University of Alberta

Tunable Diode Laser Trace Gas Detection with a Vertical Cavity Surface
Emitting Laser

by

Dragan Vujanic

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Electrical and Computer Engineering

EXAMINING COMMITTEE

John Tulip, Electrical and Computer Engineering

Wolfgang Jaeger, Chemistry Centre

Ying Tsui, Electrical and Computer Engineering

ABSTRACT

The nature of work conducted during the course of study towards a MSc degree focused on tunable diode laser absorption spectroscopy (TDLAS). This field involves the *in-situ* detection of gas constituents from low concentration samples. Specifically, I will focus on TDLAS systems utilizing practical optics, readymade electronics, and commercially available near infrared vertical cavity surface emitting lasers (VCSEL). In attempting to lower the minimum detectable concentrations of constituent gases, quantifying contributory noise sources is vital. Consequently, I seek to characterize principle noise sources of a prototypical TDLAS system in order to gain understanding of the limits that inhibit detection of trace gas concentrations. The noise sources which were focused on can be categorized as follows: source laser noise, optical noise, and detection noise. Through this work it was my goal to provide the means of achieving superior sensitivities.

ACKNOWLEDGEMENTS

Contents

List of Tables

List of Figures

List of Terms

VCSEL	vertical cavity surface emitting laser
BTJ-VCSEL	buried tunnel junction VCSEL
NIR	near infra-red
RIC	range integrated concentration
TDLAS	tunable diode laser absorption spectroscopy
WMS	wavelength modulation spectroscopy
WM	wavelength modulation
2f-WMS	second harmonic WMS
IM	intensity modulation
IM-FM modulation	intensity modulation w.r.t. frequency
ppm·m	parts per million meter
RIN	relative intensity noise
SNR	signal-to-noise ratio
OSNR	optical signal-to-noise ratio
DBR	distributed Bragg reflector
RAM	residual amplitude modulation
TEC	thermo-electric cooler
PD	photo-diode
NEP	noise equivalent power
FET	Field effect transistor
BJT	Bipolar junction transistor
LED	light emitting diode
LIA	lock-in amplifier
DPLL	digital phase-lock loop
ADC	analogue to digital conversion
CC	coherence collapse
FB	feedback

1. INTRODUCTION

Though there are many laser types, the semiconductor laser diode is a class of lasers that are electro-optic devices built using manufacturing principles borrowed directly from modern-electronics manufacturing techniques. The presence of noise is a natural phenomenon in all measured variables and because laser diodes are quantum devices they exhibit intrinsic noise levels which are easily noticed. Semiconductor lasers can be sub-categorized further into two classes; edge and surface emitters. VCSELs, vertical cavity surface emitting lasers, are surface emitters first developed to emit at optical frequencies with the development of long-wavelength single-mode emitters being more recent. These lasers are typically small, widely tunable, and power efficient. The feature sizes measure in the order of 1μm, emitting frequencies from the visible through the near infra-red (NIR) range with emitting powers around one milliWatt in magnitude. These features make the long wavelength VCSEL ideally suited to the application of in-situ trace gas detection. This dissertation, therefore, explores the characteristics exhibited by this laser source as applied in a practical, small form-factor, portable, trace gas detector.

1.1 ABSORPTION PRINCIPLE

The Beer-Lambert Law is an appropriate formulation of optical absorption by a gas given monochromatic light. The governance of absorption is predicated on the semi-classical light-matter interaction mechanism insofar as only selective energy transitions, corresponding to resonant absorptions, are permitted. As such the transmission, T, of light through an absorptive material according to the Beer-Lambert Law is expressed as:

$$T(v) = I/I_O = e^{-\int_L \alpha(v)ds}$$

Eq. 1

where,

I (mW/cm) is the detected light intensity after traversal of subject material,

L (cm) is the optical path length through subject material,

I_0 (mW/cm) the initial source intensity, that is, prior to interaction with material,

α, is referred to as the spectral absorption coefficient and is unitless, while

v is spectral frequency commonly stated in improper frequency, having units of cm^{-1}, in spectroscopy.

The absorption coefficient is in fact given by the product of total pressure (atm), mole fraction of a particular absorbing species g, line strength (cm^{-2} atm^{-1}), and line shape function (cm) as,

$$\alpha(v) = PX_g S_{g,i} \Phi_{g,i},$$

Eq. 2

respectively. The subscript i is used to denote a particular transition of a single species whereby multi-species and multi-transition measurements are indeed defined simply by the superposition of individual absorption coefficients. Line-strength and line-shape, in fact, are uniquely dependant on the parameters:

1

frequency, pressure, and temperature such that $S_{g,i} \equiv S_{g,i}(T)$, while $\Phi_{g,i} \equiv \Phi_{g,i}(T,P,v)$. This does suggest that quantitative measurements of concentration require *a priori* knowledge of system temperature and pressure. The line-shape function is indicative of the line broadening mechanism that may predominate system conditions. At low pressures, below 30 Torr, natural broadening due to thermal motion leads to Doppler broadening and so the line-shape function may be described by a Gaussian profile in this regime. At high pressures, above 300 Torr, collisional broadening elicits a Lorentzian line-shape. Between these pressures both broadening mechanisms are significant contributors such that the line shape is given by a Voigt profile, the convolution of the two. Regardless of the line-shape profile though, it is always normalized over the entire frequency spectrum such that $\int \Phi_i dv = 1$.

Though strictly the Beer-Lambert law requires the path integration over gas samples, in scenarios where it may be assumed that limited variance of pressure, temperature, and concentration are exhibited, the integration may be simplified to a multiplication over length *L*. This approximation is therefore valid in analysis where the target gas is sampled within an equilibrium-established sampling cell or over open path distances where this holds. Absorbance, defined as

$$A(v) = PLX_{CH_4}S_{CH_4,i}\Phi_{CH_4,i},$$

Eq. 3

therefore, is linearly dependant on pressure, sample path length, and concentration, inherent in partial pressure PX_{CH_4}. The formulation of absorbance, in fact, now defines what is meant by the term *detection limit*, when citing trace gas detection. Detection limit alludes to the range-integrated concentration, $RIC = LX_{CH_4}$, and will be stated in units of parts per million meter (ppm·m) for this body of work. Often throughout this thesis I will refer to the RIC, as the gas concentration, though strictly, this is a misnomer. Therefore, in the context of the Beer-Lambert Law methane concentration is given as

$$LX_{CH_4} = \frac{\int_v - \ln T(v)dv}{PS_{CH_4,i}}$$

Eq. 4

and it hints at an experimental philosophy by which to conduct gas absorption experiments.

1.1.1 Tunable Diode Laser Absorption Spectroscopy

In reality a weakening of the principle light intensity may occur by non-absorptive phenomena too, scattering for example. Though it is not possible to determine to which extent solely absorption diminishes light intensity, by tuning over an absorption feature we can ascribe the associated intensity attenuation to resonant absorption and so correlate it to concentration quantities. Consequentially, absorption measurements using tunable diode laser absorption spectroscopy, TDLAS, are based on the Beer-Lambert Law. By tuning through optical

frequencies centered about an absorption feature one may quantify the numerator defined in equation 4. This coupled, with *a priori* knowledge of pressure and temperature, and hence line-strength S, facilitates the means of quantifying gas concentration. Basing absorption solely on intensity whitewashes the effect upon the signal due to dispersion; however, employing operational parameters that place us within what is know as the wavelength modulation spectroscopy WMS regime it will be shown that this slight is of marginal consequence.

Wavelength modulation absorption spectroscopy (WMS) takes advantage of VCSEL's capacitance for fast tunability and wide tuning range. This allows for swift, dry, *in situ* gas detection. In WMS, fast sinsoidally modulated light output is slowly tuned over an absorption line. The major significance in applying this technique to trace gas detection is in shifting the detection signal to higher frequencies thus minimizing the deleterious effects of $1/f$ noise at low frequencies. The operational requirements outlined by such a detection scheme necessitate the study of VCSELs toward this task.

1.2 Wavelength Modulation Spectroscopy

In the interest of continuity, laser noise is overviewed in the next section, **Introduction to Laser Noise.** Temporarily therefore, it is enough to say that noise, particularly $1/f$ noise, degrades SNR levels significantly at baseband frequencies. Due to this noise source then, the mere scan through frequency, appropriated by the Beer-Lambert Law via equation 4, inherently extols a large degradation upon signal quality. Therefore, modulation is introduced in order to exact a shift in signal to higher frequencies. This necessitates the use of a laser that can operate so, and as previously outlined, the VCSEL is well suited toward this task. When conducting spectroscopic trace gas detection using modulated semiconductor lasers, two regimes of operation are defined: frequency modulation, FM, and wavelength modulation, WM. Wavelength modulation is characterized by modulation frequencies much smaller than the half-widths, Δv_{line}, of the absorption features ($f << \Delta v_{line}$), while the converse is true for FM, $f >> \Delta v_{line}$ [1], [5]. Absorption lines at atmospheric pressures and temperatures are broadened to Lorentzian profiles on the order of gigahertz. Though optically this represents a narrow range for nominally near infra-red, NIR, diode lasers, electronically this high boundary makes wavelength modulation much more appealing.

Poignantly, the previously introduced Beer-Lamberts Law is based on an intensity formalization, neglecting dispersion. In considering absorption as an electric-field interaction, however, modulation schemes at low frequencies implies many side-(closely-packed)-modes which must be summed over; at atmospheric conditions this is on the order of 10^5-10^6 [1]. Below, in Figure 1, the signal component due to either absorption or dispersion is shown.
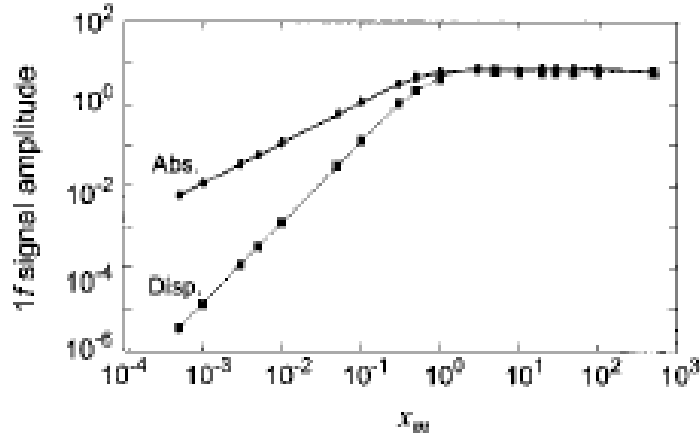
Figure 1. Respective absorption and dispersion signal strengths at the fundamental modulation frequency as dependent of the normalized frequency x=f/Δv_{line} [1].

This data supports the notion that within the WM regime, the mechanism for the light signal loss to be near-completely caused by absorption. Particularly is this true within this body of work where modulations are in the tens of kilohertz, exceedingly small relative to the several-gigahertz absorption line widths. Consequently, only absorption need be considered, absolving the need for a field model and validating a formulation based solely on absorption. Qualitatively therefore, wavelength modulation spectroscopy, WMS, is identified with the detection scheme where sinusoidally modulated light is tuned across an absorption feature of choice.

Subsequently, the interaction of the modulated source and the absorption feature produces a signal at harmonics of the fundamental sinusoidal frequency. The harmonics themselves are extractible by lock-in detection, are indicative of absorption strength, and fundamentally, offer a great bandwidth reduction, paramount for limiting noise power. This allows for greater sensitivity in the quantification of gas concentration. In this regard, the second harmonic, 2*f*, possesses several favourable properties that make it the ideal choice. These include sufficiently high frequency operation, and large signal strength. Most importantly though, the second harmonic is free of a constant and linear background signal, a degrading factor for detection sensitivity. As such, this work is incumbent on noise limitations in 2*f*-WMS detection for portable methane gas detectors.

1.2.1 Wavelength Modulation Spectroscopy Formulation

As noted, wavelength modulation spectroscopy is the technique of trace gas detection by use of a known absorption feature of the target gas which is scanned over slowly by diode laser emission. This is most commonly done in a sawtooth or triangular fashion while concurrently a much faster small-signal modulation is applied to the laser emission. Selective harmonic detection follows the interaction. With the previously validated linear model, injection current

modulation and the resulting intensity and laser frequency modulations are, respectively,

$$i(t) = i_0 + \Delta i \cos(\omega t),$$

$$I(t) = I_0 + \Delta I \cos(\omega t),$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{Eq.}$$

$$v(t) = v_0 - \Delta v \cos(\omega t + \Psi).$$

5(a,b,c)

The naught subscripted parameters signify the nominal values about which modulation is applied. Furthermore, the delta coefficients denote the significance or amplitude of modulation, while $\omega = 2\pi f$ denotes the angular frequency of modulation. Additionally, it will be demonstrated (see Figure 10) that there is a negative slope in frequency versus injection current and as such necessitates the sign reversal in relation to current modulation, here regarded as the basis of modulation. Regarding the optical frequency modulation, Ψ is the quantification of an intensity modulation versus frequency modulation, IM-FM, phase displacement. Evidence of this phenomenon is presented later, in **2.2 Vertical Cavity Surface Emitting Laser Behavior**. It is at this point prudent to state the variability of Ψ depending on the electronic frequency of modulation. Though this is true of the delta coefficients too, an exceedingly small dependence exists for these parameters at low frequencies.

Relating intensity to frequency does however necessitate the introduction of a slowly changing proportionality factor dependent on the electronic frequency of modulation. With the above outlined parameters, and a purely Lorentzian line-shape assumption, a harmonically decomposed expression for the detected signal may be given as,

$$s_{IM-FM}(v) = I_0 \cdot \left[\sum_{n=0}^{\infty} s_{np}(v)\cos(n\omega t) - \sum_{n=0}^{\infty} s_{nq}(v)\sin(n\omega t) \right]$$

$$\text{Eq. 6}$$

with,

- n the corresponding integer harmonic,
- and s the nth harmonic of in-phase p or quadrature q terms.

A full development may be found in numerous sources [1], [3], [4], [5]. Note, because we are concerned only with second harmonic detection only its terms will be focused on here. Furthermore, lock-in detection allows for the detection of not only signals in phase or quadrature but any intermediate phase Φ as well, yielding

$$s_{2,\Phi}(v) = I_0 \cdot \left[s_{2p}(v)\cos\Phi_2 + s_{2q}(v)\sin\Phi_2 \right]$$

$$\text{Eq. 7}$$

Accordingly then, the 2*f* in-phase signal profile resembles
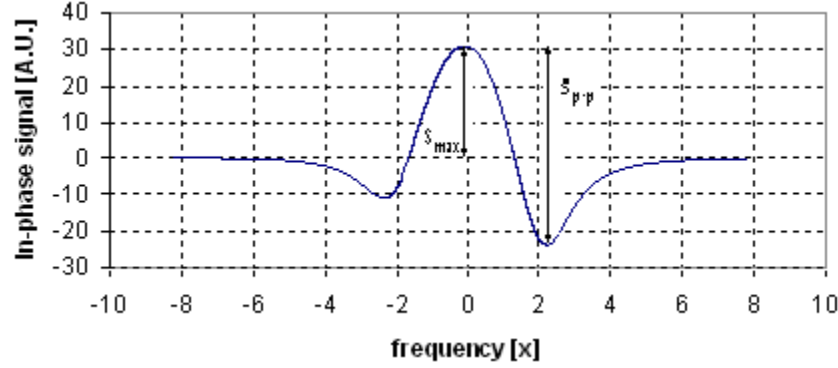
Figure 2. A typical 2f signal. Here m denotes the modulation severity, whereas optical frequency is normalized and is given by x, modeled with Matlab.

Figure 2, above, shows the profile of the 2*f* signal. Note, this and subsequent 2*f* modeling executed in Matlab, for which the script code is found in appendix B. Here several issues may be noted. First, the signal modeling incumbent upon an intensity modulation results in an asymmetry not reported in first generation models of 2*f*-WMS. The signal here also is not to be mistaken as being proportional solely to the second derivative of the absorption feature. It is composed also of higher and lower derivative fractions. Second, the resulting profile characteristics are highly dependent upon the modulation index *m*. Investigating this trend, the s_{max} and s_{p-p} quantities are shown to be highly dependant on this variable in Figure 3 below.
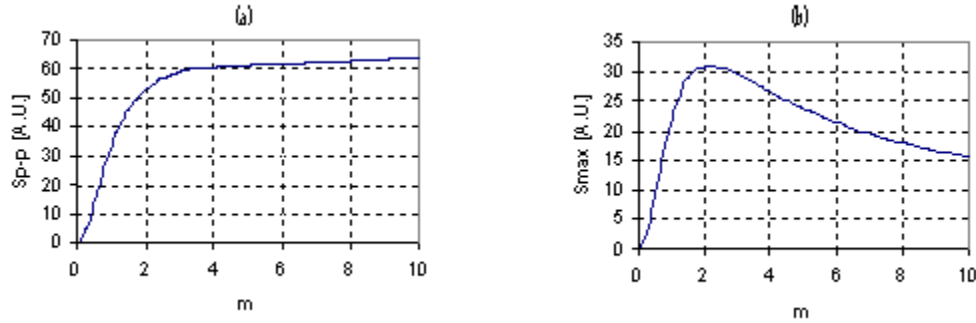
Figure 3. 2f signal sp-p (a), and smax (b), strength dependence on the modulation index m. Modeled with Matlab

Though s_{p-p} increases indefinitely with increasing modulation indices, and rather slowly after *m*~3, s_{max}~2 peaks, indicating an optimal operating-point for detection signal strength given *m*. Therefore it is this value, *m*=2.20, that is applied. Experimentally this was in fact exacted by noting the peak signal strength as the modulation index was toggled. The index was set at this peak without needing to quantify it specifically, confident in the notion provided by modeling.

