An Integrated Optical-Holographic Chemical-Vapor Sensor

By

Steven Michael Hughes

B.S. Physics, Rose-Hulman Institute of Technology, 2002

M.S. Electrical Engineering, University of Colorado, 2004

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

2007

This thesis entitled:

"An Integrated Optical Holographic Chemical Vapor Sensor"

Written by Steven Michael Hughes

has been approved for the

Department of Electrical and Computer Engineering.

Dana Z. Anderson

Robert R. McLeod

Date 6/9/07

The final copy of this thesis has been examined by the signatories, and we

Find that both the content and the form meet acceptable presentation standards

Of scholarly work in the above mentioned discipline.

Hughes, Steven Michael (Ph.D. Electrical Engineering)

An Integrated Optical-Holographic Chemical-Vapor Sensor

Thesis directed by Prof. Dr. Dana Z. Anderson

Abstract

This thesis addresses the problem of chemical vapor detection by using an integrated optical-holographic interferometer for precision interrogation of a chemical transducer. A monolithic-prism interferometer, incorporating dynamic holography in conjunction with a "sniff-locked-loop" sampling scheme, achieves high sensitivity at high speed while mitigating the deleterious effects of environmental noise and reducing costs associated with the replacement of expendable and poisonable transducers. The "sniff-locked-loop" rapidly alternates vapor sampling (at 5 Hz) between a reference and test sample, enabling synchronous detection and precision differential measurements.

The prism geometry is designed to minimize interferometer sensitivity to laser wavelength variations. To mitigate other sources of optical and electronic noise, the relative phase between the beams is modulated, mixing the low-frequency signal to a sideband of 40 kHz, using a piezoelectric-driven grating modulator. The prism system exhibits a displacement sensitivity of 180 fm/$Hz^{1/2}$, a signal-to-noise ratio of unity in a 1 Hz bandwidth. Displacement sensitivity translates into substance sensitivity depending on the transducer materials. One benchmark uses poly(*N*-vinyl pyrrolidone) as an ethanol sensor, with which we achieve a 3σ limit of detection (LOD) of 6 ppm (parts per million) in a 5s measurement time, or 1.8 ppm/$Hz^{1/2}$.

A second portable chemical-vapor sensor system is presented. It is designed for spatially independent signal processing of a linear array in a portable path-length-balanced prism interferometer. A volume-hologram-stabilized 660 nm laser enables portability but requires further support in the form of a thermoelectric-cooling system. While losing nearly an order of magnitude of sensitivity, attributable almost entirely to the linewidth stability, this prototype successfully demonstrates the proof of concept for various feature additions.

In addition to system details, this thesis examines various noise sources and the solutions implemented to mitigate the noise contributions. Supporting equipment, including the gas valve,

detection electronics, modulator components, and other designs, are presented where relevant. Presentations of several of the hurdles involving vapor delivery and concentration verification are also presented. Finally, a few general concepts necessary for an intimate understanding of the signal detection methodology are discussed. A summary of estimated noise sources is also given, followed by a performance comparison to several other transduction-based technologies.

DEDICATION

My Parents,

For their undying faith, their willingness to help wherever/however possible, and their wisdom in teaching me to keep a fire extinguisher handy.

My Instructors,

Who encouraged me to try, scrape my knees, and get back up.

Anna,

Whose spirit, confidence, and forging help me to get back up faster.

ACKNOWLEDGEMENTS

Dana. Z. Anderson:

Whose guidance, wisdom, prodding, confidence, and camaraderie helped me to get to where I am today.

Leslie Czaia:

Whose expertise with all things small, shiny, and useful enabled fabrication of the various components including the prism, modulator, and photorefractive crystals. Little of this work would have been possible without her efforts.

JILA Staff and instrument shop:

Whose expertise aided considerably in design and fabrication of many of the mechanical and electronic components.

Dimitri Veras:

Whose “outsider” viewpoint helped me asses the accessibility of the language to the general scientific community. Also, for being one of the few (as in only) members of said community who will likely ever have read this from cover to cover.

Marty, Paul, Hongke, Farhad, Matt, Evan, and the many others of the Anderson group with whom wisdom and laughter were shared.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

Trace quantities of airborne analytes, including weaponized agents, explosives constituents, and solvents, are of substantial concern to public health, safety, and national security. The problem of detecting and identifying potentially harmful vapors in a field environment holds many challenges regardless of the technology used. It is an understatement to say that the state-of-the art hazardous vapor sensors are woefully short of providing adequate detection in many real-world scenarios. Many technologies currently implemented for this task are severely limited by dynamic range, discrimination, analysis time, or costs associated with fabrication, maintenance, and disposable component replacement.

This work presents a miniature holographic interferometer that optically interrogates chemically sensitive polymers in a parallel array for trace-gas detection and discriminative identification. Though precision interferometry can offer good vapor sensitivity, precision is hampered by several factors in a system designed for field use. Factors including particulates, temperature, and interfering vapors, including water even in trace quantities, can mask analytes and degrade system performance and decision certainty. In particular, vapor concentration tends to change slowly over time scales where interferometry can be plagued by technical noise sources. To mitigate these problems we employ a double modulation scheme where the first is a low-frequency periodic comparative sampling between the test gas and a reference gas such as Nitrogen. The second modulation acts to mix the response signal away from 1/f noise while effectively amplifying the signal through heterodyne mixing with a kHz reference oscillator. The system is integrated into a compact design for robust, sensitive detection and minimization of costs associated with spent detector elements.

The need for trace gas detection has existed for centuries and was typically delegated to animals, most notably the highly adaptable canine and the single use CO/CH_4 sensor, the *Serinus canaria*. Use of animals is less desirable as currently they require substantial upkeep, often expensive training, and are prone to rapid fatigue and expiration in harsh conditions. Two research cornerstones have set the basis for the system being presented, among others. In 1961, R. W. Moncrieff established

the first cornerstone by developing a nose wherein an artificial olfactory epithelium, or olfactory membrane, of PVC, activated carbon, or silica-coated a thermistor for chemical to electric transduction. This detection process, established to be primarily one of physical adsorption and therefore easily reversible and repeatable, took one of the first steps towards the desired tireless, robust, efficient, and highly adaptable artificial nose. Twenty years later Dodd and Persaud reported their theory that odorants can be identified through use of several broadly-tuned receptor cells, or transducers, in lieu of highly specified transducers and verified their theory using semiconductor transducers in an electronic nose. Their work established the second cornerstone, a paradigm for olfaction detection and discrimination using an array of transducers with varying chemical sensitivity and selectivity.

Arrayed transducers in "Artificial nose" sensors complement analytical techniques, such as mass and optical spectroscopy, with their ability to handle complex odors, possibly consisting of many separate chemical species. Artificial noses utilize a collection of substance-selective materials that transform a substance signal into a physical one such as mass, resistance, volume, or optical refractive index. Finding and synthesizing transducer materials that are both sensitive to and selective of the substances of interest is of course crucial. Tremendous research has already been conducted on aspects of chemical and biological engineering that are necessary for enhanced chemical transducers use in various broad and specific applications. However, as each of these transducers relies on some adsorption interaction, a material index shift must occur as the material itself, even in a physical adsorption process, is modified in the interaction. As any material is optically transparent if made thin enough, any transducer material is therefore a viable component to our system. The challenge of interest here is devising a means to interrogate the transducers with high sensitivity and high speed while also keeping expendable transducer costs to a minimum.

The focus of this thesis is to present such a system capable of interrogating a given transducer with a good balance of speed and sensitivity while keeping expendable costs to a minimum. We will begin by discussing our implementation of interferometry, long accepted as fundamental tool for precision measurement. We will describe various design components, including the interferometer itself, the phase modulator, and the photorefractive BaTiO3 around which the system is constructed. A

brief discussion on transducer interrogation will motivate further discussion of various design features and how they contribute to system performance. Detail on sensitivity enhancement will then be presented in chapter 3 along with experimental performance and a comparison to an analytical fieldvector formalism model. We will then go into further detail on system size and cost reduction in chapter 4, limiting noise sources in chapter 5, and wrap up with details on analyte delivery and transducer fabrication in chapter 6. Throughout these chapter elements of the final system design are presented where they logically follow from the present discussions.

The first two appendices cover some of the finer nuances associated with detection limits both fundamental and when detecting a chemical. These are not critical to understanding the material presented, though they help clear up some common misunderstandings to newcomers in the artificial-olfaction community. Appendix C contains a table of the various noise sources with their relative contributions, measured where possible and calculated otherwise, effectively summarizing Chapter 5. Finally, Appendix D presents an overview comparison of various transducer based technologies for vapor detection and attempts to address the question of "which technology is better?"

Chapter 2 532nm Prism System V3

2.1 Introduction

We introduce the basic reference system, the third evolution in a line of 4 prism-based interferometers, designed to reduce the contribution of laser-frequency noise and environmental noise. At the system's core is a compact ($2cm^3$) prism-based holographic interferometer used to interrogate transducers sensitive to volatile organic chemicals. The system optically probes a chemical transducer, using interferometry to convert the phase information to an amplitude signal, which is then detected and measured through electronic means.

The prism's small size and monolithic construction reduces sensitivity to environmental noise sources, including temperature and vibration, while making the system robust for field deployment. Phase modulation and beamsplitting are accomplished with a single diffraction grating such that the interferometry is contained in a nearly monolithic structure. We use a photorefractive BaTiO3 crystal for dynamic holography enabling single-fringe interferometric detection and phase-drift adaptation. Phase adaptation is advantageous in two ways; it filters slow thermal, and mechanical, drift from the interferometer signal while adapting to topographical non-uniformities in the chemical transducer enabling rapid replacement of low-cost, mass-produced transducers.

In the following sections we begin with an overview of the prism interferometer and its components. We then discuss the "sniff-lock loop", a synchronous detection scheme in which the analyte gas is compared to a reference for low concentration and comparative detection. Finally, we conclude after touching on a few key experimental points for optimizing performance.

2.2 Interferometer Design

The major features of the optical nose are shown in figure 2-1. The interferometer sensitivity to environmental noise sources is minimized by using a nearly monolithic design. Beamsplitting and modulation occur at a piezo-driven diffraction grating after which one beam picks up phase

information from the chemical transducer. In this section we focus on the interferometry and how each component of the interferometer is designed to reduce noise.

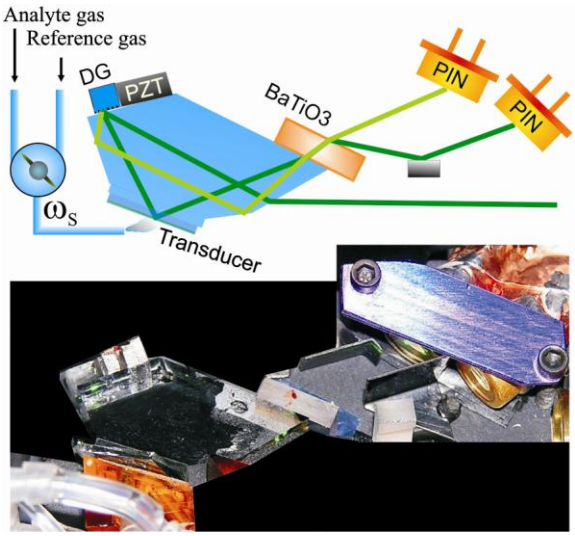

Figure 2-1: (A) A schematic of the optical noise system showing the geometry of the prism interferometer. A diffraction grating (DG) acts as a beam splitter while its transverse modulation phase modulates the diffracted beam. The polymer transducer converts a chemical vapor to an optical phase shift, through total-internal reflection (TIR), after which the signal experiences phase-to-amplitude conversion in the BaTiO3 photorefractive and is detected by PIN photodiodes. (B) A picture of the actual prism interferometer, total volume of the prism is under 2 cm^3.

2.2.1 Holographic Interferometer

Dynamic holographic interferometry is particularly useful when optical wave fronts are poor and time scales allow a holographic medium to automatically compensate for interferometer drifts. Such drifts may include thermal expansion, structural or transducer fatigue, aging, and drift due to chemical uptake or the presence of a background interferant. A photorefractive crystal, used as a beam recombiner in a Mach Zehnder interferometer as illustrated in figure 2-2, is capable of performing the operation of path-length stabilization and single fringe phase-to-amplitude conversion without additional servo equipment.

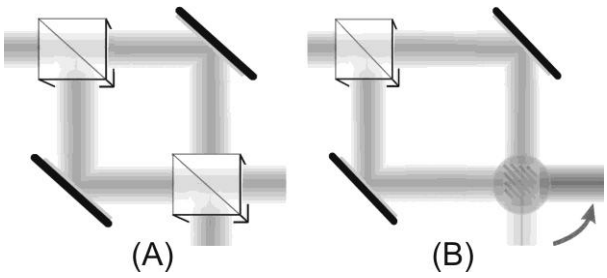

Figure 2-2: A Mach-Zehnder interferometer using (A) beamsplitter and (B) a photorefractive as the recombination elements. While the beamsplitter will recombine such that any phase variations between the beams dictates the brightness or darkness of the two output beams, the photorefractive applies gain to one beam such that a single bright and dark fringe are observed at the two output beams.

Two-beam coupling in diffusion-dominated media, including Barium Titanate, produces nominal dark and bright outputs when used as the recombining element in a Mach-Zehnder interferometer. Typical response times are on the order of fractions of a second to many seconds wavelengths in the range of 532 nm to 780 nm. In practice, measurements of small signals are easily obscured by various sources of noise, which is especially problematic using dark-fringe detection. In ref [10] we introduce a modulation scheme to improve the signal-to-noise ratio, though an interferometer design that is resilient to external noise sources would further aid sensitivity. To this end we present the prism interferometer described in section 2.2.3, after we present the components that will enable said interferometer to function.

2.2.2 Grating Splitter/Modulator

A grating phase modulator is used to simultaneously perform two tasks. It acts as a beamsplitter and as a phase modulator which provides the reference modulation for signal enhancement and enables a weak low-frequency calibration signal to evaluate the system sensitivity. Additionally, interferometer integrity is maintained within the bulk prism using index matching fluid at the grating-prism interface. Containing the interferometer within the prism helps to reduce environmental interference including fluctuations in air pressure and temperature. Modulation is achieved by applying a shear-displacement modulation along the K_d vector of the diffraction grating,

shown in figure 2-3, the 1st diffracted orders are phase modulated proportionate to the amplitude of the displacement modulation.

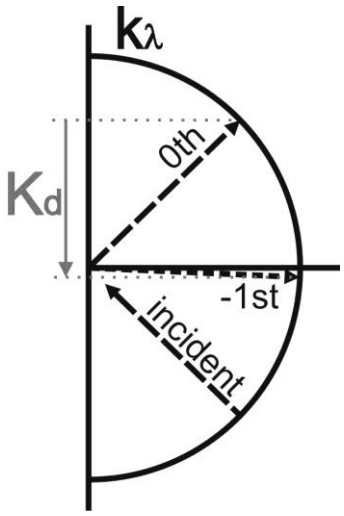

Figure 2-3: K-space diffraction picture of an incident beam reflecting and diffracting off of a grating. The -1st diffraction order is close to 0^0 to minimize angular dispersion with wavelength.

In principle, the reflected beam sees no phase shift while the diffracted orders see a phase shift proportionate to the order of diffraction. Regardless of the optical wavelength and pitch of the grating, the amplitude of the grating phase modulation directly translates to optical phase. For Piezo actuation we use a 2mm cube PICMA piezo-electric transducer (PZT). Often a preload is applied to a PZT to reduce hysteresis and allow the actuator to push as well as pull. To this end we designed a small stainless-steel flexure into which the piezo is snugly pressed. This flexure preloads the piezo, holds the diffractive grating in line with the piezo, and holds the linear actuated assembly parallel to the surface over which the grating needs to translate. Figure 2-4 shows a rendering of the flexure that was machined out of 304 stainless. Later versions, not used in this system, were fabricated using wire electro-discharge machining which yields superior tolerances (<1/10000”) and shorter fabrication times than graduate-student machining.

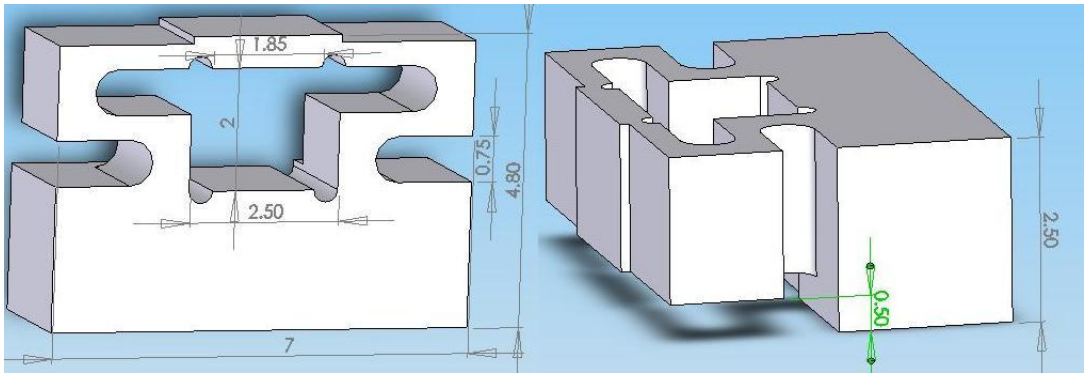

Figure 2-4: The piezo flexure designed to hold a 2mm cube piezo. The slot for the piezo has recessed material to prevent electrical shorting of the electrodes through contact with the stainless steel flexure. The view in the right shows how the flexure can be mounted onto a surface such that the piezo will push/pull parallel to the mounting surface. All units are in millimeters.

In summary, there are two parameters of the grating that are important to its acting as a beamsplitter and phase modulator. The first is that the ratio between the reflected and diffracted beam amplitudes is dependant on the phase depth of the grating and the incidence angle. The incidence angle is controlled in the prism design, while the grating depth is a wavelength-dependant parameter we have little to no control over, other than buying a different grating. Second, the optical phase shift is derived directly from the grating phase shift regardless of pitch. Therefore, smaller displacement amplitudes are necessary for higher pitch gratings to achieve the same phase modulation, though any vibration noise will have a proportionately larger contribution.

2.2.3 Prism Design

The prism-interferometer design is not only small but robust to shock, environmental interference, and other common difficulties encountered by field devices. Figure 2-1 schematically shows the prism interferometer, which has a volume of approximately $2cm^3$. A laser beam incident at Brewster's angle is split by a diffraction grating into a signal and reference beam. The transducer is deposited on a glass slide, which is mounted to the prism with index-matching fluid. After one beam interrogates the transducer, both beams intersect and write a dynamic grating in a Barium Titanate (BaTiO3) crystal, mounted directly on the prism. Precision odor detection is accomplished using the difference between the two output beams.

2.2.4 Prism Dispersion Minimization

Dispersion is a common problem in precision optical systems as many components, especially the diffraction grating and the photorefractive Bragg grating, are sensitive to wavelength. Beams diffracted from the beamsplitter experience an angle fluctuation corresponding to any wavelength fluctuation which, for non-uniform phase fronts, causes swelling/shifting of the interference pattern in the photorefractive. Furthermore, since the photorefractive BaTiO3 acts as a phase-sensitive beam recombiner, it is also a frequency-sensitive recombiner, where the degree of wavelength sensitivity scales with path-length mismatch. To reduce the affect of dispersion in the photorefractive the relative phase $\phi = |\phi_1 - \phi_2|$ between the beams should be independent of wavelength.

$$\frac{d\phi}{d\lambda} = 0 \tag{2.1}$$

This is commonly achieved with a balanced-path interferometer [1], implement in the present system to within 0.1mm.

Figure 2-5 illustrates the source of dispersion based coupling fluctuations. As wavelength changes the grating wave vector will either expand and contract or shift. Any change in alignment between the optical-interference pattern and the photorefractive-index grating will cause a change in coupling conditions and be seen as an amplitude fluctuation. Therefore, any fluctuations in wavelength will cause amplitude fluctuations in the detected beams limiting the small-signal sensitivity of the system.

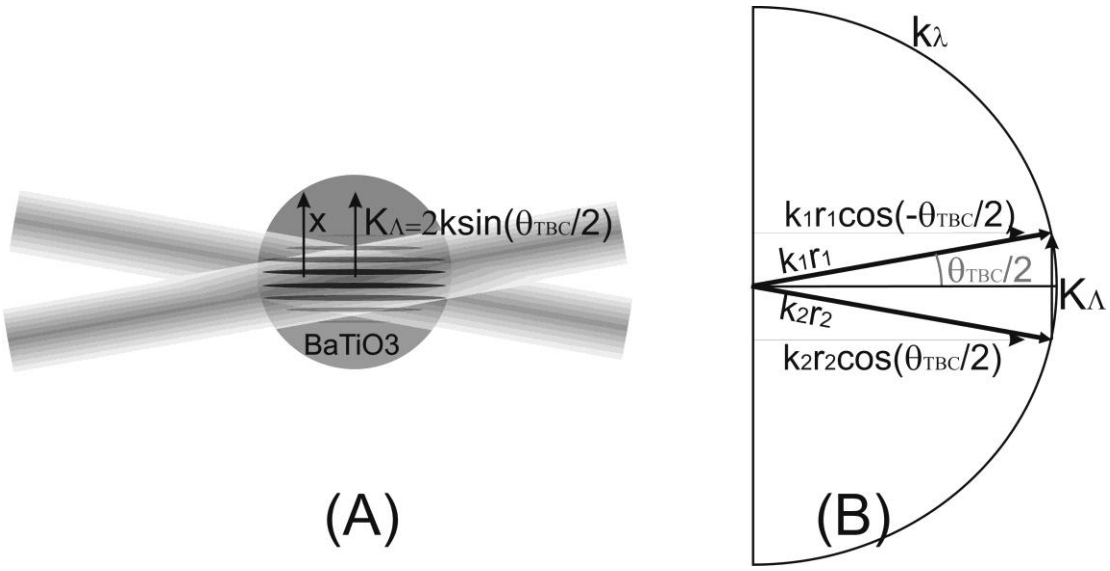

Figure 2-5: (A) Two beams interfere such that the difference in wavevectors form a grating wavevector in the photorefractive and (B) the correspond k-space representation of this interaction.

The effective displacement δx, or path length fluctuation in one beam, is found from the phase shift in the interference pattern, which is proportionate to the relative phase of the beams. The phase dependence of the optical beams therefore directly translates to the phase dependence of the grating with respect to the wavelength.

$$\delta x = \frac{\Delta\phi}{2\pi}\lambda = \frac{\lambda}{2\pi}\frac{d\phi}{d\lambda}\Delta\lambda \tag{2.2}$$

There are two dominant dispersion terms to consider. The lesser, and more obvious, is the linear dispersion term that arises from the relative phase shift between the two beams caused by a change in wavelength.

$$\frac{\partial\phi_\Lambda}{\partial\lambda} = -2\frac{2\pi n}{\lambda^2}\sin\left(\frac{\theta_{TBC}}{2}\right)(\Delta r - x) \approx -1\cdot 10^{13}(\Delta r - x) \tag{2.3}$$

Where n is the refractive index of the medium, the two-beam coupling angle θ_{TBC} is roughly 8^0, Δr is the optical path-length difference, and x is the spatial distance from the center of the two-beam coupling interference pattern.

The second, and stronger, of the dominant terms arises from the diffraction angle off of the beamsplitter. As the wavelength shifts the angle tilts and, in the limit of small tilt, this is seen as an angle change of one arm in the two-beam coupling.

$$\frac{\partial \phi_\Lambda}{\partial \lambda} = -\cos\left(\frac{\theta_{TBC}}{2}\right)\frac{2\pi}{\Lambda_d \lambda \cos\theta_m}\left(\Delta r - x\right) \approx -2.9\cdot 10^{13}\left(\Delta r - x\right) \tag{2.4}$$

where $\Lambda_d = 416nm$ is the pitch of the diffraction grating, and θ_m is the mth order diffraction angle measured at the grating. Since both terms depend on $\left(\Delta r - x\right)$ we simply sum their contribution for the remainder.

If the path lengths are balanced, the interference pattern swells and contracts about x=0. Otherwise, it appears that the entire pattern shifts, though it is still swelling about $\Delta r = x$. This means that this dispersion noise effect can be minimized by integrating over the entire beam, with a sufficiently large photodiode, as long as the arms are perfectly balanced. Additionally, spatial non-uniformities in transmission of the beams, say from scatter or inhomogeneities, increase the effective amplitude noise detected at the two-beam coupling output. Skipping to the punch line, for a reasonable path length match of $\Delta r = 0.1mm$ and a linewidth of <1MHz, as reported for the Coherent Compass 315M, we get an effective displacement noise of 320fm. Considering that specification sheets often give upper limits, as does our estimate of the path length matching, it is reasonable to expect the noise to be smaller than this.

There are effectively two points that are critical for the design of the prism. The first is that while a 0th order diffracted (reflected) beam will remain essentially unchanged, a diffracted beam of nth order (not equal to zero) will experience beam deflection as a function of wavelength. Therefore, when two beams (0th and $\pm$1st) recombine in the photorefractive, one beam tilting will change the coupling conditions and the tilt will be seen as a small signal. The second point is that the interference pattern, and over longer time scales the resulting photorefractive grating, will swell/contract if path lengths are matched. They will further experience a phase shift proportionate to any path length mismatch, which will contribute more significantly to energy transfer between the ports, for short time scale fluctuations. If the frequency noise spectra at the signal of interest is significant, this will contribute to noise on the signal. While optical path balancing will mitigate the second point, the first

is minimized in geometry, by choosing the diffraction angle similar to that in figure 2-3 such that the angle deflects as minimally as possible.

2.2.5 Brewster Coupling

Having designed the prism for balanced path interferometry the next step is to maximize energy coupling. To this end we use a Brewster window at the input of the prism. However, since we measure both beams leaving the photorefractive crystal we need to optimize the energy coupling out of the system. Since the (+) port is more normal to the surface than the (-) port we angle the photorefractive, and the prism face to which it mounts, to Brewster couple the (-) port since it is the steeper, and weaker, of the two beams. In this way the (+) port suffers a 12% hit in power while the minus port is faithfully transmitted.

2.2.6 Total Internal Reflections

In this design total internal reflections (TIR) serve two purposes. First, they act to guide the beams from the beamsplitter/modulator to the photorefractive in lieu of mirrors. Second, the reflection taking place at the polymer acts to convert chemical content information into optical phase through the refractive-index based Fresnel-phase shift inherent to TIR.

However, this same Fresnel-phase shift will occur where it is not intended if there are pressure, temperature, or other refractive index fluctuations at the glass-gas interface. To this end the housing and the vapor delivery head are designed to shield and minimize errant vapor flows over the non-transducer total-internal reflection (TIR) interfaces. Additionally, while the polymer interrogating TIR angle is close to the critical angle to maximize sensitivity, the other angles are made steeper to minimize the fundamental sensitivity to refractive index perturbations. That is, the angles are as far from critical as possible without compromising the balanced path length requirement. The number of reflections are also minimized to further reduce the contribution of index fluctuations in the surrounding atmosphere. In later designs, any non-transducer interrogating TIR is shielded from external index fluctuations with a reflective silver coating, which is protected by a SiO2 coating.

2.2.7 Additional Geometry

There are two final geometric constraints to the design of the prism. The first is that the photorefractive crystal should be as far from the polymer transducer as possible, preferably on the opposite side of the prism. The reasoning for this is that the crystal acts as a phase sensitive recombiner and so any perturbations caused by vapor exposure will affect coupling if they affect the photorefractive in a significant way. Additionally, the photorefractive BaTiO3 is both relatively expensive and brittle, warranting its isolation from potential harm.

The second geometric requirement is for an odd number of total reflections. This allows for rotation and translation of the prism to enable a beam alignment compensation for any manufacturing tolerances. If there were an even number of reflections the beam recombination would be more difficult to adjust to the center of the photorefractive to achieve optimal coupling. Additionally, an odd number of reflections minimizes the effects of incident angular instability into the prism since a tilt in the incident beam results in the center of the beam overlap translating orthogonally to the photorefractive grating wavevector, rather than along it. This, similar to the dispersion effects, will be seen as an interference pattern that is swelling/contracting about the center rather than translating, thereby decreasing associated amplitude coupling noise.

2.2.8 Assembly

The grating modulator flexure, and other permanent mounting components are attached to the prism surface using a Norland UV cured adhesive. This enables alignment of components before curing. The transducer slide, photorefractive crystal, and diffraction grating are all optically coupled to the surface of the BK7 prism using a Cargille Laboratories series A, index matching fluid (n=1.52). The capillary force exerted by a micro liter of fluid provides the only adhesion force holding the photorefractive crystal and transducer slide to the prism. The prism itself is located by 3-point contact to a base plate and adhered to the same base plate using a combination of UV adhesive and a generic 5-minute epoxy.

2.3 Chemical to optical signal transduction

2.3.1 Layer Sensitivity

Having established the interferometry, we need a means of signal transduction. More specifically, we need a way to convert the chemical content being sampled to an optical phase. Polymer-sorption kinetics have been widely studied though the details of exactly how and why certain polymers are selective to certain analytes and not others is beyond the scope of this work. We are primarily concerned that our transducers act as a good little black boxes that, when exposed to a chemical, will, through refractive index or dimension variation, alter the phase of a beam interacting with said transducer.

Chemical-to-optical phase transduction requires a means of interfacing the light with the transducer, such that a phase change can occur. There are a variety of transduction methods to choose from. Propagation through a waveguide made of the transducer depends on the quality and uniformity of the transducers. Any damage to a wave-guiding polymer could greatly affect the guiding conditions making the sensors very fragile and short lived. Multiple bounces or passes through a polymer, and possibly a gas, would improve phase response. However, any pass through vapor also subjects the beam to noises associated with air pressure and concentration inhomogeneity. Therefore, a means of signal transduction that involves minimal interaction with the vapor is preferred.

The use of TIR shields the laser from exposure to the vapor flow and any associated index fluctuations. The phase uniformity of the reflection off of the surface will be comparable to the surface quality of the polymer, which is potentially very non uniform. However, the photorefractive can naturally adapt to any surface contour, thus still performing single-fringe interference, even though surface features from one polymer to another may vary. In this way we save on the cost of making optically flat polymers to within a thousandth or less of a wavelength. TIR also improves transduction-phase gain as it takes advantage of the Fresnel-phase argument on TIR, which is more sensitive to surface index than optical path length. The mechanism for the increased sensitivity arises

from the phase slope around the critical angle (~42^0 in a glass:air interface) in figure 2-6 and the change in the critical angle for small index changes.

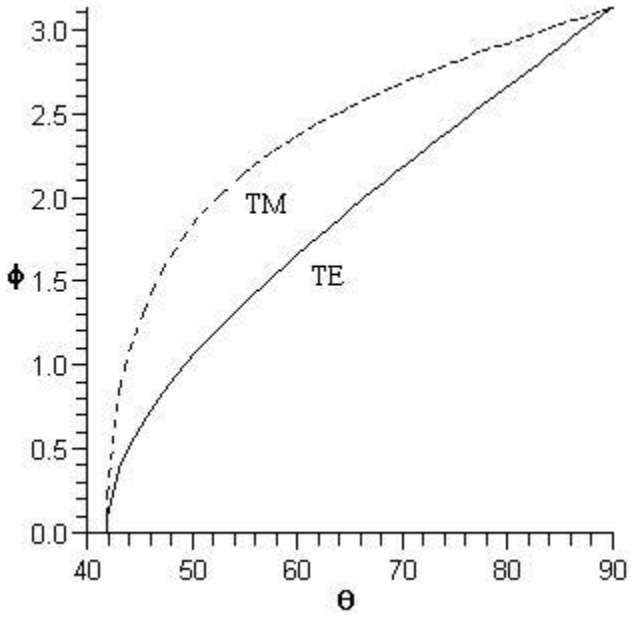

Figure 2-6: Calculated phase shift for total internal reflection (TIR) associated with the angle of reflection. The beam is oriented TM or transverse magnetic. A change in the refractive index at TIR results in the critical angle at ~42^0 to shift left or right, which in turn changes the reflected phase for a fixed angle reflection.

For TIR to occur the internal reflection must be at an angle greater than the critical angle.

$$\theta_c = \arcsin\left(n_{ext} \Big/ n_{\text{int}}\right) \tag{2.5}$$

TIR at a BK7-air interface with a 532nm index of 1.51:1, will have a critical angle of about 41.5^0. However, one might expect that since the refractive index of the transducer will rarely be equal to that of the bulk material (BK7 in this case) the critical angle will change resulting in external coupling. To alleviate this concern we use Snell's law to show that while a change in the incident angle occurs when the surface index changes, which we approximate with a slab of material placed over a pre-existing TIR interface, the beam refracts to maintain TIR.

$$\theta_f = \arcsin\left(n_0 \sin(\theta_0) \Big/ n_f\right) \tag{2.6}$$

The final internal incident angle θ_f is adjusted by the index of the slab n_f and the initial internal incident angle θ_0. We plot the final angle for a set of initial angles in figure 2-7 showing that as long as the beam satisfied TIR in the prism, before the surface index changed, it will continue to do so regardless of the material index. We also show this graphically using k-space in .

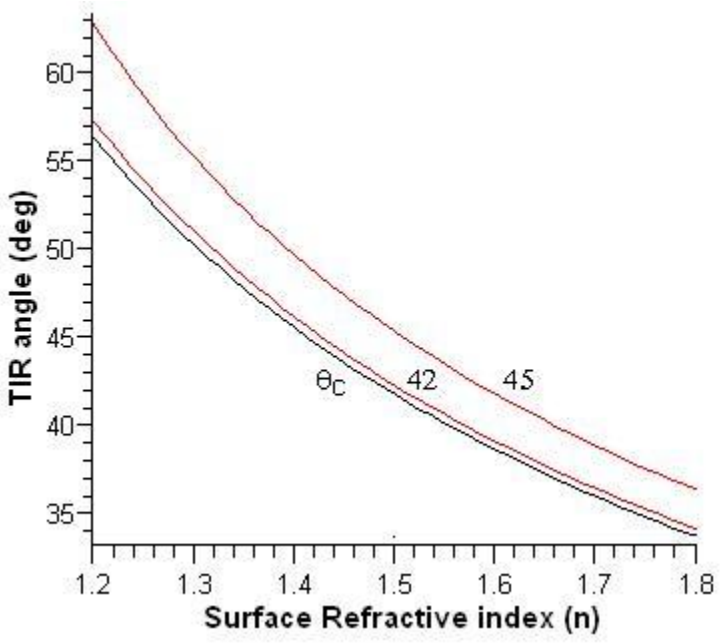

Figure 2-7: The critical angle and the TIR angles depend on the refractive index of a perfectly flat layer of index n_2 deposited on a TIR interface originally setup for BK7-air at 45^0 and 42^0. Under these initial conditions, depositing a reasonable (n=1.3..1.5) transducer on the surface should yield a viable transduction scheme without readjusting the incident angle every time a transducer material is changed.

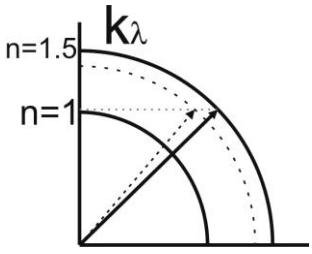

Figure 2-8: K-space depictions showing that, as long as the incident beam is within the critical angle for TIR, changing the refractive index at the surface, uniformly, will not result in a loss of TIR because the beam will refract and still TIR within the polymer.

Before proceeding further a detour is necessary to give a foundation upon which we can base analysis of the sensitivity of the TIR angle. Since slabs of vapor will not be placed upon the TIR interface we must work up an approximate model of what does occur at the interface. To this end we examine the case of a monolayer, and an incomplete monolayer, of ethanol bound to the surface. Afterwards, we will finish analyzing the sensitivity based on the TIR interface establishing the last bit of foundation necessary to discuss a basic optical noise system.

2.3.2 Ethanol Molecular Size

To aid in the determination of the theoretical detection limits we need a model of how a detection event occurs. We know that in some manner an analyte, say ethanol, is adsorbed onto the polymer. The actual mechanism is unimportant as long as, in the small concentration approximation, it can be represented as a layer on the surface causing a change in path length and/or refractive index. We construct a model based on path length and index change for light that TIRs off of a surface where adsorption events occur. Our model is based purely on the size and the refractive index of a condensed (liquid) analyte. We use ethanol as a baseline since it is a relatively safe volatile organic chemical (VOC) to handle and there are a variety of polymers exhibiting varying sensitivities to it.

We approximate Ethanol (C_2H_5OH) as an ellipsoid with a major and minor axis dictated by the longest and shortest dimension of the molecule. Q. Wei et al [2] modeled ethanol, using Chem3D, from which they concluded a minimum diameter of 420pm while the largest diameter is estimated at 610pm. As an upper estimate we will use 610pm as the effective diameter of ethanol for our calculations. It is unknown how accurate this model is, though the dimensions correspond to the reported atomic density of liquid ethanol $\rho_l = 1.03 \cdot 10^{-8}$ molecules/pm^3 [3]. Regardless, we only need an approximation to estimate the thickness of a monolayer.

2.3.3 Concentration

Having established an approximate molecular size the next step is to determine a means of estimating the phase response to the adsorbed layer. For this we need the surface concentration which

will be proportionate to the vapor concentration and the affinity, or stickiness, of the surface. Using the vapor pressure we can determine the rates of evaporation and condensation under ideal equilibrium conditions. However, rather than a simple condensate we have a surface with a certain affinity for the analyte in question and an airborne concentration well below the equilibrium vapor pressure.

Since a steady state enclosed system is not a reasonable approximation, the airborne concentration, as far as the transducer is concerned, will hardly be affected by sorption/desorption since it is constantly being replaced by continuous sampling. Furthermore, since we modulate the vapor concentration between that of a pure carrier, and that of a small fixed concentration, sorption/desorption events take place over half the modulation period. We modulate around 5 Hz making the period for sorption/desorption around 0.1 second each.

In order to optimize and characterize the response of a system, when using polymers with varying sensitivities, the mechanisms for sorption must be understood. There are various models relying on different interaction parameters to classify the binding affinities. [4,5,6] Regardless of the mechanisms in play, in the limit of small analyte concentrations, we model the sorption of a gas by the formation of a single or fractional monolayer, consisting of patchy coverage with an area of coverage proportionate to the airborne and surface concentration. However, the specific rates and sorption coefficients are difficult to determine experimentally [7]. Therefore, we acknowledge these unknown conditions (sorption rates and affinities) and simplify the problem by saying that the concentration on the surface is in some way proportionate to the airborne concentration $\left(C_s \propto C_a\right)$.

To estimate the coverage and the optical response we first review the base principle upon which the optical nose's novelty filter detects the presence of an analyte. The beam interrogating the polymer (array) interacts with a reference beam in the photorefractive writing a phase-adaptive grating, which gives rise to amplitude fluctuations that spatially correlate to phase fluctuations in the beam. This means that if a portion of the optical beam were to experience a phase shift, the same spatial portion of the beam would alter its state of coupling through the photorefractive such that an intensity fluctuation would be observed in the same spatial region of the beam. Since no imaging takes place between the polymer and the photodiodes (the photorefractive is roughly at the focus of the interrogating beam, and the distance between the polymer and diode is well within a hundredth of the

Raleigh range) the phase, and therefore intensity, cross-talk between spatial regions of the beam would be minimal (minimal diffractive effects), as described in a bit more detail in section 4.3.

2.3.4 Molecular Monolayer Estimate

Since the photorefractive causes local phase shifts to be detected as local amplitude fluctuations, the detection of a single polymer's partial response (partial in surface area) with a single photodiode means that such a response would be indistinguishable from a weaker uniform phase shift over the entire detected beam. We also assume that said patches of sub-monolayer analyte were large enough to see spatial fluctuations in phase and ignore difficulties associated with sub-wavelength sized interactions, which is admittedly one of the weakest points of this first order approximation. Therefore, with this list of assumptions, the novelty filter [13] effectively gives an intensity response proportionate to the percent coverage. This analysis only works for our optical means of interrogation and photorefractive-phase processing, and few to no other nose systems, since we use a photorefractive-dynamic hologram to spatially process the phase shifts. This also means that removing the photorefractive would likely negate this approximation as many of our assumptions would no longer be applicable.

With the above assumptions, the analyte concentration at the surface, or the surface coverage, determines the effective refractive index at the surface, assuming we average over the surface. Rather than delve into the sorption/desorption kinetics we will look simply at the end result; a certain concentration of molecules will be adsorbed onto the surface, through steady state exposure to a test vapor, thereby changing the surface refractive index. We approximate the surface concentration with the number of adsorbed molecules N_a of diameter d over area A, from which we arrive at the average refractive index at the surface.

$$n_s = \frac{n_a N_a + n_p N_p}{N_a + N_p} \approx \left(1 - \frac{N_a d^2}{A}\right) n_p + \frac{N_a d^2}{A} n_a \tag{2.7}$$

Where $N_p = A/d^2$ is total number of binding sites of the polymer, assuming $N_a << N_p$, and $\{n_a, n_p\}$ are the refractive indices of the {analyte, polymer} respectively. The monolayer depth *d* is the long axis

length of the ethanol molecule, estimated earlier. We use the long axis as this will yield the largest optical path length change to ensure we do not underestimate the optical path contribution to the phase shift. The change in index is then calculated based on the surface concentration of the analyte.

$$\Delta n = \frac{N_a\left(n_p - n_a\right)}{N_a + N_p} \approx \left(n_p - n_a\right)\frac{N_a d^2}{A} \tag{2.8}$$

This equation is useful to estimate the sensitivity limit based on surface concentration $C_s=N_a/N_p$, as long as a means of determining surface concentration from airborne concentration using sorption affinities can be found. Conversely, given a known chemical LOD and the associated refractive index change, the surface concentration can be estimated based on the measured airborne concentration. From this, a relation between surface and airborne concentration can be estimated.

When calculating the phase shift associated with a thin layer deposited at the TIR boundary one might think that the optical path added by that layer would be significant. To illustrate the negligibility compared to the phase shift, for a given fractional monolayer, we expand on the simple model above to find the path-length change. If, at the transducer TIR surface, we deposit a monolayer of the analyte, we estimate a phase shift corresponding to twice the optical path length in the monolayer.

$$\phi_{OP} = 2\frac{n_a d}{\cos\left(\theta\right)}\frac{2\pi}{\lambda} \tag{2.9}$$

The variable d is the effective diameter of the analyte molecule (610pm), and θ is the internal beam angle. If we now assume an incomplete monolayer, the phase angle is reduced proportionately by the percentage of area covered.

$$\phi_{OP} = 2\frac{n_a N_a}{\cos\left(\theta\right)}\frac{d^3}{A}\frac{2\pi}{\lambda} \tag{2.10}$$

where A is the interrogated area and N_a is the number of analyte molecules in said region. The limit of a monolayer is then achieved for $N_a=A/d^2$.

For completeness we include the change in incidence angle arising from the change in surface index.

$$\theta = \arcsin\left(\frac{1}{1+\frac{\Delta n}{n_p}}\sin\left(\theta_{TIR}\right)\right) \tag{2.11}$$

Where the initial TIR angle is given by θ_{TIR}. We now calculate the optical phase shift ϕ_{OP}, due to optical path length. We then compare this phase shift to the phase shift associated with the refractive index change Δn at the TIR interface. However, the phase shift due to the optical path ϕ_{OP}, for a fractional monolayer, is nearly 3 orders of magnitude smaller than the phase shift arising from the Fresnel equations at the TIR interface. This indicates that bulk diffusion is insignificant to the small signal detection limit other than a potential accumulated drift source. The comparison is shown graphically in figure 2-9, after we discuss the TIR phase sensitivity.

It may appear that this section discussed bulk sensitivity rather than surface sensitivity. For the case of measuring the phase difference using just the change in optical path this is true, there is effectively no distinction. For small concentrations (ppm) there will be little to no difference between this simplified partial monolayer model at the surface, and the absorption of an identical number of analytes into the bulk of the polymer. However, this analysis was indeed of a surface effect since if the analyte were absorbed into the polymer bulk, the Fresnel phase shift (to be discussed shortly) would *only* occur while the analyte molecules exist in transition at the surface. Once they diffuse deeper into the polymer the Fresnel phase shift would be lost since it is the index at the TIR interface that gives rise to the nearly three-orders-of-magnitude more sensitive Fresnel mechanism.

2.3.5 Molecular Multilayer Estimate

One issue with the previous section is that sorption may also take the form of a molecular multilayer at the surface. This effectively comes down to the difference between using the Langmuir isotherm which assumes the limit of a monolayer, and the BET isotherm which is based on the Langmuir but assumes no such limit to uptake [8]. Since the BET model is based on the Langmuir model, in the small vapor pressure approximation $P<<P_{sat}$ the two scale nearly identically. For the particulars the reader is referred to [8].

To explain this seemingly circular argument simplistically, we look at a basic case for sorption on a surface at low airborne concentrations. The molecular uptake would be governed by the affinity between analyte molecules compared to the affinity between said molecule and the transducer surface. If the affinity between molecules is low, compared to the affinity for the transducer surface, this effect is negligible. The condition under which the potential formation of a fractional multilayer is significant, compared to the fractional monolayer, occurs when the ratio of affinities (molecule-molecule vs molecule-polymer) approaches the inverse ratio in surface coverage (~1/ppm). Since the probability of an analyte molecule hitting a sorbed analyte molecule is say… 1 in a million to that of it hitting the bare surface, the contribution to detection by the formation of a fractional multilayer, for ppm airborne concentrations, is negligible.

We examine two extreme circumstances under which the ratio of affinities would approach a million, such that the fractional multilayer formation would rival that of the fractional monolayer. First, the heat of vaporization is so high that the analyte is anything but volatile when in liquid form. If such heat were liberated upon condensation, it would likely cause the "clump" of analyte molecules to be liberated from the transducer surface since the heat of desorption, was so much lower. While this rapid adsorption/desorption might yield a larger signal than the single sorption event, the probability of such an event occurring would be proportional again to the surface coverage and it would decrease the actual concentration on the surface.

Second, the affinity for the transducer is so low that the analyte will likely not stay on the transducer long enough for a second analyte to attach to it, meaning a poor transducer was selection was made for the analyte of interest. Granted, the above assumes ppm levels, but the purpose of the following sections is to determine limit of detection, hence ppm and lower concentrations.

2.3.6 TIR Phase Sensitivity

Having examined the optical phase shift arising from an increased optical path in the surface adsorbed monolayer, this section looks at the phase shift due solely to the refractive index change at TIR. Again, this section relies on the assumption made in sections 2.3.3 and 2.3.4, primarily noting that the photorefractive allows for spatially dependant phase fluctuations to give rise to amplitude

fluctuations in the same spatial region. The total detected intensity fluctuations are proportional to the percentage of coverage, and we therefore approximate the response as a uniform refractive index and therefore optical-phase response, the magnitude of which scales with the fractional coverage. Borrowing the Fresnel equation for TM reflection [43],

$$\frac{E_r}{E_i} = \frac{\cos\theta_{TIR} - n_s\sqrt{1-n_s^2\sin^2\theta_{TIR}}}{\cos\theta_{TIR} + n_s\sqrt{1-n_s^2\sin^2\theta_{TIR}}}, \tag{2.12}$$

we calculate the phase delay based on the refractive index at the surface of the polymer(n_s) in comparison to air (n_0=1).

$$\phi_{TM}(n_s) = 2\arctan\left(\frac{n_s^2\sqrt{\sin^2(\theta_{TIR}) - \frac{n_0^2}{n_s^2}}}{\cos(\theta_{TIR})n_0^2}\right) = 2\arctan\left(\frac{n_s^2\sqrt{\sin^2(\theta_{TIR}) - \frac{1}{n_s^2}}}{\cos(\theta_{TIR})}\right) \tag{2.13}$$

Using just the TIR phase shift we find the effective displacement and its dependence on an index change Δn, which is associated with a monolayer or a fractional monolayer as per equation (2.8).

$$\delta x = [\,\phi_{TM}(n_s + \Delta n) - \phi_{TM}(n_s)\,](532\text{nm}/2\pi) \tag{2.14}$$

We plot the resulting displacement in figure 2-9, from which we see that the sensitivity to the Fresnel phase shift is nearly 3 orders of magnitude over that of the propagation through the monolayer.

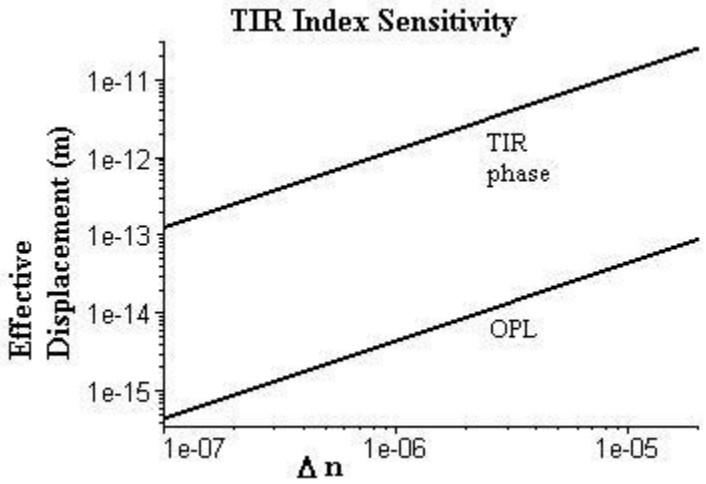

Figure 2-9: The effective displacement (for 532nm) from the Fresnel equations (top line) and the averaged optical path length (OPL) of a partial monolayer (bottom line) given an ethanol-induced–index-perturbation Δn for TIR glass:air interface at 45^0. The S/N=1 LOD of 180fm (2.1urad) is equivalent to a TIR phase-to-index sensitivity of $2.4\cdot10^{-7}$ *RIU* (refractive index units) and an optical path length to index sensitivity of $3.9\cdot10^{-5}$ *RIU* .

From here we work backwards to estimate, to first order, the surface concentration necessary to yield our achieved displacement LOD of 180fm, defined by a unity S/N. The index at the TIR surface is expressed as $n_s=n_p-\Delta n$ which, using equations (2.7) and (2.8), expands to

$$n_s \approx \left(1-\frac{N_a d^2}{A}\right)n_p + \frac{N_a d^2}{A} n_a \tag{2.15}$$

where N_a is the number of analyte particles adsorbed on the surface. The index change Δn at the TIR interface can then be found.

$$\Delta n = \left(n_p - n_a\right)\frac{N_a d^2}{A} \propto \left(n_p - n_a\right)C_s \tag{2.16}$$

For an index sensitivity of $2.4\cdot10^{-7}$ *RIU* (refractive index units) the effective surface concentration C_s is just over a ppm. The system's actual chemical LOD is roughly 514fm (for 1.8ppm/$Hz^{1/2}$ airborne concentration) corresponding to a sensitivity of $6.7\cdot10^{-7}$ *RIU* , which in turn corresponds to a theoretical surface concentration of C_s=4ppm. The loss in sensitivity is due to noise introduced by the vapor sampling at the transducer, likely associated with temperature, pressure, and analyte fluctuations. The values used for the various indices of refraction are given in the table below.

Material	Refractive index (532nm)
BK7	1.51-1.52
Poly(N-vinyl pyrrolidone)	1.53
Cellulose Acetate Butyrate	1.475
Air	1.00003
Water	1.33
Ethanol	1.36

Table 2-1: Refractive indices of important materials for the optical nose.

We have shown that the Fresnel phase shift at the TIR interface is the most sensitive fundamental transduction mechanism in this system, for the detection of chemical interactions with a transducer. Again, the transduction gain is effectively limited by the physical response of the transducer to the chemical being detected. Therefore, after attaining the most sensitive interferometer possible, detection limits can only be improved upon by improving the transducer response and selectivity. Since the advent of transducer based nose systems, entire fields of study have been developed for the purpose of optimizing transducers [9, 47, 54, 59]. Therefore, we shall not delve further into chemical transduction gain mechanisms leaving that work to the respective experts.

2.4 Optical Nose System

Having established the interferometer all that remains is to integrate it into a system for the detection of chemical vapors. This requires a means of exposing the transducer to the vapor sample and of extracting a measurand from the system in the form of a voltage. This section introduces the concept of a sniff-lock loop, a synchronous exposure and detection scheme for the detection and discrimination of small content signals. It then concludes with a functional overview of the entire sensor system followed by a discussion of the performance.

2.4.1 Sniff-Lock Loop

The last component in the chemical-to-optical transduction system is the vapor valve. When systems are running for long periods of time, or moved from one environment to another, they may

experience drift. The magnitude of a signal drift can severely limit the sensitivity to small signals as their fluctuations are easily swamped. Mammals also have a similar drift offset as they adapt to chemicals, but they can improve their small-signal sensitivity by sniffing, wherein the test gas (inhaled volume) is compared to a reference (exhaled volume). Similarly, we use a valve that actively switches between the test and reference gas, or filtered test gas. The rapid switching enables comparison or differential measurements thereby improving sensing sensitivity and long term stability.

We mentioned that the sensitivity of the chemical detection was degraded compared to the displacement detection, due to noise associated with vapor exposure. It was further noted that large pressure fluctuations potentially dominated the noise of chemical detection, and so a solenoid switch, which has a characteristic dead time, was a less than ideal valve. We developed a sinusoidal modulation valve which takes two input vapors, a test and reference, and sinusoidally mixes them between two output ports. The flow through one of these ports is incident on the transducer and the other is often used to verify the delivered concentration with an independent instrument.

