuation of one sideband relative to another, caused by the iodine absorption lines, generated FM signal. After the cell, the light beam was divided by a variable beamsplitter into two beams of balanced intensity. These beams were scattered from two white paper cards and the scattered light was collected on a fast photodetector (New Focus, Inc., Model 1601, 1GHz bandwidth). Photocurrent was amplified by two low noise broadband amplifiers (New Focus, Inc., Model 1422 traveling wave, and MiniCircuit, Inc., Model ZEL-0812LN), and filtered by RF bandpass filter (Microwave Filter Company, Inc., Model 3278RF(4)-860), centered at 860MHz, 6MHz bandpass. Thus conditioned signal was sent to a double balanced RF mixer (MiniCircuit, Inc., Model ZMY-2). The demodulated signal, after passing through a low pass filter (5MHz cut-off) was sent to a lock-in amplifier (Stanford Research Systems, Inc., SR510), synchronized to the amplitude modulation of the laser beam introduced by a chopper. The output of the lock-in amplifier was read by a PC computer, digitized by an A/D card and recorded by the standard Autoscan (Coherent, Inc.) data acquisition software.

The main goal of this experiment was to find a dependence of the FM signal on the path differences between the two beams of light. This information will allow us to understand how frequency modulated light scattered from different points in the cloud and, therefore, traveling different pathlengths interfere on the detector.

To avoid influence of other experimental conditions on the magnitude of the FM signal, special care was taken. Below, several important steps are described.

First, to be able to maintain a linear response of the detection system, the amplifiers and the mixer should be operating far below their saturation limits. With the laser tuned for the largest FM signal, the input and output RF power levels of all amplifiers were measured and verified to be below the saturation limits specified in the manufacturers' catalogs. Then, the linearity of response was verified by attenuating the input by known amounts and observing an identical attenuation of the output. A similar procedure was performed with the mixer. To avoid saturation, both the local oscillator and the signal inputs were kept below levels specified by the manufacturer. A level 17

mixer (17dBm nominal local oscillator input) was chosen to increase the output level and to improve signal to noise. All RF power and frequency measurements were conducted using an RF spectrum analyzer (Hewlett-Packard, Model 8502). The following figure (Fig. 2.7) shows the radio frequency part of the setup.

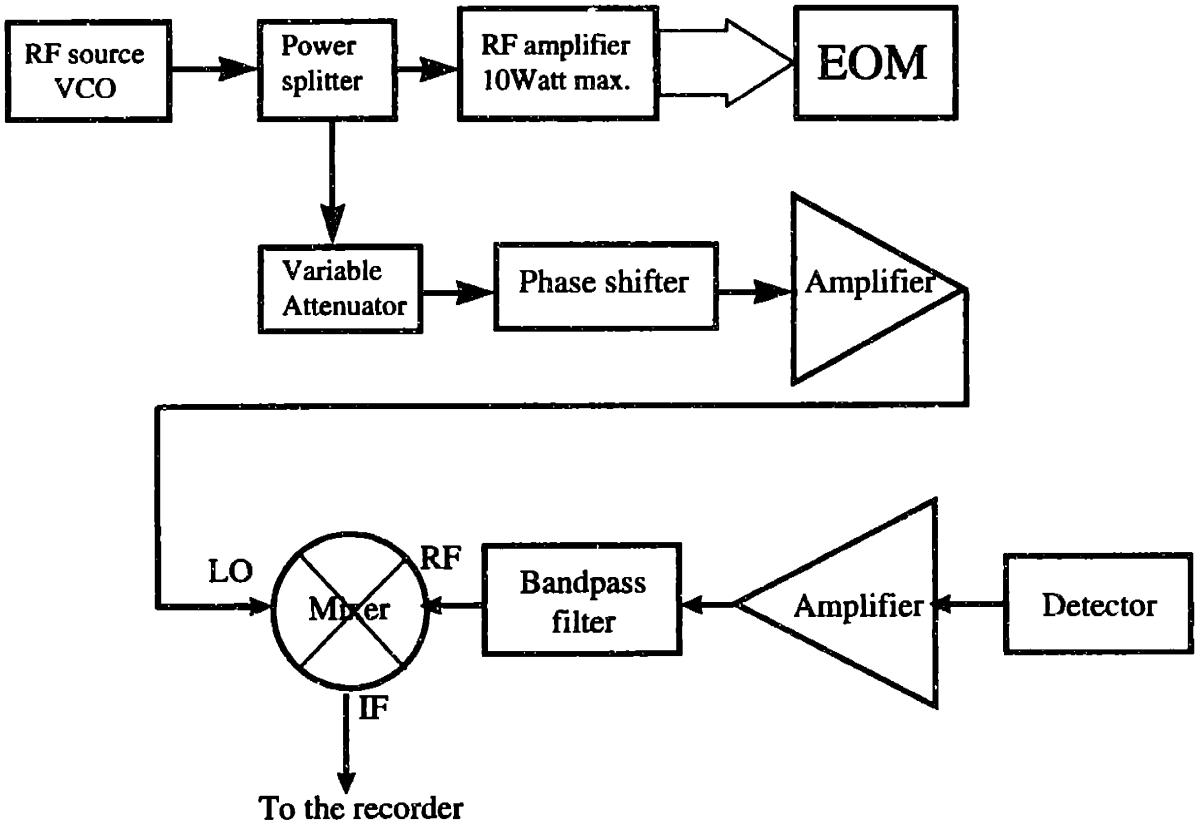


Figure 2.7. Radio frequency subsystem of the setup.

Prior to analysis, all of the data were normalized with respect to laser intensity fluctuations and the sensitivity settings of the lock-in amplifier.

2.6.2 Results and Discussion.

A typical FM scan of the iodine spectrum between approximately 16770 and 16771 cm^{-1} (or between 502680 and 502710GHz) is shown in Figure 2.8a,b. One scan was recorded at each position of the scatterers, which were white pieces of cardboard intended to model two point-scattering objects. The positions of the scatterers were changed for each scan in order that the dependence of the magnitude of the FM signal on the difference in the optical paths, that the two beams of frequency modulated light had to travel, could be found. The range of laser scan was not exactly repeated from time to time due to the laser frequency jumps. On the other hand, this interval was long enough (1 cm^{-1}), so that there were several peaks in the middle of the frequency range that were present in every scan.

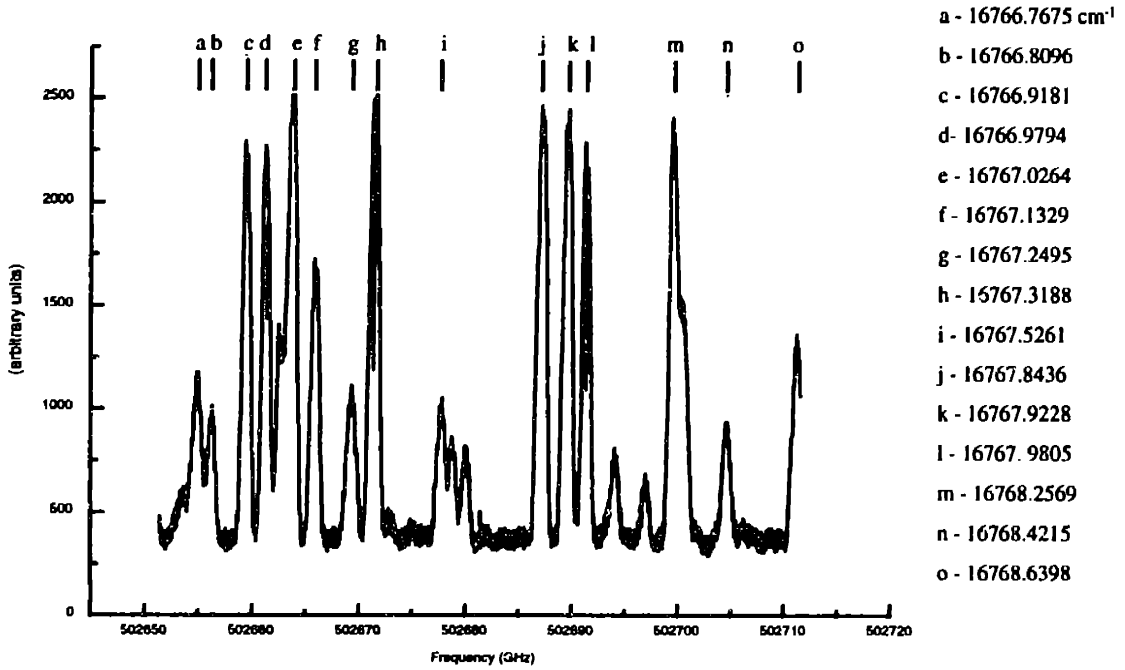


Figure 2.8a Iodine emission spectrum, recorded simultaneously with the FM spe< ım for calibration. Same spectral interval as in Fig. 2.8b.

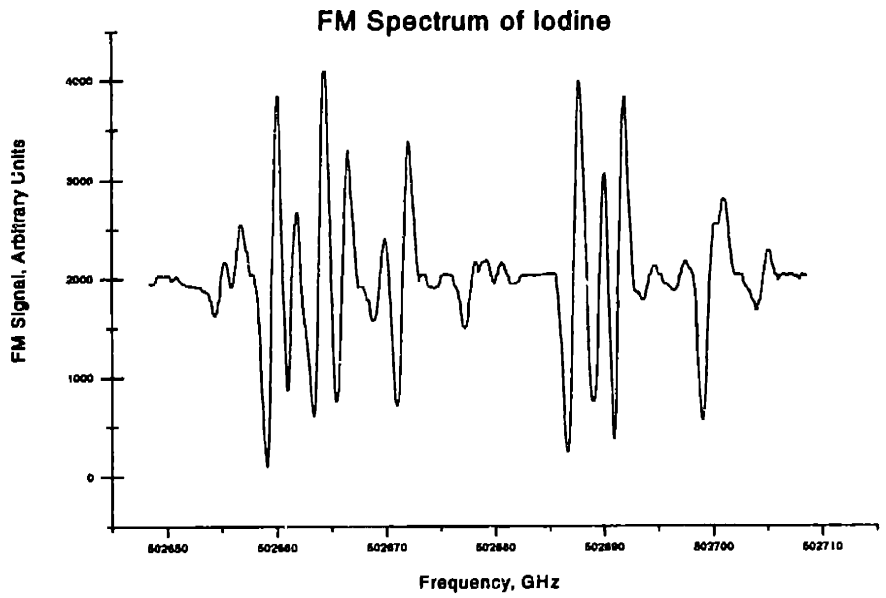


Figure 2.8b FM spectrum of iodine obtained for equal beam pathways.

The amplitude of the peak at 16767.8436cm^{-1}, after normalization and adjustment for laser intensity fluctuations, was used as a measure of the FM signal magnitude. Results are presented in Figure 2.9. It is clearly evident that as pathways are varied, the resulting FM signal (FM photocurrent from the detector) changes. Figure 2.9 shows signal level versus light pathlength difference. Within experimental error (which includes the uncertainty in pathlength determination for each of the two paths, imperfect balance of light power sent onto each scatterer and collected by the detector from each scatterer, laser power fluctuations during the experiment and residual parasitic FM signal), the FM signal exhibits an interference pattern which corresponds to the RF frequency (860MHz) used in the experiment. The error bars along the abscissa are not shown in order to minimize clutter. The uncertainty in distance measurements is estimated to be about 1 cm. The sinusoidal line in Figure 2.9 is a prediction of signal behavior based on a model which treats the two scattering points as sources of electromagnetic radiation oscillating at the FM modulation frequency.

This experiment shows that the photocurrent produced by the amplitude modulated light from two sources depends on the phase difference of the amplitude

envelopes of these two light components. This observation means that when the amplitude modulated light wave is scattered from a large number of particles present in the air, the detector current components oscillating at ω_{RF}, caused by light that traveled different distances, will interfere. There will be no interference of the light waves themselves[1].

We need to find a way to properly account for this photocurrent interference effect in order to be able to assess the plausibility of applying FM spectroscopy to the remote sensing of atmospheric molecules when the FM signal is arises from scattering by distributed atmospheric objects.

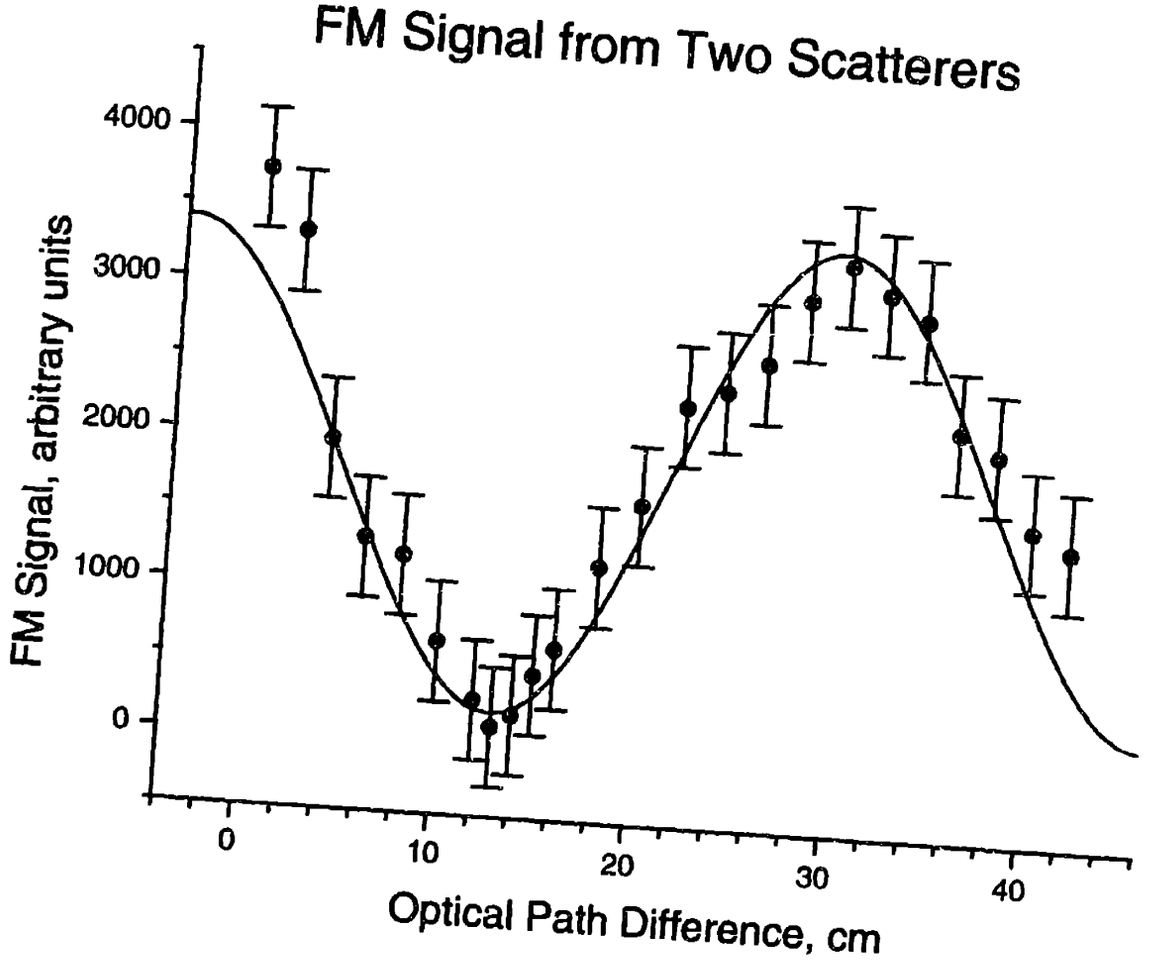


Figure 2.9 Dependence of the FM signal magnitude on the difference in pathlengths of F light scattered by two point scatterers. Solid curve is model calculation for $\omega_{RF} = 860\text{MHz}$.

References

1. Van de Hulst, H.C., *Light Scattering by Small Particles,* Dover Publications, 1981

2.7 Theory of FMRS - frequency modulation enhanced laser absorption remote sensing.

In order to understand how scattering from extended objects (such as clouds, etc.) influences the frequency modulated light we will introduce the following model. A light wave, E_{source}, is emitted from a source, in a well collimated beam, toward a large cloud positioned at x_c. We will consider what happens to light scattered toward the detector by a single particle inside the cloud. Let us assume that the air between the source and the cloud does not contain any aerosols or molecules that would cause absorption. Then, as light travels to the cloud and from the cloud back to the detector, it is attenuated as a result of Rayleigh scattering. Each leg of travel in the air will introduce an attenuation of $\exp(-A^R x_c)$, where A^R is the Rayleigh scattering coefficient. Inside the cloud, on its way to the particle scatterer which we choose to consider in our model, light is being attenuated by the process of Mie scattering. Analogously, light backscattered from the particle experiences attenuation due to Mie scattering, while traveling through the cloud on the way to the detector. This will introduce a term of $\exp[-2A^{Mie}(x - x_c)]$, where A^{Mie} is the attenuation due to Mie scattering. In general, the detector will be in a different location than the source. so the scattering efficiency should include an angle dependence as well as a dependence on the character of the scatterer: size, shape, index of refraction, etc. Lastly, the expression for the field on the detector should contain information about the phase of each component of the scattered radiation.

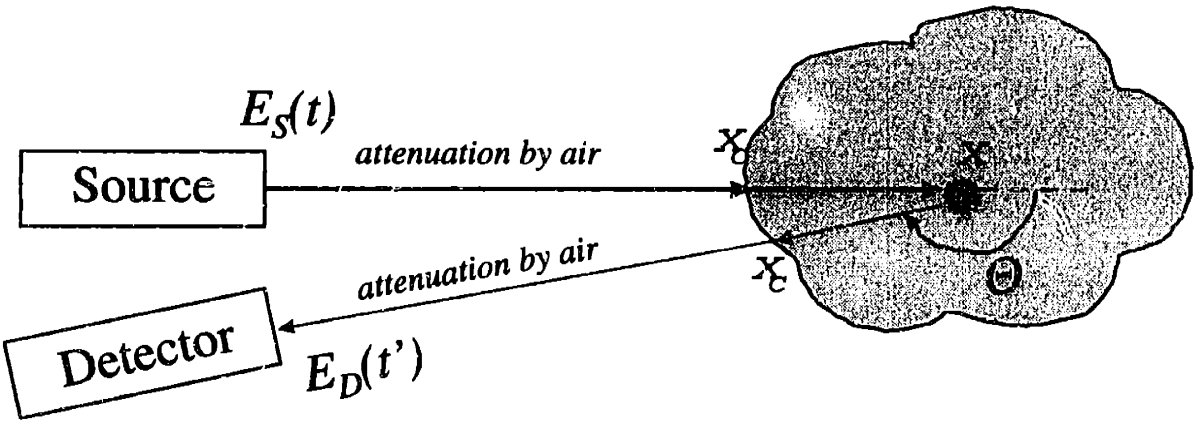


Figure 2.10 General representation of the remote sensing model.

Although absorption and dispersion of the sidebands and the carrier that produce the FM signal can happen at any time during the traversal of the light from the source to the detector, the continuum interaction case is equivalent to the case where all of the differential absorption and dispersion is localized, and assumed to occur within a small cell with equivalent concentration-times-path length. In this way we can separate our treatment of scattering from that of FM signal generation.

The "Source" shown in Figure 2.10 is a combination of a light source (narrow-band laser, amplifier), phase modulating apparatus (EOM) and a "sample cell", i.e. the source of a narrow-line absorption/dispersion which causes imbalance of the sidebands and, therefore, amplitude modulation on FM signal at ω_{RF}. Thus, we can write the electric field at the source as:

$$E_{FM}(t,x) = E_c\left[T_{-1}\frac{M}{2}e^{i(\Omega t-\omega t-k_{-1}x)} + T_0 e^{i(\Omega t-k_0 x)} - T_1\frac{M}{2}e^{i(\Omega t+\omega t-k_1 x)}\right]\cdot P(t), \quad [2.29]$$

where $P(t)$ is a pulse shape. What happens to the light field in the scattering process?

2.7.1 Scattering of Frequency Modulated Light by a Diffuse Object .

Prior to the experiment described in the preceding chapter, we had no information about how backscattered light interferes on the detector. Three plausible outcomes are: (1) coherence of the backscattered light is fully preserved, (2) backscattered light loses coherence, but the amplitude modulation it acquired on its way to the detector retains its coherence, (3) all coherence is lost. The first case can be excluded from consideration as it is well known[2] that regular interference patterns are never observed in the scattered radiation. Moreover, because in the case of atmospheric observations the scattering elements (molecules, particles) are very small, vary significantly in size, and are randomly distributed, the resulting speckle pattern is very fine and the detector surface will therefore be treated in our analysis as uniformly illuminated. The effect of speckle on noise level of the acquired signal on a detector with active area larger than one speckle cell in size will be discussed later.

Our experiment allows us to discriminate between cases (2) and (3). Not only does it prove that the FM signal coherence is preserved, but suggests a viable way of describing the backscattering of frequency modulated light. Heterodyne detection allows us to choose to see only the ω_{RF} modulated component of the photocurrent. The interference pattern obtained in our experiment can be adequately described by interference of two sinusoidal signals, with wavelengths equal to the modulation wavelength, and with a phase difference resulting from the difference in the paths traveled by light. Indeed, this observation is easy to understand if we remember our derivation of the FM signal in Section 2.3 (recall formulae [2.4]-[2.30]). In the present derivation we choose to represent light frequency components as retarded waves, dependent on both time and position in space. The resulting FM signal is a photocurrent oscillating at the modulation frequency, whose phase is related to time and position as if it were an RF wave emitted from the source and detected by the detector in the same way that laser light might have been. This means that calculation of the photocurrent caused by frequency modulated light scattered from a large diffuse scatterer is equivalent to the

treatment of interference of two monochromatic electromagnetic waves of the corresponding frequency. Therefore, we can adequately describe the behavior of light scattered by numerous particles (distributed mirror) by considering the phase shifts accumulated by the sine waves, oscillating at ω_{RF} and traveling from the source to each scatterer and back to the detector, integrated over all possible scatterer positions. The simplified mental picture illustrating this approximation is that we are just following the evolution of the amplitude modulated intensity envelope in time and space. Optical coherence of the scattered light does not matter because light itself is scattered by air and aerosols incoherently.

This allows us to assume that the source produces the wave (which is equivalent to the intensity envelope) that we can represent as $\cos(\omega_{RF}t - kx)$. Therefore, the analysis will be equivalent to the treatment of coherent scattering of the monochromatic electromagnetic wave. The ideal monochromatic cw source generates electromagnetic radiation which can be expressed as:

$$E_{source} = A\cos(\omega t - kx). \qquad [2.68]$$

On the other hand, we are dealing with a pulsed laser source and, therefore, have to incorporate this fact into the model. For simplicity, we will consider a square pulse of "light" emitted from $x=0$ at time $t=0$. This means that $E_{source} \neq 0$ if $ct - x \geq 0$, or is equivalent to multiplying E_{source} by the Heaviside function $H(ct-x)$. This pulse will be propagating along the x-axis until, at x_c, it meets the cloud. The electric field seen at time t by the detector, situated at x, will depend on the position in the cloud from which this light was backscattered. For example, if time t and detector position x are such that

$$t \leq \frac{2x_c - x}{c}, \qquad [2.69]$$

then the detector is still too far from the cloud to see any light - in other words $E_{out} = 0$. This means that the condition for the detector to see light (i.e. $E_{out}(x,t) \neq 0$) scattered from an infinitely thin slice of cloud a distance of δ away from x_c is $ct - 2x_c - 2\delta + x \geq 0$, which is equivalent to introducing $H(ct - 2x_c - 2\delta + x)$ into the expression for the field.

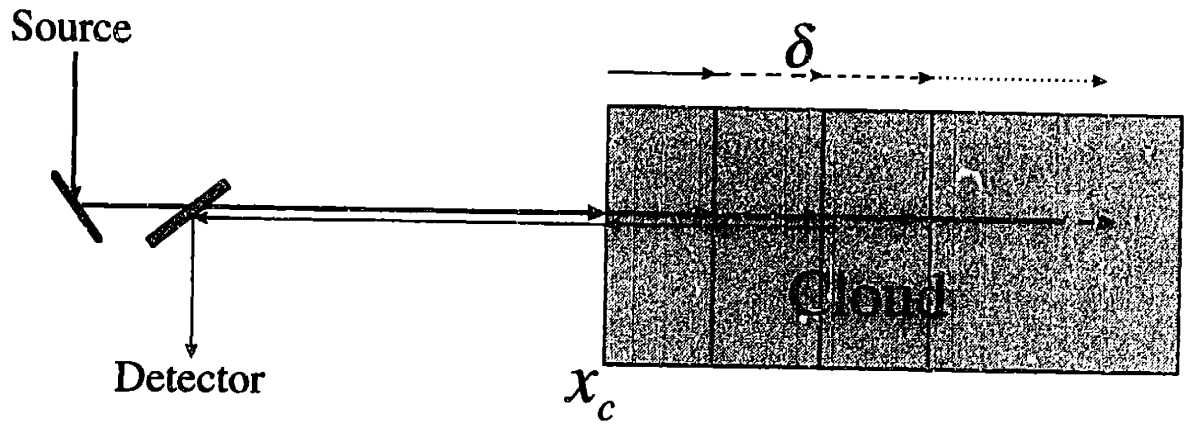


Figure 2.11 Distributed scatterer model.

This Heaviside function will set the upper limit of integration over the cloud of scatterers: the most distant point in the cloud that is a source of light seen by the detector positioned at x at time t. While the pulse of light is still entering the cloud and is not fully inside, the integration has to be performed over the values of δ ranging from zero (cloud surface) to the ones determined above. For a pulse of finite length , we must also determine the lower integration limit for the times starting from the moment when the trailing edge of the pulse reaches the cloud. Since that time, only a part of the cloud lying between the edge and the just defined upper limit of integration will be illuminated, thus producing the backscatter. If the pulse duration is τ, then the lower limit of integration is

$$t-\tau < \frac{2x_c + 2\delta - x}{c}, \qquad [2.70]$$

and is equivalent to incorporating $H(2\delta - x - 2x_c - c(t-\tau))$ into the integrand.

In order to incorporate the phase shift that the wave experiences traveling to the scattering point and back to the detector, we can imagine the scattered wave to be emitted toward the detector from the point in the cloud x_{sc}

$$E_{out}(x,t) = \cos(\omega(t - \frac{x_{sc}}{c}) + k(x - x_{sc})), \qquad [2.71]$$

where x and t are the coordinate of the detector and the time of detection. Expressing the x_{sc} via x_c and δ we obtain

$$E_{out}(x,t) = \cos(\omega t + kx - 2kx_c - 2k\delta). \qquad [2.72]$$

The equation above could also be obtained from a simple consideration that, once we assume that the phase change is solely due to light's travel through space, the distance light travels before it reaches the detector is $2x_c + 2\delta - x$.

We will consider that the cloud starts at x_c and has an infinitely remote other border. Following is the integral over such a cloud of uniformly distributed scatterers, which describes the electric field on the detector positioned at the coordinate x at time t.

$$E_{out}(x,t) =$$

$$= \int_0^\infty \left[\begin{array}{l} a(\delta)H(2\delta - x - 2x_c - c(t-\tau))H(ct - 2x_c - 2\delta + x)\cdot \\ \cos(\omega t + kx - 2kx_c - 2k\delta)d\delta \end{array} \right]. \qquad [2.73]$$

The Heaviside function $H(ct - x)$ is omitted from the integral because its only meaning is that the detector will not see any light before light reaches it, a relatively trivial fact. The coefficient $a(\delta)$ is a scattering efficiency of a cloud slice at a distance δ from the cloud edge. As we assume the cloud to be uniform, $a(\delta)$ can be taken outside the integral and substituted by $\alpha = a_{average}L$, where L is the length of interval of integration.

It is easy to see that Heaviside functions will determine the integration limits. The upper limit is therefore

$$\delta < \frac{ct + x}{2} - x_c. \qquad [2.74]$$

The lower limit will depend on whether the entire light pulse or just part of it has entered the cloud. In the second case the lower limit of integration will be $\delta = 0$, the edge of the cloud. In the first case the lower limit will be

$$\delta > \frac{(t-\tau)c + x}{2} - x_c. \qquad [2.75]$$

The integral is equal to

$$E_{out}(x,t) = \frac{\alpha}{2k}\sin(2k\delta + 2kx_c - kx - \omega t)\Big|_{lower}^{upper}. \qquad [2.76]$$

Therefore, in the case of a light pulse still entering the cloud:

$$E_{out}(x,t) = \frac{\alpha}{2k}[\sin(kct - 2kx_c + kx + 2kx_c - kx - \omega t) - \sin(2kx_c - kx - \omega t)] =$$

$$= \frac{\alpha}{2k}\sin(\omega t + kx - 2kx_c). \qquad [2.77]$$

In the case when the entire pulse is already inside the cloud:

$$E_{out}(x,t) = \frac{\alpha}{2k}[\sin(kct - 2kx_c + kx + 2kx_c - kx - \omega t) -$$
$$\sin(kct - kc\tau + kx - 2kx_c + 2kx_c - kx - \omega t)] =$$

$$= \frac{\alpha}{2k}\sin(kc\tau). \qquad [2.78]$$

This is a very important result. It shows that the FM signal will be detectable in the backscatter produced by a large diffuse object only while the light pulse is still entering this distributed scatterer. The signal will disappear as soon as the entire pulse is inside the cloud, when it will be averaged to a value depending on the pulse length. This behavior is analogous to a situation when instead of the distributed scatterer a mirror, placed at x_c, is used. The reflectivity of such a mirror (or point backscatterer) is $\frac{\alpha}{2k}$, which is the sum reflectivity (or backscattering efficiency) of all the particles in the cross section of the light beam divided by $2k$.

Knowing α we can calculate the magnitude of the FM signal in the model remote sensing experiment. Now we have to remember that, because we treated the problem of backscattering of frequency modulated light in the framework of scattering of an electromagnetic wave, we must define the relationship between α and physically meaningful quantities, such as scattering amplitude functions for the laser light, modulation depth, magnitude of imbalance of the sidebands, quantum efficiency of the detector, etc. The next sub-section will address the calculation of scattering efficiencies.

2.7.2 Light scattering in the cloud.

Understanding of radiation scattering is crucial to our understanding of the capabilities of FM spectroscopy applied to the remote sensing problem. The following sub-section is devoted to the discussion of this subject.

Any macroscopic particle can be considered as a large collection of small oscillating dipoles. These dipoles, oscillating at the frequency of applied electromagnetic field, produce a secondary field that radiates in all directions. This is what we call the *scattered* field. It is quite obvious that thus defined, particle scattering is a complex problem because the secondary waves generated by *each* dipole also stimulate oscillations in neighboring dipoles. Thus, radiation scattered from one particle is a superposition of all the waves scattered by all the dipoles and the precise form of this superposition depends on the way the dipoles are arranged in the particle. This forces us to limit the scope of the discussion and address only very few of the multitude of possible scattering problems. First, we shall limit our discussion to the cases when the scattered light has the same frequency as the incident light (i.e. effects like the Raman effect are excluded). The second, most important, limitation is that only scattering by independent particles is considered. An obvious example of scattering is that by water droplets in atmospheric air. If these particles are sufficiently distant from each other, it is possible to study scattering by each particle independently from other particles. This we will call independent scattering. The question of over what distance is sufficient for uncharged particles to act independently is beyond the scope of this discussion. Van der Hulst[2] suggests a separation of three particle radii to be sufficient. This may not be generally valid, but for most practical remote sensing situations the particles are separated by much larger distances. The third limitation we must impose on our discussion is that the effects of multiple scattering will be ignored. Treatment of multiple scattering does not introduce any new physics, because our assumption of independence implies that each particle is considered to be in free space. On the other hand, multiple scattering drastically complicates the mathematical description of the problem. As will be seen

from our discussion, for the densities of clouds and depths of propagation that are feasible in FM-remote sensing, the influence of multiple scattering will be minimal.

The radiation is scattered in all directions, although not in equal proportions and certainly in a way that depends on particle size, shape, and composition. In our discussion we should concentrate on some exemplary case and assume that the results obtained will be qualitatively valid for all other possibilities. The most widely found clouds consist of water particles. We are going to use Mie theory[1] of scattering as it most adequately describes scattering by particles of the type found in water clouds. This theory was independently developed[1,2,6] by L. Lorenz in 1890 and G. Mie in 1908 and is a result of application of classical electromagnetic theory to the problem of scattering by dielectric spheres.

The electromagnetic field scattering pattern is described in terms of $S(\Theta)$, the quantity referred to as the *amplitude function*. The physical sense of this function can be illustrated by the following experiment (Fig. 2.12).

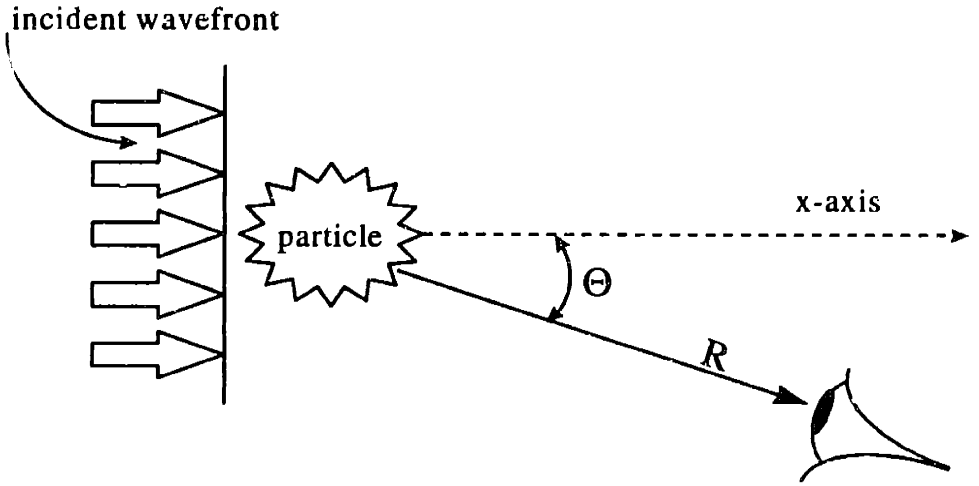


Figure 2.12 Scattering of a plane wave by an arbitrary particle.

Let us consider a particle illuminated by a plane wave which travels in the positive x direction and has the form

$$E_{incident} = E_0 e^{i\omega t - ikx}. \qquad [2.79]$$

The scattered wave a distance R away is a spherical wave whose amplitude is inversely proportional to the incident amplitude. This wave can be written as

$$E_{sca} = S(\theta)\frac{e^{i\alpha x - ikR}}{kR}E_0, \qquad [2.80]$$

which defines the amplitude function $S(\theta)$, introduced here to describe the scattering pattern as a function of scattering angle.