6

From equation 7 it is also quite evident that detection phase is of paramount importance and warrants investigation when detection limits of trace gases via 2*f*-WMS are considered.
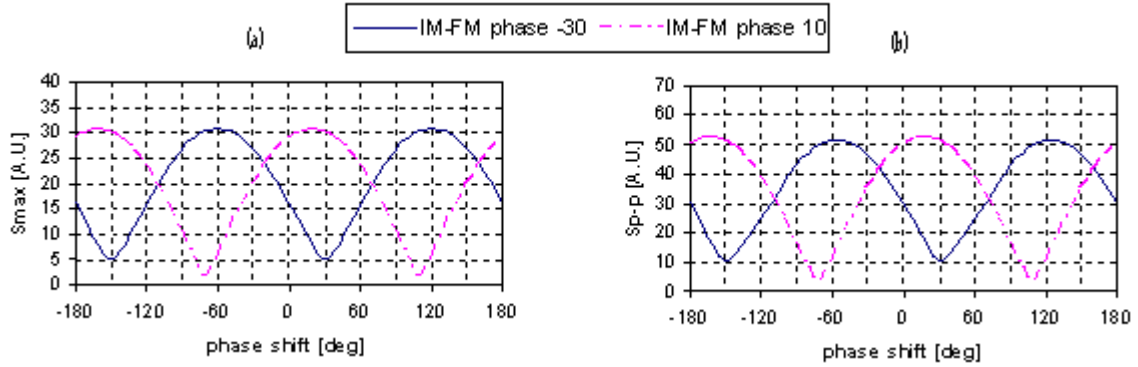
Figure 4. Signals' strength dependence on phase detection. Two differing IM-FM phase offsets were applied to each, smax (a), and sp-p (b), as detection phase is varied. Duly noted is the periodic nature, and the relation to IM-FM phase. Modeled with Matlab

Figure 4 above shows the strong dependence on detection phase, indeed for both s_{p-p} and s_{max}. The signal amplitude and range are dependent on detection in lockstep with eachother for given phases. What this modeling does reveal is the formulaic limits of these strength parameters when relating IM-FM phase to detection phase as

$$\Phi_{2,max} = 2\Psi + k \cdot 180, \qquad \text{and} \qquad \Phi_{2,min} = 2\Psi + (2k+1) \cdot 90.$$

Eq. 8

In practical experimentation this therefore quantifies the necessary lock-in amplifier phase and even suggests an alternative means of quantifying the IM-FM phase shift. Furthermore it attracts attention as a practical opportunity or oppositely a point of contention.

1.3 General Laser Operation

Diode lasers by nature are quantum devices and as such, their performance is inherently noisy when examined closely. This fact was immediately investigated by the pioneers of laser technology (eg. Schawlow and Townes) and has been extensively investigated ever since. The culmination of this work has produced a large body of knowledge on the subject and will be overviewed here. Significantly, this is relevant as a preamble and contrast to feedback induced laser noise, arguably, after $1/f$ noise, the second most significant noise mechanism responsible for the degradation of SNR. The main noise source in laser diodes originates in spontaneous emission of radiation, yielding fluctuations in the emitted laser light intensity along with variances in the emitted optical frequencies. In describing these phenomena we identify the intensity fluctuations with the so-called relative intensity noise (RIN). The frequency fluctuations are responsible for the laser emission "linewidth". The mathematical treatment of

7

these laser noise mechanisms follows and is built on laser diode theory expounded in most texts on the subject [8], [9], [10].

1.3.1 Noise Definitions and Properties

The main source of noise in laser diodes is due to spontaneous emission. This results in a fluctuation of the emitted optical intensity and of the emission frequency. In this field the intensity fluctuations are characterized by the relative intensity noise (RIN). It will be demonstrated how this is analogous to the familiar SNR quantity which pervades the telecommunications industry. Frequency fluctuations are responsible for the finite spectral linewidth of laser diodes. In order to discuss noise it is prudent to review some common definitions and so now follows the mathematical characterization of noise signals which will be used throughout this report. To start, noise here will be defined in the context of power signals and so

$$P(t) = \langle P \rangle + \delta P(t) \qquad where \quad \langle P \rangle = \frac{\lim}{t' \to \infty} \frac{1}{t'} \int_{t_o}^{t_o+t'} P(t)dt$$

Eq. 9

Evidently <P> is the average power [11]. In defining noise there are two traits that combine to give a complete picture;

- the temporal nature and,
- the statistical nature.

Because the probability distribution of noise can be difficult to determine, it is most commonly defined by the autocorrelation relation

$$\rho_P(I) = \langle \delta P(t) \delta P(t - \tau) \rangle$$

Eq. 10

This time domain description of noise, has a frequency domain representation via the power spectral density, $W_P(w_m)$, too;

$$W_P(\omega_m) = \langle |\Delta P(\omega_m)|^2 \rangle = \int_{-\infty}^{\infty} \rho_P(\tau) e^{j\omega_m \tau} d\tau$$

Eq. 11

Here $\Delta P(\omega_m)$ represent the Fourier transform of $\delta P(t)$. Worth noting is that these relations are limited insofar as they apply to stationary and ergodic processes only, the noises dealt with throughout this report being as such. Proceeding onward, the mean square of the noise signal $\delta P(t)$ may be expressed in terms of the spectral density as

$$\langle \delta P^2(t) \rangle = \rho_P(0) = \int_{-\infty}^{\infty} W_P(\omega_m) df = \int_{-\infty}^{\infty} \langle |\Delta P(\omega_m)|^2 \rangle df$$

Eq. 12

In practice we are limited by finite bandwidths of real systems and so by considering a transmission-band with unity gain and zero gain stop-band one obtains for the mean-square value [12]

$$\langle \delta P^2 |_{\Delta f} \rangle = 2\Delta f \langle |\Delta P(\omega_m)|^2 \rangle = 2\Delta f W_P(\omega_m)$$

Eq. 13

Summarily, it is worth highlighting that the above expressions for autocorrelation and spectral density form a Fourier transform pair; autocorrelation ⇔ power spectral density. The second point above is the description of noise via statistics. Toward this end a probability density function, $p(\delta P)$, is defined and is normalized such that $\int_{-\infty}^{\infty} p(\delta P) d(\delta P) = 1$.

1.3.2 Definition and Measurement of Relative Intensity Noise

In communication systems signal strengths degraded by noise signals are often sought and quantified by the signal-to-noise, SNR, ratio. It is an important parameter telling of the quality of the transmission channel and the minimum power of signals that can be detected. Implicitly it is focused on the signal, as though if to say; "what is the business end of the signal in lieu of this background – this noise?" When one is concerned with studying the noise specifically, though, we choose to highlight it and we may ask, rather; "what of the noise given such a signal?" It is a subtle distinction but it serves to place the noise, as an entity in its own rights, into the foreground of analysis, so to speak. For this purpose then we define the relative intensity noise RIN parameter:

$$RIN = \frac{\langle \delta P^2 \rangle}{\langle P \rangle^2}$$

Eq. 14

For example, if a power signal was of a sinusoidal nature of amplitude $\hat{P}$, average power <P>, m represented their ratio, and a superimposed noise δP(t), then the relation between SNR and RIN would be straight forward:

$$SNR = \frac{\hat{P}^2/2}{\langle \delta P^2 \rangle} = \frac{m^2}{2} \frac{\langle P \rangle^2}{\langle \delta P^2 \rangle} = \frac{m^2}{2RIN}$$

It is the ratio of noise power variance to signal power squared. If noise is measured with a noise bandwidth Δf, then

$$RIN = \frac{2\Delta f \langle |\Delta P(\omega_m)|^2 \rangle}{\langle P \rangle^2} = \frac{2\Delta f W_P(\omega_m)}{\langle P \rangle^2}$$

Eq. 15

Even for an ideal receiver we must consider the quantum noise at the photodiode yielding a lower RIN_Q for the detectable relative intensity noise:

$$RIN = \frac{2\Delta f \langle |\Delta I(\omega_m)|^2 \rangle}{\langle I \rangle^2} = \frac{2\Delta f e}{\langle I \rangle^2} = \frac{3.2 \cdot 10^{-16} mA}{\langle I \rangle^2} \frac{\Delta f}{Hz}$$

$\langle I \rangle$ is the mean photocurrent and the spectral density due to shot noise (or quantum noise) is given by $<|\Delta I(\omega_m)|^2>=e<I>$ with elementary charge e. Practically then, a RIN of a laser diode may thus be detected/measured as long as it is considerably larger than the detection limit RIN_Q [12].

9

1.3.3 Nominal Laser Diode Noise

When describing laser emission a common method for representing the electric field component of light is using complex notation. The actual field is then given by taking the real component,

$$E_{las} = \text{Re}\{E(t)\exp(j\omega_{th}t)\}$$

Eq. 16

with $\omega_{th} = 2\pi\nu_{th}$ the circular reference frequency of oscillation. It is also convenient to normalize the complex field amplitude $E(t)$ such that

$$E(t) = \sqrt{S(t)}e^{j\phi(t)} \quad or \quad S(t) = |E(t)|^2,$$

Eq. 17

where the phase $\phi(t)$ simply corresponds to the phase of the slowly changing amplitude and $S(t)$ is the photon count, a direct correlation of power. With such notation the laser diode field equation is then written as

$$\frac{dE(t)}{dt} = [j(\omega - \omega_{th}) + (1/2\tau_{ph})(G-1)]E(t) + E_{sp}(t)$$

Eq. 18

where,

- $E(t)$ is the slowly varying electric field component of radiation,
- $\omega \approx \omega_{th}$ the circular resonance frequency,
- τ_{ph} is the photon cavity lifetime,
- G is normalized gain,
- and $E_{sp}(t)$ denotes a complex field amplitude due to spontaneous emission [14].

The greatest noise source in laser diodes being due to spontaneous emission we include it in the field equation of diode lasers. Spontaneous emission noise extends over a large wavelength range and so may be considered "white", furthermore, it exhibits a Gaussian probability density distribution. From equation 18 above we can obtain

$$\frac{dS}{dt} = 2 \cdot \text{Re}(E*(t)\,dE/dt) = \frac{S}{\tau_{ph}}(G-1) + 2 \cdot \text{Re}(E_{sp}(t)E*(t)) \qquad \text{Eq. 19(a)}$$

$$\frac{d\phi}{dt} = \text{Im}(E*(t)\,dE/dt)/S(t) = (\omega - \omega_{th}) + \frac{1}{S(t)}\text{Im}(E_{sp}(t)E*(t)) \qquad \text{Eq. 19(b)}$$

Spontaneous emission is amplified within the laser cavity and so is not independent of $E(t)$. Petermann in [12] has shown that $<Esp(t)E*(t)>=R/2$ where R is a constant to be determined. This enables a rewriting of equation 19(a) in the form

$$\frac{dS}{dt} = \frac{S}{\tau_{ph}}(G-1) + R + F_S(t)$$

Eq. 20

with F_S representing a noise source of zero mean, commonly referred to as Langevin noise. Without getting sidetracked by going through the drudgery of its derivation [12], the noise inclusive equation for phase is simply given here as

$$\frac{d\phi}{dt} = \frac{1}{2}\alpha \upsilon_g \frac{\partial g}{\partial n}(n - n_{th}) + F_\phi(t)$$

Eq. 21

Here g is gain and α is the common linewidth expansion factor. Noteworthy is that F_S and F are independent Gaussian noise processes with zero mean. To surmise therefore, equations 20 and 21 introduce intensity and phase noise, respectively as F_S and F .

1.3.4 Intensity Noise

Many lasers emit not only in a single dominant mode but also several side modes too. Each of these modes emit with their own respective intensity noise. The VCSEL is no different but as the principle lasing mode is 30dB stronger than adjacent modes, the VCSEL can be treated as nominally single-mod for this investigation.

1.3.4.1 Single-mode Intensity Noise Analysis

In calculating the RIN of a single-mode laser we invoke equation 20 and the carrier density, *n*, equation,

$$\frac{dn}{dt} = \frac{I - I_{th}}{eV} - \frac{1}{\tau_e}(n - n_{th}) - GS/(V\tau_{ph}) + F_n(t)$$

Eq. 22

The Langevin noise term $F_n(t)$ is due to the shot noise character of the injection current within the cavity volume V. In solving these coupled equations we simplify by linearizing around stationary values *<S>* and *<n>*. This yields a solution for the RIN [12] as

$$RIN = \frac{2\Delta f\left\langle |\Delta S(\omega_m)|^2 \right\rangle}{\left\langle S \right\rangle^2} = |H(j\omega_m)|^2 \frac{(1/\tau_e + \omega_r^2 \tau_{ph})^2 + \omega_m^2}{\tau_{ph}\omega_r^4 \left\langle S \right\rangle} 4K_{tot}n_{sp}\Delta f$$

Eq. 23

where,

- ω_m is circular modulation frequency,
- *H(jω_m)* represents the small signal transfer function of current modulation to power output and has the same form as an RLC circuit; a second order transform,
- ω_r the relaxation resonance frequency,
- K_{tot} is the enhancement factor of spontaneous emission accounting for mirror reflectivities and wave guiding,
- n_{sp} is the inversion factor, the spontaneous to stimulated emission rate ratio.

RINs on the order of magnitude of 10^{-15} Δf/Hz have been measured, in agreement with equation 23 above, considering prototypical laser diode parameters.

Equation 23 is valuable in the trends in relative intensity noise that arise from its form above. It yields three major inferences that may be taken. First, at low frequencies the RIN is, thus, relatively low and so $1/f$-noise tends to dominate, more on this later. Second, ω_r^2 is proportional to photon number $<S>$ and thus makes RIN in equation 23 proportional to $1/<S>^3$. This implies a rapid decrease with emitted power. Finally, the RIN increases with increasing ω_m yielding the maximum RIN around the relaxation frequency ω_r.

1.3.5 Additional Intensity Noises

1.3.5.1 Mode Hopping Noise

This is the noise mechanism produced by the jumping between two neighbouring emission modes. The buried tunnel junction VCSEL, BTJ-VCSEL, operates with side mode emission 30dB below that of the main mode. Mode hopping noise itself is the noise perceived in the intensity output of a laser as two or more modes alternate as the single dominant lasing mode. Though this does not strictly translate to nominal VCSEL operation, we will encounter it with the introduction of feedback.

1.3.5.2 Partition Noise

Mode partition noise is rooted in the fact that even nominally single-mode laser diodes have small strength side modes. Temporal variations in the power distribution amongst these modes (while total power remains unchanged) gives rise to this noise source whose strength is comparable to above stated intensity noises. Because in fact the sum of multiple independent spontaneous emission noise sources can yield smaller net noise than any of the independents filtering laser light to eliminate mode partition noise may indeed increase intensity noise.

1.3.5.3 1/f Noise

This noise also known as "excess noise" is observed in all electronic devices that possess p-n junctions. The name is indicative of the spectral density, which possesses a $1/f$-like profile for frequencies lower than about 100 kHz. For the case of laser diodes the spectral density of $1/f$ noise has not shown to be injection current dependent above threshold. In equation 15 we see that spectral noise density is proportional to $1/P^2$. From this we infer that operating the laser diode at injection currents well above threshold will have a beneficial effect, in fact an inverse square benefit, from the perspective of noise relative to signal strength. The low frequency noise is precisely the reason for wanting to shift trace gas detection schemes to higher frequencies.

1.3.6 Frequency Noise

As noted previously the laser frequency noise is the mechanism responsible for the finite line width of laser emission. Again, to not get too bogged down with tedious derivations only the results of frequency noise will be presented here and the conclusions that they draw will be highlighted, detailed derivations are carried out elsewhere [12]. Frequency noise spectrum determines laser line shape. By

assuming that the contribution of intensity noise to line broadening is low for injection currents well above threshold we are justified in considering the frequency noise the only contributor to linewidth.

In the case in which frequency noise is assumed to be "white" the emission spectral line shape possesses the Lorentzian form,

$$\frac{\Delta\nu}{2\pi[(\nu_0 - \nu)^2 + (\Delta\nu/2)^2]}$$

while for 1/f-frequency noise the laser diode spectral line shape takes on a Gaussian form;

$$\left(\frac{4\ln 2}{\pi}\right)^2 \frac{1}{\Delta\nu} \exp\left[-4\ln 2\left(\frac{\nu - \nu_0}{\Delta\nu}\right)^2\right]$$

where in both cases $\Delta\nu$ represents the laser line "linewidth", being the spectral width at half the value of the maximum spectral component ν_0. When a combination of white and 1/f noise is purported then the resulting laser emission lineshape will in fact be a convolution of the above stated Lorentzian and Gaussian profiles, resulting in the Voigt lineshape, as it is commonly referred to.

At this point noise in laser diodes has been overviewed substantially enough that we may investigate laser diode noise properties subject to optical feedback and have a sufficient basis of comparison.