Because the design of the valve ensures that the flow never stops, and also equally balances the total impedance between the two output ports, the pressure noise sources are drastically decreased allowing for the chemical LOD to get within a factor of 3 of the system's optimum performance. Figure 2-10 shows the basic cross section of the valve. A typical flow rate through a port in the valve is 300ml/min and the valve is rotated at the signal frequency ω_s.

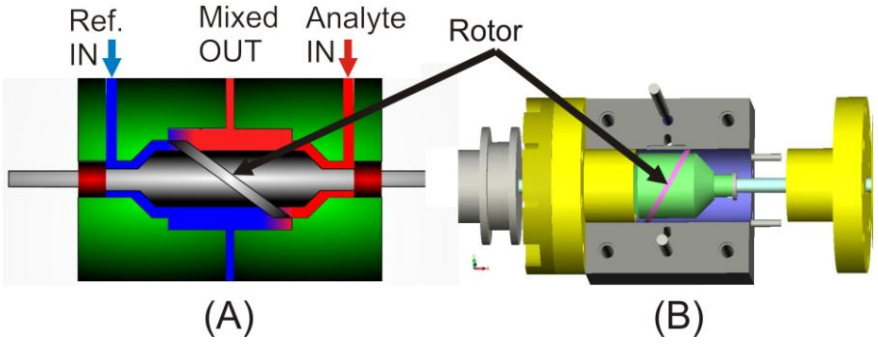

Figure 2-10: (A) A cross section of the sinusoidal valve showing two input gasses and the mixed output controlled by the position of the rotor. (B) A 3D rendering of the valve showing the rotor mounted on the outer diameter of the drum at a 30^0 angle.

2.4.2 The Optical Nose System

Having presented the chemical-to-optical signal transduction, the photorefractive phase-to-amplitude modulation conversion, and the prism-interferometer design we have the tools necessary to discuss the system as a whole. Figure 2-11 shows a function schematic in which the (-) port experiences phase modulation from both the reference modulation ω_R and the chemical signal ω_S, or its simulated PZT displacement at the same frequency. Because the photorefractive demodulates on a peak/trough the strongest amplitude modulation present occurs at the second harmonic $2\omega_R$. This second harmonic can limit the dynamic range for detecting small sum/difference frequencies $\omega_R \pm \omega_S$, as it will rail the detector output if the gain is too large. To improve the dynamic range a 25dB notch filter attenuates the second harmonic so that the fundamental ω_R dominates, enabling a signal dynamic range of over 55dB. The result is then twice demodulated, extracting the content information from the sum/difference frequency, and the DC output is read by a data acquisition board (DAQ).

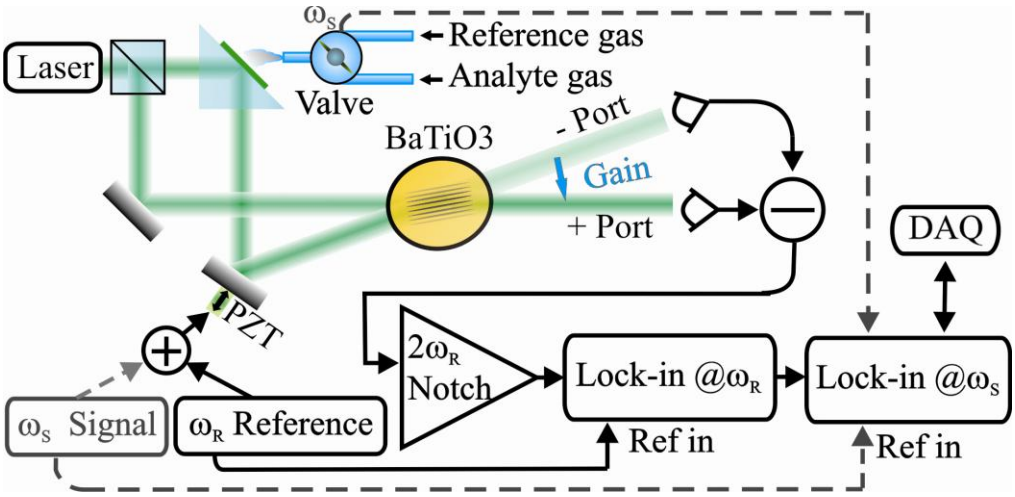

Figure 2-11: System schematic depicting both modulation mechanisms that occur on the (-) port and the detection electronics used to extract the mixed signal $\omega_R \pm \omega_S$. A notch filter suppresses the second harmonic of the reference modulation enabling a dynamic range of over 55dB.

We maximize the detected S/N by the choice of modulation depth δ_R, which in turn impacts the effective photorefractive two-beam coupling gain. The optimum modulation depth may be determined using a fieldvector formalism to analyze photorefractive two-beam coupling [10]. The experimental

optimum S/N is achieved for $\delta_R \leq 0.1 \cdot 2\pi$, or 1/10th of a wavelength, over which the noise rapidly dominates small signal detection. The amplitude of the mixed signal scales as the product of the first order Bessel functions operating on each of the modulation depths.

$$I_{sum/dif} \propto I_0 J_1\left[\delta_S(t)\right] J_1\left(\delta_R\right) \cos\left[(\omega_S \pm \omega_M)t\right] \tag{2.17}$$

where I_0 is the total intensity leaving the photorefractive. Therefore, while we gain some advantage from the photorefractive's ability to filter out slow environmental fluctuations, we also get a linear, rather than the weaker quadratic, response to small signals. The optical nose system, minus supply and demodulation electronics, is shown in figure 2-12.

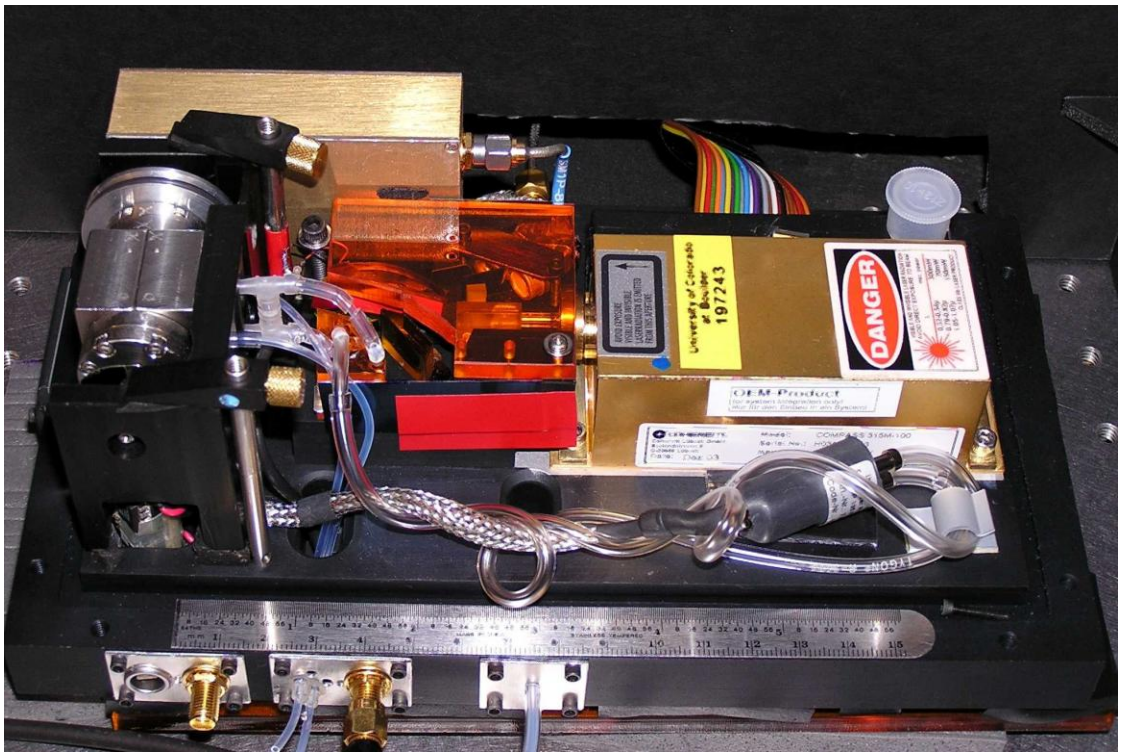

Figure 2-12: Image of nose system including compass 315M laser (right), pump (bottom right), interferometer (center), valve (left), and detector (top left).

2.4.3 Sensing Performance

The prism system exhibits an equivalent displacement sensitivity of about 180 fm/$\sqrt{\text{Hz}}$, that's a S/N of unity in one second for a displacement of about 1/1000 the size of an atom. This measurement is accomplished using a simulated 5 Hz weak signal added onto the reference signal

modulating the displacement of the PZT. We use displacement sensitivity as a benchmark for the interferometry by which we compare fundamental system performance, where a typical calibration data set is shown in figure 2-13 below.

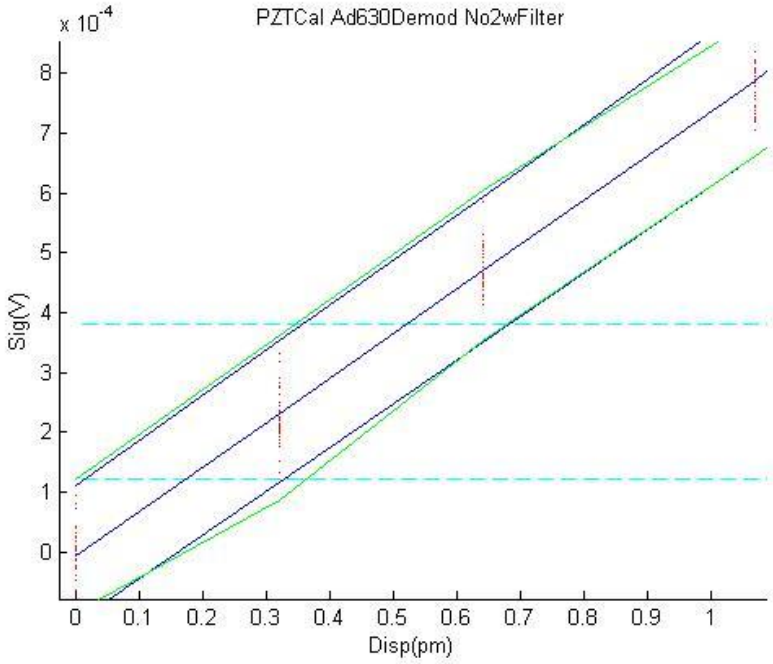

Figure 2-13: Displacement LOD of PV3. The 3-sigma LOD is a little over 350 fm while the SNR=1 limit occurs just over 50 fm. Over 35 data points were taken for each of the 4 displacements tested, where the integration time for each data point is 5 seconds. The title denotes this to be a Piezo simulated signal using onboard demodulation rather than a first stage lock-in amplifier and no notch filter.

Displacement sensitivity translates into substance sensitivity depending on the transducer materials. One benchmark uses PNVP as an ethanol sensor. Our 3-sigma LOD of 5 ppm (parts per million) in a 5 s measurement time compares favorably with the micro-cantilever system reporting a LOD of 8 ppm in 40 s [11], and the surface-acoustic wave system with a LOD of 0.23 ppm in 5 min [12], the latter improving LOD with longer integration time. Figure 2-14 shows a sample LOD calibration plot. With sensitivity comparable to other artificial nose technologies, our optical interferometer provides fast response and an ability to directly compare two odors or detect small changes in a complex odor environment.

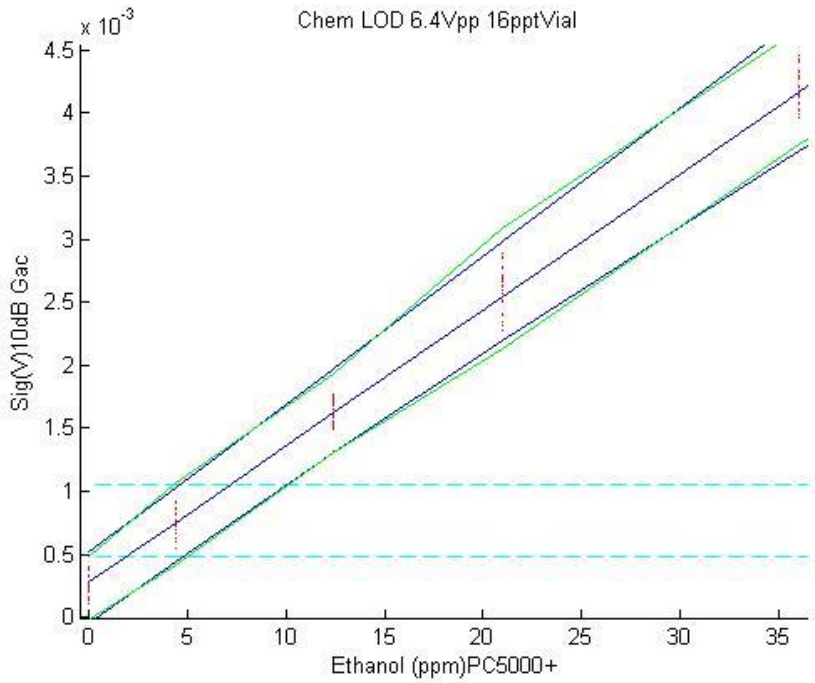

Figure 2-14: A typical chemical limit of detection plot showing the sensitivity to ethanol with a PNVP transducer. The 3-sigma LOD is 5 ppm while the SNR=1 limit is just under 900 ppb, again for a 5 second integration time. The title denotes this to be a chemical limit of detection plot with reference modulation amplitude given in Vpp and the ethanol sample is diluted from a vial that delivers up to 16 ppt. Ethanol concentrations were verified using an Ion Science Phocheck 5000+ Photoionization detector.

2.5 Summary

Small size and monolithic construction are key to attaining high precision from a fieldable interferometer. The prism geometry is designed to minimize interferometer sensitivity to laser wavelength variations. To mitigate other sources of optical and electronic noise, the relative phase between the beams is modulated at 40 kHz using a piezo-electric–sheer-driven grating modulator.

Dynamic holography tolerates imperfect optical surfaces, therefore transducer array production quality can be relatively low making transducer replacements cheap. As the polymer transducers sit on a single glass element, they are easily exchanged in the field to accommodate the odor sensing problem at hand. The novelty filter further mitigates deleterious effects of environmental noise [13], a severe limiting factor in the field.

We presented our “sniff-locked loop” detection method wherein the system input rapidly alternates (at 5 Hz) between a reference and test sample for synchronous comparative detection. This

method is arguably the best means to sense small changes in a large interferant background. The prism system exhibits 180fm/$Hz^{1/2}$ displacement sensitivity, an order of magnitude above shot noise, and an ethanol sensitivity of 1.8ppm/$Hz^{1/2}$.

Chapter 3 Modulation-Enhanced Sensitivity

3.1 Introduction

Having established the basic reference system we now take this chapter to discuss the mixing technique used to achieve sensitivity enhancement of the interferometer improving transducer interrogation. We use dynamic holography as it is particularly useful for interferometry when optical wavefronts are poor and time scales allow a holographic medium to compensate automatically for interferometer drifts. Two-beam coupling in diffusion-dominated media, including Barium Titanate, nominally produces a single dark and bright fringe at the two output ports, when used as the recombining element in a Mach-Zehnder interferometer. While this property is useful for novelty and monotony-image filtering [13], it is typically not ideal for precision interferometric measurements, because small changes in the relative optical-path length give rise to detected intensity signals that vary as the square of the change [14].

Conventional wisdom may suggest the use of drift-dominated photorefractive media, such as lithium niobate, for which detected intensity signals vary linearly with phase. Yet diffusion-dominated media can otherwise be attractive in some applications because of their high nonlinearity, appropriate speed, good optical quality, and well-understood behavior. Applications such as ultrasonic detection employ photorefractive crystals and improve sensitivity through the use of large bias electric fields [15], heterodyning with electro-optic modulation [16], and detection of non-Bragg diffraction orders from thin photorefractive materials [17].

The modulation scheme presented allows a diffusion-dominated holographic interferometry system to retain most of its advantages, including high coupling efficiency and low-frequency environmental-noise filtering while mitigating its quadratic signal sensitivity, which is especially problematic in the presence of 1/f noise. By choosing an appropriate beam splitting ratio, and phase modulating one interferometer beam with appropriate modulation depth at high frequency, we enhance the detection of a weak, low-frequency signal. The modulation up-converts the signal to a frequency range where detection sensitivity is not dominated by electronic 1/f noise or low-frequency laser-

amplitude noise. In the case presented here, our modulation scheme improves system sensitivity by over 30dB.

3.2 Theory

3.2.1 Apparatus

Modulation and synchronous detection are achieved in the relatively straightforward set-up illustrated in figure 3-1. One arm of the interferometer is phase modulated at ω_R=40 kHz; both interferometer outputs are detected with an auto-balancing noise subtraction circuit [18]. We denote the arm that receives two-beam coupling gain as the (+) port and the arm that experiences loss as the (-) port. A $2\omega_R$ notch filter suppresses the reference modulation frequency second harmonic to maintain good dynamic range. A lock-in amplifier is used to synchronously detect at ω_R. Our objective is to maximize interferometer sensitivity to a signal-induced path-length change in one interferometer arm.

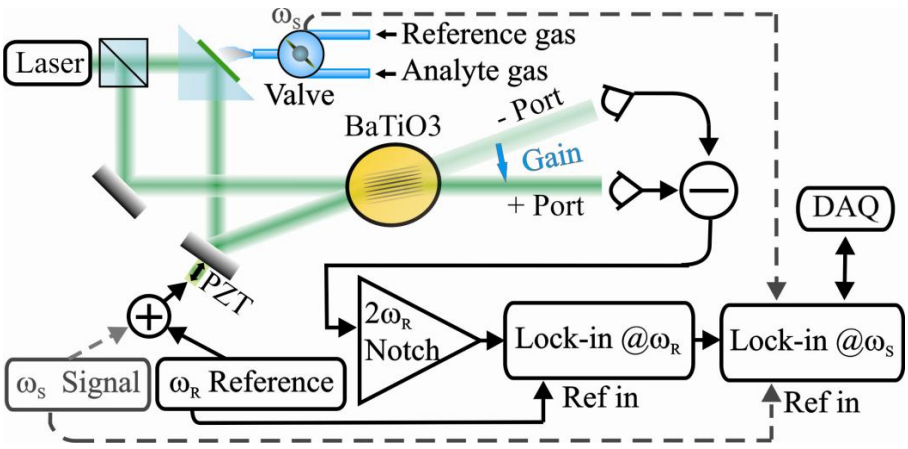

Figure 3-1: Schematic of the improved holographic interferometer using a kHz reference signal and dual synchronous demodulation of the mixed signal. The 5Hz signal generator is only for calibration purposes.

The key to maximizing the detected signal-to-noise (S/N) ratio lies on the choice of modulation depth, which in turn impacts the effective photorefractive two-beam coupling gain. The optimum modulation depth may be determined using a fieldvector formalism to analyze photorefractive two-beam

coupling [19,20]. We develop this formalism by first examining the unmodulated case. This simpler case serves as a backdrop with which to compare the modulation-enhanced result, which we analyze second.

3.2.2 Unmodulated Case

We are interested in detecting a low-frequency signal having modulation depth δ_S at frequency ω_S. The fieldvectors illustrated in figure 3-2 are written as

$$\begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \begin{bmatrix} E_{+0}\ \exp\left[\mathrm{i}\delta_S\ \sin(\omega_S t)\right] \\ E_{-0} \end{bmatrix} \tag{3.1}$$

We write the field amplitudes corresponding to the two interferometer arms in terms of a splitting angle θ_0, measured from the (+) port: $E_{+0} = E_0 \cos(\theta_0)$ and $E_{-0} = E_0 \sin(\theta_0)$. The splitting angle is directly determined by the beamsplitter ratio as $\tan^2\theta_0 = |E_{-0}|^2 / |E_{+0}|^2$.

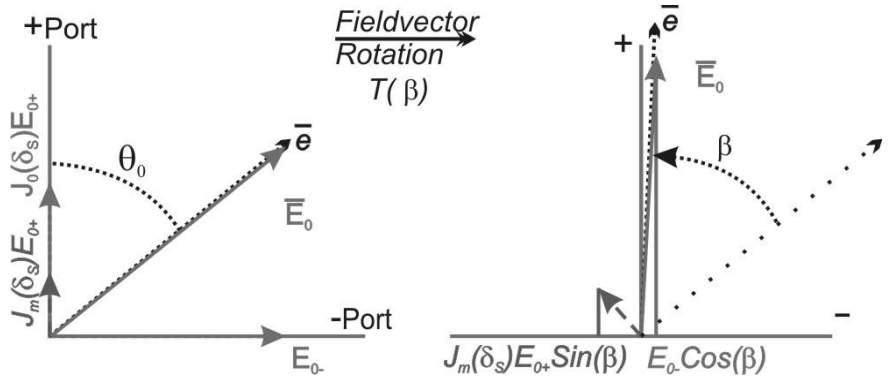

Figure 3-2: Evolution of fieldvectors with a single small signal. The unenhanced signal [Eq. (5)] arises from the |sum|^2 of the projection of the signal 2nd harmonic (m=2) and (-) port fieldvectors onto the (-) axis.

The two-beam coupling energy transfer is derived from the photorefractive gain, which is determined by the eigenvalues of a density matrix formed from the input fieldvectors [19].

$$\rho = \frac{1}{I}\begin{bmatrix} |E_{\omega+}|^2 & \sum_{\omega} E_{\omega+}\sum_{\omega} E^*_{\omega-} \\ \sum_{\omega} E^*_{\omega+}\sum_{\omega} E_{\omega-} & |E_{\omega-}|^2 \end{bmatrix} \approx \frac{1}{I}\begin{bmatrix} |E_+|^2 & E_+E^*_- \\ E_-E^*_+ & |E_-|^2 \end{bmatrix} \tag{3.2}$$

where $I = \sum_{\omega} E_{\omega}E^*_{\omega}$ normalizes the density matrix. Note the diagonals sum to 1 while the off-diagonal terms depend on the modulation depth, where stronger modulation transfers energy away from the carrier into sidebands. We use $E_{\omega\pm}$ to denote the electric field component, modulated at frequency ω, present at the (+) or (-) port. Also, the density matrix principal eigenvector, the eigenvector with the largest corresponding eigenvalue, will lie along the fieldvector $\vec{E}_0$, defined by initial beam-splitting angle θ_0, assuming $\delta_S < 0.1 \cdot 2\pi$. This formulation of the density matrix assumes that only the product of fieldvectors, whose difference frequencies are less than the reciprocal photorefractive time constant $\Delta\omega \le 1/\tau$, can write constant or slow-moving diffraction gratings [19]. This means that the photorefractive crystal works to cancel out sufficiently low-frequency signals, thereby decreasing sensitivity to interferometer drifts as well as low-frequency signals.

The output of the photorefractive interferometer is given by the fieldvectors at the output of the photorefractive medium. The output fieldvectors are in turn derived from the input fieldvectors, $\overline{E} = \overline{\overline{T}}\,\overline{E}_0$, through a transfer matrix

$$T \approx \begin{bmatrix} \cos(\beta) & \sin(\beta) \\ -\sin(\beta) & \cos(\beta) \end{bmatrix}. \tag{3.3}$$

The rotation angle β depends on the net coupling strength, which is the product of the real-valued coupling constant and the interaction length. The transfer matrix rotates all fieldvectors by the same angle. In the limit of large photorefractive gain and negligible modulation, the density-matrix principal eigenfieldvector rotates to align with the (+) axis, or $\beta \to \theta_0$. Thus, for large coupling, maximum energy is transferred to the (+) port, leaving the (-) port dark. It follows that the detected intensity, I_-, is the magnitude squared of the sum of the fieldvector projections on each port's respective axis, shown here for the (–) port.

$$I_- = \left|E_-^{'}\right|^2 = \left|-E_+ \sin(\theta_0) + E_- \cos(\theta_0)\right|^2 \tag{3.4}$$

The carrier couples to the (+) port reducing the shot-noise level at the (-) port. As a result, detection is more sensitive at the (–) port. Using an autobalancing detector between the two ports will double the signal up to a factor of 2, but it will significantly increase the detected shot-noise level. Therefore, use of an auto-balancer is only beneficial if laser technical noise is the limiting factor. For a sinusoidal small signal, the strongest-surviving signal-bearing term is the second harmonic.

$$I_{2\omega} = I\ J_2\left[\delta_S(t)\right]\sin(2\beta)\sin(2\theta_0)\cos(2\omega_S t) \tag{3.5}$$

where J_2 is a second-order Bessel function of the first kind. Note that signal strength is maximized when $\theta_0 = 45^0$ (i.e., a 50:50 beamsplitting ratio), though for reasons not relevant to the present discussion our beamsplitting is not 50:50.

At this point we should reiterate the purpose of this chapter, which is modulation enhancement of a weak low-frequency signal, ω_S. Equation (3.5) gives the largest directly-detectable signal and suffers from its second-order Bessel response, which for small signals is approximately $J_2(\delta_s) \approx \delta_S^2/8$. Ideally, only shot-noise limits the sensitivity to the signal δ_S, and is easily calculated from the total detected intensity. In practice, however, it is typically nontrivial to minimize other noise sources such as scatter in the photorefractive crystal. Post-detection electronic noise is another common limitation, especially at low signal frequencies, where 1/f noise tends to be significant. At this point we have established the backdrop of a basic system with its limitations. We now move on to modulation-enhanced detection.

3.2.3 Modulation Enhanced Detection

A reference modulation can be introduced into either arm, so we modulate the input to the (-) port for mathematical simplicity

$$\begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \begin{Bmatrix} E_{+0}\ \exp\left[\mathrm{i}\left(\ \delta_S\ \sin(\omega_S \mathrm{t})\right)\right] \\ E_{-0}\ \exp\left[\mathrm{i}\ \left(\delta_R\ \sin(\omega_R \mathrm{t})\right)\right] \end{Bmatrix} \tag{3.6}$$

where δ_R is the reference modulation amplitude and its frequency, ω_R, is chosen at sufficiently high frequency such that the dominant noise is not 1/f.

It may seem that strong modulation of the electric fields above would complicate the fieldvector picture in that phase modulation transfers energy from the carrier into a multitude of harmonics. However, analysis of the density matrix yields eigenvectors whose angles and corresponding eigenvalues are easily determined by the density matrix, as before. For this case we use the full form of the density matrix which accounts for the various fieldvector frequency components.

$$\rho = \frac{1}{I}\begin{bmatrix} \left|E_{\omega+}\right|^2 & \sum_{\omega} E_{\omega+}\sum_{\omega} E^*_{\omega-} \\ \sum_{\omega} E^*_{\omega+}\sum_{\omega} E_{\omega-} & \left|E_{\omega-}\right|^2 \end{bmatrix} \approx \frac{1}{I}\begin{bmatrix} \left|E_{+}\right|^2 & \sum_{\omega} E_{\omega+}E^*_{\omega-} \\ \sum_{\omega} E^*_{\omega+}E_{\omega-} & \left|E_{-}\right|^2 \end{bmatrix} \tag{3.7}$$

The primary change in the density matrix compared to equation (3.2) is the off-diagonal component dependence on fieldvector frequency. The significance of this is that while the trace is still unity regardless of modulation depth, coupling is dictated by the interference of frequencies that exist on both ports to the limit of $\Delta\omega \le 1/\tau$. In the absence of the same modulation signal existing on both ports before coupling, modulation depth exists in the off-diagonals in the form $E_{0+}E^*_{0-} \propto I\ J_0(\delta_R)\ \cancel{J_0(\delta_S)}^{1}$. Thus increasing modulation depth, δ_R, to the limit of $J_0(\delta_R) \to 0$, effectively reduces the correlation between the ports, thereby altering the principal eigenvector from the initial beamsplitting angle θ_0 and decreasing the difference between the eigenvalues. This in turn reduces the two-beam coupling angle β.

As before, the output fieldvectors are obtained by a rotation matrix equation (3.3) with a rotation angle

$$\beta = \left\{\arctan\left[e^{-\Lambda L}\tan(\theta_0)\right] - \theta_0\right\}\Big|_{\Lambda L>10} \to \theta_0 \tag{3.8}$$

which is a function of the eigenvalues where L is the thickness of the grating and $\Lambda = (\lambda_1 - \lambda_2)/2$ (or half the difference of the eigenvalues of equation (3.7)). To reiterate, for BaTiO3 with sufficiently high gain, equation (3.8) asymptotically goes to θ_0 (the principal eigenvector angle before coupling) for all but modulation depths that significantly deplete the carrier on one port, i.e. $J_0(\delta_R) \approx 0$.

The mixed signal, detected on the minus port, can then be expressed as

$$I_{sum/dif} = I\ J_1\left[\delta_S(t)\right]J_1\left(\delta_R\right)\sin\left(2\beta\right)\sin\left(2\theta_0\right)\cos\left[\left(\omega_S \pm \omega_M\right)t\right] \tag{3.9}$$

For small signal δ_S detection, the signal is maximized by splitting intensity equally between both initial fieldvectors E_{+0} and E_{-0}. Under ideal circumstances with sufficient grating thickness, this split results in a principal eigenvector and rotation angle of $\beta = 45^0$, as shown in figure 3-3. In light of our initial goal, we compare equations (3.5) to (3.9), ideally achieving a small signal sensitivity enhancement given by $J_1(\delta_S)J_1(\delta_R)/J_2(\delta_S) \approx 4J_1(\delta_R)/\delta_S$ over the un-modulated case. The maximized signal is therefore achieved for $\delta_R \approx 1.84 rad$.

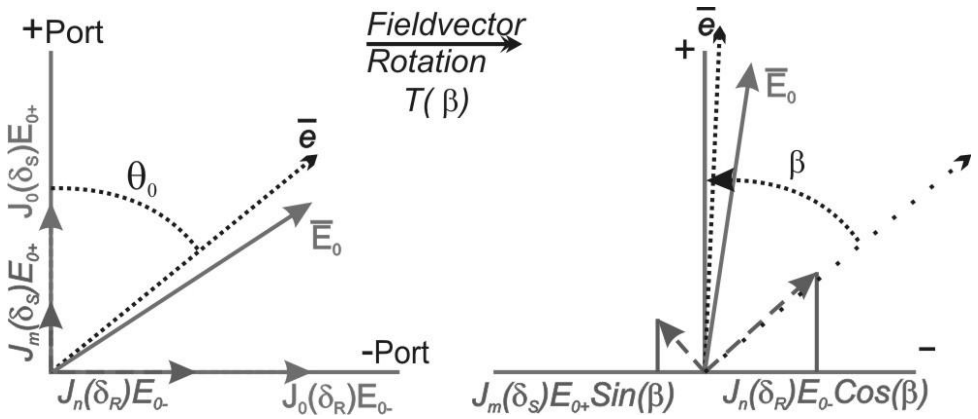

Figure 3-3: Evolution of fieldvectors with modulation enhanced sensitivity. The enhanced signal [equation (3.9)] arises from the |sum|^2 of the projection of the signal 1st harmonic and reference fieldvectors onto the (-) axis.

3.3 Experiment

To experimentally evaluate our model, we utilize the apparatus illustrated in figure 3-1. The test signal is a $\omega_S = 2\pi \times 5Hz$ phase modulation imposed onto the piezo-driven mirror. The laser is a 532 nm Coherent Compass 315M whose power is split 72:28 in the interferometer corresponding to a beamsplitter angle of $\theta_0 = 58^0$ off the (+) port. We detect 2mW of optical power at the interferometer outputs. Empirically, the detected noise increases with modulation depth. Therefore, we reference modulate with amplitude $\delta_R = 0.1\lambda$, where the optimum signal-to-noise ratio (S/N) is achieved, rather

than at $\delta_R = 0.293\lambda$, where the signal is maximized. The theoretical shift in optimum S/N is illustrated in figure 3-4.

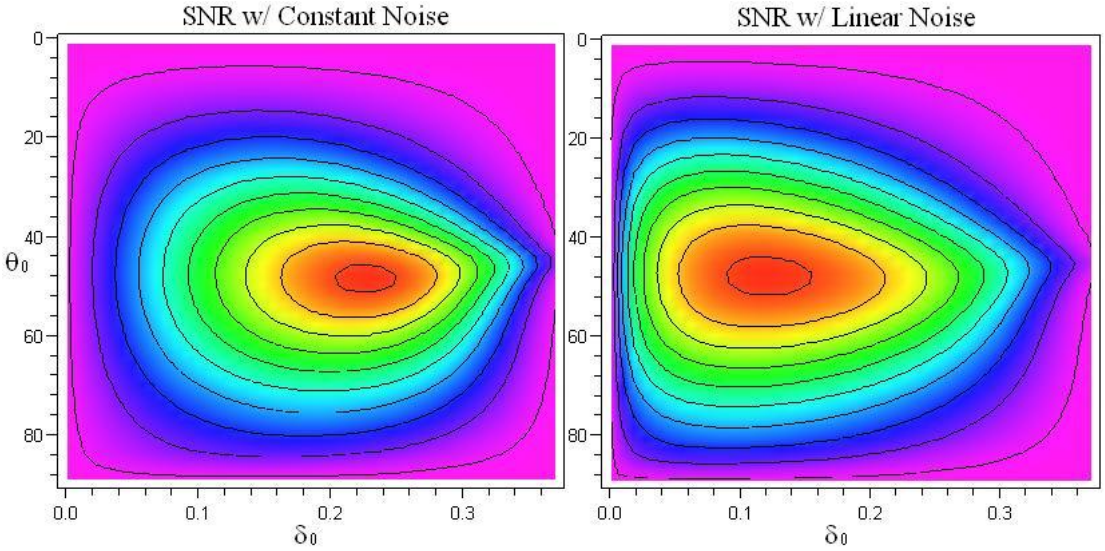

Figure 3-4: Normalized S/N detected at the output to the limit of maximum achievable coupling based on the input parameters of initial beamsplitting θ_0 and modulation depth δ. Note: noise that increases linearly with modulation depth pushes the optimum depth to lower values. This corresponds with experimental results in which a linearly increasing noise profile is observed and the optimum modulation depth is about 0.92.

Note in equation (3.6) we modulate the port that has more initial optical energy and orient the photorefractive to make this the (-) port regardless of which port the signal is on. The reason for this is that stronger modulations approaching $J_0(\delta_R) = 0$ decrease the vector rotation angle. If we begin with the eigenvector already closer to the (+) port, due to losses or other imperfections, we limit the rotation angle to $<45^0$. If however, the eigenvector is closer to the (-) port, then stronger modulations, necessary to maximize the signal, will decrease the coupling angle though 45^0, allowing us to balance these tradeoffs in signal strength to optimize sensitivity.

By varying the signal amplitude such that S/N = 1, we determine our sensitivity to be 180 fm/Hz$^{1/2}$. In contrast, the unenhanced 2nd harmonic detection sensitivity of the same 5 Hz signal is 260 pm/Hz$^{1/2}$. Thus we see an improvement of more than three orders of magnitude. For comparison, a theoretical signal whose second harmonic response [equation (3.5)] is limited by system noise to 260 pm/Hz$^{1/2}$, where 1/f noise at 5 Hz is 7.5 dB above the noise floor of the system (found experimentally),

should exhibit a maximum-response [equation (3.9)] improvement of $5.6 \times 290 \approx 1600$, also over 3 orders of magnitude improvement.

3.4 Discussion/Conclusion

We have demonstrated a mixing scheme using an additional phase modulation as a local oscillator combined with synchronous detection to improve a dynamic holographic interferometer's sensitivity. A vector representation was employed to aid in the visualization of this mixing process to allow us to gain insight into the optimum modulation parameters and adjust them as necessary for limitations in actual system implementation. Finally, experimental results were presented that are comparable to theoretical predictions.

It should again be noted that our modulation scheme will not improve the sensitivity limit for a system that is already shot-noise limited. Instead, our scheme is useful only in the case where system or electronic 1/f noise limits the detection of a low frequency signal, or where the second-harmonic response at any frequency is swamped by the electronic noise floor. Experimental adjustments in reference modulation depth may be necessary to adjust for system limitations including: noise behavior at stronger modulations, fixed beam splitting ratios, and photorefractive gain inherent to the crystal. In conclusion, by optimizing the amplitude of an applied reference modulation and the beam-splitting ratio, we demonstrate a simple means of improving sensitivity to a low-frequency–system-noise-dominated phase signal by roughly three orders of magnitude, using a diffusion-dominated holographic interferometer. For greater detail on the LOD the reader is referred to Appendix A and Appendix B.

Chapter 4 660nm Portable Prism System V4

4.1 Introduction

Having established the technology's ability to achieve a sensitive LOD, as presented in the previous two chapters, and establishing the anticipated LOD for ethanol, this chapter details several "proof of concept" modifications incorporated into a final portable prism system. The first modification uses a 660nm source since the photorefractive's response is slower at that wavelength. The slower response is preferred to adjust the photorefractive's high pass filter characteristics below the 5 Hz signal, where the previous system had the 3dB point at 5Hz. Next, the system needs to be portable, which is far easier to achieve with a diode laser system. Practical portability also requires that the detection electronics are shrunk from large expensive lock-in amplifiers to smaller demodulation boards, which we used for the first stage of demodulation in the previous system to good success. The last specification was that the system enable the use of an array of polymers. Since prism geometry would not allow a 2D array, while still satisfying the path length balancing between the interferometer arms, we decided to use a linear array of polymers. The following sections detail these changes as well as the challenges presented in their implementation.

As one final note before we present the last of the nose systems, there are 4 systems which may be referenced beyond this point. The first, built by my predecessor, uses an interferometer made of free-space optics [21]. The second is the second prism system we built (Prism V2). We exclude the first, as it performed comparably to my predecessor's system. PV2 is discussed little as it was an intermediate step to the final systems. The third and fourth are the final two respective systems Prism V3 and Prism V4, or respectively PV3 and PV4, where PV3 is the focus of most of this thesis. The prisms in both systems are designed for balanced path-length interferometry, though PV3's geometry accommodates 532nm and V4 works at 660nm.

4.2 Prism Modifications

One of the first modifications desired for the PV3 system was the elimination of all unnecessary TIR surfaces by reflective coating these surfaces with silver. The evaporated mirrors at the TIR interfaces reduce beam degradation due to dust or dirt collecting at these interfaces and eliminate the Fresnel-phase shifts due to index fluctuations in the surrounding air. The reflective silver coating was further coated with a SiO2 protective layer to both protect the silver layer and add thermal insulation over the thermally conductive silver further shielding the interface from localized temperature fluctuations. Figure 4-1 shows a drawing of the modified prism interferometer with two silvered interfaces.

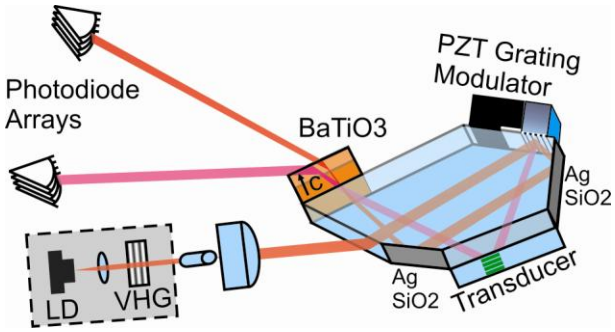

Figure 4-1: Prism interferometer of PV4, redesigned to accommodate a volume-holographic grating (VHG) stabilized laser diode (LD) at 660nm. Two TIR interfaces were mirrored with silver, and then coated with a protective SiO2 layer. The polymer array is aligned into the page, as noted by the stacked photodiodes. The only spherical lens is that which collimates the laser.

Next, the prism geometry needs adjustment to accommodate 660nm rather than the previous 532nm. Therefore, the coupling geometry out of the prism is shifted, ensuring as much power could be coupled out of the photorefractive as possible. Using figure 4-2, the steeper angled beam is set to roughly Brewster's angle for BaTiO3 to Air, and the other beam, being at a lower angle, suffers a minimal reflection (R<15%). In this way as much energy as possible is coupled out of both ports while minimizing the tilt sensitivity of the transmission amplitude.

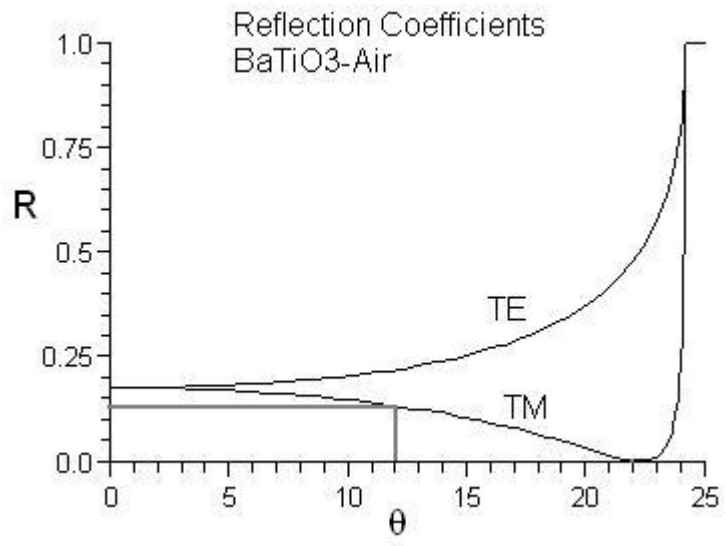

Figure 4-2: The reflection coefficients out of the BaTiO3 crystal for internal incident angles of 12^0 and 22^0 are approximately 15 and 0 % respectively for TM coupled light.

4.3 Transducer Array

Since the prism was designed to balance the path lengths of two beams in a plane, imaging a 3D transducer array would result in path-length imbalance everywhere except along the center column of polymers. Therefore, it makes sense to use a linear, rather than a planar, array of polymers, such that each polymer's interrogation beam is still path-length balanced with its interfering beam. To achieve this, a telescope in one dimension was made to expand the beam to a height of over 5mm, enabling a 1x4 polymer array to be reasonably fabricated onto a slide. Yet, somehow the polymer must be imaged onto the 1x4 photodiode array after coupling in the photorefractive.

As it is impractical to place imaging optics within the prism, the imaging of an array into the photorefractive, and subsequently onto the detector, is made more difficult. The use of imaging optics was implemented by a predecessor's tabletop design [21], resulting in an interferometer the size of a paperback book. By reducing the volume of the interferometer, the internal path length is reduced to 3cm and the distance from the polymer to the photorefractive is just over a centimeter. Using a lens just after the BaTiO3 crystal, we can image the polymer onto a diode array while still allowing the photorefractive to spatially process the individual patches. However, for a sufficiently collimated

beam, which is adjusted to be sufficient, imaging is unnecessary and so the lenses were removed from the final design.

Since the system will be limited by the number of transducers interrogated it would be prudent to determine a practical limit, of this system, to said number of polymers. The first and simplest limit may be imposed by the detector array. Using a Hamamatsu S8558 diode array, the individual PIN elements are 0.7 x 2 mm with a 0.1 mm gap between elements such that the center to center spacing is 0.8 mm. For the 11 mm beam height this means the maximum number of illuminated diodes, and therefore polymers, are 13. Again, as mentioned, the 660 nm system only uses 4 diodes per beam, as a proof of concept.

While one may be able to illuminate as many diodes and polymers as one can obtain to fit within the 11mm beam height, one must contend with crosstalk as the scaling limit. To address this we look at the Fresnel diffraction of a slit based on the polymer dimension. We take this limit since the beam height ($1/e^2$) of 11mm has a Raleigh range of over 150 meters and therefore the expansion of the beam is not an issue. The Fresnel diffraction equation of a slit is obtained from Saleh & Teich [43] and evaluated from the transducer array to the photorefractive (17mm separation) and the photodiodes (49mm separation from the transducer). The resulting cross-talk between polymers and detectors is given in table 4-1 below, from which we determine that polymer/diode sizes should not get below 0.2mm and at the current sizes there is less than 0.08 % crosstalk from one polymer onto an adjacent diode.

Polymer/Diode height	Distance	Fresnel #	Polymer-to-Adjacent-Diode Cross talk	Polymer-to-Diode detection
1 mm	17 mm	22	0.03 %	95.4 %
1 mm	49 mm	7	0.08 %	92.3 %
0.5 mm	17 mm	5.5	0.12 %	90.8 %
0.5 mm	49 mm	1.9	0.3 %	85.1 %
0.1 mm	17 mm	0.9	4.6 %	66.7 %
0.1 mm	49 mm	0.3	16 %	30.0 %

Table 4-1: Polymer-to-diode crosstalk and direct detection. From this table we see that for an array of 0.1mm polymers with no gap between, the signal on a diode will be just as strong as the sum of cross-talk from the two adjacent polymers. Therefore it is recommended that polymer dimension not be below 0.2mm. This of course assumes that there is no dead space between diodes.

4.4 Grating Modulator

The design modifications of the grating modulator were necessitated by the geometry changes associated with the prism and the wavelength change from 532 to 660nm. The incident angle had to be adjusted such that the 0^{th} and -1^{st} diffracted orders were path-length-balanced and still met in the photorefractive. The biggest modification we needed to perform on the modulator itself was to reduce the footprint to 8mm wide by 3mm long and 3.25mm thick, the critical reduction being the 3mm length. This reduction was necessary because geometrical modifications of the prism decreased the length of the face on which the modulator was mounted, necessitating the modulator shrink to accommodate. Additionally, we decided to increase width and height of the piezo used to a 3x3x2mm to reduce the possibility of the flexure expanding with some rotation component, where its expansion should be linear.

To improve the tolerances in manufacture, thereby improving the precision of performance, we had the flexure wire electro-discharge machined to yield the flexure shown in figure 4-3. The PZT is still press fit into place and driven with an external resonator tank. The diffraction grating dimensions were also increased in width to the full width of the flexure to ensure that the entire laser line is phase modulated.

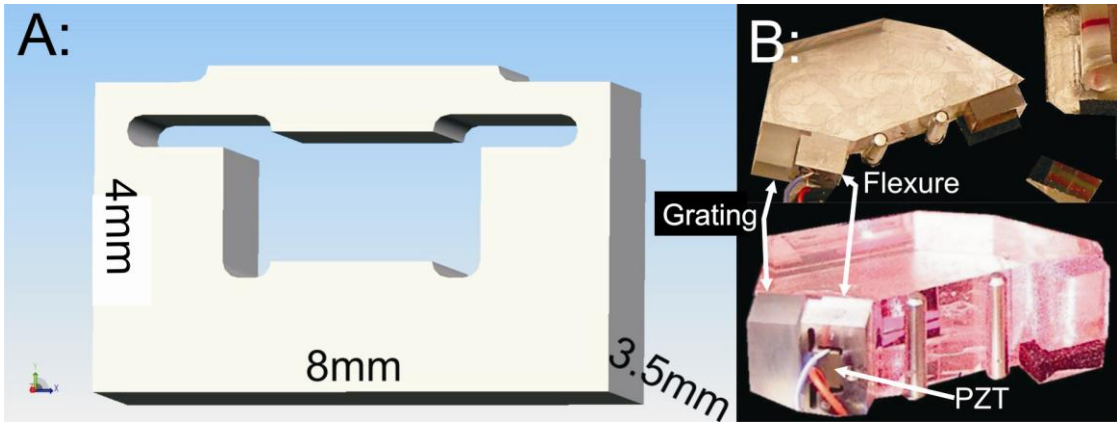

Figure 4-3: A: Piezo-electric transducer (PZT) flexure built to preload the PZT and translate the grating along the surface of the prism to which it is mounted. B: The PZT is press fit into the flexure and mounted with the direction of actuation along the surface of the prism. The rectangular slot for the PZT has recessed slits on the sides to prevent electrical shorting of the PZT electrodes through the stainless flexure.

4.5 660nm Laser Source

Since one of the goals is to show that the interferometer will work as a portable system, we investigated a laser diode source. There are other reasons for this change, primarily to slow the photorefractive response time, and reduce the size and cost of the system. This section highlights some design aspects of the laser as well as the beam dimensions and its stability.

The diode source is a Thorlabs DL7147-201, 60mW continuous wave, 660nm single mode laser that is not anti-reflection coated. The collimation optic is a Geltech 350390-B, 2.75mm focal length lens which collimates the laser to a 1.35x0.75mm diameter beam ($1/e^2$). The laser assembly is depicted in figure 4-4. Collimation is achieved with a 2.75mm focal length lens followed by a 1.5mm thick Ondax volume-holographic grating(VHG), with a center wavelength of 660nm, for external-cavity feedback. The collimation optic is mounted into an Invar sleeve which is adjusted for collimation and glued into place. The cavity is made of invar to reduce temperature dependant swelling. The VHG is mounted on a cored ruby hemisphere to reduce grating misalignment by thermal expansion, even though both the laser and the grating have their own thermo-electric cooling units and thermistor-temperature monitors.

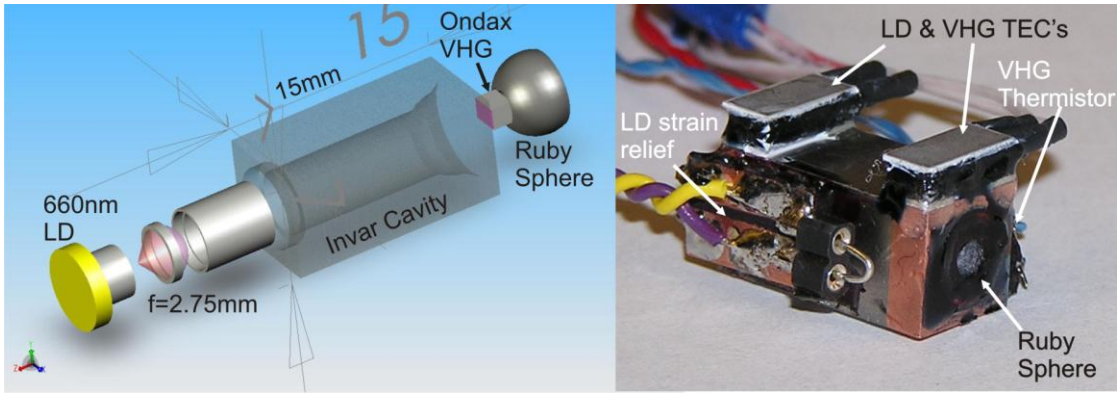

Figure 4-4: Volume holographic grating (VHG) stabilized diode laser.

The thermistors are imbedded into the copper ring that is press fit over the invar cavity to improve thermal conduction about the two ends helping to reduce the system to a one-dimensional temperature differential, again to reduce thermal misalignment. The LD and thermistors are strain relieved to PCB boards that were soldered to the invar. Additionally, the entire laser module is mounted within a second-stage-cooler housing to further aid the stability of the laser. The laser current servo was originally designed by J. Hall and T. Brown of JILA, and was modified to remove unnecessary modulation and monitoring components. Laser current fluctuations, read out by the servo's voltage monitor, are within 300 pA/s.

Next we estimate the beam waist from the collimation lens using the divergence angle $w_d = 2f\tan\theta_R$. After collimation, the beam enters a one-dimensional beam expanding telescope comprised of a 2mm cylindrical rod lens (f_{eff}~1.48mm) and an f=12.5mm cylindrical lens, giving an amplification factor of 8.4 on one axis. The effective focal length and *NA* of a perfectly spherical/cylindrical lens are approximated by

$$f_{eff} \approx \frac{nD}{4(n-1)}, \tag{4.1}$$

And

$$NA = \sin(\theta_{NA}) \approx \frac{2d(n-1)}{nD}, \tag{4.2}$$

where *D* is the diameter of the lens, *n* is the refractive index of the lens, and *d* is the clear aperture or diameter of the beam entering said lens. We summarize the results of horizontal and vertical collimation in the table below. We can estimate that, for a power of 10mW in the photorefractive and an area of 0.085cm^2, the intensity in the photorefractive is 0.1 W/cm^2. Having established the basic specifications for the laser, we now move on to its critical specifications for performing interferometry.

	FWHM	$\theta_R\left(1/e^2\right) = FWHM/2\cdot 0.59$	$w_D = 2f\tan\theta_R$ **F=2.75mm**	$w_{D_V} = w_{D_{V0}}\cdot 8.4$ **After Beam Expander**
Vert.	16^0	13.8^0	1.35mm	11.34mm
Horiz.	9^0	7.8^0	0.75mm	0.75mm

Table 4-2: Progression of beam angle and diameter of the 660nm laser through collimation and beam-shaping optics.

The laser's critical performance parameters are its amplitude and wavelength stability. Amplitude stability isn't as critical since the use of a Hobbs auto-balanced detector will reject common amplitude noise by 34dB. The 660nm laser showed an amplitude stability, measured directly with a photodiode, of approximately 0.01% while the 532nm laser's stability was within 0.005%, which was marginally larger than the instrument noise of the electronics used for this measurement. Accounting for the Hobbs subtraction the effective displacement noise is 800fm at 660nm and 300fm at 532nm.

Wavelength noise, on the other hand, is indistinguishable from the signal due to the photorefractive Bragg diffraction. It is therefore critical to have minimal wavelength noise, or achieve as narrow a linewidth as possible, to optimize the sensitivity. The Ondax VHG is spec'ed to feed back 20% of the incident power at the desired wavelength. The grating is 1.5mm thick, the center wavelength is reported to shift by 0.01nm/^{0}C, and the bandwidth is specified to FWHM=0.2nm or 137 GHz. Obviously, if this bandwidth were the linewidth of the laser, we wouldn't be able to detect nanometers. However, this does indicate that the grating does not have as high a *Q* as one might like for laser stabilization. At the time of laser fabrication we couldn't find a narrower linewidth VHG, so we used this grating hoping that sufficient stabilization of the temperature and current would yield an acceptably stable laser.