To simplify our model we are going to introduce several additional assumptions:

1. The particles that cause scattering are spheres.
2. The particles are non-absorbing.
3. The size of the particles we consider is typical of the sizes of particles found in clouds.
4. The incident light is linearly polarized.

A laser beam of intensity I_0 is scattered by a spherical water droplet in the atmosphere, and the reflected light of intensity I is collected by a detector. The intensity I depends mainly on the angle of scattering (θ), the size of the water droplet, the refractive index (1.33 for water), and the wavelength of the laser beam.

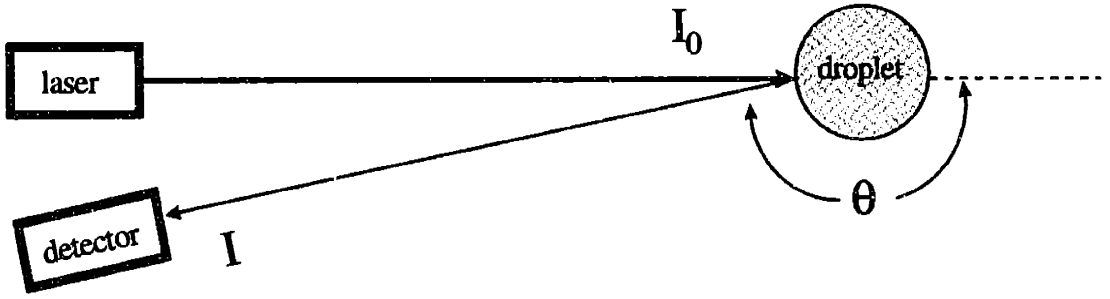


Figure 2.13 Scattering by a spherical droplet.

The following formulas give the scattered light intensity as a function of the scattering angle θ and the initial intensity I_0:

$$I^r = \frac{i_1}{k^2r^2}I_0 \qquad I^l = \frac{i_2}{k^2r^2}I_0, \qquad [2.81]$$

where I^r and I^l are the components of the light polarized parallel and perpendicular to the polarization of the incident beam, $k = \frac{2\pi}{\lambda}$, r is the distance from the scattering particle to the observer, and i_1 and i_2 are the square moduli of the amplitude function:

$$i_1 = |S_1(\theta)|^2 \quad \text{and} \quad i_2 = |S_2(\theta)|^2 \qquad [2.82]$$

(after van de Hulst)[2].

The solutions are expressed as infinite series and the rates of convergence of these series depend on the value of the size parameter $x = \frac{2\pi a}{\lambda}$, where a is the radius of the scattering sphere and λ is a wavelength of light.

To compute the amplitude function, we used the formulae given by Gumprecht et al.[3, 4]:

$$S_1 = \sum_{n=1}^{\infty} \{A_n \pi_n + B_n [\alpha \pi_n - (1-\alpha^2)\pi'_n]\} \qquad [2.83]$$

$$S_2 = \sum_{n=1}^{\infty} \{B_n \pi_n + A_n [\alpha \pi_n - (1-\alpha^2)\pi'_n]\}$$

where π_n is a function of α, and $\alpha = \cos\theta$. Upon substitution we get

$$S_1(\theta) = \sum_{n=1}^{\infty} \{A_n \pi_n(\cos\theta) + B_n [\cos\theta \cdot \pi_n(\cos\theta) - \sin^2\theta \cdot \frac{d\pi_n(\cos\theta)}{d(\cos\theta)}]\}$$

$$S_2(\theta) = \sum_{n=1}^{\infty} \{B_n \pi_n + A_n [\cos\theta \cdot \pi_n(\cos\theta) - \sin^2\theta \cdot \frac{d\pi_n(\cos\theta)}{d(\cos\theta)}]\} \qquad [2.84]$$

We can substitute the expression $\tau_n = \cos\theta \cdot \pi_n(\cos\theta) - \sin^2\theta \cdot \frac{d\pi_n(\cos\theta)}{d(\cos\theta)}$ into our equation (after van de Hulst), and obtain the forms

$$S_1(\theta) = \sum_{n=1}^{\infty} (A_n \pi_n + B_n \tau_n) \qquad [2.85]$$

$$S_2(\theta) = \sum_{n=1}^{\infty} (B_n \pi_n + A_n \tau_n).$$

Here, π_n and τ_n can be defined in terms of associated Legendre polynomials of the first kind[2]:

$$\pi_n(\cos\theta) = \frac{1}{\sin\theta} \mathrm{P}_n^1(\cos\theta) \qquad [2.86]$$

$$\tau_n(\cos\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} \mathrm{P}_n^1(\cos\theta).$$

A_n and B_n are called Mie coefficients, and are defined in terms of Riccati-Bessel functions and their derivatives. The Riccati-Bessel functions can be defined by Bessel functions of the first kind[3, 5]:

$$\psi_n(x) = \sqrt{\frac{\pi x}{2}} J_{n+\frac{1}{2}}(x) \qquad \psi_n'(x) = \frac{\partial \psi_n(x)}{\partial x}$$

$$\xi_n(x) = \psi_n(x) + i\chi_n(x) \qquad \xi_n'(x) = \frac{\partial \xi_n(x)}{\partial x} \qquad [2.87]$$

$$\chi_n(x) = (-1)^n \sqrt{\frac{\pi x}{2}} J_{-n-\frac{1}{2}}(x).$$

To compute χ_n, we first used the identity $Y_n = (-1)^{n+1} J_{-n-1}$ to transform the Bessel function of the first kind into a Bessel function of the second kind, Y_n:

$$\chi_n(x) = -\sqrt{\frac{\pi x}{2}} Y_{n+\frac{1}{2}}(x). \qquad [2.88]$$

The Mie coefficients are defined as[3]:

$$A_n = \left[\frac{(-1)^{n+\frac{1}{2}}(2n+1)}{n(n+1)}\right]\left[\frac{\psi_n'(y)\psi_n(x) - m\psi_n'(x)\psi_n(y)}{\psi_n'(y)\xi_n(x) - m\xi_n'(x)\psi_n(y)}\right] \qquad [2.89]$$

$$B_n = \left[\frac{(-1)^{n+\frac{1}{2}}(2n+1)}{n(n+1)}\right]\left[\frac{m\psi_n'(y)\psi_n(x) - \psi_n'(x)\psi_n(y)}{m\psi_n'(y)\xi_n(x) - \xi_n'(x)\psi_n(y)}\right]$$

where x is the size parameter, defined as $x = \frac{2\pi a}{\lambda}$ and $y = mx$. m is the index of refraction, and a is the radius of the scattering sphere.

Given the distance r from the scatterer to the detector, the wavelength of the laser beam λ, the size parameter x of the scatterer, and the index of refraction m, the preceding

formulas can be used to compute the intensity of scattered light as a function of the angle θ. This was done using the mathematical software package Macsyma. The equations outlined above were transformed into code readable by the computer and a batch file containing the equations was executed. (See Appendix A for the code). The batch file was executed for several values of the size parameter, and the results were compared with previously computed values[2] and were found to be in agreement.

The formal Mie theory solution does not, at least at first sight, provide a physically intuitive picture of the mechanism of scattering. However, some qualitative pictures can be drawn from it. We know that the scattered wave originates from oscillations of the electrons in the particle, excited by the incident radiation. The distribution of electric charges can be represented as a superposition of electric multipoles with arbitrary moments located at some origin point. In the case of scattering, this distribution of charges and currents oscillates synchronously with the incident wave and the scattered radiation arises from the corresponding oscillating multipoles. The amplitude and phase of each wavelet associated with a particular electric multipole is given by A_n and that for each magnetic multipole is given by B_n. The electric dipole moment is proportional to A_1 and the magnetic dipole moment is proportional to B_1, while higher order multipoles are related to corresponding higher order Mie coefficients.

Intensity functions for parameter values relevant to remote sensing were computed. As the refractive index m we chose 1.33 for water. Typical values for the size of water droplets in clouds range from about 4 μm to 100 μm.[6] Here and in the following discussion we will assume the probe laser wavelength to be 308nm, which would be representative of the experiments aimed at hydroxyl radical detection. This corresponds to a size parameter in the range from $x = 100$ to $x = 2000$. Unfortunately, with such large values of x, we encountered a difficulty in calculating the amplitude function. The results became unstable beyond $x \approx 15$. We were unable to determine the precise cause of unstability.

The efficiency of the scatterer, i.e. the ratio of $\frac{I}{I_0}$, computed for size parameter values of 0.01, 1, 5, 10, 15, and 20 are presented in Figures 2.14 through 2.19 as functions of the

scattering angle. These graphs plot the efficiency as a function of the angle. As expected, the efficiency varies with the angle of scattering, as well as with the size parameter x

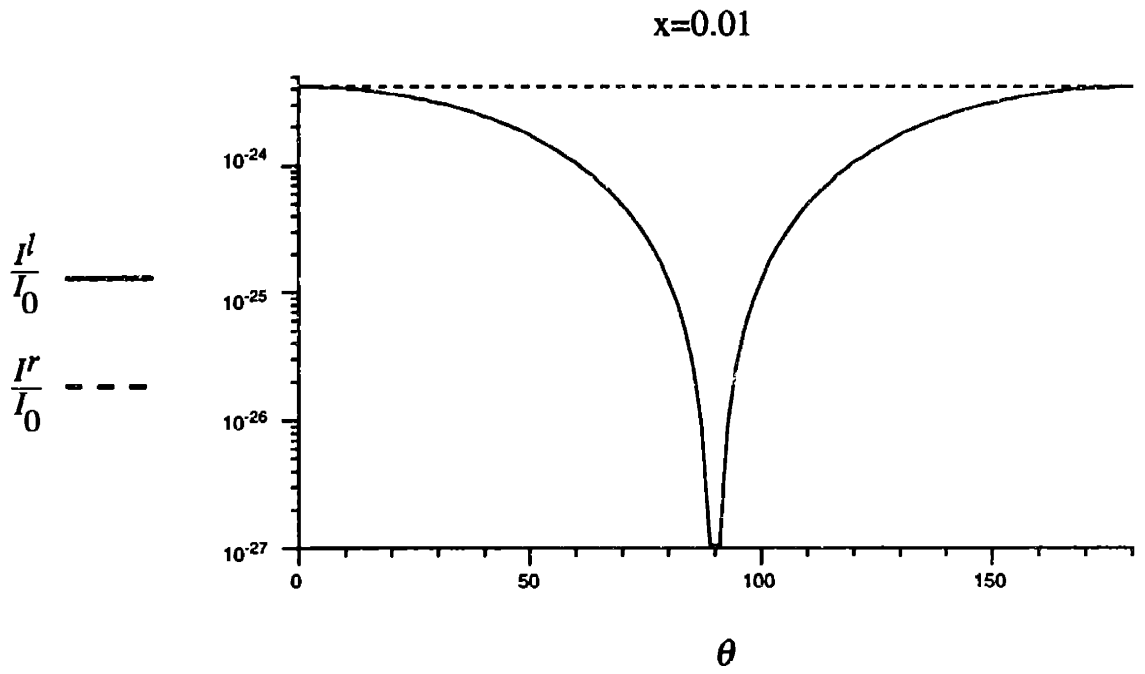


Figure 2.14 Angle dependence of scattering coefficient for x=0.01.

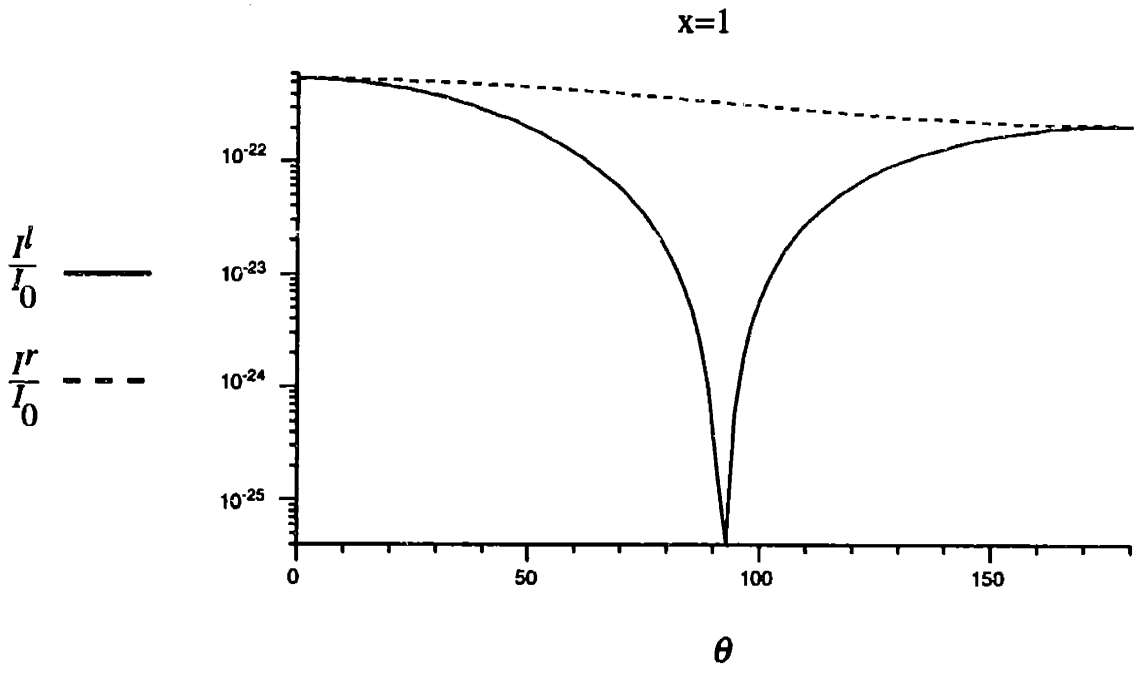


Figure 2.15 Angle dependence of scattering coefficient for x=1.

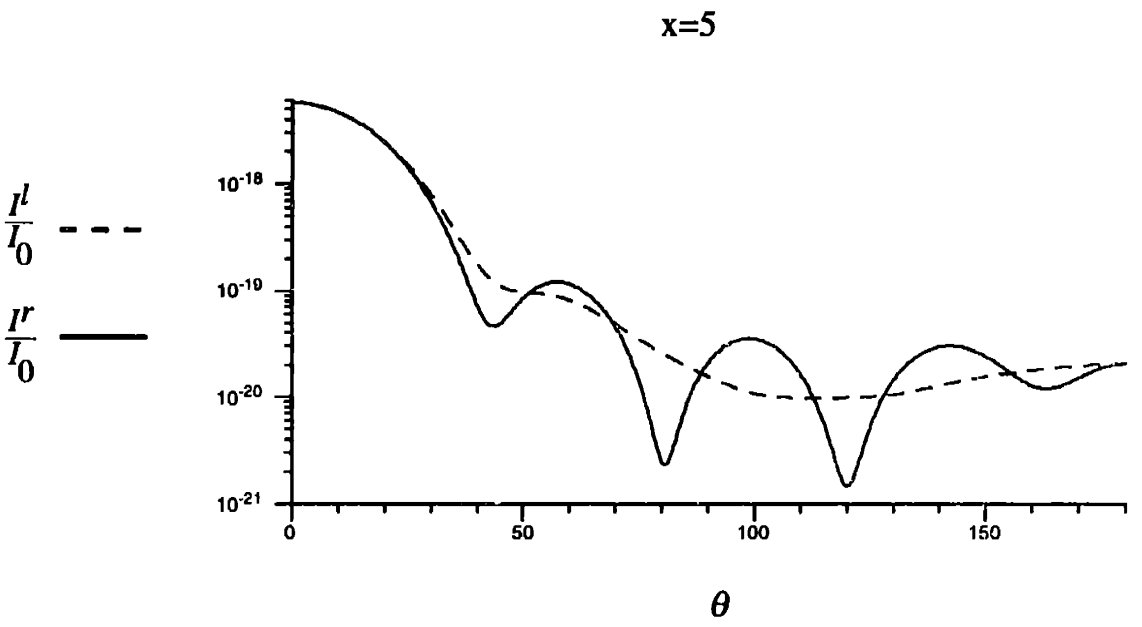


Figure 2.16 Angle dependence of scattering coefficient for x=5.

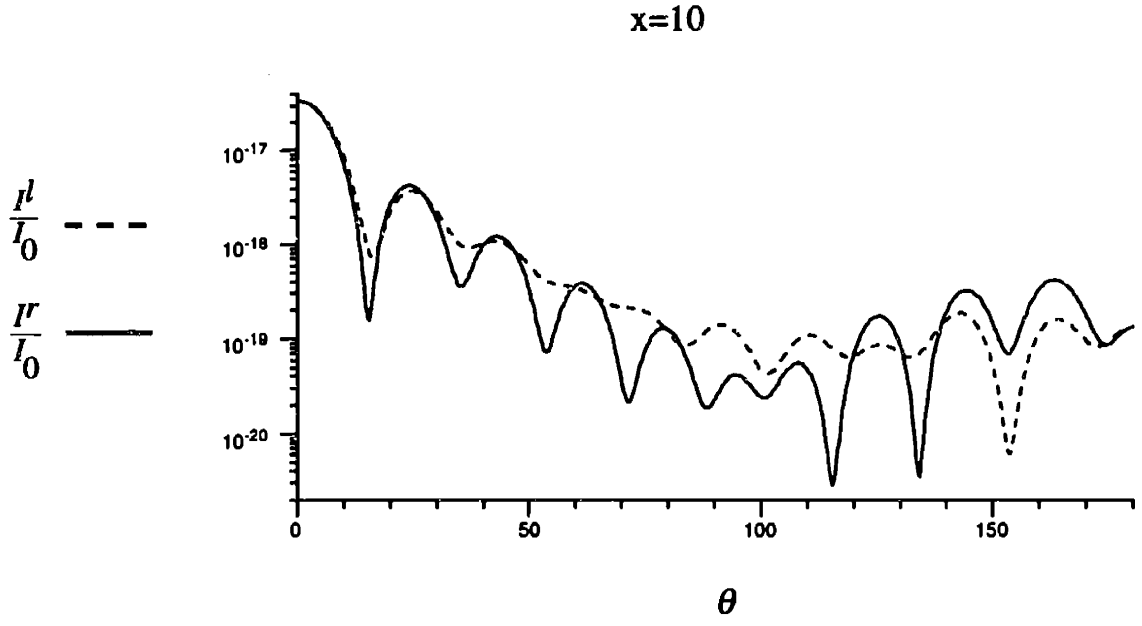


Figure 2.17 Angle dependence of scattering coefficient for x=10.

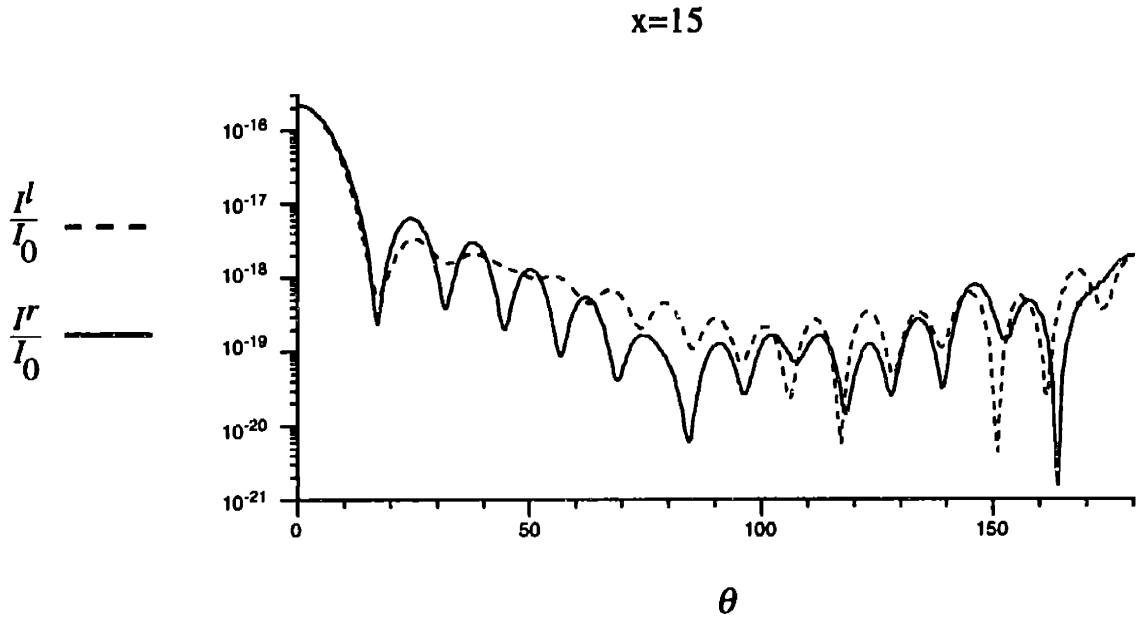


Figure 2.18 Angle dependence of scattering coefficient for x=15.

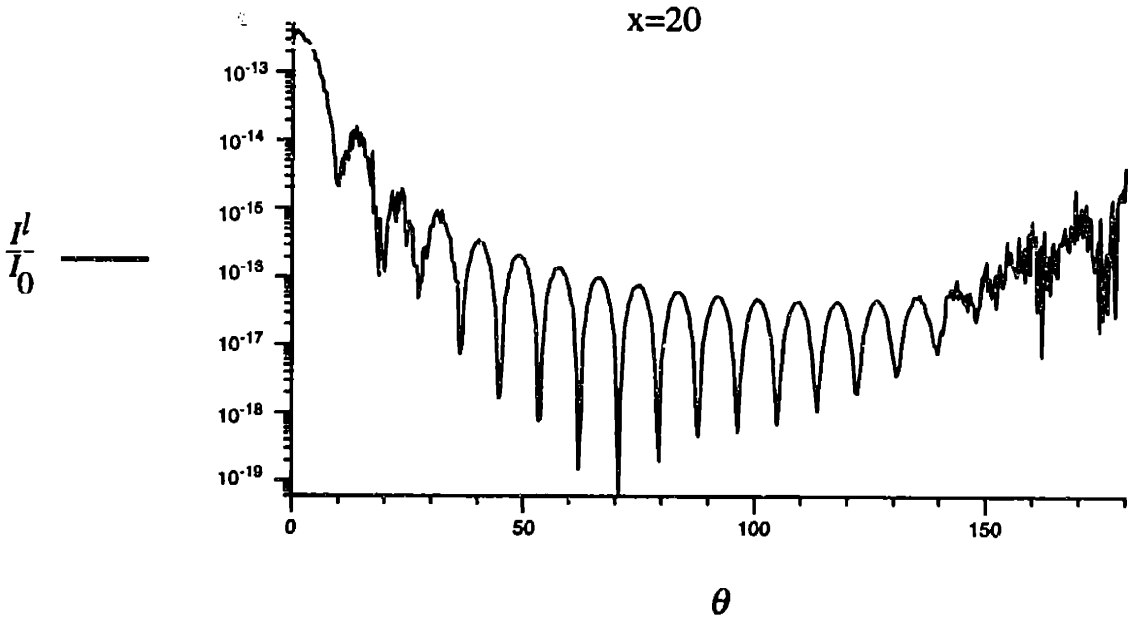


Figure 2.19 Angle dependence of scattering coefficient for x=20.

It is also easy to see that at large values of x, the angle dependence begins to develop a "fine structure" and the entire pattern begins to look chaotic and resemble a specular distribution. One possible, albeit unlikely, reason this specular behavior was not reported in previous work is that only the values for certain angles (such as 0^O, 10^O, 180^O, 170^O, etc.) were calculated and the scattering efficiency curve was assumed to be smooth for angles between. The power of modern computers has allowed us to calculate efficiency curves for very small angle increments, resulting in the unstable-looking fine structure.

Looking at the graphs of the scattering efficiencies, we can notice that they become more peaked in the forward direction as the size parameter increases. A sharp increase in scattering also occurs in the backscattering ($\theta = 180^O$) direction for the large spheres. This allows us to suggest that the monostatic configuration (i.e. when the source and the detector are in the same place) of the remote sensing setup will be both most effective and most convenient to implement.

We can also notice that, for some values of x (for example $x = 10$ in the preceeding graphs), there is an optimal angle at which scattering is greater than at $\theta = 180°$. This

information, however, cannot be used to our advantage, as particle size varies within any cloud and one can only estimate the average size of the water droplets. Instead, it would be most useful to look at values of $I(180°)$, as measurements at the location of the laser can be easily carried out.

To address the problem of estimation of backscattering efficiency for size parameters of practical interest we plotted (see Fig. 2.20) the values for $I(180°)$ for $x = 1$ through 15 in increments of 1, and added previously calculated values[3] for $x = 20$, 25, 30, 35, and 40. Since the values for $I(180°)$ level off, we extrapolated to estimate the range of values in the $x = 40$ to $x = 100$ region:

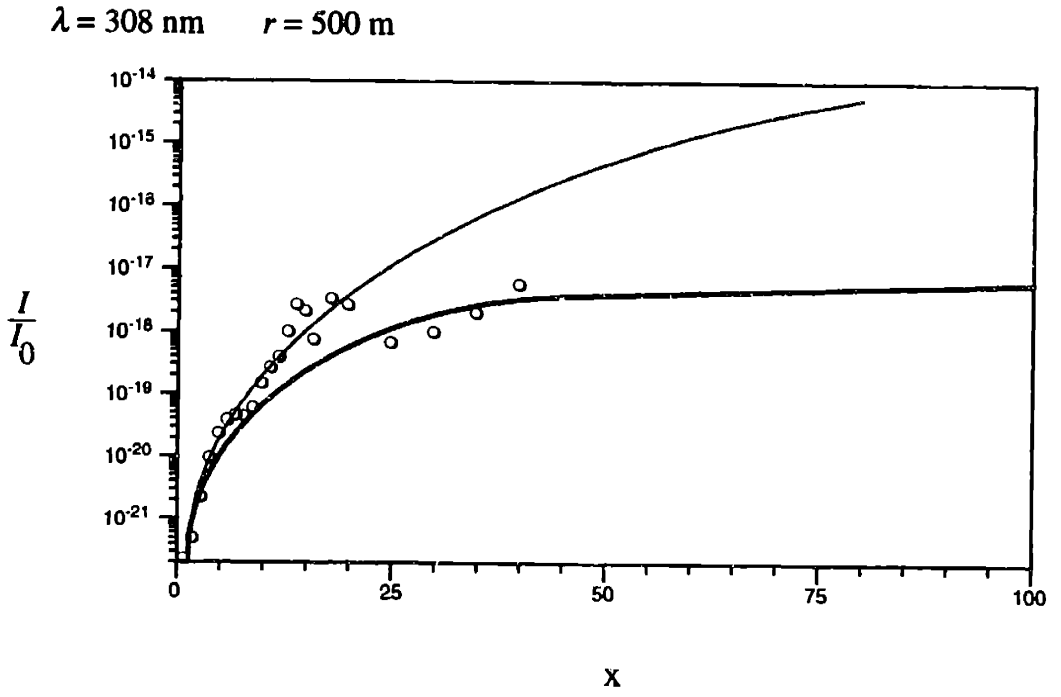


Figure 2.20 Dependence of backscatering efficiency on size parameter x. Continuous curves are extrapolations for large size parameters.

For our calculations, we chose the more conservative estimate. As a result, when x equals 50 to 100, we can expect back-scattering to be on the order from $10^{-14}I_0$ to $10^{-17}I_0$ (at a typical value of $\lambda = 308$ nm and a distance of 500 m).

The intensity I refers to the energy flux per unit area. Its units are power over area (for example, watt/m^2). I as a function of distance r (the distance between the scatterer and the detector) is given by the formula

$$I(\theta,\varphi,r) = \frac{i(\theta,\varphi)}{k^2 r^2} I_0, \qquad [2.90]$$

where I_0 is the intensity of the incident light, $k=2\pi/\lambda$, and i is a dimensionless function of direction but not of r. i/k^2 is an area, and it is inversely proportional to r^2, since the intensity of the light decreases with the square of the distance travelled. The factor of k^2 is a normalization factor and arises from the fact that the integral over all directions of the phase function has to be equal to 1, according to the law of conservation of energy.

To compute the intensity at the detector, the intensity incident on the scatterer must be expressed in units of power over area. Since $\frac{i(\theta,\varphi)}{k^2 r^2}$ is dimensionless, I also will be expressed in the same units of power over area. To obtain the intensity in units of power (or the amount of light collected) , I needs to be multiplied by the area of the collection optics. Unless noted otherwise, all calculations will be conducted and all numbers will be given for scattering coefficients recalculated to incorporate our model: laser beam diameter of 10 cm, collection optics of 1 m^2 area, and cloud 500 meters away from the source/detector.

The following section will use the model introduced above along with the expression for the FM signal resulting from the extended scattering object return and calculated efficiencies of such return to evaluate the feasibility of construction of FMRS setup.

References

1. G. Mie, Annalen der Physik 25, 377 (1908).
2. H. C. van de Hulst, *Light Scattering by Small Particles,* Dover Publications, 1981.
3. R. O. Gumprecht et al., Journal of the Optical Society of America 42, 226 (1952).
4. It should be noted that the formulae for $S_1(\theta)$ and $S_2(\theta)$ are given incorrectly by van de Hulst:

$$S_1(\theta) = \sum_{n=1}^{\infty}\left(\frac{2n+1}{n(n+1)}\right)\{A_n\pi_n(\cos\theta) + B_n\tau_n(\cos\theta)\}$$

$$S_2(\theta) = \sum_{n=1}^{\infty}\left(\frac{2n+1}{n(n+1)}\right)\{B_n\pi_n(\cos\theta) + A_n\tau_n(\cos\theta)\}$$

Specifically, the term $\frac{2n+1}{n(n+1)}$ should be omitted from the summation, and included within the expression for the Mie coefficients. Failing to do so, the Mie coefficients given by van de Hulst are missing a factor of $(-1)^{n+\frac{1}{2}}$ for A_n and $(-1)^{n+\frac{3}{2}}$ for B_n.

5. It should be also noted that van de Hulst defines the Riccati-Bessel functions incorrectly. He states the function ξ_n as $\xi_n(z) = \sqrt{\frac{\pi z}{2}}H^{(2)}_{n+\frac{1}{2}}(z)$, where he defines $H^{(2)}_{n+\frac{1}{2}}$ as a Bessel function of the **second** kind. In fact, $H^{(2)}_{n+\frac{1}{2}}$ is a Bessel function of the **third** kind.
6. G. L. Stephens, *Remote Sensing of the Lower Atmosphere*, Colorado State University, 1983.
7. For other values of r, simply multiply the given result by $(500/r)^2$. For a different wavelength, first compute the correct size parameter, obtain the intensity from the graph above, and multiply by the factor $(308/\lambda)^2$.

2.8 Feasibility Analysis of Frequency Modulation Enhanced Laser Absorption Remote Sensing.

Now that I have discussed the theory of FM spectroscopy, analyzed its sensitivity, and quantitatively understood the process of scattering of frequency modulated light, I can use the results obtained in Sections 2.3, 2.4 and 2.7 to evaluate the feasibility of using Frequency Modulation Spectroscopy in remote sensing applications. In this section I will introduce a model for a typical field remote sensing experiment. Although all calculations will be based on the parameters of this model system, estimating the performance of the FMRS in different conditions will be straightforward. Following this discussion I will address some problems and questions of design of a pilot field setup along with possible solutions and recommendations for use of currently available technologies.

2.8.1 Capabilities and Sensitivity Limits for a Typical Setup.

Our goal is to determine the applicability and utility of FM Spectroscopy for determining concentrations of various atmospheric species when no solid backreflector (such as topographic target or mirror) is available.

We shall introduce a model in which a measurement has to be done utilizing backscatter from a cloud of water droplets situated 500 meters from the spectrometer. A distance of 500 meters is chosen for two reasons. The first is that the typical collection distance of most DIAL experiments is of the same order of magnitude. The second reason is that, although there have been relatively successful attempts to measure concentrations of such important atmospheric species as OH radical using very long path absorption, it has been problematic to determine local concentrations of such trace

species. Using just 500 meters as a probe distance will allow us to demonstrate the very high sensitivity of FMRS.

In addition, the formulas provided will allow us to scale results we obtain for 500 meters for different pathlengths. For convenience we will assume a monostatic experimental configuration, i.e. a configuration in which both source and detector are in the same place.

Another choice we have to make is that of the wavelength of the laser radiation. This is an important parameter of the model because the scattering probability depends on it quite strongly (recall formula [2.89] and Figure 2.20). Our choice of 308nm reflects the importance of sensitive and accurate detection of OH radical in ambient air. Additional parameters of the model, although quite arbitrary, will reflect a model apparatus which is plausible given current technology.

A laser source will be assumed to have a well collimated 10 cm^2 beam (or a beam which will have this area when it reaches the cloud). The source will have radiation energy of one Joule and a pulse duration of 10 ns. Collection optics will be assumed to include a telescope with a $1m^2$ mirror and 100% efficiency. In the ensuing discussion we will assume that all of the backscattering coefficients are normalized for these parameters. This means that the scattering coefficients we are using, although speccified by the same symbols as the coefficients per unit solid angle, are equal to the coefficient per unit solid angle times the collection optics area divided by the beam cross section area.

Frequency modulated light is emitted from this source toward a large cloud 500 meters away. To simplify our model we will assume that there are no aerosol droplets between the source and the cloud. Thus, on its way to the cloud and back to the detector light experiences Rayleigh scattering which causes power attenuation given by

$$\tilde{A}^2 = e^{-2\int_0^{x_c} A^R(r)dr}, \qquad [2.91]$$

where A^R is the Rayleigh attenuation coefficient and x_c is the distance from the laser source to the cloud. The factor of two arises because the beam travels to the cloud and back, thus its total pathlength is $2x_c$. The attenuation coefficient $\tilde{A}$ is an electric field attenuation coefficient, hence the square in the formula above.

The factor A^R is given by the expression $A^R = N_g \cdot \sigma_R$, where N_g is the number density of gas molecules, and σ_R is the Rayleigh scattering cross section. N_g varies with altitude in the atmosphere. Table 2.1 lists several values of N_g as a function of altitude.

h (km)	N_g (cm^{-3})
0	2.547×10^{19}
1	2.311
2	2.098
3	1.891
4	1.704
5	1.531

Table 2.1 Number density of gas molecules in the atmosphere, N_g, at different altitudes, h.[3]

However, since we are concerned with altitudes of only several hundred meters, we can assume that N_g is constant over the pathlength and has the value of $2.55 \times 10^{19} cm^{-3}$ (sea-level).