2. VERTICAL CAVITY SURFACE EMITTING LASER GENERALS

Because of the surface emitting nature of VCSELs mirrors and contacts must be fabricated in such a way that epitaxial growth of the active region on top of other laser features, as in a stack, is possible. Another contrast to edge emitters whose current flow is across the cavity so that the whole of the cavity is subjected to population inversion and gain, the injected current flow in VCSELs is down the cavity so that the active region, where gain occurs, is but a small fraction of the entire cavity length. This necessitates highly reflective mirrors to offset the short active medium length relative to the cavity. Therefore, cavity mirrors are distributed Bragg reflector (DBR) structures, multiple layers of alternating index, which collectively are highly reflective to a desired wavelength. They are therefore designed to reflect light of the highest gain cavity mode. The reflectivity (R) of the DBR is given by

$$R = \left[\frac{n_0 (n_2)^{2N} - n_S (n_1)^{2N}}{n_0 (n_2)^{2N} + n_S (n_1)^{2N}} \right]^2$$

Eq. 24

Where n_0, n_1, n_2, and n_S are the respective refractive indices of the surrounding medium, the two alternating materials, and the substrate; and N is the number of repeated pairs of low/high refractive index material. Using equation 24, Figure 5 below shows the necessity of using a least 30 alternating pairs for the DBR structure to attain reflectivities of at least 0.95 regardless of substrate index.
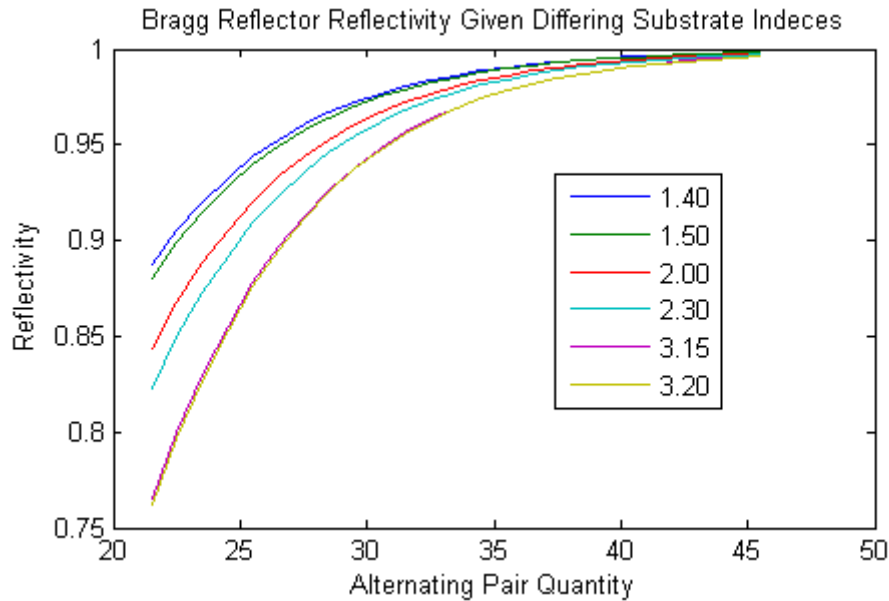
Figure 5. Reflectivity dependence of VCSELs on DBR pair quantity. Modeled with Matlab

Note, increasing the number of pairs in a DBR increases the mirror reflectivity and increasing the refractive index contrast between the materials in the Bragg pairs increases both the reflectivity and the bandwidth. Also, the true index of the substrate is considerably more complex and varies spatially, see Figure 6 (below), however, the effect can be correlated to that shown in Figure 5.

When investigating VCSEL cavities it is imperative not to forget that Fabry-Perot modes are standing waves. Short cavities, on the order of 1um, contain only a few modes and anti-modes, necessitating active medium positioning specifically at an anti-node. Placement of the active medium at a node reduces gain to near-zero until thermal effects or carrier densities possibly shift the wave distribution accordingly. Indeed even defining the cavity length of VCSELs is therefore difficult as reflection does not occur at a single hard boundary as in edge emitters but rather by multiple spatially-distributed partial reflections. Modeling the DBR structures as metallic mirrors according to Babic and Corzine [16], effective mirrors can be placed at a distance L_{eff} as seen in Figure 6 below such that phase consistency is maintained between the two models.
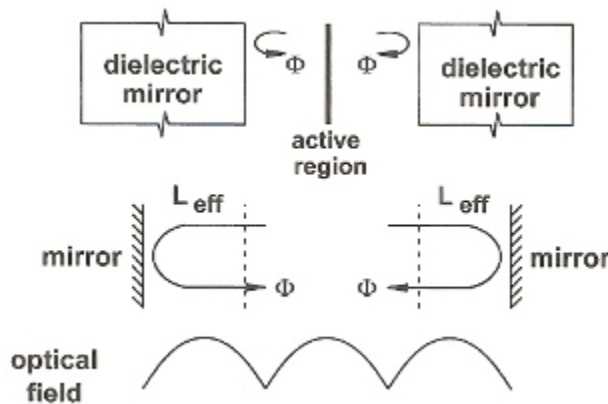
Figure 6. Conceptual VCSEL model substituting DBRs with planar mirrors.

This is an important simplification to be used in chapter 6 when laser feedback is addressed. Note the placement of the active mode at an anti-node of the standing wave pattern. Modeling work has been carried out by others on the effect of varying quantum well quantities in which it has been concluded that the optimum number of quantum wells is three when operating at 850nm; any more than this and the standing wave enhancement is reduced [16]. At longer wavelengths the quantity of material encompassed nearer an anti-node is larger so that in fact more quantum well structures are appropriately incorporated. A negative consequence in increasing the wavelength of operation is that additional considerations for the mirrors further complicate the DBR structure necessitating up to possibly 50 pairs of alternating index. Total mirror thickness increases from 2-3um for 850nm light to 11-12um for 1.5um light so that a 3-4 times increase in resistance requires greater thermal conductivity and potentially represents a large optical loss too.

2.1 BTJ-VCSEL

Long wavelength VCSELs are realized by incorporating a Buried Tunnel Junction, BTJ, shown in Figure 7, into the laser structure.
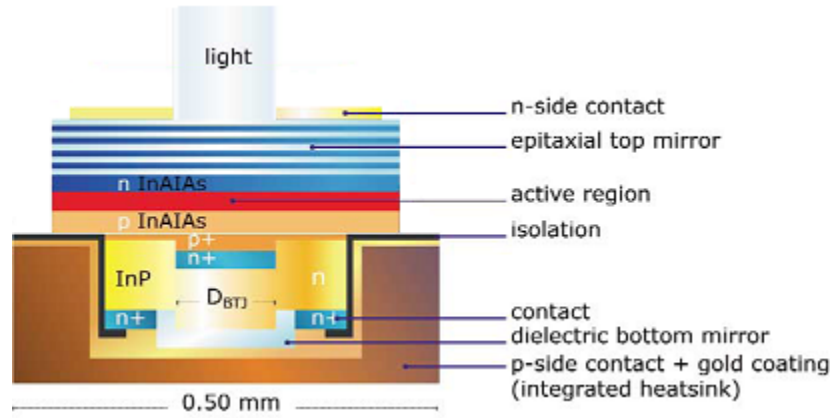
Figure 7. BTJ-VCSEL cross-sectional view.

The device consists of an epitaxial top mirror and an active region using the material system AlGaInAs grown on InP. For the top mirror, a layer stack of lattice matched AlInAs and (Al)GaInAs is used as low-index and high-index material, respectively. In the active region, compressively strained quantum wells are separated by weakly tensile strained barriers [19]. This first epitaxial step is finished by growing the heavily doped ($\sim 10^{20}$ cm^{-3}) tunnel junction layers. After etching the perimeter of n+ to create the tunnel junction between 2-20 μm, it is overgrown with n-doped InP in a second epitaxial step [19]. A dielectric mirror completes the laser structure. While the InP substrate on the top of the upside down mounted structure is completely removed by etching, an electroplated gold layer at the bottom serves as a heatsink. Optical emission occurs through an orifice that serves as the n-side contact. The BTJ design addresses excess heating and heat sinking challenges noted above by achieving lower series resistance and effective current confinement. Strong index guiding associated with the BTJ also maximizes optical power output and stabilizes single mode emission with side mode suppression in excess of 30dB, while wide tuning of output frequency is facilitated by the high current density present in the active region. Since it is the long wavelength BTJ-VCSEL that was used throughout the work conducted in this thesis it will subsequently be referred to only as the VCSEL.

2.2 VCSEL Behavior

For the VCSELs used throughout experimental work conducted for this thesis tuning ranges varied for individual diodes, the 1577nm VCSEL for example possessing a continuous tuning range -38.62cm^{-1}(9.6nm). It is worth clarifying that in these contexts tuning rate implies the frequency tunabiliy of VCSEL diodes with reference to driving current or temperature and not time. All references to time rates will be explicitly stated. Temperature tuning rates for all VCSELs varied between individual diodes with rates varying between -0.4 and -0.5 cm^{-1}/°C. Current tuning rates themselves were not constant across individual

16

diodes' injection currents and warranted a detailed examination. This correlates to principal laser operation and so is presented in tandem with intensity and frequency modulation phenomenon as well.

It is well understood that injection current modulation yields a correlated intensity output modulation, due the linear increase in gain with injection current. Prototypically this is embodied in a power-current plot such as that given in Figure 8 below.
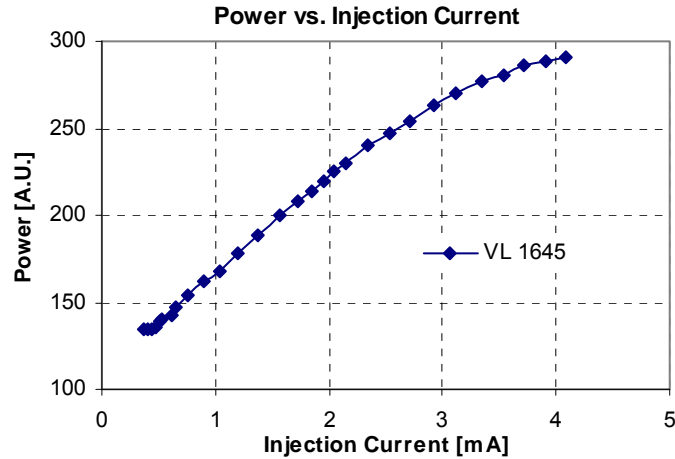
Figure 8. Characteristic power versus current profile for a 1645 nm VCSEL. Laboratory measured data points.

When modulating a semiconductor laser, through injection current or temperature modulation, a combined-effect intensity modulation and frequency modulation IM-FM occurs, generally not concurrently though. Instead, a relative phase shift between the two modulations exists that is indicative of laser structure and dependent on frequency, for the VCSEL this IM-FM phase shift is plotted in Figure 9 below.
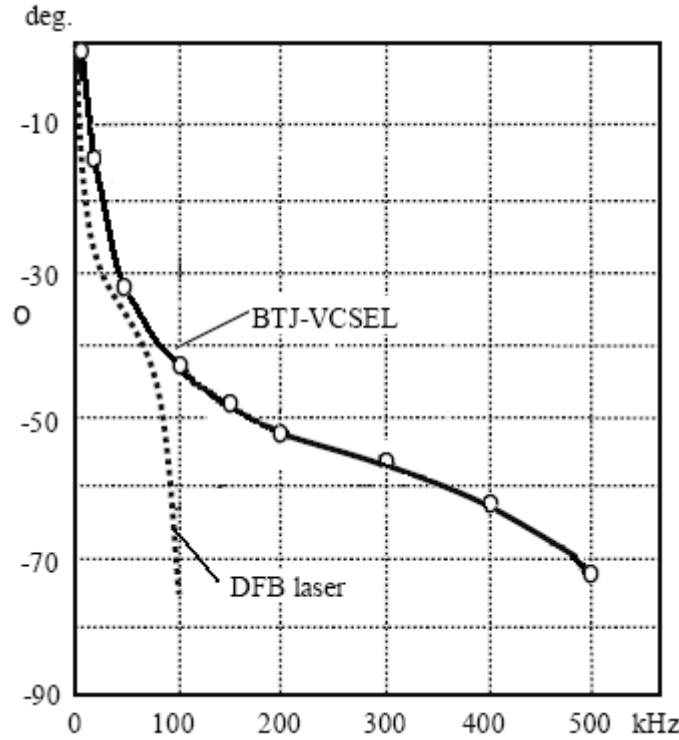
Figure 9. Comparison of VCSEL and DFB laser IM-FM phase shifts as functions of frequency. Laboratory measured data points.

The IM-FM phase shift acts to distort true absorption signals so that knowledge of this parameter is paramount in accurate concentration quantification using modulation detection schemes.

In signal detection by way of frequency modulation the intensity modulation (IM) of the signal is an undesired consequence of said modulation. As such, in literature the IM is often referred to as residual amplitude modulation RAM. Practically though, RAM is the result of non-linear frequency modulation with respect to injection current, manifested as IM due to the systematic inaccurate modeling of frequency as linearly dependent on injection current. Figure 10 shows this very effect by the linear relation exhibited between injection current and VCSEL tuning rates, inherently implying a quadratic relation between injection current and optical frequency.
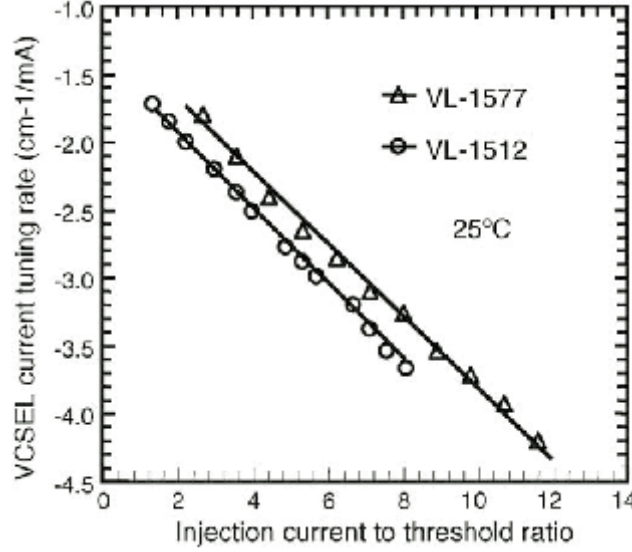
Figure 10. Frequency tuning rates for two differing VCSELs nominally operating at 15123 and 1577 nm. Tuning rates are plotted against normalized currents relative, with respect to each lasers threshold current [20].

The VCSEL's tuning by injection current therefore assumes the general second-order polynomial form

$$v(I,T) = a\frac{(I-I_0)^2}{2} + b(I-I_0) + K_T T + v_0$$

Eq. 25

where Table 1 below outlines the parameters and their associated values while I and T correspond to injection current and VCSEL heat sink temperature, respectively.

Table 1. VCSEL frequency tuning parameters, at 25°C

Parameter	VCSEL 1512	VCSEL 1577
Frequency of absorption line (v_{abs}, cm^{-1})	6621.48	6334.43
Injection current (I_0) tuning laser to v_{abs} (mA)	2.97	5.24
Temperature proportionality constant (K_T, cm^{-1} °C^{-1})	−0.51	−0.48
Reference frequency (v_0, cm^{-1})	6634.23	6346.43
Current tuning rate coefficient (a, cm^{-1} mA^{-2})	−0.30	−0.48
Current proportionality constant (b) at injection current I_0 (cm^{-1} mA^{-1})	−2.2	−3.55

The model's validity is manifest when the experimental values of table 1 were applied to
it and plotted against known absorption lines of CO_2, CO, H_2O and NH_3 in Figure 11 [21].
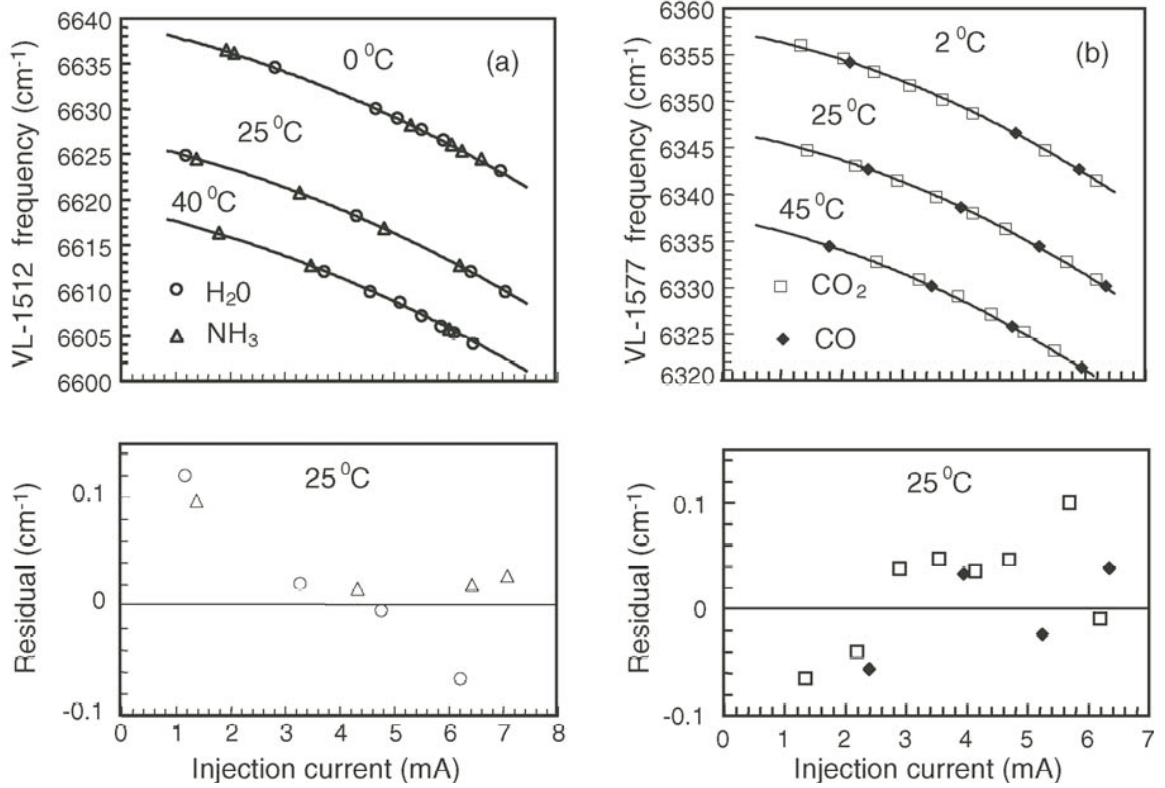
Figure 11. Fit of VCSEL tuning model to known spectra. VCSEL heat sinking temperatures and absorptive compounds are noted on the plots. Lower plots show the respective lasers' residuals at the given absorption features [20].