To test the wavelength stability two VHG stabilized lasers were built at 660nm, enabling a beat-note linewidth measurement. Tuning to find the beat note was accomplished using independent grating and laser temperature control coupled with LD current tuning. At its best, the linewidth was observed to be on the order of 100 kHz for 50ms time sweeps. However, the typical spectral stability is around 5-10 MHz in under 10 seconds. We note the limit of 10 seconds because the photorefractive has a time constant of a little over a second for a 0.1 W/cm^2, 660 nm beam.

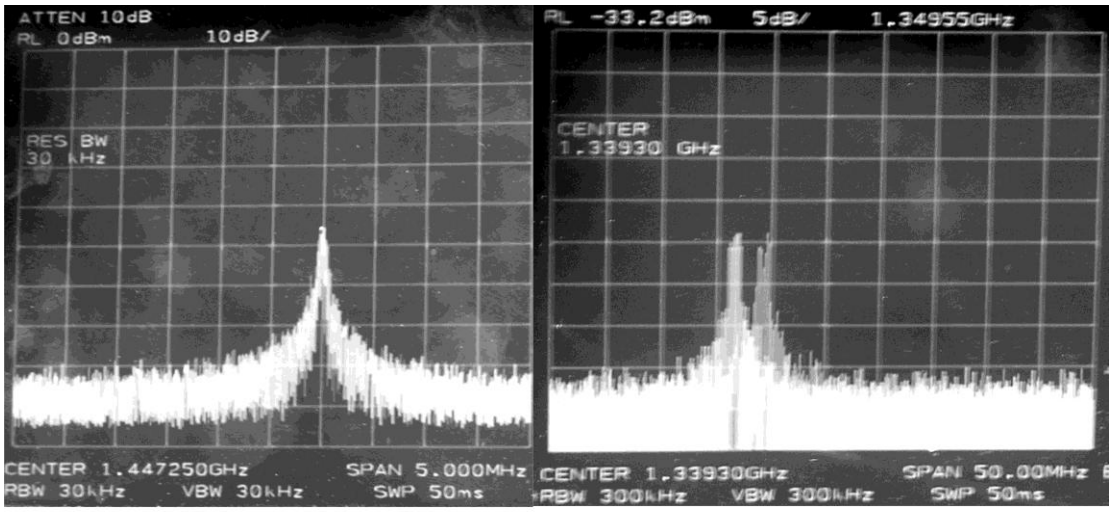

Figure 4-5: A: The linewidth of 660nm laser for a 50ms sweep is better than 100 kHz in certain sweet spots of operation while B: for longer sweep times, and without tuning the lasers to find a sweet spot, the stability of the frequency is under 10 MHz. Note, the above center wavelengths are different because the data was taken on two separate days requiring adjustments to reacquire the beat note.

This instability is believed to arise from the laser diode's lack of an anti-reflection (AR) coating and the current/temperature stability. Experiments monitoring both stabilities showed the current is stable to within 300pA while temperature is stable to within 1mK at the thermistor over several seconds. However, it is possible the diode laser exhibits temperature fluctuations faster than the thermal feedback loop can measure or compensate for, on the order of a second.

It was also observed when, using an optical spectrum analyzer, that mode hops were significant during tuning, indicating that a competition between the external cavity and the internal cavity could easily be contributing instabilities when trying to hold a single wavelength. Furthermore, we do not actually lock to a transition peak or a reference cavity, but rather we keep the linewidth

narrow by maintaining low current and temperature noise when servoed to arbitrarily fixed DC values. Therefore, it may be advantageous to use an external cavity reference or even a separate VHG to lock the laser to a specific wavelength.

4.6 Temperature Control

Since the gain spectrum and the LD cavity spacing are strongly affected by temperature, a TEC servo system controls the temperature of the LD and the secondary case built around the laser assembly. Additionally, the VHG is also temperature controlled, allowing independent manipulation of the feedback wavelength. In order to pump sufficient heat from the laser, the heat generated must be known to ensure the servo can handle the upper limit. The LD is rated to produce up to 60 mW of continuous optical power while drawing up to 120 mA at approximately 3 V. The total heat generated by the LD is roughly 300 mW of thermal power, which must be pumped from the laser to maintain a set temperature.

The temperature control of the laser diode, VHG, and secondary case are all servoed by modified variants of a JILA designed PID board designed by T. Brown. The laser and VHG servo are set at 19^0, with a 2^0 tuning range, and have a time constant of 1.2-1.5 seconds. The secondary case temperature is set to 22^0C and has a feedback time constant of about 2.2 seconds. With two stages of temperature stabilization, the temperature stability, as read from the thermistors, is under a milliKelvin-per-second.

4.7 Power Supply

The system can operate on a power supply range from +/- 7V to +/-15V, and is capable of supplying a peak value of 1.5 Amps and 1A continuous draw, due mainly to the laser, coolers, and oscillator. The laser was typically run at 90 mA while each cooler would draw around 200 mA. Since everything is designed to be battery operated, the TEC control boards are limited to 500 mA per TEC to reduce the total current draw and the laser is limited to 110 mA, just below its rated 120 mA limit.

Current limitation was necessary since it was found that current draws of an Ampere would cause the servo system to oscillate since the 1.5 amp supply could no longer satisfy the load.

The typical power-up procedure begins by activating the external cooling fan, the TEC's, and the reference oscillator. These components are powered first since they need time to stabilize and draw the largest currents at startup. When the TEC's have reached their target temperature the laser is powered up along with the rest of the detection and demodulation electronics. The power-up procedure is designed to ensure that the lasers are in minimal danger from servo malfunction associated with servo power up, and to ensure that the current draw never exceeds 1.5 Amps. The shutdown procedure is simply a reversal of the above.

4.8 Phase Discrimination

An additional feature of the arrayed detection is the ability to compare polymer response dynamics in parallel, enabling a unique signal to be extracted based on the combination of polymers. To achieve this side-by-side analysis the photodiode array was rewired such that pairs of photodiodes on the (-) port were subtracted immediately using auto-balanced detection (effectively subtracting the response of two separate polymers as a single channel), rather than the standard subtraction of the (+) and (-) port of a single polymer that was performed for all other measurements. This subtraction effectively turnes 4 polymer responses into two differential responses, where each pair had a single polymer (PNVP or CAB) subtracted from a polymer-less patch (bare glass) to help reduce the effects of noise at the TIR interface. Figure 4-6 demonstrates how 4 polymers, and their 4 corresponding photodiodes on the (-) port are converted into two parallel differential signals. Figure 4-7 then shows the X-Y trace of the oscilloscope demonstrating the potential for dynamic discrimination based on relative polymer response, phase delay associated with the vapor exposure time and the transducer response lag, and differential harmonic responses with their associated phase delays.

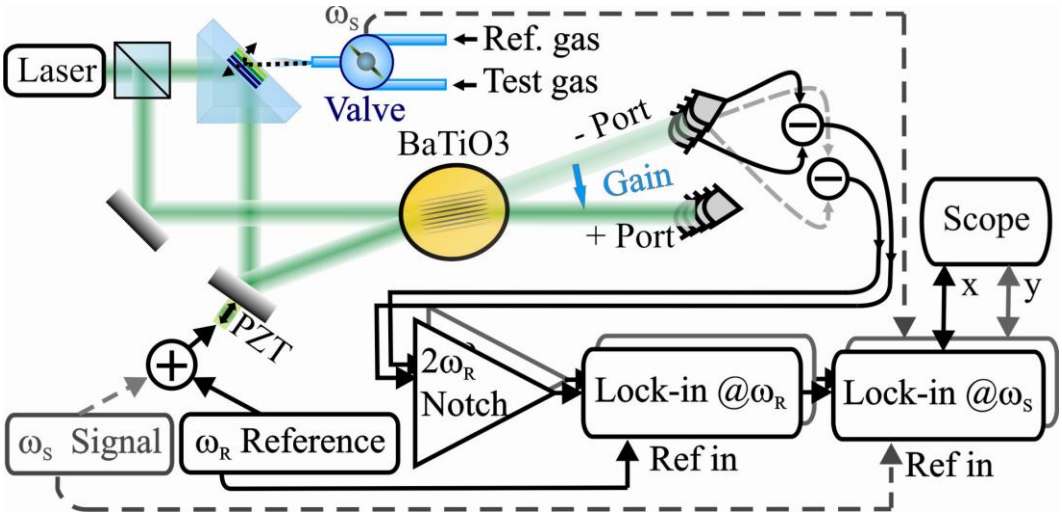

Figure 4-6: Schematic of a 4 channel system modified in parallel to provide two traces for an X-Y oscilloscope trace. The photocurrent response corresponding to 2 polymers are subtracted from 2 references (no polymer transducer) yielding two differential signals. Dynamics are compared based on phase, amplitude response, and harmonic response between the two differential signals.

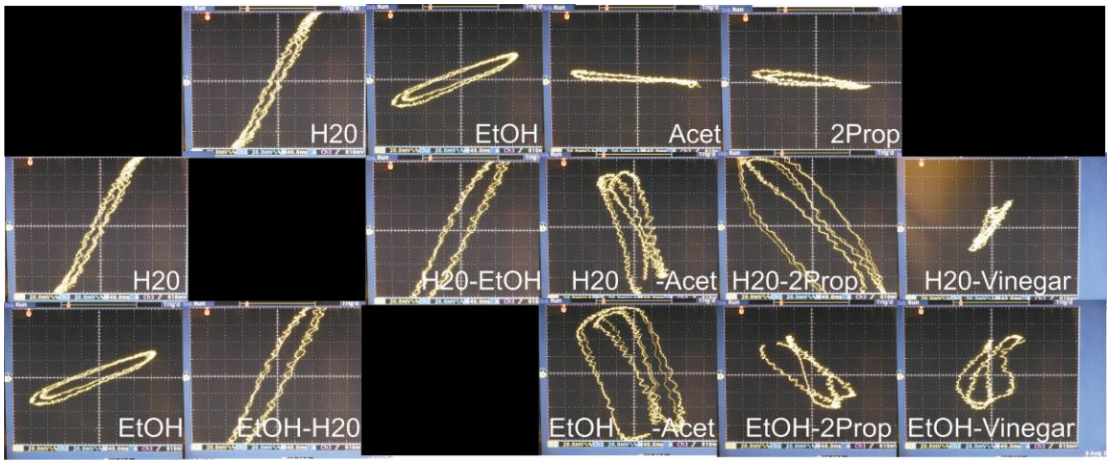

Figure 4-7: X-Y scope traces showing selectivity arising from cross-dynamics between two polymers (CAB and PNVP) when exposed to low 100's ppm levels of the indicated analytes. The comparison data is expressed as the analyte in the top-most row minus the left-most column. The high frequency wiggle is the result of a 240 Hz line noise that could often be seen on a single channel. The missing vinegar (acetic acid and water mixture) response looked almost identical to that of water.

In the X-Y traces above, angle is determined by relative transducer-response amplitude where equal amplitude yields a 45^0 oriented figure and opposite response sign yields a -45^0 orientation. Phase delay determines the area enclosed within the loop where $\pi/2$ should yield the largest enclosed area, for two fundamental sine waves. Folds or crossings of the trace, as observed with acetone and

isopropanol, can be due to larger response harmonics where the number of nodes is usually associated with the order of the harmonic. In the case of H20 the symmetric "hook" like ends arises due to third order harmonics coupled with a slight phase delay.

Finally, divergence between two overlapping traces, just noticeable in the bottom left of the ethanol curve and significantly noticeable with ethanol minus either acetone or isopropanol, is most likely associated with the asymmetry of the vapor modulator and asymmetric sorption/desorption dynamics that occur when species A is sorbing while B is present and desorbing, vs. when B is sorbing and A is present and desorbing. This phenomena has been observed when switching between analytes where a unique though short lived response is observed shortly after the sampled analyte has changed. This also explains why the patterns diverge primarily only when comparing two different analytes in the above traces. From these figure we see that the system has great potential in detecting and identifying a large variety of species based on the dynamics associated with the transducer-analyte interaction and exposure-measurement geometries, assuming significant dynamics are present at ppm-to-ppb levels.

4.9 Conclusion

Unfortunately, several tradeoffs in design resulted in decreased sensitivity, namely, the fabrication of a VHG stabilized laser, lacking the years of iterative design improvements that benefited the prism design. The displacement LOD of this systems is 1.7 pm/$Hz^{1/2}$ while the chemical sensitivity, S/N=1, is 14 ppm/$Hz^{1/2}$, owing primarily to the laser noise for the loss in sensitivity compared to PV3. An example LOD trace is shown in figure 4-8.

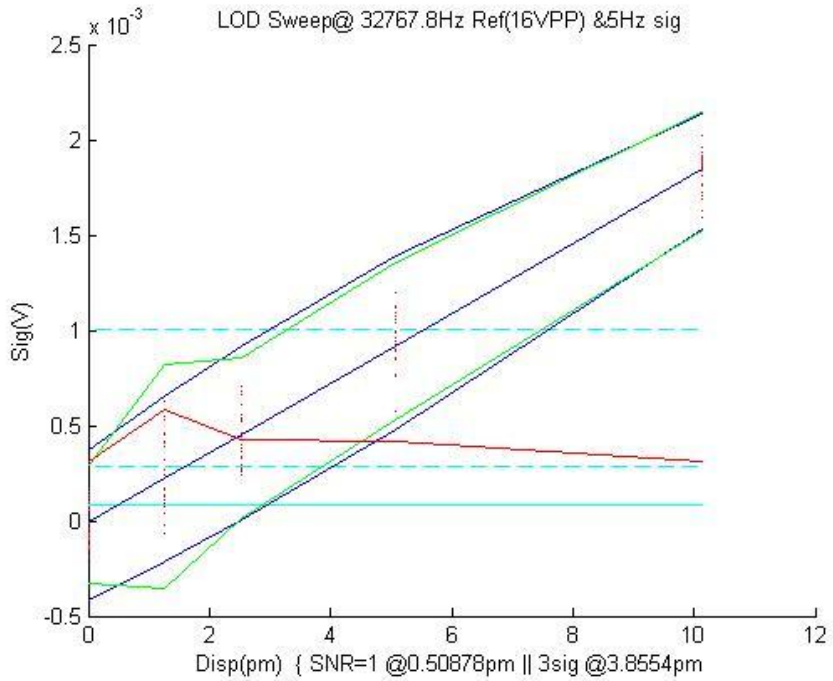

Figure 4-8: A displacement LOD data set of the 660 nm system with a SNR=1 LOD at 500 fm and a 3-sigma LOD at 3.9 pm with a 5 second integration time. The effective LOD for unity bandwidth and S/N=1 is 1.5pm/$Hz^{1/2}$

While this system is indeed portable in theory, time limitations were prohibitive to fabricating a housing to render the system completely autonomous. Essentially, with a stack of cells to accommodate the +/- 7V at 1.5A and a few panel voltage meters, the entire system would fit in a small briefcase of dimensions 30x30x10cm. If further design improvement were researched, especially in wavelength stabilization for a portable, low-volume device, this technology holds great promise as a portable, adaptable, arrayed-transducer–point-detection system. Figure 4-9 shows the compactness of the entire optical system.

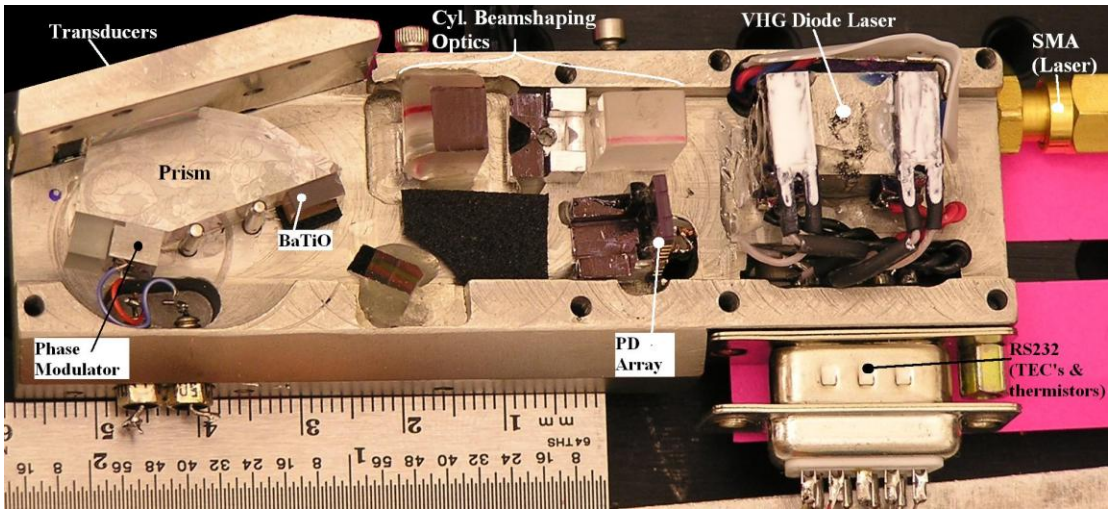

Figure 4-9: The assembled 660nm optical system mounted on an Invar baseplate. The RS232 connector allows 8 wires for the 2 TEC's and their respective thermistors, and the SMA connector provides power to the diode laser. The first cylindrical lens focuses in the plane of the page at the photorefractive wile the second and third cylindrical lenses act as a telescope to expand the beam out of the page to match the transducer array and the photodiode (PD) array.

Chapter 5 Limitations and Implementations

5.1 Introduction

This chapter covers the various fundamental and practical limitations of the optical nose systems and our chosen means of implementation. Discussions will present the reasoning behind design decisions and the necessary tradeoffs. Ultimately, it will be shown that for the reference system (Prism V3) we are potentially limited by a few noise sources, while the portable system (Prism V4) is fundamentally limited by the stability of the laser we built and the detector. For the reader's convenience we skip to the punch line and summarize the noise budget in Appendix C.

It may seem incomplete to not absolutely state a definite noise limitation, but this is done since at each point when a dominant source has been identified, steps have been taken to eliminate it. Therefore, at present, the system is limited by a combination of nearly indistinguishable noise sources. At this point it is impractical to mitigate the dominant noise source due to resource restrictions and since lowering sensitivity to one noise source risks raising another. It is reiterated that the intent of this chapter is more to aid in the understanding of the design decisions and the current dominant sources to aid in any continuation of this work.

5.2 Subtracting Optical Noise

Noise on the beams comes about from several potential sources. The laser itself may exhibit frequency and amplitude noise associated with temperature or current servoing. Vibrations in the cavity may also correlate with laser noise depending on cavity design. Though there are various sources of noise many of them can be attenuated (amplitude noise only) using an auto-balancing noise-subtraction circuit. Other than making the interferometer small, entirely within a slab of BK7, and closing the system in a box little else can be done to further remove environmental fluctuations that aren't common to both interferometer beams.

The auto-balancing subtraction circuit developed by P.C.D. Hobbs et al [22] has been well received for shot noise limited detection of optical signals, where common mode noise could be

subtracted from a signal beam and a reference (usually a tap earlier in the system). However, there are some limitations to this circuit design, especially in the optical setup used here. Some of the limitations have been worked around, but not without cost.

The auto-balancing subtraction circuit developed by Hobbs, and henceforth referred to as the Hobbs circuit, effectively subtracts the current of a "reference beam" from that of a "signal beam". Each beam is detected by its own photodiode such that the DC level and any common-mode amplitude noise, or signal, shunts past the gain stage and ideally never see amplification [22]. This subtraction process is achieved using a current splitter, comprised of a matched transistor pair, which splits the larger current from the reference diode until the DC current matches that of the signal diode. The quality of subtraction depends on the transistor gain and impedance matching, or the effective linearity of the current splitting, over the intended operation frequencies. We will discuss subtraction performance in greater detail after we motivate some necessary modifications to this circuit for our needs.

We take a moment to consider that any signal or noise source outside of the interferometer, especially those that only affect the laser amplitude, are almost inconsequential, as they will be common mode to all beams detected and therefore can be rejected with comparison circuitry. Conversely, any noise or signal imparted onto only one beam will receive no subtraction from the Hobbs circuit. One caveat is that if either signal or noise exists solely on the reference beam, their detected amplitude is attenuated proportionately to the photocurrent attenuation factor of the Hobbs circuit, to the point necessary to balance the DC photocurrents.

Amplitude noise sources occurring at beamsplitting or recombination due to scatter are subtracted as they are common, while any phase noise occurring inside the interferometer actually gets amplified in the subtraction process, especially if the phase noise source directly affects the beamsplitter or photorefractive recombiner. To illustrate the phase response we observe how a simple phase signal $\phi(t)$ is converted into an intensity signal out of the photorefractive.

$$E_{+0} = \frac{E_0}{\sqrt{2}} exp\{i[wt + \phi_\lambda(t)]\} \tag{5.1}$$

$$E_{-0} = \frac{E_0}{\sqrt{2}} exp[iwt] \tag{5.2}$$

$$\begin{bmatrix} E_+ \\ E_- \end{bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} E_{+0} \\ E_{-0} \end{bmatrix} \tag{5.3}$$

$$\begin{bmatrix} I_+ \\ I_- \end{bmatrix} = \frac{I_0}{2} \begin{bmatrix} 1+\sin(\phi_\lambda(t)) \\ 1-\sin(\phi_\lambda(t)) \end{bmatrix}_{\beta=45^0} \tag{5.4}$$

The photorefractive rotation angle is arbitrarily set to $\beta = 45^0$ for simplicity. If we now subtract the two ports, weighting the (+) port which we assume has equal or greater intensity we get a signal strength that goes as

$$I_+ \frac{|I_-|}{|I_+|} - I_- = \frac{I_0}{2}\left(\frac{|I_-|}{|I_+|}+1\right)\sin(\phi_\lambda(t)). \tag{5.5}$$

The current amplification will be between 1 and 2, 2 being when both beams are equal in amplitude and no attenuation of the reference is necessary and 1 being if one beam is a bright fringe and the second is the corresponding dark fringe. While this means that a phase signal will indeed get a gain of up to 2, noise within the interferometer will also receive such gain.

There now exists a decision to be made for detection. One can either detect the (-) port beam and use a tapped beam as a reference, suffering from any noise picked up after the tap. Or one could subtract the (-) port from the (+) and suffer up to twice the noise generated within the interferometer. One might pick the first option as it doesn't multiply noise sources, though with careful design the compact prism interferometer reduces noise due to mounted components and air convection in the interferometer. The first option also takes advantage of the gain to signal under the assumption that we aren't shot noise limited. From this point forward the (+) and (-) ports will also correspond to the reference and signal ports both optically and electronically.

The basic Hobbs circuit as shown in figure 5-1 A, and works well when a single independent diode can be committed to each beam. However, our situation calls for an array of diodes, as we wish to interrogate an array of polymers with a single expanded beam. An array presents a technical challenge in that most photodiode housing is bulky and prevents the photosensitive elements from being placed adjacent to one another with little wasted area in between. There exists the option to buy individual elements, unpackaged, though the technical challenge in wire bonding and properly mounting and shielding these was an effort we decided to avoid. We chose instead to use prepackaged

diode arrays (Hamamatsu S8558) for their robust and tightly spaced packing though nearly all manufacturers produces such arrays with common anodes, or more often common cathodes.

Though an array of common cathode diodes allows us to build an array of Hobbs circuits, we encounter the difficulty that the reference diode, that has its photocurrent split, also has its cathode connected to the current splitter, which for arrays of such circuits will not operate properly. We could let the circuit run using a single diode for the reference beam and divide its current among all of the circuits, though precision subtraction would be impossible since this would subtract the average of the polymer responses from each polymer. Additionally, this would also see the potential of electronic crosstalk by the auto-balancing feedback portions of the circuits, further decreasing noise subtraction.

Another solution includes a voltage subtraction circuit developed by Lindsay et. al [23]. Their circuit converts the photocurrent in a transimpedance amplifier and then, using four additional op-amps and a MOSFET, performs subtraction of the voltages. This design is compatible with common cathode diodes, though the system complexity is substantially increased. This circuit is reported to subtract to within a factor of 20 off shot noise compared to Hobbs reported 3dB within shot noise. We decided to avoid this design because the increased component complexity can give rise to various distortions if paralleled components aren't well matched, since each component added has its own transfer function and associated delays and noise contributions. BJT's, on the other hand, are often substantially quieter than op-amps and will subtract out to their gain bandwidth regardless of the loop feedback bandwidth. Additionally, when paralleling these circuits, we wish to minimize the component count and footprint, which is easier to do with BJT's.

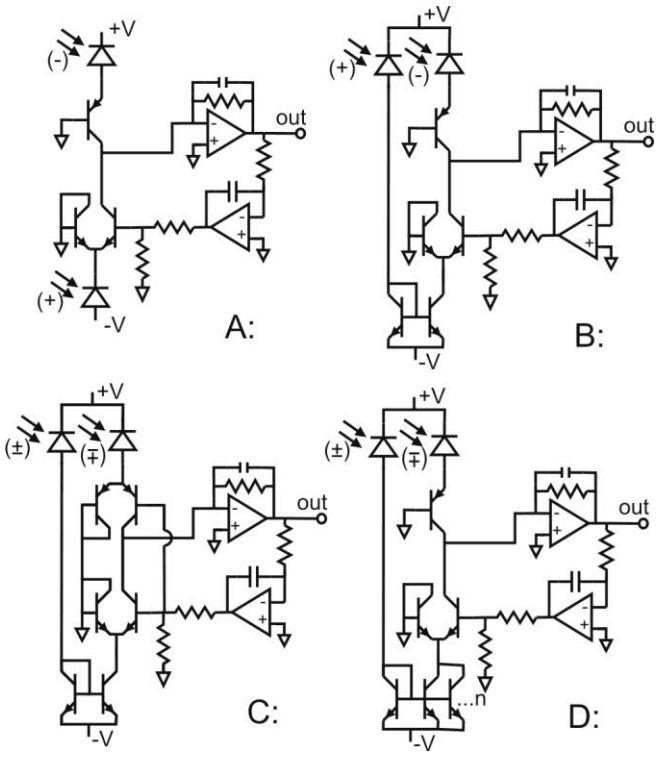

Figure 5-1: A: Hobbs Auto-balancing subtraction circuit [22]. B: With current mirror at the reference diode to allow for common cathode photodiode arrays to be used for arrays of subtraction circuits. C: With a mirror and PNP matched pair divider to enable sign-indiscriminate subtraction. D: Additional mirrors added to amplify the reference current for subsequent subtraction such that the signal current can be up to *n* times stronger than the reference circuit as a form of sign indiscriminate subtraction.

To work around the difficulty associated with a common-cathode diode array used in an array of Hobbs circuits, a current mirror is used. The mirror replicates the signal generated from the reference diode in such a way that the signal from a single diode can be properly isolated and subtracted from its counterpart, despite the use of a common cathode diode. The cost of using a mirror is degraded subtraction based on the precision of the replicated signal. Effectively, the degradation of subtraction in this case is at least twice that of the original Hobbs circuit over the ideal limit as there are now two matched pair transistors.

There is one final modification of the Hobbs circuit that was considered. One severe limitation of this circuit occurs in the regime where modulation depth is large enough to change the two beam coupling such that the beam that originally saw gain, and therefore had a greater optical intensity, is now the weaker of the two beams. This is a problem if both outputs of the interferometer were used for detection since when the ratio between beam intensities crosses unity, the Hobbs circuits

in figure 5-1 A and B will fail to subtract properly and may rail if the divider on the feedback loop is not properly tuned.

To combat this, we considered the two remaining modifications (C and D) which allow for the ratio to cross unity while continuing subtracting. Modification D works by adding 1 or more additional transistor mirrors, thereby multiplying the current from the reference diode by the number of mirrors used. It is highly recommended that the number of transistors used for the mirror does not exceed the number of well matched transistors one can buy/fabricate in a single package as mismatch will severely degrade noise subtraction. Also, this circuit can only subtract ratios (signal/reference) that range from 0 to n, where n is the number of mirrors used.

Modification C is a true sign-indiscriminate auto-balancing circuit that will subtract regardless of the beam intensity ratio using a PNP MAT03 divider on the signal diode. However, not only does the Hobbs circuit perform better if the current dividers are quality matched pairs $\Delta h_f < 2\%$ in each package, but also the matching of the PNP to NPN divider gain and impedance is necessary to improve the subtraction of common, high-frequency noise.

To show the quality of subtraction from these circuits, a series of experiments was performed to evaluate the noise suppression over the frequencies of interest (a few kilohertz to a few tens of kilohertz). Since we desire to see the effect of our modifications on both signal and noise, we needed the ability to generate optical signals with strong correlation to act as noise and negligible or anti-correlation to act as a signal. For this we used a pair of LED's to illuminate the photodiodes as in figure 5-2.

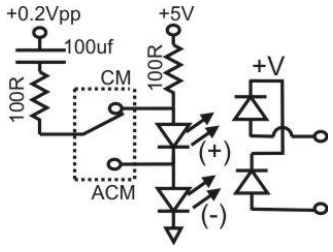

Figure 5-2: Basic test circuit to asses the modified Hobbs circuits. A 0.2Vpp sin wave is swept from 1 kHz to 50 kHz while the spectrum analyzer records the peak output from the subtraction circuit in question. With the switch in the common mode (CM) position the current fluctuations are common to both LED's simulating noise for assessing subtraction. With the switch in ACM, or anti-common mode, an increasing voltage or current corresponds to a decrease in the second diode, thus simulating the photorefractive output approximating the detected signal.

We use this test setup since we have need of an optical system that will have no extraneous phase or frequency noise associated with our photorefractive system. Though LED's are weak compared to a laser, the primary noise sources should be uncorrelated diode noise and whatever we intentionally drive the LED with. Super-bright red LED's are capable of a few mW of optical power and thus, can simulate fairly accurately the laser power levels (~2 mW per diode). The LED's are driven such that a switch controls whether the current, and thus optical signal, to be measured is noise (common to both photodiodes) or a signal (anti-common to both diodes).

First, we tested the signal response of each Hobbs variant with a 0.2 Vpp sin wave swept from 1 kHz to 50 kHz. Then we repeated the experiment under identical conditions for each circuit variant, except that we tested each circuit's ability to reject a signal common to both photodiodes. Finally, the optical setup in figure 5-2, with LED's and photodiodes, was transplanted to the demodulation board used for the final system.

Data shown in figure 5-3 indicates that use of a current mirror (modification B) has little effect on the signal response, though an array of mirrors does affect the noise suppression (figure 5-4) fairly significantly. The loss of suppression is likely due to the array not being a single quad-matched package and so higher frequency subtraction suffers. The PNP divider, however, had a substantial degradation affect on both signal and noise performance. Again, I believe this was due to the inability to match the PNP to the NPN pairs, but I did not investigate further as this modification was noncritical. The only time we would cross ratio=1 would be for calibration purposes, for which we can simply place an attenuator in front of the signal photodiode and remove the attenuator after calibration.

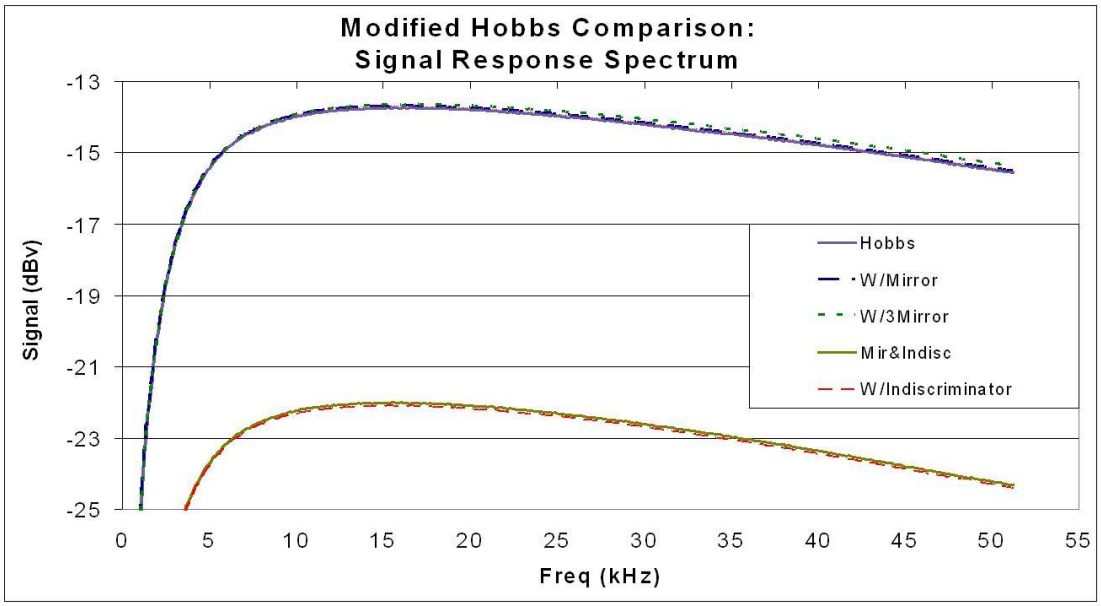

Figure 5-3 Spectrum of signal response to a pair of LEDs. Notice the use of a current mirror {MAT02 or LMH394CH} had little bearing on the response while the PNP divider caused a >8 dB drop in signal.

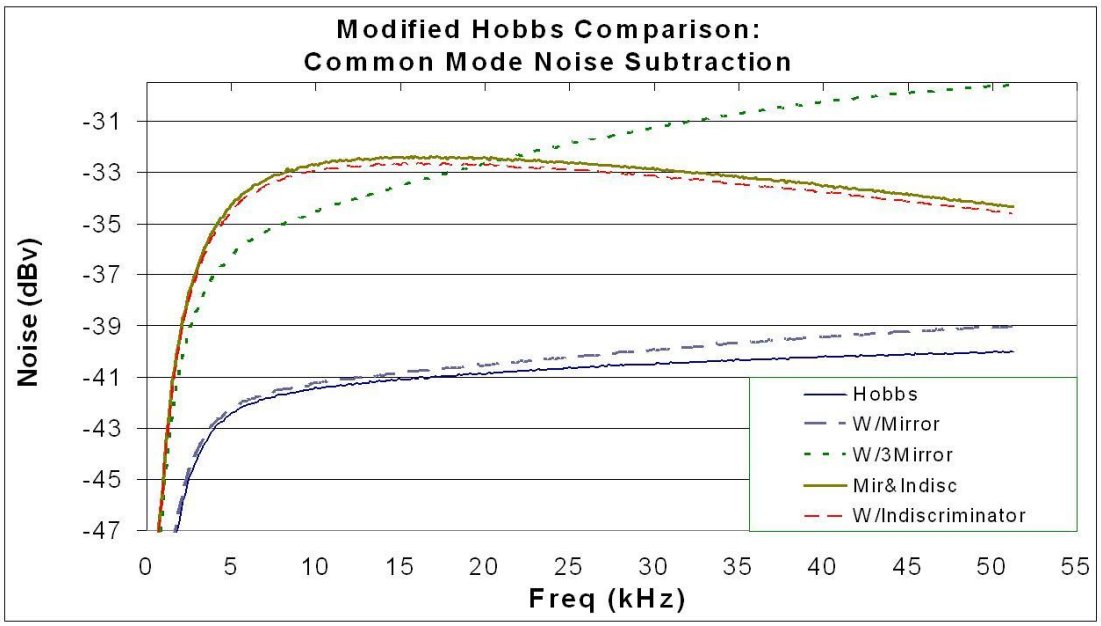

Figure 5-4: Spectrum of noise suppression for the same current on each LED illuminating its respective photodiode. The single mirror (n=1) made of a single matched pair barely affected the noise suppression <1 dB while the trio of mirrors n=3 caused more than 6 dB degradation due to the additional mirrors not being matched to the first two. Of particular interest is that the PNP divider substantially degraded low frequency noise suppression. The key point here is that use of a well-matched current mirror hardly degrades the S/N over the original Hobbs circuit, while using multiple mirrors or a PNP current divider to enable subtraction regardless of beam ratio will further degrade S/N.

Concluding from this data that a current mirror causes negligible loss in performance over the original Hobbs design, we implement modification B on a detector board. We then tested its performance, changing only between the accepted mainstay, the MAT02's and the smaller/cheaper DMMT5551 matched pairs for current division and mirroring. Figure 5-5 shows the results, indicating that, for the same optical setup under identical conditions, the board performs comparably. We notice that the signal to noise has actually improved and suspect the improvement is due to the demodulation board being a laid-out surface mount circuit while the test bed for the 4 Hobbs variants above was a rat's nest with many potential antennae. Additionally, the only other difference is the original test bed used OPA355 opamps while the newer demodulation board uses MAX412's to perform the gain and integrative feedback. However, it is unlikely that these are responsible since the Hobbs design doesn't depend on the bandwidth of the op amps [18].

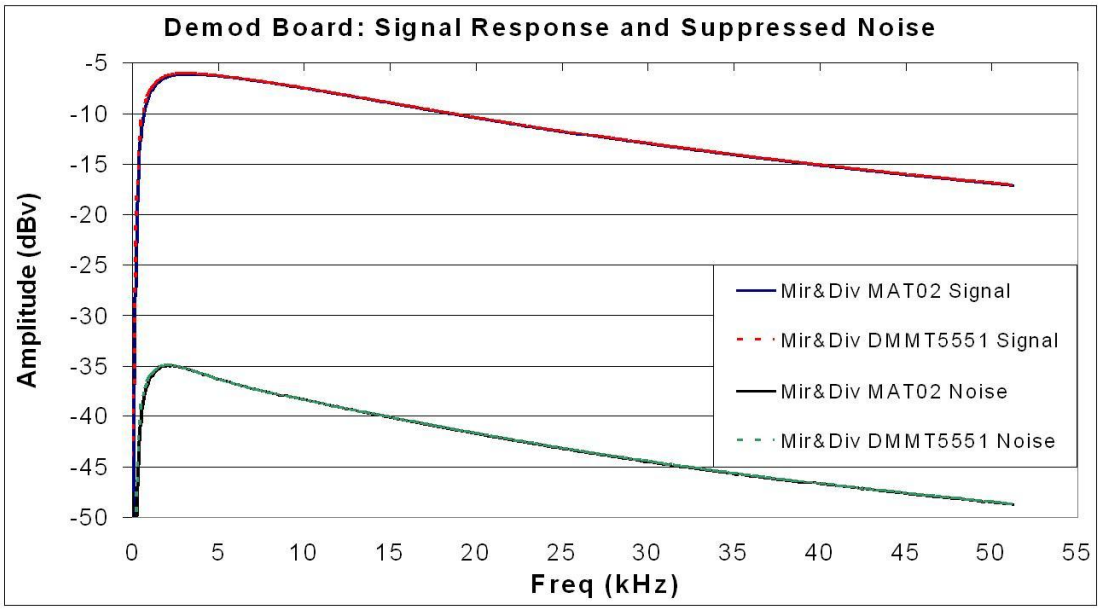

Figure 5-5: A comparison in subtraction signal response of the demodulation board used on the final system comparing the precision MAT02's vs. the cheaper surface mount DMMT5551 matched pair BJT's. A switch enabled toggling between the pairs of transistors that performed division and mirroring as per Figure 5-1 B. Little difference is seen with this weak signal on top of a large DC background as both showed subtraction of 32dB compared to the signal response. The optical and diode setup was identical to figure 5-2, verifying that S/N degradation using the DMMT5551 matched pairs over the typical MAT02's for this application is negligible.

The conclusion from this set of experiments is that it is an acceptable loss to use a current mirror to enable the use of a common-cathode photodiode array. The demodulation board behaves comparably to the detector board used for the V3 table top prism system. Since the Hobbs circuit's primary purpose is a photo detector that removes common mode laser noise, we leave a full circuit noise analysis to the next section, which deals with the entire post detection electronics.

5.3 "Leaked" Optical Noise

It has been reported that the grating of a photorefractive may be off, from the ideal 90^0 with respect to the interference pattern, by several degrees [24], which would enable a fundamental harmonic to effectively "leak" through as a signal. If such a leak occurred the amplitude noise tails at 5Hz, $V_{\omega_R+\omega_S}/V_{\omega_R} = 3e^{-6}$, on the reference modulation would be detected as noise at the sum or difference frequency. This phase shift can have several sources. If the wavelength of the laser drifts over time, there will be a phase lag due to the response time of the photorefractive, though this is effectively a dispersion effect. There may also be an external field or a bulk photovoltaic field [24,25] that causes a phase shift. Regardless of the source we proceed to estimate the effect based on measurements taken.

Using this spectral purity in conjunction with a known fundamental leak, we can estimate the noise level at the sum/difference frequency.

$$\frac{V_{\omega_R}}{V_{2\omega_R}} \cdot \frac{V_{\omega_R+\omega_S}}{V_{\omega_R}} J_2(\delta_R) = 10^{-(20+110)/20} J_2(\delta_R) = J_1(\delta_R) J_1(\delta_S) \tag{5.6}$$

where the ratio $V_{\omega_R}/V_{2\omega_R}$ describes any fundamental present at the output of two-beam coupling with respect to the second harmonic. Since the S/N is seen to suffer with increasing reference modulation depth past a point, the modulation depth is set to $\delta_R \le 0.1\lambda$, the value that yields the largest S/N.

Measurements performed to detect the fundamental frequency in the absence of a signal showed a fundamental amplitude modulation, without two-beam coupling, that was 68 dB below the amplitude of the second harmonic (with two-beam coupling). Any amplitude noise at the fundamental, before the photorefractive, must come from the phase modulator, indicating that there is some scatter

source on the grating (dust and damage) or tilt with modulation. This noise level is insignificant since, as amplitude noise, the detector will attenuate the noise by an additional 34dB.

Measurements performed after the photorefractive two-beam coupling resulted in a detected fundamental 20dB below the same second harmonic as above, indicating a fundamental "leak" due to the photorefractive. Assuming the cause is an imperfectly shifted phase grating, we can estimate the phase shift of the grating based on the measured first harmonic.

$$J_1(\delta_R)J_1(\delta_S) = \sin(2\cdot\Delta\phi_{PR})\cdot\left(J_1\left(\delta_R\frac{V_{\omega_R+\omega_S}}{V_{\omega_R}}\right)\right)J_0(\delta_S) \tag{5.7}$$

The phase shift $\Delta\phi_{PR}$ of the grating is estimated to be around ½ of a degree for a 180fm LOD. However, the cause of the leak is ultimately unknown and not of critical importance since the modulation depth is adjusted such that said leak is codominant with another noise source. The question remains: what is the other dominant noise source? To answer this we continue to discuss additional noise sources and efforts made to mitigate their contributions.

5.4 Photorefractive Fanning Noise

Another potential source of noise is the photorefractive fanning and light-scatter fluctuations. In the photorefractive there will several sources of scatter including surface roughness, defects, and inhomogeneities in addition to photo-electron shot and lattice vibrations/thermal fluctuations in the space-charge fields [26, 27]. Fanning comes about by volume scatter in the photorefractive combined with two-beam coupling gain, such that scatter in a particular cone of directions gets energy transferred from the pump beam to the cone of scattered beams. This fan of beams is the accumulation of many small gratings written between individual scatter sites and the pump beam, where the bright center of the fan notes the ideal alignment for the second beam in a two-beam coupling setup to attain maximum coupling.

In many cases the scatter from the incident beam will give rise to fanning with scatter rates from 10^{-4}/mm [27] for moderate gain to 10^{-9}/mm [28] for gains exceeding $\Gamma L \geq 5$. However, once these gratings are formed, fluctuations in scatter will not see gain unless the timescales of the

fluctuations are sufficiently slow for the photorefractive to adapt to them. Therefore, fluctuations in scatter only contribute to noise directly as scatter detected by the detectors. To estimate a worst case, if we assume that the scattered light making it to the detector acted entirely as an amplitude noise, we could estimate the limit of detection as below.

$$\left(1\cdot 10^{-4}/mm\right)\cdot L\frac{A}{4\pi\cdot R^2}\approx J_1\left(2\pi\cdot 0.1\right)\cdot J_1\left(2\pi\cdot\frac{dx}{\lambda}\right)\sin\left(2\theta_0\right)\sin\left(2\beta\right) \tag{5.8}$$

Where A=2mm x 2mm is the area of the photodiode at a distance R=3cm from the photorefractive. The interaction length L in the photorefractive is under 3mm, the initial beamsplitting angle $\theta_0 = 58^0$, the wavelength for PV3 is 532nm, and the rotation angle β is assumed to be roughly $\pi/4$. From this equation the worst case noise estimate *dx*, based on a scatter rate of 10^{-4}/mm, is roughly 67 am, well below shot noise.

We now consider in a two-beam coupling experiment, that the small gratings written from scatter must compete with the grating written between the signal and pump beam. This competition can be observed by blocking the signal beam and letting the fanning beam grow. Once the fanning is noticeable, we allow a signal beam to couple through the photorefractive along the path of the fanning, at which point one will observe that the signal beam sucks energy from the fanning. More precisely, the grating written by the signal and pump beam competes with the weaker fanning gratings diminishing the fanning gain in favor of two-beam coupling gain. This effectively turns the fanning noise problem into a multi-beam coupling problem [29]. For a signal beam whose amplitude is within an order of magnitude of that of the pump beam, the scatter can hardly compete for gain diminishing the detected noise from fanning. Granted, sufficiently powerful signal beams will experience fanning themselves depending how far from the C axis they are aligned. But this signal beam fanning will likely not reach the detector if adequate steps are taken to block said fanning.

Several routs have been taken in attempts to estimate the degree of fanning in an experiment. [29, 30, 31]. As rough close-form solutions to estimate fanning intensity, these have been shown to closely approximate experimental observations in fanning amplitude. Without such estimates, and often with them [28, 29, 32], the primary means of estimating fanning is to numerically simulate the crystal itself with all known material properties and random scatter sources [33]. However, this would

require far more knowledge about one's particular crystal than we currently know about ours. And there is still the problem of obtaining the effective noise level at the frequencies of interest and combining this with the non-uniform speckle-like pattern overlaid on the detected signal.

In short, a significant amount of work would be necessary to acquire the scatter rate from the photorefractive, independent of the absorption rate, estimate the competitive gain for the signal and fanning beam, and all of this information would only be useful for a particular crystal, or perhaps for those made from the same crystal bule. Any attempts to get at the fanning noise level without the presence of two-beam coupling would be incomplete since the signal beam will bleach the fanning, particularly in a cone around the signal beam. Such a calculation would essentially be little more than an exercise in tedium, rather than a useful means of getting at the optical noise.

There is one bit of experimental evidence which suggests fanning may be a significant noise source and that is the fact that if I increase the optical power in the 532nm system the S/N goes down. Although it is believed more likely that the increase in noise is due to the laser linewidth, increased fanning may also be the culprit. However, separating these two noise sources would be quite difficult unless one could change the laser noise independent of the fanning amplitude. One way to get at the laser noise without fanning, which was not thought of till after the system was assembled, is to rotate the polarization of the laser such that two-beam coupling cannot occur in the photorefractive [34]. Unfortunately, to minimize convective noise the system (PV3) was built such that there is no room to insert an optic for polarization rotation. Another attempt to show the presence of fanning noise used a white, incoherent erasure beam, a super-bright LED or light bulb, to actively wash weak gratings in the photorefractive thereby diminishing fanning [35]. However this experiment showed that, for a detected optical power of about 4mW, where 2mW achieved the optimum S/N in PV3, the S/N was unaffected suggesting, that the fanning noise was not significant for the setup of PV3. This experiment was not performed for the PV4 system.

5.5 Detection Electronic Noise

In order to get a useful signal from an optical beam, a photodiode is used to generate photoelectrons that correlate with the photons detected from the interferometer. The current then

conducts through BJT's in the Hobbs circuit to a transimpedance amplifier, converting the non-common current into a voltage. After photo-electron conversion, there are several sources of noise that we briefly discuss.

5.5.1 Detector Noise

Optical signal level and noise levels are detected and transmitted as faithfully as the design allows, such that electronic noise ideally never exceeds optical noise thus, preserving the photocurrent S/N as closely as possible. However, Johnson noise from the resistors, input noise levels for the op amps and BJT's, and noise from the demodulation circuitry, all accumulate and potentially establish a noise floor in which picometer or smaller signals may be lost. An effective optical gain is established (through modulation-enhancement) using a technique nearly identical to heterodyne beating that allows for small optical signals to be strong enough that the electronic noise floor isn't as critical, but we still must know the noise level of our electronics.

Upon detection a signal level and noise level are established.

$$S/N_e = \left(S/N_{volt}\right)^2 = \frac{i_{sig}^2 R}{i_{nse}^2 R} = \left(\frac{i_{sig}}{\sqrt{2qBi_{tot}}}\right)^2 \tag{5.9}$$

The electronic or voltage signal-to-noise depends on the current fluctuations associated with the signal i_{sig}, and a noise level derived from the electron charge q, bandwidth of detection B, and the total photocurrent i_{tot}. Notice the transimpedance feedback resistor R drops out of the equation. The resistor will contribute to Johnson noise but otherwise is unimportant to the fundamental S/N limit. This S/N cannot be improved upon further as any electronic amplification will amplify signal and noise together while introducing further noise sources.

Since we detect voltage levels and calculate S/N based on said levels we will use the voltage S/N from this point forward. The expression for S/N immediately after the photodiode is

$$S/N_{volt} = \left(\frac{\left\{\Re G P_{opt} \cdot J_1\left(\delta_{ref}\right)\frac{2\pi(\Delta x)}{2\cdot 532nm}\sin(2\theta_0)\sin(2\beta)\right\}}{\sqrt{\left(2Bq\Re P_{opt}\right)+\sum\left\langle i_n\right\rangle^2}}\right) \tag{5.10}$$

Where $\Re$ is the responsivity in amps/watt and the remaining variables have previously been discussed in the modulation-enhancement section. Since we are effectively summing the output of two photodiodes in the Hobbs circuit, the current gain function goes as

$$G = \frac{P_- + P_+ \frac{P_-}{P_+}}{P_- + P_+} = \frac{2P_-}{P_- + P_+}, \tag{5.11}$$

where $P_\pm$ is the average optical power on the (+) and (-) port respectively. Therefore, if the beams are of equal intensity, the Hobbs circuit can theoretically achieve the shot-noise limit, otherwise the circuit will ideally incur a maximum error of 3dB off [18]. The detected optical powers are not balanced, but rather have a ratio of approximately 3:1.

We now examine contributing noise sources $\sum \langle i_n \rangle^2$ from the circuitry to estimate their impact on the detector. The contributing noise sources include Johnson from any resistor, the effective input noise of the op-amps, and photoelectric shot noise. The front end noise sources, as well as their anticipated contributions, are summarized in the table below.

Noise Source	Equation & relevant values	Effective noise level
Johnson	$Vn_J = \sqrt{4k_B TR}$ $In_J = \sqrt{4k_B T / R}$	$Vnj = 13nV\_(10k\Omega)$ $Inj = 1.3pA$ $Vnj = 29nV\_(50k\Omega)$ $Inj = 0.58pA$
Photoelectron shot	$In_S = q\sqrt{2B\Re\left(\frac{\lambda}{hc}\right)P_{tot}} \approx \sqrt{2q(\Re P_{tot} B)}$	$8.3pA/\sqrt{Hz}\,(2mW\_532nm)$ $\rightarrow 83nV/\sqrt{Hz}\,(10k\Omega)$ $12pA/\sqrt{Hz}\,(2mW\_660nm)$ $\rightarrow 620nV/\sqrt{Hz}\,(50k\Omega)$
Photodiode (PIN5DI) (V3)	$\Re_{532} = 0.25A/W$ $C = 15pf$ $I_d = 120pA(1.2nA)$ $NEP = 1.4e^{-14}W/\sqrt{Hz}$	$3.5fA/\sqrt{Hz}$
Photodiode (S8558) (V4)	$\Re_{660} = 0.45A/W$ $C = 10pf$ $I_d = 20pA(0.4nA)$ $NEP = 5.6e^{-15}W/\sqrt{Hz}$	$2.5fA/\sqrt{Hz}$
Op amp (MAX410/414)	$Vn = 4nV/\sqrt{Hz}$ $In = 2.6pA/\sqrt{Hz}$	$Vn = 4nV/\sqrt{Hz}$ $In = 2.6pA/\sqrt{Hz}$
BJT (DMMT5551)	$h_{fe} = 50-250(match2\%)$ $GBWP = 100-300MHz$ $NF = 8dB$	Insignificant [18] and unknown
BJT (MAT02)	$h_{fe} = 400-600(match2\%)$ $GBWP = 200MHz$ $Vn = 1.6nV/\sqrt{Hz}$	Insignificant and unknown

Table 5-1: Noise sources in Hobbs detection electronics. The photodiode levels due to dark current or NEP generate noise comparable to a few attometers of displacement. The dominant noise term thus far is the optical shot noise. The effective displacement corresponding to 8.3 pA and 12 pA is 11 fm and 12 fm respectively.

From the various noise sources in the front end we see that the optical noise should be dominant and that, for a 5V rail and a minimum signal of 1.4uV, we should have a front-end dynamic range of over 120dB, or six orders of magnitude for a voltage signal, ignoring the second harmonic. Allowed to propagate through the system, the second harmonic will potentially limit the system to 1-2 orders of magnitude depending on the modulation depth of the reference. A notch filter is used to improve dynamic range, though it may potentially add to phase and amplitude noise. The following AD630 synchronous demodulators used as lock-in amplifiers effectively cut the voltage S/N by at least a half with each gain stage, and have a dynamic range of up to 110dB. The entire demodulation board

schematic for PV4 is shown in figure 5-6 with exception of the TTL variable duty phase shifter depicted in figure 5-7. The implemented board is shown in figure 5-8.