The scattering cross section coefficient, σ_R, is dependent on the wavelength of light, λ. It is given[4] (in a simplified form) by

$$\sigma_R = 4.17 \cdot \left(\frac{1}{\lambda[nm]} \right)^4 \times 10^{-16} cm^2 . \qquad [2.92]$$

Now we can write an expression for A^R. Substitution and simplification give

$$A^R = \frac{1.064 \times 10^6}{(\lambda[nm])^4} m^{-1} . \qquad [2.93]$$

Since, in this approximation A^R is constant, we can rewrite the attenuation function as $e^{\frac{-2.128 \times 10^6}{(\lambda[nm])^4} x_c[m]}$. Therefore, for a wavelength of 308 nm and a distance of 500m, we get a value of 0.888; i.e. on its way to the cloud and back, light is attenuated due to Rayleigh

scattering by 21.2%. The electric field is attenuated by 5.8% (or E=0.942E_0, where $\tilde{A} = 0.942$)

While propagating in the cloud, the laser beam is attenuated in two ways. First, there is attenuation by the Rayleigh scattering due to air, which is given by $e^{-2A^R d}$, where d - is the light penetration depth into the cloud. Since, as we found out in section 2.7.1, d is equal to the pulse length and is very small (on the order of a few meters), this contribution to the attenuation can be safely neglected.

The second part of the attenuation is caused by forward Mie scattering by water droplets that make up the cloud. To analyze its contribution, we will represent part of the cloud interacting with the light by a cylindrical volume, with the diameter equal to the diameter of the laser beam and the length equal to the penetration of the laser beam.

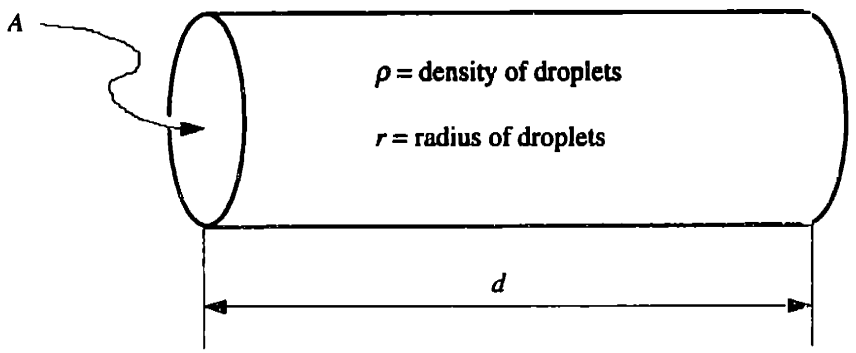


Figure 2.21 Model cloud parameters.

We denote the cross-sectional area as A, the distance the beam travels (or penetration) as d, the concentration (or density) of droplets as ρ ,and the radius of droplets as r. For simplification, we will assume the radius for all droplets to be equal to the mean radius of droplets in a particular cloud (see below). Another simplification we introduce is that the droplets are distributed in the cloud so that they do not shade one another. Given values for all the parameters, we can calculate how many times a photon within the total area of the laser beam will be forward scattered by droplets.

Let us define:

medial cross section of a droplet = πr^2 [2.94]

volume of the laser beam cylinder = $A \cdot d$ [2.95]

total number of particles in the cylindrical volume: $\rho \cdot$volume = $\rho \cdot A \cdot d$ [2.96]

For simplicity, the particles can be placed in a single layer and we can then calculate te probability that they scatter each photon out of the beam:

$$\frac{\text{Total scattering area of particles}}{\text{Area of the beam}} = \frac{\rho \cdot A \cdot d \cdot \pi r^2}{A} = \rho \cdot d \cdot \pi r^2 = q \qquad [2.97]$$

The physical sense of this equation is that a beam traveling a distance d into the cloud on average will be scattered q times.

For an altostratus, a Cumulus, or a Stratus cloud[1]:

ρ = 200-400 cm^{-3}

$r = 5 \cdot 10^{-4}$cm

If we take d = 10m, which is a typical maximum value for our calculations, then q = 0.24, which means that 24% of the beam will be scattered once, while the remaining 76% of the beam will not be scattered at all. The beam would have to travel a distance of 42m before all of it is scattered on the average once.

For a Cumulonimbus or Cumulus Congestus cloud[1]:

ρ = 60-70 cm^{-3}

$r = 20\text{-}25 \cdot 10^{-4}$cm

If we take d = 10m again, then q = 0.91. In fact, for a distance of 11m, we can expect the beam to be scattered on average once.

Droplets of the size listed above scatter light predominantly in the forward direction (recall our discussion of light scattering in Section 2.7.2). In fact, for a particle with a size parameter of 20, 99.5% of the light is scattered in the forward direction. For a

particle of size parameter of 6, 99.7% is scattered forward. Thus, even if the beam is scattered on the average once, the attenuation effect due to Mie scattering by the droplets in the cloud is negligible and can be ignored.

One of the most important modifications of the standard FMS detection scheme, which we propose to introduce in FMRS to improve its sensitivity, is based on a simple fact illustrated by the expression for the FM signal (e.g. [2.34]):

$$I_{FM}(t,x)=\frac{Q_{det}E_c^2}{2}e^{-2\delta_0}\cdot\begin{bmatrix}[1+M(\delta_{-1}-\delta_1)\cos(\omega t-kx)+\\+M(\phi_{-1}+\phi_1-2\phi_0)\sin(\omega t-kx)]\end{bmatrix} \qquad [2.98]$$

The FM signal (the photocurrent oscillating at the modulation frequency ω) is, among other factors, proportional to the intensity of the carrier:

$$I_{FM}(t,x)\sim E_c^2 e^{-2\delta_0}. \qquad [2.99]$$

This fact manifests itself even in traditional FM spectroscopy - if the modulation frequency is small and the carrier and the sideband are both attenuated by the molecular absorption feature, a significant decrease in FM signal occurs. This has a very simple explanation: the FM photocurrent is proportional to the product of the light fields of the carrier and sidebands, while sideband fields are themselves proportional to the carrier field (recall [2.8]).

This can be used to our advantage. In the remote sensing measurement, both carrier and sidebands are equally attenuated by the scattering in the atmosphere and by the non-efficient backscattering target, such that only a tiny amount of light reaches the detector. We can significantly increase the FM photocurrent by seeding the photodetector with cw laser radiation at the carrier frequency. Then, the major component of the FM signal will be caused by the product of the field of the backscattered sidebands and seeding laser field. Such seeding is necessary for sensitivity enhancement when the photocurrent produced by the backscattered light alone is smaller than or comparable to the dark current or thermal current of the detection system. In essence, this is an introduction of an additional heterodyning stage into the detection scheme and results in amplification of a weak signal by a strong seeding laser (which functions as a local oscillator in this case). It is easy to derive the expression for the FM signal for this

experimental arrangement. Following equation [2.29] we can express the frequency modulated light field on the detector as:

$$E_{FM}(t,x) = E_c \begin{bmatrix} \tilde{A}S(180^\circ)T_{-1}\frac{M}{2}e^{i(\Omega t-\omega t-k_{-1}x)} + T_0 e^{i(\Omega t-k_0 x)} \\ -\tilde{A}S(180^\circ)T_1\frac{M}{2}e^{i(\Omega t+\omega t-k_1 x)} \end{bmatrix}, \qquad [2.100]$$

where the term oscillating at Ω describes the field of the carrier and $S(180^\circ)$ is a field scattering coefficient (see Section 2.7.2). For brevity we will denote $S(180^\circ)$ as S_b.

T_0 has a different meaning than in equation [2.29]. While a characteristic of interaction of the carrier with the sample molecules in equation [2.29], T_0 is introduced here to facilitate adoption of formulas for the FM signal and sensitivity obtained earlier, so that they will have the same functional form as in previous sections. Thus,

$$T_0 = e^{-\delta_0}e^{-i\phi_0}, \qquad [2.101]$$

where δ_0 and ϕ_0 are chosen in such a way that if $\tilde{E}$ is the amplitude of the seeding laser electric field:

$$\tilde{E} = E_c e^{-\delta_0}, \qquad [2.102]$$

while ϕ_0 is the phase of the seeding laser. We are neglecting the electric field term corresponding to the carrier light emitted toward the cloud and scattered back, as its cross product with the sideband electric field will be insignificant compared to the products of sidebands and seeding field terms. Expressing the photocurrent (recall equation [2.30]), we have:

$$I_{FM}(t) = \frac{Q_{det}}{2}E_{FM}E_{FM}^* =$$

$$= \frac{Q_{det}E_c^2\tilde{A}S_b}{2}\begin{bmatrix} -T_{-1}T_0^*\frac{M}{2}e^{-i\omega t-i(k_0-k_{-1})x} - T_{-1}^*T_0\frac{M}{2}e^{i\omega t+i(k_0-k_{-1})x} + \\ +T_0T_1^*\frac{M}{2}e^{-i\omega t-i(k_1-k_0)x} + T_0^*T_1\frac{M}{2}e^{i\omega t+i(k_1-k_0)x} + \frac{1}{\tilde{A}S_b}|T_0|^2 + \\ +\frac{M^2\tilde{A}S_b}{4}\left(|T_{-1}|^2+|T_1|^2\right)+\{(2\omega)\ term\} \end{bmatrix} \qquad [2.103]$$

$$SNR = \frac{i_{FM}^2}{i_{SN}^2 + i_{BFM}^2 + i_{FM-L}^2} = \frac{M^2\delta^2}{\xi + \frac{M^2\left(\delta^2 + \Delta^2\right)}{LNR}}, \qquad [2.53]$$

and we can obtain the expression for signal to noise ratio for background limited FMRS:

$$SNR^B = \frac{LNR}{1 + \frac{\Delta^2}{\delta^2}}, \qquad [2.109]$$

which is equivalent to the result obtained earlier for the general case of background limited FMS. This result shows that, if enough backscattered photons fall on the detector to enable heterodyne detection and to create a photocurrent which is larger than the fluctuations in the dark current and thermal noise currents in the detection electronics, and if the sensitivity is background limited, the sensitivity will not depend on the amount of light returned. Taking a SNR of unity as acceptable, we get the familiar expression for the smallest detectable absorption:

$$\delta_{\min}^{FM,B} = \frac{\Delta}{\chi\sqrt{LNR - 1}}. \qquad [2.110]$$

If modulation frequency has been chosen correctly (see discussion in Section 2.4.3), the attenuation factor will be close to unity. Although the magnitude and form of laser intensity fluctuations are characteristic of the particular laser system used, typical values for LNR can range from unity, for 100% laser intensity fluctuations, to 10^4, for 1% laser noise. A value of 25 we choose for LNR (which corresponds to 20% laser noise) is not unreasonable for a pulsed laser based system. Background levels (see discussion in Section 2.4), Δ, can be on the order of 10^{-2}-10^{-5}. Thus, taking the worst case scenario for background signal, $\Delta=10^{-2}$ and LNR=25, a minimum absorption of $\delta_{\min}^{FM,B} = 0.002$ can be detected. Needless to say that any reduction of the laser source intensity fluctuations or the FM background signal will lead to significant increases in sensitivity.

Even with moderate requirements to the apparatus discussed above, the sensitivity of the technique is quite impressive (at least compared to the existing detection systems). Within the limits and parameters of the proposed model, an absorption of 2×10^{-8}cm^{-1} can be detected by FMRS in a pathlength of 1000 meters (to the cloud and back).

In the case when no seeding laser amplification is used, the FM current from the detector caused by the FM signal is given by:

$$i_{FM} = \frac{\chi Q_{\det} E_c^2 \tilde{A}^2 |S_b|^2 M\delta}{2}. \qquad [2.111]$$

To calculate the FM signal corresponding to the minimum detectable absorption, we need to introduce some additional parameters into the model. A modulation index of 0.8 will be used along with detector parameters corresponding to Hamamatsu Microchannel Plate Photomultiplier tube series R3809U. This choice of detector is dictated by the MCP's fast time response (typical rise time of 150 picoseconds), high gain (on the order of $2x10^5$) and very broad spectral response (detectors covering the spectral region from 115nm to 1.2 μm are available). For this type of detector, Q_{det} is (from Eq. [2.36]) is around $5x10^3$ Amperes/Watt. The intensity scattering coefficient, recalculated for the beam size and collection area size, is on the order of 10^{-16} per scatterer. Equation [2.111] gives the photocurrent caused by light backscattered by a thin slice of the cloud. Therefore, for the total photocurrent we can write

$$I_{FM} = \int_0^d i_{FM}(x)dx \qquad [2.112]$$

which is equivalent to Eq. [2.77]. Thus, analogously with the derivation following Eq. [2.77], the amplitude of the total FM photocurrent is

$$I_{FM} = \frac{\chi Q_{\det} P_0 \tilde{A}^2 M\delta\Phi}{2}, \text{ for a setup not using seeding laser amplification} \qquad [2.113]$$

and

$$I_{FM}^{seeding} = \frac{\chi Q_{\det} \tilde{P}\tilde{A} M\delta\sqrt{\Phi}}{2}, \text{ for a setup using seeding laser amplification,} \qquad [2.114]$$

where is the intensity scattering efficiency of a thin slice of cloud. The thickness of this slice is going to be determined by the bandwidth of the detector. If the detector can distinguish between the electric fields coming from two different points, then the points belong to different slices. The MCP-PMT detectors we have incorporated into our model have a rise time of 150 picoseconds. It is safe to assume that a light intensity

change that occurs on the timescale of 100 picoseconds is not going to be resolved. Therefore, it is convenient to specify a slice thickness of $3\text{cm} = 3 \cdot 10^{10}\ cm/s \cdot 10^{-10}\ s$.

$$\Phi = \left|S(180^{\circ})\right|^2 nAl. \qquad [2.115]$$

The concentration of droplets, n, will be taken as 70 cm^{-3} (characteristic of a Cumulus Congestus cloud), the laser beam cross sectional area A is equal 78.5cm^2 (10cm diameter beam), and slice thickness l=3cm. Once again, the intensity scattering coefficient $|S(180^{\circ})|^2$ has been recalculated for the distances and sizes of our model system.

Thus, in the case when no seeding amplification is used, the photocurrent corresponding to an absorption of δ=0.2% is $I_{FM} = 5.86 \cdot 10^{-4}\ Amp$, which corresponds to 29.3 mV (into 50 Ohms).

When seeding is present, same absorption will amount to $I_{FM}^{seeding} = 8.35 \cdot 10^{-6} \cdot \tilde{P}$, where $\tilde{P}$ is the power of the seeding laser (we also have to remember that seeding power is related to the probe laser power via Eq. [2.111]). It is quite obvious that in order to use a seeding arrangement to our advantage, we must have a pulsed seeding laser. This requirement is caused by the fact that the highest cw current an MCP-type detector can withstand is on the order of 100nA. Such a current corresponds to an optical power of only 10^{-11} Watts. Using such a seeder in this setup would be counterproductive. On the other hand, it is quite easy to create a seeding pulse of any (reasonable) power necessary. It is interesting to note that even without the use of a seeding laser for signal amplification, the sensitivity of an MCP-PMT detector will be sufficient to detect an FM signal corresponding to 0.2% absorption within the limits of the specified model. Still, in the case when the scattering object is too far away and the collection efficiency of backscatter is much worse than in the case we considered, one may benefit from using the seeder as an amplifier. The only constraint one has to be aware of when designing such a system is that the seeding laser itself should create a photocurrent which is less than the specified limiting photocurrent of the detector.

2.8.2 Design recommendations for the field setup.

We have outlined most of the requirements and main parameters of a generic remote sensing setup. Let us now discuss in more detail the technical solutions and available technology that would allow us to create such a setup.

Again, we shall divide the system into subsystems and will discuss each separately. The subsystems include the Source (which in turn include source laser, frequency modulator, and amplifier) and the Detection electronics.

The source subsystem must be based on a single longitudinal mode cw laser with a very narrow bandwidth. This limits the choice to very few options. The system which is currently feasible would be based on cw dye or Ti:Sapphire laser (analogous to the well known Autoscan 899 series by Coherent, Inc.) Positive features of these lasers include relatively high output powers (on the order of 1W single mode), very high frequency and power stability, ability to scan without mode jumps over a very broad spectral range, ability to cover a very broad spectral range (from near IR to near UV). Negative features include very high cost, very high cost of support (including the hiring and training of a highly qualified laser technician), very stringent requirements on the working environment (clean air, cooling water, highly stable optical table are very important). In addition, these lasers require optical pumping which nowadays can be accomplished only by a water cooled powerful ion laser, that is both very expensive and demanding of maintenance. In short, in addition to the very high cost of such a system, its adaptation to field working conditions will certainly provide major challenges to the designers. The cw laser light must be frequency modulated. The major difficulty of this part lies in the fact that the lithium niobate or magnesium doped lithium niobate crystals that are mostly used in EOMs, are prone to photorefractive damage if more than 100mW of light at visible wavelengths is input to the crystal. One possible solution would be to use quartz based EOMs. That would require building resonant standing wave modulators, but this might improve the durability of the system. Companies such as Conoptics, Inc. expressed interest in designing such electro-optic modulators for trial experiments. Frequency modulated light is then passed through pulsed amplifiers to obtain output power levels

which will provide backscatter return sufficient for detection. For a dye laser based system either Bethune cell amplifiers or amplifiers in which amplified light experiences total internal reflection off of the front surface (to minimize ASE) should be used. For a Ti:Sapphire laser based system, a special Ti:Sapphire amplifier chain should be developed. In both cases, optical pumping of the amplifier will require a relatively long pulse (10-20ns), injection seeded YAG laser. Such a pump laser is important because the optical spectrum of the amplifier output will be a convolution of the input (frequency modulated narrow band) and the pump (recall Section 2.5). Such pump lasers are almost a commonplace now, but they are quite expensive and difficult to maintain. Finally, if wavelengths outside the range of dye/Ti:sapphire lasers are desired, sum-frequency mixing of amplifier output (possibly with second or third harmonics of the pump laser) will provide a convenient source of UV wavelengths[5].

Another choice would be to design a system using a cw single mode diode laser instead of a Dye/Ti:Sapphire laser. Recent advances in diode laser technology give hope that within several years, such a system will be feasible. A very promising feature of such a system is that it is quite possible to fully integrate at least the first part (cw source-frequency modulator) of the system into a single compact maintenance-free "black box". This is exemplified by the current use of frequency modulated diode lasers in optical communications. Although the communications equipment cannot be directly transferred to our application (communications devices work at a fixed wavelength), much of the technology is applicable. Especially interesting is the possibility of using optical fiber based EOMs, and Diode pumped amplifiers. The main questions that must be addressed before a diode FM enhanced remote sensing system becomes feasible are those of access to broader spectral range and improved fiber amplification technology, leading to chirp-free amplifiers and higher output powers.

Detection electronics includes detector, amplifiers, RF mixer, and a recorder. The choice of the detector is governed by three considerations: it has to be sensitive in the spectral region of interest and have a sufficient bandwidth and sensitivity. It is quite clear that a detector with internal gain would be desirable for field measurements, especially

because the source of the signal is backscatter from a diffuse object. Until recently, such a detector was not available, or at least was not available cheaply, as PMTs are relatively narrowband and avalanche photodiodes have gain of only 10-50 times. Currently, the best candidates are Micro-Channel Plate (MCP) PMTs. MCPs (for example Hamamatsu's R3809U series) are relatively fast, with typical rise time of 150 picoseconds, sensitive over a broad spectral range (various models exist, covering from 115nm to 1200nm) and have current amplification on the order of 10^5. Moderately priced (under $1000), they are good candidates for FMRS detectors. Alternatively, use of avalanche photodiodes or bi-planar photoelements can be suggested for some applications.

Amplifier choice must also receive very serious consideration. It has to be low-noise, as the noise level of the first amplifier will determine the noise floor of the detection electronics. It must also be pulse tuned. This means an amplifier which does not distort the wavefront of the pulse. Such amplifiers are more expensive than cw tuned amplifiers, but are necessary to avoid generation of parasitic signals.

To facilitate demodulation of the FM signal in the mixer, the phase of the local oscillator (LO) must be matched to that of the radio frequency component of the photocurrent. This can be accomplished by using a variable length waveguide in the LO line. The disadvantage of this method lies in the necessity to tune the LO phase continuously, as the distance between the detector and the scattering target changes during the measurement. An elegant solution to this problem would be use of an *I and Q demodulator* (I&Q) (Figure 2.22). This device splits the input signal into two (50% each) and shifts one of them by 180°. Then, the signals are demodulated by beating against the same phase of local oscillator signal in two identical, well balanced mixers. The square root of the sum of squares of the resulting signals will always give the current equivalent to the current that would be obtained if the local oscillator phase were perfectly matched to the FM photocurrent oscillation. In other words, an I&Q demodulator effectively takes projections of the RF signal on the LO signal. An equivalent scheme would involve keeping two FM currents the same phase and shifting one of the LO currents by 90°.

Choice of the recorder will depend on the preferences of the designer, but any fast digitizing oscilloscope, fast sample-and-hold circuit with A/D board or boxcar-A/D board will suffice.

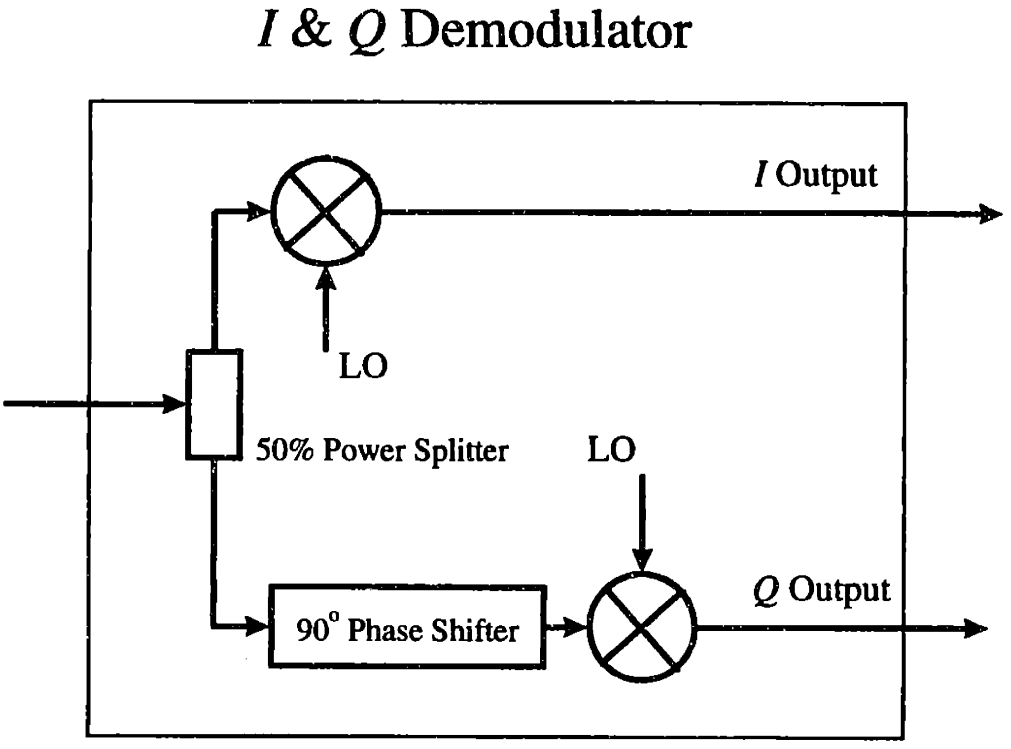


Figure 2.22 Schematics of I&Q demodulator.

Another important feature that such a detection system should have is the ability to normalize the FM signal relative to the intensity of the backscattered light arriving at the detector. Such an ability would be crucial to extracting quantitative information about the concentrations of species of interest, because the magnitude of the FM signal after such normalization will be independent of the intensity of the returned light. This can be accomplished by the use of directional RF splitter/combiner. This device, which is commercially available from many sources and is rather cheap, separates the signal on its input into high and low frequency components. After the high frequency component goes through the usual detection procedure, giving the magnitude of the FM signal, this signal

value can be divided by the magnitude of the low frequency part (scaled for convenience). This procedure can remove significant shot-to-shot fluctuations that will be present due to continuous changes in the environment.

References:

1. Clark, R.J.H., Hester, R.E. ed. by, "Spectroscopy in Environmental Science", *Series: Advances in Spectroscopy*, 24, 1 (1995).
2. Clark, R.J.H., Hester, R.E. ed. by, "Spectroscopy in Environmental Science", *Series: Advances in Spectroscopy*, 24, 223 (1995).
3. Elterman, L., "UV, Visible, and IR Attenuation for Altitudes to 50 km, 1968", *Environmental Research Papers,* **285,** Air Force Cambridge Research Laboratories, Bedford, Mass.
4. Hinkley, E.D., *Laser Monitoring of the Atmosphere.* Springer-Verlag, 1976.
5. E.Eyler et all. "Pulsed FM"

2.9 Conclusion

As we have shown earlier (recall Section 2.8.1), sensitivity to absorption on the order of $2x10^{-8}cm^{-1}$ can be detected over a pathlength of 1000 meters (to our model cloud and back). As a conclusion we will list the cross sections and concentrations for some molecules of atmospheric importance and determine whether FMRS is capable of detecting them.

One of the most important molecules of atmospheric interest is hydroxyl radical, OH. Reported ambient concentrations (which of course vary widely with time of day and with location) is on the order of $5x10^{6}$ cm^{-3}. If one uses the $Q_1(2)$ line (308nm) for detection, $\sigma_{max} = 2.25x10^{-16}cm^2$. This means that in a single shot, FMRS is capable of detecting concentrations on the order of $10^{8}cm^{-3}$. This means that FMRS (at least in our model configuration, using a very noisy laser source, etc.) is still below the detection limit. Of course, averaging over 10^4 shots improves the chance of detecting OH at this concentration as will improvements in reducing laser source fluctuations and improving the efficiency of collection optics.

Another test system would be the SO_2 molecule. The reported cross section of the absorption around 308nm is $\sigma \approx 2x10^{-20}cm^2$. This means that FMRS would be capable of detecting on the order of 10^{12} molecules. Ambient concentrations of SO_2 were reported to span the range between $5x10^{9}cm^{-3}$ and $1x10^{12}cm^{-3}$. Again, we should not forget that the sensitivity is estimated for a single shot, for our very non-ideal setup.

Thus, one of the immediate conclusions is that a very thorough and thoughtful design of the actual pilot system will have to be made in order to create a universally useful system capable of detecting a broad range of substances at their ambient concentrations.

Another important application is to use FMRS to detect air pollutants in locations where use of a backscattering target is not convenient, but concentrations of the species of interest are slightly higher than their typical atmospheric values.

Finally I would like to mention that this work is but the first step toward development of this very promising technique, and hopefully better technologies and

materials in the future will facilitate to increases in both its sensitivity and the spectral interval accessible.

3. Synthesis and Study of Isocyanogen.

3.1 Motivation for the Study of Spectroscopy of Isocyanogen.

One of the most fundamental questions of chemistry as a science is *how* chemical transformations occur. Statistical rate theories can accurately predict rates of reactions, but the dynamics of making or breaking of a bond for any system beyond simple diatomic molecules has proven to be very difficult to understand. On the other hand, chemistry and spectroscopy are very intimately connected to each other because any chemical transformation that the species undergoes is governed by the same potential energy surface (PES) that determines its spectrum. Spectroscopy helps us understand chemistry because every chemical change that the species can undergo is encoded into its spectrum.

Near local minima of the PES, restoring forces for sufficiently small displacements are quadratic. From this (Hooke's law) model, one is able to use normal mode analysis and solve for certain linear combinations of displacement coordinates which correspond to "perfectly uncoupled" harmonic oscillators. As one advances up in energy toward the top of a barrier or along some dissociation coordinate, this harmonic description begins to fail. Different vibrational modes couple to each other and the dynamics of a molecule containing a chemically significant amount of energy is qualitatively different from that for a molecule in the normal mode limit. Anharmonic and Coriolis resonances cause intramolecular vibrational redistribution (IVR) of energy in highly vibrationally excited molecules, so that spectroscopic study of resonances is central to a qualitative and quantitative understanding of vibrational dynamics[1]. Resonances break down both the physical significance of vibrational and rotational-projection quantum numbers as a description of the wave functions. Their practical utility is that they serve as labels, which are simply related to the energy and spectroscopic selection rules. On the other hand, a multiple-resonance Hamiltonian may still have "approximately conserved" quantities if the number of resonances is small

enough[2]. Quantum mechanically these approximate constants of motion correspond to approximate quantum numbers, valid on the timescale corresponding to the resolution of the spectrum. In the spectrum this situation will be reflected by the appearance of groups of levels (polyads) capable of interacting via anharmonic resonances. When the curvature of the PES starts to change significantly (and we would certainly expect this to happen near the barrier top) it will lead to restructuring of the resonances which will immediately be reflected by changes in the polyad structures. Understanding these changes and connecting them to the dynamics of the molecule is the ultimate goal of a series of experiments which will hopefully follow the study, the first steps of which are reported in this section (Section 3).

Isocyanogen (NCNC) appears to be a particularly valuable system for such a study. Linear in its ground electronic state ($X\ ^1\Sigma^+$) [3], isocyanogen has a bent structure in the first electronically excited state ($A\ ^1A''$) [4], which allows Franck-Condon access to the barrier to isomerization region.

Franck-Condon pumping followed by stimulated emission (SEP) would allow one to explore systematically the region close to the maximum of the barrier both in terms of geometry (the transition state has a very distorted shape) and energy. Additionally, isocyanogen represents a step up in complexity of reactional dynamics, compared to such an IVR "workhorse" as acetylene, because all four atoms of isocyanogen are heavy.

Isocyanogen belongs to the rare class of electron-deficient isonitriles and was thought to be a fairly unstable molecule[5]. Although its isomer, cyanogen (NCCN), has been known since 1815, when Gay-Lussac reported its synthesis, its isomer NCNC has not been considered until recently. The first *ab initio* theoretical study of NCNC was done in 1980 by Haese and Woods, who predicted its structure in the ground state and its dipole moments. Following this study, Sana and Leroy[6] predicted vibrational normal modes, infrared spectra and thermodynamic properties of all three dimers of $(CN)_2$ and of the CN radical. The stability order they obtained was NCCN>NCNC>CNNC.

In 1988 pericyclic reactions were proposed as a source for synthesis of isonitriles[5]. This pathway was used in the fragmentation of norbornadienone azine, the first reported synthesis of CNCN[7]. The product was originally misidentified as a di-isocyanogen

(CNNC), on the basis of NMR and mass spectroscopy. However a year later, Stroh and Winnewisser[9] repeated the synthesis and determined from rotational and vibrational spectra that the product was a mixture of cyanogen and isocyanogen present in a ratio of 7:3. In that work, the authors reported results of their study of the rotational spectra of CNCN and its single ^{13}C substituted isotopomers in the ground vibrational state and of CNCN in the $(v_3 v_4 v_5) = (001)$, (002), (010) and (011) vibrational states, and analyzed ν_1, ν_2 and ν_5 fundamental bands of CNCN around 2302.00, 2059.00 and 197.53cm^{-1}.

Following the work by Stroh and Winnewisser[9] isocyanogen has become a subject of a number of experimental[8,11-17] and theoretical[3,10,18-24] investigations. An alternative synthesis, by pyrolysis of N-cyano-2,3-diphenyl-cyclopropeneimine, was reported by Yamada et al. in 1989[8].

Matrix-isolation infrared investigation of products of norbornadienone azine[11] led to detection of the ν_4 fundamental vibrational band around 463.5cm^{-1}, as well as ν_1, ν_2 and several combination bands. In addition, a weak band around 1996.4cm^{-1} was assigned to the antisymmetric stretching mode, ν_3, of diisocyanogen.

In their *ab initio* investigation, Botschwina and Sebald predicted a strong Fermi resonance between the vibrational states ν_3 and $2\nu_4$ of CNCN. A weak band due to the lower component of this Fermi resonance pair has been found in the matrix IR spectrum around 889cm^{-1}. Consequently, the stronger absorption observed around 980cm^{-1}, which was originally assigned to the ν_3 fundamental band, has been reassigned to the upper component of this Fermi resonance pair[15].

Yamada and coworkers performed a microwave and millimeter-wave study of CNCN in the ground and $(v_3 v_4 v_5) = (001)$, (002), (010) and (011) excited vibrational states and determined improved rotational and centrifugal distortion constants for those states[8].

Gerry and coworkers[12] used microwave Fourier Transform and millimeter- and submillimeter-wave spectroscopy of the $(v_3 v_4 v_5) = (002)$ and (011) levels to determine nuclear quadrupole and spin-rotation coupling constants and precise rotational and centrifugal distortion constants up to the sixth order in *J*.

In more recent papers[16,17], Winnewisser and coworkers published rovibrational analyses of the bands due to the bending modes ν_5 (194.9cm^{-1}), ν_4 (463.7cm^{-1}), combination band system $\nu_4 + \nu_5$ (659.0cm^{-1}), the Fermi resonance pair of ν_3 and $2\nu_4$ (883.9cm^{-1}and 974.4cm^{-1}), and of the stretching bands of ν_1 and ν_2.

All but one of the theoretical investigations of isocyanogen were directed at the study of the ground electronic state. Botschwina and coworkers have published results of high quality calculations of the geometry and vibrational frequencies of CNCN[21,24,25]. Nguyen[18], Sunil, Yates and Jordan[19], and Sherrill[10] recently reported on the results of optimizations of geometries of the transition states for the isomerization between isocyanogen, cyanogen and diisocyanogen.

As we have seen from the previous discussion, all the experimental and theoretical (with one exception which will be addressed shortly) work has been done on the ground electronic state. The first (and only) theoretical paper determining the geometry and vibrational frequencies of the $\tilde{A}\,^1A''$ state of CNCN was published by Sherrill and Schaefer in 1994[4]. In this paper the authors reported that the first singlet excited state of isocyanogen has a strongly bent geometry, suggesting the possibility of its use for SEP-type experiments to study isomerization on the ground state potential energy surface.

Until now, no successful experimental investigation has been done on the first excited state ($\tilde{A}\,^1A''$). Vaccaro and coworkers[26] tried to use a standard grating UV spectrometer, but were unable to detect any transitions that could be securely attributed to isocyanogen in the spectral region between 260nm and 220nm.

Thus, exploration of the $\tilde{A}\,^1A''$ state of CNCN logically becomes our most immediate goal. This information is needed to locate and characterize the intermediate levels for potential PUMP-PROBE SEP experiments.