The small residuals serve to validate the general polynomial tuning model used to describe the VCSELs used in experimentation. Furthermore, from Figure 11 above we note the near-linearity of our VCSELs over the short tuning ranges, those less than about $1 cm^{-1}$. This therefore defines the limit to which the linear approximation may be used, a significant assertion to for subsequent modeling of wavelength modulation spectroscopy.

It is noteworthy to mention that the equation above points to the fact that frequency may be tuned by injection current or temperature separately. Though equation 25 was formulated with a linear dependence of frequency on temperature this suggestion was not substantiated experimentally owing to the slow time-tuning rate by temperature. This time rate of change is a practical limitation imposed upon us by the slow tuning rate of the thermo-electric-cooler TEC. As such, electronic tuning via injection current affords us the only practical means of tuning diode lasers at rates necessary in implementing harmonic phase detection. Thus, inherently, slow temperature tuning precludes its use for our objectives, and so, was not investigated beyond the first degree approximation.

Therefore, the tractability of the WMS linear model and the near-linearity of VCSELs when modulated over short current ranges behooves their use, whereby we accept the finite error this approximation as RAM.

3. EXPERIMENTAL SETUP

The use of wavelength modulation spectroscopy and detection at the second harmonic goes a long way in predetermining the mode of operation. Necessary for the front end, active in the generation of optical emission are, a modulation source, a tuning source, VCSEL along with temperature and pump (current) controller, and power supply. All this while staying true to the philosophy of a design for a portable methane detector.

The laser source used was the previously overviewed long wavelength BTJ-VCSEL from Vertilas[®]. Most often this was specifically the VL-1645-4-SP-A-T4 whereas substitutes' use is dully noted where applicable. The laser diode was housed in a Thorlabs[®] thermo-electric-cooler model TCLDM9 connected to a Thorlabs[®] TED350 controller. A Thorlabs[®] VITC002 current control board was used to drive the laser. The control board injection current was coupled to the VCSEL through the TEC. The current supply and control board provided a variable DC offset, invariably used to tune the diode output to line centres or as experimentation precluded.

In order to ramp-tune the spectral output of the VCSEL an Aligent[®] 33220A function generator was used. This provided the valuable capability of using various tuning schemes throughout experimentation, be it sinusoidal, triangular, or sawtooth. Modulation was supported by a purpose built circuit operating at 10kHz with variable output signal strength, circuit lay out is shown in appendix C. The modulation signal, along with the tuning signal, was impressed upon the DC bias through signal-support channels on the VITC002 control board. Power was supplied by a long-lasting 12V battery which was regulated to 5V before powering all but the function generator and temperature controller. Schematically, the front end components are shown in the upper left portion of figure 12.
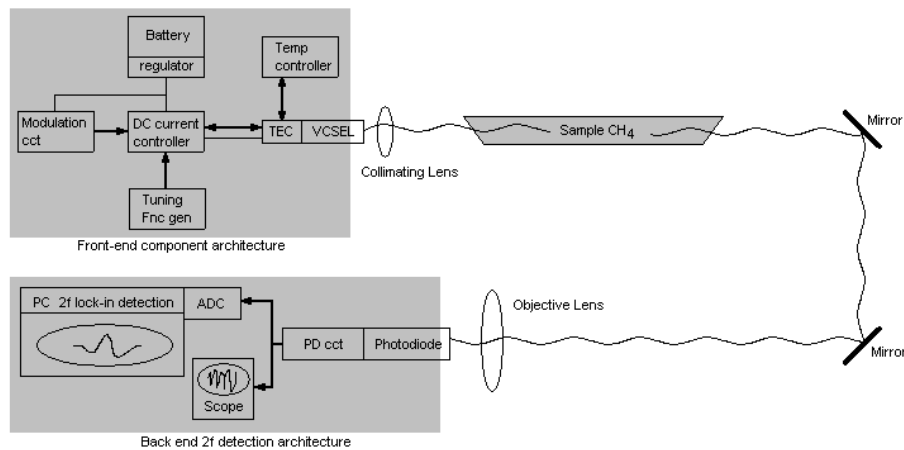
Figure 12. General experimental layout. Components dedicated to same task are within same shaded region. Sample cell is 50cm in length with windows at Brewster Angles.

In experimentations that needed ascertaining of optical frequency an etalon was also introduced into the path. The undulations characteristic of the free spectral range therefore provided a frequency detuning gauge relative to a benchmark absorption line.

On the back end, a Thorlabs [®] FGA310 InGaAs photodiode, PD, provides a photocurrent to a purpose built photo-detection circuit. Because detection limits of trace gas suffers invariably to sporadic noise, the major importance of light detection schemes on signal-to-noise cannot be understated, therefore, this circuit is subsequently examined in detail. Intermittently though, the PD circuit signal is digitized with a 16-bit National Instruments[®] PC interface. The $2f$-harmonic signal is then obtained with a fully digital lock-in amplifier designed in Labview[®] 8.0, covered in chapter 4.

3.1 Optical Path

In this portion of the thesis topics pertaining to optical factors relevant to the detection limit of the target gas are to be presented. Of primary concern is of course the manner in which the principal reaction is to take place, that is, the engagement of the target by the optical source. The basis of the detection scheme being the Beer-Lambert Law it is worthwhile to consider the target gas and the absorption line to be engaged.

In regard to the latter issue, methane was chosen. Its importance lies in being a largely man-made, global-warming gas, this along with being the principle constituent of natural gas. Though the first harmonic overtone absorption band in the near infra-red, NIR, around 1.65μm is 2% less than that of the absorption band around 3.39 μm, it provides for a practically tangible detection design using modern solid state laser and electronic technology. VCSELs nominally emitting light near absorption lines in this band were complemented on the detection side by InGaAs semiconductor PIN photodiodes.

Having measured the VCSEL beam divergence at 17 degrees the first necessary component to be integrated into the propagation design was a collimating lens. For the eventuality that long open path experimentation would be carried out a lens that was able to provide long spatial coherence was necessary. For this an aspheric 5 mm lens with a focal length of 6.3mm was used (Thorlabs [®] C110TM-C lens). With the use of this lens it was possible to achieve collimations to tens of meters. As all of the experimentation was conducted with path lengths not longer than 5 meters this sufficed.

The gas cell was designed to be a transmissive single-pass holding cell. This was implemented with a stainless steel tubing with bored flanges welded at the extremes. Melles Griot[®] BK7 windows with 5arcsecond off-parallel surfaces were used to couple light in to and out of the holding cell. The windows were set in seal-forming o-rings, sandwiched within the flange and bored faceplates.

Flanges were set at counter-parallel Brewster Angles for the same plane of polarization, Figure 13.
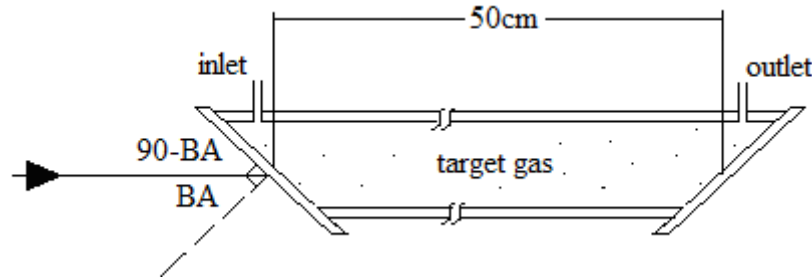
Figure 13. Cell design. Windows at cell entrance and exit set such that incident ray at Brewster Angle, BA. For the window index 1.5 BA = 57.3°.

Angling the windows prevented macro optical fringing due to etalons when tuning, while micro etalon fringes were countered with the slightly off parallel window surfaces. Furthermore, this arrangement maximized the transmission of light in to and through the cell. The holding cell, 50cm in length, along the body also possessed an inlet/outlet valve pair to allowing for the piping of gas in or out.

Though beam divergence was negligible, with a collimating lens, at these tabletop path lengths an objective lens was used to maintain continuity with future open path experimentation. The objective lens, as the collimating lens, was an antireflective coated aspheric short focus lens. This afforded tight concentration onto the small 1mm^2 InGaAs photodiode to be used. To limit background signal noise the photodiode was placed at the rear of a black encasement. At short-path table-top lengths scintillation and speckle effects were negligible though at larger open path lengths, studies, as applied to trace gas detection limitations, of these effects may be warranted. Commonly in portable gas detector design optical beams are steered back toward the emission package where the detection process takes place too. Beam paths are usually redirected back at the transceiver via mirrors, retro-reflectors, or arbitrary scatterers. This exposes the source VCSEL to potential optical feedback, possibly a substantial noise inducing phenomenon. For this reason this mechanism is of high interest in the investigation of noise sources limiting trace gas detection. Prior to investigating feedback in chapter 3 though, the primary issue of light detection in general needs be addressed. This interlude therefore follows in the next chapter.

4. OPTICAL DETECTION NOISE

The potential for improving SNR in trace gas detection applications is always a foremost thought. Noise limits the threshold of signal detectivity employing electronic circuitry and no different is it in photodetection applications. Optical spectroscopy entails a complex arrangement which lends itself to noise susceptibilities from a multitude of varying sources which may be difficult to isolate and strictly identify. However, noise contributed by optical detection presents a lower boundary and therefore a full treatment is warranted. In this chapter characterizing and profiling noise found to limit signal identification, in a setup using semiconductor photodetectors under various modes of operation, is undertaken. Because of the vast diversity of possible detection circuits the work embodied within this thesis employs both photoconductive and photovoltaic modes with but a few circuit architectures. By this study we endeavor to find the most suitable one, it being chosen as the most appropriate for $2f$-WMS detection

4.1 Photodiodes

In many instrumentation applications photodetection is facilitated through semiconductors operating on principles of the photoelectric effect. A semiconductor device with a p-n junction where light is absorbed in a depletion region generates electrical carriers and thus a photocurrent. Indium Germanium Arsenic (InGaAs) diodes are best suited for detection in the near infrared, the band of interest for our spectroscopic application, given the bandgap energy correlation to these wavelengths. Ideally each photon striking the PD would generate one hole-electron pair, and resultantly contribute to current generation. Photodetection is, fundamentally, a quantum process and so discrete units of charge will be generated [29]. Responsivity, *R*, the ratio of the short-circuit photocurrent generated by the photodiode (in miliAmperes mA) divided by the power of the incident light (mW) in this ideal case is therefore:

$$R = I / P = q\lambda / hc = 0.807 \cdot \lambda_{\mu m} (mA / mW),$$

Eq. 26

where,

- q is fundamental electronic charge,
- c the speed of light, and,
- h is Planck's constant.

This above derivation shows that the maximum expected responsivity at 1.555 μm would be 1.2 mA/mW while at 1.654 μm it would be 1.33 mA/mW.

Naturally this raises the question of necessary light detection levels, the answer provided by the following section. Later, evidence will be presented that shows how noise performance of the PD circuit is heavily dependent upon the irradiance befalling the PD.

4.2 Photodiode Equivalent Circuit

With a short analysis of the photodiode its use in this application is validated and operational guidelines are outlined. This subject is of great importance as effects such as non-linearity result in false-positive 2f signals in our application. In order to be able to analyze the performance of varying photodiode circuits a relatively easy means of modeling the PD itself must exist. To this end there exists a silicon photodiode equivalent circuit model [31]. With impunity I treat the already existent Silicon model as a general model and use it freely in the analysis of InGaAs photodiodes, a wholly valid undertaking considering the similarity in the underpinning mechanism responsible for photocurrent generation. This allows the concentration of work to remain on studying minimum detectable signal levels. The model itself is illustrated in figure 14 and its simplicity allows for quick integration within any already well understood circuit.
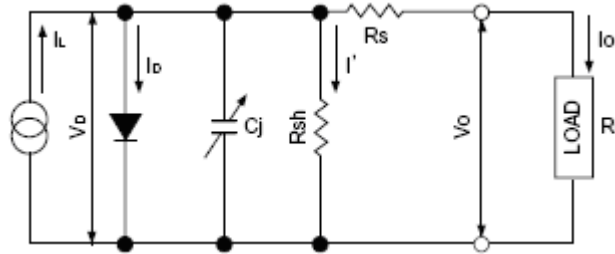
Figure 14. Equivalent circuit model.

- I_L : Current generated by the incident light (proportional to the amount of light)
- I_D : Diode current
- C_J : Junction capacitance
- R_{SH} : Shunt resistance
- R_S : Series resistance
- I' : Shunt resistance current
- V_D : Voltage across the diode
- Io : Output current
- Vo : Output voltage

We must always remember that this is merely a model of a whole two terminal diode, and thus, I the "manipulator" have no direct access to the internal elements. The only points of access are those across Vo, while all other component parameters are defined by particular performance specifications. The equivalent photodiode circuit (Figure 14) treats the photodiode as a perfect source of photocurrent in parallel with an ideal conventional junction diode, recalling, the junction diode differs slightly from the ideal diode in that its through current it is given by

$$i = I_{SAT}(e^{qV_D/KT} - 1)$$

   Eq. 27

where KT/q = 0.026 @ 25°C [33]; different behavior than a simply ideal diode. Ideally a photodiode would have R_S = 0 and R_{SH} = infinity. In this instance

$$I_O = I_L - I_D$$
    Eq. 28

If the external terminals are shorted together a short circuit photocurrent I_{SC} will flow. When this happens, for the ideal case, $V_D = 0$ and $I_{SC} = I_L$ (ideal case). However, for the real world case:

$$Io = I_L - I_D - I'$$
    Eq. 29

$$I_{SC} = I_L - I_{SAT}(e^{qV_D/KT} - 1) - I_{SC}R_S/R_{SH}$$
    Eq. 30

and since Vo = 0, thus: $V_D = I_{SC}R_S$

$$I_{SC} = I_L - I_{SAT}(e^{q(I_{SC}R_S/KT)} - 1) - I_{SC}R_S/R_{SH}$$
    Eq. 31

For the non-ideal case the second and third term of the above equation limit the linearity. In order to achieve good linearity R_S should be made as small as possible and R_{SH} as large as possible (indeed, this is the ideal assumption). Despite these shortcomings the short circuit light current is often quite linear over a wide range of light intensities, in fact, excellent linearity over several decades (of light intensity) can be expected [29]. The reader should infer that there is an operation point about which the diode is best for, and which great deviations from will result in highly non-linear behavior or saturation. At low light levels linearity is limited by the dark (or shunt) resistance and the noise current. At high levels of irradiance linearity is limited by the internal series resistance of the photodiode. In the writing here this will not be dealt with, however, it is something any user of photosensitive equipment must be aware of. Use of the photodiode in the linear region, for the trace-gas detection application, is therefore a terrific tool for making absolute light intensity measurements. This argumentation is provided therefore as a fundamental criterion for quality signal acquisition in gas detection. Namely, in order that one may ascertain range-integrated concentrations with a high degree of confidence, intensities detected must not be too low, nor too high. Both cases present vary real possibilities due to the high directionality of laser emission. In the former case low levels of detection occur with even slight misalignment whereas good alignment over short distances reflects the latter scenario whereby non-linear saturation results.

4.3 Modes of Photodiode Operation

As noted previously a photodiode is, fundamentally, a current generator. Practically thought, PD operation is governed by the circuit architecture which the PD is a part of. As a result there are three principal modes of operation; biased, zero-biased, and short circuit mode. Comparison of each of the modes will highlight the benefits and pitfalls of each and lead into the experimental techniques used.

4.3.1 Photovoltaic or Zero Bias Mode

When photons strike the PD photocurrent flows through the diode's shunt resistance. The voltage across the resistor opposes the band gap potential of the photodiode junction, and consequently forward biases it. The shunt resistance

value drops exponentially as the illumination increases giving rise to a photo-generated voltage that is a logarithmic function of incident light intensity, Figure 15.

Photovoltaic P-I Profile
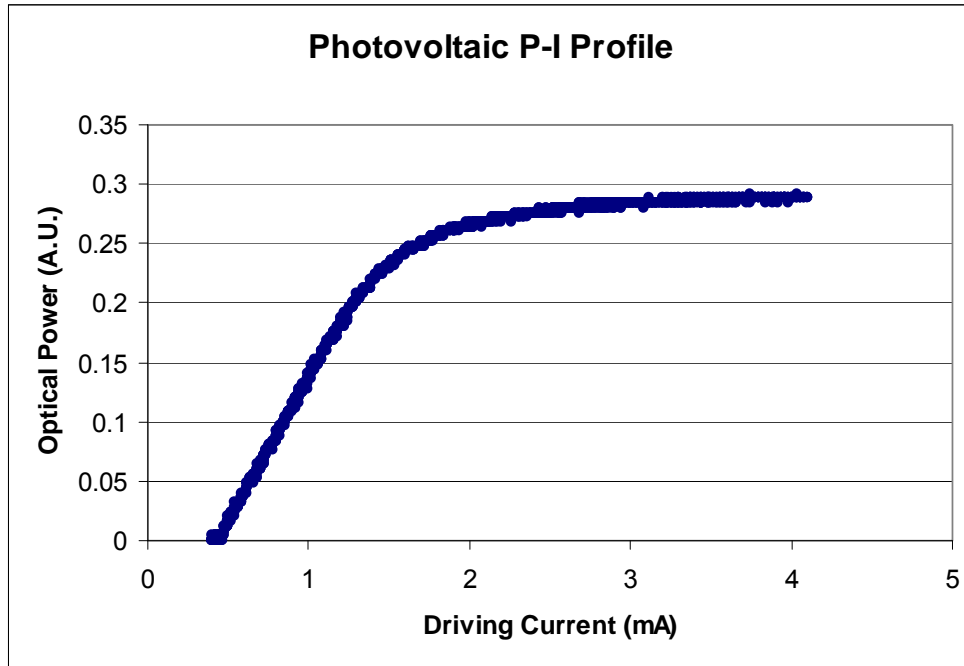
Figure 15. Non-linear response of a photodetector circuit. That to be outlined in figure 16. Laboratory measured data points.