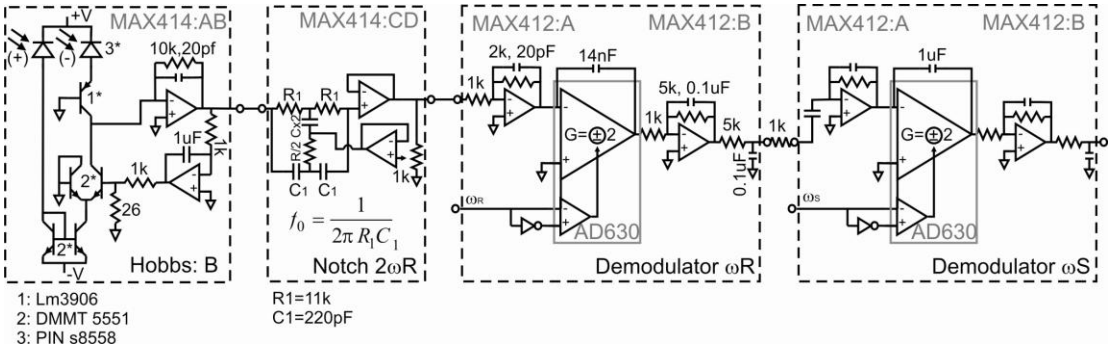

Figure 5-6: Demodulation board for Prism system V4. The Hobbs circuit uses two DMMT 5551 matched pairs in the Hobbs modification B configuration. Diodes are Hamamatsu PIN S8558 arrays. The depth of the notch filter is adjusted to maximize the signal to the limit of letting the second harmonic swing 90% of full scale. The synchronization signal for demodulation effectively flips the gain in the demodulator circuit and the low-pass filters clean out higher harmonics and switching artifacts. There are no component values for the second demodulation stage because this never successfully worked. A DC offset in the output of the prior stage required a HPF modification that is indicated here but was never implemented. All measurements of LOD were made using an external lock-in which read the output of the first demodulation stage.

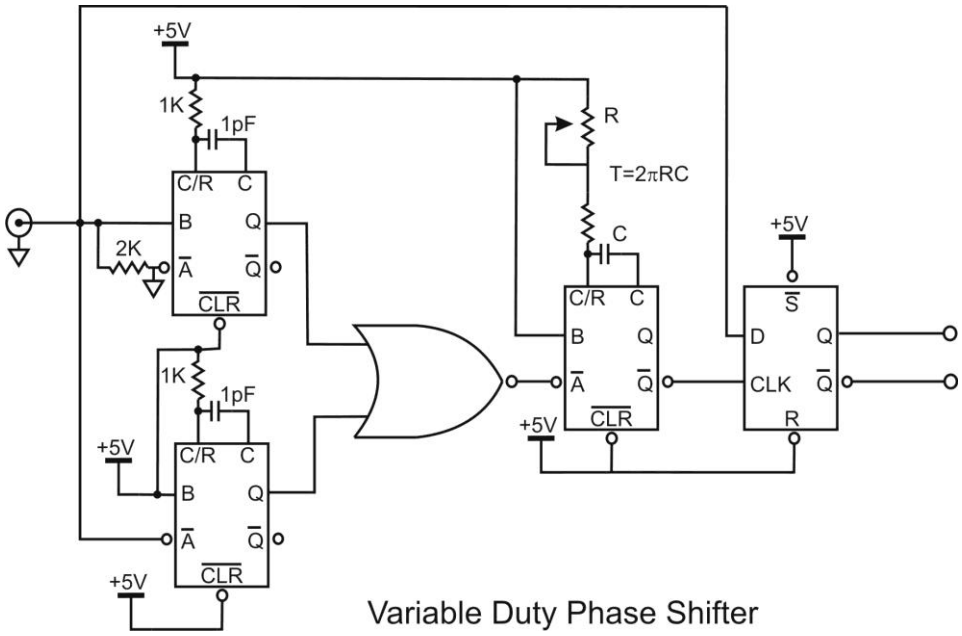

Figure 5-7: Variable duty TTL phase shifter helps to track slow or accumulative jitter. Design courtesy C.Sauer, JILA

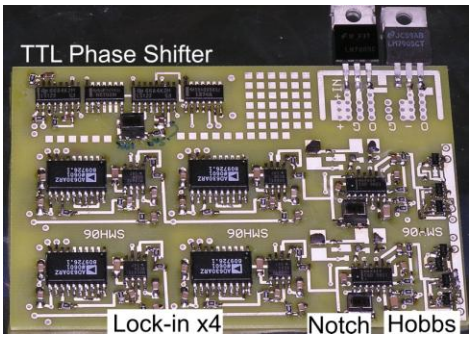

Figure 5-8: Board implementation. Each board contains two complete detector/dual-demodulation channels and a single TTL phase shifter. Therefore, one board's phase shifter is tuned for 10's of kHz while the other is tuned for 5Hz. Manual adjustment is required to adjust the TTL phase shift when changing signal or reference frequency as this board is designed to be adjusted for operation at a fixed frequency and modulation depth. When determining the optimum operation conditions the input to either demodulator can be tapped to an external lock-in amplifier for rapid frequency and amplitude sweeps.

It is anticipated that the shot noise will stay dominant through the entire circuit assuming only moderate gains and similar low-noise components are used. However, the largest anticipated source of detection/demodulation noise will come from the synchronous demodulators in the form of phase-to-amplitude noise. The notch filter potentially makes this problem worse as its phase response is not flat, further compounding phase and frequency fluctuations.

It should be stressed that the purpose of the notch filter is to improve the dynamic range of the system. The dynamic range is limited by the largest signal, which is the second harmonic,

$$I_{2\omega_R} \approx I_0 J_2\left[\delta_R(t)\right]\sin\left(2\theta_0\right)\sin(2\beta)\cos\left(2\omega_R t\right) \tag{5.12}$$

while the sum/difference frequency contains the signal of interest.

$$I_{sum/dif} \approx I_0\left[\delta_S(t)\right]J_1\left(\delta_R\right)\sin\left(2\theta_0\right)\sin(2\beta)\cos\left[(\omega_S \pm \omega_R)t\right] \tag{5.13}$$

The notch filter doesn't actually improve the limit-of-detection, it simply allows the system to operate more than one to two orders of magnitude above the limit of detection without railing due to the second harmonic. It also requires less filtering by the demodulators/lock-in amplifiers that would otherwise be needed to dampen the second harmonic at the output.

The notch filters shown in figure 5-9 operate on two principles. The filter on the left, designed by Dana Z. Anderson, applies a phase shift using an LC filter that is 180^0 shifted at the desired notch frequency. The notch response arises when the bypass signal is added to the phase

shifted signal, which, by tuning the potentiometer to balance the signal levels, will perfectly cancel at a particular frequency. The width of the notch is dictated by the slope of the phase while adjustment of the potentiometer will adjust the depth. This version has the advantage that the phase at half the notch frequency has a relatively gentle slope, decreasing the contribution of phase related noise at demodulation. Unfortunately the physical size of the inductors and low drift capacitors makes this notch filter prohibitively large when attempting to build an array of detectors.

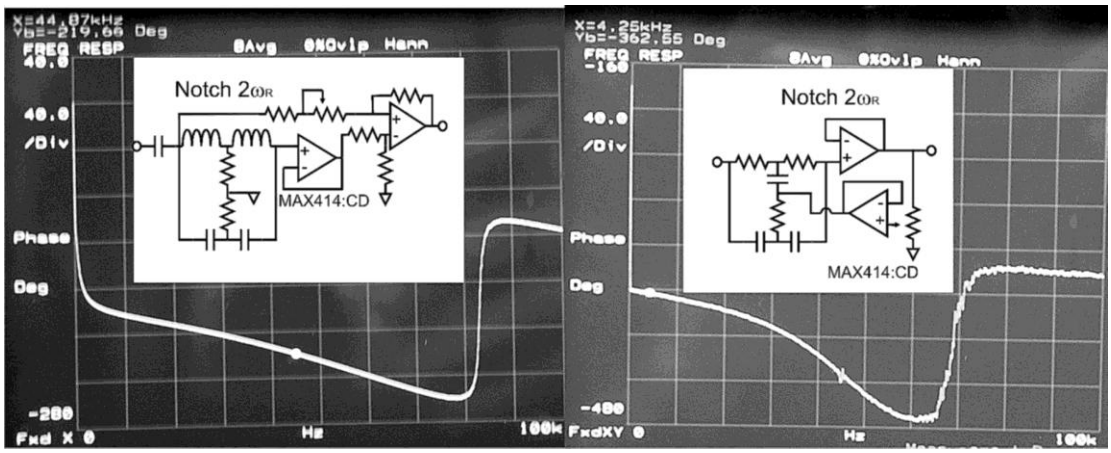

Figure 5-9: Notch filters used in Left: Prism system V3 and Right: PV4. PV3's filter has a lower phase slope and the signal is retarded by a smaller angle than that for PV4. However PV4's architecture allows it to be built on a surface mount board occupying nearly an order of magnitude less area for similar or superior Q performance.

The notch frequency of the left can be found by

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{5.14}$$

The depth of the notch as well as the Q are dictated by the matching between inductor pairs and between the capacitor pairs. A typical notch depth of around 25dB is attained with component matching of around 1 percent by tuning the inductors. The depth is optimized by the potentiometer, which is adjusted to perfectly balance the amplitude of the source and the 180^0 shifted component.

The second filter on the right of figure 5-9 is a National Instruments design capable of high Q's despite having passive components with low inherent Q and a center frequency defined by

$$f_0 = \frac{1}{2\pi R_1 C_1} \tag{5.15}$$

This design's notch depth is defined by the matching quality of the capacitors and resistors, for which it is significantly easier to get 0.1% resistors and 1% capacitors due to industry mass production. Because the capacitors for this design aren't as critical, and there are no bulky high Q inductors, this design leaves a footprint easily an order of magnitude smaller than a filter of the former design with similar Q. The major drawback of this design is that the signal is phase-delayed by a larger angle and the phase slope, at half the notch frequency, is over twice that of the former design. Table 5-2 summarizes the RCL values used and the matching tolerances.

System (2 ω_R notch)	**Value**	**Match**
PV3 (84 kHz)	L=1.2mH	variable
	C=2.9nf	1%
	R=450	1%
PV4 (64 kHz))	C=220pF	1%
	R=11kOhm	0.1%

Table 5-2 Summary of values used in notch filter for both the 180^0 LC subtraction filter and the high-Q RC filter. These are example frequencies and values since the reference modulation was adjusted several times between 40 and 45kHz.

The total noise level expected at the output of the front end, using the values from table 5-1, is 84 $nV/Hz^{1/2}$ and 620$nV/Hz^{1/2}$ for PV3 and PV4 respectively. This excludes RIN which was attenuated by adjusting the reference modulation depth and optical power until it no longer dominated the S/N. The corresponding expected sensitivity limits are 23$fm/Hz^{1/2}$ and 20$fm/Hz^{1/2}$ respectively. These values are insignificant compared to the achieved limits of detection. Assuming increasing optical power did not hurt the S/N, which is an erroneous assumption because this optical power (~2mW) was chosen for the optimum experimental S/N, the optical power would have to be increased by a factor of 60 and 7000 fold , for the PV3 and PV4 respectivly, for the shot-noise to dominate the LOD.

5.5.2 Lock-in Detection Jitter Noise:

Two demodulation stages are required to extract a low frequency signal off of a high frequency signal which is impressed onto the optical beam. This mixing is done to move the signal of

interest away from low-frequency electronic noise that may be introduced after the Hobbs auto-balanced subtraction circuit. However, the act of lock-in demodulation introduces noise due to phase jitter in the reference signal since the lock-in typically generates an internal phase-locked loop reference. A synchronous demodulator should alleviate jitter noise as long as a duty-cycle independent phase shift can be applied to the reference.

Using synchronous demodulation, the jitter of an oscillator is a noise source that can be mitigated if the jitter is identically transferred, through the detected optical signal and the demodulation reference, without relative phase delays. However, as the phase response of the above notch filters indicates, there is a phase delay of at least 220 and 400 degrees through the detector electronics making jitter noise an issue. While one might argue that 180^0 can be accounted for by an inverting amplifier, that would still mean a phase shift of over 40^0 and 220^0. The reported RMS jitter value of the Agilent 33250A function generator used to drive the reference modulation is on the order of $\Delta\phi_R \le 0.01\%+525\text{ps}$, which corresponds to approximately 3.6ns for a 32kHz oscillation and 2.7ns for a 45kHz oscillation.

To estimate the noise due to jitter, the function of the lock-in amplifier, or synchronous demodulator, can be modeled as a simple mixing of a reference modulation with the detected modulation. While the reference input is often TTL, its fundamental harmonic sinusoid exhibits the same phase noise [36]. Thus, we treat the reference as a sinusoid with phase noise.

$$V_{osc}(t) = V_R \sin\left(\omega_R t + \Delta\phi_R(t)\right) = V_R \sin\left(\omega_R\left[t + \frac{\Delta\phi_R(t)}{\omega_R}\right]\right) \tag{5.16}$$

where V_R is the voltage modulation depth. Equation (5.16) is essentially just the output of the oscillator with phase fluctuation, thus we are interested in the jitter stability of our oscillator.

For sufficiently small phase fluctuations $\Delta\phi_R(t) << \pi/2$ we can simplify the above relation to an additive noise relation.

$$V_{osc}(t) = V_R \sin\left(\omega_R t\right) + V_R \Delta\phi_R(t) \sin\left(\omega_R t\right) \tag{5.17}$$

We can further reduce the noise expression to a component existing at the detected sum/difference frequency.

$$V_{osc}(t) = V_R \sin(\omega_R t) + V_N \sin(\left[\omega_N + \omega_R\right]t) \tag{5.18}$$

The expression above effectively expresses the spectral purity of the oscillator defined by an additive noise source. This relation can be used to describe how the spectral purity of the oscillator will be detected as noise at the sum/difference frequency assuming either the photorefractive is imperfect or that there is a random phase-noise fluctuation between the arms in the interferometer. If neither of these is true, this additive term, as with all the other fundamental frequencies, will not be detected at the fundamental and therefore won't be a noise issue.

It has been reported that the grating of a photorefractive may be off by several degrees, which would enable a fundamental harmonic to effectively "leak" through. We determined above that the photorefractive likely has a grating phase shift of over ½ of a degree.

$$P_{\omega_R} = 2\left\{P \cdot \sin(2 \cdot \Delta\phi_{PR}) \cdot J_1(\delta_R) J_0(\delta_S)\right\} \tag{5.19}$$

Knowing that some fundamental exists, any jitter in the reference modulation will be converted to an amplitude noise when mixed with the "leaked" fundamental.

To estimate this mixture, we mix the oscillator reference with the detected optical signal. The mixed signal is then low pass filtered, approximated by an averaging over *n* periods, from which we see the detected signal.

$$\begin{aligned}\langle V \rangle &= \frac{\int_{t=0}^{2\pi n/\omega_R} \left[V_{osc}(t)\right] \cdot \left[V_{opt}(t)\right] dt}{2\pi n/\omega_R} \\ &= 2P\left\{\sin(2\Delta\phi_{PR}) \cdot J_1(\delta_R)\right\} \frac{1}{T} \int_t^{t+T} \left\{\sin(\omega_R t + \phi_R)\right\} \cdot \sin\left[\omega_R t\right] dt\end{aligned} \tag{5.20}$$

We then differentiate the result with respect to phase to determine the effective demodulation noise.

$$\frac{d\bar{V}}{d\phi_R} = P\left\{\sin(2\Delta\phi_{PR}) \cdot J_1(\delta_R) J_0(\delta_S)\right\} \sin(\phi_R) \tag{5.21}$$

Finally the differential power can be compared to the effective power of a detected signal to find the corresponding noise level.

$$\left\{\sin(2\Delta\phi_{PR}) J_0(\delta_S)\right\} \sin(\phi_R) \cdot \Delta\phi_R = 2\left\{\frac{2\pi(\Delta x)}{2 \cdot 532nm}\right\} \tag{5.22}$$

Before finding the effective noise level we note that if the reference frequency is perfectly phase matched to the detected optical signal, the effective displacement will be zero. Experimenting with successive auto-phase operations of the lock-in amplifier reveals a typical phase uncertainty of $\phi_R < 0.2^0$ over 5 minutes. For such a phase mismatch, the effective displacement noise, given a jitter level of $\Delta\phi_R < .01\%$, is 11fm. Therefore, using high quality lock-in amplifiers and function generators, the jitter noise should be insignificant. The resulting implications are that when tuning the demodulation board, being off by as much as 2 degrees will allow jitter noise to begin to dominate the LOD.

Later experiments, with the notch filter removed, showed no change in the LOD. Instead, they simply exhibited the anticipated reduced dynamic range, indicating the phase delay of the notch filter was not the dominant noise source. Two possibilities include either my tuning was within the 2^0 phase tolerance (<1/10th of a turn) or the jitter noise reported on the specifications sheet is overestimated.

5.6 Chemical Transduction Gain

For the moment we will briefly discuss chemical transducers, devices that can be used to interact with chemical constituents in the air in such a way that optical interrogation picks up a phase shift characteristic of the transducer/chemical interaction. Various transducers exist from polymers, enzymes, metal oxides, bacteria, and even immobilized antibodies. As our system only requires a total internal reflection to pick up phase information, any transducer that can be made thin enough, and therefore optically transparent to some degree, can be used.

The mechanism for detection is irrelevant, as detection with an analyte will require some chemical interaction which will in turn affect the local electronic and atomic structure in some way which will impart a phase shift to the interrogating beam. What is relevant is the gain factor associated with this chemical to optical signal. As some materials will have a greater affinity or chemically react stronger, a larger phase shift can be expected. Therefore, in many ways, this system is limited by the

sensitivity of the transducers. Even if the interferometry is shot noise limited, it is for naught if the system isn't complimented with high performance chemical transducers.

Since the point of interest in this section is limitations, there is also the issue of transducer noise. Noise will come about due to the transducer itself reacting to something other than the chemical content in the air. Should the transducer have more plastic and elastic properties, it may exhibit mechanical noise from particle shot noise, individual gas molecule impacts, and variances in particle velocity. More compressible polymers may compress more easily under such variant impacts. The transducer itself may exhibit some chemical or structural instability which will contribute to noise. Though little detail is explored in these problems, they are never-the-less a potential severe limitation with more exotic transducers and worthy of study for the evolution of this technology.

5.7 Optical Dispersion

As mentioned earlier, the stability of the laser source is critical to performing good interferometry. The line width can be viewed as the effective span in wavelength over which a laser will statistically fluctuate. This fluctuation will have its own spectra, which can effectively be viewed using an Allan variance. We don't show the Allan variance of the laser itself, but rather we show the variance when coupled through the output of the detection demodulation circuitry later on in section 5.10.

As the purpose of this thesis is not laser design, laser fabrication was only investigated in the last half year of research, which is why we only present the most cursory of laser design overview and performance analysis. The problems are presented along with potential solutions that have shown promise. However, full implementation, though successive redesigns, could not be performed. Instead, we will focus more on measures in the design of the interferometer to reduce its sensitivity to laser noise as these design aspects also coincide with several issues in reducing sensitivity to environmental noise and other interferants.

We take a moment for a brief review of the refractive index dependence on wavelength, as this shows itself frequently in dispersion equations. As wavelength increases in the visible spectra, or

more precisely the wavelength region bound by 532nm and 660nm, the index dependence (based off of the Melles Griot Glass catalog for BK7) goes as,

$$\frac{\delta n_{532nm}}{\delta \lambda} = -0.000056$$
$$\frac{\delta n_{660nm}}{\delta \lambda} = -0.000030 \tag{5.23}$$

Not only does index decrease with increasing wavelength but effects based on index change will see nearly twice the effect at 532nm, assuming similar laser stability.

Wavelength (λ)	Refractive Index (n)	$\Delta n / \Delta \lambda$
514.5	1.52049	
532.0	1.51947	-.00005601
546.1	1.51872	
643.8	1.51472	
656.3	1.51432	-.00002970
694.3	1.51322	

Table 5-3: Refractive indices of BK7 at various wavelengths of interest in addition to the effective variation in index with respect to wavelength. The refractive indices were acquired from the Melles Griot glass catalog for BK7, the material of the prism interferometer.

5.7.1 Path Length Dispersion

Fundamentally, if all else is ideal for an interferometry system, the phase resolution, or sensitivity, is dependant on the linewidth of the laser source. The linewidth can be viewed as the uncertainty or instability of the wavelength. Due to the photorefractive Bragg-matched grating, wavelength fluctuations directly translate to the fluctuations in coupling, thereby limiting the phase sensitivity. The measure of frequency instability is generally given by laser manufacturers in MHz, which translates to a wavelength uncertainty.

$$\Delta \nu = \frac{-c}{\lambda^2} \Delta \lambda \tag{5.24}$$

For this equation c is the speed of light in m/s, $\Delta \nu$ is the frequency uncertainty, and $\Delta \lambda$ is the resultant uncertainty in wavelength. The 532nm Coherent Compass 315M reports a frequency stability of <1-2MHz, yielding a line width of 0.9 to 1.9fm. Empirically, the worst-case frequency stability of the 660nm diode lasers we built is <10MHz over 3s. A three second period is quoted because the

photorefractive time constant is on the order of tenths of a second and thus three seconds allows the photorefractive to nearly completely adapt. Using the information from our laser sources, the line width contribution to noise can be determined.

Recall that phase modulation of the diffraction grating position results in optical phase modulation of the diffracted orders by the same phase angle, regardless of wavelength. The same is essentially true for phase shifts in the photorefractive grating. Should the phase difference between the beams suddenly shift, by say 90^0, this would result in the same amplitude fluctuation as if the Bragg grating shifting by the same phase. Therefore, phase fluctuations between the arms, due to wavelength fluctuations faster than the photorefractive time constant, yield coupling fluctuations proportional to the displacement in one arm that would yield the same phase difference. With this understanding about noise due to phase or grating fluctuations, the effective displacement noise can be calculated.

The displacement limit of detection δx is based on the uncertainty in the phase $\Delta\phi$ and the path difference Δr.

$$\begin{aligned}\delta x_{S/N=1} &= \frac{\Delta\phi}{2\pi}\lambda = \left(\frac{\Delta r}{\lambda} - \frac{\Delta r}{\lambda+\Delta\lambda}\right)\lambda \\ &= \Delta r\left(\frac{\Delta\lambda}{\lambda+\Delta\lambda}\right) \approx \Delta r\,\frac{\Delta\lambda}{\lambda}\end{aligned} \tag{5.25}$$

By converting to a frequency uncertainty, as this is the form most often obtained from specification sheets and spectrum analyzers, the displacement limit of detection goes as

$$\delta x = -\Delta r\,\frac{\lambda}{c}\Delta\upsilon = -\Delta r\,\frac{\Delta\upsilon}{\upsilon} \tag{5.26}$$

Increasing line width yields a loss in the limits of detection as predicted. However, the above does not explicitly account for various dispersion related noise factors arising from the geometry in the photorefractive interferometer. To ensure that the dispersion is properly modeled, we examine the interferometer instability more explicitly.

5.7.2 Dispersion Effects on Two-Beam Coupling

The following exercise assumes the wavelength at any instant is spectrally perfect and centered within the line width given above, though from one instant to the next wavelength can shift

within said line width. Effectively, the time period of interest is the difference of optical time of flight between the interfering arms from where they are split to recombination in the photorefractive.

The path length difference between each beam gives rise to a phase difference between each beam, resulting in a change in the detected intensity.

$$E_1 = \frac{E_0}{\sqrt{2}}\exp\left(-I(\vec{k}_1 \cdot r - \omega t)\right)$$
$$E_2 = \frac{E_0}{\sqrt{2}}\exp\left(-I(\vec{k}_2 \cdot (r + \Delta r) - \omega t + \phi_0)\right) \tag{5.27}$$

$$I = \left|E_1 + E_2\right|^2 = I_0\left(1 + \cos(K_\Lambda\left(\Delta r - x\right) + \phi_0)\right) \tag{5.28}$$

The path length of the arm from beamsplitter to photorefractive is r, ω is the carrier frequency, and the difference in *k* vectors is given as

$$K_\Lambda = 2k\sin\left(\frac{\theta_{TBC}}{2}\right). \tag{5.29}$$

Figure 5-10 illustrates the origin of the grating wave vector based on the wavevector subtraction of the two interfering beams. In effect, changing wavelength results in a rescaling of the individual wavenumbers, which in turn rescales the photorefractive grating wavenumber K_Λ .

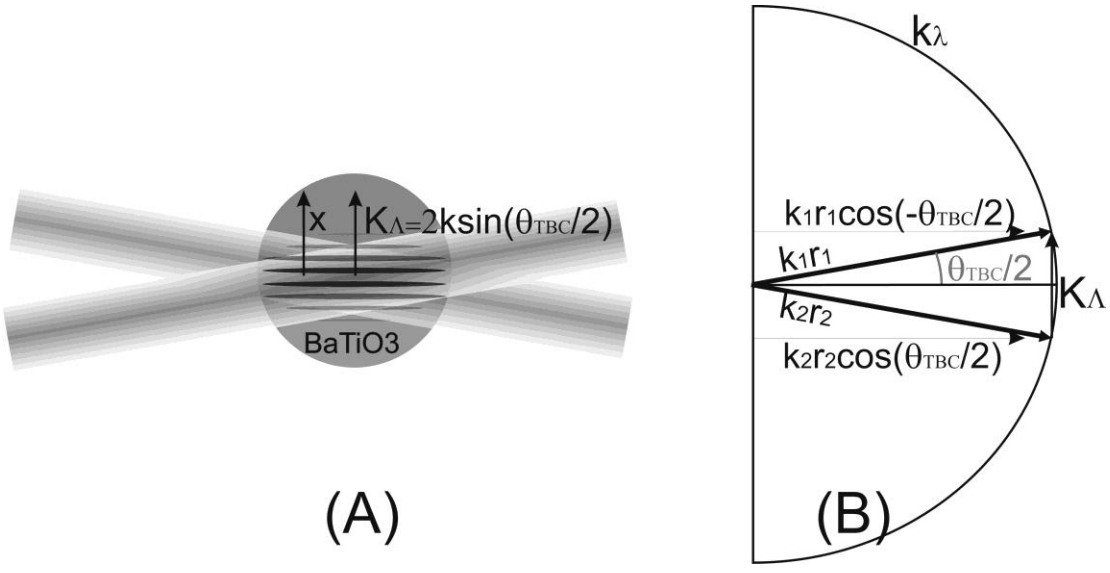

Figure 5-10: Two beams interfere such that the difference in wavevectors forms a grating wavevector in the photorefractive. θ_{TBC} is 16^0 and 18^0 for the 532nm and 660nm systems respectively.

The above model assumes a phase difference ϕ_0 between the beams while *k* is the wave number and the component which will give rise to the uncertainty in phase resolution. Also θ_{TBC} is the angle between the beams in the photorefractive and approximately 16^0 in the photorefractive and 18^0 in the prism. The position x along the photorefractive grating is measured along the vector $\vec{K}_\Lambda$, Δr is the difference in path length between the arms, and the wave number is defined simply as $k = 2\pi n/\lambda$. Differentiating with respect to wavelength yields the phase dependence on linewidth $\Delta\lambda$.

$$\begin{aligned}\frac{\partial\phi_\Lambda}{\partial\lambda} &= \frac{\partial}{\partial\lambda}\left(K_\Lambda\cdot\left(\Delta r - x\right)\right) = \frac{\partial K_\Lambda}{\partial\lambda}\left(\Delta r - x\right) + \frac{\partial\Delta r}{\partial\lambda}K_\Lambda = \frac{\partial K_\Lambda}{\partial\lambda}\left(\Delta r - x\right)\\ &= \left(2\frac{2\pi n}{\lambda}\frac{1}{2}\cos\left(\frac{\theta_{TBC}}{2}\right)\frac{\partial\theta_{TBC}}{\partial\lambda} + 2\frac{2\pi}{\lambda}\sin\left(\frac{\theta_{TBC}}{2}\right)\frac{\partial n}{\partial\lambda} - 2\frac{2\pi n}{\lambda^2}\sin\left(\frac{\theta_{TBC}}{2}\right)\right)\left(\Delta r - x\right)\end{aligned} \tag{5.30}$$

Should the path lengths be balanced, the phase dependence at $x = \Delta r$ disappears. This null result is due to the phase, relative between both beams, advancing equally, resulting in zero differential phase. Otherwise, a swelling and contraction of the grating about the center occurs due to a rescaling in the grating wavenumber K_Λ. Light coupled near the center of the grating will see no coupling change, due to such swelling/contraction, while as |*x*| increases, more photons will experience a change in coupling. Therefore, larger beam waists contribute to wavelength-dependant noise. This noise is only a severe issue if the entire beam is not detected. Otherwise, if the phase noise or wavelength noise is spatially uniform across the beam, integrating the intensity fluctuations over the area results in a net fluctuation of zero, for equal path lengths.

We now inspect individual components of equation (5.30) for dominant terms starting with the refractive index dependence.

$$\frac{\partial\phi_\Lambda}{\partial\lambda} = 2\frac{2\pi}{\lambda}\sin\left(\frac{\theta_{TBC}}{2}\right)\frac{\partial n}{\partial\lambda}\left(\Delta r - x\right) \tag{5.31}$$

Given the value of the index dispersion in the table below, and having only a single λ in the denominator, this term is insignificant compared to the remaining terms and will subsequently be ignored.

5.7.3 Angular Dispersion

The phase dependence of the photorefractive grating with respect to angular dispersion is given as

$$\frac{\partial \phi_\Lambda}{\partial \lambda} = \frac{2\pi n}{\lambda} \cos\left(\frac{\theta_{TBC}}{2}\right) \frac{\partial \theta}{\partial \lambda} (\Delta r - x). \tag{5.32}$$

Where n is the refractive index in the photorefractive or prism and θ_{TBC} is the angle between the beams in the photorefractive or prism respectively. Needing only to define the angular dispersion term to complete this expression, this section explores the sources that may contribute to this angular dispersion.

The angularly-dependant component has three main contributors, or sources of angular instability. The three angular uncertainties include: 1) fundamental laser pointing stability, which may or may not be coupled to wavelength; 2) refraction angle upon entering the prism; and 3) diffraction angle off of the diffractive beamsplitter. The angular dependence on refraction angle is

$$\frac{\delta \theta_r}{\delta \lambda} = \frac{\sin \theta_{inc}}{n_g \cos \theta_r}, \tag{5.33}$$

where $n_g \approx 1.51$ is the index of glass, $\theta_{inc} \approx 56^0$ is the angle incident into the prism, and $\theta_r \approx 33^0$ is the refraction angle. This term is negligible compared to that dependant on the diffraction angle

$$\frac{\delta \theta_m}{\delta \lambda} \approx \frac{m}{n \Lambda_d \cos \theta_m} \tag{5.34}$$

The diffracted order is given by m=-1, making $\theta_m \approx -14^0$ the diffracted angle, and Λ_d the diffraction grating pitch (416nm for a 2400 l/mm grating). This design sets $\theta_m \approx -14^0$ to reduce the contribution of noise from the diffraction angle dependence on wavelength. A -1st order diffracted beam interfering with a 0th order beam results in an angular dispersion term

$$\frac{\partial \phi_\Lambda}{\partial \lambda} = -\cos\left(\frac{\theta_{TBC}}{2}\right) \frac{2\pi}{\Lambda_d \lambda \cos \theta_m} (\Delta r - x) \tag{5.35}$$

The remaining component of equation (5.30) expresses a linear phase relation to wavelength variations regardless of the two-beam coupling angle or refractive index variations.

$$\frac{\partial \phi_{\Lambda}}{\partial \lambda} = -2\frac{2\pi n}{\lambda^2}\sin\left(\frac{\theta_{TBC}}{2}\right)(\Delta r - x) \tag{5.36}$$

Since the path length in the prism is over an order of magnitude larger than in the photorefractive, the conditions in the prism are used to estimate the contribution of this equation.

If balanced properly in the design of the interferometer, by using the proper diffracted order to control the sign, equations (5.35) and (5.36) could be made to compete, thereby diminishing the overall effective noise. However, to get balanced path lengths and to achieve a workable geometry, our design uses the -1st diffractive order, making this tactic unattainable. We now move on to the last significant dispersion term.

5.7.4 TIR Dispersion

The fact that TIR is advantageous due to its phase response to Δn means that same sensitivity can be a burden where dispersion is an issue because the refractive index varies with wavelength. The phase response to a change in index is

$$\phi_{TM}(n_s) = 2\arctan\left(\frac{n_s^2\sqrt{\sin^2(\theta_{TIR}) - \frac{1}{n_s^2}}}{\cos(\theta_{TIR})}\right) \tag{5.37}$$

where θ_{TIR} is the prism/air total-internal reflection angle within the prism and n_s is the refractive index at the TIR surface, which is assumed to be BK7 for simplicity.

Analysis of the wavelength dependence of each term yields a dispersion term that has two of our prior perturbation relations.

$$\frac{d\phi_{TM}(n_s)}{d\lambda} \approx 25\frac{d\theta}{d\lambda} + 15\frac{dn}{d\lambda} \tag{5.38}$$

The details of the full derivative are left to the adventurous reader as the transcribing of it is an exercise in tedium that will likely not benefit the reader to see. Instead, the result will be given assuming 42^0 TIR angle. The two dispersion factors are functions of $d\theta_{TIR}/d\lambda$ and $dn_s/d\lambda$. Since the refractive index dependence on wavelength may vary from transducer to transducer, this term is

estimated based on the calculation for glass $dn_s / d\lambda = dn_{BK7} / d\lambda$, which was shown to be negligible. The angular dependant term is coupled with the diffraction angle's dependence on linewidth, as this was the largest angular dispersion relation. The TIR angle is assumed to be 42^0 (just shy of the BK7/air critical angle) for maximum effect.

The effective TIR dispersion term can now be shown to be a weighted dependence on the diffraction angle.

$$\frac{d\phi_{TM}}{d\lambda} \approx 25\frac{d\theta}{d\lambda} = 25\frac{m}{n_s\Lambda_d \cos\theta_m} \tag{5.39}$$

Note there is no $(\Delta r - x)$ dependence, as this term affects the phase of a single arm, the arm exposed to the chemical transducer, and therefore this is a fixed phase shift that cannot be escaped by path length balancing. All that's left regarding dispersion would appear to be the determination of which of the above terms dominate.

5.7.5 Dominant Dispersion Term

To calculate the displacement noise, or effective limit of detection, we need the path length difference and a relation between phase resolution and displacement sensitivity.

$$\delta x_{S/N=1} = \frac{\Delta\phi}{2\pi}\lambda = \frac{\lambda}{2\pi}\sum_n \frac{d\phi_n}{d\lambda}\Delta\lambda \tag{5.40}$$

The absolute limit of detection can now be calculated, assuming that wavelength instability is the limiting factor. While other noise factors may dominate the instability making noise worse, then assuming the conditions of frequency stability, pointing stability, and path length difference are correct, one cannot detect displacements smaller than this. The various components that in some way depend on wavelength dispersion are summarized in table 5-4.

Dispersion Source	**Phase relation**	**Relative Impact (**rads/ $\Delta\lambda$ **)**
Refractive Index of the Prism	$\frac{\partial\phi_\Lambda}{\partial\lambda} = 2\frac{2\pi}{\lambda}\sin\left(\frac{\theta_{TBC}}{2}\right)\frac{\partial n}{\partial\lambda}(\Delta r - x)$	$\approx 2\cdot 10^2(\Delta r - x)$
Refraction Angle Into the Prism	$\frac{\partial\phi_\Lambda}{\partial\lambda} = \frac{2\pi n}{\lambda}\cos\left(\frac{\theta_{TBC}}{2}\right)\frac{\sin\theta_{inc}}{n_{BK7}\cos\theta_r}(\Delta r - x)$	$\approx 1.1\cdot 10^7(\Delta r - x)$
Diffraction Angle from the Beamsplitter/ Modulator m=-1	$\frac{\partial\phi_\Lambda}{\partial\lambda} = -\cos\left(\frac{\theta_{TBC}}{2}\right)\frac{2\pi}{\Lambda_d\lambda\cos\theta_m}(\Delta r - x)$	$\approx -2.9\cdot 10^{13}(\Delta r - x)$ 532nm $\approx -2.3\cdot 10^{13}(\Delta r - x)$ 660nm
Linear Dispersion Relation	$\frac{\partial\phi_\Lambda}{\partial\lambda} = -2\frac{2\pi n}{\lambda^2}\sin\left(\frac{\theta_{TBC}}{2}\right)(\Delta r - x)$	$\approx -1\cdot 10^{13}(\Delta r - x)$ 532nm $\approx -6.8\cdot 10^{12}(\Delta r - x)$ 660nm
Fresnel Phase at TIR	$\frac{d\phi_{TM}}{d\lambda} \approx \ldots 25\frac{m}{n\Lambda_d\cos\theta_m}$	$\approx -4.1e\cdot 10^7$

Table 5-4: Relative contribution from dispersion effects. While TIR is not as sensitive to wavelength as previously thought, depending on the path length balance in the interferometer, it provides the detection floor that no amount of tweaking can get us past. To be fair, alignment issues, fabrication tolerances, and transducer thickness variations among other imperfections will likely make path-length matching better than $\Delta r = 0.1mm$ impractical.

Table 5-4 shows that the two dominant dispersion-driven noise sources are the diffraction angle and the linear dependant term. The two dominant terms are used to find the effective displacement noise due to linewidth using equation (5.40) above.

$$\begin{aligned}\delta x_{532nm} &\approx 3.1\cdot 10^{-15}\Delta\nu(\Delta r - x)\\ \delta x_{660nm} &\approx 4.5\cdot 10^{-15}\Delta\nu(\Delta r - x)\end{aligned} \tag{5.41}$$

This is not so dissimilar from the simplistic and shorter derivation at the beginning of this section 5.7.1.

$$\delta x_{532nm} = -\Delta r\frac{\lambda}{c}\Delta\upsilon = 1.8\cdot 10^{-15}\Delta\upsilon\Delta r \tag{5.42}$$

Indeed, the earlier estimate was within a factor of three, though a little simplistic. Ultimately, the effective limit of detection due to wavelength fluctuations is within reach, though we take a moment to investigate a potential complication.

5.7.6 Spatial Average of Dispersion

It was mentioned earlier that to be more accurate in the determination of the limits due to laser stability, an averaging over the beamwidth in the photorefractive is necessary. The overlap of the two beams, at the widest point, is estimated to be D=0.8 mm ($1/e^2$). Let us then define the normalized Gaussian profile over which we shall integrate.

$$I(x) = \frac{2}{\pi\omega^2}\exp\left(-2\frac{x^2}{\omega^2}\right) \tag{5.43}$$

The average limit of detection over the entire profile is calculated since this is what a single photodiode effectively detects, assuming the beam is centered and wholly contained on the photodiode. We then integrate over the photodiode.

$$\delta x_{S/N=1} \approx \int_{-2\omega}^{2\omega} I(x)\cdot\delta x_{S/N=1}dx = \int_{-2\omega}^{2\omega} I(x)\cdot 3.1\cdot 10^{-15}\Delta\nu\left(\Delta r - x\right)dx \tag{5.44}$$

The normalized intensity profile negates the need to normalize the result of the integral. Again, for matched path lengths, this term will average out to negligible fluctuations in intensity. As the path lengths diverge, we see that the limit of detection scales linearly. Due to the weighting of the Gaussian in the integration over the beam waist, the integration is proven unnecessary as it yields the same result as just finding the effective displacement noise at x=0.

Using either equations (5.44) or (5.41) the displacement uncertainty can be calculated. The table below shows a list of optical path length mismatches corresponding to nose systems, and the resulting resolution limit of the interferometer for the expected range linewidth for the laser used. It should be considered that linewidth specs given are generally the noisiest possible and likely based on drifts over long periods of time (hrs). While the 532 nm Compass 315M may have an instantaneous linewidth in the 10's kHz range as a Coherent technician stated these were, "cherry picked for small linewidths," it is possible that the center drifts on the order of the photorefractive response time (0.1-0.5 s) increasing the effective linewidth. The actual effective linewidth is unknown since in various tests we were unable to get two separate lasers to within 2 GHZ of each other to measure a linewidth.

We assume a maximum linewidth of 1Mhz, as quoted by the specifications, and a minimum of 100 kHz. Based on this range it reasonable that the path length error approaches 0.1 mm, since we

detect down to 180 fm with PV3. PV4, on the other hand, showed a linewidth that is likely under 5 MHz, which would also correspond to an alignment error of 0.1mm. The uncertainty in path length presented is due to the inability to actually measure the path length in the prism. It may be possible using white-light interferometry to determine if a path length balance is possible however, since beam alignment is sensitive it would be challenging to then replace the white light source with a perfectly aligned laser to take advantage of the white light alignment. The values presented are estimates of the optical path length mismatch based on manufacturing tolerances and a CAD software package.

System / Line Width	~10mm Briefcase Crystalaser	5mm Prism V2 Compass 315M	*0.1mm* Prism V3 Compass 315M	*0.1mm* Prism V4 VHG LD 660nm
10 MHz	--	--	--	*4.4 pm*
1 MHz	31 pm	15 pm	*0.31pm*	*0.44 pm*
0.1 MHz	3.1 pm	1.5 pm	0.031 pm	--
Measured Limit S/N=1	**2 pm/Hz$^{1/2}$**	**0.7 pm/Hz$^{1/2}$**	**0.18 pm/Hz$^{1/2}$**	**1.7 pm/Hz$^{1/2}$**

Table 5-5: Calculated displacement limits-of-detection values for several nose systems based on linewidth and the optical path mismatch between the interferometer arms. The bottom row gives actual measured values for limit of detection.

Based on the above calculations it would seem reasonable to make the following conclusions. The linewidths of the Coherent laser and the 660 nm diode laser are likely about 0.5-1 MHz and 2-4 MHz respectively. While spectrum analysis results of the 660 nm beat note taken suggested it would be closer to 10 MHz, this was an absolute worst case. Experiments show the mean noise was closer to 3 MHz over 3 seconds and easily under 100 kHz in 50 ms. Based on the differences in linewidth, and the similarities in the rest of the interferometry, it is reasonable to expect that the portable system behaves a factor of 10 worse than prism system V3. It is also reasonable to assume that the chief limiting factor for the systems is the linewidth, though other sections in this thesis will contend that there are a few close contenders.

5.8 Motivation Behind Vapor Modulation

Several sources of noise have been explored that would limit the fundamental detection sensitivity. However, it has only briefly been mentioned why we modulate the vapor at 1-10 Hz, which was the source of several restrictions in the systems design. While reference modulation mixing would have still been useful for amplifying the signal against a strong reference frequency just to get over electronic white noise, avoiding low frequencies altogether would have been easier. Unfortunately, the low frequencies of our signals cannot be avoided.

This system is limited in frequency by two primary components, the photorefractive and the vapor delivery tubes. The photorefractive BaTiO3 performs dynamic holography, which acts as a high-pass filter (HPF) that potentially attenuates signals below 10 Hz. The sampled vapors experience laminar flow through reasonable tube diameters (<5mm) at flow rates around 100-300 ml/min. A characteristic laminar flow acts as a low-pass filter(LPF) for any periodic chemical signal flowing through said tubes. Additionally, the use of a Hobbs circuit means there is an effective band-pass response as the auto-balancing circuitry acts as a HPF while the GBP and additional filtration limits the upper frequency response. Since the detection electronics have already been discussed we take the following sections to discuss the affective optical/chemical band-pass filter characteristics of the system.

5.8.1 Photorefractive High-Pass Filter Dynamics

A photorefractive writes a dynamic hologram which, even for non-uniform phase profiles, will write a Bragg matched grating that couples and phase matches the two beams such that an effective single fringe demodulation is observed on the output beams. As a photorefractive whose carrier migration is dominated by diffusion, BaTiO3 writes a grating that is phase shifted by 90^0 such that the interference at the output ports results in a single fringe bright and dark field.

It is a common practice, where resources permit, to increase optical power in order to improve S/N, assuming the noise limitation is optical shot noise or other non-optical noise sources like electronic white or 1/f noise. However, when using photorefractives, especially for signal frequencies

such that $\omega_{signal} \leq 1/\tau_{PR}$, the photorefractive acts as a high pass filter, partially attenuating the signal on the output ports. To better understand what, if any, gains we can expect to achieve from increasing the optical power we must understand the time constant associated with the photorefractive more fully.

The photorefractive grating formation time for BaTiO3 is inversely dependant on the incident intensity. Empirically, a sub-linear frequency response to beam intensity is observed in thick, photorefractive crystals [37]. In the limit of thick dynamic gratings, the literature reports space-charge time-constant dependencies on intensity between $1/I$ and $1/\sqrt{I}$. This sub-linear photoconductivity is due to the charge-migration speed which is in turn dictated by the concentrations of free-carriers and effective empty traps of the particular crystal [38]. Put more simply, increasing the intensity causes the grating to adapt faster to phase changes between the beams and thus the effective 3dB point of the "photorefractive high-pass-filter" shifts by $1/I^x$ where x=0.5-1.

.

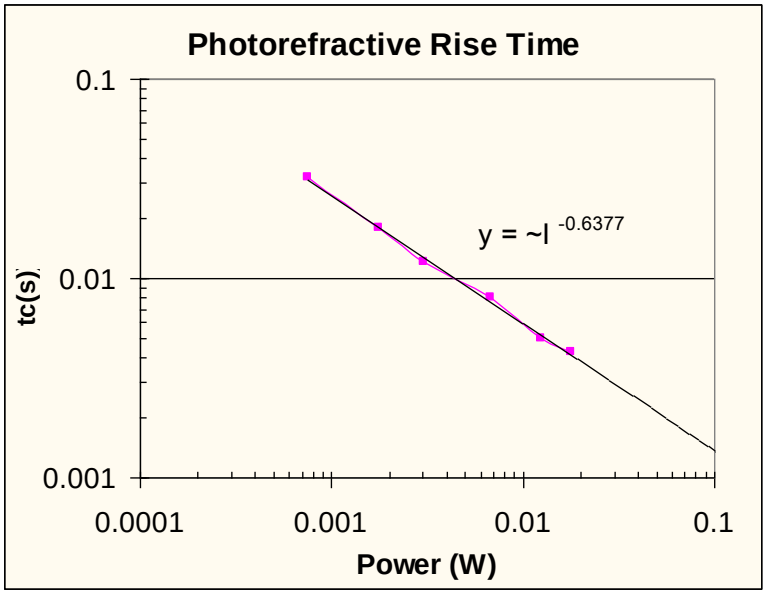

Figure 5-11: The Photorefractive response time (1/e) vs incident optical power of a 532nm beam in BaTiO3. The photorefractive responds with a sub linear power of -0.64 for the fixed internal coupling angle of 8^0. Note this data is accumulated for a constant beam diameter of 0.32mm.

As figure 5-11 shows, the BaTiO3 used for our experiments exhibits a sub-linear response, meaning that an increase in optical intensity will only increase the signal response proportionate to

$$\frac{\omega_{HPF}}{\omega_{HPF\_0}} \approx \left(\frac{I}{I_0}\right)^{0.64} \tag{5.45}$$

where I_0 is the initial optical intensity, I is the final, and the ratio of the high-pass filter point is expressed on the left.

Effectively, the photorefractive acts as a HPF, thereby attenuating low-frequency phase fluctuations by continually and slowly rewriting the grating. Attempts at increasing the intensity result in a signal increase by virtue of more photons being detected, as well as a decrease due to the effective shift in filter frequency. The photorefractive crystal response goes as [39,40],

$$g(t_s) \propto \frac{1}{\tau_{PR}} \exp\left(-\frac{t_s}{\tau_{PR}}\right) \tag{5.46}$$

Where t_s is the time after a perturbation or the effective period of a signal and τ_{PR} is the photorefractive response time constant. Since the grating adapts as a low pass filter the optical response, optical signals that are not attenuated by grating adaptation, goes as a high-pass filter or $I(\omega) \propto 1 - G(\omega)$. We then convert the optical response to the frequency domain in which we multiply the result by the transform of the sinusoidal signal. Since the delta functions of the cosine transform are equally spaced about the origin at $\pm\omega_s$ the transform of the effective high-pass function acts as a scaling coefficient on the cosine back in the time domain where the coefficient $C(\omega_s)$, as a function of signal frequency ω_s, is given by

$$C(\omega_s) = \frac{\omega_s^2 \tau_{PR}^2}{1 + \omega_s^2 \tau_{PR}^2} = \frac{1}{1 + \frac{\omega_{HPF}^2}{\omega_s^2}}, \tag{5.47}$$

where ω_s is the phase-modulated signal frequency and $\omega_{HPF} = 1/\tau_{PR}$ is the effective 3dB filter point of the photorefractive.

Combining the transfer function, based on the filter frequency, and the fundamental increase in photon count yields the first-order approximate gain G in signal amplitude based on intensity.

$$G \approx \left(\frac{I}{I_0}\right)^x \frac{\omega_{HPF\_0}^2 + \omega_s^2}{\omega_{HPF}^2 + \omega_s^2} \underset{x=1}{\overset{x'=0.64}{\rightarrow}} \frac{I\left(\omega_{HPF\_0}^2 + \omega_s^2\right)}{I\left(\frac{I}{I_0}\right)^{0.28} \omega_{HPF\_0}^2 + I_0 \omega_s^2} \tag{5.48}$$

Where *x'* is the sublinear coefficient and *x* is the fundamental improvement in S/N having a value between 0 and 1. Values of 1 occur when the dominant noise is constant, values of ½ occur when the signal is shot noise limited, and values of 0 occur when the noise scales linearly with the intensity. For simplicity we assume the noise is constant for the following equation. This equation indicates that there is an optimum optical intensity. The optimum intensity is found by differentiating and setting the above equation to zero.

$$I \approx I_0\left(-\frac{x}{\left(x-2x'\right)}\frac{\omega_s^2}{\omega_{HPF\_0}^2}\right)^{1/2x'} \underset{x=1}{\overset{x'=0.64}{\rightarrow}} 2.7I_0\left(\frac{\omega_S}{\omega_{HPF\_0}}\right)^{2/1.28} \tag{5.49}$$

This simple equation assumes a single noise source and that said source remains dominant regardless of the intensity. Realistically, if a system were marginally electronic noise limited, then increasing power past a point would make the system optical or shot-noise limited, thereby changing the exponent. We now give an accumulated signal to noise equation accounting for the effect of intensity on the photorefractive response.

$$S/N_{volt} \approx \frac{\Re G \cdot P \cdot J_1\left(\delta_{ref}\right)\frac{2\pi(\Delta x)}{2\cdot\lambda}\sin(2\theta_0)\sin(2\beta)}{\sqrt{\left(2Bq\Re G\left\{P/P_0\right\}\right)+\sum\left\langle i_n\right\rangle^2}\left(1+\frac{\omega_{HPF}^2}{\omega_s^2}\right)} \tag{5.50}$$

While the above equation is more accurate, it leaves out two limiting noise sources. The first is noise that scales with the power, including noise due to the PTZ grating modulator or the function generator driving the reference with noise at ω_S or $\omega_R \pm \omega_S$. The second is noise from the laser at the same frequencies (ω_S or $\omega_R \pm \omega_S$). If the system is limited by either of these, then one should acquire better components or give up. Some work is done in the prism design to mitigate laser noise but there is a practical limit, which is discussed in the section 5.7.

As an example of the intensity dependant signal, upon exposure to a 532nm beam at 2mW, the photorefractive exhibits a HPF 3dB point of approximately 5Hz. Figure 5-12 shows the HPF behavior for a BaTiO3 crystal irradiated with 532nm at 2mW detected power for frequencies ranging from 1Hz to 1kHz. The 3dB point occurs roughly at 5Hz and a peak response occurs around 20-50Hz. Successive experiments show that my optimum LOD occurs between 1 and 3 miliwatts of detected power for a 5 Hz signal. As I increase optical power, the noise rises super-linearly, likely due to the

reference oscillator and the phase stability of the carrier suppression in the photorefractive. As I decreased optical power, the signal simply dropped relative to background noise, again hurting the S/N, and indicating that either this was a transition point for noise source dominance or that there was insufficient intensity to maintain efficient two-beam coupling. Before we delve further into the signal frequency vs. S/N, we look at the other filter source, which, when combined with the photorefractive, establishes a band-pass transfer function.

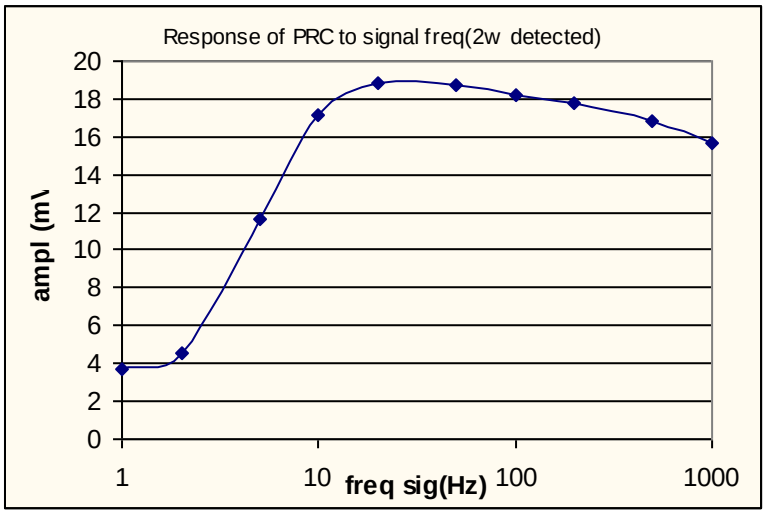

Figure 5-12: Spectral transfer function of the photorefractive for phase modulated signals imposed in two-beam coupling. Note the 3dB point for 2mW at 532nm is approximately 5Hz.