Section 3.4 of this thesis gives a progress report on our work on $\tilde{A}\,^1A''$ state spectroscopy of CNCN.

References

1. Nesbitt, D. and Field, R.W. *J. Phys. Chem.* **100**, 12735 (1996).
2. Jonas, D. M. et al., *J. Chem. Phys* **99** (10), 7350 (1993).
3. Botschwina, P. *Chem. Phys. Lett.* **225,** 480-485 (1994).
4. Sherrill, C. D. and H.F.Schaefer III, *J. Chem. Phys.* **100** (12), 8920 (1994).
5. Herges, R. Tetrahedron Comput. Method. **1,** 15 (1988).
6. Sana, M., Leroy, G., *J. Mol. Struct.* **76**, 259 (1981).
7. Van der Does, T. and Bickelhaupt, F. *Angew. Chem.* **100,** 998 (1988).
8. Yamada, K. M. T. et al., *Chem. Phys. Lett.* **160**(2) (1989) 113.
9. Stroh, F. and Winnewisser, M. *Chem. Phys. Lett.* **155,** 21-26 (1989).
10. Sherrill, C.D., *personal communication.*
11. Stroh, F., et al. *Chem. Phys. Lett.*, **160**, 105 (1989); **162**, 253 (1989).
12. Gerry, M.C.L., et al. *J. Mol. Spec.*, **140**, 147 (1990).
13. Goede, S.J., et al. *J. Am. Chem. Soc.*, **113**, 6104 (1991).
14. Bickelhaupt, F.M., et al. *Int. J. Mass Spectrom. Ion Process.*, **103**, 157 (1991).
15. Winnewisser, M., et al. *J. Mol. Spec.*, **153**, 635 (1992).
16. Stroh, F., Winnewisser, M., Winnewisser, B. *J. Mol. Spec.*, **162**, 435 (1993).
17. Stroh, F., Winnewisser, M., Winnewisser, B. *Can. J. Phys.* **72**, 1251 (1994).
18. Nguyen, M.T. *Chem. Phys. Lett.* **157** (5), 430 (1989).
19. Sunil, K.K., Yates, J.H., Jordan, K.D. *Chem. Phys. Lett.* **171** (3), 185 (1990).
20. De Almeida, W.B., *J. Mol. Struct.*, **206**, 77 (1990).
21. Botschwina, P., Sebald, *P. Chem. Phys.*, **141**, 311 (1990).
22. Scheller, M.K., et al., *J. Electron. Spectroscop. Relat. Phenom.*, **51**,75 (1990).
23. Pyykko, P., Runeberg, N., *J. Mol. Struct.*, **234**, 269 (1991).
24. Botschwina, P., Flugge, J. *Chem Phys Lett.*, **180**, 589 (1991); *Chem Phys Lett.*, **210**, 495(E) (1993).
25. Botschwina, P. *Chem. Phys. Lett.*, **225**, 480 (1994).
26. Vaccaro, P., *private communication.*

3.2 Synthesis of Isocyanogen.

The first report of the synthesis (in which isocyanogen was mistakenly identified as diisocyanogen, CNNC) was published in 1988 by van der Does and Bickelhaupt [1]. Isocyanogen was obtained as a result of flash pyrolysis of the precursor, nonbornadienone azine at 500 C. A little less than a year later, Stroh and Winnewisser[2] repeated the synthesis and conducted microwave and high resolution infrared (ν_1, ν_2 and ν_5) studies of the product and unambiguously determined it to be isocyanogen.

For preparation of CNCN we chose to use a slightly modified alternative synthetic method, which was reported by Yamada[3] et al. in 1989. This method uses a different precursor, N-cyano-2,3-diphenyl-cyclopropeneimine, which was selected because it already contains the CNCN chain, which is easily "ejected" during the flash pyrolysis stage of the synthesis.

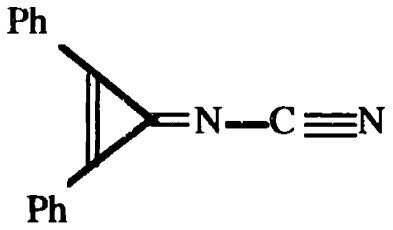


Figure 3.1 N-cyano-2,3-diphenyl-cyclopropeneimine.

In contrast, norbornadienone azine has the symmetric CNNC chain that corresponds to the unstable isomer, diisocyanogen. Thus, during the pyrolysis, fast isomerization must take place, which leads to formation of both CNCN and NCCN. Possible reaction pathways are discussed in references [2,3].

Synthetic procedure

1. Synthesis of N-cyano-2,3-diphenyl-cyclopropeneimine.

Unfortunately, the synthesis referenced in the Yamada et al. paper was not described in sufficient detail to repeat it. We were able to communicate with Prof. Vaccaro (Yale Univ.) who "reinvented" the same synthesis with the help of colleagues in the Organic Chemistry Department. The precursor synthesis reported below is loosely based on Vaccaro's guidelines.

The precursor, N-cyano-2,3-diphenyl-cyclopropeneimine, was prepared in the reaction of 2,3-diphenylcyclopropenone with $TiCl_4$ and N,N'-bis-(trimethylsilyl)-carbodiimide (Aldrich). 420mg of diphenylcyclopropenone was dissolved in approximately 20ml of dehydrated CH_2Cl_2. 0.336 ml of titanium tetrachloride (1.5 eq., Aldrich, ShureSeal bottle) was transferred by airtight syringe to the stirred solution. 1.36ml of N,N'-bis-(trimethylsilyl)-carbodiimide (3 eq., Aldrich) was added and the resulting mixture was refluxed for approximately 2 hours. At this point the mixture has a dark brown color. The reaction was quenched with a cold aqueous solution of (more than) 504 mg of KOH (I used 700mg). The mixture turns into transparent yellow solution (our product dissolved in CH_2Cl_2) with the top layer of water containing a white solid residue.

The product of interest is extracted with diethyl ether and dried over anhydrous $MgSO_4$. The ether was evaporated and the crude product was redissolved in 45ml of ethyl acetate. The solution was applied to a SiO_2 column and eluted with additional ethyl acetate. Although the original description of the synthesis spoke about collection of the first fractions and discarding later ones, using paper chromatography we have been able to identify only one substance in the output of the chromatographic column. Therefore, we assumed that the SiO_2 is used primarily as a filter and utilized a relatively short (2cm) column to speed up the process. Aliquots were collected, placed into a 100ml round-bottom flask and the solvent was vacuum-evaporated. The white-yellow residue is very well purified by recrystallization. Ethyl ether is found to be a poor solvent, so ethyl

acetate was used. Unfortunately, ethyl acetate is too good a solvent, so after waiting for a several days and seeing no crystals, the "solvent diffusion" method was used. Pentane, a highly non-polar solvent with vapor pressure similar to that of ethyl acetate, was placed into paper a chromatographic development vessel along with the vial containing very little (less than 1ml) of precursor in solution (see Figure 3.2). The vessel was closed with an air-tight lid and placed under the hood for a few days. The transparent crystals that formed in the vial are crystals of N-cyano-2,3-diphenyl-cyclopropeneimine.

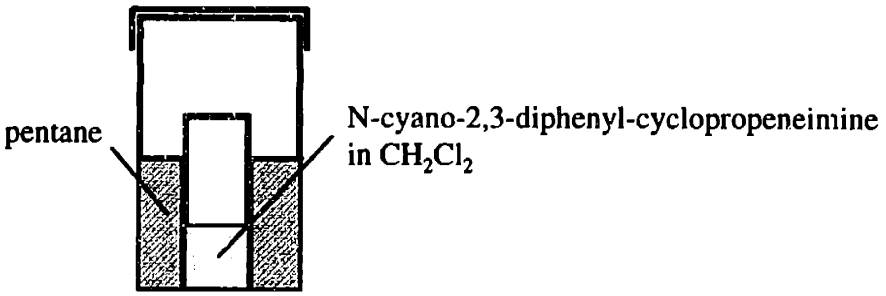


Figure 3.2 Recrystallization setup.

Hydrogen and carbon NMR spectra as well as elemental analysis confirmed that the product was indeed N-cyano-2,3-diphenyl-cyclopropeneimine.

Mg. sample	**%N**	**%C**	**%H**
1.1233	**12.12**	**83.37**	**4.20**
Theoretical per cent	**12.17**	**83.46**	**4.38**

Table 3.1 Elemental analysis report. Completed by *E+R Microanalytical Laboratory, Inc. (718)699-6555.*

2. Pyrolysis of N-cyano-2,3-diphenyl-cyclopropeneimine.

The pyrolysis is carried out in a 40cm long, 1" inner diameter quartz tubular reactor, heated to a temperature of about 500°C by an electric oven (Figure 3.3) The high temperature area is about 30cm long. Prior to the synthesis, the reactor and the liquid nitrogen cooled vials, used for collection and distilling, are pumped overnight and/or dehydrated by heating under vacuum . The precursor imine is placed in the quartz boat, which, after the reactor has reached the necessary temperature, is moved into the hot zone by pulling the magnet in the boat's handle with the external magnet. The precursor undergoes a very rapid decomposition in the high temperature region of the reactor, and volatile products are collected in the liquid nitrogen cooled trap. Some other decomposition products, which are not so volatile, deposit on the walls of the reactor as a dark brown/black film.

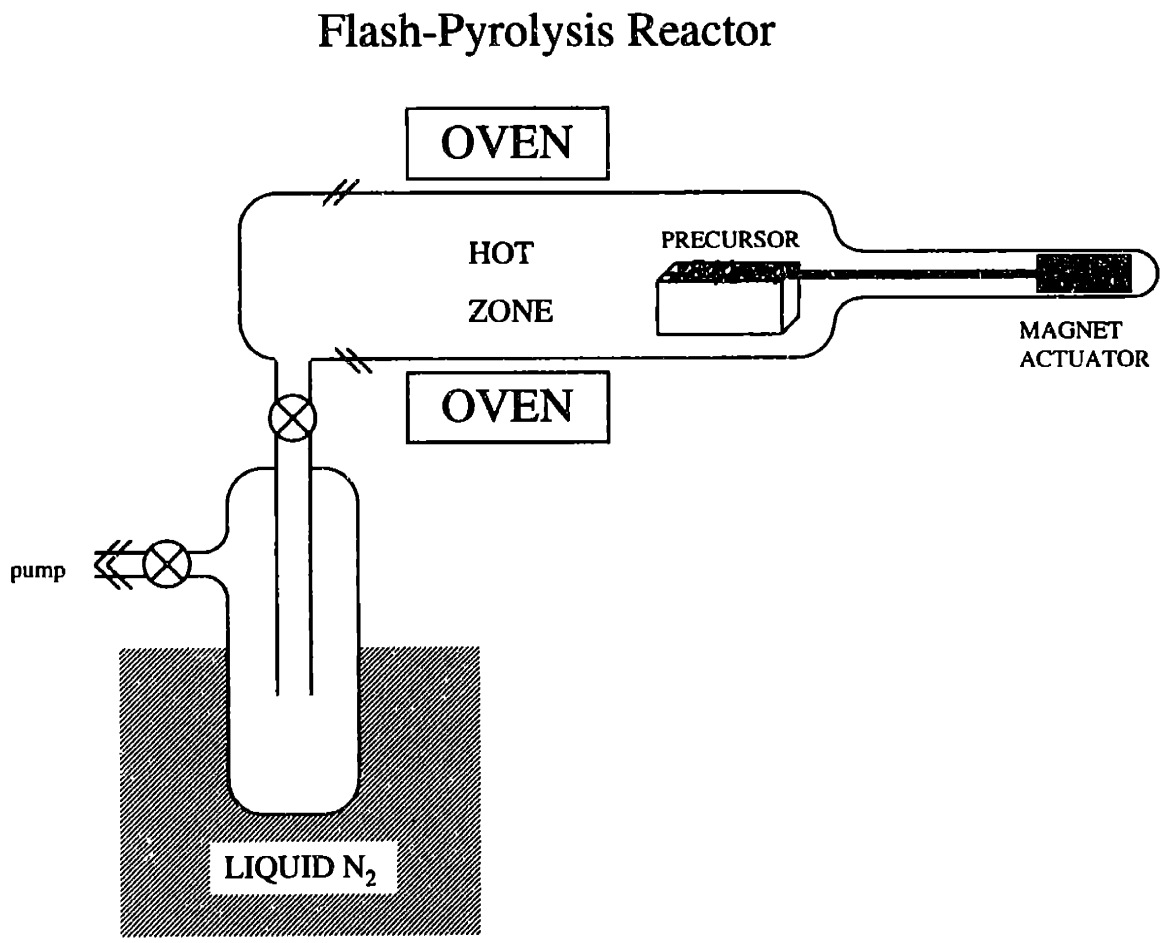


Figure 3.3 Flash-pyrolysis apparatus.

An FTIR spectrum of the raw product recorded to identify the necessity of further purification by freeze-pump-thaw cycles. The main undesirable contaminant is NCCN which has a slightly higher volatility. Therefore, following the collection of the raw product in the second of the two liquid nitrogen cooled vials (the second vial was chosen here to avoid contamination by the less volatile products of the precursor's decomposition, which are usually deposited on the walls of the first vial), a distillation has been performed. The receptor vial was kept at liquid nitrogen temperature, while the vial containing raw product was immersed in the n-pentane/N_2 bath (131°C). Distillation was done for 3 hours, with several interruptions to thaw/freeze the raw product vial. An FTIR spectrum of the distilled product was recorded (see Figure 3.4) During the first syntheses, the distillation procedure had not been perfected, which resulted in contamination of the sample. The thin line on Figure 3.4 shows a spectrum of such a contaminated sample. A strong CO_2 line overlapping with a CNCN line at about 2060cm^{-1} is easily seen in addition to some lines attributed to other by-products of the synthesis. The thick line on Figure 3.4 shows an essentially pure sample. The spectrum was recorded with a different, less sensitive photodetector. Still, the CNCN signal is much stronger and fewer contaminants are present.

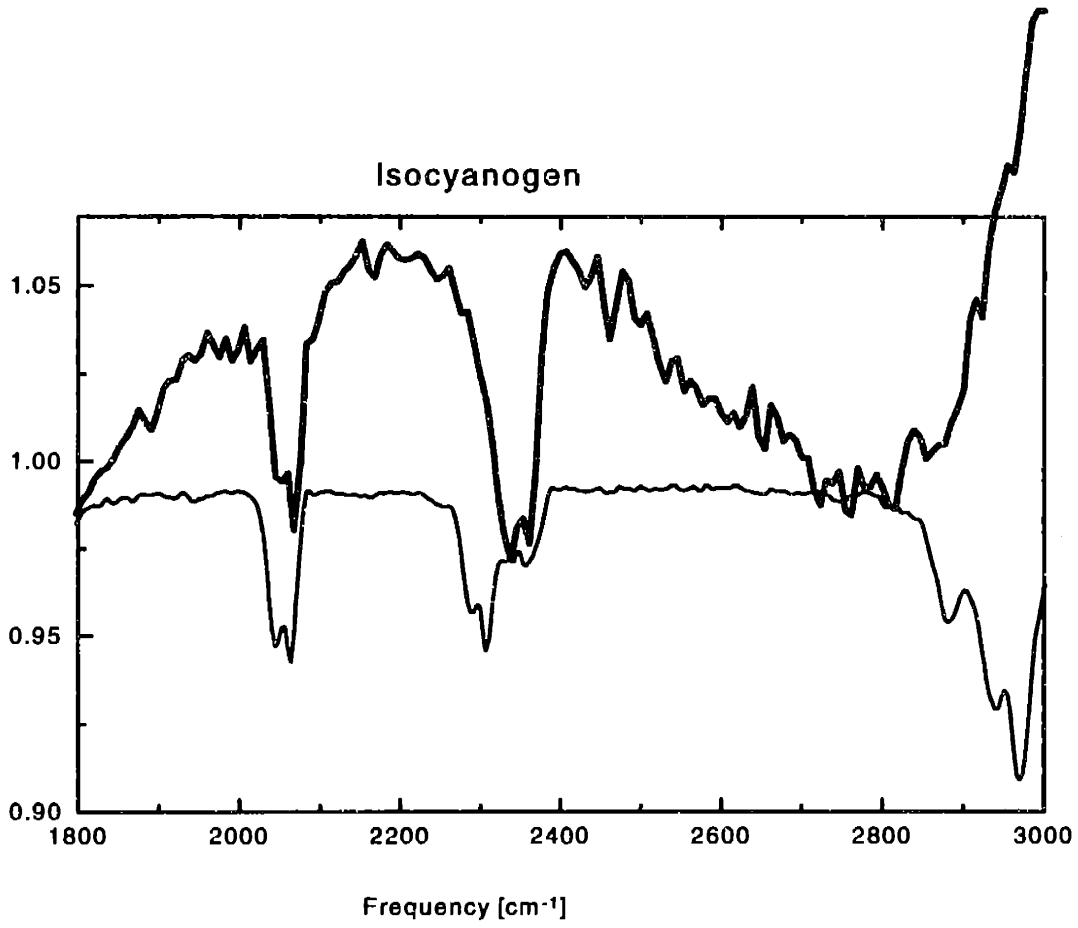


Figure 3.4 FTIR spectrum of CNCN.

Isocyanogen bands are be clearly seen at 2080cm^{-1} and 2380 cm^{-1}, confirming the success of our synthesis. For comparison we recorded the spectrum of cyanogen (Figure 3.5). It is quite clear that no trace of cyanogen could be noticed in any FTIR spectra of the product of our synthesis after purification by distillation.

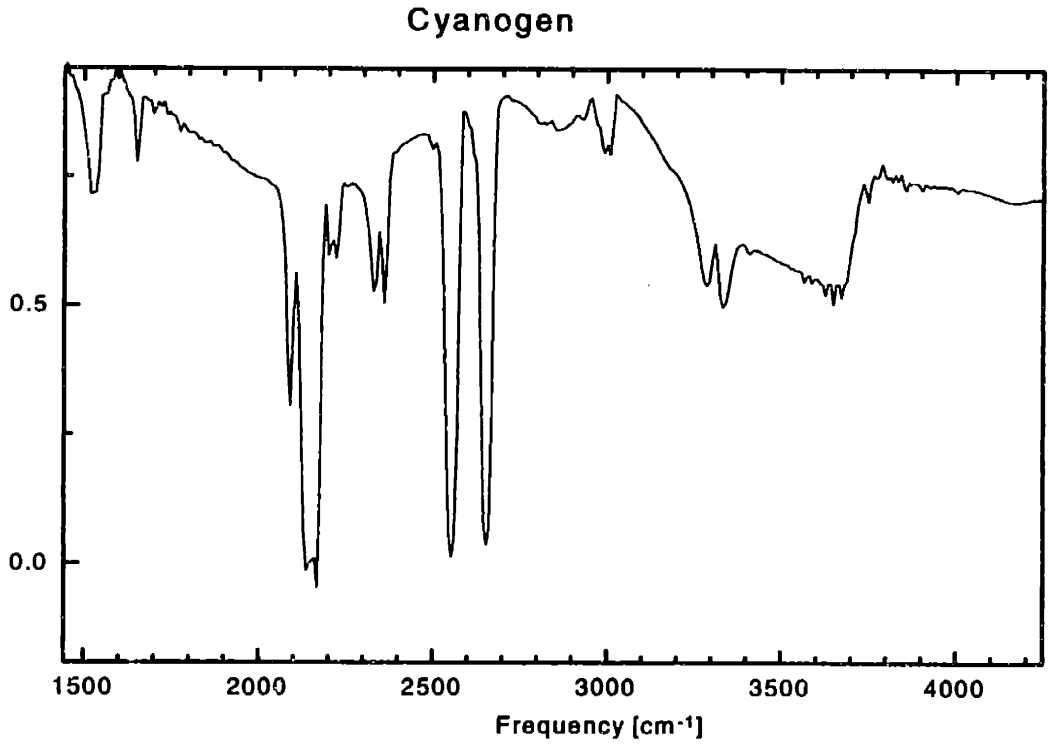


Figure 3.5 FTIR spectrum of cyanogen

3. Safety issues relevant to isocyanogen synthesis.

The reactor is pumped by a roughing pump with its exhaust connected to lab air exhaust, and the products of reaction are collected in two traps cooled with liquid nitrogen. The whole setup is placed in the hood, so that even if the reactor or any other part breakes, people in the room are not exposed to CNCN. Iso-cyanogen's toxicity owes to its conversion to HCN when it reacts with atmospheric water. Other possible products of reaction are identified as much less poisonous than iso-cyanogen.

Some of the work on the iso-cyanogen (spectroscopic study) must be conducted outside the hood. To reduce the danger of exposure to the substance, in case of leak or container breakage, the gas-containing parts of the setup are placed inside a plastic (transparent Plexiglas) box which is kept under negative pressure by connecting it to

vacuum exhaust lines. Whenever transporting the isocyanogen, secondary non-breakable container is used.

An HCN monitor is available from the MIT Industrial Hygiene Office every day until 5pm, so all the changes to the setup involving gas handling system must be completed within this time and in the presence of this monitor. An HCN antidote kit, containing amyl nitrite, is kept in the lab in a lockable container, and must be present at the workplace when working with CNCN. The treatment by the Cyanide Detoxification Kit from Lilly Pharmaceuticals, which includes injectable doses of sodium thiosulfate and sodium nitrate, is administered by MIT Health Center in case of emergency. Administration of hyperbaric oxygen is also indicated in cases of acute poisoning, but it should be noted that O_2 by itself has little detoxification ability.

Necessary protective clothing - neoprene gloves, lab coat and goggles - must be worn at all times. If clothing gets contaminated it must be removed. All the work must be done in the presence of another person and the other person has be fully informed about the use of the antidote kit and other first aid procedures.

As a precaution, work is done with quantities of isocyanogen that will be smaller than dangerous ones.

If we assume that CNCN is twice as poisonous as HCN (one molecule of CNCN makes two of HCN), and knowing that PEL(OSHA) is set at 11mg/m^3 , and estimate the volume of the room where the synthesis will be conducted as 792 m^3 , then the maximum amount of CNCN we plan to have at any time is limited to (roughly) 8 gram. We limit the amount of CNCN we synthesize and store to less than 100mg.

References

1. Van der Does, T. and Bickelhaupt, F. *Angew. Chem.* **100,** 998 (1988).
2. Stroh, F. and Winnewisser, M. *Chem. Phys. Lett.* **155,** 21-26 (1989).
3. Yamada, K. M. T. et al., *Chem. Phys. Lett.* **160**(2), 113 (1989).
4. Vaccaro, P., *private communication.*

3.3 Theoretical Considerations of Isocyanogen Spectroscopy.

In this section we would like to predict, at least semi-quantitatively, the wavelength location of the expected isocyanogen $\tilde{A}\,^1A'' \leftarrow \tilde{X}\,^1\Sigma^+$ transition. The frequency corresponding to this transition can be calculated from the difference between the upper state and lower state energy levels:

$$\bar{\nu} = T_0 + G'(v') + F'(J') - G''(v'') - F''(J''), \qquad [3.1]$$

where T_0 is the energy difference between the zero-point energies of two electronic states, G's are vibrational energies, and F's are rotational energies.

There is, on the other hand, a need to determine, at least qualitatively, which transitions are likely to be the strongest, in order to narrow the focus of our search. It is reasonable to assume that the transitions between the vibrational states that have good Franck-Condon overlap will be the most intense. In the absence of reliable FC calculation we will use quantitative arguments to identify these states. Let us compare the geometries and vibrational normal modes of the $\tilde{X}$ and $\tilde{A}$ states(after Sherrill and Schaefer[1]). Figure 3.6 shows the structure of the ground state of isocyanogen.

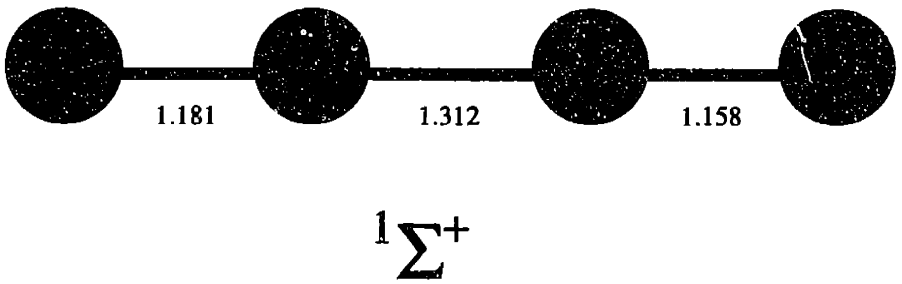


Figure 3.6 Equilibrium geometry of the ground electronic state of isocyanogen. Bond lengths are in angstroms and angles are in degrees.

Isocyanogen in its ground electronic state is a linear molecule. It has five normal modes, three nondegenerate(stretches) and two doubly degenerate (bends). The vibrational frequencies of these modes are given in Table 3.2.

Table 3.2 Fundamental vibrational frequencies for CNCN $\tilde{X}\,^1\Sigma^+$. Vibrational frequencies are in cm^{-1}.

Mode	CISD	Expt.
ν_1'' (C_2–N_2)	2333	2303
ν_2'' (C_1–N_1)	2076	2060
ν_3'' (N_1–C_2)	902	975
ν_4'' (N_1–C_2–N_2)	473	464
ν_5'' (C_1–N_1–C_2)	187	195

The first excited singlet state of isocyanogen was predicted to have a bent geometry (see Figure 3.7).

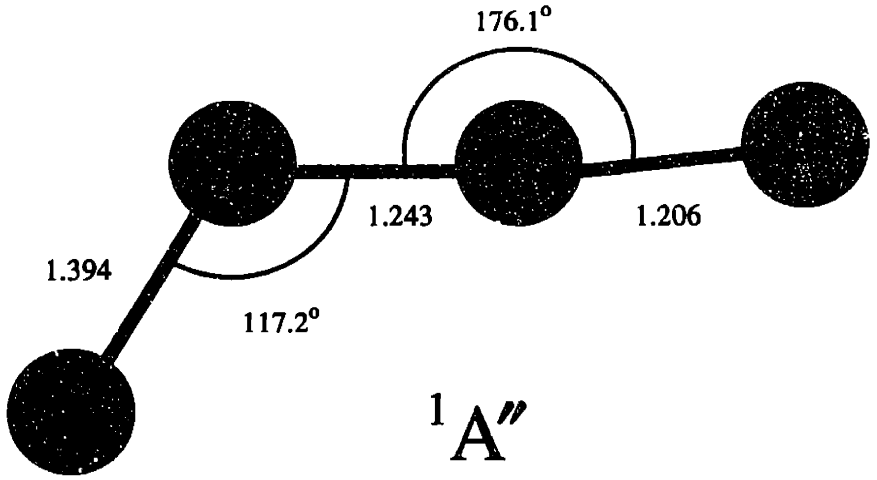


Figure 3.7. Predicted equilibrium geometry of the $\tilde{A}\,^1A''$ excited electronic state of isocyanogen. Bond lengths are in angstroms and angles are in degrees.

In the excited state isocyanogen has six normal modes. Vibrational frequencies are presented in Table 3.3.

Table 3.3 Fundamental frequencies calculated for the first excited singlet state of isocyanogen. Vibrational frequencies are in cm^{-1}.

Mode	CISD
ν_1' (C_2–N_2)	1841
ν_2' (C_2–N_1)	1239
ν_3' (N_1–C_1)	979
ν_4' (N_1–C_2–N_2)	605
ν_5' (C_1–N_1–C_2)	207
ν_6' (torsion)	502 (SCF)

By simple inspection we can see that the largest changes in geometry in going from the ground state to the excited state are in the length of the C_1-N_1 bond and in the C_1-N_1-C_2 bond angle. We will assume that the vibrational modes that correspond to these changes in geometry will give rise to the Franck-Condon (FC) active transitions. In addition, if we compare the size of the vibrational quantum of the ν_5'' bend with kT at room temperature (roughly $208cm^{-1}$), we will see that transitions corresponding to hot bands should be observable as well. In fact, they may be even stronger than cold band because of more favorable FC factors.

To calculate frequencies of (rotationless) vibronic transitions we will T_0=41,100cm^{-1} as given by Sherrill and Schaefer and calculate vibrational energies as:

$$G_v' = \sum_i \omega_i' \cdot v_i'; \quad G_v'' = \sum_i \omega_i'' \cdot v_i''. \qquad [3.2]$$

Using formula [3.1], and taking into account only vibrations of the FC active modes, we obtain the frequencies of these FC-favored transitions. They are given in Table 3.4.

Table 3.4 Frequencies and Boltzmann intensity factors for Franck-Condon active absorption transitions.

ω_5 "	ω_5'	ω_3 "	ω_3'	frequency (cm^{-1})	Boltzmann intensity factors
0	0	0	0	41100	1
0	1	0	0	41307	1
0	2	0	0	41514	1
0	3	0	0	41721	1
0	4	0	0	41928	1
0	0	0	1	42079	1
0	0	0	2	43058	1
0	1	0	1	42286	1
0	2	0	1	42493	1
0	3	0	1	42700	1
1	0	0	0	40905	0.39
1	1	0	0	41112	0.39
1	2	0	0	41319	0.39
1	3	0	0	41526	0.39
1	4	0	0	41733	0.39
1	1	0	1	42091	0.39
1	2	0	1	42298	0.39
1	3	0	1	42505	0.39
2	0	0	0	40710	0.15
2	1	0	0	40917	0.15
2	2	0	0	41124	0.15
2	3	0	0	41331	0.15
2	4	0	0	41538	0.15
2	1	0	1	41896	0.15
2	2	0	1	42103	0.15
2	3	0	1	42310	0.15

For better visualization these transition frequencies are displayed in Figure 3.8.

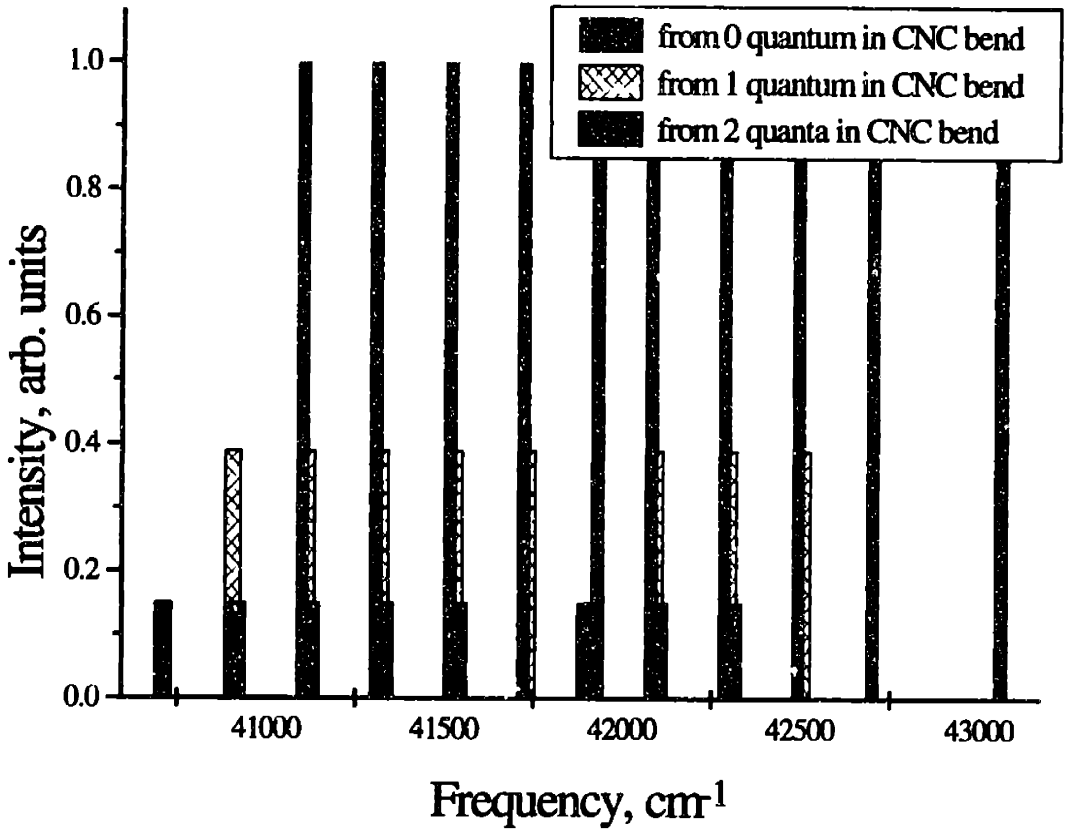


Figure 3.8 Frequencies of rotationless $\tilde{A}\,^1A'' \leftarrow \tilde{X}\,^1\Sigma^+$ electronic transitions starting with zero, one and two quanta in the C-N-C bend vibration. "Intensity" of the lines indicates only the population of the level from which the transition originates, due to the Boltzmann factor, and does not reflect the actual strength of the transition..

It is also interesting to note, by comparing the equilibrium CNCN geometries for the ground and first excited singlet electronic states with that of the transition state (TS) complex on the ground potential energy surface (isomerization to cyanogen pathway), that the very same vibration that we found to be FC active in the absorption will, most likely, be giving good FC access from the $\tilde{A}$ state to the isomerization barrier region on $\tilde{X}$ the state. Figure 3.9 shows the Transition State geometry calculated by Sherrill and Schaefer[1].

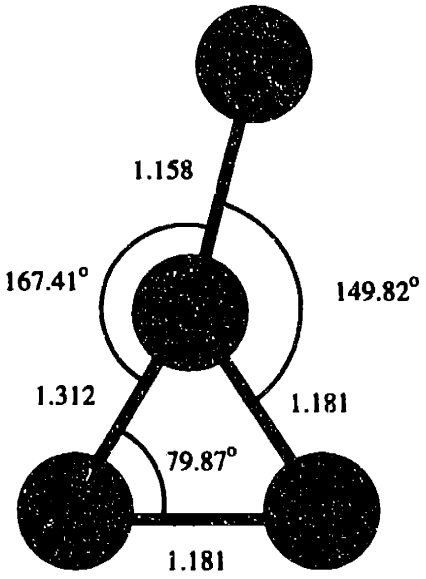


Figure 3.9 Ground electronic surface CNCN↔NCCN isomerization Transition State Geometry (after Sherrill and Schaeffer[1]). Bond lengths are in angstroms and angles are in degrees.

Now, that we have, semi-quantitatively, suggested y the form of the rotationless electronic spectrum of isocyanogen to look, we can compute the rotational structure of these transitions.