It is perhaps the easiest mode to implement as minimal circuit complexity is warranted, however, the major disadvantage of this circuit is the non-linear behavior described above. It is pictured in Figure 16 below. Additionally, the output signal depends largely on the shunt resistance, which typically has a wide spread of values over different production batches.
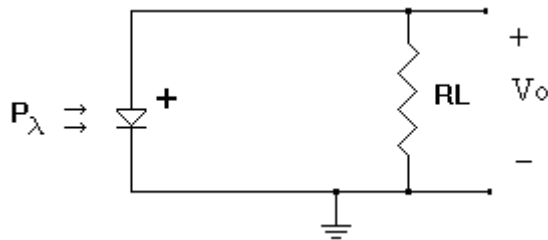
Figure 16. Zero Bias PD Mode.

Rather the arrangement shown in figure 17 is recommended.
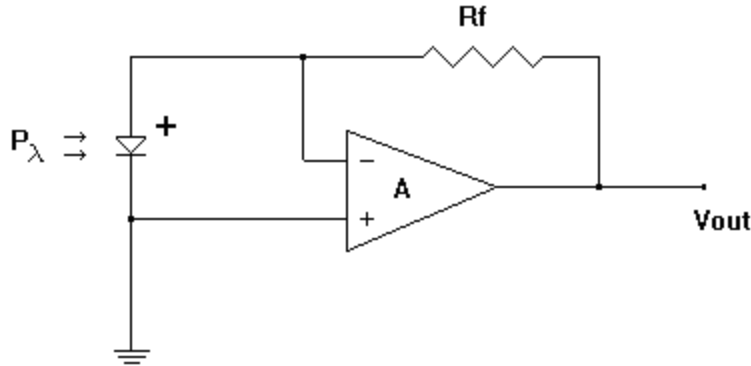
Figure 17. Photovoltaic mode in transimpedance configuration.

One way to achieve sufficiently low load resistance, and an amplified output voltage, is by feeding the photocurrent to an operational amplifier. The generated photocurrent flows through R_f which is fixed. The resultant voltage is therefore linearly dependent on the incident radiation level. So, the circuit has a linear response and has low noise due to the almost complete elimination of leakage current.

4.3.2 Photoconductive or Biased Mode
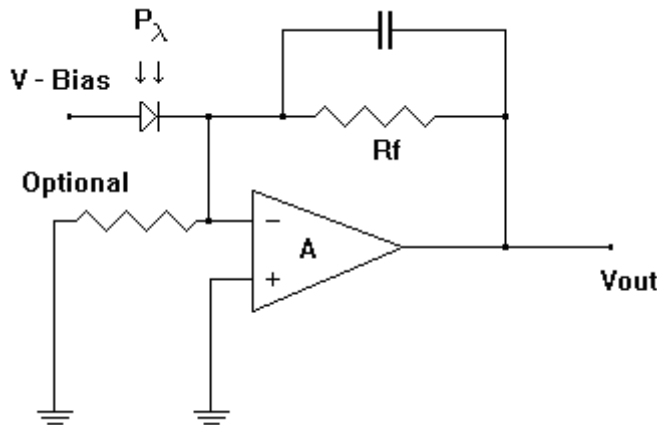
Figure 18. Reverse bias photoconductive circuit.

In the photoconductive mode, the generated photocurrent produces a voltage across a load resistor in parallel with the shunt resistance. Since in the reverse biased mode R_{SH} is substantially constant, large values of R_f may be used still giving a linear response between output voltage and applied radiation intensity. This form of circuit is required for high speed of response. The main disadvantage of this mode of operation is the increased leakage current due to the bias voltage, giving higher noise than the other circuit modes already described. A typical photoconductive mode circuit is shown above, Figure 18. Of major concern when dealing with photodiodes that are reverse biased is the potential to destroy the diode by driving too much current through it. With a small battery supply (5-10V in the experimental setup) the reverse bias breakdown should not be exceeded, nevertheless, individual diode's spec sheets were always cross-referenced to

28

assure this. In addition, under excessive illumination, damaging current can result with too small a load resistance so this condition was summarily avoided.

One of the simplest configurations of a detector circuit consists of a reverse biased PD in series with a resistor, Figure 19, and it is certainly the simplest bias configuration.
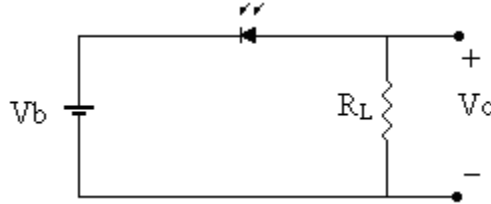
Figure 19. Simple reverse bias configuration.

This simple arrangement lent itself to easy alterations in sensitivity by varying R_L, with lower value resistances yielding lower sensitivities. Evidently, for a constant power and wavelength it can easily be realized how greater R_L values produce greater sensitivities where Vo is governed by Ohm's law, with R_L behaving simply as a weighting factor to the passing current. Thus, Vo may be expressed as

$Vo = P*R(\lambda)*R_L$
           Eq. 32

Importantly, also, in this circuit arrangement the output responds linearly to power changes, Figure 20.

Photoconductive P-I Profile
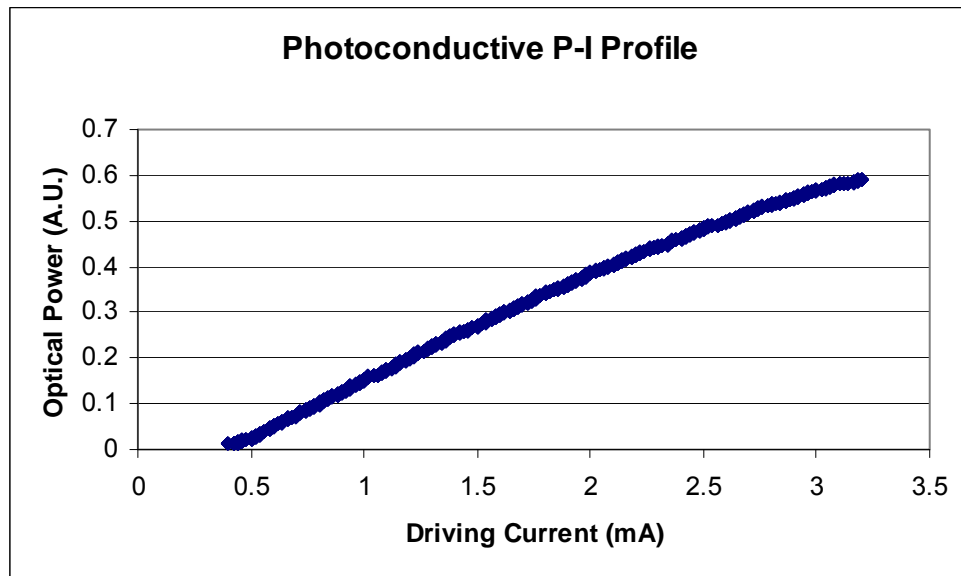
Figure 20. Linear response of such circuit as depicted in figure 19. Laboratory measured data points.

Because this circuit does not consist of semiconductors (barring the photodiode itself), or energy storage components this circuit will not oscillate, a phenomenon that does occur in the transimpedence amplifier configuration. Recalling the

presence of the diode's capacitance, the idealized operation of this circuit yields the following bandwidth

$$f_{BW} = \frac{1}{2\pi C_j R_L},$$

Eq. 33

dependent, on both the load impedance and the diode [29]. Using the Thorlabs FGA10 our bandwidth profile was as shown in Figure 21.
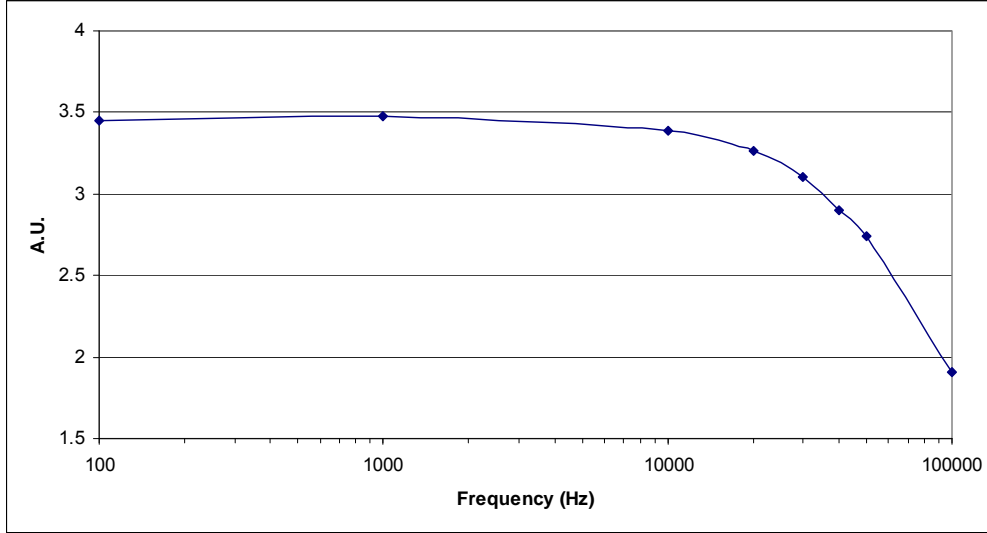
Figure 21. Photoconductive bandwidth profile at 1mA offset photocurrent. Using photodiode FGA10 with 5V bias. Laboratory measured data points.

Obviously a minimum diode capacitance and load resistance enhances the bandwidth, though lowering R_L diminishes the achievable sensitivity. This presents the first opportunity to emphasize the importance of minimizing the intrinsic photodiode capacitance. As noted previously, larger bias voltages are an effective means of lowering said capacitance.

In lieu of these benefits there are certainly characteristics that limit this circuit's usefulness. One such limitation is the sensitivity limit even with very high load resistances, exceeding, say, 1MΩ. One problem is that a 1MΩ resistor may not in fact posses that value at high frequencies due to capacitive effects. Another problem is coupling. If a 1MΩ oscilloscope is used to measure the output voltage the shunting of the scope impedance with the load yields a lower output than expected. Also, problems may occur with voltage offsets, even in total darkness, caused by leakage currents driven by the reverse voltage. Therefore, this arrangement's usefulness has shown to be limited and using an optimized amplifier detection circuit I believe is preferable.

4.4 Analysis of transimpedence performance

One excellent method of speed enhancement is to use the transimpedence amplifier configuration shown earlier in figures 18 and 19. The op amp is connected with resistive feedback provided by the load resistor while the noninverting input is grounded, see Figure 22.
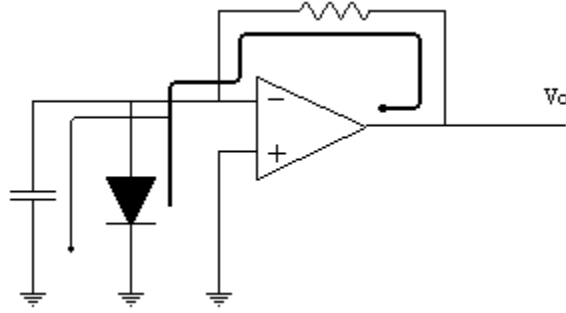
Figure 22. Transimpedence amplifier current flow.

This configuration is typical of opamp inverting-feedback configurations. The inverting input is taken to be a virtual ground, while the input resistance is very low. Because the amplifier's input current is very low, a few tens of picoamperes in some cases, the bulk of the photocurrent has to flow, as shown, through R_f to the amplifier's output. With the polarities shown, the output current must therefore become negative to pull current out of the photodiode anode. With the capacitance connected to the virtual earth, changes in photocurrent barely change the voltage on Cp. If the voltage does not change then neither does the charge Q=CV, and its apparent capacitance is greatly reduced. In the ideal case the diode capacitance is effectively shorted out, making it invisible to the photocurrent and feedback resistor. Consequently, the response time of this circuit, a concern given the nature of 2f-WMS, is significantly improved compared to the simple zero bias circuit. Without delving into rudimentary op-amp theory the upper frequency limit or bandwidth of the photodiode-transimpedence amplifier configuration is:

$$f_{\lim it} = \frac{GBW}{\sqrt{2\pi R_f C_j}}$$

Eq. 34

This formulation is derived though approximate considerations, and as such, should not be relied on for exacting accuracy but, rather, to provide a guide to the systems bandwidth. One practical reality of using the transimpedence configuration is that it oscillates. This is because the extra phase shift caused by the low-pass R_fCj feedback is added to the amplifier's own phase shift. At some high frequency it will probably lead to positive feedback and if the gain is above unity it will oscillate. One solution is to add a capacitance in parallel with R_f to reduce the transimpedence at high frequency. Note, this oscillation was observed on several detectors built and even varied between different instances of identical circuits.

Key issues are the capacitance of the photodiode, limiting speed through its interaction with the load resistor, and the difficulty of making large value load resistors which look resistive out to high frequencies. The common 100MΩ, 0.4-W component has an impedence of 100 MΩ only out to 8kHz. Beyond that its capacitance will dominate performance, and the physical size of the resistor and the details of the circuit layout play a bigger role than the active components. By understanding the principles, though, we were able to make receivers with a bandwidth and sensitivity sufficient for the application. Although it is not really a

separate topic at all, but should be handled as if intimately bound with the provision of adequate bandwidth, the calculation of noise is handled next.

4.5 Photodiode Noise

Ultimately system performance is graded by the final signal-to-noise ratio (SNR) it provides. Considerable time has already been spent on looking at the signal (photocurrent), how to maximize it, and how to measure it in a useful bandwidth. Now a shift is exacted with a look at the fundamental sources of noise as far as photodiode detection is concerned. All detection systems are limited, at least, by shot noise and thermal (or Johnson) noise. The noise generated by a photodiode, operating under reverse bias, is a combination of shot noise, due to dark leakage current, and Johnson noise due to the shunt resistance of the device and the ambient temperature.

4.5.1 Shot Noise

Plainly put, shot noise is the uncertainty in determining the magnitude of current. If a large number of precision measurements are made of a nominally constant current and the results are plotted, the results should be distributed evenly around the nominal value. If the obtained values are plotted in a histogram, a Gaussian distribution will manifest with a variance equal to the mean. A constant photocurrent (Ip) exhibits a current shot noise power spectral density ($I_S{}^2$) given by:

$$I_S{}^2 = 2qI_dB \quad or \quad I_S = \sqrt{2qI_dB}$$

where:
- I_S = shot noise current
- q = electronic charge ($1.6*10^{-19}$ coulomb)
- I_d = dark leakage current (A)
- B = bandwidth of system (Hz)

Of importance to note here is that the current-noise power spectral density is proportional to the measurement of the photodiode, such as the reciprocal of the transit time of electrons through the detector junction. The spectral density of shot noise is also "white" meaning that the power per unit bandwidth is independent of frequency. The current noise power spectral density can be interpreted as the variance of the current Ip, assuming the distribution of current values looks Gaussian. In reality the probability P(I) of a given current cannot be Gaussian, which has the form:

$$P(I) = e^{-(I-I_{mean})^2 / \mathrm{var}iance}$$

because this would give a finite probability of our measuring negative currents. The true distribution is a Poisson one and comes from the analysis of photon statistics. However, for situations where the mean detection rate of photoelectrons is large, the Poisson distribution becomes near-identical to the Gaussian distribution. In practice, even in weak-power detections, the quantity of electrons comprising current are so large that assuming a Gaussian distribution

about the mean current is logical.   In this treatment therefore, the variance in the photocurrent gives rise to the fundamental noise contribution we call shot noise.

It is desired that shot noise dictate the ultimate precision with which the photocurrent can be measured (this will be explained later), and typically it is stated in units of pA/√(Hz).
A big issue is what currents actually show full shot noise? This is a contentious issue, Netzer (1981) suggests that shot noise is seen in situations where charge carriers cross a barrier independently of one another, such as pn-junction diodes where the passage occurs by diffusion, a vacuum-tube cathode where electron emission occurs as a result of thermal motion, and photodiodes.

4.5.2 Thermal (Johnson) Noise

The second fundamental source of noise, thermal noise, was investigated by Johnson and Nyquist in the 1920s, hence the moniker. It is present in all resistors at a temperature above absolute zero and is characterized by internal current fluctuations and fluctuations in voltage across their open circuit terminals, even when no external current is flowing. If connected across into an external circuit, these will cause external current fluctuations. Note, however, that even though a warm resistor acts as a little generator, it is not possible to extract power from it. Analogous to the treatment of shot noise currents, a resistor of R ohms will show a noise power spectral density given in voltage or current by [29];

$e_n{}^2 = 4kTBR$ $\qquad$ (in V²)
$\qquad$ Eq. 35

As with shot noise, this thermal noise power is proportional to the measurement bandwidth. To ease calculation the above may be rewritten in two handy forms;

$$e_n = \sqrt{4kTR} \qquad \text{(in } V/\sqrt{Hz} \text{)}$$
$\qquad$ Eq. 36
$$I_{jn} = \sqrt{4kT/R} \qquad \text{(in } A/\sqrt{Hz} \text{)}$$
$\qquad$ Eq. 37

where,

- I_{jn} is the Johnson noise current,
- k being Botzmann's constant,
- T the absolute temperature in Kelvin,
- R the resistance giving rise to noise (ohms), and
- B the bandwith of the system, in Hz.