5.8.2 Laminar Flow Low-Pass Filter Dynamics

To actually measure the concentration of an analyte in a vapor, it must first be sampled from the environment. The sample is then exposed to the sensitive transducer element, often at a predefined sampling rate, in this case measured by volumetric gas per second. To move a sample, a pump is used to provide a positive or negative pressure and therefore tubing is used to isolate and guide the test gas and reference gas. However, the use of tubing on a scale that is practical for our system introduces yet more challenges. Section 6.6.3 discusses the low pass filter characteristics of laminar flow in a tube in greater detail.

The punch line is that a modulated chemical in a tube under laminar flow conditions will effectively be low pass filtered by the following transfer function.

$$H = \frac{\bar{C}|_{pp}}{C_0} \approx \frac{1}{1+\kappa\frac{\omega_s L}{\bar{v}R}} \tag{5.51}$$

The original contrast C_0, a measure of content-to-blank contrast, in a tube of length L and radius R, under flow conditions of average velocity $\bar{v}$ is low pass filtered for the modulation frequency to the new contrast, assuming the entire flow out of the pipe is measured equally, of $\bar{C}|_{pp}$. The fitting constant $\kappa \approx 2\cdot 10^{-4}$ is used to adjust this first approximation to the excessively complicated solution.

5.8.3 PRC and Laminar Bandpass Characteristics

By combining the two filter characteristics we find an effective band pass filter.

$$H(\omega) \propto \left(\frac{1}{1+\frac{\omega_{HPF}^2}{\omega_s^2}}\right)\left(\frac{1}{1+\kappa\frac{\omega_s L}{\bar{v}R}}\right) \tag{5.52}$$

Just as in a band pass circuit consisting of a capacitor and inductor, the photorefractive and delivery tube time constants may be modifiable by changing their effective impedance. If we model the pair as a band pass, we note that it would be beneficial to detection limits if the 3dB point of each were moved apart, lowering the Q, but allowing the band to peak closer to unity. Ideally, one would simply adjust the time constants of each to shift the center frequency to the desired chemical modulation frequency. However, other restraints, such as optimum valve frequency, system optical power, minimum tube length, and optical geometry all place limits on the adjustments of various components. Rather than calculate the ideal components, we simply aimed to use the shortest tubes possible of reasonable diameter such that the flow volume was less than that exchanged in a half cycle at 5Hz, the photorefractive 3dB frequency.

The low pass point of the laminar flow can also be pushed to higher frequencies by increasing the pressure, or effectively the average flow velocity, though doing so also increases the amplitude of

any differential pressure noise inherent to valve switching. Likewise, the photorefractive filter characteristics can be pushed to lower frequencies by decreasing the response time of the photorefractive. Lower response times can be achieved by lowering the intensity of the beams writing the internal grating, or otherwise lowering the response time of the space charge field. Since lowering the laser beam intensity will then hurt the signal to noise by $1/\sqrt{P}$, this factor will work against the intensity dependant response of the grating.

However, the photorefractive can be slowed down in other ways without changing the beam intensity. One such way is to increase the wavelength to one that the photorefractive responds less efficiently and, in effect, more slowly for the same optical power. If all other parameters were held constant, changing the wavelength decreases the shot noise for the same power level while also slowing the response time of the photorefractive. Figure 5-13 and figure 5-14 show the frequency dependant transfer function of the same photorefractive for two different wavelengths {532nm, 660nm}. Unfortunately, for the system design, changing wavelength will necessitate a redesign of every wavelength-dependant component such as the diffraction grating beamsplitting angle and the prism geometry for TIR and Brewster angle coupling.

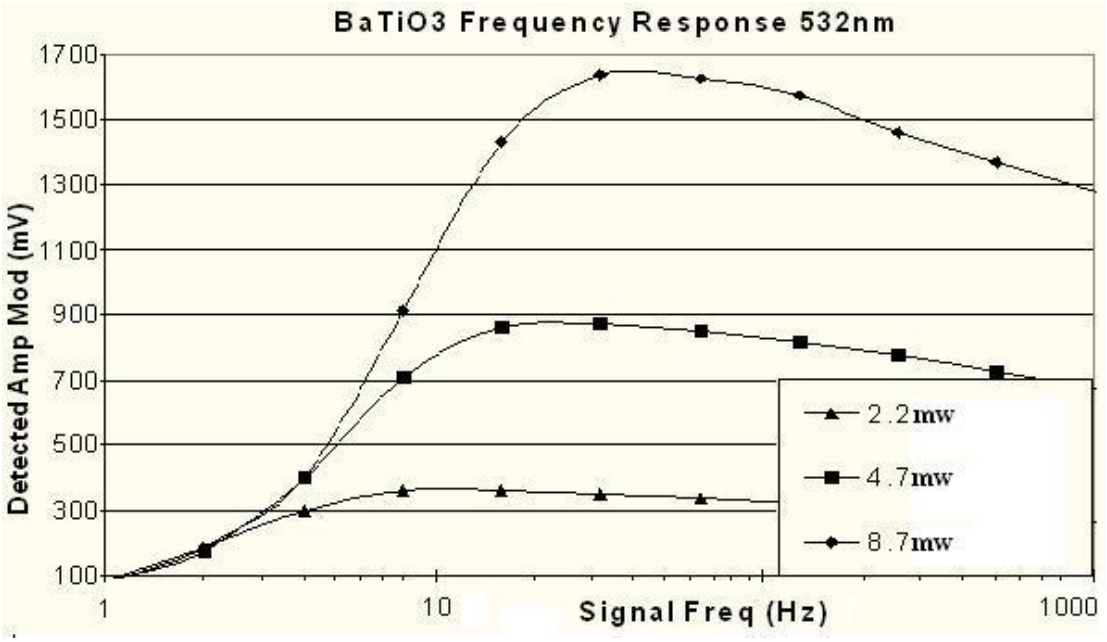

Figure 5-13: . Phase-to-amplitude response of BaTiO3 to a 532nm phase modulated signal. This plot shows signals levels for 3 different detected power levels, given in the legend in milliwatts.

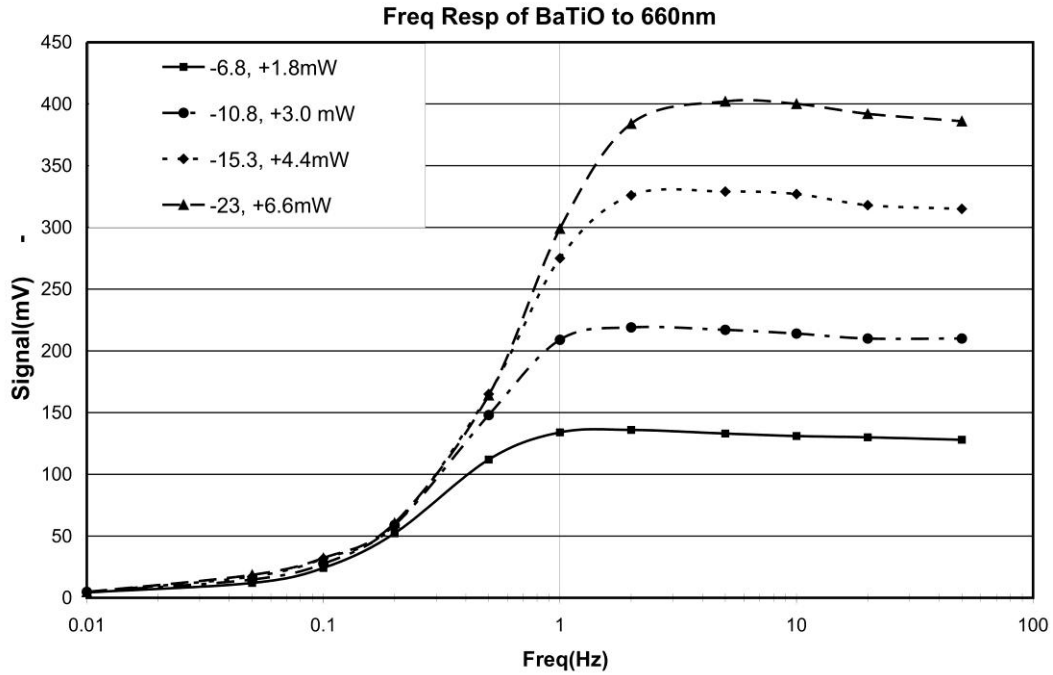

Figure 5-14: The signal amplitude detected from a small phase modulated signal of frequency .01 Hz to 50 Hz. The legend indicates the trace and the optical power on the minus and plus ports for each trace. The plot is not normalized, as it merely illustrates the progression of the peak response, 3dB High-pass filter point, and response amplitude. Note the 3dB point ranges from 20-60 mHz and a peak response of 1-7 Hz based on power level, which better matched to the chemical sampling frequencies we have been using (2-5 Hz).

With the effective bandpass characteristics of the optical system understood, optimum conditions can be found for a given wavelength, and valve design to maximize the signal of interest. Typically, this is more easily done by running a series of tests and noting the trend in S/N when the signal is held at a fixed amplitude.

5.9 TIR Noise Sources

Since the TIR interface is sensitive to minor perturbations in index, we must consider any perturbation at the surface a potential noise source. The site of such perturbations occur in two regions of interest. First, an index change in the atmosphere occurs due to pressure, temperature, or the introduction of different vapors including water, in the form of humidity. Second, physical changes in the polymer itself yield an optical path length and index change. These changes also include thermal

fluctuations, differential pressures in the atmosphere, and possibly even a kind of mechanical shot noise dependant on the number and mass of gas particles bombarding the transducer.

Since fluctuations in refractive index may be a substantial noise source due to the TIR phase response, we present the relevant subset of the modified Edlen formulae [41].

$$(n-1)_{\lambda} = 10^{-8}\left[8091.37 + \frac{2,333,983}{130 - \left(1/\lambda(um)\right)^2} + \frac{15,518}{38.9 - \left(1/\lambda(um)\right)^2}\right] \tag{5.53}$$

$$(n-1)_{tp} = \frac{(n-1)_{\lambda}\cdot P(Pa)}{93214.6}\cdot\frac{1+10^{-8}\left(0.009,876\cdot T(^0C)\cdot P(Pa)\right)}{1+0.003,661\cdot T(^0C)} \tag{5.54}$$

These equations approximate the refractive index of air based on pressure, temperature, and wavelength. We exclude humidity and CO2 based on W.T.Estler's paper discussing precision interferometry in air concerning uncertainties in the refractive index due to pressure, temperature, and humidity [42]. He found that while temperature and humidity indeed contributed to index perturbations, pressure was the largest contributing factor.

A few experiments were conducted to determine the temperature and pressure fluctuations of the carrier flow immediately out of the guiding tube, as this would be similar to the conditions the polymer experiences. The temperature was measured using a YSI 43141 glass-bead thermistor monitored by an HP34401A multimeter set to measure the resistance. Granted, a thermistor produces heat and therefore its temperature will be less stable if the airflow is at all turbulent. Hence, this situation can only be considered a worst case limit. After letting the system stabilize for a few minutes, the fluctuation in temperature, when isolated from air currents, was <0.9 mK/s, while air conditioning currents caused local fluctuations up to 1.8 mK/s. The N2 carrier, with a flow rate of 300ml/min, was then monitored, which yielded fluctuations up to 0.9 mK/s, corresponding to 530 fm, or mean drifts up to 0.12K/min corresponding to a displacement of 70fm. Since the fluctuations corresponded to the stability of a thermistor in a closed chamber control experiment, the temperature stability may be superior to this, but we can't measure it.

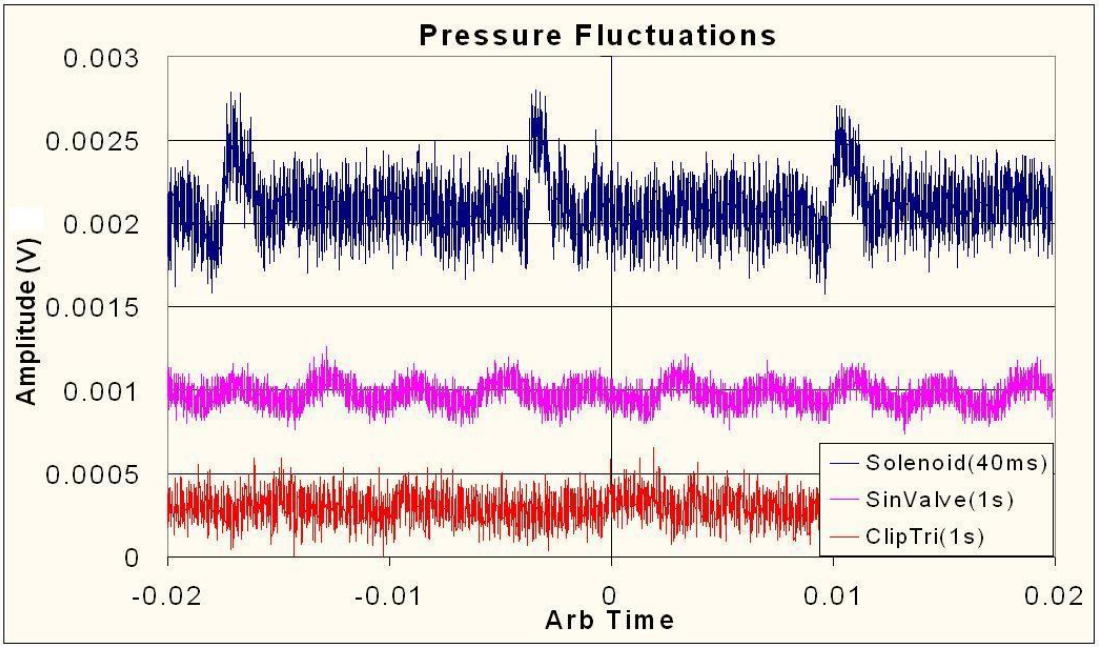

Figure 5-15: Pressure Fluctuations of the 3 fluid valves. The time axis is appropriate for the solenoid switch to show the pressure spikes of ~3ms width and 0.6mV(29Pa) amplitude. The solenoid and clipped sin valve have pressure noise around 0.1mV(5Pa) and <0.1mV respectively. Conditions for all three were nearly identical for a 300ml/min flow rate with similar tubing loads at their outputs. The width of the sample period is given next to each label in the legend. The solenoid valve switch frequency was increased to get multiple spikes on a single trace. The conversion for the vertical scale is 48Pa/mV and DC levels were offset for clarity.

To monitor pressure, an omega PX26-001DV differential pressure sensor was used with one port measuring the pressure in the tube and the second open to air. The tube was connected to each valve, which ran under normal conditions. The pressure fluctuations due to the valve were measured immediately upstream. We measured a maximum 28Pa fluctuation from the flow tube for the solenoid valve while the sinusoidal valve caused fluctuations less than 5Pa. Using the modified Edlen formula should yield a displacement noise of approximately 52fm. While these noise levels are superior to the shot noise level, they aren't as significant as are other sources.

Next, we look at the pressure and temperature effect on the surface refractive index. A pressure perturbation can be approximated using the strain-optic coefficient p_s and its contribution to index [43].

$$\Delta n(x,t) = -\frac{1}{2} p_s n^3 s(x,t) \tag{5.55}$$

where n is the refractive index at STP and $s(x,t)$ is the strain applied. Table 5-6 shows a couple relevant strain-optic coefficients at the TIR interface.

Material	**Young's Modulus(E)**	p_{11}	p_{12}
"Glass"	65 GPa		
Polystyrene (633nm)	3 GPa	0.3	0.31

Table 5-6 Material properties of glass and polystyrene, acting as an "average" representative polymer.

To determine the degree of compression due to the pressure fluctuation, we solve for strain using Young's Modulus of the material E and the applied stress.

$$E = \frac{stress}{strain} = \frac{F/A}{\Delta L/L} \tag{5.56}$$

where A is the area of interrogation over which a force F is applied and ΔL is the resulting compression of an initial length L. This equation assumes the force applied is uniformly over the interrogated area because otherwise a spatial dependence will arise which we ignore for simplicity.

$$\Delta L(x,y) = \frac{L * P(x,y)}{E} \tag{5.57}$$

The index variation with compression is assumed to occur uniformly through the polymer depth, for a sufficiently thin polymer in steady state. thus, rather than just a TIR interface perturbation, an additional phase shift due to optical path length compression is picked up by the laser. However, as we found earlier by examining the sensitivity of the TIR interface to a monolayer and the associated index change, we find that the index change dominates. For pressure fluctuations of 24 Pa compressing polystyrene, we see an effective displacement noise of 3fm from the phase shift at the polymer, which is insignificant compared to shot noise. The refractive index change arising from equation (5.55) is $4 \cdot 10^{-9}$ RIU, or $1/50^{th}$ of the systems experimental sensitivity.

We can conclude that the noise due to pressure fluctuations, under the circumstances measured here, are insignificant compared to other noise sources. However, altering the flow conditions by increasing the flow rate or changing the modulation frequency or the valve used can make the noise at the TIR interface more severe. It is likely that the above experiment is a gross

underestimate of the pressure noise since the chemical LOD for PV3 corresponds to roughly 3 times the displacement LOD. However, with the pressure sensitive equipment on hand, experiments to find the actual fluctuations at the surface were not easily attainable.

5.10 Long Term Stability

The ultimate measure of performance for an electronic or optical nose is limit of detection. Since this limit is effectively a signal-to-noise question it must be given in terms of $ppm\sqrt{W/Hz}$. Since the limitation associated with adjusting the optical power to improve the limit of detection has been discussed, this section focuses on the measurement bandwidth or integration time. The signal-to-noise measure inherently includes integration time. A single 5 second averaged measurement must be compared to the LOD curve associated with data collected at that integration time. From the curve one can ascertain how this measure corresponds to the likelihood of a false detection. By averaging many such samples, or increasing the averaging time, the variance of the mean, between points, decreases improving confidence in the content that is associated with the signal measured.

To this end one should be able to integrate to infinity. The resulting limit of detection will depend solely on how long one is allowed to sample the air before making a decision, to the limit of the dominant noise term. Unfortunately, most systems will experience a drift in signal level, even for a blank, such that it will appear that the content level is changing when it is not. Therefore, there is a fundamental limit to integration time beyond which there is no further improvement.

An Allan variance of the system output can be used to determine the optimum time for integration. The Allan variance measures the variance of the mean over an integration time or window.

$$\sigma_A{}^2(k) = \frac{1}{2m}\sum_{s=1}^{m}\left(A_{s+1}(k) - A_s(k)\right)^2$$
$$A_s(k) = \frac{1}{k}\sum_{l=1}^{k} x_{[(s-1)k+l]} \tag{5.58}$$

where the total number of data points x taken is given by N and k is a subset such that there are $m = \lfloor N/k \rfloor$ subsets or windows of integration and s denotes the particular window where $s = 1..m$ and

$s \in \square$. The optimum integration time is the point at which the variance of the mean is minimal before it begins to rise again.

In an Allan-variance plot, a minimum will be dictated by the 1/f noise, which is essentially a flat-band variance since increased integration time merely integrates more 1/f noise, yielding no gain. The variance due to White noise drops off with increased integration time while drift increases with longer integration [44]. Therefore, the minimum variance is often established by the 1/f noise before drift becomes dominant. Figure 5-16 show two Allan variances for the same system.

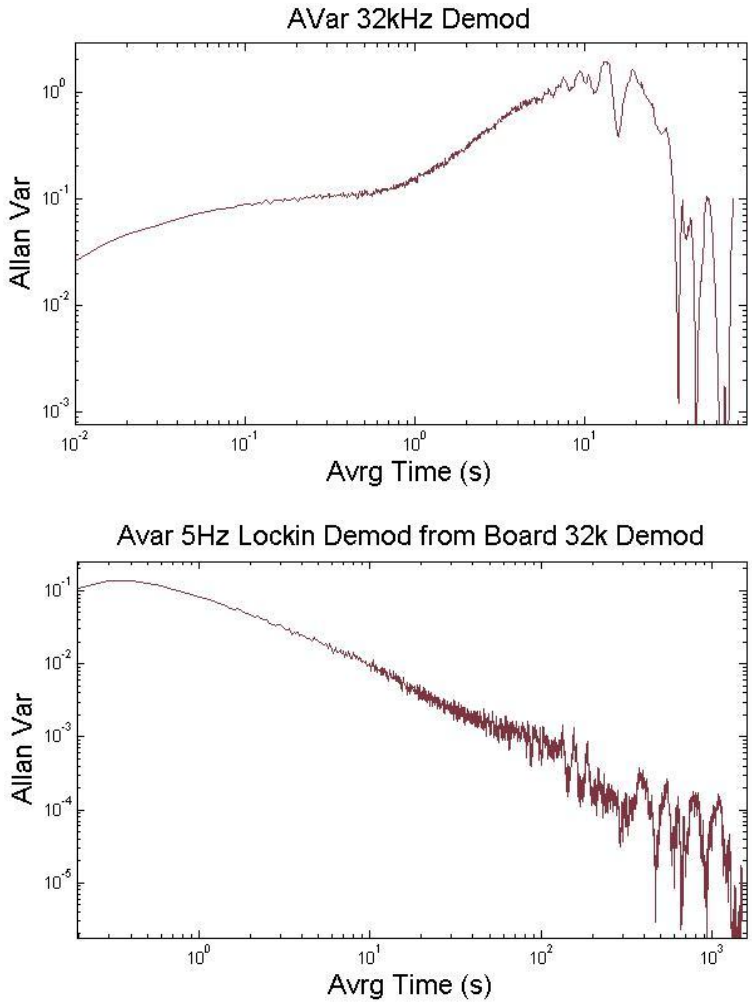

Figure 5-16**:** Top: Alan Variance out of the high frequency reference demodulator (PV4) and Bottom: out of the subsequent signal frequency demodulator. The high frequency demodulator shows that increasing integration time will result in a loss in S/N, and thus the time constant should be in the 10's of milliseconds, which is appropriate for a 32kHz demodulation. Granted, a time constant over 40s would yield similar stability though it would also eliminate a 5 Hz signal. The subsequent low frequency demodulation indicates that the system integration time constant should be as long as the user is reasonably willing to wait. In this case, the user considered 50 minutes well past reasonable. In

both cases the Alan variances becomes noisier for longer integration times due to poorer statistics arising from fewer averaging windows over the 10,000 sample set.

From these plots we have already established the first demodulator's integration time at 20ms and set the integration time constant to 5 seconds, though data is reported in $m/Hz^{1/2}$, already accounting for the improvement. At this point there is little further to do to the system to improve its fundamental detection limit. A displacement sensitivity of $180fm/Hz^{1/2}$ isn't too shabby and yet somehow other technologies are doing better than Prism Version 3's $2ppm/Hz^{1/2}$ ethanol detection limit. So what can be done to improve the LOD?

We have already discussed the polymer content-to-signal transduction gain and the fact that a transducer that achieves higher gain can improve sensitivity. The selectivity of a transducer helps to improve certainty in a detection event. Metal-oxide sensors achieve better gain by running at high temperatures (300^0C) with a potential bias to assist single oxidation events and improve the signal over background gas interactions [45]. Biological sensors using antibodies and biological amplification mechanisms work well and theoretically would work for our system. However, they would have to maintain viability suspended on glass and be made sufficiently thin, which is unlikely to be a problem as most surface-plasmon-resonance (SPR) bio sensors are a single layer of binding agent and antibody [46].

There are a handful of techniques that can be used by nearly any technology to improve sensitivity, including increased volumetric flow sampling, repeated comparison against a reference gas, or the ever popular preconcentration. Increasing the flow rate has its own problems associated with noise due to a higher flow and the question of whether a higher flow rate actually increases the exposure counts per second at the transducer. If the transducer isn't very "sticky," as characterized by a low ratio in adsorption:desorption rates, then this won't improve the signal significantly. If the transducer is very sticky, then the steady state content on the surface of the transducer might increase assuming the increased airflow doesn't perturb the sorption kinetics appreciably.

A comparison against a reference gas is done automatically as our design requires it to prevent the photorefractive from adapting to a constant signal. In this way, the photorefractive already helps to adjust for some drifts in the system that manage to influence the phase of the optical beam.

However, we still benefit from the sampling because our signal is a comparison between a flow with an analyte and a blank flow. Therefore, drift noise before detection is minimized, unless it is on the order of the sample frequency (2-10 Hz).

Finally, there is preconcentration [47]. This is a term used to describe the collection of analytes over a long period of time with typically high flow rates followed by the rapid desorption into a standard sampling rate which is fed immediately to the olfaction system. Preconcentration is typically performed using a tube packed with a sorbant, often of an activated carbon variant, which at or below room temperature (or the desorption temperature) has a high affinity and capacity for the analyte of interest. At hotter temperatures (above the desorption point), the sorbant exhibits orders-of-magnitude less affinity and capacity for the analyte. Therefore, to the limit of the capacity of the packed sorbent volume and the time one is willing to wait/integrate, a preconcentrator can sample the air collecting analyte. Afterwards, through flash heating, the collection tube can rapidly desorb the collected sample into a small volume of air, which is passed over the sensor system. Ultimately however, preconcentration can perform no better than a system that has no drift and can run for the same period of time, under the same conditions as the preconcentration sampling.

Using a less-than-ideal preconcentrator, one must calculate two critical values and compare them to the expected analyte volume to be sampled. The first is the breakthrough volume associated with the analyte at the desorption temperature, which is important because this is the volume of analyte that cannot be retrieved on desorption, assuming the preconcentrator was completely purged. For one that is frequently cycled, this low-temperature breakthrough volume should not be a problem as it will be preloaded with said volume. The second is the volume associated with breakthrough at the cold adsorption temperature. In actuality, the preconcentrator charges similar to a capacitor and thus waiting to load the sorbant to capacity yields decreasing returns with time. Thus, an effective impedance match is necessary for loading time such that the system performance diverges minimally from the normalized limit in $ppm\sqrt{W/Hz}$ that would be obtained if drift and volumetric airflow were not an issue.

The issue of preconcentration comes down to two key points. First, most preconcentrators operate at over 200^0 C to get adequate desorption of small concentrations (sub ppm). This requires a

system that is robust enough for the interferometry to handle 200^0C. Since the system is already sensitive to temperature, the large gradients this would impose would limit sensitivity and potentially damage the photorefractive without adequate temperature isolation and regulation to better than a few mK. Second, preconcentration was experimented with at room temperature, and while data was found to show that it had the potential to work, significant development would be necessary to optimize the sorption dynamics, sorbant chemistry, and other impedance matching problems that would enable improvement beyond simply integrating for longer periods. Preconcentration is discussed in a bit more detail in section 6.5. We conclude from the above highlights that preconcentration would require more work than was reasonable to ensure the entire optical system could withstand, or be sufficiently isolated from, the large thermal fluctuations.

5.11 The Piezo-Grating Modulator

A Piezoelectric stack transducer, or actuator, is a series of thin sheets of material that expand or contract along a primary axis based on an applied electric field by the piezoelectric effect. Stacking these allow for larger displacements to be achieved (1-2 um) with low voltages (<150V). This system employs such a device to translate the grating beam splitter along the surface of the prism to phase modulate the diffracted beam. To accomplish this phase modulation, the grating's position oscillates at the reference frequency ω_R, while the low-frequency small-amplitude calibration signal at ω_S is often added for displacement LOD measurements. For further reference, table 5-7 gives the specs of the PZT's used in the final two prism systems.

	Prism V3: PL 022.30	**Prsim V4: PL033.30**
Dimensions	2x2x2mm	3x3x2mm
Capacitance	25nf	160nf
Displacement (@100 V)	2.2um	2.2um
Resonant Frequency	>300 kHz	>300 kHz
Maximum Operating Temp.	150^0C	85^0C

Table 5-7: PZT properties. PZT's are both manufactured by PICMA.

At low frequencies the Piezo is expected to behave according to spec. (i.e., its displacement will be a direct function of its voltage). However, at higher frequencies, the piezos exhibit resonances and roll offs to their response, which is further compounded by the resonances of the preload structure and when pushing an external mass. Most PZT's are given specifications of a resonant frequency (*fo*) without an external mass. The base resonant frequency of a piezo can be calculated using its mass *m* and spring constant c_t.

$$fo = \frac{1}{2\pi}\sqrt{\frac{c_t}{m/2}} \tag{5.59}$$

However, calculating the resonance with an external spring flexure(non-ideal) and additional mass is nontrivial.

Since displacements up to $\lambda/4$ are required for maximum signal, a resonator is necessary to improve the impedance matching between the high frequency reference driver and the PZT. To drive 532 nm/4 with a response of 2.2 um/100V, a peak-to-peak voltage swing of 7.5 V is necessary. The supply current needed to charge the capacitor at this rate would then be estimated as a factor of π over that necessary to drive a triangle wave of the same amplitude.

$$I_{pp} \approx 2\pi f \cdot CV_{pp} \tag{5.60}$$

In the worst case scenario, to drive up to 20 Vpp at 45 kHz for a 160 nF load, the supply must be able to handle up to 0.9 Amps. Hence, without a beefy driver, a resonator is necessary.

When discussing resonators, the quality factor (*Q*) is used, and represents a general measure of the ability of a system to retain its oscillatory energy. More specifically, *Q* is defined as the ratio of the systems stored energy W_s to that lost per cycle ΔW .

$$Q \equiv \left.\frac{W_s}{\Delta W}\right|_{1\ cycle} \tag{5.61}$$

As it may be difficult to measure the energy stored and lost in a system, one can look at the resonant frequency f_0 and bandwidth Δf to arrive at the *Q*.

$$Q = \frac{f_0}{\Delta f} \tag{5.62}$$

The Q of a system is typically used to define the tendency for a system to oscillate, whether or not it was designed to do so, where $Q > 1/2$ indicates an under-damped system. The quality factor is of importance since many components, mechanical or electrical, may be exposed to impulse or oscillatory excitations that cause sympathetic ringing which acts as noise. Indeed, even thermal vibrations, airflow, or power supply ripple may lead to undesired oscillatory noise. To examine the role of noise in an oscillator, we must discuss the resonator itself and then the signal-to-noise ratio measured when using said resonator.

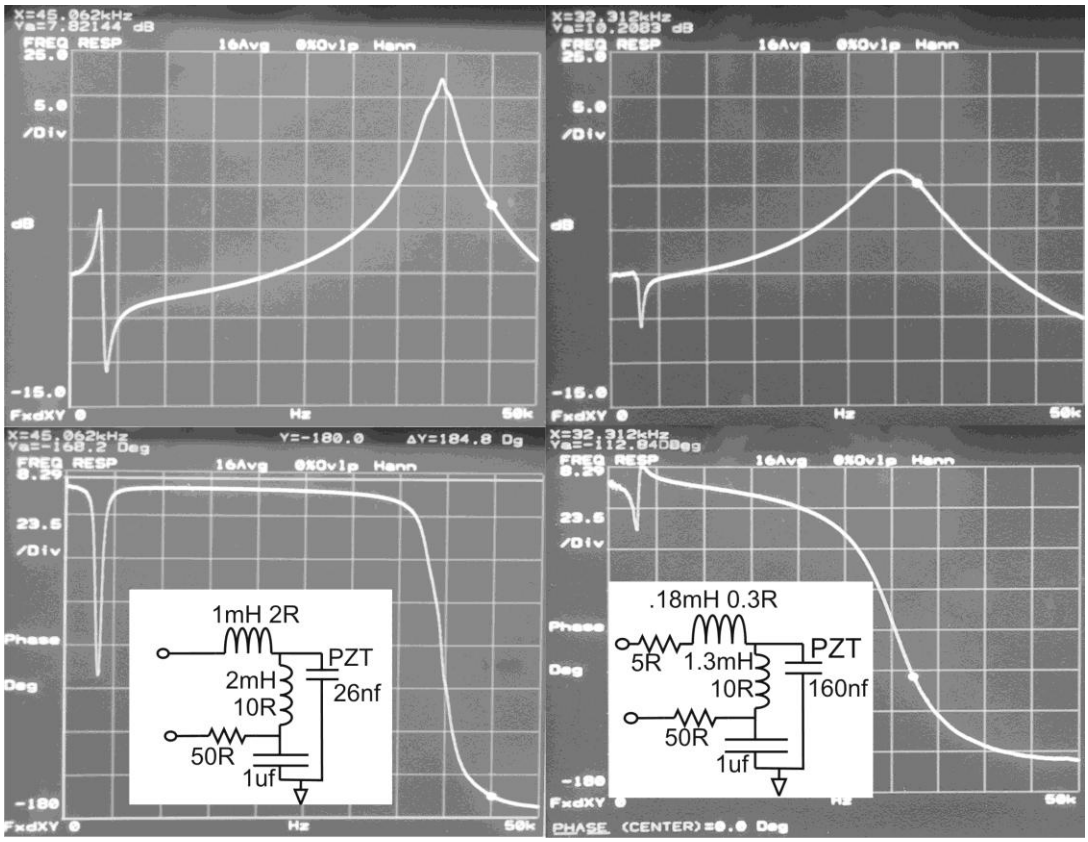

Figure 5-17: Amplitude (Top) and phase (bottom) response of a resonant PZT driver for prism systems V3 (Left) and V4 (Right). The values are different because the piezo has different capacitances and the desired driving frequency was 45kHz Q=13(left) and 32kHz Q=4.3(right). The 45kHz drive was chosen to be near the tail of the PZT phase curve while still receiving adequate impedance matching. The 32kHz was dictated by a crystal oscillator and the resonance of the tank was chosen to be close enough to allow adequate amplitude gain at the sacrifice in phase slope.

The resonant tank circuit is designed to accomplish two tasks. First, it improves the impedance matching to allow more power to be coupled at the desired driving frequency without the need for a beefy oscillator. Second, the circuit acts as an adder, allowing a small low-frequency signal to be added onto the high-frequency reference oscillator, where the low frequency signal is attenuated heavily to simulate a small signal for calibration purposes. Because this tank has the added bonus of acting as a decent notch filter, the tank can be driven with a square wave, say from a quartz crystal oscillator, and will output a nearly clean sine wave. However, we wish the reference oscillation to have high phase and amplitude stability, which requires a tradeoff between operating on resonance (higher amplitude stability) and away from resonance (higher phase stability).

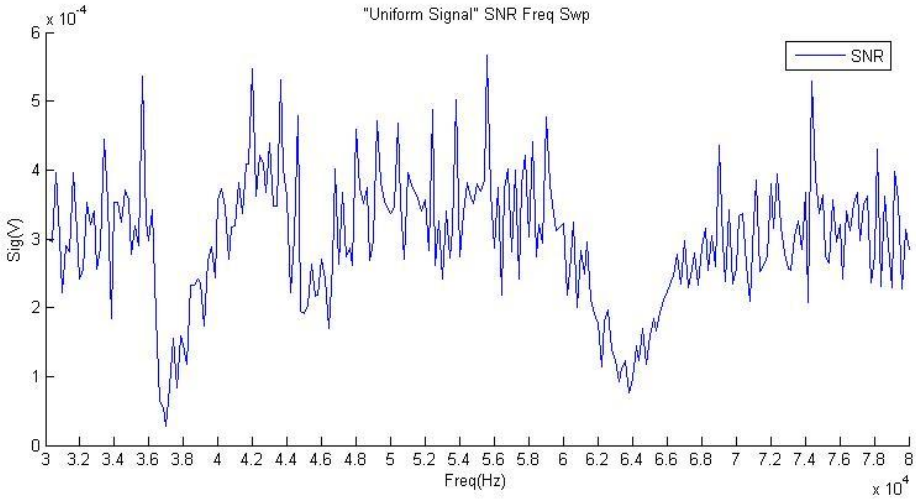

Figure 5-18: S/N measured at the sum/difference frequency during a reference frequency sweep using PV3. If we compare this trace to the electrical resonance curve, we see that the optimum S/N is achieved off resonance by about 4kHz.

Figure 5-18 shows the S/N response for a constant modulation depth while the reference frequency is swept above resonance of the PZT oscillator. The figure shows that the optimum S/N is not achieved at resonance but rather 4 kHz away from the 39 kHz resonance. Therefore, the phase stability may be of critical importance for the resonator since any such phase noise is carried through the system affecting the detection limits. The other dips in this figure are due to mechanical resonances of the PZT flexure spring..

To describe the behavior of the PZT electrical resonator, we examine the transfer function, which is simply the voltage over the PZT divided by the driving voltage.

$$H(\omega) = \frac{L_2}{4L_1L_2C_{PZT}\omega^2 - L_1 - L_2} \tag{5.63}$$

where L_1 is the top inductor, L_2 is the parallel tank inductor, and C is the capacitance of the PZT. The resonant frequency is found by setting the denominator to zero.

$$\omega_0 = \frac{\sqrt{L_1L_2C_{PZT}(L_1 + L_2)}}{L_1L_2C_{PZT}} \tag{5.64}$$

Small correction factors for the 1 uf cap and the resistances of the elements will adjust the center frequency and the *Q*, but we leave them out for simplicity.

Having introduced the resonator we now examine the driver. For the reference oscillator we use an Agilent 33250A. The portable system was deigned to use a 32 kHz quartz crystal oscillator SG-3030 JC with square wave output driving an APEX PA02A amplifier, but the oscillator was not used for LOD experiments in favor of the Agilent. The Agilent showed spectral purity of >116dB separation between the peak and $\pm 5Hz$ while the quartz oscillator had a minimum purity of 102dB. Both measurements were made using a Standford Research signal analyzer (model SR785) set to a 1Hz bandwidth. Using the oscillator spectral purity we can find the ratio between the driving modulation depth and the effective noise modulation depth.

$$\Delta x_{\pm 5Hz} = \delta_R \frac{V_{\omega_R + \omega_S}}{V_{\omega_R}} \tag{5.65}$$

where $\delta_R = 0.1\lambda$ and the ratio is that of the carrier to the effective noise sidebands.

$$dB = -20\log(V_{\omega_R+\omega_S} / V_{\omega_R})$$

The effective modulation depth of the sideband noise is 85 fm for the Agilent while the output of the Agilent-Oscillator pair is 170 fm and 210 fm for the 532 nm and 660 nm systems respectively. The crystal-resonator pair contribute the worst noise, causing an effective displacement noise proportionate to 530 fm at the 5 Hz sidebands of the carrier.. This result is rather interesting in that it suggests that not only is the linewidth the limiting factor in sensitivity, but the purity of the reference modulation is also a significant factor. Experimental results lend more weight to this conclusion in

that while decreasing modulation below 0.1 λ does not improve the signal to noise, increasing it makes the S/N worse, as seen in figure 5-19.

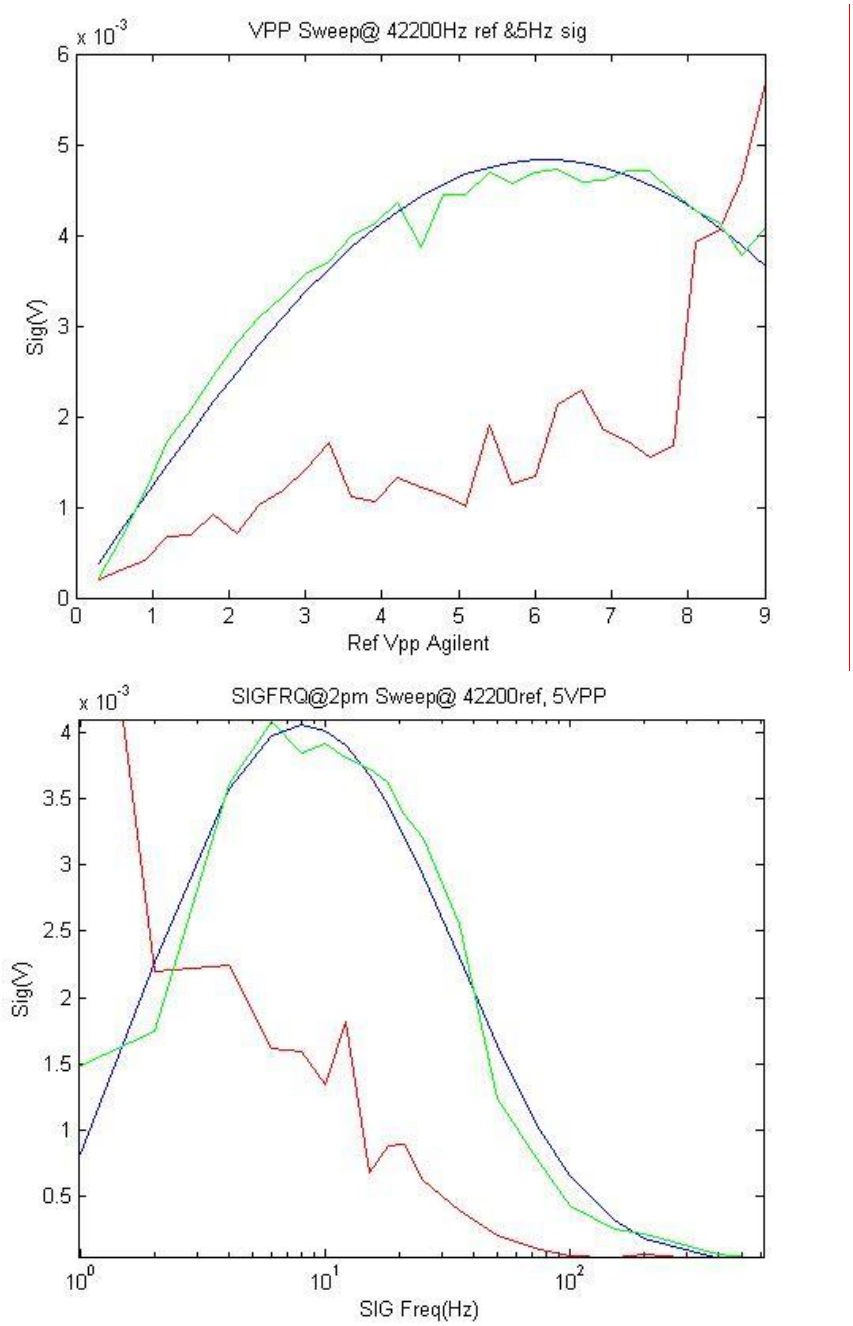

Figure 5-19: Top: Signal (with theoretical prediction overlaid) and noise curve as the amplitude of the reference generator is swept. The peak signal occurs near the anticipated 6.6V while the noise appears to be linearly increasing. It is for this reason that the system is never run above 1/3rd of the voltage that would yield the peak signal as the S/N suffers greatly for modulation depths above 0.1λ . The fit curve was adjusted using only the modulation depth and an amplitude scaling factor applied to equation (3.9) Bottom: Signal and noise curve with increasing signal frequency. This plot indicates that the optimum

S/N occurs for signal frequencies around 20-100 Hz, where low frequency noise sources are minimal. The high frequency cutoff is due to electronic low-pass filtering of the demodulated signal.

5.12 Conclusion

Indeed, after many design iterations to reduce various noise sources, the reference system (Prism V3) is limited by several noise sources including laser phase noise, fluctuations of the fundamental ω_R leaked through the photorefractive, and index perturbations at the TIR interface during vapor exposure. The portable system (Prism V4) is primarily limited by noise sources associated with the laser itself including phase and amplitude fluctuations to a more severe degree.

Recommendations for future work would be to improve the linewidth and stability of the 660nm source and continue to redesign the valve to achieve smaller mixing volumes, lower pressure fluctuations, and higher rotation frequency stability. Achieving the 50-70dB suppression reported by Hobbs in the auto-balancing circuit would also help to mitigate laser amplitude noise. If a prism design can be found that achieves all of the desired attributes (Brewster windows, near critical TIR, silvered internal mirrors, optimum two-beam coupling) and also manage to balance the dominant dispersion effects against each other, then the linewidth/wavelength sensitivity of the system would be further reduced improving the displacement sensitivity.

Chapter 6 Chemical Delivery

6.1 Introduction

This chapter will discuss the means of chemical delivery and detection. Since the analyte of interest is a volatile organic chemical, a means of generating and delivering a precise vapor phase concentration is discussed. After the technical details are presented, along with tradeoffs and solutions, we will move into a brief discussion on chemical transducers, the elements that convert a content signal into an optical phase signal. Though polymer transducers are one of the mainstays for array based detection, we also discuss alternative transducers and the feasibility of incorporating non-polymer sensors.

When performing a chemical sampling test any uncertainty in the actual content level further compounds the measurement and calibration uncertainty of the system. Therefore we require a means of generating and sampling precise concentrations of analyte, which introduces its own mess of problems. The means of producing a concentration vary from bottles, to evaporation, to diffusion tubes. All of these may work well under certain circumstances and become highly unreliable under others. Faithfully delivering the sampled concentration to the transducer while minimizing sampling memory, due to contamination of the sample lines, is necessary for reliable small-signal detection. Additionally, the response of the system may change over time due to analyte accumulation/loading (inhibition), the permanent chemical degradation of sensitivity (poisoning), or a permanent mechanical degradation of the transducer.

6.2 Dilution for Delivery

One fundamental problem with calibration and testing of an olfactory system is the reference against which its performance is measured. Measuring displacement is relatively straightforward but measuring, let alone generating, a flow of uniform concentration is less straightforward. There are a few standard methods in the field for generating an analyte concentration, which we briefly introduce.

The first is to place a known mass or volume of analyte into a chamber filled with nitrogen or air, give the sample sufficient time to evaporate and uniformly diffuse, and then sample the volume [45]. This process has potential problems with the material of the container, which may absorb some of the analyte changing concentration, desorb analyte from prior experiments, and, as the volume is finite, the container must decrease volume during sampling or take in air to displace the volume removed for sampling (slowly decreasing the concentration).

The second is to purchase pressurized gas bottles with fixed concentrations of a known analyte on a carrier gas. These are often calibrated to a few ppm and can be easily diluted but can't be easily concentrated. Bottles are one of the mainstays in precision references as they are produced by various labs. However, bottles are heavy, relatively expensive to keep several concentration levels on hand, and potentially hazardous, requiring extra safety and room.

Finally, the third is to actively evaporate from a reservoir which typically takes on two forms: permeation tubes and evaporation canisters. The permeation tubes are references that, for a fixed sealed volume containing the analyte in liquid form, the semi-permeable barrier allows analyte diffusion at a known rate as long as there is liquid in the reservoir. A flow is required over the membrane to pick up the analyte, though increasing and decreasing the flow rate will respectively decrease and increase the concentration from the calibration point. We purchased a couple G-Cal permeation tubes which we used to verify the optical nose response at 10ppm ethanol against evaporation, as well as to verify that the PID can actually detect 100 ppb, though only as a differential measurement.

Evaporation tanks are operated on the principle of bubbling, headspace, or diffusion. In bubbling, the carrier inlet tube is submerged, often yielding parts-per-thousand and higher concentrations. In the headspace method, the concentration is dictated by the height of the inlet and outlet tube over the liquid level and generates concentrations in the ppm, depending on flow rate and height. The sample concentration from headspace diffusion is dictated by the height of the tube over the volume and the length of said tube until it meets at a T with the carrier. For greater detail on these methods we refer the reader to the Handbook of Machine Olfaction [48].

We originally used the diffusion headspace to generate sub-ppm levels and then dilution to achieve 10's of ppb levels, but abandoned this procedure as it proved unreliable if the flow over the diffusion tube was not constant. It was theorized that pressure spikes, caused by switching with solenoid switches, would cause a periodic exchange of gas within the diffusion tube, as the volume was pressurized and depressurized, resulting in concentrations above the steady state diffusion method. For this reason we stepped back to the headspace method and/or used short diffusion lines relying on flow dilution for lower-concentration generation. Having discussed the methods of concentration generation, we move on to measuring the concentration level.

6.3 Vapor Concentration Generation

The concentration of a vapor in a carrier gas is often given simply as a ratio of volumes or the partial volume of the analyte gas divided by the total gas volume. To determine vapor concentration we take a moment to review some basic principals associated with ideal gasses. The ideal gas law states $PV = nRT$, or that the product in pressure and volume is equal to that of the temperature, the number of moles *n*, and the Gas constant (8.314 m^3Pa/K·mol). The Avogadro constant defines the number of particles or molecules of an ideal gas in a mol as $A_{\#} = 6.022e^{23} / mol$. The atomic weight of an atom is an effective measure of the mass in grams contained within a single mol, from which we can estimate the molecular weight, M_{mol}, by the sum of the atomic constituents. Finally the volume occupied by an ideal gas is 22.4 L/mol at STP or approximately 27 L/mol in Boulder Colorado.

The concentration *C* from an evaporating source under steady state evaporation conditions can be determined by the loss in mass ΔM over time Δt and the total carrier flow *F* (m^3/s) in which the vapor is mixed.

$$C = \frac{\Delta M}{M_{mol}\Delta t}\left(\frac{27L}{mol}\right)\frac{1}{F} \tag{6.1}$$

Similarly, by the ideal gas law, we find the ratio between particles is proportionate to the ratio in volume, $n_1/n_2=V_1/V_2$, (a similar expression exists for pressure). Using this equation we simply

measure the mass loss over a period of time at fixed flow rate and calculate the delivered concentration on the carrier gas.

To verify the concentration calculated by mass loss, a Phocheck 5000+ photo-ionization detector (PID) was acquired. This device samples at 300 ml/min and is completely indiscriminate, but will report a concentration of analyte if told what it being observed. While the response time is roughly a second, this instrument can also be used to deduce the concentration of a modulated source where the signal reported is the average concentration multiplied by a scaling factor.

$$C = S_F\left(\frac{C_1 + C_2}{2}\right) \tag{6.2}$$

Where S_F is the scaling factor based on flow rate which is 1 for flows over 300 ml/min and otherwise proportionate to 300 ml/min divided by the flow rate the intake of the PID. The scaling factor assumes that the inlet to the PID is not sealed such that excess flow can escape and is not forced through the PID or insufficient flow is augmented by the atmosphere.

Finally, to alleviate doubt with yet a third reference, we acquired a couple of G-cal permeation devices from VICI metronics that release ethanol at a permeation rate P_R of 221 ng/min and 13.6 mg/min. These devices were designed to operate at 25^0C and simply required a carrier flow be passed through their ¼" Swagelock T-connector. To calculate the effective concentration the permeation rate is multiplied by the molar constant K_m and divided by the flow rate F.

$$C = \frac{P_R K_m}{F} \tag{6.3}$$

With a flow rate of 300 ml/min and a molar constant for ethanol of 0.531 the effective concentration of the devices is 400 ppb and 24 ppm respectively. For calibration we introduced these flows into a dilution flow and achieved dilutions of a factor of ten. With this range of concentrations we verified the systems sensitivity to similar concentrations achieved from evaporation and also the performance of the PID. We only used the permeation devices for a short time as they were rated for 6 months to a year and their primary purpose, to ensure our accurate generation of ethanol vapor, had been fulfilled.

6.4 Concentration Dilution

Since the concentration generation can be sensitive to changes in flow rate, and to ensure evaporation was as stable as possible, we decreased concentration primarily by dilution. The evaporation apparatus used incorporated several flow meters/regulators, a switching valve to alternate between test and reference gas, and evaporation canisters as depicted in figure 6-1. The canisters were originally 125ml flasks, but to decrease headspace volume were reduced to 4-dram vials. The reason we didn't simply flow the entire dilution flow rate over the bottles was that we had noted that at higher flow rates, the turbulence and disturbances created at the surface of the fluid in the evaporation canister make the concentration levels unstable and unreliable.

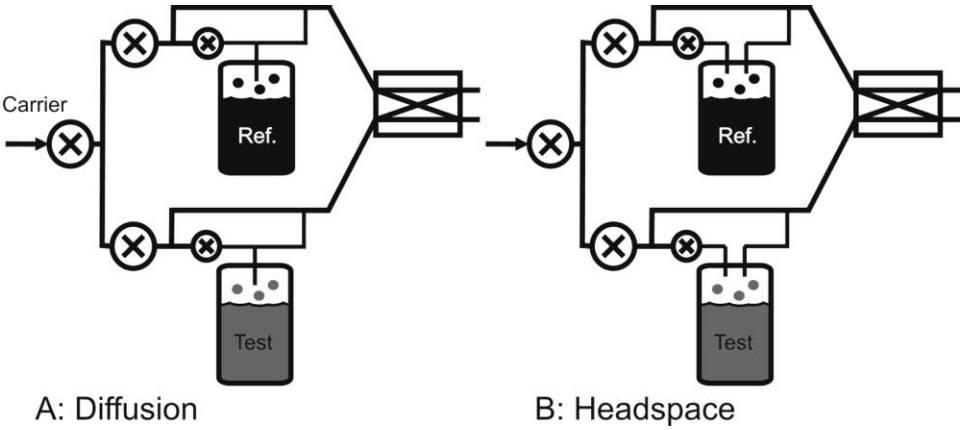

Figure 6-1: Our basic dilution scheme using A: a diffusion sample injection and B: headspace sample injection.