The ground state is linear, so its rotational energy levels can be expressed as[2]

$$F''(J'') = B''J(J+1) + gl^2. \qquad [3.3]$$

The upper state is a near prolate symmetric top, so its energy levels can be expressed as[2]

$$F'(J,K) = \overline{B}'J(J+1) \pm \frac{q'}{2}J(J+1) + (A' - \overline{B}')K^2, \qquad [3.4]$$

where

$$\overline{B}' = \frac{B'+C'}{2}, \text{ and } q' = \frac{B'-C'}{2}. \qquad [3.5]$$

To find the rotational energies we have to calculate values of the the rotational constants for the ground and excited electronic states. Using the geometry provided by Sherrill and Schaefer[1] we find for the ground state I_B=97.67 amu $\cdot$ Å^2, so $B'' = 0.1726\ \text{cm}^{-1}$.
For the excited electronic state (using the geometry provided by Sherrill and Schaefer[1]) I_A=6.173 amu $\cdot$ Å^2, I_B=86.62 amu $\cdot$ Å^2 and I_C=93.44 amu $\cdot$ Å^2, therefore $A' = 2.731\ \text{cm}^{-1}$, $B' = 0.1946\ \text{cm}^{-1}$, $C' = 0.1804\ \text{cm}^{-1}$, $\overline{B}' = 0.1875\ \text{cm}^{-1}$, and $q' = 0.0071\ \text{cm}^{-1}$.

Therefore, for F' and F'' given by formulas [3.4], [3.5], and taking into account that $\Delta K = \pm 1$ ($K = l$, because $K = l + \Sigma$, and Σ=0 for the ground state of CNCN), we can write the energies of transitions for P,Q and R branches as:

$^rR_0$ branch:

$$\left(\overline{B}' - B'' \pm \frac{q}{2}\right)J^2 + \left(3\overline{B}' - B'' - \frac{3q}{2}\right)J + 2\left(\overline{B}' - \frac{q}{2}\right) + (A' - \overline{B}')K^2. \qquad [3.6]$$

$^rP_0$ branch:

$$\left(\overline{B}' - B'' - \frac{q}{2}\right)J^2 - \left(\overline{B}' + B'' - \frac{q}{2}\right)J + (A' - \overline{B}')K^2. \qquad [3.7]$$

$^rQ_0$ branch:

$$\left(\overline{B}' - B'' + \frac{q}{2}\right)J(J+1). \qquad [3.8]$$

The rotational intensity distribution in the absorption band can be obtained from the expression[4]

$$I_{KJ} \propto A_{KJ}\, g\, e^{-F(J)/kT}, \qquad [3.9]$$

where A_{KJ} is the Honl-London factor, g is the statistical weight of lower state and $F_l(J)$ is the rotational energy of the lower state for a particular J.

The expression for calculation of Honl-London factors is given by Zare[5]. The statistical weight for the lower state is just $2J''+1$. For simulation of the spectrum we have chosen to use the "SpecSim" program[3]. Result of the simulation is given in Figure 3.10.

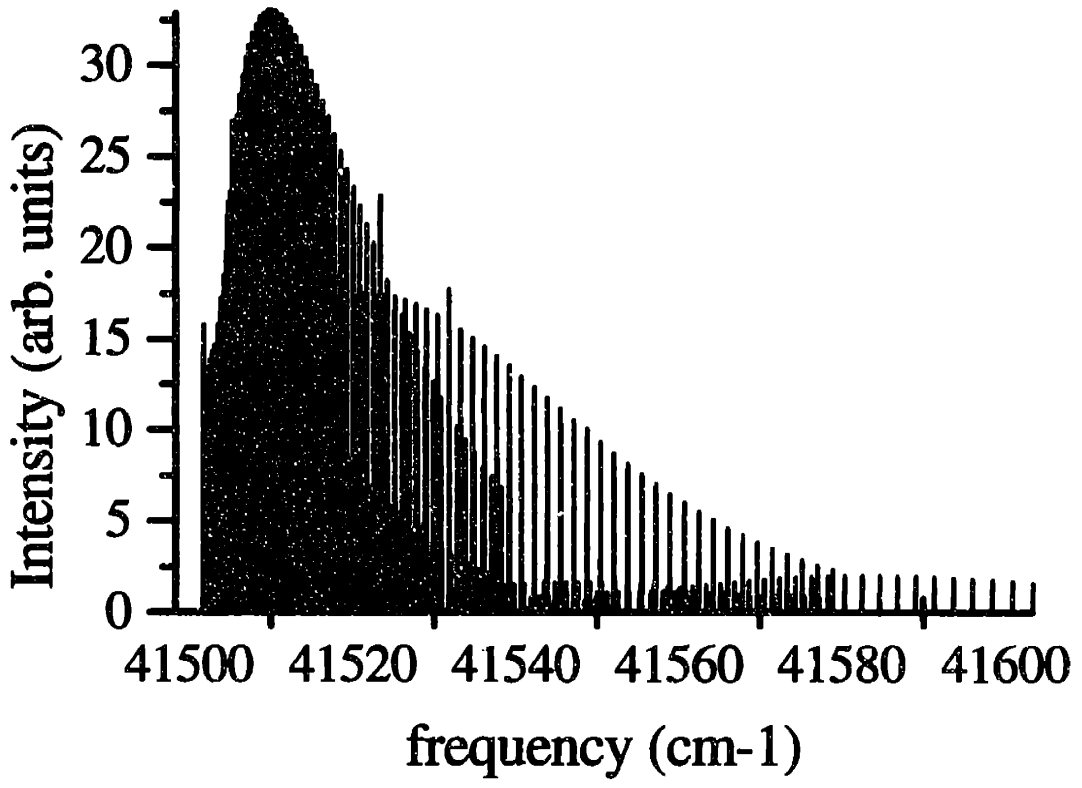


Figure 3.10 Simulation of the rotational spectrum of CNCN for the origin band transition. Calculated for *J* values up to 60.

Now that we have a prediction of what spectral region we need to search for CNCN spectrum and what we expect to see, we can proceed to the next section, describing our experimental efforts directed at detection of this spectrum.

References

1. Sherrill, C. D. and H.F.Schaefer III, *J. Chem. Phys.* **100** (12), 8920 (1994).
2. Herzberg, G., "Molecular Spectra and Molecular Structure", Vol. II, 2nd ed., Krieger Publishing Co., 1991.

3. Panov, S.I., "SpecSim - spectrum simulating program", private communication.
4. Bernath, P., "Spectra of Atoms and Molecules", 1st ed., Oxford University Press, 1995.
5. Zare, R.N., "Angular Momentum: Understanding Spatial Aspects in Chemistry and Physics", 1st ed., John Wiley & Sons, 1988.

3.4 Experimental Investigation of Isocyanogen. Conclusions.

Our first series of attempts to record the spectrum of isocyanogen was conducted in collaboration with Prof. Yoshino of Harvard Smithsonian Astrophysical Laboratory. The 6.65 meter grating spectrometer (McPherson, Inc.), with detection on glass photographic plates was used in the experiment. Prior to use, the spectrometer was refocused and calibrated with CO emission spectra. During the experiments the sample was kept in a 130cm long static quartz cell. The low sensitivity of the technique (smallest detectable absorption would correspond to 1% fractional absorption) was a topic of concern from the very beginning of the experiment. The hope was, though, that it would be sufficient to see the spectrum.

After each synthesis, before the sample was transported to the Smithsonian laboratory for the study, FTIR spectra were recorded to verify the presence of isocyanogen and its purity. The distillation process described in the Section 3.2 had not been perfected during the first syntheses, which, we believe, led to the detection of the transitions reported in the bold font in the Table 3.5. FTIR spectra of the sample were also recorded after each UV experiment in order to study stability of isocyanogen under UV irradiation and storage. Spectra recorded before the UV experiment and after it were essentially the same - contained the same amount of isocyanogen (qualitatively), the same amount of impurities (in the first experiments) and never contained any detectable trace of cyanogen.

At this point it is important to explain the contents of the Table 3.5 in greater detail. The first two columns are cyanogen lines from Pearse and Gaydon[1]. Wavelengths are given in Angstroms and the intensity of each line is given in arbitrary units that range from 0 to 10.

The next column contains the lines in the spectrum of cyanogen that we recorded in order to have one more calibration for the experiment and to have information about the strength of cyanogen transitions. Almost all of the lines we observed can be matched with the lines in the spectrum from Pearse and Gaydon[1].

The next two columns are presumably spectra of isocyanogen. The first spectrum, recorded on June 16, was underexposed in the shorter wavelength region, and hence no lines can be discerned there. The second spectrum, recorded on June 25, was weaker in the longer wavelength region, and hence weaker lines do not show up on it.

In both cases, the purified product of the isocyanogen synthesis did not show any of the lines reported in the table in bold. These lines resulted when we introduced the byproducts of the synthesis together with the purified product into the cell. In other words, both syntheses yielded products that absorbed in the region from 2300Å to slightly beyond 2400Å. These lines correspond to neither known lines of cyanogen nor CO.

Although the isocyanogen synthesis does yield cyanogen, we were able to separate the two isomers by distillation (see section 3.2, Figures 3.4, 3.5).

On the other hand, as we can see by the matching lines on the spectra we obtained when only isocyanogen was in the cell, the spectrum of cyanogen was obtained whenever we attempted to record the spectrum of isocyanogen in the 210-230nm spectral region. On the other hand, when, after each spectroscopic study the sample was collected back into the vial and transported to MIT, where an FTIR spectrum was taken, no trace of cyanogen was detected - only isocyanogen, and in the approximately same amount as before the UV experiment.

Table 3.5 Spectral lines observed while recording spectrum of CNCN. See discussion in the text.

P+Q λ(Å)	I	NCCN λ(Å)	Q16 λ(Å)	Q25 λ(Å)
2007.0	10	2007.0		
		2021.0		2021.0
2035.2	7	2034.7		2034.7
2054.6	7	2054.2		2054.2
		2056.0		2056.0
		2074.3		2074.3
2093.1	10	2093.0		2093.0
2107.4	7	2107.6		2107.6
2125.0	7	2125.3		2125.3
2137.7	4	2137.7		
2145.6	4	2146.2	2146.2	
2164.5	4	2165.2	2165.2	
2188.2	9	2188.5	2188.5	2188.5
2200.0	5	2200.8	2200.8	2200.8
2226.4	5	2226.8	2226.8	2226.8
2237.7	3	2237.8	2237.8	
			2313.4	
			2324.9	
			2367.7	**2367.7**
			2375.6	
			2416.3	**2416.3**

One possible explanation of these strange observations is that due to the peculiar design of the spectrometer, light from the source (hydrogen discharge, 3 Torr, 300mA current) had to first be directed through the cell with the sample. Only after it passed the sample cell and entered the spectrometer, was it dispersed on the grating and recorded on the photoplate. This led to exposure of isocyanogen in the cell to very intense broadband UV radiation, which could cause some of it to isomerize to cyanogen. Because the cyanogen bands in the UV are relatively strong, even small amounts of cyanogen could be detected in our experiment. On the other hand, it might mean that the amount of cyanogen necessary for UV detection is, apparently, below the detection capability of our FTIR setup. This means that cyanogen has to have a much stronger UV absorption than isocyanogen and/or a much weaker IR spectrum. The latter is very unlikely.

For easier comparison, stick spectra summarizing the four spectra presented in Table3.5 are plotted in Figure 3.5.

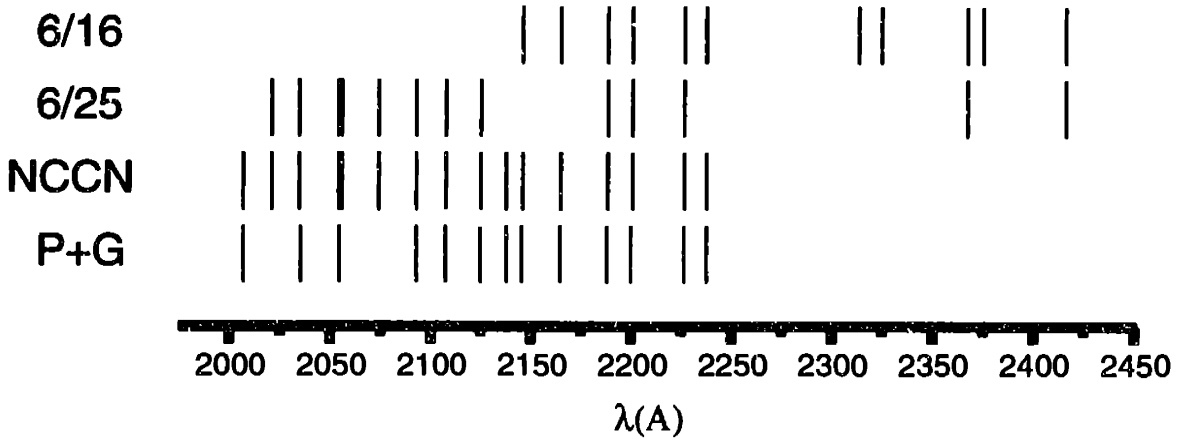


Figure 3.5 Comparison of the spectra recorded in the 6.65meter grating spectrometer with the spectra of cyanogen tabulated by Pearse and Gaydon[1].

The 5 lines from 2313Å to 2416Å are spaced regularly in energy and were first suspected to be a signature of isocyanogen. The energy differences between them, in cm^{-1}, are 213, 706, 212, and 709, as shown in the Table3.6:

Table3.6. Energy spacing of the unassigned lines originally assumed to belong to isocyanogen.

λ(Å)	wavenumber (cm^{-1})
2313.4	43,226
	↑
	213
	↓
2324.9	43013
	↑
	706
	↓
2363.7	42,307
	↑
	212
	↓
2375.6	42,095
	↑
	709
	↓
2416.3	41,386

After the distillation of the product of the pyrolysis had been done properly and the spectrum was recorded between 230nm and 250nm, no lines were observed. The very first experiments were conducted with less then 2 Torr of CNCN in the cell, while the final ones had about 6 Torr of CNCN in the sample cell.

During the time of these studies isocyanogen proved to be a relatively stable gas, surviving for several days in a glass vial at room temperature and under intense UV radiation.

All the aforementioned led us to believe that a technique, capable of higher detection sensitivity, was necessary to study the spectroscopy of the $\tilde{A} - \tilde{X}$ transition of

isocyanogen. Section 4 will report on our development of an improved version of Cavity Ring Down spectroscopy, which will be instrumental in our further studies of CNCN.

References

1. Pearse, R.W.B., Gaydon, A.G., "The Identification of Molecular Spectra", 4th ed., John Wiley & Sons., 1976.

4. Cavity Ring Down Spectroscopy in the Ultra-Violet

4.1 Introduction

As we have discussed earlier (Section 2.1), direct detection of laser light absorption by molecules is preferable to other spectroscopic methods since it can be successfully applied even when molecular excited states do not fluoresce (which is necessary for LIF) and cannot be easily ionized (which is required for REMPI detection). Furthermore, when absorption spectroscopy is applied to the study of high pressure samples, such as flames and plasmas, where molecules fluoresce and LIF can be used, absorption based techniques are capable of extracting quantitative absolute concentration data, which is difficult by LIF due to difficulties associated with collisional deactivation of the fluorescing state. Additionally, unlike some other highly sensitive techniques (e.g. four-wave mixing), direct absorption is a linear spectroscopic technique both in molecular concentration and in absorption strength.

A very important drawback of conventional absorption spectroscopy comes from the fact that in order to make sensitive absorption measurements, one must be able to distinguish between source intensity fluctuations and those introduced by molecular absorptions. This problem is particularly serious for pulsed lasers for which pulse-to pulse fluctuations can easily by as large as 10%. Although it is possible to normalize these fluctuations to some extent, detection of an absorption resulting in less than a part in 10^3 reduction of intensity is highly problematic[1].

Cavity Ring Down spectroscopy (CRD) is a relatively new, versatile, high sensitivity absorption technique, which in recent years has gained a lot of popularity in the spectroscopic community. CRD was introduced, as a sensitive gas phase absorption technique employing pulsed lasers for measuring of molecular absorptions, by O'Keefe and Deacon[2] in 1988. The technique has a sensitivity comparable to that of photoacoustic spectroscopy (PAS) and intracavity laser absorption spectroscopy (ICLAS). Its major advantage over PAS is that it does not require indirect and inaccurate calibration

procedures to extract the absorption coefficient of the sample. The strongest advantage CRD has over ICLAS is in the spectral coverage it can achieve. While CRD is limited by the existence of high reflectivity mirrors which cover particular spectral range, ICLAS depends on the availability of a gain medium suitable for creation of a laser working in that spectral range. Additionally, one of the most attractive features of CRD is its conceptual and practical simplicity. CRD may be considered a multipass technique with pathlengths of several orders of magnitude higher than those in traditional multipass cells. The method is immune to laser source intensity fluctuations because it is based on the measurement of the rate of decay (absorption) of light field, confined within a stable optical cavity, rather than on the magnitude of this decay (absorption). This decay time is determined by cavity losses due to absorption by gas sample present in the cavity, scattering by this gas sample (which at low concentrations can be neglected), and losses on the mirrors.

One significant limitation CRD has had until recently was in its reliance on very high reflectivity mirrors, typically R=0.9999 or better. This seriously impaired the technique's use in the UV region of spectrum (mirrors with R>0.997 are not available) and far IR (high reflectivity mirrors are even more difficult to make).

In the following sections we will briefly discuss previous work (Section 4.2), give theoretical background of the technique (Section 4.3), describe experimental setup constructed to perform a spectroscopic study of the first excited singlet electronic state of iso-cyanogen (Section 4.4) , and discuss sensitivity considerations and improvements that have been done to enable us to use CRD in the ultraviolet for quantum noise limited detection.

4.2 Previous Work.

To my knowledge, the technique in its most widely used form (operating with pulsed laser sources) was introduced by O'Keefe and Deacon[2] in 1988, when they detected forbidden transitions of gaseous molecular oxygen in the visible region. In that work the authors showed that absorption losses, due to the sample, on the order of 10^{-6} per pass can be detected by CRD. Although, before that publication, there was experimental work employing similar principles[3,4,5,6], previous efforts were concentrated on accurate characterization of resonator mirrors and the importance of the CRD technique for spectroscopic applications was not recognized.

Shortly after the original CRD work, the technique was applied to optical spectroscopy in a molecular beam by Saykally and co-workers[7]. Vibronic spectra of a variety of transition metal clusters were obtained. These studies have since been expanded to include copper dimers and trimers and various metal silicides[8]. Although that work encountered significant difficulties caused by large pulse-to-pulse intensity fluctuations of laser generated ablation molecular beams, it demonstrated the feasibility of the use of CRD with molecular beams.

Benard and Winker applied CRD to detection of either optical gain or loss in laser induced chemical reactions[9]. In their experiment, a flow reactor was transversely pumped by a CO_2 laser. Both pump beam direction and flow direction were perpendicular to the confocal CRD cavity. The probe laser was an eximer pumped dye laser. Population inversion created by the pump laser in BiF was recorded.

The first application of ringdown spectroscopy to study of chemical kinetics was done by Lin and co-workers[10] in 1993 when they studied decomposition of phenyl radical. The concentration of phenyl radical, which was produced by photolysis of nitrobenzene, was followed using CRD. To obtain the rate constant for this reaction, the absorption term in the CRD signal expression (see Section 4.3) was substituted for the exponentially decaying concentration function describing pseudo-first order kinetics of phenyl radical decomposition. A second order rate coefficient was found from the slopes of plots of pseudo-first order rate coefficients versus concentration. The validity of this

approach was discussed and it was demonstrated that CRD can be used for kinetic studies as long as the characteristic time of the chemical process under study is significantly longer than the ring-down time. A number of works using CRD has since been published by M.C. Lin and other groups[11,12,13].

Romanini and Lehmann have used CRD to obtain overtone spectra of HCN[14,15]. Those experiments were aimed at finding signatures of isomerization, but surprisingly have not found any even for relatively high vibrational excitations. The transitions they observed were for the 205, 106, 007, 305, 206 and 107 overtone and combination bands (first number is the CN stretch, second is the bend and the third is CH stretch). One possible explanation is that with no excitation in the bend, the likelihood of isomerization is relatively small. These works were also interesting in that the authors obtained overtone intensities that were different from those obtained previously by photoacoustic spectroscopy, but very similar to those obtained as a result of *ab initio* calculations, thus demonstrating the validity of the claim that CRD is better suited for absolute measurements. The authors also used Brillouin scattering in a methanol cell to eliminate amplified spontaneous emission (ASE) from their pump laser beam. In the first of their two papers the authors discussed the quantum noise limitation on the sensitivity of the CRD method. Although the authors claimed that the detection sensitivity of their experiment was quantum noise limited, they made this claim on the basis of an incomplete analysis of the noise sources in their experiment (this will be discussed in Section 4.5). Additionally, in a more recent paper[16] the authors gave a thorough description of the physical principles of the method and applied stable resonator theory (see e.g. Yariv[17] or Siegman[18]) to CRD.

Meijer et al[19], and later Cheskis et al[20], have used CRD in the ultra-violet region to study OH radicals in flames. Even though only low reflectivity mirrors were available, sensitivity of $2x10^{10}$ molecules/cm^3 was reported[20]. The authors[20] also discussed the use of spectroscopic data for accurate measurements of concentration and temperature profiles in a flame. Meijer and co-workers also experimentally studied the interaction of a narrow-band laser with the cavity modes, in support of the discussion by Lehmann and Romanini[16].

In another recent work, Meijer and co-workers[21] managed to extend their measurements deeper into the UV, using CRD for the determination of the absolute transition strength of the Cameron band of CO (at 206nm). The authors reported sensitivity of $5x10^{-5}$, even though only relatively low reflectivity mirrors (R=98.5%) were available.

Engeln and Meijer[22] have recently reported a technique which is a result of combining CRD and FT spectrometry. In that work broadband laser radiation is coupled into the cavity and then directed into a standard FT instrument. An interferogram is obtained in combination with the decay of the signal characteristic of the losses in the cavity. The authors have achieved a sensitivity which is comparable with those of multipass FT systems. Another very interesting development from the same group is the coupling together polarization spectroscopy (PS) and CRD[23]. Just as in traditional CRD spectroscopy, where the rate of light absorption is measured, in PS-CRD the rate of polarization rotation is determined. In addition to the study of electro-optic and magneto-optic phenomena in gaseous species, the method has been demonstrated to apply to the study of such phenomena in transparent solids as well.

Zare and co-workers[24] has been using CRD to detect methyl radical in a hot-filament flow reactor, observing the 216nm band. The spatial concentration profiles of methyl radical relative to the filament were obtained from the intensity of the absorption spectrum recorded. Vibronic band spectra obtained in that work exhibited a signal to noise ratio comparable to that obtained by the degenerate four wave mixing method reported earlier by Sick and Farrow[25].

Recently CRD has been extended further into the mid-IR spectral range (1.6 μm and 3.3 μm) using a single mode OPO laser system[26]. A comparison between photoacoustic and mid-IRCRD methods was reported[26]. Paul et al[27]. have used infrared CRD to study water clusters, obtaining vibrationally resolved spectra for jet cooled $(H_2O)_n$, for n up to five.

Finally, it is interesting to note that one of the first applications[6] (dating back to 1984) of CRD was based on a cw radiation source. Anderson and co-workers exploited occasional coincidences of a He-Ne laser with the cavity modes of their cavity to

characterize the reflectivity of their mirrors. These authors could not tune their laser nor did they try to obtain any absorption spectra. Almost twelve years passed before CRD spectroscopy utilizing cw-laser sources was done[28]. As a demonstration, overtones of acetylene were measured around 570nm. Obtained spectra are Doppler limited and display a rms noise-equivalent absorption of 10^{-9} cm^{-1} (or $5x10^{-8}$ per pass through the sample). The most promising applications include high resolution spectroscopy at low pressures, sub-Doppler spectroscopy in a supersonic jet and various trace detection applications using inexpensive diode laser technology.

All of the aforementioned studies clearly show the wide applicability and high sensitivity achievable by CRD spectroscopy. In the following sections I will describe our application of CRD along with improvements we deem important for sensitivity improvements in the spectral regions where very high reflectivity mirrors are not available.

References

1. Weisman, R.B., Selco, J.I., Holt, P.O., Cahill, P.A., *Rev. Sci. Instr.*, **54**, 284 (1983).
2. O'Keefe, A., Deacon, D.A.G., *Rev. Sci. Instr.* **59**, 2544 (1988).
3. Herbelin, J.M. et al. *Appl. Optics* **19**(1), 144 (1980).
4. Herbelin, J.M, McKay, J.A. *Appl. Opt.*, **20**(19), 3341 (1980).
5. Kwok, M.A., Herbelin, J.M., Ueunten, R.H., Opt., 21, 979 (1982).
6. Anderson, D.Z., Frish, J.C., Masser, C.S., *Appl. Optics* **23**, 1238 (1984).
7. O'Keefe, A. et al. *Chem. Phys. Lett.*, **172**(3), 241 (1990).
8. Scherer, J.J., Paul, J.B., O'Keefe, A., Saykally, R.J. *Chem Rev.*, **97**(1), 25 (1997).
9. Benard, D.J., Winker, K.B. *J. Appl. Phys.*, **69**, 2805 (1991).
10. Yu,T., Lin, M.C., *J. Am. Chem. Soc.*, **115**, 4371(1993).
11. Yu,T., Lin, M.C., *J. Phys. Chem..*, **98**, 2105(1994).

12. Yu,T., Lin, M.C., *J. Phys. Chem..*, **98**, 9697(1994).

13. Zhu, L., Johnson, G. *J. Chem. Phys.*, **99**, 15114 (1995)

14. Romanini, D., Lehmann, K.K. *J. Chem. Phys.*, **99**,6287 (1993).

15. Romanini, D., Lehmann, K.K. *J. Chem. Phys.*, **102**,633 (1995).

16. Lehmann, K.K., Romanini, D., *J. Chem. Phys.*, **109**,10263 (1996).

17. Yariv, A., *Quantum Electronics*, 3rd ed. John Wiley & Sons, 1989.

18. Siegmann, A.E., *Lasers*, Univ. Science Books, 1986.

19. Meijer, G., Boogaarts, M.G.H., Jongma, R.T., Parker, D.H., Wodtke, A.M. *Chem. Phys. Lett.*, **217**, 112 (1994)

20. Cheskis, S., Derzy, I., Lozovsky, V.A., Kachanov, A., Romanini, D. to be published.

21. Jongma, R.T., Boogarts, M., Holleman, I., Meijer, G. *J. Mol. Spec.*, **165**, 303 (1994).

22. Engeln, R., Meijer, G. Paper presented at the 51st Ohio State University Symposium on Molecular Spectroscopy, 1996.

23. Engeln, R.,Berden, G., van der Berg, E., Meijer, G. accepted for publication in *J. Chem. Phys.*, (1997).

24. Zaliski, P., Ma, Y., Zare, R.N., Wahl, E.H., Owano, T.G., Harris, J.S., Kruger, C.H. *Chem. Phys. Lett.*, **234**, 269 (1995).

25. Sick, V., Farrow, R.L. *Opt. Lett.*, **11**,111 (1995).

26. Scherer, J.J., et al. *Chem. Phys. Lett.* **245**, 273 (1995).

27. Paul, J.B., et al. *Science*, submitted.

28. Romanini, D., Kachanov, A.A., Sadeghi, N., Stoeckel, F. *Chem. Phys. Lett.*, **264**, 316 (1997)

4.3 Theoretical Background of Cavity Ring Down Spectroscopy.

A discussion of the principles of CRD and a detailed technical description of the method in its standard form can be found elsewhere[1,2,3,4]. They will be outlined here for the sake of completeness.

We consider a cavity ring down experiment in which the laser pulse length is shorter than cavity round trip time. This allows us to neglect interference effects inside the cavity. In addition, this is the most usual experimental arrangement used by the majority of laboratories practicing CRD.

If a light pulse of intensity I falls on the "entrance" mirror of the cavity, the light intensity that passes through the cavity will be $I \cdot T^2 e^{-\alpha(\omega)l}$, where T is a mirror transmittivity, $\alpha(\omega)$ is a frequency dependent absorption coefficient of the sample inside the cavity, and l is the cavity length. Because the absolute intensity of the first pulse is not important for recovering the value of the sample absorptivity (only the rate of change of intensity is important), we will temporarily assume that the intensity of the first pulse that passes through the cavity is I_0. With each roundtrip in the cavity, the light intensity decreases by a factor of $R^2 e^{-2\alpha l}$, where R is the reflectivity of a mirror. Therefore, after n roundtrips, the intensity of light falling on the detector is

$$I_n = I_0 \cdot R^{2n} \cdot e^{-2n\alpha l}, \qquad [4.1]$$

where I_0 is the light intensity on the photodetector after the first pass. We can rewrite this as

$$I_n = I_0 e^{2n(\ln R - \alpha l)}. \qquad [4.2]$$

Diffraction losses, Rayleigh scattering and absorption by the mirror substrates can be introduced as generic *mirror losses*, $\mathcal{L}$. All three characteristics of the *empty* cavity (R, T, and $\mathcal{L}$) are linked together by the energy conservation law:

$$R + T + \mathcal{L} = 1. \qquad [4.3]$$

Usually losses in a cavity are of the same order of magnitude as the transmittivity, while reflectivity, for very high quality dielectric coated mirrors is much higher than both.

Therefore we can expand the logarithm of reflectivity in a Taylor series, and retain only the most significant term:

$$I_n = I_0 e^{2n(\ln R - \alpha l)} \approx I_0 e^{-2n(T+\mathcal{L}+\alpha l)}. \qquad [4.4]$$

Now we will assume that after each laser shot we measure the magnitudes of peaks for each value of n (we will discuss the question of the design of a data acquisition system for a real-life CRD experiment and give a comparative analysis of different methods used in section 4.5). Then, rewriting formula [4.4] as

$$\ln\left(\frac{I_n}{I_0}\right) = -2(T+\mathcal{L}+\alpha l)n, \qquad [4.4a]$$

we can see that the combined losses of the cavity can be found from the slope of the plot of the logarithm of the ratio of the n-th peak intensity to the first peak's intensity versus twice the peak number $2n$. The sample absorptivity can be found from the difference in slope values for the signal decays recorded with and without the sample in the cavity. In other words, the ringdown signal is described by the first order decay envelope $e^{-t/\tau}$, with time constant τ. This decay constant is connected to the characteristics of the cavity as

$$\frac{1}{\tau} = \frac{c}{l}(T+\mathcal{L}) + c\alpha(\omega), \qquad [4.5]$$

where c is the speed of light. As we can see, this decay constant is a sum of a weakly frequency dependent term (the mirrors can be considered relatively broadband - 20-50nm) $c\alpha_c = \frac{c(T+\mathcal{L})}{l}$ and the sample absorption term $c\alpha(\omega)$. The first term, creating the baseline in the measurement of the spectrum, can be independently measured either before introducing the sample into the cavity or after removing the sample. Sample absorption is obtained in the units of cm^{-1} from the expression [4.5]. This expression is also convenient because it allows one to specify the sources of experimental error. For small signals we can write

$\frac{\delta\tau}{\tau} = \frac{\delta\alpha_{total}}{\alpha_{total}}$, where α_{total} is simply a sum of $\alpha(\omega)$ and α_c. Then, the noise equivalent absorption (absorption that causes the signal that is equal in magnitude to the noise) can be written as

$$\delta\alpha_{tot}^{noise} = \frac{\delta\tau}{c\tau^2}. \qquad [4.6]$$

In the following section, I will discuss the design of the experimental setup for CRD spectroscopy for the first excited singlet state of iso-cyanogen.

References

1. Romanini, D., Lehmann, K.K. *J. Chem. Phys.*, **99**,6287 (1993).
2. Lehmann, K.K., Romanini, D., *"The superposition principle and CRD spectroscopy", to be published).*
3. Zalicki, P., Zare, R.N. *J. Chem. Phys.*, **102**(7), 2708 (1995).
4. Romanini, D., Kachanov, A.A., Sadeghi, N., Stoeckel, F. *Chem. Phys. Lett.*, **264**, 316 (1997)

4.4 Experimental Setup: Design considerations and implementation.

Our experimental setup (see Figure 4.1) is very similar to that of Cheskis et al., except that we used a static glass cell instead of the low pressure burner and our cavity is 212.5cm long.

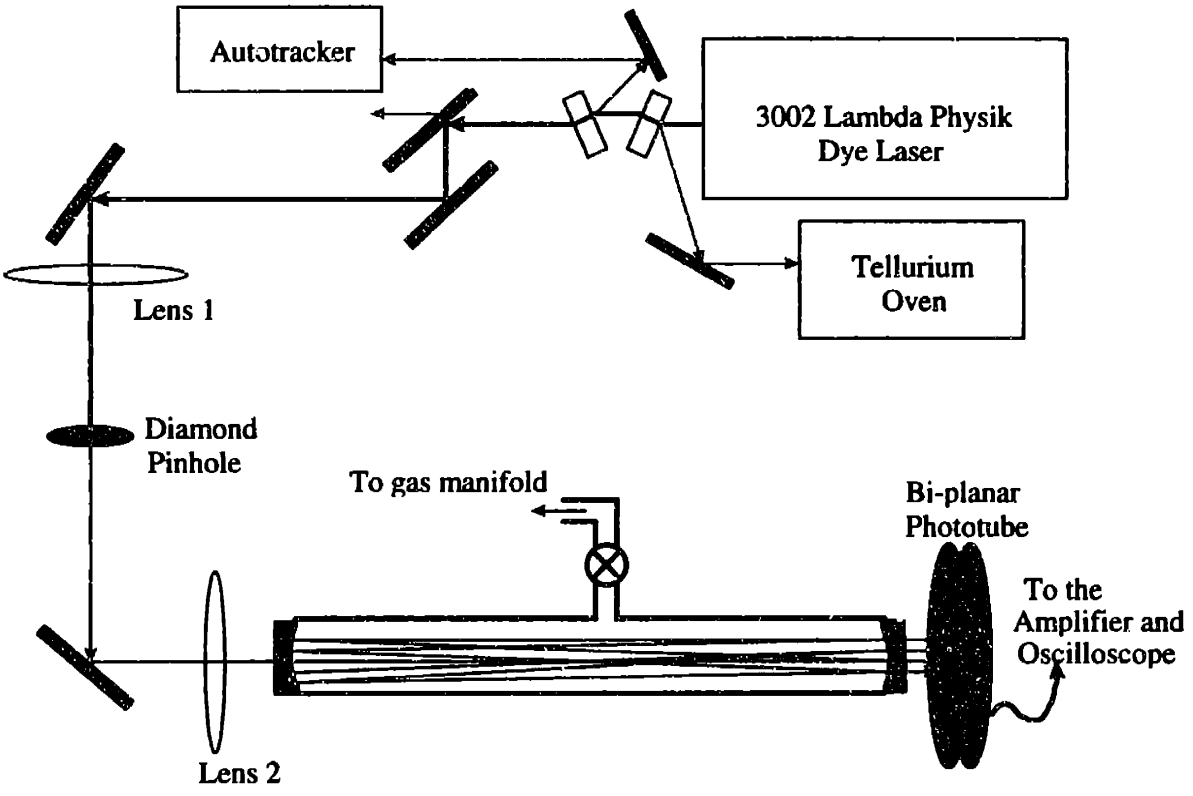


Figure 4.1 Simplified schematics of the CRD experimental setup.