The Johnson noise contribution is provided by the shunt resistance of the device, series resistance and the load resistance. The total noise current is the root mean square sum of the individual noise current contributions.   Shot noise is the dominant component of the noise current of a reverse-biased photodiode. This is particularly true at higher voltages. If devices are operated in a photovoltaic mode with zero bias, the Johnson noise dominates, as dark current (appendix D) approaches zero. When operating in the zero bias mode the noise current is reduced such that the NEP, and hence the minimum detectable signal, is reduced in spite of some loss of absolute sensitivity [29], [33].

4.5.3 Short Circuit Noise

It is useful to estimate the detection noise and performance of the simple bias photoconductive system described earlier. The noise sources to be considered are the thermal noise of the resistive load, the shot noise of the leakage current, and the shot noise of the signal photocurrent.

It is clear that a higher value of load resistor gives more thermal noise. However, the larger load also gives a much larger signal, with the result that the SNR improves with the square root of the increasing load. The greater the photocurrent, the greater the noise but the greater the SNR. It is generally, our goal to design the measurement system to be shot noise limited. Why? Given that the total noise due to shot and thermal noise contributors is

$$\sqrt{(RI_S)^2 + (RI_{jn})^2} = \sqrt{\left(\sqrt{2qI_d BR^2}\right)^2 + \left(\sqrt{4kTR}\right)^2}$$

we see that while the thermal noise is the dominant contributor, increases in R increase noise $\sqrt{R}$-like, while the signal strength, $I \cdot R$, increases linearly with R. Therefore the return on SNR increases per decade increase in R, is great, $\sqrt{R}$ like in fact. On the other hand, however, if shot noise dominates, noise increases like the signal, linearly that is (just about for $I_S \gg I_{jn}$), and so the benefits of increasing R further diminish. So, by pushing the signal into the shot noise limited domain the best use has been made of the photocurrent to yield high SNR [29]. This concept is of such importance that it is worth repeating: operating in the shot noise limited domain makes best use photocurrent for yielding high SNR. No better can be done than that! For example, if Johnson noise dominates so much that it is assumed the only noise source, then increasing R by a factor of 10 will increase the SNR by $\sqrt{10} = 3.16$ times. If the shot noise is dominant enough to act as the sole noise contributor, then in this case, increasing R by a factor of 10 would not change the SNR.

In order to determine whether the circuit is optimized to be shot noise limited we can equate the thermal noise voltage density with the shot noise voltage density:

$$\sqrt{4kTR} = R\sqrt{2qI_P}$$

Eq. 38

And obtain

$$V_0 = I_P R = 2kT/q = 52mV$$

Eq. 39

Therefore, if the V_0 observed is greater than approximately 52mV we can be confident that we are limited by the shot noise statistics. If $V_0 < 52mV$, thermal noise in the load resistor should define the limit to SNR. I say "should" because other noise types (e.g. $1/f$ noise) or interfering signals may actually form the limit to SNR.

4.5.4 Operational Amplifier noise

All amplifier systems can be described in terms of two frequency-dependent noise sources, a voltage noise generator e_n, and a current noise generator i_n, figure 23.
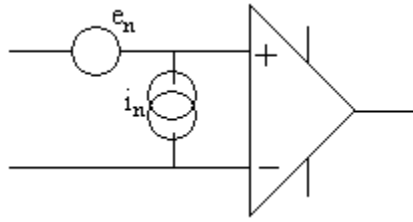
Figure 23. Op amp noise source depiction.

Essentially, the current noise is just the shot noise of the bias currents in the amplifier stages. FETs require much smaller gate currents than their BJT counter parts while both posses bias currents whose magnitude is very dependent on temperature. The equivalent voltage noise is usually due to thermal noise in resistive components. For example the base/emitter resistors of input stage transistors. They are often similar for BJT and FET opamps of the order of 5 to $50\,nV/\sqrt{Hz}$. Not all noise sources are "white". The current and voltage noise parameters which are so necessary to photodetector analysis are therefore not scalar quantities. Inspection of most opamp data sheets indicates that below a certain frequency, called the lower corner frequency f_L, both these noise parameters increase. The frequency variation of noise density is different for FET and bipolar opamps. This is especially marked in the case of the equivalent voltage noise generator. For example, the data sheets for the LMC7101 give $37\,nV/\sqrt{Hz}$ at 10kHz, increasing to $80\,nV/\sqrt{Hz}$ at 100Hz, $200\,nV/\sqrt{Hz}$ at 10Hz, and $600\,nV/\sqrt{Hz}$ at 1Hz. This is assumed to be an instance of so-called 1/f noise. Current noise densities also show some 1/f character, but usually this is less pronounced and starts at a lower frequency. The 1/f character means that the noise power per decade is proportional to the reciprocal of the frequency. The noise varies as:

$$i_n^{\,2} = i_{no}^{\,2}(1 + f_L / f)$$

Eq. 40

Here i_{no}^2 is the "white" contribution to total noise. The high noise power density at low frequencies strongly suggests that measurements should be carried out at a higher frequency, preferably at audio frequencies or higher. This point was made previously in the context of laser excess noise and is now being reiterated here. Therefore it represents the modus operandi fundamental to trace gas detection by two alternate accounts. Light sources should therefore be modulated! Either taking only the high-frequency noise spectral densities, or including the full variation with frequency, system noise calculations are "simply" down to determining the output noise contributions from these two noise generators, modified if necessary by any connected circuitry. In practice, this process is sometimes far from simple.

4.5.5 Transimpedence Amplifier Noise Peaking

Feedback resistors exhibit stray capacitance, photodiodes have parasitic capacitance, and amplifiers have input capacitance and frequency-dependent gain

characteristics, all of which combine to modify the overall noise as a function of frequency. In this section I address the main features which can significantly affect dynamic system performance. The approach is simply to apply well-known feedback amplifier gain expressions, including all the significant frequency-dependent components. The input load seen by the photodiode current generator is the parallel combination of the input capacitance of the amplifier Ci, and the parasitic capacitance of the photodiode Cj [29]. The easiest way to treat the thermal noise of the load resistor is to consider it as another current source, $4/\sqrt{R(k\Omega)}$, in $pA/\sqrt{Hz}$, in parallel with the photocurrent and shot noise sources. All three current sources then show the same low pass characteristic due to the falling shunt impedence Z_{sh}.

We now lump together all the elements of the photodiode equivalent circuit, input parasitics, and those of the feedback elements into general input and feedback impedances Z_{SH} and Z_f, respectively. If the opamp is ideal noise flows through the feedback impedance Z_f, which is the parallel combination of the load resistor R_f and its parasitic capacitance C_f. Above a break frequency f_1 the signal output decreases at -20 dB/decade. Writing in the complex frequency (s) domain the impedance of C_f:

$Z_{cf} = 1/sC_f$
Eq. 41

$Z_f = R_f Z_{cf} /(R_f + Z_{cf})$ (parallel combination of Z_{cf} and R_f)
Eq. 42

$f_1 = 1/(2\pi R_f C_f)$
Eq. 43

That is, the photocurrent, feedback resistor thermal current noise, and amplifier current noise outputs decrease together above f_1 [35].
The effect of the circuit configuration on the amplifier's voltage noise is more complicated. The voltage noise generator sees an inverting amplifer with a gain given by the ratio of feedback to input impedance. Hence this noise is amplified to:
$e_n(1+ Z_f / Z_{SH})$ where: $Z_{SH} = R_{SH}Z_{cp}/(R_{SH}+Z_{cp})$ (parallel combination of Z_{cp} and R_{SH})
C_p as before contains all the input capacitances. This combination to noise therefore increases at 20 dB/decade above a break frequency f_2 where the impedances Z_f and Z_{sh} are equal [29],[35]. With large transimpedance resistors and/or high-capacitance photodiodes this will be typically $f_2=1(2\pi R_f C_p)$.

signal gain

-20 dB/dec : feedback impedance dropping due to Cf

f1=1/2piRfCf

limit due to opamp open-loop gain

1+Csh/Cf

Cf/Csh

log |A| dB

+20 dB/dec

f2=1/2piRfCsh

log frequency
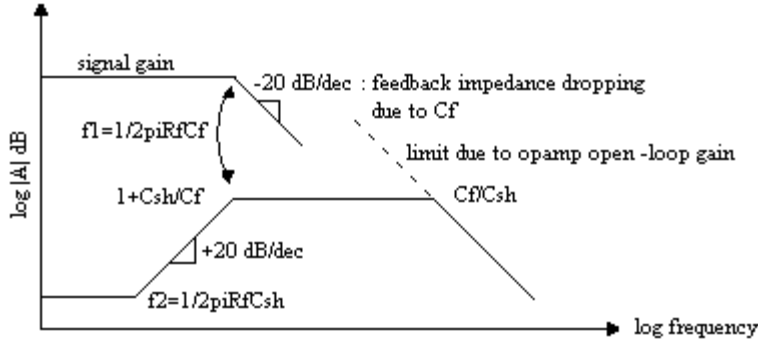
Figure 24. Gain representation of noise peaking phenomenon.

The various characteristic frequencies are shown in figure 24 as a schematic Bode plot. Where large detectors are used with their high capacitance and consequently low impedance and high value transimpedances, the magnification of the voltage noise caused by this effect can easily become the dominant noise source. As can be seen from figure 8, the noise gain increase is checked by the parasitic capacitance of R_f and eventually reduced by the dropping open loop gain of the amplifier [35]. In between a region of high excess noise is often seen, which is termed gain peaking. Adding extra capacitance across R_f can lower the f_1 break frequency to reduce the peak, but only at the expense of signal bandwidth. This is a serious consideration in this system. Given a particular photodiode, large enough to collect most of the signal light and a transimpedance to provide adequate signal voltage, there is not much we can do about gain peaking except look for amplifiers that contribute the lowest possible noise. It is one of the surprising aspects that I found in this study, that despite working with nA photocurrents and MΩ resistors, it is often not the current noise but the voltage noise density e_n that defines our overall performance. The majority of IC opamps exhibits and $e_n = 20 \, nV / \sqrt{Hz}$ in the flat region above the onset of $1/f$ noise, but there are exceptions. The Burr-Brown OPA121 is a FET-input 2 MHz gain-bandwidth device with 1 pA bias currents and e_n typically $6 \, nV / \sqrt{Hz}$. This is about the limit now for integrated FET amplifiers and it by this study that their use is employed.

4.6 Simulation

The above coverage of noise performance in the transimpedance amplifier configuration is difficult to envision. The following section further clarifies how noise behaves in this circuit.

In order to analyze this I have simulated a transimpedance PD receiver using Matlab (appendix A). It is a sensitive receiver for 1nW detection at low audio frequencies. A 100 MΩ transimpedance has been chosen, while across it is a feedback capacitance of 1.6pF. The photodiode is a 10mm² device with a capacitance of 400pF. It is with the results of this work with light emitting diodes, LEDs, that subsequent designs for VCSEL emission-detection drew upon.

Table 2 summarizes the calculated results and figure 25 shows the calculated frequency responses.

Table 2. Transimpedence Analysis of a 1nA Reciever

Signal photocurrent Ip	1nA
Feedback resistance Rf	100Mohms
Feedback capacitance Cf	2.7pF
Photodiode capacitance Cp	400pF
Input noise voltage density en	10nV/√(Hz)
Input noise current density in	1.1fA/√(Hz)
Signal, current noise break frequency f1	1000 Hz
Voltage noise break frequency f2	10 Hz
Thermal noise of feedback resistor at 1kHz	0.63 μV/√(Hz)
Current noise contribution at 1kHz	0.05 μV/√(Hz)
Voltage noise contribution at 1kHz	0.5 μV/√(Hz)
Signal shot noise voltage	0.9 μV/√(Hz)
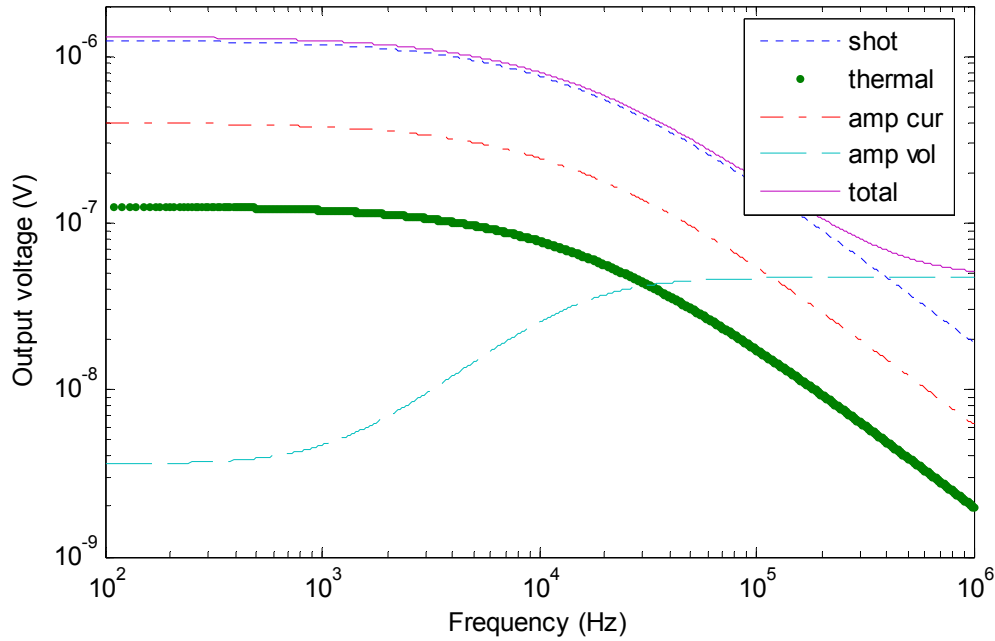
Figure 25. Noise profile for a PD transimpedance amplifier. Modeled with Matlab.

It can be seen that although at 10kHz the amplifier current noise density is greater than the voltage noise density contribution, there is a crossover at about 30kHz. Up to 400kHz the transimpedance thermal noise is greater than both amplifier noise contributions, but just less than the signal shot noise for this 1nA photocurrent. Above about 400kHz the receiver is no longer shot-noise limited, due to the increasing amplifier voltage noise gain peaking. The design highlights the severe limitations of even small amounts of stray capacitance on the high-value feedback resistance. Even 1.6pF causes a drop in signal gain above 1000Hz. Equally, the 160pF photodiode capacitance causes voltage noise to be magnified starting at 1000Hz. Between this frequency and the second break, this noise peaks by almost 40dB.

4.7 Experimental Results

Experimentally, separation of LED and photodiode was arranged to reflect power level changes due to absorption at trace levels. This corresponded to powers on the order of micoWatts. This was overlaid with an extremely weak sine-wave modulation to mimic signals that result due to absorption by gas compounds. In the experimental setup the photoconductive mode was implemented with the architecture shown in figure 26a.
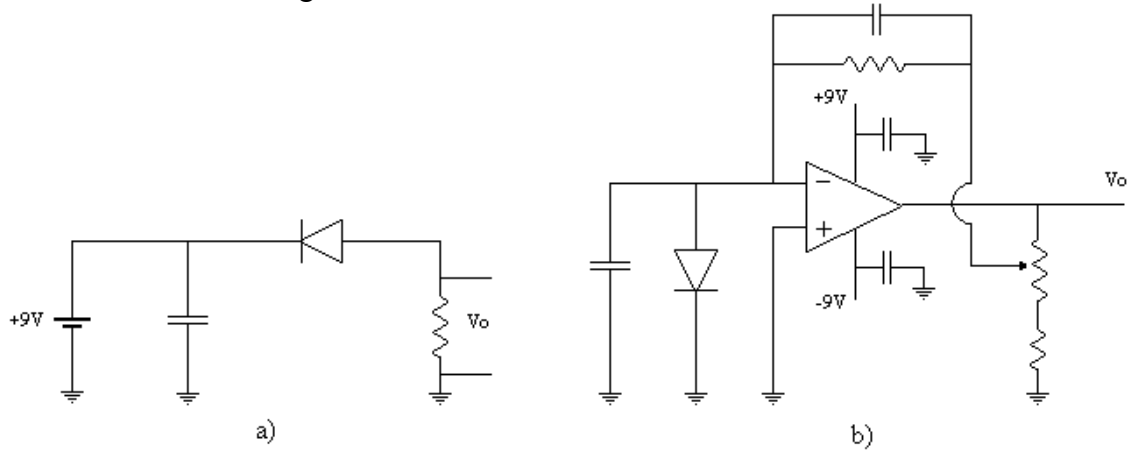
Figure 26. Circuit layout a) photoconductive mode b) photovoltaic mode.

Figure 26b shows a transimpedence circuit layout, a alteration from the ideal one. Both circuits include filters to retard any noise that may occur in the power sources. All of the pertinent circuit parameters are provided below in Table 3.