We show both a diffusion and evaporation injection method. Diffusion relies on the continual evaporation in the sample canister, which, at a fixed headspace, begins diffusion through a diffusion tube. The rate of diffusion *S*, which is comparable to P_R from the permeation tubes, is given as,

$$S = \frac{D \cdot M \cdot P \cdot A}{R \cdot T \cdot L} \ln\left(\frac{P}{P - p}\right) \tag{6.4}$$

Where *D* is the diffusion coefficient, *R* is the Gas constant, *T* is temperature, *L* is the length of the diffusion tube, *A* is the tube cross-sectional area, *M* is the molar mass, *p* is the partial pressure of the analyte and *P* is the total pressure in the sample tank [48]. The measured concentrations achieved from

diffusion were in the 10's to 100's of ppb. However, we abandoned diffusion injection early on due concentration uncertainty issues that will be discussed shortly.

The headspace method involves a low-flow carrier displacing the headspace vapor inside the canister pushing it into the output tube. Though this method will gradually change concentration until equilibrium is reached [48], we circumvent this by letting the sample reach equilibrium over half an hour. The time to reach equilibrium depends on the flow rate, temperature of the carrier gas vs. the analyte gas, and the mass of the analyte liquid. To minimize time, we use a 4-dram vial filled with roughly 3cc of ethanol, the outside of the glass vial is insulated with Styrofoam, and a headspace of 1-2 cm was used.

The sample flow rate was typically held to 20-50mL/min while the total flow per line was around 1 L/min, giving a dilution ratio of up to 50:1. The delivered concentrations using headspace were between 1 ppm and 50 ppm. For the few experiments where we hoped to obtain higher concentrations, we simply increased the sample flow until the desired concentration, measured by the PID, was achieved.

6.5 Preconcentration

6.5.1 Preconcentration Motivation

After spending time discussing vapor dilution to attain small concentrations and the efforts taken to ensure the accuracy of this measurement, it is only appropriate that we now discuss preconcentration, or a means of amplifying small concentrations to detectable levels. Preconcentration in its basest of forms is the long-term integration or collection of analyte molecules followed by rapid release of said concentrate such that, in a standard measurement, the instrument sees a substantially amplified signal. Before discussing the means of preconcentration we first motivate its existence in the field of chemical olfaction.

There are four primary reasons to use a preconcentrator. First, the measurement system suffers from long term drifts which would manifest as drifts in the signal mean. If this drift did not exist, greater sensitivity could be obtained simply by letting the system integrate for a longer period of

time. A longer integration would lower the variance of the mean, thus pushing the limit of detection down so that the detection limit is proportional to $\tau_{\text{int}}^{-1/2}$. The preconcentrator allows for a collection time that is limited by the concentrator's capacity and the sample rate. After loading, the preconcentrator can rapidly release all collected particles into a volume of air that is easily sampled within the optimum integration time of the system.

The second possibility is that the system that will not increase sensitivity with flow rate. In some systems a higher flow rate may increase the sensitivity if the detector has a strong affinity for the analyte such that sorption or uptake has a substantially higher rate constant than desorption. Factors that can limit the flow rate include an increase in noise with higher flow rates, transducers that have very low sorption to desorption rates, or transducers whose size is sufficiently small that there is significant shielding of the transducer by a fraction of the airflow such that increasing airflow does not increase exposure. Preconcentration allows the sample tube to be loaded at flow rates substantially higher than those capable by the system followed by desorption flow rates that are within system operational tolerances.

Third, the preconcentrator matrix is often made of transducer materials such as carbon black, activated carbon, and porous polymers. Just as each transducer has its range of selectivities, so does each preconcentrator tube. Therefore, if one has a transducer that is sensitive to two chemicals but a preconcentrator that will primarily only collect one of them, this combination can be used to increase the contrast in signal to interferant.

Finally, one may be incapable or unwilling to move the system to the sample area. A portable preconcentration tube allows for sample collection at various places and then analysis at a location and distance far removed from the source. This sampling method is useful for high precision nose devices that are too large to be moved through not for rapid reporting of dangerous pollutants.

Since many believe preconcentration is the fix-all of nose systems, as is evidenced by many conversations with funding agents and colleagues, we take a moment to analyze the advantage one would yield. The gain associated with preconcentration can be estimated through rate equations.

$$\left(\frac{dn}{dt}\right) = \left(\frac{dn}{dt}\right)_{sorb} - \left(\frac{dn}{dt}\right)_{des} = k_a P\left(N_T - N_o\right) - k_d N_o \tag{6.5}$$

Where P is the probability of a free collision with a vacant site, k_a and k_d are the rates of sorption/desorption, and N_T, N_o are the number of total and occupied sorption sites. The number of molecules on the sorbed is then

$$N_t(t) = \frac{k_a P N_T}{k_a P + k_d}\left\{1 - \exp\left[-\left(k_a P + k_d\right)t\right]\right\} \tag{6.6}$$

For a well made preconcentrator, the rate equations are irrelevant at the specified volumetric flow rate for small concentrations. A comparison in breakthrough volumes is sufficient under two conditions. First, the difference in breakthrough volume at the loading/sorption and desorption temperatures is sufficiently larger (order of magnitude) than the maximum volume of analyte to be sampled. Second, the preconcentrator must be preloaded to the breakthrough volume corresponding to the desorption conditions, or said volume must be at least an order of magnitude lower than the anticipated volume of analyte to be sampled. Otherwise, the desorbed concentration will be smaller than expected as it will be effectively trapped. This loss in desorbed analyte is a non-issue when desorption temperatures are over 200^0 C but under 100^0C they can severely limit the performance of common preconcentrators.

To estimate the actual gain a preconcentrator yields over simple integration, we first assume that the core system is running at its optimum flow rate f_0 and integrating over its standard time constant t_0. The system can then detect a concentration C_0. Increasing the integration time constant to t_f would improve the LOD by the root of the ratio in time constants.

$$LOD_{tot\_\mathrm{int}} = LOD_{t_0}\sqrt{\frac{t_0}{t_f}} \tag{6.7}$$

Preconcentration for the same time t_f at a flow rate of f_f with minimal delay time t_d between preconcentration and desorption would then yield a new limit of detection.

$$LOD_{precon} = LOD_{t_0}\frac{f_0}{f_f}\frac{t_0}{t_f} \tag{6.8}$$

The net gain over simple integration for the total preconcentration and desorption time would then be.

$$LOD_{precon} \approx LOD_{t_f}\frac{f_0}{f_f}\sqrt{\frac{t_0}{t_f}} \tag{6.9}$$

This is the absolute smallest limit of detection possible, assuming $t_0 \approx t_f$, and that everything behaves ideally. Note, the larger gain comes not from the preconcentration time, but rather the flow rate over this time.

Unfortunately, not all is ideal as, for instance, there is an impedance matching associated with preconcentration. Often when reporting on preconcentration merely a small side note gives the sorbant material, the preconcentration/desorption time, and the temperatures used. The sorbants in these tubes are typically classified by breakthrough volume, which is the volume of analyte in the air necessary before analyte can make it through the preconcentrator without being completely absorbed. This assumes a certain flow rate and packing density of sorbant, such that the analyte will be completely adsorbed up to a certain volume of analyte. Increasing flow rates beyond the specified values may decrease the adsorption rate, making the preconcentrator adsorb less efficiently. Additionally, the breakthrough volume depends on the concentration and can vary by an order of magnitude over several orders of magnitude of concentration [49].

6.5.2 Commercial Tenax Concentration Columns

As a small side project, I attempted to fabricate a preconcentrator to determine its viability for our system. Though enough progress was never completed to make a firm conclusion I present the details with observations as they may be useful for future work. We could have purchased a commercial preconcentrator unit from another system although, as mentioned above, the impedance matching may have been a severe issue and likely not worth the expenditure. Instead, I attempted to fabricate one.

I began by obtaining two ¼ inch diameter, 7 inch long stainless steel preconcentrator tubes from Sigma-Aldrich. They contained Tenax TA and Tenax GR packed with approximately 300mg. Both are reported to have a low affinity for water and methanol, making them ideal discriminating sorbants when using transducers that react to water, methanol, and ethanol.

Due to temperature issues with the interferometry, I decided to attempt to make a system that would operate at room temperature. Therefore, rather than wrapping the tubes with Ni-Chrome wire, I

fabricated a simple copper conduction plate to couple thermal energy between the tubes and two thermo-electric coolers. The idea was to cool the tubes by 10-20 degrees Celsius below room temperature, thereby increasing their breakthrough volume, sample a known concentration, and then desorb at room temperature.

Searching the literature further I discovered the desorption efficiency of Tenax TA for ethanol is reported to be roughly 15-20% at 65 ^{0}C and 100% around 250 ^{0}C [49]. Further searching revealed the range of breakthrough volumes vs. temperature for the sorbant materials shown in Table 6-1. Unfortunately, room temperature desorption for both would require breakthrough volumes above 1.8 L and 0.8 L of ethanol to be sampled in a short (<1 minute) time.

Temperature	**0**	**20**	**40**	**60**	**80**	**100**	**120**	**140**	**160**	**200**
Water in Tenax TA	0.13	0.065	0.035	0.018	0.01	*0.006*	*0.004*	*0.002*	*0.001*	
Ethanol in Tenax TA	**7.9**	**1.8**	0.481	0.152	0.055	0.021	0.01	*0.005*	*0.003*	*0.001*
Ethanol in Tenax GR	**2.00**	0.8	0.3	0.116	0.045	0.019	*0.009*	*0.005*	*0.003*	*0.001*
Ethanol in Carbotrap C	0.03	0.013	*0.006*	*0.004*	*0.002*	*0.001*				

Table 6-1: Breakthrough volumes in Liters-per-gram of sorbant materials. The preconcentrator tubes came with 300mg of sorbant packing. Data was acquired from Scientific Instruments Services, inc. [50].

A perfectly clean 300 mg trap can load 0.54 L ethanol on Tenax TA or 0.24 L ethanol on Tenax GR. To load the trap at 50 ^{0}C in 1 minute would require a concentration of 80ppt on a 3 L/min nitrogen carrier. At 200^{0}C the trap would load in 1 minute with a 100ppm concentration. From this we conclude that Tenax, at room temperature with low L/min flow rates, is unviable as a preconcentrator. To make such a trap viable, the mass of the sorbant should be decreased to a couple micrograms such that a concentration below 1ppm would load the trap in the required time. A gas micro-fluidic channel is recommended with a small diameter (<1mm) to maximize breakthrough volume for such a small mass of sorbant. The reason such a large mass is used is that these tubes were designed for large tabletop preconcentrators which have the power to operate at higher temperatures and with larger flow rates.

To check the above calculations I performed a few sanity tests loading at 15 ^{0}C and desorbing at 50^{0}C. For a 300 ml flow of 4 ppm in a 3 minute collection period, the desorption concentration (continuous flow at 300 ml/min) peaked at 11 ppm after the system began heating. It takes a cool system 2-3 minutes to heat to 50 ^{0}C and 4-5 minutes to cool to 15 ^{0}C. If the system has been running a while the housing loads thermally making cooling less efficient with the TEC's. Ideally the volume would be heated to the desorption temperature with no flow, or a flow in a closed loop, and then the desorbed volume would be sampled. As this was a side project, time was not spared to construct such a circulator.

A change in loading procedure was implemented to improve response time and to enable the rapid cycling of the preconcentrator for these tests. Since the Tenax trap has a tremendous capacity for ethanol, it is pointless to ever stop exposure as well as to not expose during cooling. The trap was hooked up to the Phocheck 5000+ photo-ionization detector and the pump was never turned off. A continuous concentration of 2 ppm was injected through the preconcentrator, allowing it to eventually load to capacity at 50^0 C through constant use. The breakthrough concentration at the end of the Tenax GR tube was continuously sampled, with the tube at 50 ^{0}C, until a steady state of 2 ppm was measured. The tube was then cooled to 15 ^{0}C in 10 minutes and then immediately heated once the cold temperature was reached. While cooling, the concentration dropped and held constant around 0.4 ppm and after roughly 100 seconds of heating the concentration peaked at 16.6 ppm as shown in figure 6-2. Additionally, preconcentrator data was collected using a Phocheck 5000+ PID to eliminate the potential of thermal noise in the interferometer, because the PID is more sensitive (though less selective, but we knew the analyte was EtOH), and the purpose was to test if the preconcentrator would even work.

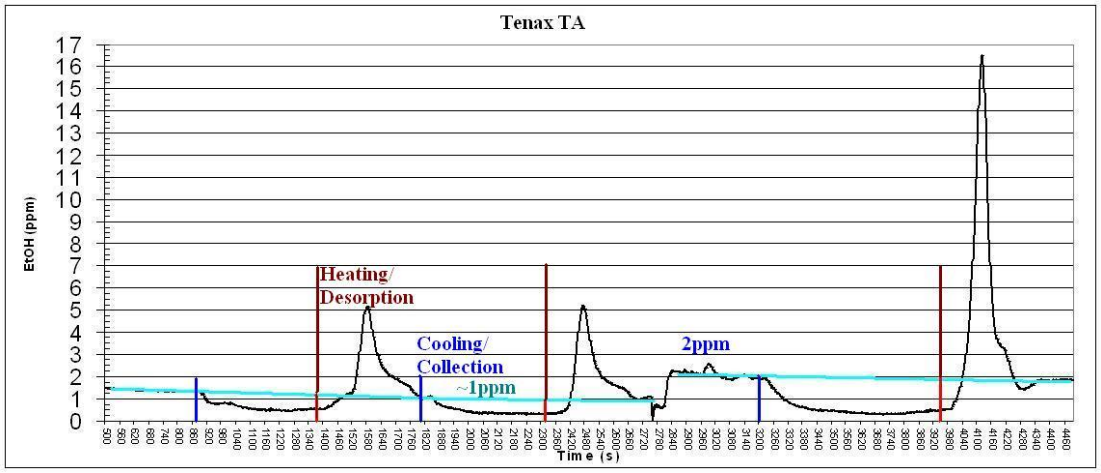

Figure 6-2: Preconcentrator performance data using Tenax TA. The column was operated between 25 and 50 ^{0}C. The first two runs show the preconcentration of approximately 1ppm continuous flow while the third shows a run of a 2ppm flow. The collection, cooling, period ran until the cold target temperature was reached and then immediately switched to heated desorption. The longer integration time and collection volume of the third trial was due to the increased flow rate, necessary to double the sampled concentration, which increased effective load on the TEC's by convectively heating/cooling the tubes to room temperature. Vertical bars denote a switch between heating (red) and cooling (blue) while the horizontal bar is the sampled concentration for the particular run.

The conclusion of the above experiments was that the breakthrough volume was too large at the desired desorption temperatures for the mass of sorbant. It should be noted the temperatures were limited to 50^{0}C because the TEC's would be damaged over 80^{0}C and to use as small a temperature differential as possible to reduce thermal noise at the interferometer TIR interface. Additionally, any polymer used would likely degrade faster at higher temperatures, though it may be useful to use higher temperature desorption if a metal-oxide chemical transducer is in use, to facilitate transducer recovery.

6.5.3 Metal-Oxide Mesh Concentrator

Since some papers report the use of transducer compounds as preconcentrator compounds, it seemed only reasonable to attempt a metal oxide preconcentrator since I knew that, to some extent, these worked at room temperature. Aluminum oxide was chosen since aluminum wool was easily obtained, easily manipulated, and inexpensive. Additionally, it readily grows a corundum oxide layer in air making it easier to grow an oxide layer, and aluminum oxide beads can be used as a desiccant. A fine 0.002" grade aluminum wool was purchased from Palmer Engineering Products, Inc and oxidized

over 24 hours in an oven at 500^{0}C. To facilitate oxidation water was sprayed over the wool 4 times at 1 hr intervals. A noticeable gray matt coloring was observed on the finished product, indicating the presence of an oxide. I then packed 500g of wool into a stainless tube similar to the commercial Tenax tubes, and capped both ends with stainless steel screens pressed into place to minimize shifting during use, as shown in figure 6-3.

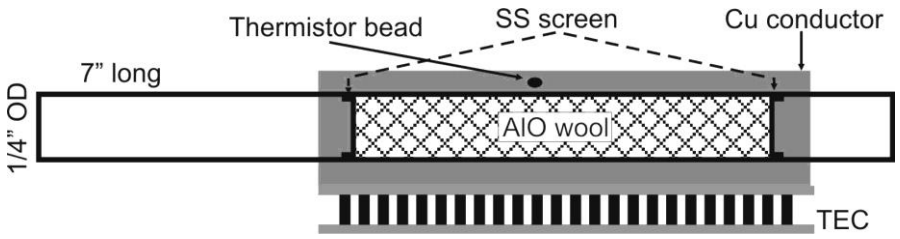

Figure 6-3: The Aluminum oxide preconcentrator tube cross-section. The TEC pumps heat to a copper sleeve that fits over the concentrator tube to more uniformly heat and cool the tube and its contents.

I reran the experiment at 13-25^{0}C cold-hot using the aluminum oxide wool. The system takes 1.5-2.5 minutes to cool, depending on external case temperature, and <50 seconds to heat. Again the system was allowed to run continuously sampling a 2 ppm source. The 25^{0}C steady state read 1.8 ppm, while the concentration during cooling was around 1.4 ppm and a peak of 2.7 ppm was read 10 seconds after heating began. Measuring a control, no ethanol, the 25^{0}C steady state read 0.16 ppm, while the concentration during cooling was around 0.13 ppm and a peak of 0.18 ppm was read 10 seconds after heating began.

Finally, since such a poor result was obtained for the 10 ^{0}C temperature differential the Aluminum oxide was retested using a swing from 15 to 50 ^{0}C. The results showed that the AlO2 shows promise as a preconcentrator though, the oxide on the aluminum wool used had a significantly smaller sorption capacity. The data shown in figure 6-4 shows the response of the anodized aluminum sample. However, after 8 trials the tube appeared to have been poisoned, in that it lost capacity and no longer maintained any appreciable sorption capacity. No attempts to bake the tube out at higher temperatures were made due to time constraints.

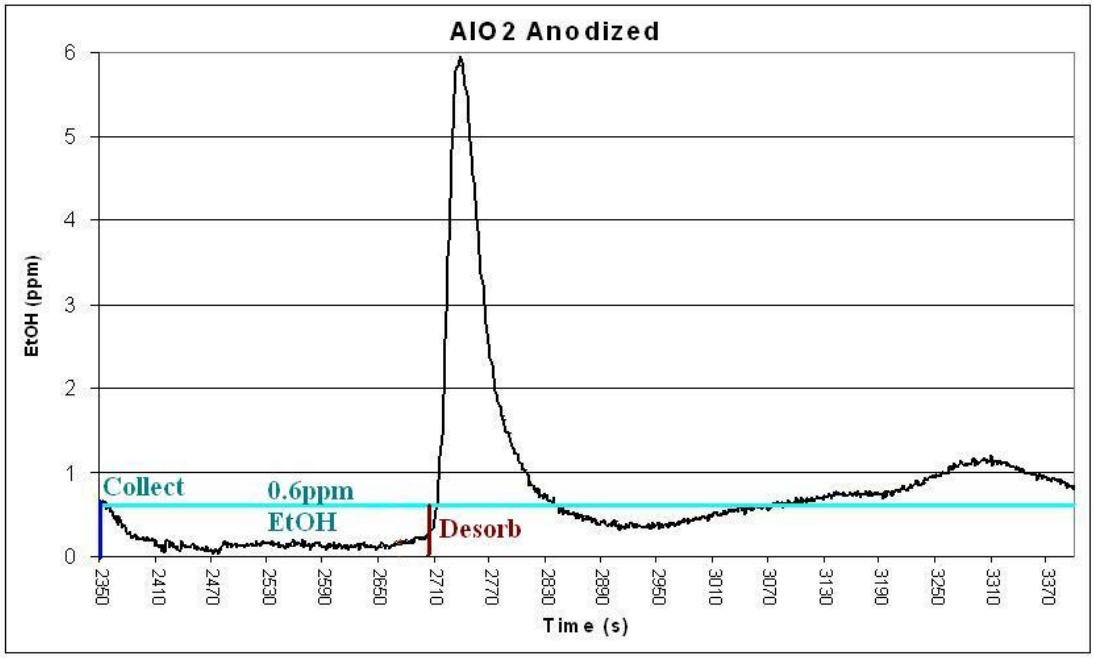

Figure 6-4: Preconcentrator performance using Aluminum oxide. The AlO2 showed a significantly smaller capacity for EtOH, though it responded faster to heating and cooling. This plot shows a sorbed capacity of approximately $3 ppm \cdot \min$, at the fixed flow rate of 200 ml/min, which was roughly the capacity of the tube.

The conclusion drawn from this last experiment was that the aluminum oxide performs faster, likely due to suspension on aluminum rather than glass. An aluminum oxide powder may be used, though it will likely suffer slower response times where the wool helps to conduct temperature more rapidly, thereby improving response time. The wool appeared to work well enough though the best performing batch, the anodized sample, stopped working after only a few cycles bringing serious doubt on the abilities of the oxide as a preconcentrator material.

6.5.4 Preconcentration Conclusion

An alternative mode of operation was presented in which the preconcentrator is a front-end through which the environment is constantly being sampled. During cooling the trap is loading as fast as its increasing breakthrough volume will allow while also enabling the system to sample a nearly analyte free background. After the trap has cooled to its set point the system is set to desorb as rapidly as possible while continuing to sample. While this method looses the benefit of higher flow rates for

preconcentration, it enables a contrast measurement to be made where the preconcentrator is attenuating species it has an affinity for and otherwise leaving the rest unperturbed. Such contrast allows the preconcentrator to add another level of selectivity to the system, while improving sensitivity. However, to improve cycle time it may be advantageous to load the tube at a high flow rate, then rapidly desorb the analyte into a collection circuit using nichrome to rapidly heat the column, and then sample the concentrated content of the collection tube.

Preconcentration ultimately can aid small-signal detection, assuming sufficient design is put into the balancing of breakthrough volume, and response time. We do not use a preconcentrator for any further experiments as these were meant to be a proof-of-principal set of experiments. It is recommended that aluminum oxide not be investigated further as all batches were found to poison after long periods of cycling, while the Tenax samples continued to operate. A Tenax or Carbotrap tube with 10-50 mg would likely be sufficient for concentration of sub ppm levels in the times desired.

6.6 Valves

6.6.1 Laminar Flow Basics

One of the fundamental requirements of this system is a concentration-modulated vapor flow. Modulation is necessary to prevent the photorefractive from adapting to a constant flow, and to compare against a reference gas. To achieve content modulation, a valve is used to switch between the test and reference flow lines at the modulation frequency, resulting in a flow whose concentration is modulated down the sample tube. However, limited by practical obtainable flow rates for miniature systems and reasonable tubing sizes, the fundamental flow physics restrict the speed of vapor modulation.

Vapor flow through the delivery tube will exhibit a Laminar flow for pressures, flow rates (low 100's milliliters-per-minute), and dimensions (1-2mm diameter tubing) used in this system. One fundamental characteristic of laminar flow in a tube of radius R is a radial flow velocity profile [51].

$$v(r) = 2\bar{v}\left[1 - \frac{r^2}{R^2}\right] \tag{6.10}$$

The flow profile can be described as uniform velocity layers which approach zero velocity near the channel guide ($r = R$) and are faster closer to the center of the tube. The average velocity $\bar{v}$ is found by averaging over the total flow area, which empirically is just the volumetric flow rate times the tube area. Figure 6-5 illustrates a cross-sectional flow profile of two alternating gasses injected into the tube.

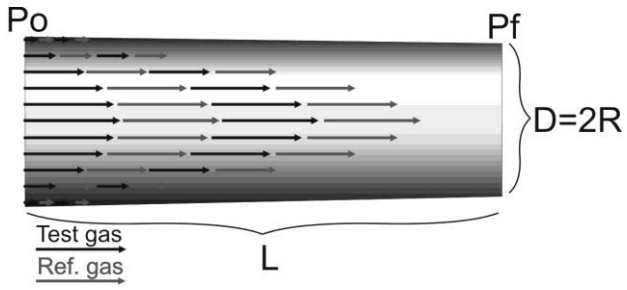

Figure 6-5: A concentration-modulated gas exhibiting laminar flow through a tube. Radial velocity dependence causes mixing near outer radius resulting in effective low pass filtering of the modulation.

The average velocity can also be estimated using the differential driving pressure ΔP between two ends of a tube of length L.

$$\bar{v} = \frac{\Delta P R^2}{8\mu L} \tag{6.11}$$

The effective impedance of the tube scales as the viscosity, the length, and inversely with the radius to the fourth power.

$$Z_f = \frac{8\mu L}{\pi R^4} \tag{6.12}$$

Finally, the Reynolds number for flow in a tube is defined as

$$R_{\#} = \frac{\rho v D}{\mu} \tag{6.13}$$

where μ is the kinematic viscosity($Pa \cdot s$), ρ is the fluid density(kg/m^3), v is the characteristic velocity(m/s), and D is the characteristic distance or tube diameter(m). The critical boundary between laminar and turbulent flow is defined by the Reynolds number as $R_L < 2000 < R_T$. Air has a typical viscosity of 18 $\mu Pa \cdot s$ at 25^0C.

We briefly fill in the remaining values for calculation purposes using fundamental chemistry concepts. Nitrogen gas, the carrier for most experiments, has a mass of 28 g/mol, which will occupy a volume of 27 L. The volume is arrived at using the Boyle's Law, $P_1V_1 = P_2V_2$, where the air pressure in Colorado is approximately 630 mmHg compared to 760 mmHg (sea level), and the volume occupied by a mole of gas is 22.4 L at sea level. The density of nitrogen gas is therefore 1.03 kg/m^3.

The vane pump used (Schwarzer Precision: SP 135 FZ) can only supply a differential pressure of 6 kPa, or 45 mmHg. Therefore, the density will change by no more than 0.08 kg/m^3 with maximum flow impedance. The experimental flow rate is measured to be between 100-300 ml/min for a 2 mm diameter tube, indicating a mean velocity of 1.6m/s and a differential pressure on the 10cm long tube of 23 Pa. These values indicate that fluctuations in density are negligible. The Reynolds number can also be estimated at 60-180, well within the laminar flow regime (<2000). For convenience, a pair of useful conversions are given as $1\ ft^3/hr = 471\ mL/\min$ and $1\ mL/\min = 1.26\cdot 10^{-5}\ m^3/s$.

The vapor modulation serves a few key functions. First, it allows continual comparison between two vapors, helping to eliminate noise due to system drift. Second, it helps the system resist adapting to a trace gas, similar to the way animal sniffing aids in detecting faint smells. Finally, the transducer dynamics associated with sniffing, or modulated exposure, can aid in the identification and discrimination between two analytes of similar response sensitivity. We don't employ this last method but its proof-of-principle has been demonstrated by H. Ye [53].

6.6.2 Valve Switching/Chemical Modulation

Next we expand on the idea of chemical modulation. By using two gases of different content one can model the periodic mixing with basic periodic functions. The two basic forms of modulation would then take on a sinusoid and a rectangular function.

$$C_S = \frac{C_0}{2} + \frac{C_0}{2}\sin\left(k_r z - \omega_s t\right) \tag{6.14}$$

$$C_R = C_0 rect\left(k_r z - \omega_s t\right) \tag{6.15}$$

where ω_s is still the signal modulation frequency, k_r is the radial dependant fluidic wavenumber, and C_0 is the concentration, which we set to 1, when indicating pure test gas concentration, and 0, when indicating pure reference gas concentration. The wavenumber is found based on the modulation frequency and the speed at the particular radial layer.

$$k_r = \frac{\omega_s}{v(r)} = \frac{\omega_s}{2\bar{v}\left[1 - \frac{r^2}{R^2}\right]} \tag{6.16}$$

Since the polymer sees a jet of air from the output of the tube, we integrate the concentration function over the entire area at that end of the tube to determine the effective contrast by observing the mean concentration at $z=L$.

$$\bar{C}(t) \approx \frac{2\pi \int_0^R C_s(L,t) r \cdot dr}{\pi R^2} \tag{6.17}$$

This equation assumes that modulated layers mix evenly at the end, though this assumption is not true since the velocity profile indicates that the more central region of the tube will have the momentum necessary to reach the polymer. The boundary layers are more likely to shed once they leave the tube, due to their lower velocities and pressure differentials at the end of the tube. Still, this equation gives the worst-case mixing and any correction factor simply limits the outer radius of integration R.

6.6.3 Laminar Low-Pass Filter

Having established the flow conditions to be laminar and a means of modulation, we take a closer look at the effective low pass filtering that laminar flow performs on a chemically modulated signal. Since the fluid has layers of constant velocity, by the time the fluid reaches the end of its tube there will mixing near the outer radius, if the input flow concentration is modulated. When integrated though a revolution inside the pipe, mixed layers account for a larger percentage of the volume substantially affecting the contrast between the test and reference gas. This radial dependence is best observed in figure 6-6, which shows a cross-section of the flow that is periodically modulated between the reference and test gas.

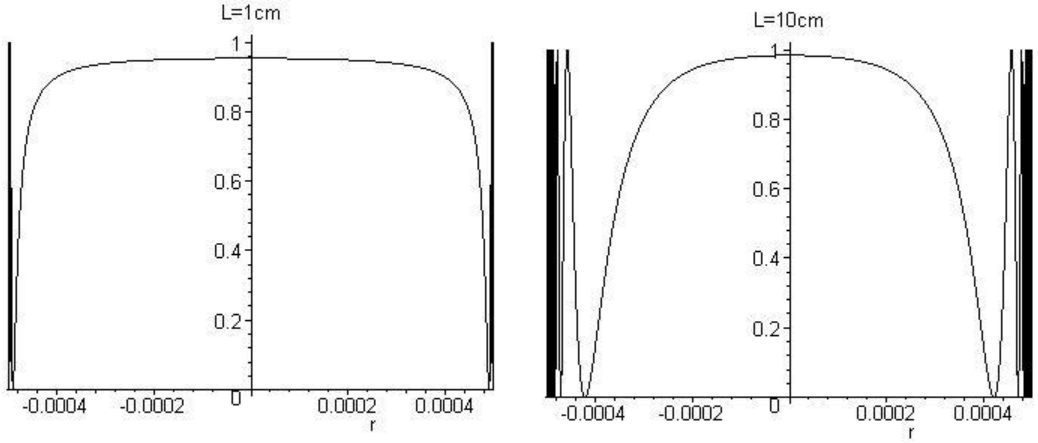

Figure 6-6: A cross-section of a 10 cm long 1mm diameter tube showing the concentration of the test and reference gas for sinusoidal modulation at L=0. A value of one indicates pure test gas while a concentration of zero indicates pure reference gas.

The contrast at the output of the tube decreases with tube length and modulation frequency. We describe the contrast behavior with a modulation frequency similar to a low pass filter.

$$H = \frac{\bar{C}\,|_{pp}}{C_0} \approx \frac{1}{1+\kappa\frac{\omega L}{\bar{v}R}} \tag{6.18}$$

where $\kappa \approx 2.4e^{-4}$ is a constant appropriate to our flow conditions. This is a fitted function, analogous to a capacitive filter, that is accurate for long characteristic lengths L/R>10 at our flow conditions (0.3 to 3 L/min). The function in its full form is impractically complicated for our purposes and amounts to a minor correction over the above approximation. Figure 6-7 shows the resulting filter characteristics for a laminar tube indicating that, for lengths of 2 mm diameter tubing beyond 1cm, significant signal attenuation occurs. Having characterized the tubing filter properties we have everything needed to deliver a sample of gas to the optical nose for analysis.

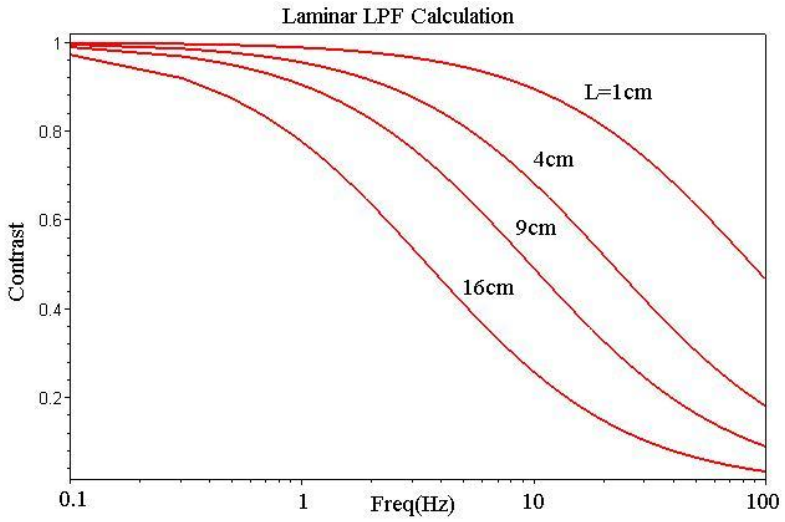

Figure 6-7: Attenuation curves vs. chemical signal frequency ω_s. These curves indicate that the shortest tubing possible along with the lowest modulation frequency that's reasonable will yield the best response to a low concentration of analyte.

6.6.4 Ethanol Diffusion Length

We quickly detour to analyze the diffusion for ethanol in a nitrogen environment to estimate any potential mixing due to diffusion. The diffusion coefficient *D* for ethanol in nitrogen is between 0.1147 and 0.153cm^2/s, which are the coefficients for propane and carbon dioxide respectively [52]. We use this range since propane is close in size to ethanol while CO2 is comparable in mass. Assuming the lower coefficient for ethanol, we examine the characteristic diffusion length for *t=1ms* and *250ms*, the time of the pressure spike due to solenoid dead time, and the half-period dead time for a 2 Hz switch frequency, respectively.

$$L = \sqrt{Dt} \tag{6.19}$$

The diffusion lengths for these times are 0.1 mm in 1 ms and 1.7 mm in 250 ms.

While the diffusion length of 100 um is fairly insignificant during the short pulse, the length of 1.7 mm is very significant since, in a 2 mm diameter tube any laminar flow profile will radially average out by diffusion in 250 ms. The reasoning behind the choice of these two arbitrary times will be made apparent in the next section.

6.6.5 Solenoid Valve and Problems

While it may be simple to modulate theoretically a chemical carrier, in practice a means of switching or mixing two separate lines is needed. Originally, solenoid valves made by Pneutronics were used to switch between gasses in a square wave fashion. However, the typical solenoid valve design presents a problem with dead time. It was noticed early on, depending on the flow rate and diffusion-tube length, that adding a vent before the switch reduced the detected concentration by up to a factor of 3 compared to a similar flow-dividing vent after the switch. Additionally, increasing the diffusion tube length did not appear to have the linear effect on concentration predicted by equation (6.4). To try to explain this discrepancy, we developed a simple model for what may have been occurring.

The fluid dilution system is schematically depicted in figure 6-8 below. The original system A: lacked vents at the output such that during switching, a brief dead period would cause a minute pressure hiccup through the system. This hiccup was easily observable on the flow regulator/indicators as the glass beads would bounce with each switching event, which we originally set to a 2 Hz cycle. We modified the system to include a pair of solenoids and vents which ensured that air was always flowing through both lines, and that there was minimal pressure ripples through the vapor system.

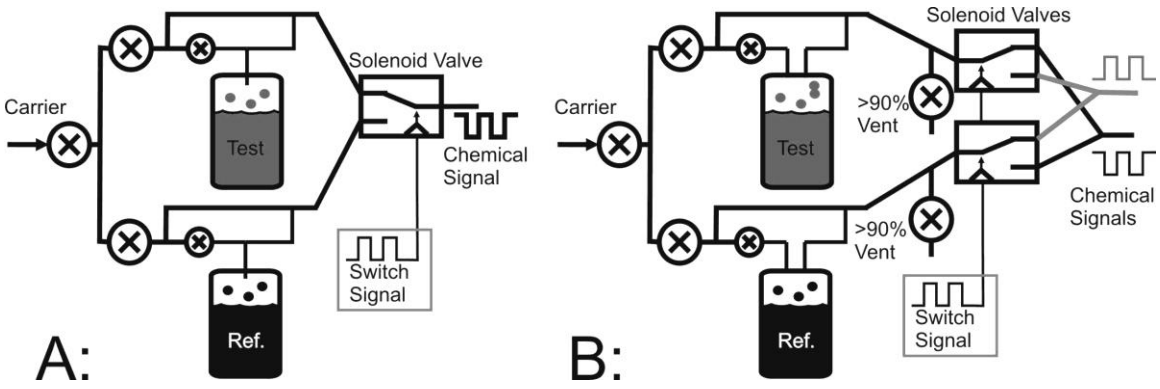

Figure 6-8: Vapor dilution and modulation lines A: in its original form which experienced large pressure hiccups on switching and B: with balanced switching lines and vents to reduce hiccups and pressure buildup, ensuring both lines see a nearly constant flow over the sample chamber.

To illustrate the need for the vents and balanced switching we take a moment to model the fluid system as a basic RC circuit. First, we refer to table 6-2 for the basic conversion of base units and components from which we can create an electrical analog of the dilution scheme. The carrier is modeled as a large current *Ic* which picks up a smaller current, representing the signal contribution of the sample bottle, and exits through the valve when the switch is closed. The concentration can effectively be viewed as the ratio of the current entering at the "T" to the carrier current *It/Ic*. Additionally, the current leaving the tank in steady state is equal to the current entering (It=Is), which is to say, that in steady state, the evaporating sample is the only source of molecules leaving the system, mixed with the carrier gas. A figure of the circuit diagram is shown in figure 6-9.

Electrical	**Fluidic**
V: Voltage	P: Pressure (Pa)
I : Current	F: Volumetric Flow Rate (m^3/s)
R: Resistance	Z: Flow Impedance ($Pa*s/m^3$)
Q: Charge	V: Volume (m^3)
C: Capacitance	C: Fluidic Capacitance (m^3/Pa)

Table 6-2: Analogous conversion between electrical and fluidic variables

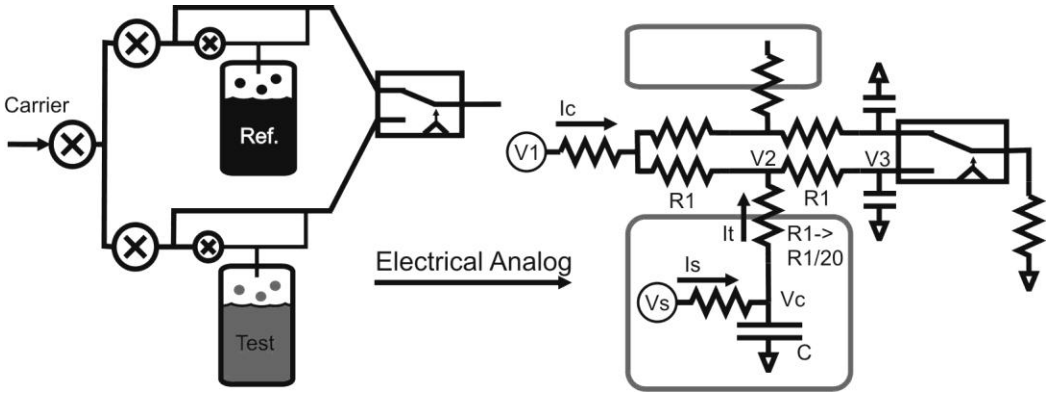

Figure 6-9: Vapor dilution system with a single solenoid and its equivalent circuit for pressure pulse analysis.

The capacitance of the sample bottle, due to the compressibility of vapor in its headspace, will allow a slow buildup of pressure when the switch is closed. Additionally, should the solenoid have a dead time, a time during switching in which no flow is possible, then the voltage sources would begin to charge all capacitances at an increased rate, which is to say the entire system would begin to

increase in pressure. When the valve finally opens on either side the pressures will drop back to their steady state limit for the particular valve position. Evidence of the pressure spike during switching is observed by both glass-bead flow meters showing a notable, though not practicably measurable, hiccup in flow rate when switching between their flow and no-flow state.

Should the sample bottle have no headspace, system pressurization would make little difference. However, since the headspace gas volume in the early system was on the order of 50-100 ml, in a 125 ml flask, this volume is substantial compared to that of the entire tube lines. Therefore, given adequate charge time and a sufficiently large pressure fluctuation, the change in pressure in the bottle would be a substantial portion of the volume in the supposed diffusion line that connects the sample tube to the "T", or voltage point *V2*. Thus, the current *It* would be many times that of *Is*, increasing the concentration sampled.

6.6.6 Sinusoidal Valve

While proper dilution design can mitigate or remove detrimental effects of pressure pulses, noise due to pressure pulses can be further reduced by simply removing the pulses. If the valve had zero dead time (continuous flow through the valve), combined with a constant vent to atmosphere, pressure fluctuations upstream due to modulation should be eliminated. To achieve the above characteristics, several design iterations of a sinusoid valve were implemented. The fundamental principal is that a rotating barrel with a chopper or divider will sweep along a slot, resulting in sinusoidal mixing of the input flows. Figure 6-10 illustrates the component and rotor in better detail.

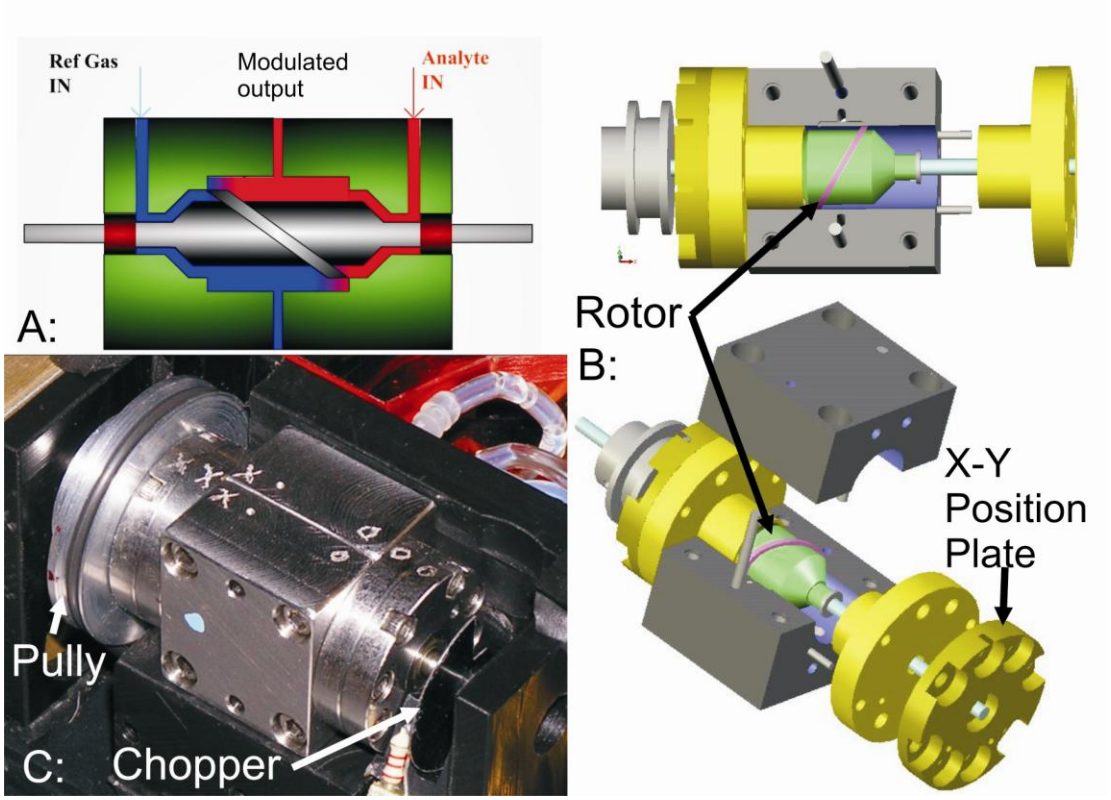

Figure 6-10: Sinusoid Valve. A: 2D cross section of valve performing sinusoidal division of two input flows among two output flows by sweeping a rotor back and forth. B: Two viewpoints of a Solid Works exploded model, showing the division rotor mounted at 45^0 about the circumference of the rotating barrel. The barrel's position in the cylindrical cavity is dictated by the two X-Y position plates which were fabricated to allow ½ mm positional adjustment for fine tuning of rotor's center position and orientation. C: A photograph that shows the valve housing along with the optical chopper used to synchronize the signal demodulation with the rotation position of the valve.

A problem associated with a sinusoidal valve is mixing in the dead space around the rotor. While decreasing the size of the rotor and barrel would decrease the dead space, fabrication and alignment tolerances of the sinusoid valve are prohibitive. The third valve design, as shown in Figure 6-10, has adjustments built in to allow the rotor/barrel assembly to be tilted and translated for perfect alignment. However, the brazing process used to attach the rotor, such that it can be machined, causes deformations that can be difficult to overcome. Simply using an adhesive to mount the rotor risks it dissolving off with strong organic solvent concentration sampling, which would potentially destroy the inside of the valve, especially when spinning at 300-600 RPM.

Additionally, it was later found that the rotor swinging by the division slot at higher speeds causes pressure pulses at the output of the valve. It is believed these pulses are due to leading-edge

compression and trailing-edge, drag which is coupled straight to the output port when the rotor swings by the output port hole centered in the division slot. In the TIR-noise section it was established that pressure sensitivity is quite high and indeed the spikes due to these were significant, though less so than for solenoid switching. Fortunately, reversing the valve such that the input fed to the outer diameter of the barrel while the output was near the bearings reduced this problem, allowing 2ppm/$Hz^{1/2}$ detection sensitivities. Also, the laminar low-pass filtering decreases the contrast amplitude of a sinusoid valve. Ultimately, while this valve is a substantial improvement over solenoid valves, something with the pressure stability of a sinusoid valve but a higher harmonic division similar to the solenoid, to combat the laminar low-pass filtering, would be preferable.

6.6.7 Clipped Triangle Valve

Rather than attempt yet another more expensive design to incorporate a rotor that would yield higher harmonics in the modulation while still minimizing the leading and trailing edge pressure fronts, a new path was taken. The sinusoid valve operated essentially 3 dimensionally, but required a large internal volume and a high manufacturing precision for many different parts. To reduce volume but still maintain the same principal of flow division, we folded the three dimensional design into a two dimensional design. Figure 6-11 shows this new design, which utilizes a sealed ball bearing and ruby ball for rotation and location. The inner drum is made of brass and the outer is stainless steel, since these materials slide well when they come into contact. While the precision of the rotor head is still extremely important, the reduction in the degrees of freedom and necessary alignment make this design simpler to fabricate and align. Internal volume is also drastically reduced though a rim along the outer diameter can be increased to add capacitive dampening for pressure spikes in future designs.

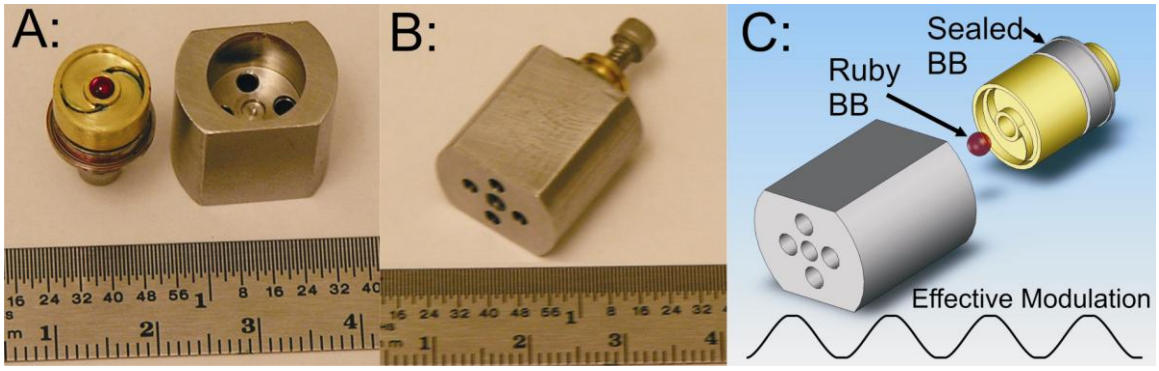

Figure 6-11: Clipped triangle valve. A: Shows the inner drum with curved rotors and center fixed ruby ball bearing. B: Exterior view of valve. C: Exploded model of assembly. The effective modulation is that of a slight LPF clipped triangle wave.

This new design is referred to as a clipped triangle valve since about 10% of the rotation should see no change in concentration, while a nearly linear division between the ports occurs as the rotor swings past the holes. The division rate is "nearly linear" since the curve of the rotor effectively rounds off the edges of the transition. This design helps combat the LPF of the fluid flow lines by holding the maximum and minimum concentration for a fraction of the rotational duty cycle while any pressure spikes should occur at 2X the modulation frequency and be rounded off since the curve of the rotor eases the beginning and end of each transition. Additionally, the rotor is razor-edged to minimize pressure fluctuations that may arise from the impedance changes associated with the aspect ratio of the rotor with respect to the hole it is dividing.

We have discussed various problems associated with low volume vapor sampling and some solutions. Several valves were introduced with known strengths and weaknesses though the immediate problem of pressure noise on the transducer was minimized through the use of lower harmonic valve modulation. The combination of atmosphere vents just up-stream of the valves, and low harmonic valve modulation with continual flow, enabled the chemical LOD to get within a factor of 3 of the system's absolute limit. Further reduction of pressure fluctuations may enable the system to improve by up to a factor of 3 in chemical LOD.

6.7 Chemical transducers

6.7.1 Basic Transduction Overview

A small amount of work was performed to study the ideal polymer chemistry leaving a more indepth investigation to those with an appropriate background. The base knowledge in polymers used for this thesis was acquired by Dr. Hongke Ye before his departure from this project and essentially one a handful of the polymers he explored were used. Of this handful, PNVP or Poly (n-vinyl pyrrolidone) was used as the workhorse to compare system performance. PNVP is essentially an alcohol sensor that has a high sensitivity to many VOC's as well as water. The polymer is easy to apply and clean while also shows one of the higher sensitivities of Dr. Ye's array of polymers. To compliment PNVP's sensitivity, ethanol is used as a baseline VOC, since social & cultural activity has shown its long term effects to be relatively benign unless ingestion or inhalation at high concentrations (>ppt) are coupled with high velocities. Since we care about ppm and lower concentrations this is a fairly safe chemical for everyday tests.

As mentioned earlier, polymers aren't the only transducers available or viable for this system. Dr. Ye has demonstrated biological yeasts and peptides as transducer elements. Essentially anything made thin enough to be optically transparent will suffice as a transducer for TIR sensing. While there has been some controversy over the chemical processes involved, essentially any interaction with a surface that doesn't result in the incident analyte simply bouncing off the surface will cause a localized index perturbation. For a more detailed view, we refer the reader to Dr. Ye's thesis [53], and another paper [54].

6.7.2 Polymers as Transducers

A polymer is typically a complicated molecule consisting of thousands monomers. In the progression of selectivity, polymers are a step below proteins and antibodies. Polymers often have a broad range of sorption selectivities, while antibodies usually only bind to a specific molecule or analogs that are extremely close in molecular structure. Yet, where antibodies must be immobilized

and suspended to maintain reactive viability as a transducer, polymers are relatively easy to deposit onto a sensor surface and are often quite robust.

Used as a transducer, the basic function of a polymer is to selectively adsorb an analyte such that the refractive index of the polymer can then be interrogated to determine if a sorption event, i.e. a detection event, has occurred. Knowledge of the mechanism for adsorption is important to design a system that is optimized for the analytes of interest. To this end, we briefly explore the mechanisms for sorption of a molecule like ethanol onto a polymer.

In a simplistic view the concentration in the air and at the surface satisfy a sorption rate equation.

$$\left(\frac{dN_s}{dt}\right)_{liq} = \left(\frac{dN_s}{dt}\right)_{ads} - \left(\frac{dN_s}{dt}\right)_{des} = k_a P\left(N_{air}\right) - k_d N_s \tag{6.20}$$

The surface concentration N_s is given as the number of surface molecules, N_{air} is the number of airborne molecules, P is the probability of a collision with the surface and k is the probability of a adsorption/desorption event a given second. As this model is simple, we assume no shielding and that the probability of an adsorption event is unaffected by the current load of the polymer in the small airborne concentration limit. Then assuming that we start with zero adsorbed molecules at the surface, the surface concentration over time is given as

$$N_s(t) = \frac{k_a P N_s}{k_a P + k_d}\left\{1 - \exp\left[-\left(k_a P + k_d\right)t\right]\right\} \tag{6.21}$$

Since it is difficult to determine the actual sorption rates and binding affinities, we will only go so far as to overview the various theories explaining adsorption. One does not necessarily need to understand the sorption kinetics in detail to use polymers successfully, though such knowledge combined with appropriate resources may enable more sensitive and selective transducers. A basic starting point to determine if a polymer would be a good transducer for a particular volatile organic chemical is to determine the solubility of the polymer in the solvent. High solubility often indicates that the transducer-VOC interaction would yield a favorable signal.

There are a range of binding mechanisms that may occur when an analyte gas or sorbate come into contact with a polymer or sorbent. The physisorption (Van der Waals) mechanisms are, in order

of increasing strength; dispersion, dipole-dipole, and the hydrogen bond. Physisorption denotes the temporary binding of an analyte to the transducer surface which may form multiple layers on the surface and transport, by diffusion, into the surface. Chemisorption notes the occurrence of a chemical (covalent) bond and typically results in the formation of a single monolayer via bonding to surface sites. Such binding is rarely reversible without additional energy added to the system and may lead to loading or poisoning of the transducer. The basic behaviors of the binding mechanisms are expressed in the models below.

To describe the sorption of analytes onto polymers, the basic models presented make a few assumptions about behavior which may describe several mechanisms. There are two main forms of sorption uptake; adsorption, where the analyte sticks to the surface, and absorption, in which the molecule enters inside the sorbent or polymer. The uptake can typically be modeled as one or a combination of the following sorption models: the Henry, Langmuir, Flory-Huggins, and the BET Isotherms. The explanations of the various physical sorption isotherms and mechanisms are borrowed heavily from two sources [55,56].