A Neodymium-YAG (NY61, Continuum) pumped pulsed dye laser (Model 3002, Lambda Physik) provides the source of tunable radiation in the visible- near UV spectral region. Immediately after the dye laser light frequency is doubled by a β-BBO crystal (Castech, Inc.). The resulting radiation, tunable within a range of ± 10 nm around a central wavelength of approximately 240nm, is separated from the fundamental frequency

by a pair of dichroic mirrors (CVI, Inc.). The light is, then steered into the ringdown cavity by several high reflectivity, high damage threshold mirrors (CVI, Inc.) and fused silica prisms. The ringdown decay is detected by a fast bi-planar phototube (Model R1193U, Hamamatsu). The output of the phototube is either coupled directly to a port (50Ω DC coupled) of the digitizing oscilloscope, or first pre-amplified by a low noise broadband amplifier (ZFL-1000LN, ~ 20dB gain, 2.9dB NF, Mini-Circuits).

On its way to the ringdown cavity, light is conditioned by the optical filtering system depicted on the Figure 4.2. The main purpose of this system is to clean the spatial mode of the laser beam entering the cavity and to couple the light into the cavity most effectively. All the dimensions of the optical system for optimal mode matching, done by imaging the pinhole into the cavity, have been numerically calculated using programs given in the Appendix 1 (written in MathCad, courtesy of Dr. A.A. Kachanov). The distance between the first lens and pinhole is 300mm, between the pinhole and second lens is 360mm, and between the second lens and the entrance mirror of the cavity is 50mm. The pinhole's diameter is 100μm, the focal lengths of the first and second lenses are 300mm and 200mm, respectively.

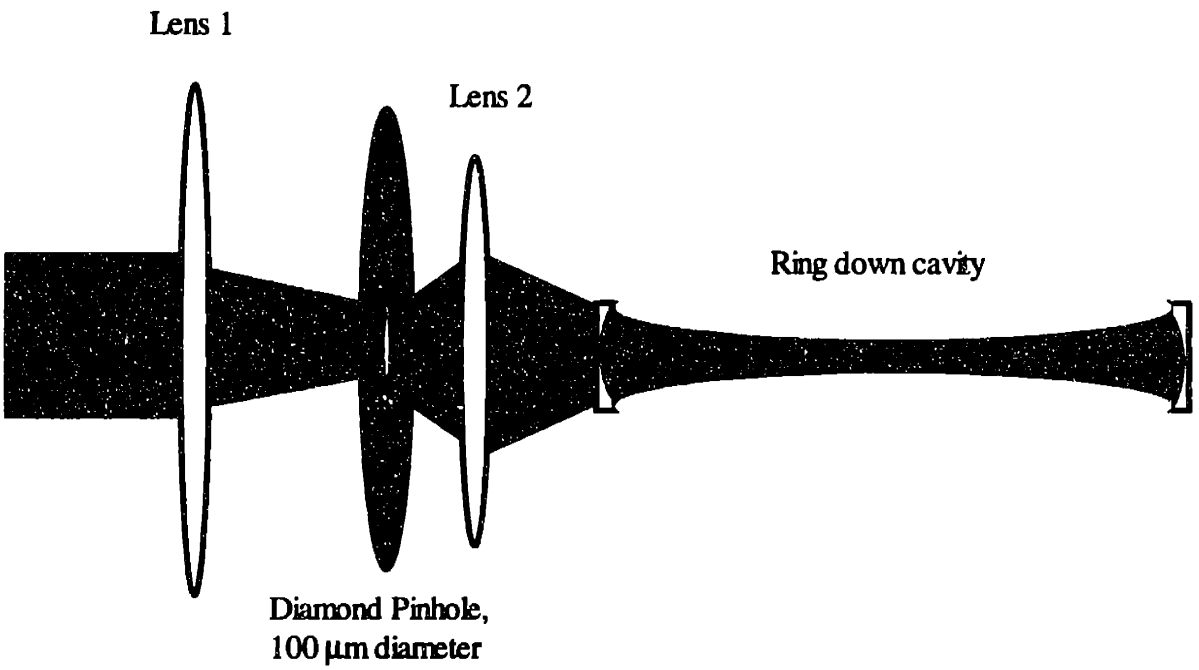


Figure 4.2 Mode coupling of light into ring-down cavity.

Special care was taken to design a cell that would not have any metal parts in contact with the sample, as it was believed that contact with metal would catalyze the isomerization of iso-cyanogen to cyanogen. No windows can be used inside the CRD cell (each glass surface would introduce loss into the cavity!), so the cavity mirrors were attached to the specially designed holders in such a way that they maintain a vacuum seal. Mirror holders are held by optical mounts (New Focus) and attached to the rest of the cell by sliding "Ultratorr"-type connectors made out of Delrin.

The dye laser and oscilloscope are interfaced to a PC computer via GPIB. Data from the oscilloscope, after averaging over 10-500 pulses, is downloaded into the computer and processed, while the laser is scanned to a new wavelength.

Details of data processing are discussed in the following sections.

4.5 Vignetting Effects in Cavity Ringdown Spectroscopy. "Mode Beat Patterns": Their Origin and Effect on the Accuracy of the CRD Experiment.

In a number of previously published papers describing the CRD experiment, the authors attributed the imperfect exponential character of the decay to the so-called "mode beats", referring to the chaotic energy redistribution among the spatial modes of the cavity. These mode beats were considered an unavoidable evil and were not given too much thought. The widespread sentiment was that, to minimize the interference of this effect with the ringdown decay, one should try to couple light into the cavity such that only the TEM_{00} mode is excited, which is, of course, impossible to do with pulsed lasers.

In our very first experiments using CRD, we had a unique opportunity to observe and study the cause and effect of these beats and, as a result, came up with a method allowing us to eliminate them. Our good fortune came from a combination of several factors. The first was the use of relatively low reflectance mirrors, which made the use of a PMT for detection impossible without attenuation. Instead, we chose to employ a bi-planar phototube (BPPT) R1193U-02 (Hamamatsu). The BPPT is used for detection of relatively high fluxes of pulsed light and can be thought of as, essentially, a very fast (300 psec. rise time) PMT without any internal gain. The combination of this detector with a digital oscilloscope capable of 5Gsamples/sec (LeCroy 9360) allowed us to completely time resolve the individual ringdown curve peaks. To be able to separate these ringdown peaks from one another, we are using a cavity with roundtrip time significantly longer than the laser pulse duration. Another very important feature that sets our experimental setup apart from most and that allowed us to notice strange decay patterns was that we have designed the cavity length, L, to be equal to $1.5R$, where R is the radius of curvature of the cavity mirrors. This last condition, if satisfied exactly, will make the pulse, ringing in the cavity, exit the back mirror in three spots[1]. This means that we would expect to see

the ringdown decay break into three groups of peaks, having the same decay constant, but different initial amplitudes.

This is best illustrated by the not-so-complicated case of a confocal cavity (i.e. a cavity which has a length equal to the radius of the mirrors). In a perfectly aligned confocal cavity, the laser pulse, once inside, will be tracing a stable path, depicted on Figure 4.3, and exiting the cavity in two points. These two beams fall onto the detector at two different locations, which, if we remember that no detector has a uniform quantum efficiency (or, in other words, ideal zone characteristics), led us to conclude that in the ringdown decay there will be two groups of alternating peaks, having the same decay constant, but different amplitudes.

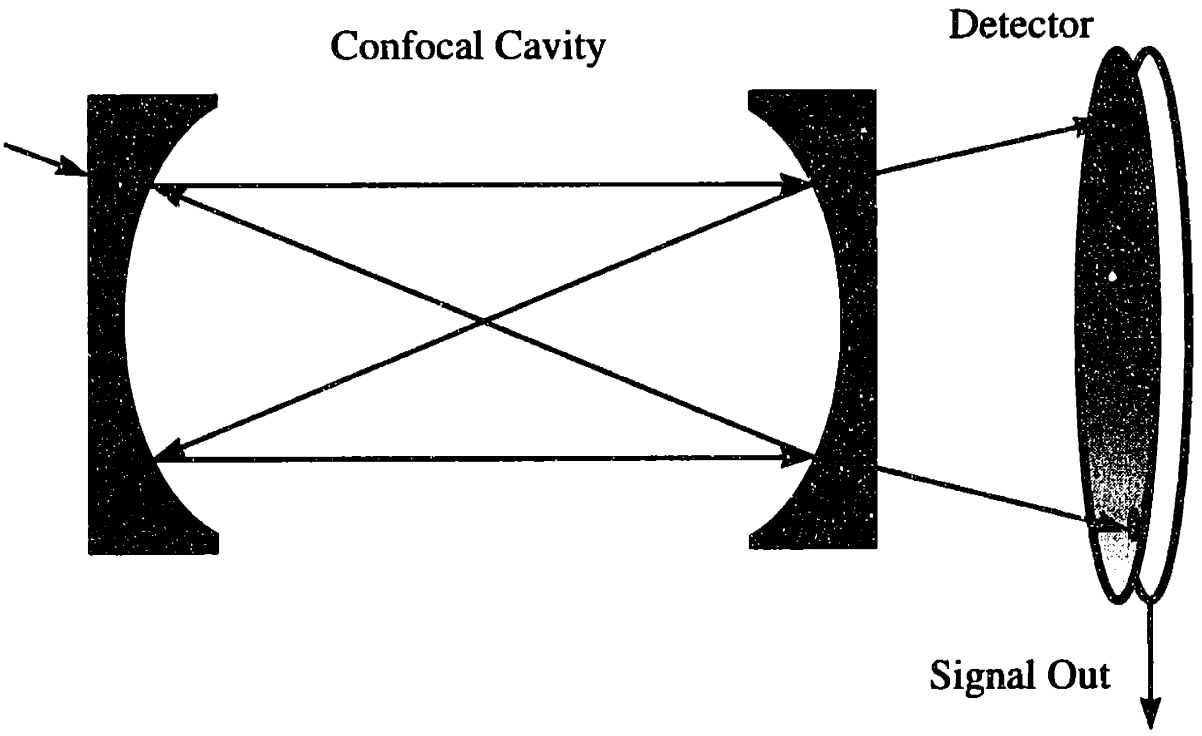


Figure 4.3 Schematic depiction of light propagation in a confocal cavity. Uneven shading of the detector illustrates the non-ideal zone characteristic (i.e. unequal quantum efficiency of different parts of the detector).

Figure 4.4 shows a ringdown decay signal which was recorded for the cavity having mirrors with R=750mm, and cavity length, L=1125mm. We would expect to see

the three groups of peaks mentioned above to have the same decay constants. Instead, we notice that the behavior of each of the groups cannot be described by a single exponential function. Additionally, all three groups behave differently from each other. In fact, as the cavity was slightly tuned, curves were obtained in which one or two of the groups of peaks even exhibited temporary growth in intensity, rather than decay! Surprisingly, the reason for such behavior does not have anything to do with cavity modes. We found that we can introduce variations, analogous to those caused by cavity tuning, by simply moving the photodetector! The explanation is quite simple and lies in the fact that if the cavity length is not exactly matched to one and a half times the radius of curvature of the mirrors, the spots at which the beam exits will be shifting from one full pass to another, producing Lissajous figures on the mirrors and the detector. The uneven quantum efficiency of the photodetector is then convoluted with the exponential ringdown decay, resulting in reduced accuracy of the experiment. On the following few pages we will try to give a more qualitative justification to this statement.

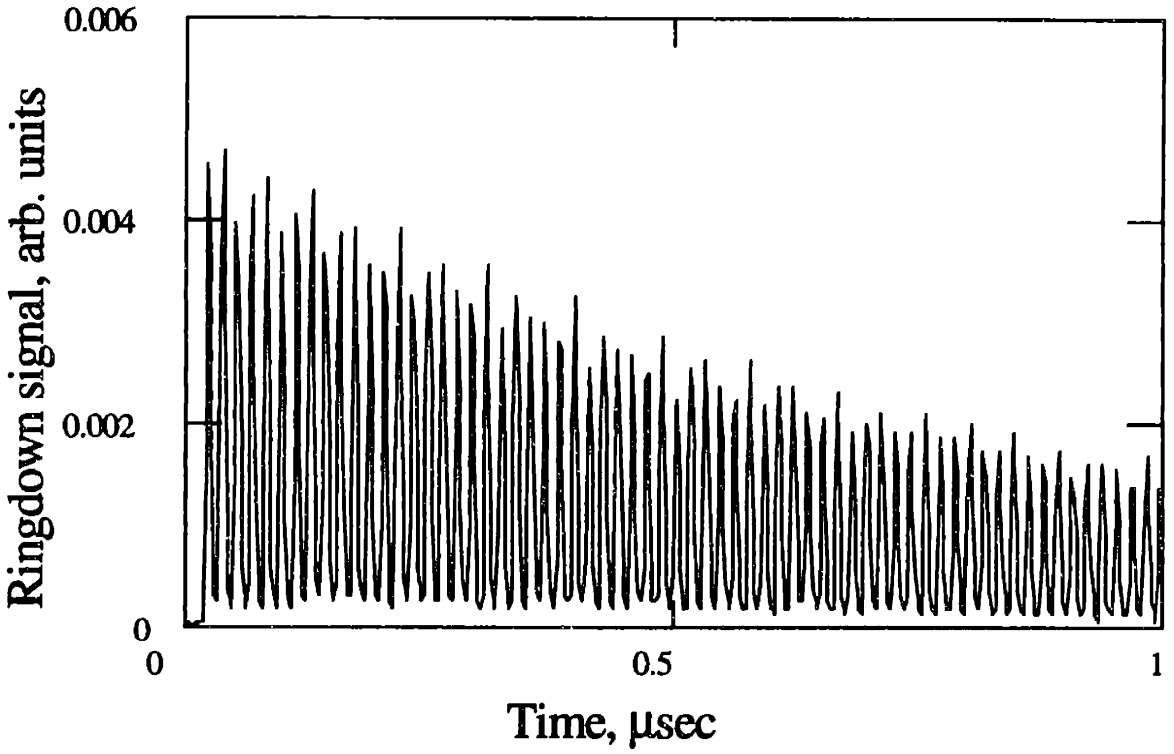


Figure 4.4 Ringdown signal recorded for a cavity with radius of curvature of the mirrors, R=750mm, cavity length, L=1125mm and laser pulse duration 6ns.

Because in our cavity, where the mirrors are separated by 1.5 radii, the pattern of light travel is more difficult to visualize than in the confocal cavity, we will use a simple numerical model to calculate the motion of the beam on the exit mirror for different cavity "detuning" values. In addition, the mirrors in our model will have a predefined loss to simulate CRD decay. The resulting coordinate of the beam exit along with its intensity will then be fed into a detector model used to simulate the actual bi-planar phototube geometry. The theory of ray optics and ray matrices, used in this simulation, is given in a number of easily available texts[1,2,3] and will not be discussed here. The program used for the calculation, written in MathCad, is given in the Appendix 2. The short description of the program follows. In the following discussion, all sizes, if not specified, are in millimeters.

The model cavity is equipped with the mirrors having radii of curvature of R=750mm, and is of the length of $L = 1.5 \cdot (R + \delta)$, where δ is a "detuning factor". The mirror losses are $\gamma = 0.02$. The roundtrip matrix **M** written for the reference plane just before the exit mirror of the cavity is given by:

$$\mathbf{M} = \mathbf{T}(L) \cdot \mathbf{F}\left(\frac{R}{2}\right) \cdot \mathbf{T}(L) \cdot \mathbf{F}\left(\frac{R}{2}\right), \qquad [4.7]$$

where **T** and **F** are ray matrices for propagation through free space and reflection by a curved mirror:

$$\mathbf{T}(\alpha) = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}, \text{ and } \mathbf{F}(\beta) = \begin{pmatrix} 1 & 0 \\ \frac{-1}{\beta} & 1 \end{pmatrix}. \qquad [4.8]$$

Then, the ray coordinates in the cavity reference plane for the n^{th} passage are given by the matrix **R** in which the first column is the coordinate and reduced slope along the abscissa axis, and the second column is the coordinate and the reduced slope along the ordinate axis. Columns of matrix **R** are obtained by acting n-times with the cavity roundtrip matrix **M** on the initial ray position vectors r. The initial values of the position vectors we chose are not of importance, but for the sake of completeness they were

$$r_{0x} = \begin{pmatrix} 1 \\ 0.0025 \end{pmatrix}, \text{ and } r_{0y} = \begin{pmatrix} 1 \\ 0.0001 \end{pmatrix}.$$

The pulse intensity after the n^{th} passage is obtained by iterative multiplication of matrix **R** by coefficient $\chi = (1-\gamma)^2$ per roundtrip.

The ringdown signal exiting the cavity at the coordinates defined above is "recorded" by a photodetector with a square active surface, having $1cm^2$ area, and located 200mm behind the exit mirror. To model a biplanar phototube that has an anode wire grid in front of the photocathode (which introduces a high level of inhomogeneity into the detector's zone function), we introduced a wire passing through the center of the detector (along Y-axis). We can also "shift" the detector relative to the cavity by a distance Δ, to simulate changes in detector positioning between different experimental runs necessitated by procedures requiring removal of the detector (e.g. mirror cleaning). Details of detector function definition are not important and are given in the Appendix 3.

Figures 4.5 -4.7 show the simulation of the evolution of the exiting beam position. It is easy to observe that the farther the cavity length, L, is detuned from 1.5 .times the radius of curvature of the mirrors, the faster the beam "walks" across the mirror (and detector surface.

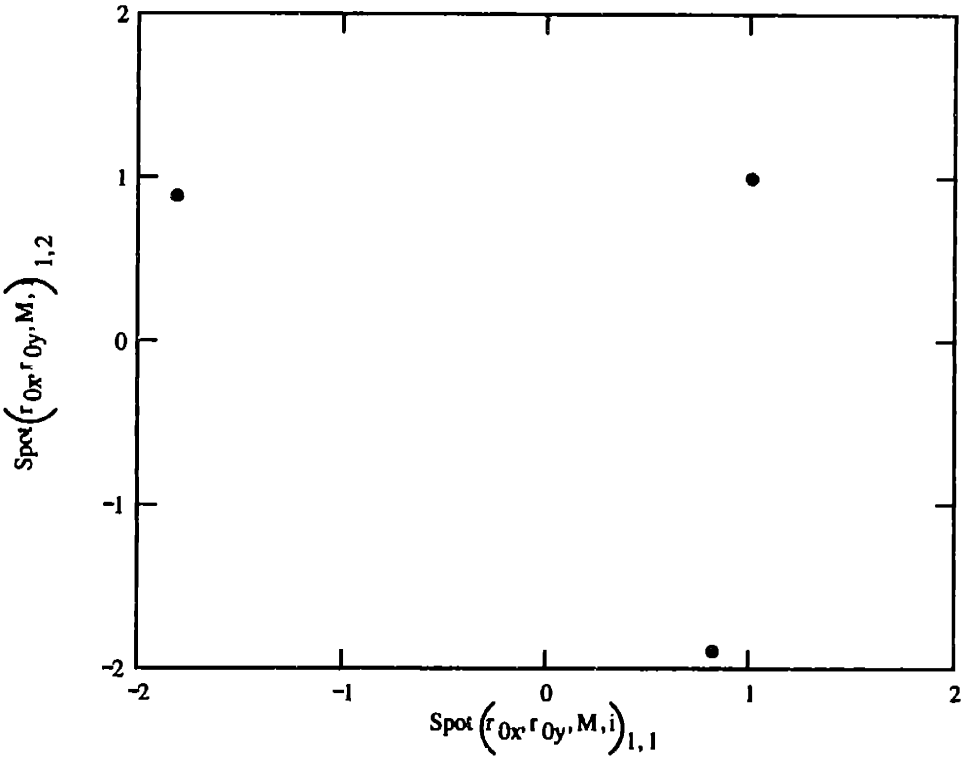


Figure 4.5 Exit spot positions for 80 roundtrips. Cavity is "tuned" to be exactly 1.5R.

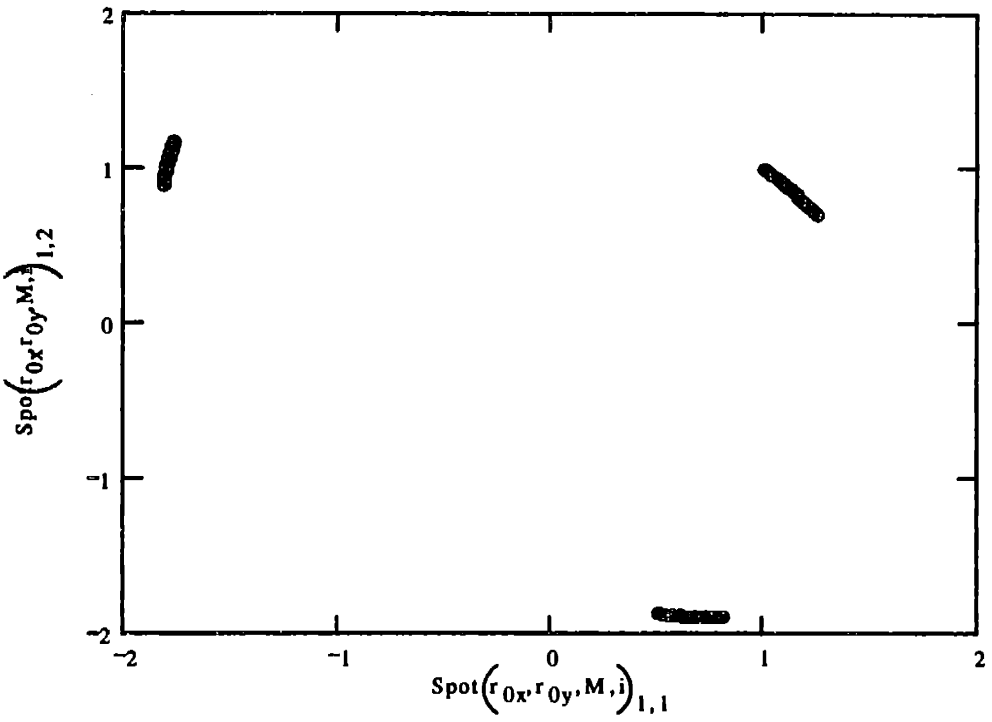


Figure 4.6 Exit spot positions for 80 roundtrips. Cavity length is 1.5(R+0.001). Shift of exit points is easily observable.

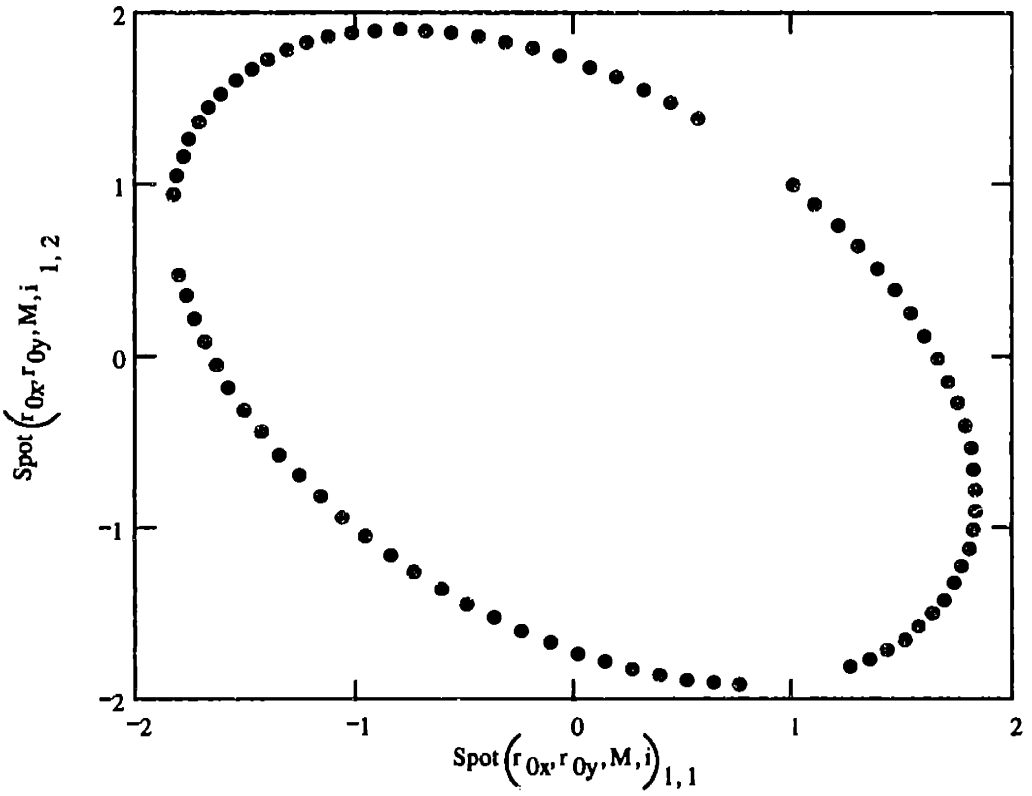


Figure 4.7 Exit spot positions for 80 roundtrips. Cavity length is 1.5(R+0.01). Exit points have moved very far from their original positions.

Figures 4.8-4.13 display the detector signal as a function of time for various detuning factors as well as for several detector positions (relative to the position of the exit mirror of the cavity). Each dot on a plot represents the maximum photocurrent from the detector for each ringdown pulse detected.

In the case of exactly tuned cavity (Figures 4.8-4.10), all three groups of peaks, if treated separately, result in exponential decays having the same decay constant. This is not influenced by the fact that the light exiting in different spots falls on the different areas of the detector, converting to photocurrent with different efficiency. As long as the ray exit spots do not move in time, each group of peaks should follow the exponential decay curve with decay coefficient depending only on the cavity losses.

On the contrary, in the case of a detuned cavity (even as slightly as we have!) all three groups of peaks display very different behavior. One may notice that the "decay" trace originated from a moving exit spot, which, apparently, moved on and then off the "wire" on the detector, not only does not decay exponentially, but even goes up in intensity during some time interval. Needless to say, such a behavior would be capable of drastically decreasing the accuracy of the determination of the sample absorptivity from the fit. Most probably, all the CRD experiments done to date have suffered from this defect to a certain extent. The most probable reason this effect has been unnoticed is that the majority of researchers using the CRD technique did not try to match the cavity length to the mirror curvature, except to make sure that the cavity is a stable one.

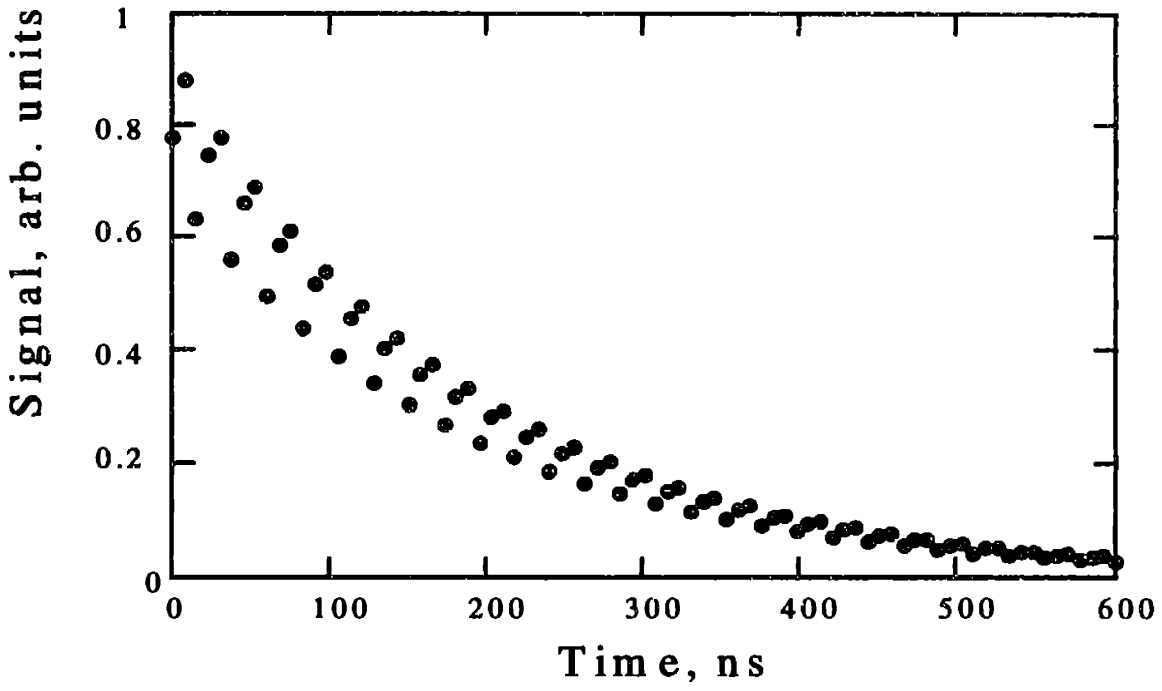


Figure 4.8 Simulated detector signal for cavity length tuned to be exactly 1.5 times the mirror radius, detector displacement, $\Delta = 0$. All the decay curves have the same decay constants.

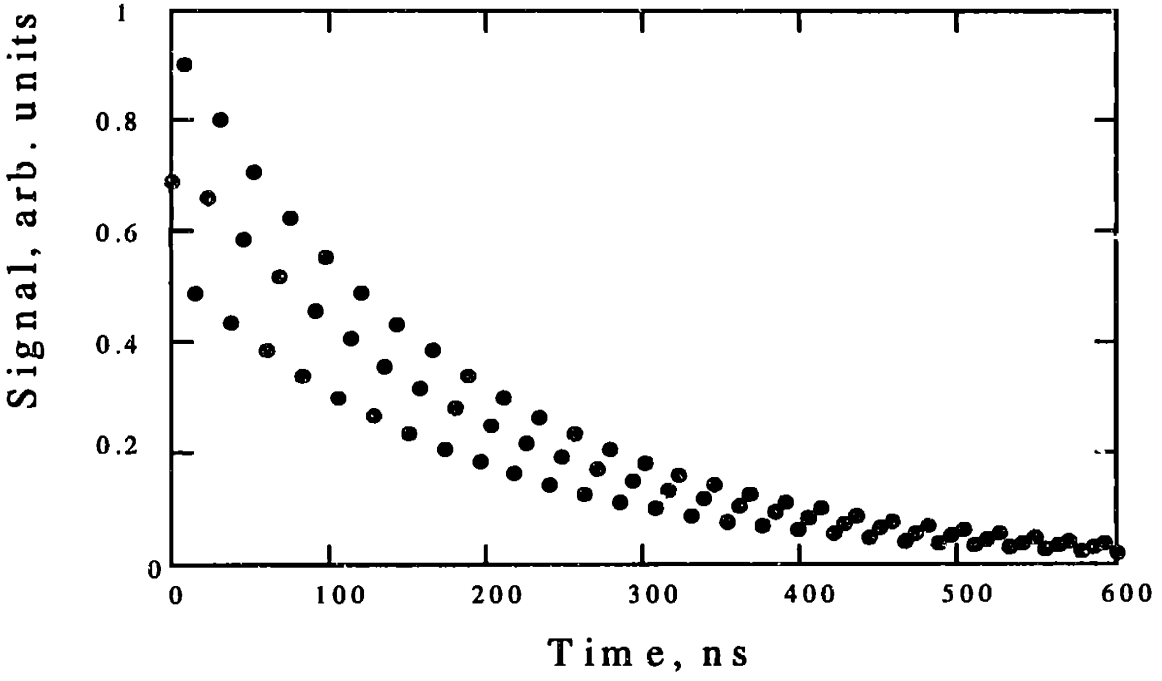


Figure 4.9 Simulated detector signal for cavity length tuned to be exactly 1.5 times the mirror radius, detector displacement, $\Delta = 0.5$. All the decay curves have the same decay constants.

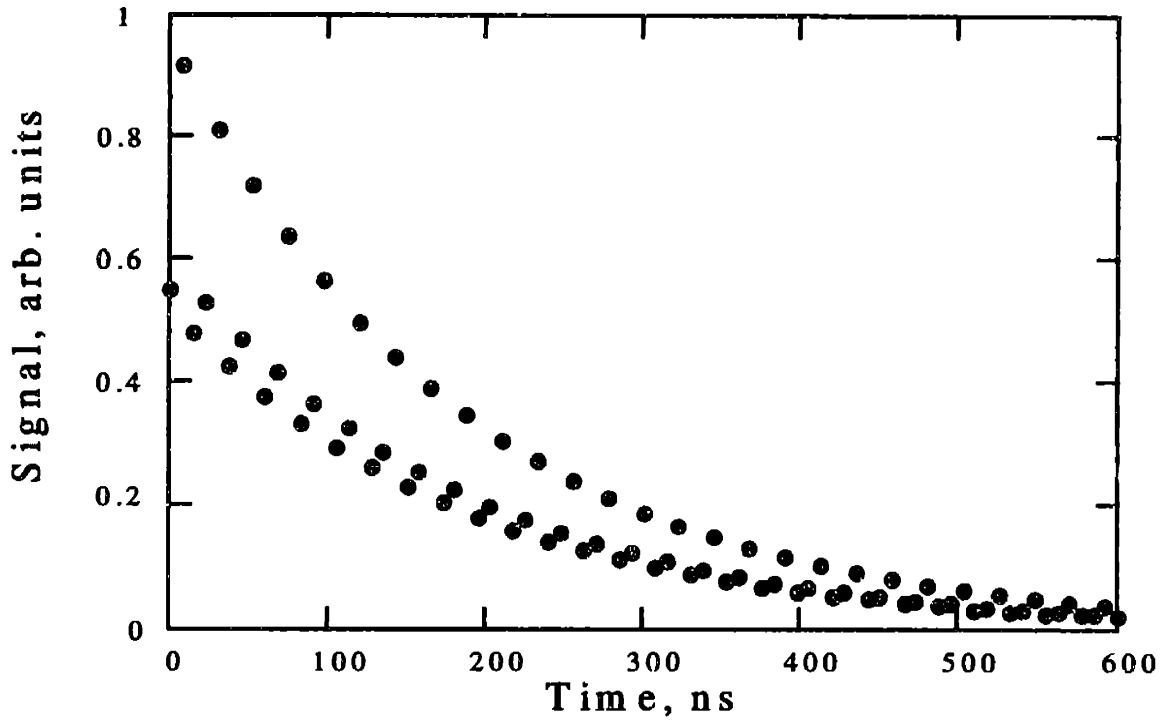


Figure 4.10 Simulated detector signal for cavity length tuned to be exactly 1.5 times the mirror radius, detector displacement, $\Delta = 1.0$. All the decay curves have the same decay constants.