Table 3. Experimental circuit components

Photoconductive		Photovoltaic	
LED	0.008 mW	LED	0.008 mW
LPF Capacitor	0.1 uF	LPF Capacitor	0.1 uF
LPF Resistor	1kohms	Op Amp	OPA 128
PD	Hamamatsu PD	PD	JDS Uniphase EXT 1000 T
RL	10kohms	RF	1-10 Mohms
		CF	2.7 uF
		CSH	400 uF

The photodiode used in the photoconductive mode is a Hamamatsus$^®$ visible PD while the photovoltaic mode employs a JDS Uniphase EXT 1000 T. Though they are differing photodiodes the differences between the two operating modes are nevertheless easily identified. Much of the focus from here on in is on the photovoltaic circuit because it requires more analysis due to its greater complexity relative to the elementary photoconductive circuit.
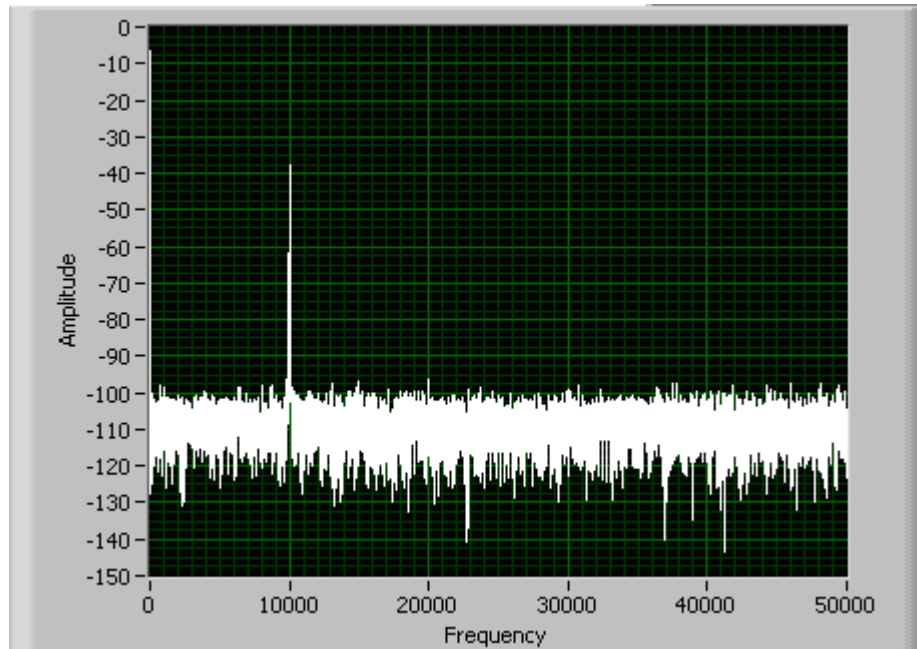
Figure 27. 10kHz modulation signal. Laboratory measured data.

Figure 27 above shows the detection of a 10kHz modulated -38dB signal. In the figure we can also see the ever-present noise level for the transimpedance circuit, here taken to be about -100dB. Comparing the signal level to one like in Figure 28 we see that at 60kHz the signal has droped 3dB. This would indicate that the bandwidth of this circuit configuration is 60kHz. In fact this is consistent with what may be expected using $f_1 = 1/(2\pi R_f C_f)$ and the circuit parameters given in table 2.
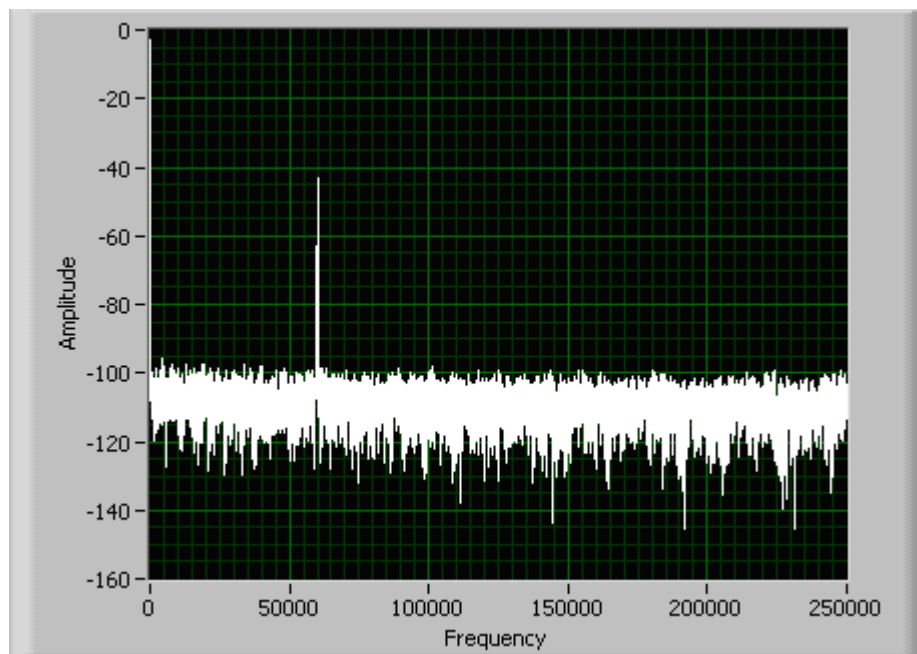
Figure 28. Detection of 60kHz modulation signal. Laboratory measured data.

Certainly of extreme importance is to note the SNR difference between the two circuit types. In the photovoltaic configuration shown in figure 27 the signal strength of -38dB with a noise floor of -100 dB is evident. In the photoconductive mode with the same input figure 29 shows the signal strength to be roughly -82dB with the noise at -109 dB.
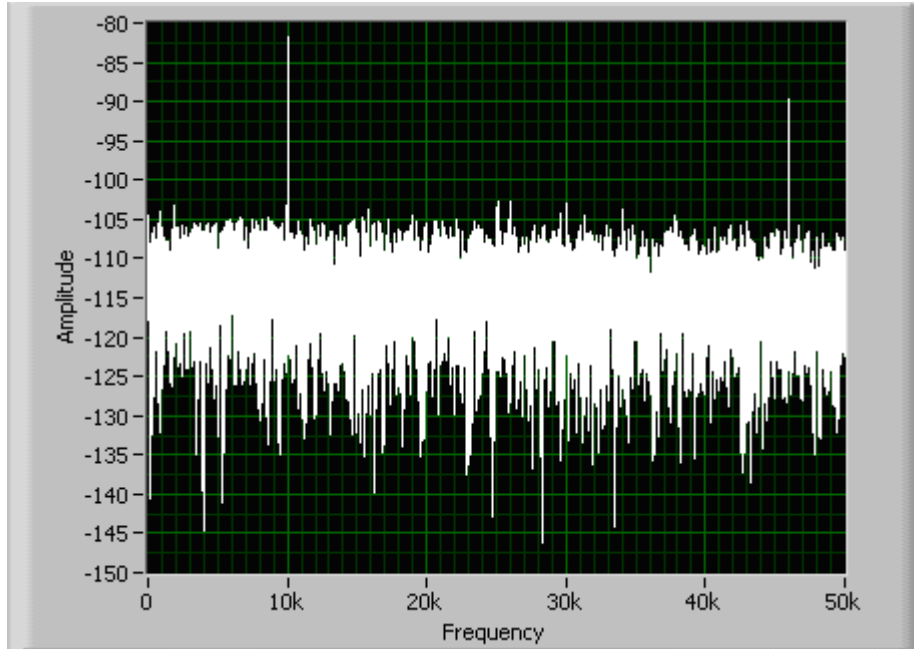
Figure 29. Signal at 10kHz (implemented in photoconductive mode). Laboratory measured data.

So, even though the noise has decreased by a factor of 8 the signal in the photoconductive mode has decreased by 5 orders of magnitude resulting in a drastic decrease in SNR.

As discussed earlier the noise behavior in the transimpedance configuration is highly dependent upon the component parameters used. In addition, differing noise sources contribute differently as we sweep through the available frequency range. Though a keen observer may perceive noise roll off in, say, Figure 29, Figure 30 shows the effect clearly.
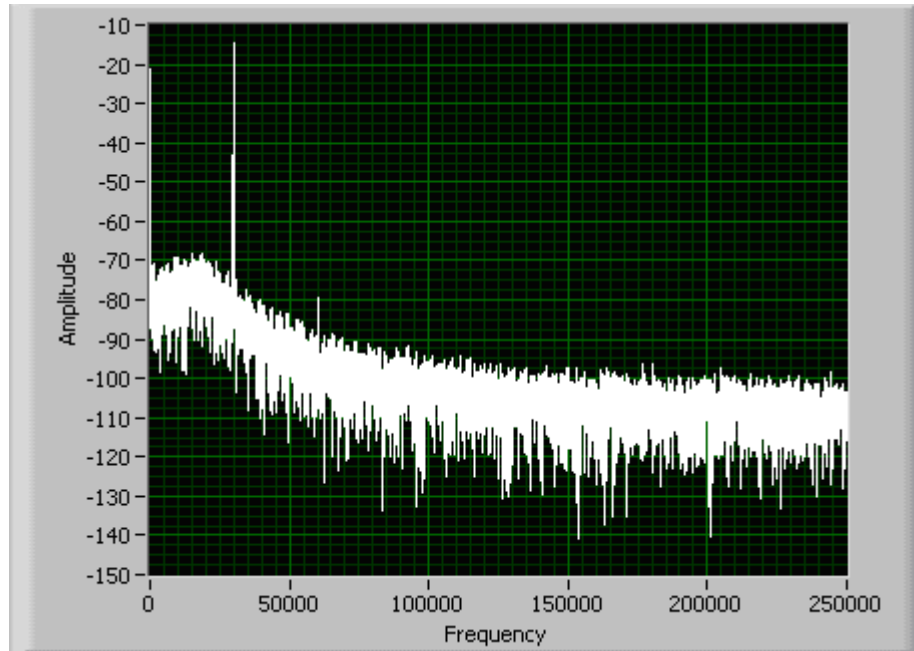
Figure 30. Noise roll off with extreme gain. Laboratory measured data.

A peak of the noise term is also evident. Ascertaining the noise profile shown in Figure 30 analytically is difficult because of the circuit configuration itself. As shown in Figure 26 the feedback capacitor is across the feedback 1MΩ resistor, namely, not across the entire feedback loop. In fact, a better setup would have been the complete replacement of the feedback gain structure with a simple 10MΩ resistor. These considerations now provide the basis by which to design the detection schemes for VCSEL emission.

4.8 VCSEL Light Detection

Though this chapter has limited noise to circuit noise it would be incomplete without touching upon interference noise. Within the lab a strong source of noise was found to be contributed by the overhead lighting. In the laboratory the fluorescent lights contributed noise components at approximately 47kHz and 92kHz, shown in Figure 31. The ballast necessary for the fluorescent lights is hypothesized to be the cause of this phenomenon. Readings have indicated that in fact ballast can contribute to frequency components anywhere between 40 and 100 kHz [22]. Ultimately, the florescent lighting contributed an additional 3mV atop of laser detected signals; a considerable quantity when often raw detected VCSEL signals were no greater than 100mV.
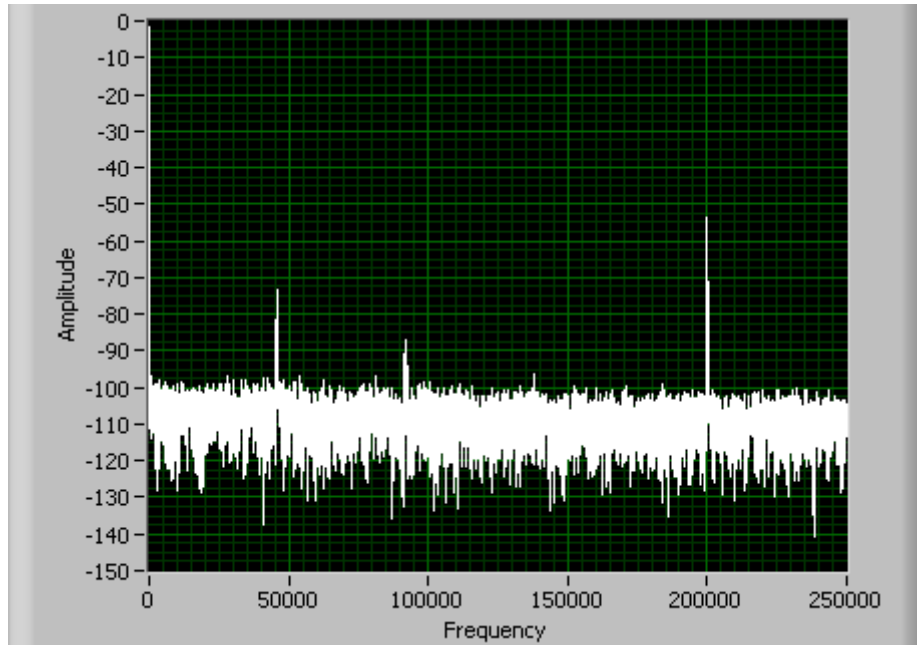
Figure 31. Flourescent lighting noise at 47 and 92 kHz with a 200kHz signal. Laboratory measured data.

Because of these noise components all of the experimental results were obtained with strictly no ambient lighting. As a whole though, of all noise that contributes to the debilitation of the SNR the most important one may in fact be the photodetector circuit layout and composition. The care and prudence with which it is constructed influences important factors such as parasitic capacitances and electrical interferences. It is certainly the easiest measure of improvement that can be enacted upon any circuit mode, or architecture chosen for operation.

Also figures 28 through 32 all contain a strong DC component that has not been previously alluded to. In reality it is the very reason a LED source was used as opposed to a laser diode. The laser would provide a truly magnificent DC component with the amount of gain that the transimpedance arrangement produces. As such, photodiode saturation was an issue. One solution to this is the incorporation of a capacitor between the photodiode and opamp, effectively, AC coupling the two. Another solution may lay in the implementation of an active load for feedback. Whichever route is taken, certainly further investigation is necessary.

Mentioned previously was the possibility of redesigning the entire circuit to alter the feedback as opposed to using the variable resistor. In addition to this, with more time, it may prove beneficial to conduct some experiments under natural lighting outside. It could provide preemptive knowledge of some issues that eventually must be dealt with.

Though the high resistor values providing large gain used here are impractical for the strong impingent laser light the conceptual framework laid in these investigations points to the ultimate design parameters for 2f-WMS signal detection:

- Low capacitance and fast response photodiodes
- Low-noise front-end amplifiers
- A transimpedance configuration
- Gain accommodating 20kHz bandwidth and DC signal suppression
- Detection without background artificial lighting.

In answering these criteria therefore, the Thorlabs® FGA10, with 40pF junction capacitance and 5ns rise-time, was used. Furthermore, the FGA10 InGaAs photodiode has a responsivity of 0.9 mA/mW at 1555 nm from which point it falls off so that at 1.654 μm it is 0.8 mA/mW. Front end AC coupling was exacted while using low noise FETs at the front end. Below, Figure 32, shows the bandwidth performance profile of the photodetection circuit built to these criteria, while Figure 33 shows the SNR profile over several tens of kilohertz.
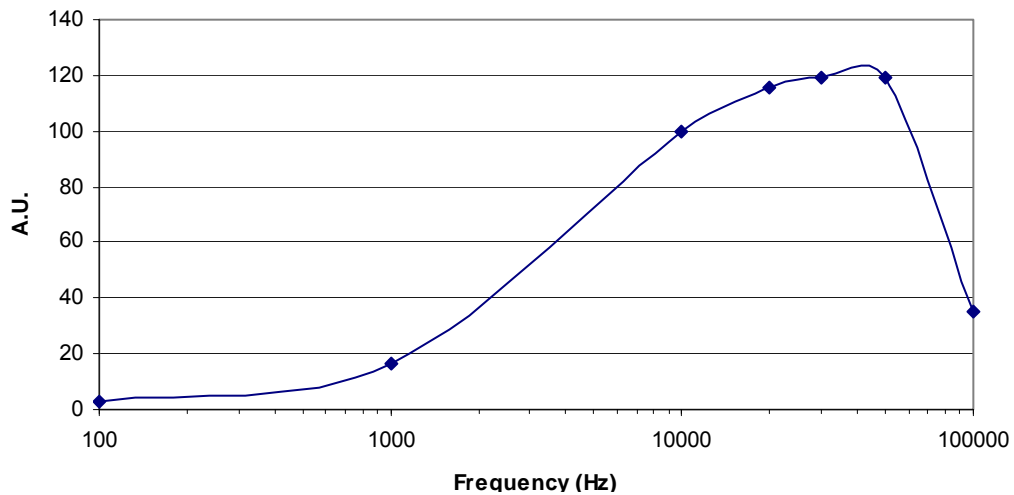
Figure 32. AC/DC PD BW at 1mA bias. Bandwidth profile of transimpedance photodetection circuit with high pass filter. Laboratory measured data.
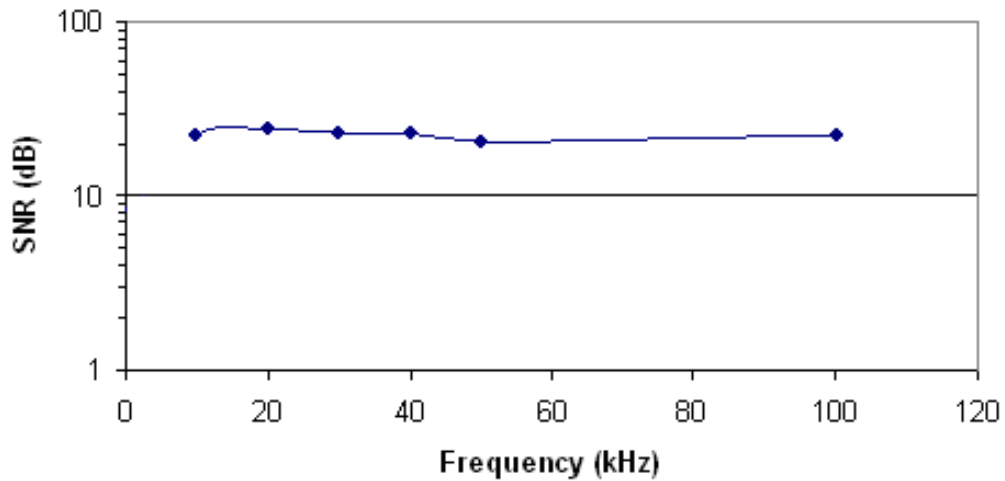
Figure 33. Transimpedene detector SNR. The largely flat transimpedance SNR of the used photodetection circuit. VCSEL-1654 was used in this study. Laboratory measured data.