The Henry isotherm describes a linear sorbed concentration with respect to the vapor concentration. The isotherm generally assumes a non-specific absorption in which the binding strength is relatively weak between both sorbate-sorbent and even weaker between sorbate-sorbate. The Henry model is mainly useful where the mechanism for sorption takes the solvent into the bulk of the material and the rates of sorption and desorption are not affected by the sorbed concentration.

The Flory-Huggins Isotherm is a model describing sorption when there are a fixed number of ways that a solvent can fit into a lattice structure upon bulk sorption. This model then assumes that the sorbate-sorbate uptake is significantly stronger than that the sorbate-sorbent, meaning that the uptake concentration vs. vapor phase partial pressure is super linear, i.e. once some of the analyte is sorbed, the polymer has a higher affinity for additional analyte. However, since it has been shown that our system is nearly 3 orders-of-magnitude more sensitive to adsorbed rather than absorbed solvent, due to the TIR phase shift vs. bulk optical path sensitivities, these models, which primarily describe absorption, are less likely to be applicable to the reaction of our system.

Moving on to kinetic models of sorption, we look at the Langmuir model, which describes the absorption to specific finite binding sites or the adsorption to the limit of a single monolayer of sorbate. This model saturates at high partial pressures due to the limitation of sorption sites. However, it also assumes that all sites exhibit the same probability of sorption regardless of the occupational state of neighboring sites. Modifying this model the Frumkin Isotherm includes polar or ionic interactions but still assumes a limited number of sorption sites available.

The Brunaver-Emmett-Teller or BET isotherm takes a different approach in modifying the Langmuir model by assuming multilayered adsorption. This isotherm therefore has two affinity constants to consider, the first being the sorbate-sorbent interaction and the second describing that of the sorbate-sorbate. Regardless of which of the above mechanisms are at work or which model(s) best describes the analyte interaction with the transducer, the goal when studying sorption kinetics is to select the polymer, and therefore the kinetics, that achieves the best sensitivity and selectivity possible.

For further detail in sorption dynamics and the potential of polymers as a transducer the reader is urged to seek the work of other Ph.D. dissertations whose focus is that or a nearly parallel goal. For a more detailed review of the binding models the reader is referred to chapter 2 of the dissertation by M. Harbeck [55]. One polymer found to be useful in conjunction with ethanol sensitive polymers is Cellulose Acetate Butyrate (CAB), a polymer that is less sensitive to ethanol than PNVP and almost insensitive to water, making CAB a good discriminator against humidity which, in conjunction with PNVP, enables the sensitive detection of ethanol. Unfortunately, as will be discussed below, its deposition was substantially trickier than PNVP.

6.7.3 Fabricating Polymer Transducers

The fabrication of polymers is relatively straight-forward. A polymer (say PNVP) is dissolved in a solvent, one having a high solubility (water or ethanol), and the solution is dropped or spin-coated onto a clean slide that has been cut down in dimension. The concentration of the solution is typically 50 mg solute per milliliter solvent. An alternative is to deposit the polymer onto a Mylar film or similar robust base sheet, and the resulting solid layer is cut into sensor patches and adhered, Mylar side down, to a slide in an array fashion with other polymers.

Though simplistic in principal, polymer-transducer deposition is a bit of an art in that layers can be cured in clean chambers for hours to achieve nearly perfect polymer sheets, or they can be cured with other techniques that, under the right circumstances, appear nearly identical to the unaided eye. The method used early on was to drop the solution onto a perfectly clean transducer substrate and seal it in a chamber that had a reservoir of the solvent off to the side. The idea was that the reservoir would maintain a high partial pressure of the solvent to slow its evaporation out of the polymer, helping to ensure the polymer dried with a nearly perfect surface layer.

One reason for the slow cure method was that the surface tension of the solvent on the substrate often would cause the deposited solution to bead up to a slight degree, causing thicker layers out near the edges and otherwise making a uniform coat more difficult to obtain if allowed to dry fast. The high vapor concentration in the curing chamber helped to slow this process and make the layer more uniform. However, we found that surface treating the glass for more polar solvents like water and ethanol, which exposes hydroxyl groups, would decrease the contact angle and help the polymer to cure more uniformly at shorter cure times.

Surface treatment was achieved with some success in two ways. First, a strong base-like sodium or potassium hydroxide was used to clean and scrub the surface immediately before the solution was deposited. Second, oxygen plasma was used to ash the surface, exposing hydroxyl groups and increasing the surface affinity to polar solvents. However, neither of these methods has proven to yield noticeably superior polymer sensitivities, as they just help make the layers more uniform. It was also noted that these last two techniques are more useful with more hydrophobic surfaces, but this is irrelevant for the system at hand since microscope slide glass works well as transducer substrates.

The deposition of CAB by dissolution requires stronger solvents such as acetone, and is substantially more difficult to cure in a chamber than PNVP and other polymers. Because of the difficulty involved in solvent-deposition methods, melt deposition in a fume hood was developed as a viable alternative. A sample of CAB, cut from a sheet, is heated to about 210-250^{0}C, at which point the polymer becomes molten. Caution is advised as just beyond this temperature, CAB begins to react with the atmosphere, browning rapidly. Once the transition has occurred, a glass slide is used to squeegee or smear a thin layer of CAB over the substrate surface, which is then allowed to cool. Once

cooled the substrate can be sliced, pealed up, and glued into place with an index matching glue. Otherwise, if the substrate is a glass slide, the transducer is ready for use. While these samples are often thicker than solvent-solution deposition, the fast turnout time enabled a much more rapid tuning of the deposition recipe.

6.7.4 Metal-Oxide Transducers

Metal oxide sensors have, in many cases, become the dominant resistive sensor used for VOC detection. Various elements show a wide range of selectivity and systems have been demonstrated with sensitivities in some cases going below ppb for ethanol in 50 seconds [45]. Though we state that our system can use any transducer, the only non-polymer transducers attempted were baking yeasts and a peptide 1,3-phenylenediamine [53]. We therefore introduce a short series of experiments whose purpose was to demonstrate additional transducers that are viable and discover some of their characteristics.

It is important to note that the selectivity of these sensors has been shown to be directly dependant on their temperature. This presents a problem in that heating the interferometer between 200 and 400^0C may be detrimental to the photorefractive, let alone damaging to any adhesive layers present. Temperature fluctuations would once again become a significant concern since any pressure fluctuation will then potentially give rise to temperature convective cooling fluctuations. For this reason, we do not increase the temperature of the oxide, but rather operate at room temperature simply to determine if the transducer has potential for our system. If it does then a means of working around the temperature sensitivity can be found later.

The most common elements used include oxides of Tin (Sn), Platinum (Pt), and Tungsten (W). Since a sub ppb ethanol sensitivity was reported using a Pt-doped SnO_2 sensor, we begin with this material on the hopes that it will be sufficiently sensitive at room temperature to respond to 10's of ppm of ethanol. A 20 nm layer of Sn was sputtered onto a 5mm by 1cm glass slide after being thoroughly cleaned. Half of the sputtered samples were then coated with a 5nm sample of Pt. The Sn layer appeared yellow while the Pt-Sn layer took on a shade of purple. To grow an oxide, the samples were then place in an oven at 570^0C for 2 hours with a 2 hour ramp on either side. The only reagent

oxygen was that naturally available from the atmosphere. Both samples turned a greenish purple after baking. Additionally, where the surfaces once appeared perfectly flat to visual reflective inspection, they now show significant scattering. While this is presumably the effect of a non-uniform oxide structure, we cannot rule out other potential surface contaminants.

The actual mechanism for detection using a metal oxide is the oxidation/reduction reaction that occurs between the transducer and the gas it is exposed to. The hydroxyl groups are believed to adsorb via hydrogen bonding, after which they react with the surface ion releasing an electron [57]. Higher temperatures facilitate this process as well as provide enough energy to encourage further oxidation of the surface after the analyte has desorbed thereby improving the recovery time. At least, this is the mechanism used in current sensing applications though it follows that for a finite charge, the same reduction reaction can occur on any such oxide surface. The actual mechanism that gives rise to the detected phase shift is indeterminate since each bonding/debonding action will cause a localized refractive index shift.

A series of experiments were run using each of the two oxidized transducers along with a control of just the bare Sn and Pt transducers with no oxide layer, other than that grown at room temperature. From this data (figure 6-12) we see that the transducer with a grown oxide is significantly more sensitive than that without an intentionally grown oxide (figure 6-13). Further growth tests would be useful to determine if there is an upper limit to the optimum thickness and what limits the optimum. The transducers also appear to be far less sensitive to water than their PNVP counterparts even though an actual LOD for water was not performed.

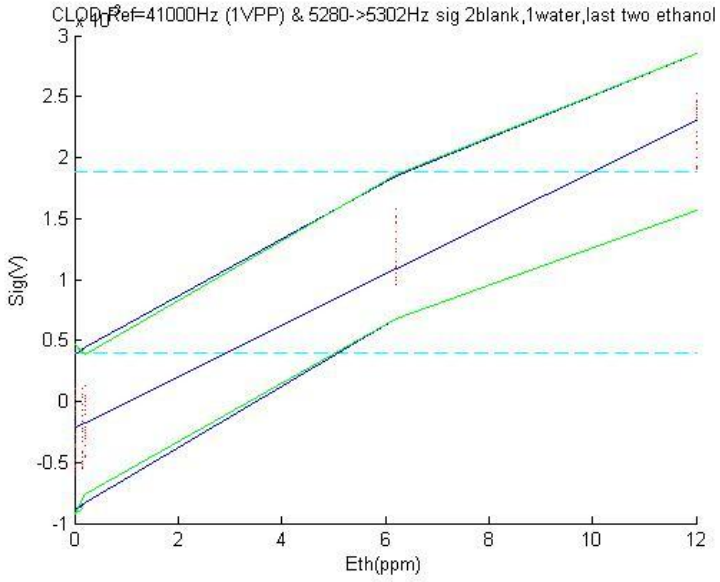

Figure 6-12: Sensor LOD with SnO2 along with two blank runs and a water vapor run at 20 ppm which is indistinguishable from the blank. The sensor is significantly less sensitive to water than the standard PNVP, though no data to calibrate the water LOD was performed.

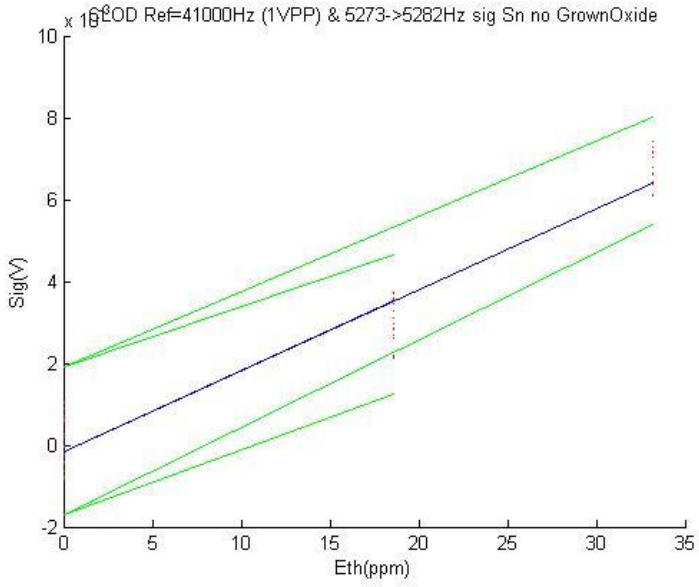

Figure 6-13: Sensor LOD with no oxide growth showing some sensitivity with a LOD of approximately 25 ppm.

From these tests we can conclude that the oxide sensors which are viable though the transducers appear to age or poison faster than PNVP. No data could be acquired due to the settling time but it was noted that within the first couple minutes the transducers would loose substantial sensitivity, appearing to stabilize after about 10 minutes of exposure, and continuing to slowly loose sensitivity over the course of hours of exposure at the 10's of ppm level. Some of the sensitivity could be regained by baking for 30 minutes at 300^0C, presumably desorbing the molecules that have adsorbed with strong affinity or reoxidizing reduced sites on the surface. Since metal oxide sensors are typically run around 200-300^0C [58], it is likely that the higher temperatures help the desorption process thereby decreasing the loading of the transducers.

6.8 Future Work

It would be recommended that improving the transducer sensitivity and selectivity be the focus of any following work, though such efforts would be more suited to chemists and biochemists. The system itself is about as stable as the interferometry will get for an interferometer meant to be used in the field. Since the bulk of the noise limit for chemical transduction comes form the flow itself, it would be useful to put more effort into a sampling and modulation scheme that would introduce little to no pressure fluctuations, perhaps with active servoing and a superior valve design.

One may investigate the possibility of curing the polymer to optimize sensitivity and selectivity to the solvent used. This would be analogous to the technique known as molecular imprinting in which a molecule's "footprint" is essentially stamped into a normally low selectivity polymer to make it more selective. This stamping typically involves the use of a crosslinker, during exposure to the analyte, to lock the base polymer into a shape that, similar to an antibody, is very selective to the unique structure of the analyte [47,59]. A cruder form of imprinting may involve curing with the desired analyte present, or using the desired analyte as the solvent during deposition, to form sites that will be more selective to the analyte, assuming the polymer holds its form well. Such a technique could allow for a very sensitive polymer to also become a more selective polymer. Preliminary experimentation was performed curing low selectivity UV-cured polymer in an environment with the desired analyte, ethanol, which normally shows almost no sensitivity when

cured. Though some results were quite promising, the sum of the results was mixed, likely due to the poor quality of the flash curing procedure, such that no firm conclusions to the viability of this method could be made.

Metal-oxide sensors show promise as a new transducer material though there are problems with poisoning and loading at low temperatures. Effort into developing lower temperature metal-oxides or changing the reaction potential may help these transducers to achieve the sub-ppm sensitivities exhibited by specialized resistive metal oxide experiments run at higher temperatures [60]. One may even fabricate an array of metal-oxide transducers with polymer and biological transducers as long as they will all operate under similar circumstances.

Chapter 7 Summary

7.1 Summary

This thesis presented an adaptive holographic interferometer system for the point detection of airborne volatile-organic chemicals and other hazardous analytes. This system compares favorably to many of the other core technologies (without the use of preconcentrators), with its primary strengths laying in the rapid replacability of cheap transducer elements and fast response and recovery times. This is significant since many systems have the core interrogation components integrated with the transduction elements requiring the core of the sensor system to be replaced when the transducers fatigue, become dirty, or are poisoned.

When considering the various transducer based technologies available, if put to the question "which technology is best?" the answer would have to be "it depends." The cheapest, most readily available transducers are individually packaged conducting polymer and metal-oxide transducers. However, commercial products exhibit less sensitivity and an array is necessary for good discrimination, requiring arrayed or multiplexed electronics. Basic spectroscopic techniques have shown to be capable of fairly rapid detection and identification of ethanol, 125ppb within 30 seconds [78], where the main tradeoff is price and size. If cost and size were not an issue spectroscopic techniques would likely win using a combination of comb sampling in a ring-down configuration.

Photo Ionization detectors have been developed with high sensitivity and response times, <100ppb in 15 seconds, but are completely indiscriminate. Essentially these would act as a useful cheap first-responder unit to identify when something dangerous is present and warrant further examination with a discriminative instrument.

Requirements on portability and low cost while still allowing for discrimination would then suggest that a surface-acoustic wave (SAW), optical nose, or SPR based device be used as they can be adapted in the field to anticipated analytes, and each unit is easily portable with reasonable sensitivities. While SAWs see have high sensitivities (100's ppb in ~min) in laboratory experiments [12], a commercial units testing for ethanol exhibited a practical limit of detection of a few ppm in just over a minute [69]. Additionally, due to some of their high frequencies of operation, SAWs often have rather expensive/complicated electronics [71] and, either due to poisoning or fatigue of the transducer, it is non-trivial to replace the transducers since this replaces the SAW core as well. Yet, once mass-produced, the SAW technology may overcome many of these features due to its high compatibility with IC fabrication technologies.

The Optical nose and SPR technologies are nearly identical in many of their strengths and weaknesses. The transducers in both can be easily and rapidly replaced, both achieve high sensitivity and noise rejection, and both systems respond in a few seconds. On the down side, both are sensitive to wavelength and angle fluctuations in the beams, both measure refractive index and therefore pick up any noise in said index at the transducer interface, and both require relatively thin transducer materials. However, where the optical nose can tolerate low-quality transducer-surface uniformity from batch to batch, requiring only that they are transparent, SPR response depends largely on the thickness and precision of the transducers [61] increasing the cost of replaceable transducers. SPR is also more sensitive to environmentally coupled noise to the transducer, at low frequencies, where the optical nose inherent high-pass filters the detected signal, removing a lot of environmental noise. However, common path SPR techniques have also minimized noise sensitivity by exposing two polarizations to the same transducer reducing noise sensitivity [7]. Where the optical nose can adapt rapidly to any transducer-array configuration, similar adaptability exists for SPR using CCD detection at the output enabling some spatial detection coupled with angular SPR detection.

At this point more research has gone into SPR technology and so it is more developed. The highest reported index sensitivity (2.9e-7) [61] is very close to that of the optical nose (2.4e-7) [Section 2.3]. One of the largest hurdles for the optical nose is the photorefractive crystal, which comprises the bulk of the cost. This is due largely to the low availability of photorefractive crystals, though costs can be minimized using thinner smaller crystals enabling larger yields from a boule. If the cost of the crystal were substantially negated, the optical nose, due to its adaptability, environmental rejection, low-cost transducers, and response speed would be a solid contender in the transducer-based sensor market.

7.2 Thesis Contributions

The primary contributions of this thesis are listed below.

- Design and integration of a prism interferometer based optical nose system for precision, low-signal level measurements
    - Design of piezo-flexure-grating phase modulator
    - Design of a dispersion minimized prism interferometer
- Design and integration of a portable, 660nm diode based, linear arrayed prism interferometer system.
- Design of a sinusoid and triangle wave modulated, continuous flow, vapor-mixing valve.
- While the theory/technique was initiated by Ye et. al [21], this thesis thoroughly explores the modulation enhancement technique used to improve interferometer sensitivity by several orders of magnitude.
- I improved the reliability of vapor concentration delivery and data collection for the optical nose system. This included statistical based data calibration using automated electronic data collection based on the foundation presented by Curie [63].

7.3 Future Work

The underlying theme of this work was to determine the reasonable lower limit of sensitivity and move the technology towards a commercializable, portable technology. Our displacement sensitivity is within an order of magnitude of the shot-noise limit (18 fm) resulting in a strong transducer interrogation base, though, this was achieved only for the 532nm system. We implemented a 660nm diode based system to reduce costs, size, and enable portability of the nose technology as a step towards making a commercializable product. The largest foreseeable hurdles are listed below.

[1] The laser linewidth must be stabilized to within 500 kHz to have any hope of achieving the optimum sensitivity for a portable system that was achieved by the 532nm system. As this work only shows a first generation design, plenty of additional work remains to be completed. It would be advisable, if BaTiO3 remains the recombination element of choice, to not exceed 660nm in wavelength as the photorefractive would respond too slowly and therefore coupling performance would suffer. However, the more developed 780nm would be easier to achieve good stability with meaning that a shift to a rhodium doped BaTiO3 or a different photorefractive may be beneficial. It may be advantageous to laser stability to side-lock to a second VHG or to use the side-mode artifacts that only appear when the VHG has strong cavity feedback. Also, an AR coated diode would likely drastically improve laser stability as it would nearly eliminate competition between cavities.

[2] The valve must be miniaturized further and stabilized such that variances in its rotation rate and in its division ratio are minimal, as these variances will become a large contributor to phase noise and small signal uncertainties. The triangular modulated valve is a good start but one to two more iterations of this design would be necessary.

[3] The detection and processing electronics need to be further developed and miniaturized. For a 4 channel system with 8 analog synchronous detection circuits, the processing footprint of the latest system is a significant improvement above the previous array of lock-in amplifiers.

[4] Transducer development should be borrowed from other transducer-based nose technologies. The basic selection criteria include optically thin/transparent, viable in a continuous air flow, and easily deposited onto glass. However, one can borrow from SPR technology and use an intermediate

nanometer thick layer of gold or other material to assist in binding and immobilization. It may even be possible to use layers requiring heating or an applied electric field as long as the supporting elements do not interfere with the TIR interrogation.

[5] With the reduction of the laser and electronic systems, the photorefractive BaTiO3 is now the single most expensive and fragile components in the system. Since the crystal cannot endure temperatures below around 9^0 C without suffering a catastrophic phase change, this element must either be temperature stabilized at all times or replaced with a more robust recombination element. While it may seem prudent to remove the photorefractive all together, this would negate the prism design and require substantially more expensive transducer elements as the precision of their manufacture would have to increase substantially to achieve comparable interferometry from one array to the next. The simplest solution is to cut the photorefractive crystals smaller which not only would reduce the prohibitive cost, but also decrease the thermal mass that must be temperature controlled in cold environments.

Bibliography

[1] K. Petermann, E. Weidel, "Semiconductor laser noise in an interferometer system," IEEE J.of Q. Elect. **17**, 1251–1256 (1981).

[2] Q. Wei, M. Nieuwenhuyzen, F. meunier, C. Hardacre, S. L. James, "Guest sorption and desorption in the metal-organic framework [Co(INA)2] (INA=isonicotinate) – evidence of intermediate phases during desorption," Dalton Trans. 1807–1811 (2004).

[3] L. Saiz, J. A. Padro, "Structure and Dynamics of Liquid Ethanol," J. Phys. Chem. B, **101**, 78–86 (1997).

[4] J. M. Slater and J. Paynter, "Prediction of Gas Sensor Response Using Basic Molecular Parameters," Analyst. **119**, 191–195 (1994).

[5] M. H. Abraham, "Application of salvation equations to chemical and biochemical processes," Pure & Appl. Chem. **65**, 2503–2512 (1993).

[6] G. Phillips, "An electronic method of detecting impurities in the air," J. of Scientific Instr. **27**, 341–347 (1951).

[7] J. L. Hall, S. Petropavlovskikh, O. Nilsen, B. Hacioglu, "Detection and Discrimination of Low Concentration Gas Contaminants by means of Interferometrically-Sensitive Polymers," IEEE Sensors, **30**, 1366–1369 (2005).

[8] S. Brunauer, P. H. Emmett, E. Teller, "Adsorption of Gases in Multimolecular Layers," J. Am. Chem. Soc. ***60***, 309–319 (1938).

[9] K. E. Geckler, B. Muller, "Polymer materials in biosensors," Naturwissenschaften, **80**, 18–24 (1993).

[10] S. M. Hughes, D. Z. Anderson, "Modulation Enhanced Sensitivity of Holographic Interferometry," Appl. Opt. (2007).Submitted

[11] L. Fadel, F. Lochon, I. Dufour, and O. Français, Chemical Sensing, "Millimeter size resonant microcantilever performance," J. Micromech. Microeng. **14**, S23–S30 (2004).

[12] M. Penza, G. Cassano, P. Aversa, A. Cusano, M. Consales, M. Giordano, L, Nicolais, "Acoustic and Optical VOCs Sensors Incorporating Carbon Nanotubes," IEEE J. Sensors **6**, 867–875 (2006).

[13] D. Z. Anderson, and J. Feinberg, "Optical novelty filters," IEEE J. Quantum Electron. **25**, 635–647 (1989).

[14] P. Delaye, L. A. de Montmorillon, and G. Roosen, "Transmission of time modulated optical signals through an absorbing photorefractive crystal," Opt. Comm. **118**, 154–164 (1995).

[15] Alexei A. Kamshilin and Alexander I. Grachev," Adaptive interferometer based on wave mixing in a photorefractive crystal under alternating electric field," Appl. Phys. Lett. **81**, 2923–2925 (2002).

[16] B.R. Pouet, R. Ing, S. Krishnaswamy, and D. Royer, "Heterodyne interferometer with two-wave mixing in photorefractive crystals for ultrasound detection on rough surfaces." Appl. Phys. Lett. **69**, 3782–3784 (1996).

[17] V. Petrov, C. Denz, J. Petter, and T. Tschudi, "Enhancing the sensitivity of an adaptive holographic interferometer using non-Bragg diffraction orders," Opt. Lett. **22**, 1902–1904 (1997).

[18] P.C.D. Hobbs, "Ultrasensitive laser measurements without tears," Appl. Opt. **36**, 903–920 (1997).

[19] D. Z. Anderson, R. W. Brockett, and N. Nuttall, "Information dynamics of photorefractive two-beam coupling," Phys. Rev. Lett. **82**, 1418–1421 (1999).

[20] V. Damiao, and D. Z. Anderson, "The geometry underlying photorefractive two-beam coupling," J. Opt. A: Pure Appl. Opt. **5**, S536–S549 (2003).

[21] H. Ye, O. Nilsen, V. M. Bright, and D. Z. Anderson, "Holographic chemical vapor sensor," Opt. Lett. **30**, 1467–1469 (2005).

[22] P.C.D. Hobbs, "Shot Noise Limited Optical Measurements at Baseband with Noisy Lasers," SPIE **1376**, 216–221 (1990).

[23] C.M.Lindsay, R.M. Rade, Jr., and T. Oka, "Survey of H_3^+ transitions between 3000 and 4200 cm^{-1}," J.Mol.Spectrosc. **210**, 51–59 (2001).

[24] McMichael, P. Yeh, "Phase shifts of photorefractive gratings and phase-conjugate wave," Opt. Lett. **12**, 48–50 (1987).

[25] P. Yeh, "Photorefractive two-beam coupling in cubic crystals. II.," J. Opt. Soc. Am. B. **5**, 1811–1813 (1988).

[26] T. Y. Chang, J. H. Hong, F. Vachss, and R. McGraw, "Studies of the dynamic range of photorefractive gratings in ferroelectric crystals," J. Opt. Soc. Am. B, **9**, 1744–1751 (1992).

[27] M. Segev, Y. Ophir, and B. Fischer, "Nonlinear Multi Two-Wave Mixing, the Fanning Process and its Bleacing in Photorefractive Media," Opt. Com. **77**, 265–274 (1990).

[28] P. Xie, P. Y. Wang, J. H. Dai, and H. J. Zhang, "Effect of random volume scattering on image amplification and beam fanning in photorefractive materials," J. Opt. Soc. Am. B, **15**, 1889–1894 (1998).

[29] P. Xie, Y. H. Hong, J. H. Dai, and Y. Zhu, "Theoretical and experimental studies of fanning effects in photorefractive crystals," J. Appl. Phys. **74**, 813–818 (1993)

[30] J. Feinberg, "Asymmetric self-defocusing of an optical beam from the photorefractive effect," J. Opt. Soc. Am. **72**, 46–51 (1982).

[31] P. P. Banerjee, and H. L. Yu, "Noise analysis of two-beam coupling in a photorefractive material," J. Opt. Soc. Am. B. **11**, 1809–1812 (1994).

[32] M. Bunsen and A. Okamoto, "Theoretical and experimental studies of holograpm multiplexing that uses a random wave front generated by photorefractive beam fanning," Appl. Opt. **44**, 1454–1463 (2005).

[33] A. Zozulya, M. Saffman, and D. Z. Anderson, "Propagation of Light Beams in Photorefractive Media: Fanning, Self-Bending, and Formation of Self-Pumped Four-Wave-Mixing Phase Conjugate Geometries," Phys. Rev. Lett. **73,** 818–821 (1994).

[34] J. Joseph, P. K. C. Pillai, and K. Singh, “High-gain, low-noise signal beam amplification in photorefractive BaTiO3,” Appl. Opt. **30**, 3315–3318 (1991).

[35] Q. B. He, and P. Yeh, “Fanning noise reduction in photorefractive amplifiers using incoherent erasures,” Appl. Opt. **33**, 283–287 (1994).

[36] “Jitter Measurements for CLK Generators or Synthesizers,” Dallas Semiconductor/Maxim, Application note 2774 (2003).

[37] G.A.Brost, R.A.Motes, “Origin of the sub linear photorefractive response time in BaTiO3,” Opt. Lett. **15**, 1194–1196 (1990).

[38] D. Mahgerefteh, J. Feinberg, “Explanation of the Apparent Sublinear Photoconductivity of Photorefractive Barium Titanate,” Phys. Rev. Lett. **64**, 2195–2198 (1990).

[39] M. Lesaffre, F. Jean, F. Ramaz, A. C. Boccara, M. Gross, P. Delaye, and G. Roosen, "In situ monitoring of the photorefractive response time in a self-adaptive wavefront holography setup developed for acousto-optic imaging," Opt. Express **15**, 1030-1042 (2007)

[40] P. Delaye, L. A. de Montmorillon, and G. Roosen, “Transmission of time modulated optical signals through an absorbing photorefractive crystal,” Opt. Com. **118**, 154-164 (1995).

[41] G. Bonsch, and E. Potulski, “Measurement of the refractive index of air and comparison with modified Edlen’s formulae,” Metrologia **35**, 133–139 (1998).

[42] W. Tyler Estler, “High-accuracy displacement interferometry in air,” App. Opt. **24**, 808–815 (1985).

[43] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (Wiley, 1991).

[44] P. Werle, R. Miicke, and F. Slemr, “The Limits of Signal Averaging in Atmospheric Trace-Gas Monitoring by Tunable Diode-Laser Absorption Spectroscopy (TDLAS),” Appl Phys. B. **57**, 131–139 (1993).

[45] P. Ivanov, E. Llobet, X. Vilanova, J. Brezmes, J. Hubalek, X. Correig, “Development of high sensitivity ethanol gas sensors based on Pt-doped SnO2 surfaces,” Sensors and Actuators. B, 201–206 (2004).

[46] K. R. Rogers, “Recent advances in biosensor techniques for environmental monitoring,” Analytica Chemica Acta. **56**, 222–231 (2006).

[47] Y. Fu, *Synthesis and Characterization of Molecularly Imprinted Polymers and Their Application in Preconcentrators for Gas Phase Sensors.* Dissertation. 2003.

[48] T. C. Pearce, S. S. Schiffman, H. T. Nagle and J. W. Gardner, (eds), Handbook of Machine Olfaction: Electronic Nose Technology. (Wiley-VCH, 2002).

[49] C. Feng, S. Mitra, “Breakthrough and Desorption Characteristics of a Microtrap,” J. Microcolumn Separations, **12**, 267–275 (2000).

[50] Scientific Instrument Services, Inc, “Alcohol Breakthrough Volumes for Adsorbent Resins,” http://www.sisweb.com/index/referenc/bv-alc.htm

[51] P. Fenster, W. K. Kahn, “An Optical Technique for Measurement of Gas Flow Profiles Utilizing a Ring Laser,” **7**, 2383–2391 (1968).

[52] R. F. Barr, H. Watts, “Diffusion of Some Organic and Inorganic Compounds in Air,” J. of Chem. And Eng. Data. **17**, 45 (1972).

[53] H. Ye, *Chemical Vapor Sensing With Dynamic Holography.* Dissertation. 2005.

[54] K. A. Schult, D. R. PAUL, "Water sorption and transport in blends of poly (vinyl pyrrolidone) and polysulfone,” J. Polym. Sci. B **35**, 655–674 (1997).

[55] M. Harbeck, *New Applications of Organic Polymers in Chemical Gas Sensors*. Dissertation. 2005.

[56] N.F.E.I. Nestle, R. Kimmich, “Concentration-Dependant Diffusion Coefficients and Sorption Isotherms. Application to Ion Exchange Processes As an Example,” J. Phys. Chem. **100**, 12569–12573 (1996).

[57] J. Liu, X. Wang, Q. Peng, Y. Li, “Vanadium Pentoxide Nanobelts: Highly Selective and Stable Ethanol Sensor Materials,” Adv. Mater. **17**, 764–767(2005).

[58] B. B. Rao, V. J. Rao, “Sensitivity behavior of tin oxide based semiconducting sensor for fluorocarbons,” J. of Materials Sci. Lett. **18**, 1493–1495 (1999).

[59] M. Yoshikawa, “Molecularly imprinted polymeric membranes,” Bioseparation. **10**, 277–286 (2002).

[60] B. Esfandyarpour, S. Mohajerzadeh, A. A. Khodadadi, M. D. Robertson, “Ultrahigh-Sensitive Tin-Oxide Microsensors for H_2S Detection,” IEEE Sensors Journal, **4**, 449–454 (2004).

[61] K.H. Chen, J.H. Chen, J.Y. Lin, “Comparison of Kretschmann-Raether configuration angular and thickness regimes with phase-difference shift for measuring changes in refractive index,” Opt. Eng. **45,** 023803-1–023803-5 (2006).

[62] Hubaux, and G. Vos, “Decision and Detection Limits for Linear Calibration Curves,” Anal. Chem. **42**, 849–855 (1970).

[63] L.A.Currie, “Detection: International update, and some emerging di-lemmas involving calibration, the blank, and multiple detection decisions,” Chemometrics and Intel. Lab. Sys. **37**, 151–181 (1997).

[64] H. Butt and M. Jaschke, “Calculation of thermal noise in atomic force microscopy,” Nanotechnology **6**, 1–7 (1995).

[65] J. R. Barnes, R. J. Stephenson, C. N. Woodburn, S. J. O’Shea, M. E. Welland, T. Rayment, J. K. Gimzewski, and C. Gerber, “A femtojoule calorimeter using micromechanical sensors,” Rev. Sci. Instrum. **65**, 3793– 3798 (1994).

[66] P. Maddaloni, G. Gagliardi, P. Malara, P. De Natale, “Off-axis integrated-cavity-output spectroscopy for trace-gas concentration measurements: modeling and performance,” J. Opt. Soc. Am. B, **23**, 1938-1945 (2006).

[67] M. J. Thorpe, K. D. Moll, R. J. Jones, B. Safdi, J. Ye, “Boadband Cavity Ringdown Spectroscopy for Sensitive and Rapid Molecular Detection,” Science **311** 1595–1599 (2006).

[68] D. James, S. M. Scott, Z. Ali, W. T. O’Hare, “Review: Chemical Sensors for Electronic Nose Systems,” Microchim. Acta. **149**, 1-17 (2005).

[69] E. J. Staples, "Detection of Ethanol in Water and Air Using the zNose," Electronic Sensor Technology (Estcal). Technical Paper.

[70] P.Ivanov, E. Llobet, X. Vilanova, J. Brezmes, J. Hubalek, X. Correig, "Development of a high sensitivity ethanol gas sensor based on Pt-doped SnO2 surfaces," Sens. Actuators. B **99**, 201–206 (2004).

[71] K. Arshak, E. Moore, G.M. Lyons, J. Harris, S. Clifford," A review of gas sensors employed in electronic nose applications," Sensor Rev. **24**, 181–198 (2004).

[72] J. Ge, J. Wang, H. Zhang, X. Wang, Q. Peng, Y. Li, "High ethanol sensitive SnO2 microspheres," Sens. Actua. B. **113**, 937–943 (2006).

[73] R. K. Joshi, F. E. Kruis, "Influence of Ag particle size on ethanol sensing of SnO1.8: Ag nanoparticle films: A method to develop parts per billion level gas sensors," Appl. Phys. Let. **89**, 153116 (2006).

[74] A.J. Haes, R.P. Van Duyne, "Preliminary studies and potential applications of localized surface plasmon resonance spectroscopy in medical diagnostics," Expert Rev. Mol. Diagn. **4**, 527–537 (2004).

[75] T. C. Pearce, S. S. Schiffman, H. T. Nagle, J. W. Gardner, *Handbook of Machine Olfaction: Electronic Nose Technology.* (Wiley, 2003).

[76] Burl MC, Sisk BC, Vaid TP, Lewis NS," Classification performance of carbon black-polymer composite vapor detector arrays as a function of array size and detector composition ," Sen. Actuators B, **87**, 130–149 (2002).

[77] J. Shi, Y. Zhu, X. Zhang, W. R. G. Baeyens, A. M. Garcie-Campana, "Recent developments in nanomaterial optical sensors," Trends in Analytical Chemistry, **23**, 351–360 (2004).

[78] A. Kosterev, R. F. Curl, F. K. Tittel, C. Gmachl, F. Capasso, D. L. Sivco, J. N. Baillargeon, A. L. Hutchinson, A. Y. Cho, "Effective utilization of quantum-cascade distributed-feedback lasers in absorption spectroscopy," Appl. Opt. **39**, 4425–4430 (2000).

[79] S. M. Cho, Y. J. Kim, Y. S. Kim, Y. Yang, S. C. Ha, "The application of carbon nanotube-polymer composite as gas sensing materials," IEEE Proc. 701–704 (2004).

Appendix A Shot-Noise Limited Signal-to-Noise Ratio

Any measurement of an optical signal will have an inherent noise due to the quantization of the photons and electrons being measured. The signal is a mean number of detection events per second while the noise, or degree to which this number varies, is defined as the standard deviation of the detection events. In this appendix I work through various aspects of S/N when limited by shot noise, ignoring all other sources, especially in regards to several conventions for measurement I've frequently encountered in optics and electronics.

The shot-noise limited S/N is defined as the average signal count $\bar{n}$ over the standard deviation (std) of that signal σ_n.

$$S/N = \frac{\bar{n}}{\sigma_n} = \frac{\bar{n}}{\sqrt{\bar{n}}} = \sqrt{\bar{n}} \tag{A.1}$$

Since it is tedious, if not absurd, given the number of events per second in may laser systems, to count individual photons, the S/N can be expressed as the ratio in the power of the signal P_s over the std of the detected power, or the effective noise power P_n, given by.

$$S/N = \frac{P_s/2B\hbar\upsilon}{\sqrt{P_{opt}/2B\hbar\upsilon}} = \frac{P_s}{\sqrt{2B\hbar\upsilon P_{opt}}} = \frac{P_s}{P_n} \tag{A.2}$$

The energy per photon is given by $h\nu$, the optical power for a given number of counts n is $2B(h\upsilon)n = P_{opt}$, and B is the bandwidth over which our measurement is taken where $2B=1/T$, where T is the measurement interval [43]. We discriminate between signal power P_s and total optical power P_{opt} because if the interferometer's contrast is not 100% or if the signal of interest is only a portion of the total power, then the entire power detected contributes to quantum noise. Since one issue with our system is what happens when we converted from 532 to 660 nm, we express the S/N in wavelength.

$$S/N = \sqrt{\frac{\lambda P_{opt}}{2Bhc}} \tag{A.3}$$

We examine the wavelength dependence, as this shows that if all else remained the same, the S/N will improve with the root of the wavelength. This a useful point in previous sections which noted that to improve the S/N, wavelength was increased to slow the photorefractive thereby, improving its signal response to low frequencies, and, as just shown, to also fundamentally improve the S/N, Again, this

method assumes all else remained the same, which it didn't because the linewidth of the new laser was over an order of magnitude worse than the previous linewidth.

Note that we can use two definitions for sigma based on what is being detected, though they are identical.

$$\begin{aligned} \sigma_n{}^2 &= \bar{n} \\ \sigma_{n\_pow}^2 &= 2Bh\nu / P_{opt} \end{aligned} \tag{A.4}$$

where the first is based on counts and the second is based on power. The definition for sigma must be defined and held consistent based on what is being measured. This difference does not affect the S/N as shown above. The optical S/N, if the total power detected is the signal, scales with the square root of power or of the number of counts.

There is a second means of defining S/N that is useful, and in wide use due to its practicality. The electric, rather than optical, S/N is given as the ratio between the electronic signal power (P_{s_e}) vs. the standard deviation of this power or the effective noise power (P_{n_e}).

$$S / N_e = \frac{P_{s_e}}{P_{n_e}} \tag{A.5}$$

Essentially there is no fundamental difference because both are the uncertainty associated with counting statistical events that behave nearly identically. To convert from optical to electrical we must be familiar with both the quantum efficiency η (electrons-per-photon) and the responsivity $\Re$ (Amps-per-Watt) of our conversion, typically performed by a photodiode. With these we can convert between photon events *n*, photoelectron events *m*, optical power incident on the detector P_{opt} , and electrical power P_e.

To detect the optical information above, we measure a photocurrent.

$$i_{opt} = 2Bqm = q\eta \frac{P_{opt}\lambda}{hc} = \Re P_{opt} \tag{A.6}$$

Since the responsivity $\Re$ is typically the only value quoted on photodiode datasheets, one will most commonly use the relation of photocurrent based on the detected optical power, $i_{opt} = \Re P_{opt}$. For completeness, we give a few relations between these variable components that occasionally prove useful.

$$\eta = \frac{m}{n} = \frac{photoelectrons}{photon} < 1 \tag{A.7}$$

$$\Re = \frac{i_{opt}}{P_{opt}} = \eta \frac{q}{h\nu} = \eta \frac{\lambda}{1240(nm)} = \frac{Amps}{Watt} \tag{A.8}$$

$$m = \frac{i_{opt}}{2Bq} = \eta \frac{P_{opt}\lambda}{2Bhc} \tag{A.9}$$

Continuing, if we then apply a gain to the photocurrent, an effective responsivity is defined including gain.

$$\Re_G = \frac{Gi_{opt}}{P_{opt}} = G\eta \frac{\lambda}{1240(nm)} = G \cdot \frac{Amps}{Watt} \tag{A.10}$$

This is useful if the current from a photodiode were amplified or divided by some circuitry in a detector. With current gain, which is <1 for a current divider in a Hobbs circuit, the detected electrons scale proportionately.

$$m = G\frac{i_{opt}}{2Bq} = G\eta \frac{P_{opt}\lambda}{2Bhc} = Gn\eta \tag{A.11}$$

Before proceeding, we provide a brief discussion on noise. Noise, if represented by fundamental shot noise is defined by a Poisson distribution, is a function of the randomness of detection events. The randomness is noise because for a constant power we may receive 95 photons in one second, and then 105 photons the next, which is within one standard deviation and easily expected. Since the detector just saw a 10% increase in optical power, or detection events, even though the total average power is constant, this statistical variance directly translates to a power noise. It is the randomness that acts as an effective power fluctuation over finite bandwidths.

The variance of the photocurrent $\sigma_i^{\,2}$ is given based on photoelectron counts $\sigma_m^{\,2}$ as

$$\sigma_m^{\,2} = m = \frac{\bar{i}_{opt}}{2Bq} = \eta n = \eta \frac{P_{opt}\lambda}{2Bhc} = \frac{P_{opt}\Re}{2Bq} \tag{A.12}$$

$$\sigma_i^{\,2} = 2Bq\bar{i}_{opt} = (2Bq)^2 \sigma_m^{\,2} = q^2 2B\eta \frac{P_{opt}\lambda}{hc} = 2BqP_{opt}\Re \tag{A.13}$$

Unlike optical power, which is proportional to the number of photons, electrical power is proportional to the square of the number of electrons due to the definitions of optical and electrical power. The S/N for electric current is then expressed, assuming a portion of the total detected power is that of the signal $i_{sig} < i_{opt}$, as

$$S/N_e = \frac{\bar{m}^2}{\sigma_m^2} = \left(\frac{2Bq\bar{m}}{\sqrt{(2Bq)^2 \bar{m}}} \right)^2 = \left(\frac{i_{sig}}{\sqrt{2Bqi_{tot}}} \right)^2 \tag{A.14}$$

Where the total detected current i_{tot} may also include the amplifier and resistor Johnson noise $i_{tot}^2 = i_{opt}^2 + i_{amp}^2 + i_R^2 + \ldots$ (for values see section 5.4), but for the remaining calculations we still focus just on the shot limit. In terms of electrical power, when measuring electric current, the S/N_e is expressed as follows.

$$S/N_e = \left(S/N_{current}\right)^2 = \left(\frac{i_{sig}}{\sqrt{2qBi_{tot}}} \right)^2 = \frac{i_{sig}^2 R}{i_{nse}^2 R} = \frac{P_{s_e}}{P_{n_e}} \tag{A.15}$$

Where i_{nse} is the effective current noise. And for measurements of electrical voltage signals, assuming a transimpedance factor R(ohms), we find

$$S/N_e = \frac{P_{s_e}}{P_{n_e}} = \frac{V_{sig}^2/R}{V_{nse}^2/R} = \left(S/N_{volt}\right)^2. \tag{A.16}$$

Again, this is shown for completeness, illustrating that there are no discontinuities when converting from one convention to another. At this point we show a simple conversion between electronic and optical S/N based on detection counts and the quantum conversion efficiency.

$$\frac{\bar{m}^2}{\sigma_m^2} = \frac{(\eta\bar{n})^2}{\sigma_n{}^2\eta} = \eta\frac{\bar{n}^2}{\sigma_n{}^2} = \eta\bar{n} \xrightarrow{gain} G\eta\bar{n} \tag{A.17}$$

$$S/N_e = G\eta\left(S/N_{opt}\right)^2 \tag{A.18}$$

To summarize this rather long exercise in redundancy, we give a relation for the signal-to-noise ratio in the forms that are most applicable to the experiment. The first is a relation of optical power

$$S/N_{opt} = \frac{P_s}{P_n} = \frac{P_s}{\sqrt{2B(h\nu)P_{opt}}}, \tag{A.19}$$

which can only be measured in terms of power… which we detect via electronic means… so it's rather worthless, but a good starting point. Its use comes in when calculating the power in the signal arising from phase modulation, as this is a starting point to determining the limit of detection. The second is a relation of the electronic signal-to-noise, which is a little more useful. It shows the base power relation followed by detected current and voltage for calculations, and then gives a relation to optical S/N explaining why you can never do as well as the optical limit.

$$S/N_e = \frac{P_{s_e}}{P_{n_e}} = \left(\frac{i_{sig}}{\sqrt{2qBi_{tot}}} \right)^2 = \left(S/N_{volt} \right)^2 = \eta S/N_{opt}^2 \qquad \text{(A.20)}$$

Finally, I give a voltage based S/N, as this is what I detect from the demodulation circuitry as recorded by the computer in voltage. It is this relation (S/N_{volt}) that is referred to in the entire thesis by S/N in any experimental result. A factor of two is added to the signal because the subtraction from both photodiodes increases the effective signal by up to a factor of 2, or more precisely, 1+G where G<1. We therefore also assume that the total power detected is the sum of the average power on both diodes after photorefractive two-beam coupling with resulting coupling angle β.

$$S/N_{volt} = \frac{2i_{sig}}{\sqrt{2qBi_{tot}}} = \frac{2\left\{ P_{opt} \cdot J_1\left(\delta_{ref}\right) \frac{2\pi(\Delta x)}{2 \cdot 532nm} \sin(2\theta_{Powsplit}) \sin(2\beta) \right\} \sqrt{\frac{\eta q \lambda}{hc}}}{\sqrt{2BqP_{tot}}} \qquad \text{(A.21)}$$

$$S/N_{volt} = \frac{2P_{sig}\sqrt{\Re}}{\sqrt{2BqP_{tot}}} = \frac{2\left\{ P_{opt} \cdot J_1\left(\delta_{ref}\right) \frac{2\pi(\Delta x)}{2 \cdot 532nm} \sin(2\theta_{Powsplit}) \sin(2\beta) \right\}}{\sqrt{\frac{2BqP_{tot}}{\Re}}} = 1 \qquad \text{(A.22)}$$

The second expression is the most useful for our work as it allows the theoretical detection shot-noise limit (S/N=1) based on experimental parameters, to be easily compared to the achieved limit.

If you've made it this far I may as well give the punch line to this story. The tabletop reference system V3 has an achieved displacement LOD of 180 fm/Hz$^{1/2}$. The sensitivity is for the conditions of P_{opt} = 2mW, $\Re \approx 0.33 A/W$, $\delta_{ref} \approx 0.1 \cdot 2\pi$, $\theta_{Powsplit} = 58^0$, $\beta \approx 40^0$, B=1, and q is the electron charge. The theoretical shot noise limit is 8-9 fm/Hz$^{1/2}$. See Chapter 5 and Appendix C for the actual limiting factors. We never achieved the shot noise limit because other noise sources dominated. One might wish to increase the optical power to make the system shot-limited however this power level, 2mW detected out of the photorefractive, was selected because for a given signal amplitude the signal-to-noise was optimum.

Appendix B Limit of Detection

The limiting factor in small signal detection is the certainty in a detection event. The standard signal-to-noise S/N has been discussed, though it approached limits of detection from counting statistics. The effective count is determined based on the optical power. However, here we take a look at how the actual signal generated from the system is the detection decision is based upon this signal. This exercise is necessary because a simple signal-to-noise calculation is inadequate to discuss limits-of-detection in the field of trace gas detection, where heteroscedasticity is often observed. Signal-to-noise assumes the noise is relatively constant, regardless of the noise source. That is to say, the standard deviation is the same for a blank (no analyte present), as it is for a strong analyte presence. A brief overview follows of the methodology that yields the smallest detectable content level within a degree of confidence that basic S/N overlooks. For a more complete analysis of limits-of-detection refer to the works by A.Haubaux [62] et all and L.A.Currie [63].

The standard deviation is defined based on the expected value of a signal, assuming the content giving rise to said signal is constant. Regardless of the dominant noise source, assuming normally distributed fluctuations, its standard deviation is calculated based on the difference between each data point (x_n) and the mean ($\bar{x}$) of the set of size N.

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{B.1}$$

The standard deviation σ can then be defined based on the mean value of the population.

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x_n - \bar{x}\right)^2} \tag{B.2}$$

On inspection it becomes apparent that the standard deviation will change slightly when sampling a different population set unless the sample population is unreasonably large (unreasonable in time necessary to collect such a sample). A small correction factor is often included when sampling populations effectively yielding a sample standard deviation.

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left(x_n - \bar{x}\right)^2} \tag{B.3}$$

The signal-to-noise is then simply the ratio of the mean to the standard deviation.

$$S/N = \frac{\bar{x}}{\sigma} \tag{B.4}$$

One of the difficulties of defining a limit of detection is the decision for a detection event is desired for the shortest time span possible. Otherwise, people die while waiting to discover that they should have evacuated. The societal impact of a false positive and false negative are just as severe as a decision that comes too late. From this point forward, a simple view is taken which requires an explicit measure of the blank in the calibration data to ensure adequate representation variations in the standard deviation with content level. Given a set of normally distributed data, the probability that a data point will fall within *n* standard deviations from the mean is graphically shown in figure b-1.

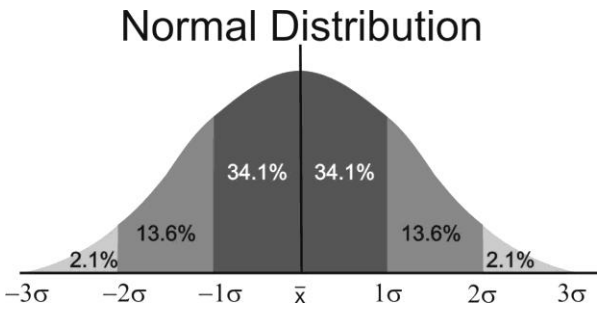

Figure B-1: Normal distribution showing probability that a data point will lie within *n* standard deviations of the mean.

Since a decision as to whether a detection event occurred must be made, and within an acceptable confidence level, a decision point must be defined. When the decision point is defined for a point *n* standard deviations above the signal corresponding to the blank, then effectively *S/N=n*. This is often interpreted as a S/N of 3 with a 99.7% confidence level that the signal is due to a detection event. However, comparing the actual spread of data points due to a blank versus the spread in signal whose content, and therefore mean value, corresponds the 3 sigma cutoff, we see that there is a 50% likelihood of an erroneous detection event as shown in figure b-2. To achieve a 99.7% confidence that a *content* will be detected above the critical signal level, the content should correspond to a signal mean 3-standard deviations, of the signal, plus an additional 3-standard deviations, of the blank, above the blank.

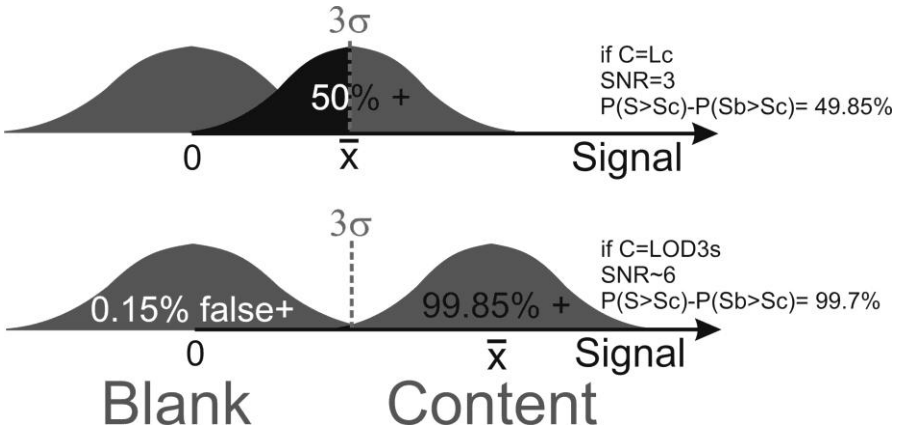

Figure B-2: Top: The confidence that a signal above the 3 sigma cutoff is due to a content at the 3 sigma limit is roughly 50%. Bottom: The confidence that a signal above the 3 sigma cutoff is due to a content at the LOD=3 sigma, or roughly the 6 sigma, is roughly 99.7%. Therefore to have traditional 3 sigma confidence, 99.7% of the limit of detection must be established as the content corresponding to a signal 3 standard deviations (measured of the signal) above an additional 3 standard deviations (measured of the blank) above the blank. A homoscedastic dependant variance, meaning the variance of the blank and of the signal are identical, this implies S/N=6.