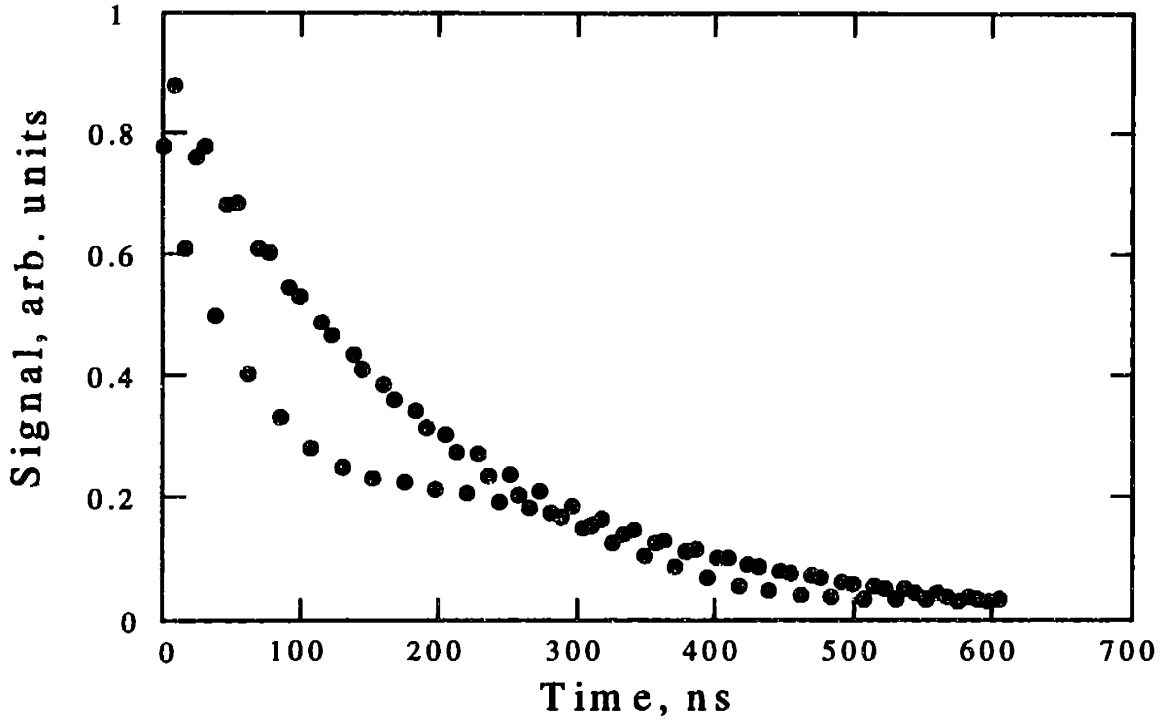


Figure 4.11 Simulated detector signal for cavity length detuned from exact match to 1.5 times the mirror radius. Detuning factor δ=0.01. No detector displacement, $\Delta = 0$.

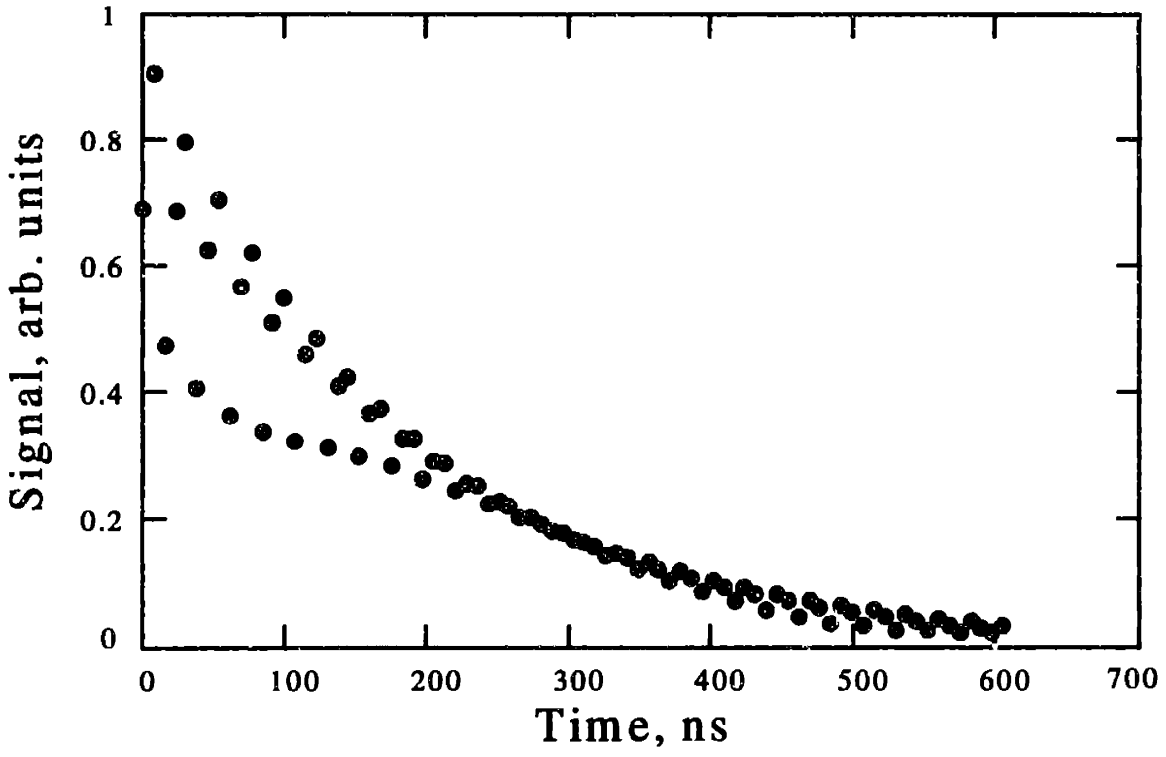


Figure 4.12 Simulated detector signal for cavity length detuned from exact match to 1.5 times the mirror radius. Detuning factor δ=0.01. Detector displacement, $\Delta = 0.5$.

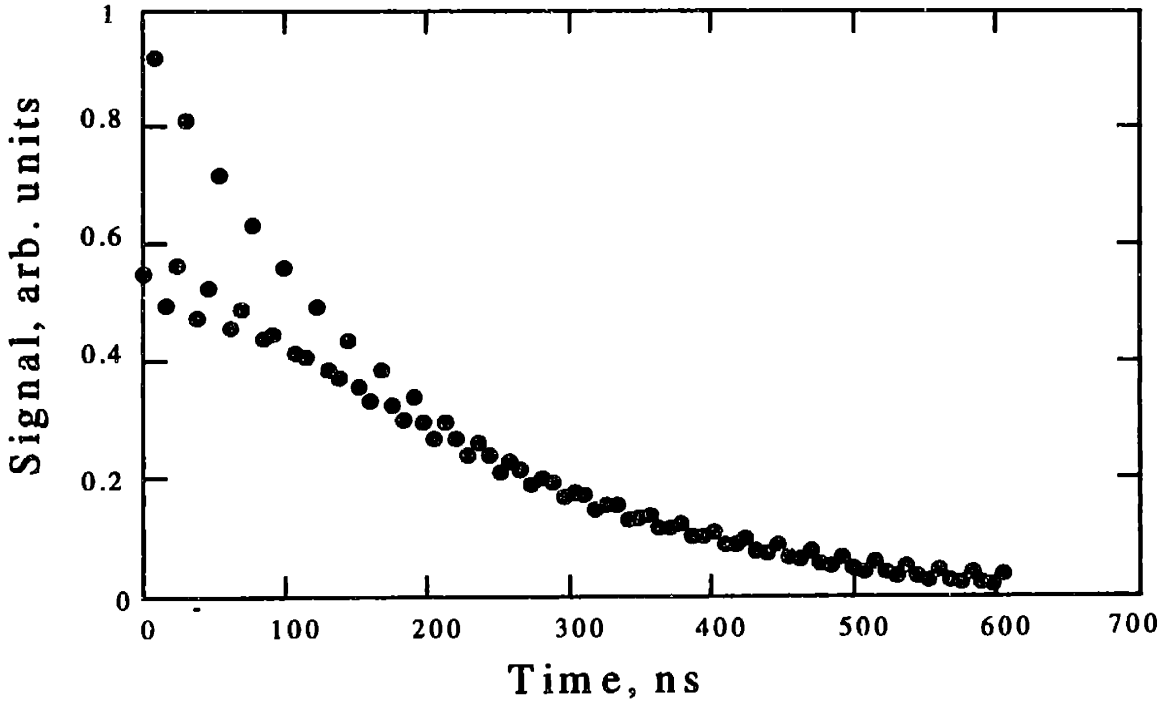


Figure 4.13 Simulated detector signal for cavity lenght detuned from exact match to 1.5 times the mirror radius. Detuning factor δ=0.01. Detector displacement, $\Delta = 1$.

(The only notable exclusion to this has been the work by Cheskis et al. where the authors used a cavity with L=1.5R for easier diagnostics and alignment, but did not notice this effect because of a relatively slow PMT).

Based on the understanding of this source of error, we have modified our experimental procedure. It is clear that the best recipe would be to place the mirrors at a separation of exactly 1.5 times their radius of curvature. The following is the description of how this is done.

One of the cavity mirrors (the front mirror in our case) is affixed to a translation stage driven by a micrometer screw . After the cavity mirrors are placed as close to their desired positions 1.5 (roughly 1.5 times the radius of curvature) and light is coupled into the cavity, we can start moving the front mirror with a micrometer screw. Light is detected by the fast bi-planar photoelement and the photoelectric current (sometimes after amplification) is sent to the fast digitizing oscilloscope. A data acquisition program was written that instructs the oscilloscope to collect a number of CRD decay waveforms, average over them, and to download the average into the computer. The second part of the data acquisition program fits the waveform (details of the data acquisition and fitting will be discussed in section 4.6) to extract the maxima of the ringdown peaks and sort them into the three groups discussed above. For each group, the logarithm of each signal is taken and fitted with a linear function. The slopes of all three fits are displayed and compared. Having this information in "real time" allows us to change the position of the front mirror until the three exponential decays have the same decay constant.

The importance of this "fine tuning" of the cavity length is hard to overestimate. It allows one to dramatically increase the accuracy of the determination of the absorptivity from the decay time. Without fixing this problem, the decay time is a complicated (and unknown!) function of both cavity losses and detector spatial response.

Figures 4.14 *a*, *b* and *c* show an example of the result of such tuning. The plots on these figures represent actual decay curves recorded after the position of the entrance mirror of the cavity has been adjusted following the procedure specified above. Differences in decay constants calculated for the three groups of peaks is 0.02% of the decay constant value. One manifestation of this stability can be observed directly on the

plots. The pattern of the peaks' relative heights repeats unchanging for as long as we could detect. To facilitate easier comparison, figures *b* and *c* are "blow-ups" of two pieces of the waveform at different times.

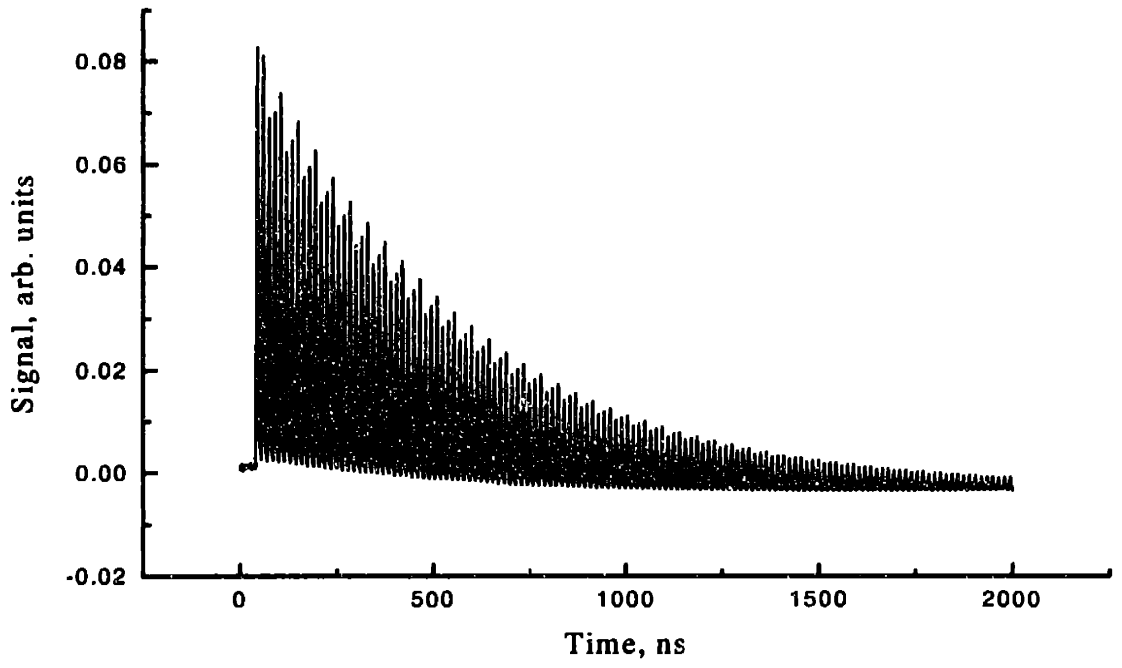


Figure 4.14a Cavity Ringdown decay obtained for a cavity, whose length was very closely matched to 1.5 times the radius of curvature of the mirrors.

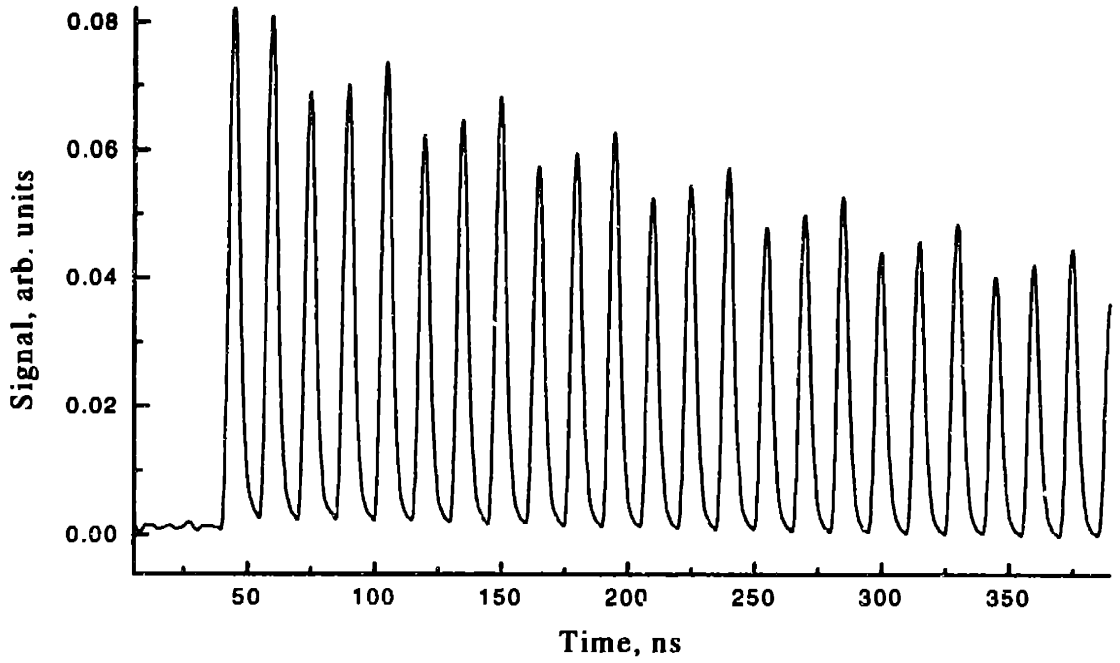


Figure 4.14b Enlargement of the first 350ns of the decay given in the Fig. 4.a12a.

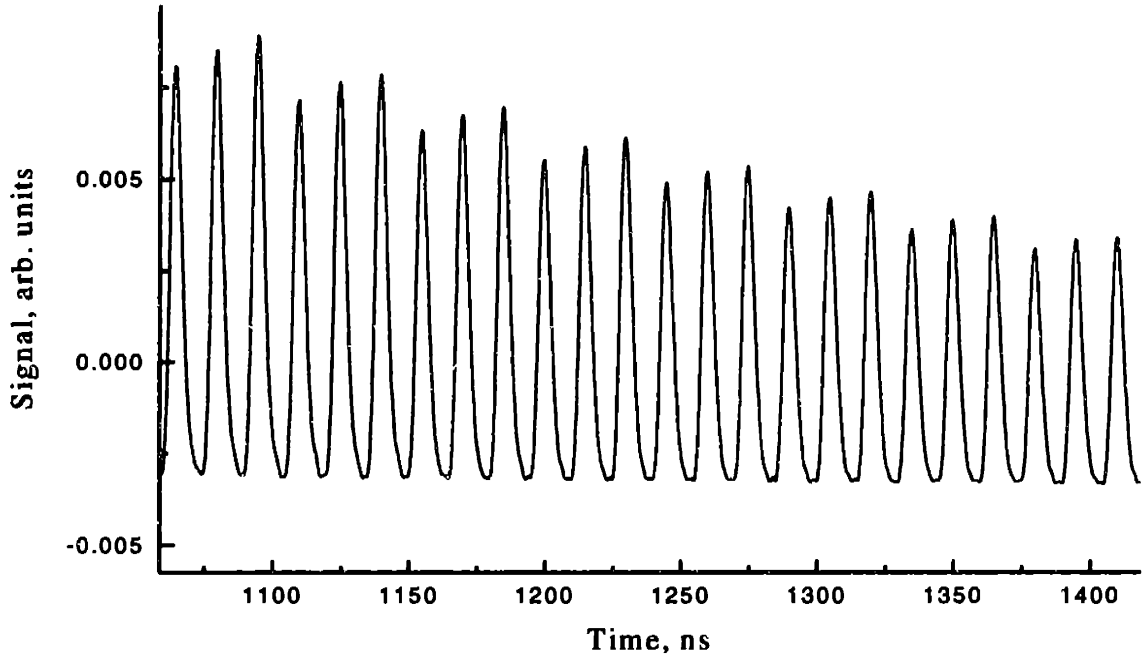


Figure 4.14c Enlargement of the decay curve given in Fig. 4.a12a in the time interval between 1μs and 1.5μs.

An additional improvement can be achieved by tuning the cavity mirrors so that the optical paths of all the beams inside the cavity lie as close to each other as possible. This will ensure that the exit beams will fall on the detector very close to each other, thus reducing possible fluctuations due to the detector's uneven zone characteristics.

Finally, we should notice that $L=1.5R$ is not the only "magic" combination[1]. Our choice to use this particular ratio was caused by reasons of mere convenience.

References

1. Siegman, A.E. "Lasers", University Science Books, 1986.
2. Luneburg, R.K. "Mathematical Theory of Optics", University of California Press, 1964.
3. Yariv, A. *Optical Electronics*, 1991, Saunders College Publishing.

4.6 Sensitivity Analysis of CRD Spectroscopy.

Our next goal is to establish the theoretical limits of the sensitivity of CRD spectroscopy. A very good discussion of this problem is presented in the work by Romanini and Lehmann[1]. The authors established the theoretical sensitivity limits of the CRD apparatus for four different methods of data acquisition. Although they made excellent points providing results for quantum noise limited detection, the authors failed to correctly determine the sensitivity of their own (real) experimental setup, and to address other sources of noise in the detection. The authors also avoided mentioning the fact that comparing the sensitivity of the CRD experiments by simply stating how much worse than the quantum noise limit they perform puts experiments that used lower reflectivity mirrors (or higher light fluxes) in a disadvantaged position. (This can be illustrated by the fact that σ_x/x, where x is a variable whose value obeys Poisson statistics, is equal to $1/\sqrt{x}$, and will become smaller as x grows.)

In this section I will discuss main results derived for shot noise limited detection by Romanini and Lehmann (R&L), point out the inadequacy of their analysis for application to a real experimental setup, define the sensitivity limits of our CRD experiment, and the possible changes and improvements.

Before proceeding with our discussion, I would like to point out that the main goal is to determine the standard deviation of the absorptivity σ_α that can be obtained from the CRD experiment. R&L have, instead obtained the standard deviations for the decay time τ, σ_τ. The two are very simply connected. If the decay of the CRD signal can be described as

$$I = I_0 e^{-t/\tau}, \qquad [4.9]$$

then, from [4.5] :

$$\frac{1}{\tau} = \frac{c}{l}(T + \mathrm{L}) + c\alpha(\omega) \qquad [4.10]$$

The first term describing the quality of the cavity mirrors, can be represented as an effective absorptivity:

$$c\alpha_c = \frac{c(T+\ell)}{l}, \qquad [4.11]$$

while the second term, $c\alpha(\omega)$, reflects the sample absorptivity. Then, the standard deviation of the absorptivity $\alpha_{tot} = \alpha_s + \alpha_c$ is easy to express as

$$\sigma_\alpha = \sqrt{\left(\frac{\partial\alpha}{\partial\tau}\right)^2 \sigma_\tau^2} = \frac{\sigma_\tau}{c\tau^2}, \qquad [4.12]$$

so that coefficient of variation for the absorptivity, expressed though the coefficient of variation for the decay time is (compare with essentially same result in [4.6])

$$\frac{\sigma_\alpha}{\alpha} = \frac{\sigma_\tau}{\tau}. \qquad [4.13]$$

In their paper, R&L assumed that the only source of noise in the electrical signal, which is recorded and processed to extract absorptivity of the sample, was quantum noise. Validity of this assumption will be addressed later in the section. For now we will just accept it.

The first method used to measure absorptivity that we will discuss utilizes a two channel box-car averager with internal analog signal processing. For simplicity let us suppose that the detector does not resolve the individual ringdown pulses (which is true of nearly every previously reported experiment but ours). In that case the ringdown signal is expressed as

$$I = I_0 e^{-t/\tau}, \qquad [4.14]$$

where I_0 is the initial intensity, and τ is the ringdown time. Amplified output of the photodetector is input into a two channel boxcar averager, so that the output of each channel is an integral of the ringdown signal within the same integration interval ΔT, with the origins at times T_1 and T_2 separated in time as $T_2=T_1+\delta T$.

The analytical expression for the integral value is:

$$\int_T^{T+\Delta T} I_0 e^{-t/\tau} dt = I_0 e^{-T/\tau}\left(1 - e^{-\frac{\Delta T}{\tau}}\right). \qquad [4.15]$$

Therefore, for the outputs of each of the two channels, A and B, we can write

$$A = I_0 \cdot \tau \cdot e^{-\frac{T_1}{\tau}} \cdot \left(1 - e^{-\frac{\Delta T}{\tau}}\right), \text{ and} \qquad [4.16]$$

$$B = I_0 \cdot \tau \cdot e^{-\frac{T_2}{\tau}} \cdot \left(1 - e^{-\frac{\Delta T}{\tau}}\right). \qquad [4.17]$$

Their ratio is given by:

$$\frac{A}{B} = \exp\left(\frac{T_2 - T_1}{\tau}\right) = e^{\frac{\delta T}{\tau}}. \qquad [4.18]$$

Then the absorption spectrum can be found (after appropriate subtraction of baseline due to the losses of the cavity itself) from the expression

$$\alpha = \frac{1}{c\tau} = \frac{1}{\delta T} \ln\left(\frac{A}{B}\right). \qquad [4.19]$$

To obtain the coefficient of variation (sometimes called fractional noise), one has to express the variance of absorptivity via those of the variables of the equation[4.20] using the generalized rule for propagation of errors[2]:

$$\sigma_\alpha^2 = \left(\frac{\partial \alpha}{\partial(\delta T)}\right)^2 \sigma_{\delta T}^2 + \left(\frac{\partial \alpha}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial \alpha}{\partial B}\right)^2 \sigma_B^2, \qquad [4.20]$$

since we assume that all of the variables of the expression [4.19] are statistically independent. Variances of A and B should, in turn, be re-expressed via the variables of formulas [4.16] and [4.17] as:

$$\sigma_A^2 = \int_{T_1}^{T_1+\Delta T} \left[\left(\frac{\partial A}{\partial I(t)}\right)^2 \sigma_{I(t)}^2 + \left(\frac{\partial A}{\partial T_1}\right)^2 \sigma_{T_1}^2 + \left(\frac{\partial A}{\partial(\Delta T)}\right)^2 \sigma_{\Delta T}^2\right] dt, \qquad [4.21a]$$

σ_B^2 – analogously for channel B. [4.21b]

In their derivation, R&L assumed that the errors in δT and in ΔT can be neglected (which means that the errors in the determinations of T_1 and T_2 are also neglected), then the measurement errors are determined by the error in the measurement of the sum of the signal that has been integrated in one channel of the boxcar. The next important assumption the authors made was that the signal they integrate has a quantum noise limited error and that the box-car integration and mathematics do not introduce any additional error in the expression of $\sigma_{I(t)}$.

The resulting equation for the coefficient of variation of the absorptivity is

$$\frac{\sigma_\alpha}{\alpha} = \frac{1}{\delta T\sqrt{I_0\tau}} \frac{\tau\sqrt{1+e^{\frac{\delta T}{\tau}}}}{\sqrt{1-e^{-\frac{\Delta T}{\tau}}}}. \qquad [4.22]$$

One can find the optimal value for the delay between the two integration gates, δT. The minimum of coefficient of variation occurs for δT between 2τ and 2.5τ, more precisely at $\delta T_{min} = 2.22\tau$. For this delay, the coefficient of variation is

$$\frac{\sigma_\alpha}{\alpha} = \frac{1.44}{\sqrt{I_0\tau}} \frac{1}{\sqrt{1-e^{-\frac{\Delta T}{\tau}}}} \geq \frac{1.44}{\sqrt{I_0\tau}}, \qquad [4.23]$$

for $\Delta T \geq \tau$.

The second method, discussed by R&L in their paper, has previously been used to find absorptivities in the CRD experiments (essentially the method used in the first CRD paper by O'Keefe et al.[3]). In it, the decay signal is digitized in a time interval [0,T], linearized by taking the logarithm, and fit to a straight line using a least squares fit (see [4.4], [4.5]) . The slope of the line is directly proportional to the absorptivity $a = -c\alpha$, where a is the slope of the fitting function $f(t) = at + b$. It is easy to see that the coefficient of variation for the absorptivity is the same as for the slope of the fitting function, a.

The linearized signal is

$$S(t) = \ln I(t) = \ln I_0 - c\alpha t. \qquad [4.24]$$

The least squares fit used in all previous works involves minimizing the expression

$$\chi^2 = \int_0^T \left(S(t) - f(t)\right)^2 dt. \qquad [4.25]$$

with respect to a and b. Following R&L we can obtain the value of a:

$$a = \frac{12}{T^3}\int_0^T S(t)tdt - \frac{6}{T^2}\int_0^T S(t)dt, \qquad [4.26]$$

from which we can write the propagation of errors expression:

$$\sigma_a^2 = \int_0^T \left(\frac{\partial a}{\partial S(t)} \right)^2 \sigma_{S(t)}^2 dt, \qquad [4.27]$$

where

$$\frac{\partial a}{\partial S(t)} = \frac{12}{T^3} t - \frac{6}{T^2}, \text{ and} \qquad [4.28]$$

$$\sigma_S^2 = \left(\frac{\partial S(t)}{\partial I(t)} \right)^2 \cdot \sigma_I^2(t) = \frac{1}{I(t)}, \qquad [4.29]$$

where the last equality is valid in the case of quantum noise limited detection.

From formulas [4.27]-[4.29] it is easy to obtain the variance for the slope of the fitting function[1]:

$$\sigma_a^2 = \frac{36}{I_0 \tau^3} \frac{(x^2 - 4x + 8)e^x - x^2 - 4x - 8}{x^6}, \text{ where } x = \frac{T}{\tau}. \qquad [4.30]$$

When this function is minimized with respect to T to provide the smallest possible coefficient of variance, the best achievable value is found to be

$$\frac{\sigma_\alpha}{\alpha} = \frac{\sigma_a}{a} = \frac{1.84}{\sqrt{I_0 \tau}}, \qquad [4.31]$$

which seems to be slightly worse than that obtained for the boxcar case (remember, in both cases the assumption was that quantum noise the only noise source of the signal!). The reason why choosing T very large is not going to improve detection sensitivity is that an unweighted least squares fit is used. The fact that the quantum noise on the logarithm of the ringdown signal increases exponentially with time explains why, if the values of ring down signal for all times have the same weighting, the fit becomes noisier as time T increases beyond T_{min}.

Of course, it is also possible to find the absorptivity value by integrating over the digitized signal. The most significant drawback in this approach is that the error in the determination of the maximum of a pulse is much smaller than that of the integral, because integration involves summation over all signal values in a certain time interval, and the smaller the value of the signal itself, the larger the coefficient of variation. So,

even if the integration does not introduce any error itself, it degrades the coefficient of variation of the result.

The obvious improvement to the second case then would be (R&L's third case) to use a weighted least squares fit (the sum is replaced by an integral for convenience):

$$X^2 = \int_0^T \left[\frac{S(t) - f(t)}{\sigma_i} \right]^2 dt, \qquad [4.32]$$

where X^2 must be minimized with respect to a and b, and $\sigma(t)$ is the standard deviation of the measurement. The result obtained by R&L, assuming that quantum noise is the only source of noise in the system, in the limit of $T \to \infty$ is

$$\frac{\sigma_\alpha}{\alpha} = \frac{1}{\sqrt{I_0 \tau}}, \qquad [4.33]$$

which, probably represents the best fractional noise (equal to the quantum noise!) obtainable in any signal processing scheme.

The final case considered in the paper[1] was a hypothetical signal processing scheme employing analog electronics (time to amplitude converter, integrators, ramp generators and multipliers) plus a box-car signal averager. The idea was to calculate

$$\alpha = \frac{1}{c\tau} = \frac{\int_0^\infty I(t)dt}{c \cdot \int_0^\infty t \cdot I(t)dt}. \qquad [4.34]$$

This scheme will have the same noise limit as the weighted least squares fitting method. On the other hand, it has the disadvantage of precluding any test for the purity of the exponential character of the decay.

All previous discussion focussed on determining and comparing the sensitivities of various methods of signal processing, assuming that the predominant noise source in the system is quantum noise. Even more so, R&L have made a claim about the sensitivity of their experiment being shot noise limited based on the fact that the lowest noise level

they obtained has been recorded when the delay between two gates of the box-car averager was about two times the exponential decay time (a claim that is not very well supported by the plot of the decay time versus delay between gates to which they refer).

One of the main problems of their sensitivity analysis becomes apparent when one notices that the authors did not include, or analyze in any way, the errors introduced by various pieces of detection electronics through which electric signal passes to the "processor". Neither they include the analysis of the errors introduced by the box-car integrator and processing functions they used to extract decay time from the ringdown signal. In the following discussion I will try to address these questions. This will help us to define and estimate the sensitivity limits of the optimal experimental arrangement we can use.

Photodetector as a source of noise.

All the photodetectors featuring internal gain have (in addition to the shot noise due to the photon flux, thermal noise and noise caused by dark current) noise caused by the amplification process itself. The detector used by R&L, as well as by most of the researchers using CRD is a photomultiplier tube (PMT). In this detector, photons eject electrons from the photocathode (which is held at a large negative voltage relative to the anode). The electrons ejected by the cathode are accelerated and focused by electric fields from additional electrodes until they strike the first dynode, which is made of a material that emits a number of electrons when struck by a single high energy electron. These secondary electrons are accelerated toward the second dynode, which in turn produces more electrons, with the process repeating down the chain of dynodes. The PMT output signal can be described in terms of number of protons striking the cathode, n_0, in time t, quantum efficiency of the cathode η, and amplification by a dynode, d. The number of the electrons at the entrance of the first dynode is

$$n_1 = \eta n_0 , \qquad [4.35]$$

with the coefficient of variance given by

$$\frac{\sigma_1}{n_1} = \frac{1}{\sqrt{\eta n_0}} . \qquad [4.36]$$

This is the quantum noise of the primary photoelectrons which was used by R&L in their paper as a dominant source of detection noise. On the other hand, it is known that the signal-to-noise ratio is reduced by the amplification process in the PMT[4]. If the first dynode has amplification d_1, the electron current leaving it is

$$n_2 = \eta d_1 n_0 = d_1 n_1. \qquad [4.37]$$

Assuming that the electron multiplication obeys Poisson statistics, we can express the noise of the current leaving the first dynode as

$$\sigma_2 = \sqrt{d_1^2 n_1 + n_2} = \sqrt{d_1^2 \eta n_0 + d_1 \eta n_0}. \qquad [4.38]$$

Then, the coefficient of variance for the current emerging from the first dynode is given by

$$\frac{\sigma_2}{n_2} = \sqrt{\frac{d_1 + 1}{d_1}} \frac{1}{\sqrt{\eta n_0}}. \qquad [4.39]$$

Repeating this calculation for a chain of dynodes we obtain the following expression for coefficient of variance

$$\frac{\sigma_{out}}{n_{out}} = \sqrt{1 + \frac{1}{d} + \frac{1}{d^2} + \ldots} \cdot \frac{1}{\sqrt{\eta n_0}} \approx \frac{1}{\sqrt{\eta n_0}} \cdot \sqrt{\frac{d}{d-1}}, \qquad [4.40]$$

where, for simplification, all the dynodes are assumed to have the same amplification, d. A "standard" PMT has ten dynodes and amplification on the order of 10^5 times. The net signal after the amplification by m dynodes is

$$n_{out} = d^m \eta n_0, \qquad [4.41]$$

so for our hypothetical PMT we can write $d^{10} = 10^5$, or $d \approx 3$. This factor alone increases the coefficient of variance of PMT relative to that of the photoelectron shot noise by a factor of 1.22 (or by 22%).