Mathematically, this is expressed in terms of S/N=1 as a function of the signal S corresponding to a content C and the standard deviation of the signal σ_S and of the blank σ_B.

$$\bar{S}_{S/N=1} = \frac{\bar{x}}{\sigma_B} = \frac{\bar{S}_{S/N=3}}{3} = \frac{\bar{S}_{LOD3\sigma} - 3\sigma_S}{3} \tag{B.5}$$

Or in terms of the 3 sigma limit of detection.

$$\bar{S}_{LOD3\sigma} = \frac{3\sigma_B + 3\sigma_S}{\sigma_B} \approx 6\bar{S}_{S/N=1} \tag{B.6}$$

where the limit of detection is only comparable to 6 times the S/N=1 limit for a homoscedastic system.

One significant difficulty with a heteroscedasticity is that the function dictating the standard deviation may be nontrivial and must be determined empirically. One way to determine a limit of detection is to perform a linear regression with upper and lower confidence levels as shown in figure b-3. This figure shows how, using a data taken for calibration, the limit of detection can be determined by the intercept of the lower -3σ error bar with the critical level defined at $3\sigma_B$. The figure additionally shows how variation in the standard deviation can influence the LOD.

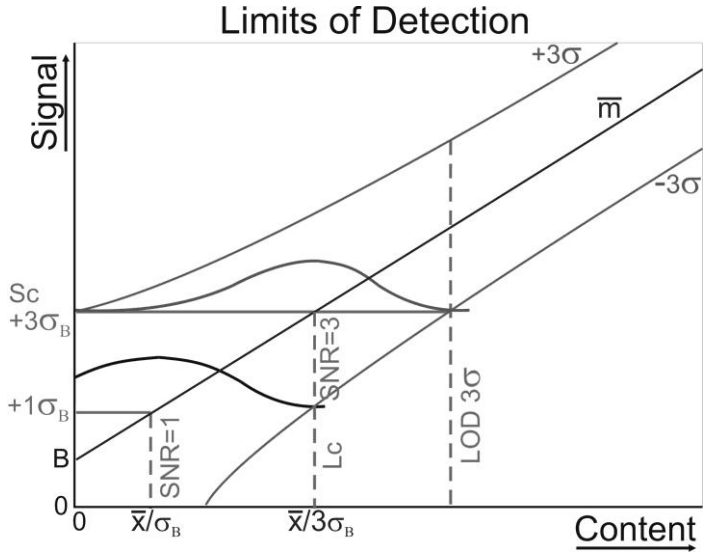

Figure B-3: Limit of detection values derived from a calibration curve using the intercepts of the 3 sigma error bars. Since the standard deviation is not constant, the 3 sigma Limit of detection is not equal to twice the content that corresponds to an average S/N=3.

From this information one can easily determine S/N of 1, 3, or LOD of 3-sigma using the calibration curve. In many cases it is a reasonable estimate to convert between LOD and S/N=1 by dividing by six, though this conversion is not done in this thesis as such a conversion is often less than honest. The primary reason for this discussion is to bridge the gap, which is often glossed over, between engineers, who are used to S/N, and those involved in Chemometrics, who may be more likely to quote a LOD. This also reveals a questionable way that some may choose to report their data as they will pick the value that sounds most impressive in their community, such as S/N=1 while such a content level is not reasonably detectable and says nothing about the heteroscedastic nature, or lack there of, of their system.

When taking data for a LOD plot, one must collect enough data to accurately estimate the standard deviation of the data and concurrently calculate a stable mean. The question is, how many samples are sufficient to accurately estimate the mean? A simple way to discover the answer is to run an experiment and plot the standard deviation with respect to the number of data points collected for a fixed content. Figure B-4 shows just such an experiment from which we can say that the mean is stable to within 20% of its value by about 30 data points while 70+ would be preferred for more

accurate traces. The majority of LOD plots will therefore contain at least 30 signal data points per content point.

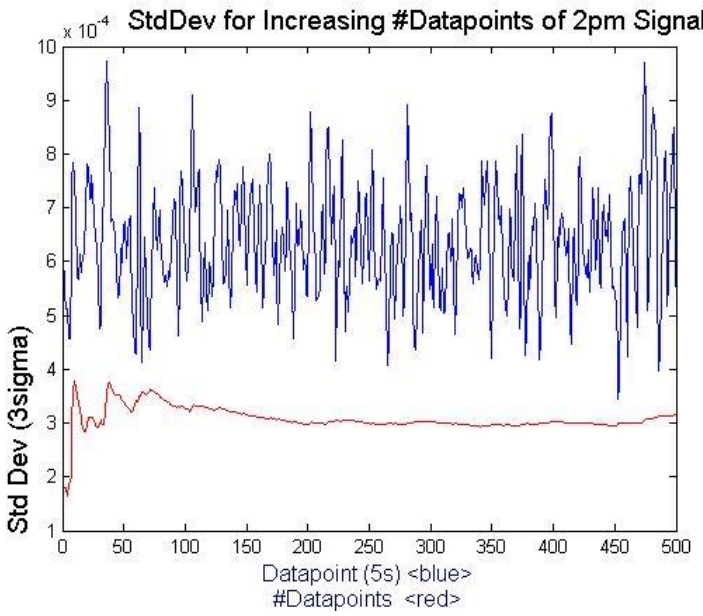

Figure B-4: Evolution of the Standard deviation with increasing sample counts. Each point represents a 1 second integrated sample point with a sampling frequency of 1Hz. The data start point was intentionally shifted to begin at the noisiest region ensuring a reasonable upper limit is established.

We present a few LOD traces for displacement and chemical content, as well as a table showing various limits of detection taken at a few points over a year. We begin with a basic displacement calibration in which the PZT has a 5Hz signal added to the reference oscillator. Figure B-5 shows two LOD traces taken early in the life of PV3. The first looks at the 10 Hz second harmonic response without modulation enhancement, and the second looks at the mixed frequency signal. The LOD for the first is estimated to be as bad as 300pm though later data led us to conclude that the LOD was as bad as 600pm. The enhanced detection scheme yielded a LOD of 350 pm or a S/N=1 at 180 fm/$Hz^{1/2}$.

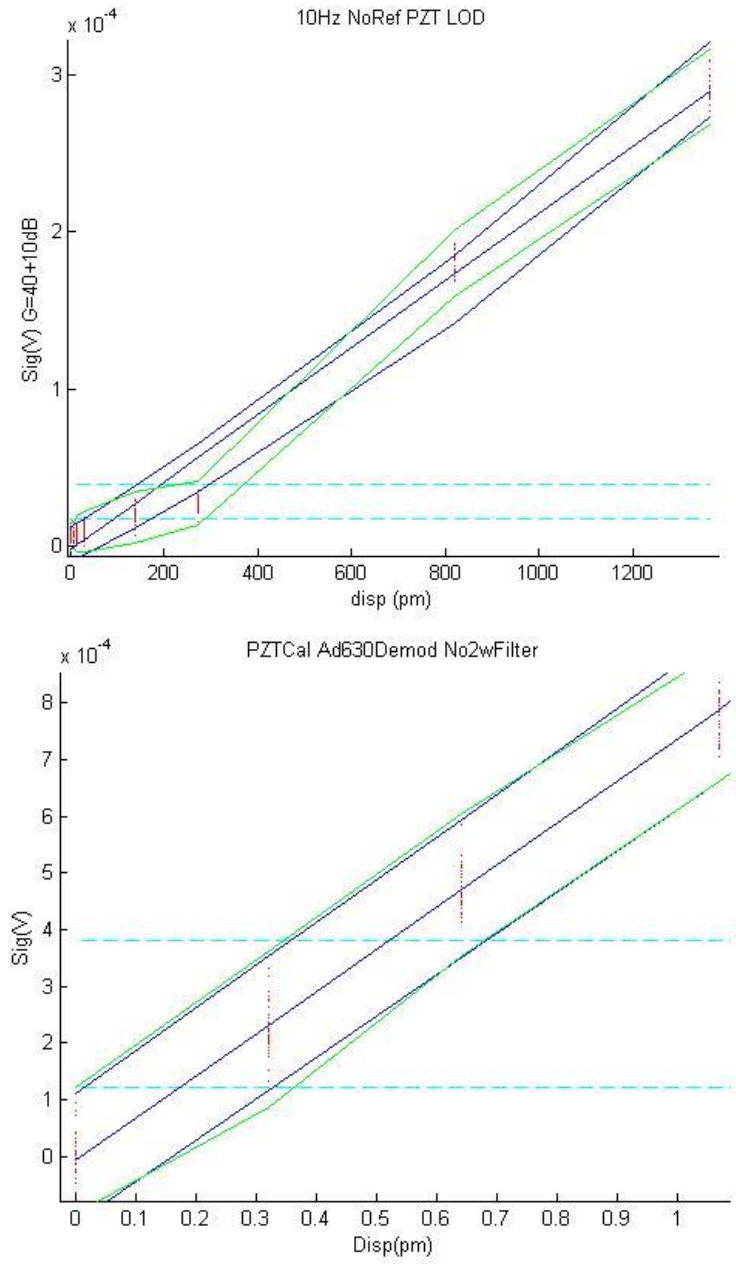

Figure B-5: LOD calibration traces for the Top: the unenhanced system detecting the 10Hz second harmonic and Bottom: the enhanced system detecting the mixed signal. Little data was taken of the unenhanced system though an upper limit of sensitivity is estimated at twice what's shown due to the uncertainty in the standard deviation. The enhanced system shows a stable variance for the 50 data points collected per content point.

In this next experiment, an ethanol source was diluted from 16ppt to a few ppm and the concentration level was checked with an Ion Science Phocheck 5000+ Photoionization detector. The plot shows a S/N=1 LOD of around 1.8 fm/Hz$^{1/2}$, for a 5 Hz modulation and 5 s integration time

constant. Running at 3Hz managed to get a LOD of roughly half this value though an unstable valve and other technical issues made a 5Hz modulation a more stable long term modulation frequency.

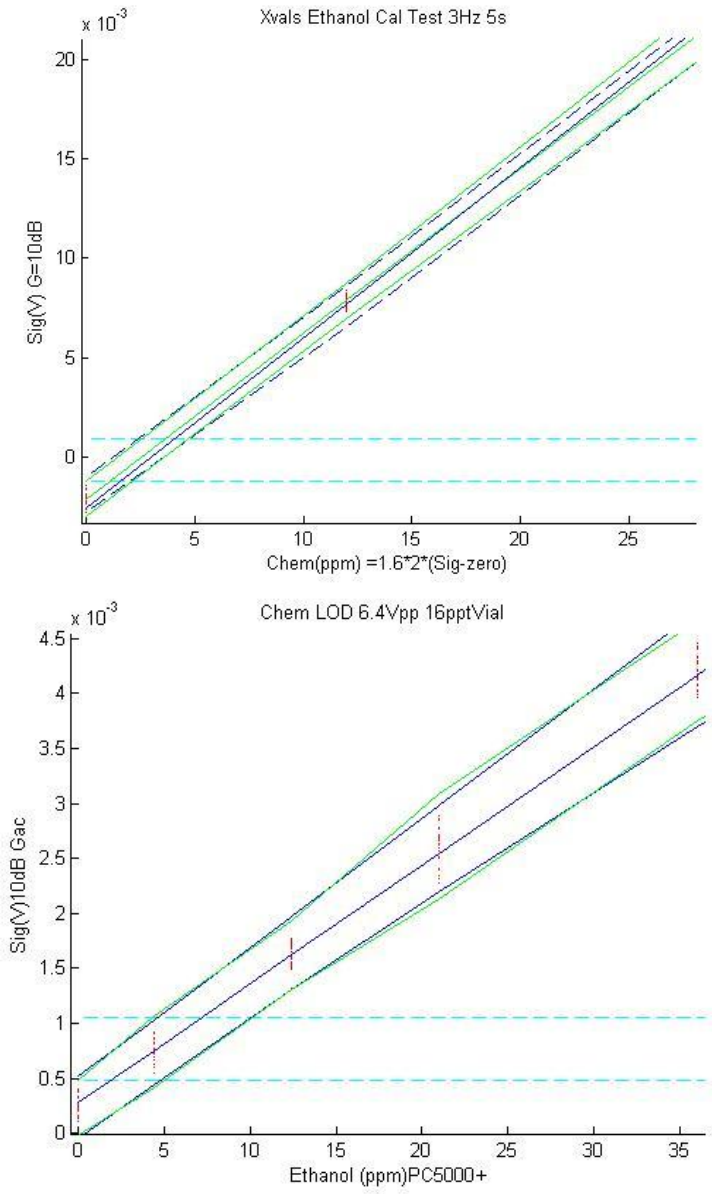

Figure B-6: . Example LOD calibration data taken for Prism system V3 at Top: 3 Hz signal frequency and at Bottom: 5 Hz signal frequency. Ethanol was detected using a PNVP transducer to levels of 4.5ppm with a 99.7 % confidence that a detection event occurred. The lower horizontal dashed line corresponds to 3 sigma of the blank while the upper corresponds to 10 sigma. There are 35 data points per content level, taken in order, starting with the blank. The 3 Hz signal is superior to that of the 5 Hz signal due to the low pass filtering of the laminar flow tubes. However, 5 Hz is the standard presented because the rotary valve brushless motor was unstable around 3 Hz, making this data extremely hard to obtain without glitches (hence only 3 content points used). Additionally, the layout in PV3 made it impractical to increase the gearing ratio by 40%.

The LOD calibration data in table b-1 was collected for PV3 for 4 points over half a year. The data is shown because the reported values are derived from the average of these sets. Additionally, this data shows that the system ages with time, which was difficult to show before automated data collection was established. However, this table does not represent enough data sets to show conclusively the trend of the decay. It only shows that performance ageing does occur.

LOD 3s for 5Hz signal					**LOD 3s for 100Hz signal**		
3-15-06	3-28-06	8-2-06	8-20-06		3-15-06	8-2-06	8-20-06
LOD3s 5hz	LOD3s 5Hz	LOD3s 5hz	LOD3s 5hz		20-100hz	100Hz	100Hz
400	350	630	*440*		100	270	*160*
300	450	560	*460*		120	220	*200*
350	390	650	*600*		110	240	*190*
300	380	670	*510*		110	290	*160*
350	340	520	*410*			170	*190*
350	350	670	*530*			200	*180*
370	550	700	*600*			200	
320	400	600	*500*			260	
420	500	710	*420*			220	
330	400	660	*530*			250	
370		650				240	
		540				200	
351 fm	411 fm	630 fm	500 fm		110 fm	230 fm	180 fm

Table B-1: Limit of detection for a 5s integration time. This table shows decay of the system with age except on 8-20-06, in which the algorithm was modified to discriminate against signals proportionate to the phase angle. The system was also cleaned just prior to 8-20-06.

Figure B-7 shows the affect of weighting on a set of data where the STD at low content levels decreases. The algorithm used is simply a weighting that is unity when the phase angle is 0 and zero when the angle is 90^0.

$$X' = X\left|\frac{2}{\pi}\arctan\frac{X}{Y}\right| = X\left|\frac{2}{\pi}\theta_{x/y}\right| \tag{B.7}$$

The phase discrimination effectively lowers the standard deviation of signals that are dominated by noise and typically only adjusted the LOD by 10-20%. At best this algorithm will only lower the STD of the blank to $\pi/4 \approx 0.79$ for random noise.

$$\frac{\sigma}{\sigma_0} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{2}{\pi}\theta\right|d\theta\right)^2} = \frac{\pi}{4} \tag{B.8}$$

Where N is the number of samples and the equation assumes the phase angle of the noise of the blank is randomly distributed between 0 and 2π.

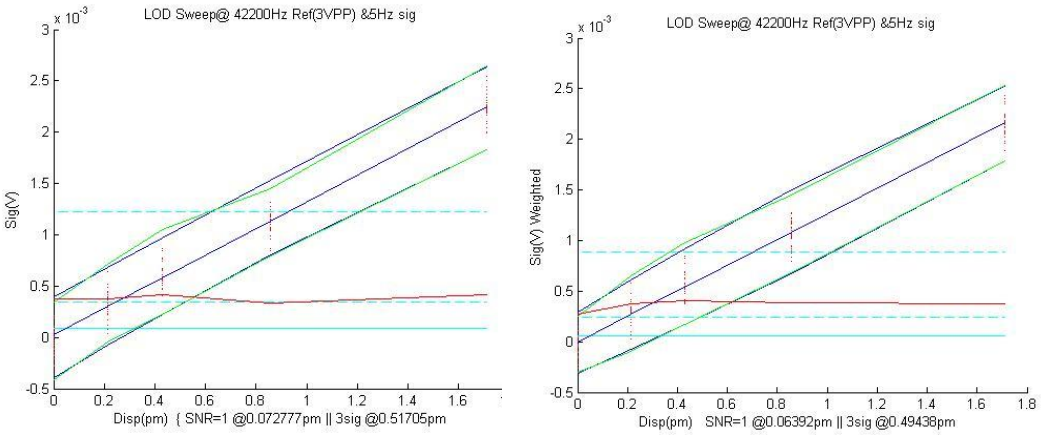

Figure B-7: Weighted LOD calibration curve of PV3 with the latest analysis algorithm. The latest algorithm calculates the LOD 3s intercept as well as the S/R=1 intercept. The nearly horizontal red line is the 3 sigma noise level while the solid horizontal line is the level of $1\sigma_B$ used to find S/N=1.

We have shown how the LOD is determined with 99.7% confidence and how this value relates to the S/N ratio. We have further shown that LOD data for displacement sensitivity to be 180 fm/Hz$^{1/2}$ and Ethanol sensitivity to be 2 ppm/Hz$^{1/2}$ for a 5 Hz signal. More data will be presented elsewhere as is appropriate to its section, but this section effectively gives the average performance of the optical nose system and how that performance is estimated. Furthermore, a basic algorithm was introduced that improves noise discrimination by up to 20% by decreasing the weight of noisier signals and in turn lowering the standard deviation of the blank.

Appendix C Noise Table

LOD Data from Prism system V3

5Hz, 2mW: S/N_1~85fm = 180fm/(Hz)$^{1/2}$

100Hz, 2mW: S/N_1~30fm= 67fm/(Hz)$^{1/2}$

5Hz, ethanol: S/N_1~220fm=510fm/(Hz)$^{1/2}$

LOD Data from Prism system V4

5Hz, 1.8mW: S/N_1~600fm = 1.7pm/(Hz)$^{1/2}$

Source	**Reported Values/ Real values**	**Equations/Explanations**	**Effective Displacement: Worst case**	**Concern**
Laser				
Linewidth/ Frequency Stability	$\Delta\nu_{532} < 0.5-2MHz$ $\Delta\nu_{660} < 3-10MHz$ $\Delta r = 0.1-0.5mm$	$\delta x_{532nm} \approx 3.1\cdot 10^{-15}\Delta\nu(\Delta r - x)$ $\delta x_{660nm} \approx 4.5\cdot 10^{-15}\Delta\nu(\Delta r - x)$ [Section 5.7]	**Max** 532nm=> 3.1pm 660nm=>22.5pm **Min** 532nm=> 150fm 660nm=>1.3pm	High
Shot Noise	$2mW(2mW)$ $532nm(660nm)$ $\theta_0 = 58^0$ $\beta = 45^0$ $\Re = .25(.45A/W)$ $\delta_{ref} = .092\cdot 2\pi$ $B = 1$	$\left(\frac{P_{opt}\cdot J_1(\delta_{ref})\frac{2\pi(\Delta x)}{2\lambda}\sin(2\theta_0)\sin(2\beta)}{\sqrt{\frac{2BqP_{tot}}{\Re}}}\right) = 1$ [Appendix A]	532nm=>18fm/Hz$^{1/2}$ 660nm=>17fm/Hz$^{1/2}$	SHOT

Amplitude Noise/ Relative Intensity Noise	$P = 2mW$ $\Delta P/P < 0.005\%_{532}$ $\Delta P/P < 0.01\%_{660}$ $E.Suppression$ $\approx 34dB$	$0.0001 \cdot 10^{-34/20} = 2\left\{ J_1\left(\delta_{ref}\right) \frac{2\pi(\Delta x)}{2 \cdot 532nm} \sin(2\theta_0) \sin(2\beta) \right\}$ [equation 3.9]	530nm=>0.3pm 660nm=>0.8pm	High
660nm LD: Temp/ current	$\Delta T < 0.4mK/s$ $dP/dI = 1.1W/A$ $\Delta I < 300pA/s$ $d\lambda/dT = 130nm/{}^0C$ $L_{LD} = 200um$ $L_{ExtCav} = 1.5cm$	$\frac{dP}{dI}\Delta I \cdot 10^{-34/20} = 2\left\{ P \cdot J_1\left(\delta_{ref}\right) \frac{2\pi(\Delta x)}{2 \cdot 532nm} \sin(2\theta_0) \sin(2\beta) \right\}$ $\Delta x \approx \frac{L_{LD}}{L_{ExtCav}} \left(d\lambda/dT \, \Delta T \right)$ [Above] This calculates amplitude and frequency stability based on current/temperature stability. Noise due to actual measured amplitude and frequency stability is given in prior rows.	1 fm due to current. 700 fm due to temp.	High
PZT Phase Modulator				
Heating: Drift in grating position	$CTE \approx -4 \cdot 10^{-6}/K$ $x_0 = 2mm$ $\Lambda = 416nm$ $\Delta T_{typ} < 3e^{-4}\left(K/s\right)$	$\Delta x = -4 \cdot 10^{-6} \Delta T \left(\frac{\lambda}{\Lambda} \right) x_0$	3pm/s Until steady state is reached	Low

Mechanical Thermal Noise:	$E_{PZT} = 0.6GPa$ $K_B = 1.3e^{-23} J/K$ $T = 300K$ $B = 5Hz$ $L = 2mm$ $A = 2x2mm$ $Q_{ElectMech} < 4$ $\omega_0 = 50kHz$	$\Delta P = \frac{E}{L}\Delta L$ $\frac{1}{2}mv^2 = \frac{k_s \Delta x_{th}{}^2}{2} = \frac{AE}{L}\frac{\Delta L^2}{2} = \frac{3}{2}\Big\vert_{3D} k_B T$ $\omega_0 = \sqrt{k_s / m} = \sqrt{AE / Lm}$ $\Delta x_{RMS}\left(m/\sqrt{Hz}\right) = \sqrt{\frac{3k_B TBL}{AE\omega_0 Q}}$ $NSE = \Delta x_{RMS} \cdot \left(\frac{\lambda}{\Lambda}\right)$ [64]	0.7fm	Low
Capacitive/ Johnson Thermal Noise:	$PZT = 2.2um/100V$ $C = 0.025uF$ $k_B = 1.3e^{-23} J/K$ $T = 300$ $R = 1e^6$ $B = 5Hz$	$V_n = \sqrt{4k_B TRB}$ $V_n = \sqrt{\frac{k_B T}{C}}$ $\Delta x_{RMS} = \left(\frac{\lambda}{\Lambda}\right)\left(\frac{2.2 \cdot 10^{-6} m}{100V}\right)\sqrt{\frac{k_B T}{C} + 4k_B TRB}$ [Section 5.4]	6fm	Low
Grating Scatter: AM mod	$\Delta P / P\Big\vert_{45kHz} < 0.01\%$ $V_{\omega_R+\omega_S} / V_{\omega_R} = 3e^{-6}$	$2mW \cdot 3e^{-6} \cdot 1e^{-4} \cdot 10^{-34/20} = 12\,fW$ $12\,fW = 2\left\{P \cdot J_1(\delta_{ref}) \frac{2\pi(\Delta x)}{2 \cdot 532nm} \sin(2\theta_0)\sin(2\beta)\right\}$ [Equation 3.9]	2am	Low
Ref Mod Frequency Purity	Voltage Purity +5Hz 110dB Agilent/Osc 102dB Crystal/Osc First Harmonic vs 2nd Harmonic -20dB	$\frac{V_{\omega_R+\omega_S}}{V_{\omega_R}} = 10^{-dB/20}$ $10^{-(20+110)/20} J_2(\delta_R) = J_1(\delta_R) J_1(\delta_S)$ [Equation 5.6]	9fm$_{\text{AgilentOsc532nm}}$ 11fm$_{\text{AgilentOsc 660nm}}$ 26fm$_{\text{Crystal/Osc 660nm}}$	Med-set

BaTiO3				
Imperfect 90^0 grating shift	First Harmonic vs 2nd Harmonic -20dB I set $\delta_R \le 0.1\lambda$ since higher values were seen to limit the LOD	$10^{-(20+110)/20} J_2(\delta_R) = J_1(\delta_R) J_1(\delta_S)$ $= \sin(2 \cdot \Delta\phi_{PR}) \cdot \left(J_1\left(\delta_R \frac{V_{\omega_R+\omega_S}}{V_{\omega_R}} \right) \right) J_0(\delta_S)$ [Equation 5.7]	180fm For which the grating phase shift must be $<0.5^0$	High-set
High-Pass Filter	@5Hz	$H(\omega_s) = \frac{1}{1+\frac{\omega_{HPF}^2}{\omega_s^2}}$ $\omega_{HPF} = \tau_{PR}^{-1} \approx \frac{1}{\tau_0}\left(\frac{I_0}{I}\right)^{-0.64}$ $P_{sig} \approx H(\omega_s) P_{sig_0}$ [Section 5.8.1]	Attenuates the low frequency optical signal and noise equally. Is not a noise term.	None
Thermal Noise: Vibration of grating	$\theta_{\text{int}} \approx 8^0$ $\Lambda_{BaTiO} \approx 15-19um$ $L = 3mm$ $A = 5mm \times 2mm$ $E = 67GPa$ $Q < 1$ $m_{PR} < 1g$ ω_{0PR} □ $470kHz$	$\Lambda_{BaTiO} = \frac{2\lambda}{\sin\frac{\theta_{\text{int}}}{2}}$ $\Delta x_{RMS} = \sqrt{1.56*10^{-20}(J)\frac{BL}{AE\omega_0 Q}}$ $NSE = \Delta x_{RMS} \cdot \left(\frac{\lambda}{\Lambda_{BaTiO}}\right)$ [65]	27am	Low
Photo-refractive Fanning		Not practical to obtain analytically. Experiments using an incoherent erasure beam failed to show S/N improvement suggesting fanning is not one of the dominant noise sources.		

Prism				
Thermal Noise: =>OPD	$E_{BK7} = 65GPa$ $K_B = 1.3e^{-23} J/K$ $B = 5Hz$ $L = 20mm$ $A = 8x10mm$ $n_{BK7} = 1.51$ $CTE = 86e^{-7}/K$ $\rho = 6g/cm^3$	$\Delta x_{RMS} = n_g \sqrt{1.56*10^{-20}(J)\frac{BL}{AE\omega_0 Q}}$ Unless orthogonal, both arms should effectively cancel since they start and end at the same point and are roughly common path.	=26fm for wo=5,Hz Q=1	Low
Index Noise: bulk	$p_{str.opt.} = .12 - .3$ $\Delta r < 0.5mm$ $\Delta L = 26\,fm$	$\Delta n(x,y,t) = -\frac{1}{2} p_s n^3 \frac{\Delta L}{L}$ $\Delta n^*(\Delta r) = dx$ [43]	<1/70th of above result	Low
TIR				
Index of air: Pressure Thermal	$\Delta T_{N2} < 0.12K/\min$ $\Delta P \sim 24Pa$	$(n-1)_\lambda = 10^{-8}\left[8091.37 + \frac{2,333,983}{130 - \left(1/\lambda(um)\right)^2} + \frac{15,518}{38.9 - \left(1/\lambda(um)\right)^2}\right]$ $(n-1)_{tp} = \frac{(n-1)_\lambda \cdot p(Pa)}{93214.6} \cdot \frac{1+10^{-8}\left(0.009,876 \cdot t(^0C) \cdot p(Pa)\right)}{1+0.003,661 \cdot t(^0C)}$ [41]	T=0.12 ^{0}C [Δn =9^{-8}] =>70fm 24Pa, [Δn =6.6e^{-8}] =>52fm	Med-High with direct airflow

Index of Surface: Pressure	L=10um Ps~0.3$_{Polystyrene}$ E=.6-6GPa$_{Polmer}$ P=24Pa n=1.51	$\Delta n_p = -\frac{1}{2} p_s n_p^{\,3} \frac{\Delta P}{E}$ $\phi_{TM} = 2\arctan\left(\frac{n_p^2 \sqrt{\sin^2(\theta_{TIR}) - \frac{n_a^2}{n_p^2}}}{\cos(\theta_{TIR}) n_a^2} \right)$ $\Delta x = \frac{\lambda}{2\pi} \frac{d\phi_{TM}}{dn_p} \Delta n_p$ [43]	[Δn =2e^{-8}] =>15fm	Med
Polymer				
Thermal noise	L=10um A=.2mmx.2mm E=.6-6GPa$_{polymer}$ n=1.54 $_{PNVP}$	$\Delta x_{RMS} = n\sqrt{1.56*10^{-20}(J)\frac{BL}{AE\omega_0 Q}}$ $NSE = \Delta x_{RMS} \cdot 2_{worstcase\_normal\_incidence}$	80-240 fm Q=1,B=5,wo=5Hz Yet removing the polymer doesn't improve the displacement LOD so this must be an overestimate	Med
TIR thermal noise:	L=10um Ps~0.3$_{Polystyrene}$ E=.6-6GPa$_{Polmer}$ P=240Pa n=1.54 dL =80-240fm	$\Delta n(x,y,t) = -\frac{1}{2} p_s n^3 \frac{\Delta L}{L}$ [43]	dL=22fm=>dn=1.3e-9=>0.6fm dL= 300fm=>dn=4e-9=>1.8fm	Low

Detector:				
Diode I_{dark} and NEP	A=5.1mm^2 R=0.24A/W(.45) C=40pf (-3V) Id=3nA NEP=280fW/Hz$^{1/2}$ (5.6fW/Hz$^{1/2}$)	$2\left\{P\cdot J_1\left(\delta_{ref}\right)\frac{2\pi(\Delta x)}{2\cdot 532nm}\sin(2\theta_0)\sin(2\beta)\right\} > 280\,fW$	$\Delta x \approx 2am$	Low
Opamp(x2) MAX410	4nV/rtHz 2.6pA/rtHz R_{fb}=10kohm (50kohm) P_-/P_+~1/3 P_{tot}=2mW $\Re$ =.25A/W (.45A/W)	$\langle i_n\rangle^2 = i_{JR_f}^2 + i_{d\_tot}^2 + i_{n\_opamp}^2 + \left(\frac{V_{n\_op}}{Rf}\right)^2$ $\langle i_n\rangle^2 = \frac{4K_BTB}{R_f} + 2q(\Re P_{tot}B)_{shot} + i_{n\_opamp}^2 + \left(\frac{V_{n\_op}}{R_f}\right)^2$ $\frac{\langle i_n\rangle_{rms}\cdot 2\sqrt{2}}{\Re} = 2\left\{P\cdot J_1\left(\delta_{ref}\right)\frac{2\pi(\Delta x)}{2\cdot 532nm}\sin(2\theta_0)\sin(2\beta)\right\}$ $\langle V_n\rangle = \langle i_n\rangle R_f$ [Section 5.5.1, 43]	$13\,pA/\sqrt{Hz}\,(2mW\_532nm)$ $\rightarrow 23\,fm/\sqrt{Hz}$ $17\,pA/\sqrt{Hz}\,(2mW\_660nm)$ $\rightarrow 20\,fm/\sqrt{Hz}$	Low
Demod. Noise: Reference Jitter	Agilent33250A Jitter $d\phi_R < .01\%$ P=2mW $\omega_R = 45kHz$ $\delta_R = 2\pi\frac{20\,pm}{V}V_{R_{PP}}$ $\phi_R < 0.2^0$	$fund\_leak = 2\left\{P\cdot\sin(2\cdot 4^0)\cdot J_1(\delta_R)J_0(\delta_S)\sin\left[\omega_R t\right]\right\}$ $\bar{V} = 2P\left\{\sin(8^0)\cdot J_1(\delta_R)\right\}\frac{1}{T}\int_t^{t+T}\left\{\sin(\omega_R t+\phi_R)\right\}\cdot\sin\left[\omega_R t\right]dt$ $\frac{d\bar{V}}{d\phi_R} = P\left\{\sin(8^0)\cdot J_1(\delta_R)J_0(\delta_S)\right\}\sin(\phi_R)$ $\left\{\sin(8^0)J_0(\delta_S)\right\}\sin(\phi_R)d\phi_R = 2\left\{\frac{2\pi(\Delta x)}{2\cdot 532nm}\right\}$ [Section 5.5.2]	0^0 Synch Error 0 fm 0.2^0 Synch Error 11 fm	Med

Appendix D Machine Olfaction Technologies

D.1 Introduction

With so many technologies in existence it may be difficult to determine the best vapor detector for a given task. While focusing on a particular technology may yield exceptional results regardless of the technology chosen, it may be advantageous to use technologies with orthogonal fundamental detection principles. Orthogonalization helps to ensure that the strengths of each dissimilar technology covers the others' weaknesses. For this reason we present an overview of the varied technologies in existence with their inherent fortes and foibles. This is by no way a complete listing as new variations and innovations are occurring rapidly in a post 9-11 world and while some technologies have many branching development lines, others hybridize in the search of the nose technology to solve various problems.

Chemical olfaction detectors can effectively be broken into 3 categories; biological, spectroscopic, and transduction based. Biological includes animals with known sensitivities such as canines, which can be trained to detect a variety of analytes in the ppb level, and canaries, for CO and CH_4 detection. Biological can also include many labeled detection methods that, via a chemical reaction, will indicate the presence or absence of an analyte. To some extent these can be considered transduction-based detectors since the means of sensing is often a flourophore or visual inspection. However, by the same argument a canine and canary are chemical-to-audio/visual transducers, so they still fit in the same category with different readout methods. With proper chemistry and sufficient reagents for anticipated analytes, rapid and very precise detection is attainable.

Spectroscopic techniques use characteristic electronic or nuclear vibrations and mass characteristics to analyze and identify a specific species. These methods often rely on probing the analyte with radiation and looking for either absorption or radiation characteristics. Ionization is also used as the energy of ionization and the effective charge to mass ratio is very unique for molecules. The strength of these methods is that with prior knowledge of probably analyte responses, the ability to identify an analyte is limited by the spectral database and the ability to probe or measure multiple wavelength responses.

Transducer methods include any system through which a transduction material interacts with the analyte, altering a fundamental property which is measured via electronic or optical interrogation of the transducer. Transduction-based detection is often very non-specific with the exception of a few specifically designed transducers. Their strength is that extensive signal processing is seldom needed and arrays of transducers with varying sensitivities enables rapid identification over a broad solution space. On the down side, transducer based detectors are notorious for poor drift compensation, their low selectivity, minimal redundancy and the potential for poisoning and loading of the transducer.

D.2 Performance Criteria

Justifying the production of any commercial product necessitates that it has some advantage over pre-existing and imminent technologies. Electronic noses have been in development for the past few decades as potentially tireless and reliable replacements for human and canine olfactory detection to fill a growing need for continual monitoring and safeguarding. In order to mark progress and note improvements a set of standard attributes are examined.

- Sensitivity
- Selectivity/ Environmental interference rejection
- Response time
- Reproducibility
- Dynamic range
- Size
- Cost

Selectivity is the measure of a system's ability to discriminate and selectively identify an analyte, especially in the presence of other similar or interfering analytes. A misidentification can have drastic consequences, such as an innocent suspect being arrested for possession of a biological weapon or a crop of contaminated food escaping into public consumption. A sensor that has little-to-

no selectivity can be used as no more than a warning that something of import may be present, necessitating closer and more specific examination.

The sensitivity of an olfactory system must be sufficient to detect dangerous or important analytes before they reach dangerous concentrations. Such critical points may be defined as an unhealthy exposure limit or the flash point of a leak around a chemical reactor. High sensitivity allows for time critical response to be possible before levels surpass a critical point. However, dynamic range must be coupled to sensitivity, as some sensors may be capable of detecting ppb yet they saturate at ppm. While it may be imperative to detect a minor leak before it becomes a severe problem, the inability to give warning to the severity of a major leak could be catastrophic. For example, detecting a methane leak before it becomes a big leak is good, but knowing that the leak is well beyond explosive concentrations may change the response procedure to minimize the risk to life and equipment.

The time from the initiation of sampling to the output of a reliable response must be coupled with sensitivity to get a real world important response. Response time can be affected by several factors. The electronic processing of a system may take a few seconds before a conclusive response signal is generated. A system may take time for the reaction mechanism to interact with analyte to the limit of its ability at a particular concentration level. Longer electronic integration times may improve the S/N proportionate to the square root of the integration time. Finally, many systems will have a means of preconcentrating a chemical sample, integrating/collecting particles over time and then releasing them quickly to improve response proportionately to the total volume sampled. During the preconcentration time, these systems typically never see the analyte and so rather than seeing a growing signal output, the output is null until preconcentration is complete, after which the additional time delays still play their role. Many systems can integrate for minutes or hours to improve their sensitivity proportionately, yet such integration does no good if a critical level or exposure limit was exceeded in that time. A balance is therefore required between response time, sensitivity, and often selectivity.

For a system to be reliable, it must be able to reproduce results within acceptable tolerances. Reproducibility can suffer from variances between sensor and transducer batches, especially in

systems where either of these are replaced. Many systems will age over time, causing a baseline drift, which, without differential detection methods or constant recalibration, negates the small signal sensitivity. While it may be possible to overcome such problems with recalibration, the frequency of such a procedure may significantly add to the cost of the systems maintenance.

Many systems will experience a drastic drop in performance between the lab and a real world environment. A system's ability to ignore environmental influences, including dust, humidity, and temperature, can be a critical impediment to its usefulness in the real world. Some systems may perform amazingly in a controlled clean lab with pure NO2 carrier flows, but introduce a bit of humidity or sometimes even a speck of dust and performance can degrade and possibly even cease.

The size and cost are often decision points for any system. Size, price, and performance are the big three for the bottom line of any product. Pick any two, but choosing the third requires a careful balancing act with potentially significant tradeoffs. A system may seem perfect as it fits on a watch and can detect sub ppb levels of anything, but the price will be astronomical.

For the rest of this chapter we will primarily focus on sensitivity and response time since, for many systems, selectivity can be influenced by a variety of factors including constituents of an array or integration time. Additionally, selectivity is difficult to standardize between technologies since few papers and technical documents report a system's ability to discriminate between two similar analytes.

D.3 Olfaction Technology Overview

Having established a basic starting point, we will survey several common nose technologies, in no particular order. A brief description of the prominent technologies is given although for detailed information, the reader is referred to the citations in each section. The following list is by no means complete.

D.3.1 Spectroscopy Based Detection

The complex combination of subatomic particles comprising each atom and the forces between these particles gives rise to the electron energy levels uniquely available to an atom or

molecule. From this basic principle the concept of spectroscopy, the identification of a molecule by its characteristic absorption, emission, or scatter, is born. However, as gas absorption, as described by Beer's law, is relatively small at one atmosphere, one of the largest weaknesses of this system is the requisite long interaction length necessary to obtain decent signal to noise. A simple solution in many cases uses a multipass cavity in which a shorter volume is probed often up to 10,000 times, improving effective path length. This solution can suffer from saturation in the absorption spectrum, in which further exposure of a volume of gas no longer absorbs the laser, yet the multi-pass cavity still adds noise iteratively and can potentially broaden the apparent absorption such that the line appears broader, reducing spectral resolution.

A resonant multipass can be used in what is referred to as Cavity Ring-Down Spectroscopy (CRDS) [66,67]. The resonant Q is typically dictated by the mirror reflectivity R and the absorption rate within the cavity. The beam is walked in such a cavity so that the same volume is not excessively sampled, thus decreasing saturated absorption affects. The multipass sampling allows for very long path length interaction on the order of km. CRDS does not suffer from laser intensity variations as the measurement is only of the cavity decay time. Unfortunately, the cost of the high reflectivity mirrors and the laser systems can be substantial due to the quality required. Excitation of more than a single mode can be achieved using pulsed or modelocked lasers. However, the mirrors for such cavities are even more expensive, as the high reflectivity must now be combined with dispersion balancing.

Mass spectroscopy often involves the use of radiation exposure to ionize or break a molecule into smaller atomic and molecular ion fragments. Laser-induced breakdown spectroscopy or LIBS does just this with picosecond to femptosecond beams. However, LIBS suffers from the inability to discriminate between the detection of complex molecules, and the detection of molecules with similar atomic constituents but slightly varying ratios or bond orientations. Femtosecond excitation enables such discrimination without breakdown, through the laser is not eye safe and must be well contained.

There are a plethora of other spectroscopy based systems that operate on niche improvements, though many of which suffer the same limitations. Spectroscopy in general requires long interaction lengths or some means of analyte condensation to obtain decent signal to noise, which then requires large volumes or expensive high-Q resonators. The quality of the laser source can be critical

depending on the techniques used, especially if narrow absorption features close in wavelength must be discriminated. Various modifications show immunities to some error sources, including fluctuations in laser amplitude or wavelength, temperature, pressure, humidity, and interferant analytes. Typically, as the systems are made less sensitive to more of these sources, the system complexity increases dramatically, requiring more stringent control of temperature, wavelength, and mode spacing.

Source: [78]

Ethanol LOD: 125 ppb

Integration time: 30 s

D.3.2 Photo-Ionization Detection

While technically this section could fall under mass spectroscopy, implementations of this technology typically lacks the ability to discriminate based on the ionization energy, other than to say a particular analyte's ionization energy was low enough to be excited sufficiently by the UV source. Otherwise, since these detectors primarily measure the current of the ionized species, they can measure small 10's ppb concentration levels given knowledge of the analyte being detected. The primary reason we include this device is because it is commercially available, and it is one of the references with which we measure our ethanol concentrations as a second check against mass loss.

Not only do PID's lack discrimination but, even though it can report 10's ppb in under 5 seconds, its drift limits the practical detection by a few ppm. We have found, using an Ion Science Phocheck 5000+, that we can detect down to 100 ppb of ethanol in 15 seconds. However, this is only true if we have a purge or reference gas with which we test for the first and last 5 seconds, of the 15 second test, establishing the contrast necessary to see past the system drift.

The drift is likely due to battery power as the lamp power will scale with the charge of the batteries. This detector is an excellent first responder unit in that it will report if something is in the air, but it cannot report if it has detected chlorine gas or ethanol.

Company/Product: Ion Science/ Phocheck 5000+

Ethanol LOD: 100 ppb

Integration time: 15 s

D.3.3 SAW and BAW Devices

A gravimetric sensor is one which responds to the mass of the analyte interrogated. One of the more dominant technologies in this area is that of acoustic wave detection, of which there are two primary types, bulk acoustic wave (BAW) and surface acoustic wave (SAW). The most common BAW devices are called quartz-crystal microbalances (QMB or QCM).

The basic acoustic wave sensor uses a piezoelectric substrate to conduct an RF electric input between electrodes via a mechanical stress wave. Sensing is performed by an analyte sensitive layer, often a polymer, deposited in the wave propagation region. The QCM devices produce a wave traveling through the bulk of the material operating in the 10's f MHz. SAW's use a surface acoustic mode, allowing the structure to be thinner and operating frequencies to reach the low GHz. This increase in frequency yields a larger frequency response for the same sensor area. However, the electronics often exhibit higher noise at higher frequencies [68].

SAW devices can be fabricated en masse by standard batch lithographic processes since the piezo substrate needs to support a surface wave and electrodes are easily deposited on the top surface. This allows for arrays of SAW devices to easily be batch processed and die mounted. BAW devices often require more complicated MEMS fabrication processes to achieve the desired structures. The use of MEMS fabrication makes production of arrays more difficult than SAW devices making SAW's a simpler choice to implement in large arrays.

Acoustic wave devices are sensitive to several interferences, including temperature, humidity, acceleration, and drift. Temperature and humidity are significant since introducing an air sample over the transducer element may introduce noise from these fluctuations in the environment. Paralleling identical devices with different transducer layers can help to mitigate some of these interference contributions while improving selectivity via arrayed detection.

Quality in coating thickness is also critical since non-uniformities increase noise and thin layers are needed for fast transduction response [68]. SAW devices are noted to have response times on the order of seconds to minutes. Additionally, the batch to batch uniformity is critical since, as the chem-sensitive layer ages, fatigues, or is covered in dust. Replacement of the transducer requires replacement of the entire acoustic wave device. While SAW devices in batch lithographic processes can be made relatively cheaply, the batch-to-batch continuity may require tuning of the detection electronics or otherwise limit the overall performance.

Product: zNose, [12]

Ethanol LOD: 5ppm [69] 0.7ppm [12]

Integration time:15s_ 60s recovery

D.3.4 Metal-Oxide-Semiconductor Sensor

Metal oxide semiconductor sensors operate by measuring the resistivity of the semiconductor as it is exposed to the volatile organic chemical. Often operated at temperatures between 200 and 500^0 C, these sensors show a range of selectivities based on temperature and dopants used in the semiconductor oxide. For instance, Platinum doping of Tin-Oxide has enabled a laboratory experiment to achieve ppb ethanol detection limits in 50 seconds [70]. However the selectivity of these transducers is still comparable to that of polymers requiring an array of transducers to discriminate against humidity and interfering analytes.

The oxide sensor responds to ion absorbed oxygen at the surface by oxidizing, removing oxygen from the air, or reducing, letting the surface oxygen react with the volatile organic chemical. Upon reduction, electrons are released into the conduction band, increasing the conductivity of the sensor and also reducing the oxygen surface density, while oxidation traps electrons from the conduction band, thereby increasing resistance.

The higher temperatures enable the reversibility of the transducers while improving sensitivity by providing more energy for the redox reaction kinetics. Higher temperatures also improve reaction times, enabling SnO sensors to cycle within 20 seconds to a minute [71]. Commercially, to maintain

operation temperature, these devices often have power consumptions around 800mW per sensor, leading to large power requirements that scale with array complexity. At lower temperatures, the reversibility is hindered since the transducer is more prone to chemisorption of the ambient humidity, which effectively loads the transducer [68]. The transducer can be further poisoned by exposure to sulfur-containing-compounds and ethanol uptake, which can reduce sensitivity to other VOC's [71].

Laboratory experiments yield limits of detection from the ppb's to 10's of ppm. However, many of the tests achieving below 1ppm require synthetic air and are in extremely well-controlled conditions with have integration times over a minute [70,72,73]. Commercial models typically have LOD's in the ppm such as Figaro's TGS2602.

Product: Figaro TGS2602

Ethanol LOD: 1-10 ppm

Integration time: 15 s

D.3.5 Surface Plasmon Resonance

In a Surface plasmon resonance (SPR) system, a laser bounces off of a thin metallic interface, usually gold or silver, in the transverse magnetic (TM) orientation. The evanescent electric field interacts with the metallic surface, generating a surface plasmon and, depending on the refractive index at the metal surface, may be resonantly absorbed. The output beam intensity is then monitored, often with respect to reflection angle, to determine if a detection event has occurred. Localized surface plasmon resonance can achieve refractive index sensitivities down to 1ppm with typical integration times between 0.1 and 1000s [74].

In many ways, this device has similar strengths and weaknesses as our system. To reduce the contribution of phase noise sources, the beam can be split into two polarizations before surface plasmon interrogation such that a TE mode, which has minimal sensitivity to the analyte, enables common mode subtraction. The sensitivity of the device is fine-tuned by angle and thickness of the metallic surface and deposited sensor such that there is a particular angle/thickness optimum. Since many sensor systems operate with a small angle spread being monitored, the angle of resonance shifts

as the refractive index shifts. However, refractive index has high sensitivity to temperature, and pressure. SPR is most often used for fluid sensing rather than air sensing since the standards application uses monolayered biosensors such as antibodies. Arrayed detection is possible although difficulties can arise in crosstalk between angular resolution and spatial resolution if angle detection of individual array sections is performed.

Company: AlphaSniffer [7]

Ethanol LOD: < 1ppm

Integration time:5s

D.3.6 Fiber Optic

The optical interrogation of a transducer can operate on several principals. If a transducer is coating an optical fiber with reduced or stripped cladding, then the optical interaction can yield information through absorption, fluorescence, and optical path length [75]. A simple setup would involve a photodiode at the end to measure the drop in intensity as the optical interaction with a transducer absorbs or couples light from the guiding mode. Similar coupling or evanescent interaction with the transducer can yield fluorescent reactions when an analyte is present, allowing wavelength filtration to reduce the optical background. The path length change, due to evanescent interaction with a transducer of changing refractive index, can be used for interferometry. The sensitivity of the device is limited only by the sensitive length of the fiber, though noise would likewise accumulate.

Fiber optic sensors have several strengths and weaknesses, although one of their largest strengths is the backbone established by the telecom industry in making system fabrication at certain wavelengths simple and affordable. Fiber sensors can be paralleled, making common mode subtraction easy, and industry fiber couplers make replacing fibers quick. However fiber sensors also suffer from vibrations and environmental effects through the fiber which compound as fiber lengths increase. Fiber lasers for telecom help establish low-noise laser sources that are industry standard, though only in a few select wavelengths. Fluorescence measurements often require shorter wavelengths that are more difficult and expensive to get fiber-coupled. Since transducer coatings are

often polymers, they suffer from polymer-limited selectivity with a single fiber, necessitating multiple fibers for discrimination. However, while bundles may be easy to fabricate, launching into a bundle and selectively measuring the output from a bundle can be challenging.

D.3.7 Conducting Polymer

The transducer-based technology of a conducting polymer measures the conductivity of a bulk polymer/carbon mixture between two electrodes [76]. As analyte particles absorb into the polymer bulk, the bonds interfere with the transfer of electrons changing the resistivity of the sensor. A current is then applied to measure the resistance. Resistivity depends on the geometry of the device, and composition of the conducting polymer mixture. Thinner devices will yield higher sensitivities, and yet are more noisy. These devices often suffer from humidity and drift associated with temperature and polymer creep. These devices often yield detection limits in the 10's of ppm, though are capable of lower limits in well-controlled environments, and are built with precision fabrication.

Source: [79]

Ethanol LOD: 10 ppm

Integration time: 10 s

D.3.8 Optical Nose

For perspective we rehash the relevant features of the optical nose. As with the SAW, conducting polymer and other transducer based systems, an array of transducers is used for chemical to optical signal transduction. The dimensions and seize of the array are limited solely by the beam dimensions, power, and the photo-detector array. However, unlike many of these systems the transducer slide of the optical nose is easily replaceable and the interferometer will adapt to any phase differences between polymer arrays owing to surface feature variations. The optical nose is also

capable of accepting any transducer material as long as its transduction mechanism works without extra reagents and the material can be made sufficiently thin.

Regarding sensitivity, again for the workhorse of our polymers, Poly (n-vinyl-pyrrolidone) allows for the detection of around 2 ppm/$Hz^{1/2}$ of EtOH, This compares favorably to many of the other technologies when their sample/recovery times are taken into account assuming the use of a preconcentrator is excluded. The most prohibitive component of this technology is the photorefractive which sets requirements on temperature and base price.

Source: This work

Ethanol LOD: 2 ppm

Integration time: 5s

D.4 Summary

Primarily considering transducer based technologies, the optical nose holds its own in the competition. Systems such as SAW's and SPR, based on the literature presented, are the strongest competition since they exhibit superior sensitivities but with comparable to longer response times. The SAW has the disadvantage of it being difficult to replace the transducer elements since this requires the core be exchanged. However, IC fabrication technology may make the replacability of the transduction cores competitively cheap if a system were mass-produced in sufficient quantities.

SPR in many respects is nearly identical to the optical nose and therefore suffers many of the same disadvantageous. The photorefractive improves immunity to environmental drifts, though common path SPR likewise reduces environmental sensitivities. SPR is, however, primarily an aqueous technology used for biosensing and is rarely used as a vapor sensor. Many of the other sensor technologies, including the conducting polymer, conducting metal-oxide, and photo-ionization detection , are simply indiscriminate or insensitive compared to the optical nose.

Spectroscopy has been discussed little as it is essentially a class in its own, with many branching technologies, and inclusion would not benefit any insight into the performance of transducer based technologies. Spectroscopy often has the drawbacks of requiring expensive and stable laser

systems and must acquire an adequate spectrum or be tuned to look for very specific transition lines. Yet, given time spectroscopy is the most versatile as the limit in what can be detected is dictated by the database and the employed spectroscopic technique. Transducer based systems are primarily limited in what they can detect by both the specificity of the transducers available and the capacity of their arrayed processing. To compare the various technologies presented we include the table below.

	EtOH LOD (ppm)	**Response Time**	**Cons.**	**Ref**
SAW	5ppm* 0.7ppm**	75s* mins**	Core Replacement, Dust, batch-batch reproducibility.	[69]* [12]**
Chem-Luminescence	0.25ppm	Few min	Consumable Reagents	[77]
Spectroscopy	125ppb	30s	Database, Stable Laser, Specific spectral feature.	[78]
Optical Nose	2 ppm	5 s	Stabile Laser Wavelength Pressure BaTiO3	This work
CP	>10ppm	10's-100's s	Indisc, drift	[79]
MOX	0.1 ppm* 1-10ppm**	300 s	Indisc, Hot, drift	[70]* Figaro TGS 2602**
PID Phocheck	0.1 ppm	15 s	Indisc, drift	This work
SPR AlphaSniffer	<1 ppm	5 s	Drift, Angle dependant, primarily aqueous	[7]

Table D- 1: Sampling of performance parameters of various vapor olfaction systems.