Stability and linearity of the amplification are two additional factors of prime importance in CRD applications. The voltages required by the PMT are typically supplied by a divider circuit. The dynode amplification has an almost linear dependence on the applied voltage. Assuming total gain of d^m we can derive a dependence between the stability of the supply voltage and phototube gain[4]:

$$\left(\frac{d+\Delta d}{d}\right)^m = \frac{G+\Delta G}{G} = \left(\frac{V+\Delta V}{V}\right)^m, \qquad [4.42]$$

where G is a total gain of the PMT and V is supply voltage. Taking the logarithm of the expression and simplifying via $\ln(1+x) \approx x$, for small x, we obtain:

$$\frac{\Delta G}{G} = m\frac{\Delta V}{V}. \qquad [4.43]$$

which means that to maintain gain stability of 1% requires voltage stability of 0.1% for a typical PMT. It is especially important because in CRD, the PMT is used in current mode, so that gain stability directly affects the accuracy of the measurement.

Linearity of amplification is a necessity in CRD experiments. Among other factors, it depends on the design of the voltage divider chain and will not be discussed here due to our lack of knowledge of the particular design used by the authors. On the other hand, it should be noted that care must be taken to insure that the signal current flowing in the phototube does not alter the amplification of the dynode stages.

However, maybe the most significant input to the error of determination of the absorptivity comes from the uneven zone characteristics of the detector. Due to the lack of reported experimental details, it is very hard to quantitatively estimate its impact in the experiments reported by R&L. On the other hand, from our discussion in section 4.000, it is quite clear that this error can be quite large.

We will be neglecting errors caused by the dark current of the photodetector and those of the thermal noise, assuming their relative insignificance.

Amplifier as a source of noise.

The amplifier Model 502 (Tektronix, Inc.) used by R&L in their work has the following characteristics. If the PMT is coupled into 50Ω, the input noise is specified at 10 μV rms., while gain stability and linearity have uncertainty of $\pm 0.5\%$. The authors reported that the peak signal on the output of PMTwas on the order of 50 mV. This limits the coefficient of variance due to input noise of the amplifier to 0.02%, which can be neglected compared to the amplification noise. On the other hand, the amplification

noise standard deviation is comparable to that stated by the authors for the shot noise of their experimental conditions, and must be included in a signal-to-noise analysis.

Box-car averager as a source of noise.

An ideally functioning box-car averager would have a very low input noise relative to a several volt input signal, stated by R&L. On the other hand, the error of the signal processing can be very significant. The manual for the Model 162/165 box-car averager (Princeton Applied Research, Inc.) specifies that for $\log\left(\frac{A}{B}\right)$ function, used in the box-car based acquisition scheme, the error in the output voltage will range from 0.25%, for larger signal input in the beginning of the ringdown decay, to above 1%, in the tail of the decay, when the signal voltage level is low. Needless to say, these constitute a very significant input into the total noise of the experiment.

The sum of the variances of all the noise components will give the variance that has to be used in determination of the smallest detectable absorption. It is quite clear that in the box-car integrator approach taken by Romanini and Lehmann, shot noise will not the limiting factor in the sensitivity of the detection, even though, due to very high reflectivity of the mirrors used and, therefore, to low photon fluxes on the detector, the coefficient of variance of the shot noise is relatively high.

On the contrary, in our experiment we have to work relatively far in the ultraviolet, where typical mirrors reflectivities are not higher than $R \leq 0.997$, and photon fluxes on the detector are much higher, allowing for lower values of shot noise coefficient of variance. In these conditions it is much more difficult to reach the shot noise limit, but it is still possible to have very high sensitivity.

The last note is on the validity of comparison of the experimental noise averaged over N measurements with the shot noise of a single measurement. It is quite clear that averaging, although improving the signal-to-noise ratio of the experiment, cannot change the relationship between the signal to noise ratio achieved in a single measurement and that of a single measurement shot noise limited experiment. Obviously, just as the signal-

to noise of the experiment improves proportionally to the square root of the number of measurements taken, the shot noise does too.

References

1. Romanini, D., Lehmann, K.K. *J. Chem. Phys.*, **99**,6287 (1993).
2. Shoemaker, D.P., Garland, C.W., Nibler, J.W., "Experiments in Physical Chemistry", 5th ed., McGraw-Hill, 1989.
3. O'Keefe, A., Deacon, D.A.G., *Rev. Sci. Instr.* **59**, 2544 (1988).
4. Rieke, G.H. "Detection of Light: From the Ultraviolet to the Submillimeter.", reprint of the 1st ed., Cambridge University Press, 1996.

4.7 Optimization of the Experimental Procedure and Signal Processing in the CRD Spectroscopy of the First Excited Singlet State of Isocyanogen.

This section serves to summarize findings presented in previous section and to introduce some additional ideas on the improvement of sensitivity and accuracy of the CRD experiment. Sensitivity limits of our experiment will be estimated.

Two first requirements to our experimental setup are imposed by the necessity to resolve ringdown peaks in time. This is caused by the fact that the baseline of the ringdown signal is often distorted by the electro-magnetic pick-up in the detection electronics and by the amplification process (we have to remember that even when the external amplifier is not used, the signal processor has its own amplifier). This problem can be circumvented by subtraction of the of the baseline from the signal. In our setup, having the laser pulse duration of only six nanoseconds and cavity roundtrip time of fifteen nanoseconds we can record both signal (peaks) and baseline (space between peaks) in one waveform, providing our detector, amplifier (if used) and signal processor (oscilloscope) are fast enough to resolve it. Bi-planar phototube (R1193U-02 , Hamamatsu) and LeCroy digitizing oscilloscopes we use (Model 9360, 5Gs/sec and Model 3430A, up to 2Gs/sec) are sufficiently fast to satisfy our requirements. Cheap, low noise, high bandwidth amplifiers are available from a number of manufacturers (e.g. Mini Circuit, Inc.).

Once the requirements to the oscilloscopes are mentioned, it would be advantageous to be able to collect and digitize very long ringdown decays with high time resolution. Our new oscilloscope (Model 3430A, LeCroy) is capable of collecting 200K samples at digitizing rate of 1Gs/s.

To improve signal-to-noise ratio averaging over a hundred pulses in computed in the oscilloscope in real time and then sent to the computer. There, the averaged waveform is fitted with the special interpolation routine to calculate exact maxima and

minima for each ringdown peak and baseline spaces in between, respectively. The routine is based on the Savitzky -Golay algorithm[1]. Baseline for the ringdown waveform is assumed to go through the such calculate minima. The peak height for each peak is then calculated by subtraction of the baseline value from the corresponding peak maximum value. The resulting decay curve consisting of peak hight values is sorted into three "sub-curves" (recall discussion in section 4.4), and logarithm of each of them is fitted, using weighted least squares fit (recall discussion in section 4.5), to a straight line.

It is very important to use the values if true maxima of ringdown peaks for such a fit. If, instead, a ringdown curve is detected with lower time resolution (or smoothed by the use of low pass filters), such that there is still a "wiggle" superimposed onto exponential decay, and the whole waveform is used in the least squares fit to obtain the decay time of the ringdown, the result of such a signal processing will have little in common with the true value of the ringdown decay time.

Another interesting method to improve signal to noise ratio when very long waveforms are recorded is to record the ringdown on two channels, having different sensitivities. Due to the decaying exponential character of the ringdown curve, the end of the curve will always have much smaller signal level than the beginning. On the other hand, it is accepted that the error which oscilloscope introduces into the measurement is equal to one bit, or $1/256$ of the full scale at current sensitivity setting. It is easy to see that as signal goes down, the coefficient of variance goes up. Therefore, recording the tail of the decay on a channel with higher sensitivity improves signal to noise of the experiment. Figure 4.15 *a* and *b* show the waveforms of the same signal recorded in such a way. Baseline fluctuation is due to the amplifier used in that experimental run.

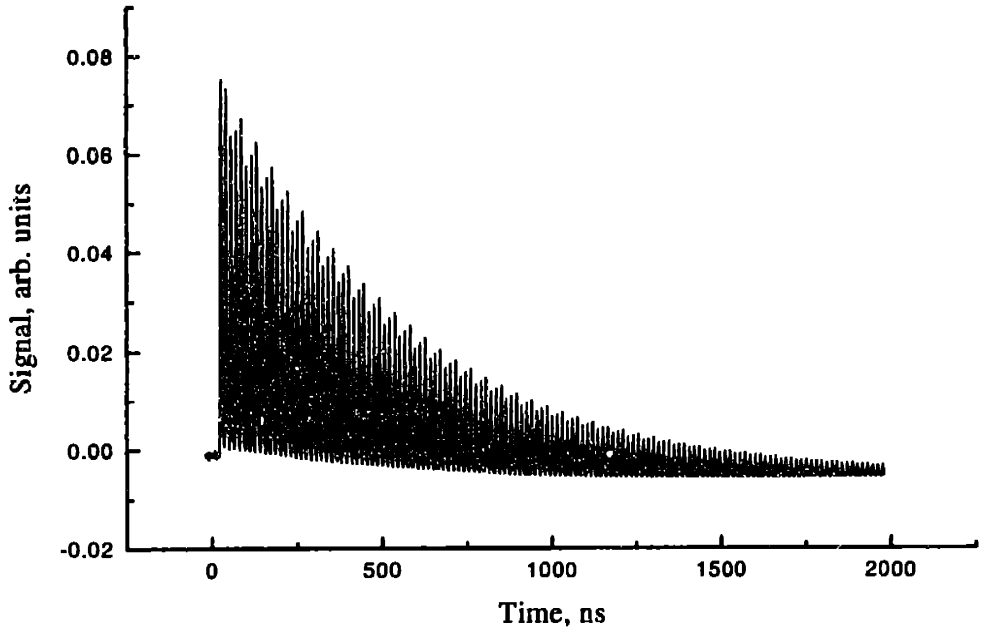


Figure 4.15a Waveform recorded at low sensitivity. Vertical scales give values in the same scale.

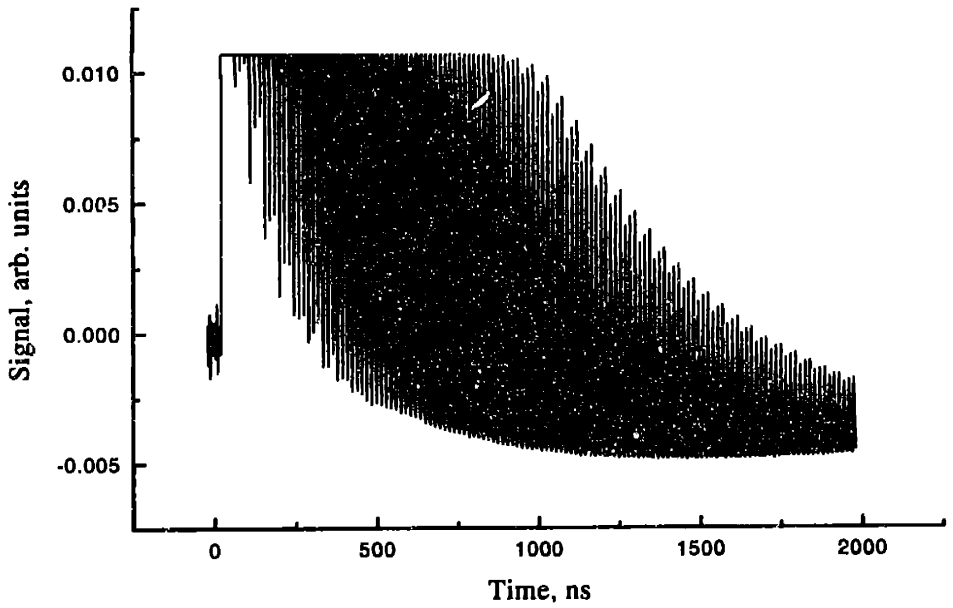


Figure 4.15b Waveform recorded at higher sensitivity. As a result the first microsecond of the waveform saturated the oscilloscope, but the second microsecond has a superior signal-to-noise ratio relative to waveform presented on Figure 4.z1a.

A very important issue that has a very strong influence on the accuracy of the absorptivity calculation has been discussed in section 4.4. Therefore, I shall limit myself to a simple reminder that it is very important to tune the cavity to be 1.5 of mirror radii long.

Finally, we can make a rough estimate of the sensitivity of out CRD setup and compare it to the shot noise limit. Neglecting dark current and thermal noise in the detector and connecting its output directly to the input of the oscilloscope we limit signal to noise consideration to two major sources - shot noise of the photocurrent and noise of the oscilloscope. We do not to include amplifier noise into consideration because bi-planar phototube does not have internal amplification. Oscilloscope's largest source of noise is digitizing process. For 8 bit oscilloscope, the noise is defined as a value of the lowest bit over the value of the full scale. Thus oscilloscope coefficient of variance is

$$\frac{\sigma_S}{S} = \frac{1}{256} \approx 3.9 \cdot 10^{-3}. \qquad [4.44]$$

To calculate the shot noise we have to remember that in our oscilloscope (Model 3430A, LeCroy) digitizing speed of 1Gs/sec is achieved by coupling two digitizers, each capable of 500Ms/sec. Therefore, for each point taken by oscilloscope signal is collected for 2ns. The lowest coefficient of variance (highest signal to noise ratio for the shot noise) is achieved for the first, most intense peak. A value of 10mV maximum peak height was not uncommon in our experiments. This voltage into 50Ω corresponds to the photocurrent of 200μA. To produce this current in two nanoseconds a charge of approximately $4 \cdot 10^{-13}$C. This charge corresponds to the number of photoelectrons, N_e, equal to $2.5 \cdot 10^6$. Therefore, the coefficient of variance of the shot noise is approximately equal to $6 \cdot 10^{-4}$. Comparing the values for the digitizing error and for the shot noise error we see that we will be able to perform no better than at a noise level six times higher than the shot noise for the light flux characteristic to our experiment.

The sensitivity limit of our experiment is obtained by dividing the standard deviation of the signal (dominated by the oscilloscope digitization noise) by the effective

pathlength (defined as twice the decay time of the cavity times the speed of light). This means that with averaging over a hundred laser shots per each point in the spectrum

$$\sigma_\alpha \approx \frac{3.9 \cdot 10^{-4}}{1.1 \cdot 10^{5}} \approx 4 \cdot 10^{-9}\, cm^{-1}. \qquad [4.45]$$

This is, essentially the sensitivity achieved by the CRD experiments conducted in the visible with much longer cavity decay times, due to better mirrors. We hope that our developments in the Cavity Ringdown method, allowing highly sensitive measurements even with relatively lossy cavities, will not only enable one to study spectroscopy of weakly absorbing molecules (such as iso-cyanogen) deeper in the UV, but will facilitate broader acceptance of CRD for the study of spectroscopy of relatively short lived phenomena (triplet states of various molecules, metastable molecules, double resonance experiments), that require faster cavity decay times to be resolved.

References

1. Press, W.H., et al. "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press, 2nd ed. 1994.

Appendix 1

(courtesy of Dr. A. Kachanov)

Mode matching in CRD by a pinhole and an imaging lens. The position and size of the beam at the pinhole are found numerically.

The program is written in MathCad.

NOTE: all distances in mm!

Values for some fundamental parameters:

$\lambda := 230 \cdot 10^{-6}$

$n_a := 1.0$

$n_m := n_Q(\lambda \cdot 10^3)$

$\lambda \cdot 10^3 = 0.23$

$n_m = 1.528$

For a symmetric resonator of a length

$L := 1125$

and mirrors of curvature

$R_m := 750$

one has

$$z_0 := \sqrt{\left| 2 \cdot R_m - L \right| \cdot \frac{L}{4}}$$

$$\omega_0 := \sqrt{\frac{\lambda \cdot z_0}{\pi \cdot n_a}}$$

$$\omega_m := \left(\frac{\lambda \cdot L}{2 \cdot \pi \cdot n_a}\right)^{\frac{1}{2}} \cdot \left[\frac{2 \cdot R_m^{\,2}}{L \cdot \left(R_m - \frac{L}{2}\right)}\right]^{\frac{1}{4}}$$

which give

$z_0 = 224.76$

$\omega_0 = 0.154$

$\omega_m = 0.308$

The Gaussian beam parameters at the CRD mirror are therefore defined as (note the sign of R_m !):

$$a := \frac{\lambda}{\pi \cdot n_a}$$

$$q_m := \left(\frac{1}{-R_m} - i \cdot \frac{a}{\omega_m^{\,2}}\right)^{-1}$$

Plugging ABCD law for the optical chain matching the ringdown cavity to the pinhole.

First of all, the cavity mirror itself makes a diverging lens effect:

using $d << R$

and $n_a = 1$, it gives :

$$\begin{bmatrix} 1 & 0 \\ \frac{n_a - n_m}{n_a \cdot (-R_m)} & \frac{n_m}{n_a} \end{bmatrix} \cdot \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_a}{n_m} \end{bmatrix}$$

$$\begin{bmatrix} 1 & \frac{d}{n_m} \\ \frac{n_m - 1}{R_m} & 1 \end{bmatrix}$$

$d = 6.35$

where

$$d := \frac{25.4}{4}$$

So the beam coming out of the RD cavity is given by

$$q_a := \frac{R_m \cdot (q_m \cdot n_m + d)}{n_m \cdot (q_m \cdot n_m - q_m + R_m)}$$

On the other hand, a pinhole, of diameter D, has far field which is well approximated by a Gaussian transverse mode with infinite radius and waist given by+

$D := 0.10$

$$\omega_p := \frac{D}{2.6}$$

which gives

$$q_p := \frac{i}{a} \cdot \omega_p^2$$

Now, q_p and q_a must be related by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & d_2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & d_1 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{1}{f} \cdot \begin{bmatrix} (f - d_2) & (d_1 \cdot f - d_1 \cdot d_2 + d_2 \cdot f) \\ -1 & (f - d_1) \end{bmatrix}$$

+ see worksheet "pinhole" on page 185.

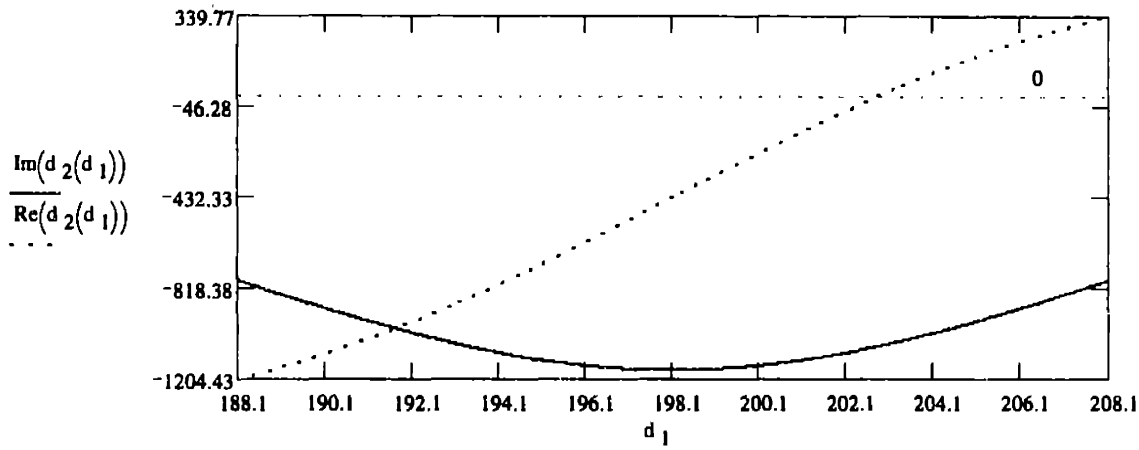
Finally, from

$$q_a = \frac{A \cdot q_p + B}{C \cdot q_p + D}$$

we can obtain the relationship between d_1 and d_2

$$d_2(d_1) := \frac{q_a \cdot q_p - q_a \cdot f + q_a \cdot d_1 + q_p \cdot f + d_1 \cdot f}{q_p - f + d_1}$$

$d_1 := f - 10, f - 9.98 .. f + 10$



f≡198.1

$d_1 := f$

So

given

$Im(d_2(d_1))=0$

$Re(d_1)>0$

$Re(d_2(d_1))>0$

$d_1 := minerr(d_1)$

$d_1=222.8$

and

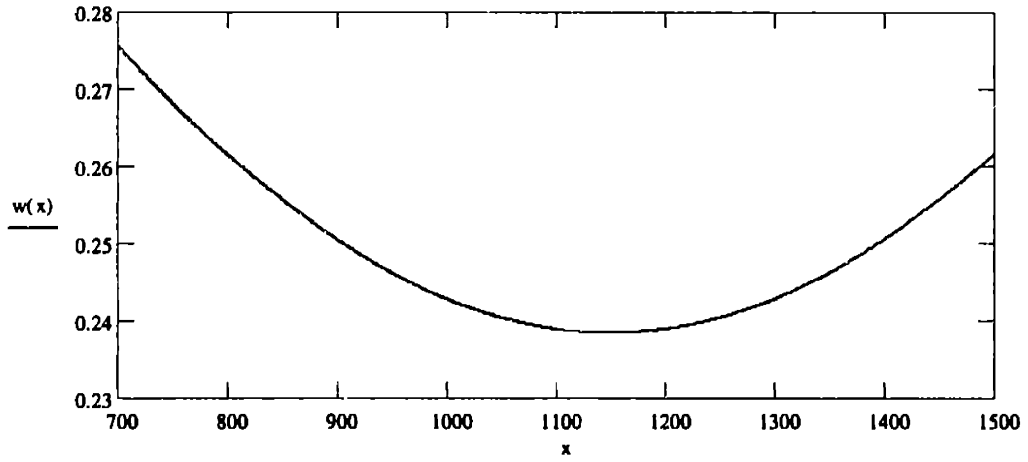
$d_2(d_1)=519.2$

$d_2(d_1)+d_1=742$

$$q_a(x) := \frac{(f-x)\cdot q_p + (d_1\cdot f - d_1\cdot x + x\cdot f)}{-q_p + (f-d_1)}$$

$$w(x) := \sqrt{\frac{\lambda\cdot(|q_a(x)|)^2}{\pi\cdot Im(q_a(x))}}$$

$x := 700..1500$



Refraction coefficient of fuzed silica at 20 C

wavelength is an air wavelength in microns

$$n_Q(\lambda) := \sqrt{\left[2.104682 - \left(9.601184\cdot10^{-3}\right)\cdot\lambda^2 + \frac{8.445395\cdot10^{-3}}{\lambda^2} + \frac{1.799233\cdot10^{-4}}{\lambda^4} - \frac{1.122795\cdot10^{-5}}{\lambda^6} + \frac{6.560106\cdot10^{-7}}{\lambda^8}\right]}$$

Example:

$n_Q(0.7) = 1.4552932$

$n_Q(0.7000) = 1.4552929$

Worksheet Pinhole

Here an Airy function is graphically compared with a gaussian least-square fitted function to show that these functions are indeed qualitatively quite close. This is to justify that the far field transverse profile of light beam propagating from a pinhole is rather well approximated by a gaussian beam.

As a practical result, the waist size of the gaussian beam from a pinhole is calculated.

$\sigma := 2.7$

gives the best LSQ fit shown below

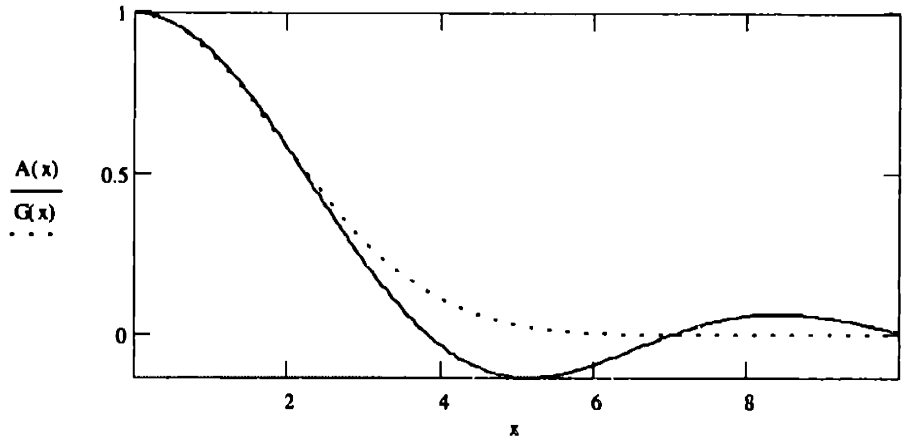
Airy and Gaussian functions

$$A(x) := 2\cdot\frac{J1(x)}{x}$$

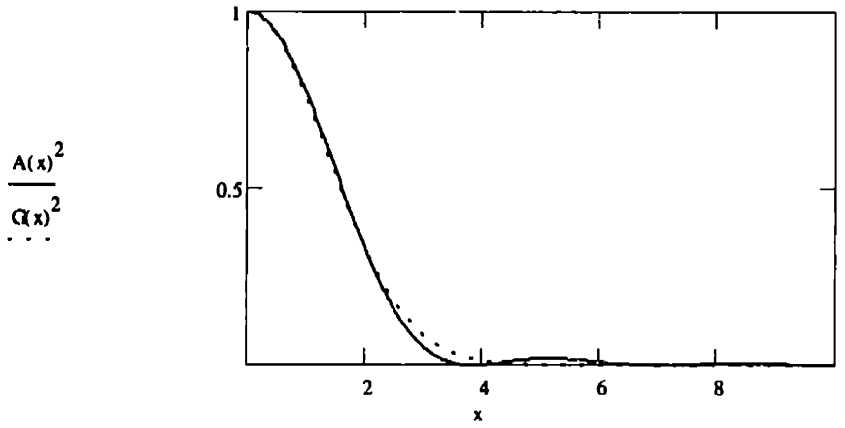
$$G(x) := \exp\left(-\frac{x^2}{\sigma^2}\right)$$

$x := 0.001, .05 .. 10$

$(x \equiv k \cdot a \cdot \theta)$



This comparison is quite favorable, but the situation gets even better when the intensity (square of the field) is considered:



$n := 1$

$\lambda := 550 \cdot nm$

$k := \frac{2 \cdot \pi \cdot n}{\lambda}$

$\omega_0 := 50 \cdot \mu m$

$nm \equiv 10^{-9} \cdot m$

$\mu m \equiv 10^{-6} \cdot m$

$$z_0 := \frac{\pi \cdot n \cdot \omega_0^2}{\lambda}$$

$$\omega(z) := \omega_0 \cdot \sqrt{1 + \frac{z^2}{z_0^2}}$$

At large z

$$\tan(\theta) = \frac{\omega_0}{z_0}$$

which gives, for small angles

$$\theta = \frac{\lambda}{\omega_0 \cdot \pi \cdot n}$$

Now, given a pinhole of a radius a, the far field diffraction pattern (Airy function) is well approximated by a Gaussian (especially when one considers the intensity) with angular width given by

$$\sigma_\theta = \frac{2.6}{k \cdot a}$$

which corresponds to a gaussian waist size

$$\omega_0 = \frac{\lambda}{\sigma_\theta \cdot \pi \cdot n} \quad \text{or} \quad \omega_0 = \frac{a}{1.3}$$

Appendix 2

(courtesy of Dr. A. Kachanov)

Vignetting effects in Cavity Ringdown Spectroscopy producing "mode beat patterns"

Note: all sizes are in mm. The program is written in MathCad.

Let us take two cavity mirrors with the radius of curvature $R = 750$, and the cavity of the

length $L = 1.133 \cdot 10^3$. The mirror losses are $\gamma := 0.02$. Cavity roundtrip time is

$\tau := \frac{0.2 \cdot L}{c}$.

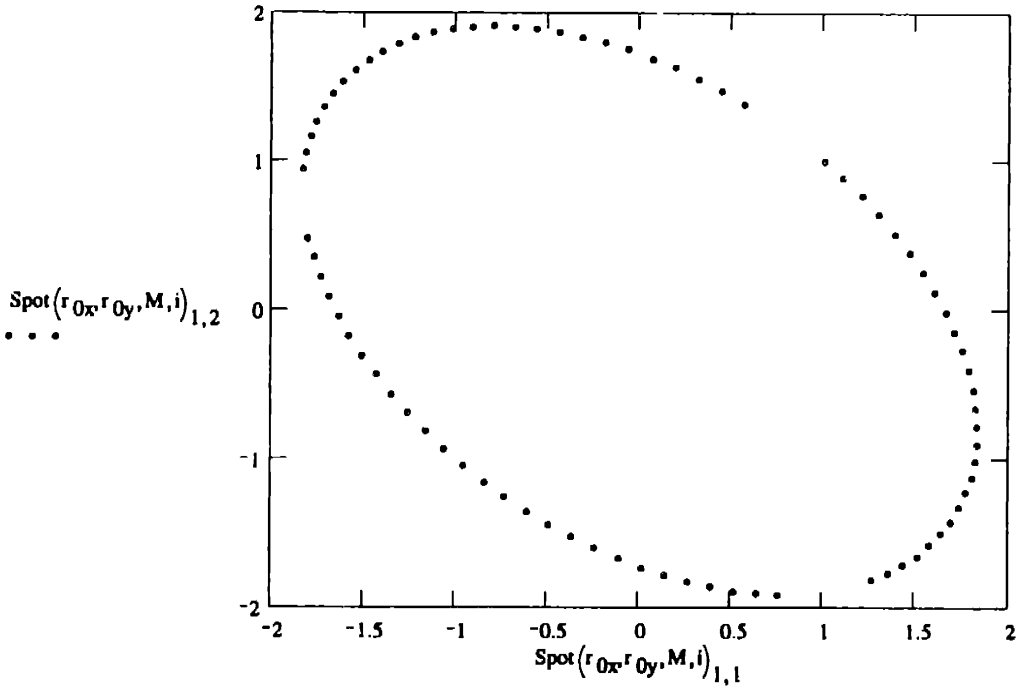
Roundtrip matrix, starting from the point just before the end mirror of the cavity, is

$M := T(L) \cdot F\left(\frac{R}{2}\right) \cdot T(L) \cdot F\left(\frac{R}{2}\right)$

The *half trace* for this cavity is $\frac{M_{1,1} + M_{2,2}}{2} = -0.48$

The value of the *eigenvalue angle* for this cavity, defined as $\theta := \operatorname{acos}\left(\frac{M_{1,1} + M_{2,2}}{2}\right)$ is

$\frac{3 \cdot \theta}{\pi} = 1.978$.



The exit spot pattern on the end mirror of the RD cavity, depending on the entrance beam parameters r0x and r0y, and on the cavity length is given on the plot above.

$$r_{0x} := \begin{bmatrix} 1 \\ 0.0025 \end{bmatrix}$$

$$r_{0y} := \begin{bmatrix} 1 \\ 0.0001 \end{bmatrix}$$

R:=750

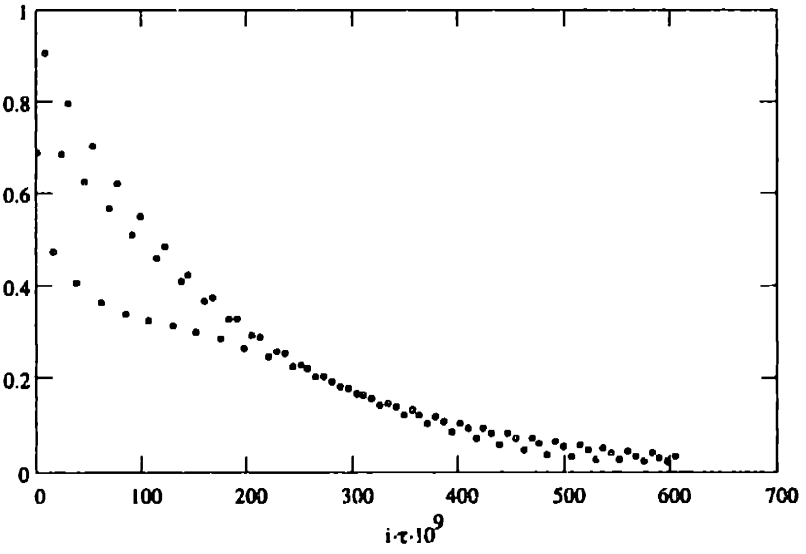
L:=(1.5 + 0.01)·R

i:=0.. 80

The ringdown signal recorded by a photodetector with the square surface of the size D := 10 , placed at a distance d := 200 from the end mirror is given on the plot below. The detector has a wire passing through its center vertically (in Y axis direction). The wire has a width w := 1 . The "wire transmission" $T_w := 0.5$.

The detector is shifted with respect to the cavity center in X direction by the distance Δ := 0.5 .The detector signal function, DS, is defined in the Detector Worksheet on page 191.

$DS(r_{0x}, r_{0y}, M, i, \gamma, d, \Delta, w, T_w)$



$\Delta \equiv 0.5$

$w \equiv 1$

Detector worksheet:

Speed of light:

$c \equiv 3 \cdot 10^{10}$

1. Beam propagation matrixes:

$$T(\alpha) \equiv \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}$$

$$F(\beta) \equiv \begin{bmatrix} 1 & 0 \\ \frac{-1}{\beta} & 1 \end{bmatrix}$$

2. Beam coordinates in the cavity reference plane on the n-th passage

rx,ry - entrance beam vectors in X and Y directions, M - cavity matrix, n - passage number

$$Spot(r_x, r_y, M, n) \equiv \left| \begin{array}{l} R \leftarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ R^{\langle 1 \rangle} \leftarrow M^n \cdot r_x \\ R^{\langle 2 \rangle} \leftarrow M^n \cdot r_y \\ R \end{array} \right.$$

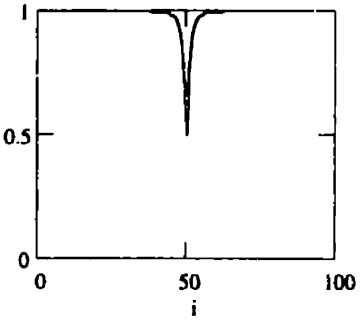
3. Pulse intensity after the n-th passage:

$$I(\gamma,n)\equiv \left|\begin{array}{l} R\leftarrow 1 \\ \text{for } i\in 1..n \\ \quad R\leftarrow R\cdot(1-\gamma)^2 \\ R \end{array}\right.$$

4. Detector response for the n-th passage:

$$DS(r_x,r_y,M,n,\gamma,d,\Delta,w,T_w)\equiv \left|\begin{array}{l} V\leftarrow Spot(r_x,r_y,M,n) \\ Rn_x\leftarrow V_{1,1}+d\cdot V_{2,1} \\ I_n\leftarrow I(\gamma,n) \\ I\leftarrow I_n\cdot\left[1-\dfrac{T_w}{1+\left(\dfrac{Rn_x-\Delta}{w}\right)^2}\right] \\ I \end{array}\right.$$

$$1-\frac{T_w}{1+\left(\frac{50-i}{w}\right)^2}$$



$T_w=0.5$