Though this circuit yielded the greatest ultimate performance concerning 2*f*-WMS detection, the forthcoming optical feedback analysis requirements differed slightly. There the study of degenerate feedback noise across a larger frequency band and particularity at low frequencies saw fit the use of the simple photoconductive detection circuit, represented in Figure 26a. This circuit was tested also whilst shorting the supply bias to ground with the performance SNR shown in Figure 34 below.
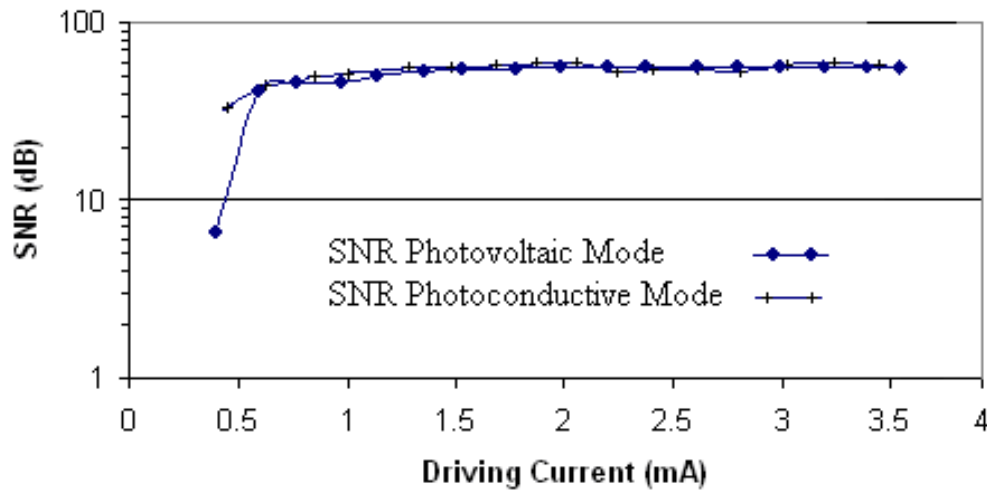
Figure 34. Photocurrent SNR of a simple photoconductive and photovoltaic circuits. Note the almost identical noise profiles over most of the VCSEL-1654 driving range. Laboratory measured data.

The significantly lower SNR of the transimpedence amplifier circuit once more validates its use for methane detection.

This chapter covered the principle factors governing photodetetion circuits. The first topic included an overview of modern-day photodiodes and their governing parameters. Subsequently multiple circuit modes were examined so that an understanding of factors pertaining to performance could be obtained. In spectroscopic gas analysis extremely weak signals mean that limiting noise and designing circuits toward improving SNR is of paramount importance. Noise contributors in both photoconductive and photovoltaic modes were shown and noise theory itself was covered. The amalgamation of all these factors influenced the experiments that were undertaken and performance characteristics evidenced. The shear breadth of the topic of study limited the scope of research conducted; nevertheless, the information presented here served well as an excellent stepping stone for studies in the future which may be more involved.

5. SECOND HARMONIC WAVELENGTH MODULATION SPECTROSCOPY DETECTION

The remedial work conducted thus far permits the focus to fall now on to the detection and observation of methane. The style employed for gas detection using second harmonic wavelength modulation, 2f-WMS, spectroscopy will be digital. Tracing the signal thus far we find an analogue control driving laser emission along with an analogue detection circuit. Signal processing, however, at the receiver end will be wholly digital. Signal analysis, as noted previously, will be executed with a lock in amplifier LIA, allowing weak signals mired in noise to be detected with confidence. Signal acquisition in this manner necessitates synchronization and this too was done digitally with the implementation of a digital phase lock loop, DPLL. The digital system has advantages over analogue systems due to long term accuracy; analogue systems being dogged by aging components producing performance drift, or thermal susceptibilities. The ease by which parameters, such as filter bandwidths for example, are augmented also makes digital systems time-efficient experimentation tools. In this chapter therefore, the means of signal processing are delineated and trace levels of methane are finally detected.

5.2 Lock-in Amplifier design

The preamble to digital signal processing is the obvious consideration of the way in which signals are digitized and its role in affecting signal quality. Due to the limited scope of this thesis work this issue will be bypassed here. It is however noted that a commercial National Instruments® 250kHz 16-bit ADC was used in this capacity having outperformed others tested. The NI sampling rate was sufficiently high given the interest of the 2f signal at 20kHz while the 20 V_{P-P} detection range provided a range complimentary to the amplification provided by the photodetection circuit. Absorption tuning was done with a 50Hz trinagular wave and given the sampling speed of the ADC was 250kHz 5000 samples were obtained at a time, tantamount to obtaining one complete trangular wave period.

As noted previously high-frequency modulation and detection provides a means of shifting the signal to sufficiently high frequencies such that noise powers at low frequencies are avoided. Given this system we can further enhance our acquired SNR by exacting phase sensitive detection. The philosophy of this technique is supported on the basis of the detected range-integrated concentration, RIC, signal being mostly represented at the phase of the modulation signal (this is inferable from Figure 4). By applying synchronous detection, at a moderate cost to system simplicity, it is possible to reduce the in-band noise power as noises in phase-quadriture are parted and dropped from the signal. Such architecture is depicted below in Figure 35.
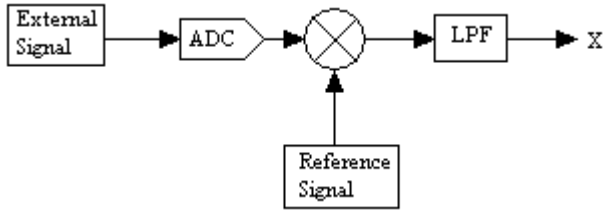
Figure 35. Simple design concept of a lock in amplifier with low pass filter LPF.

For this thesis the entire work of this digital signal analysis was conducted using Labview® 8.0. After the laser emission engages the target gas and is detected the generated and amplified photocurrent is digitized. Mixing with a synchronous reference signal, in our case one of twice the modulation frequency, and low pass filtering, choking off noise and extraneous signal, we obtain the desired $2f$ signal. In practice the difficulty in such a design is underscored by the challenge of providing a synchronous reference signal. Addressing this task is done in the next section with the design of a digital phase-lock loop, DPLL.

5.2.1 Design of a Digital Phase Lock Loop

Lock-in amplifier detection necessitates down conversion of high frequencies to the DC level via synchronous rectification. The reference signal for this task may be obtained from the modulation driving circuit itself, pulled from some other external source, or generated within the lock-in amplifier by some means. In this body of work the latter approach is used with the creation of a synchronous reference signal by way of a digital phase lock loop, DPLL [28]. The archetypical phase lock loop block diagram is shown in Figure 36.
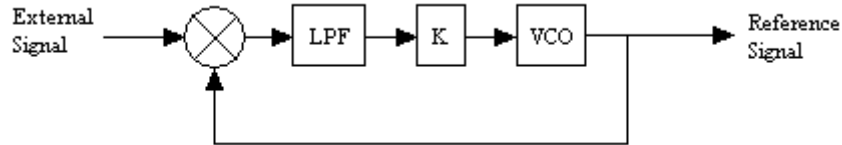
Figure 36. First order PLL. A low pass filter, LPF, loop gain, K, and voltage controlled oscillator, VCO, comprise the signal fed back into the phase-comparator, a mixer. This PLL is sufficient under the small-signal approximation.

This was therefore extended to the LV design given in Figure 37.
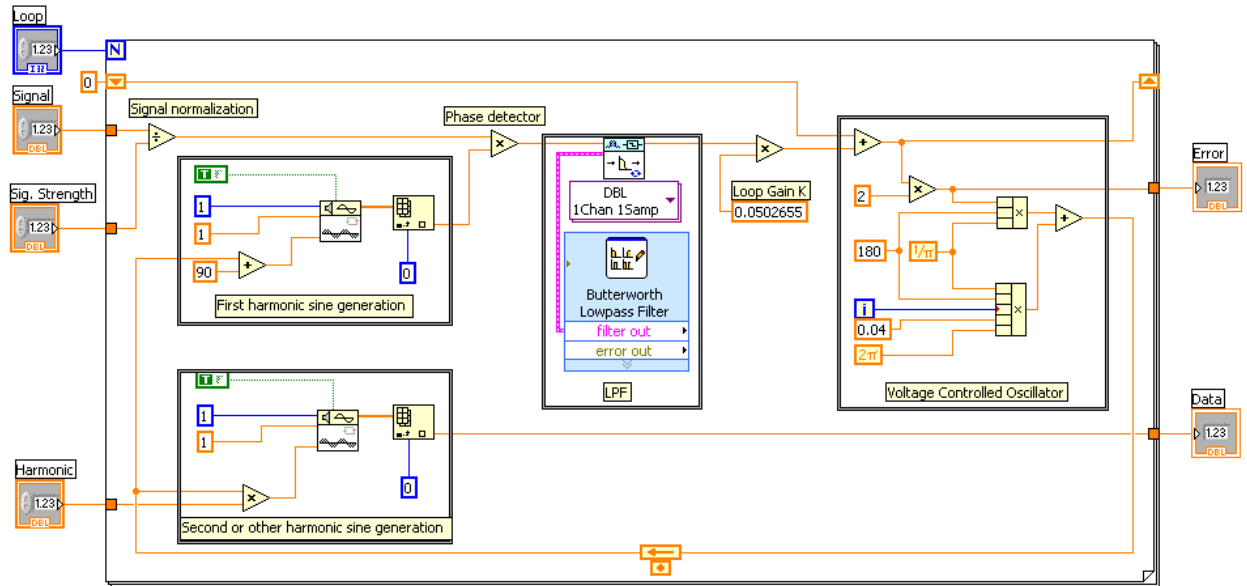
Figure 37. Digital realization of a phase lock loop. Additional input lines like the harmonic output control, error signal and external signal line strength line are seen. The input signal and controllers are along the left while outputs appear down the right side. "Loop" executed the iteration as many times as there are data point. Modeled with Labview.

After mixing the low pass filtering was affected by a second-order, Butterworth, low pass filter with 1.8kHz pass band, consistent with PLL rule of thumb that. Filters of higher order showed to be too slow, while higher band pass filters proved to pass too much noise power. The loop gain, K, was set to 0.0502655. Given a detection signal, tracking error showed not to be more than 0.5 degrees at all times.

With this, the second harmonic synchronous signal generation makes feasible the implementation given in Figure 35. Referencing Figure 37, the line identified as the data line (output line) provides the synchronous reference signal to that of the (input line) signal line, shown in the redrawn block of Figure 38.
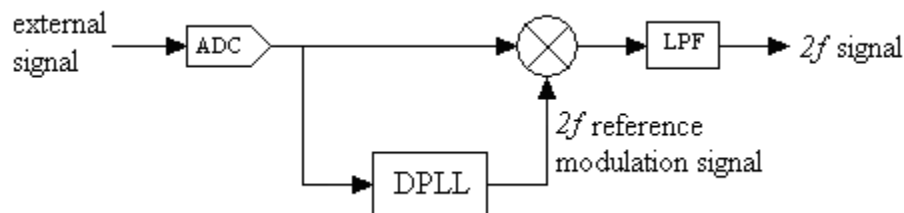
Figure 38. Implementation philosophy for 2f lock-in detection. LPF to be discussed.

All the aspects of the lock-in amplifier have thus far been addressed bar one; the low pass filter. Revisiting second harmonic detection theory the signal lies in band of …. This was made the case by again utilizing the Labview® filter block, implementing an elliptical filter, the specifications of which are listed in table 3.

5.2.1.1 Phase-insensitive and phase-sensitive detection

The lock-in amplifier may further be evolved to yield a kind of phase insensitive R signal. By adding the two quadriture signals a mathematical analysis shows irrelevance of reference phase upon the signal.

Therefore, augmenting the lock-in detector diagram once more to that shown in Figure 39 has alleviated the need for reference phase deduction all together.
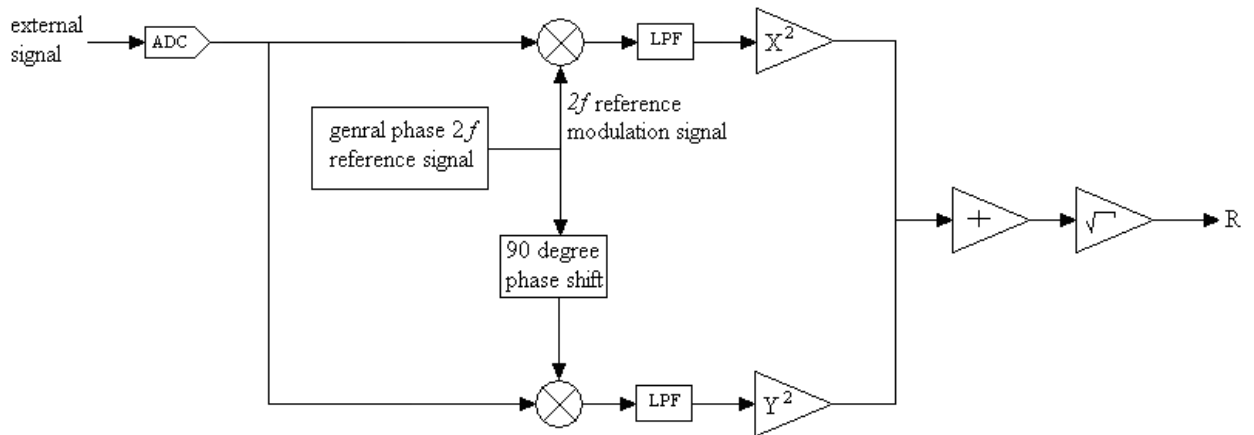
Figure 39. Lock-in amplifier detector.

This does come at a marginal price to performance as Figure 4 shows that the ideally out of phase signal does not nominally reach zero, to say nothing of additional noise. This is moderately offset by the noise that is introduced by the generation of the synchronous reference signal, whereas the reference generated for R signal acquisition is a mathematical construct of an ideal sinusoid at the relevant frequency.

Touching upon one more issue prior to observing the data generated by these implementations, that of signal averaging. This is the most powerful tool that we posses in our arsenal in order to filter our high frequency fluctuations. Without diverting our focus from 2f WMS singal detection I matter-of-factly assert that averaging was implemented such that signal integration time was no larger than 20 seconds. Typically this involved 200 scan averages.

Having put all the previously noted concepts together, Figure 40 shows one such result.
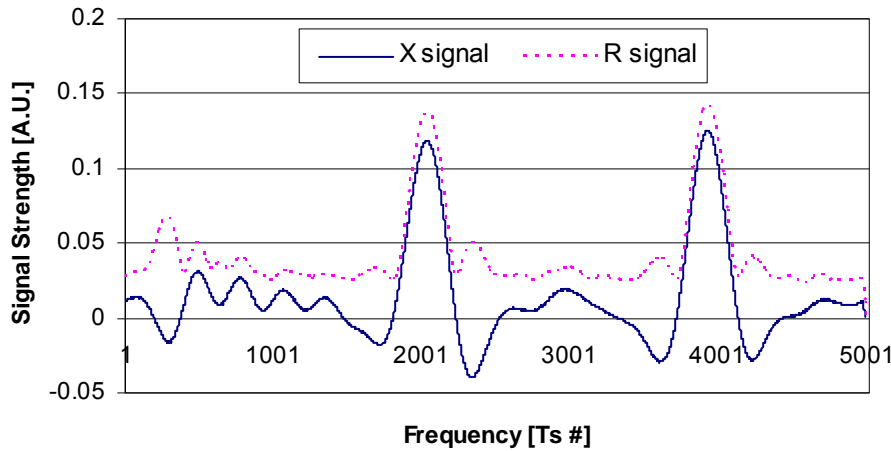
Figure 40. Here both the X signal and R are the result of methane at 50ppm within the holding cell. Tuning frequency 50Hz and modulation 10kHz. Laboratory measured data.

In the above figure the translation of the R signal off of zero points to the point raised earlier concerning the penalty of implementation of R signal lock-in detection. For the signal strengths shown the R signal SNR~5.4dB whilst that of the X signal is 8.1dB. Extrapolation of this result suggests the feasibility of detection of approximately 5.4ppm*m RIC given the R signal SNR at 1, the atmospheric concentration of methane being approximately 1.6ppm.

6. LASER DIODE FEEDBACK OPERATION

Many independent noise sources contribute to the degradation of detected optical signals and so limit the scope of applications. Assuming they are caused by independent forces each noise type in its own right can be studies as a justifiable cause of limited signal quality. The phenomenon of emitted laser light being reintroduced into the laser cavity; a phenomenon termed solely as "feedback" from hereon in, has been identified as one such noise source. This topic is of interest given the portable methane detector operates as an optical transceiver. The focus in this chapter then will be in qualifying inhibitory noise properties of long wavelength vertical cavity surface emitting lasers (VCSELs) in the presence of optical feedback. This subject has been well documented for semiconductor lasers as applies to the telecommunications industry. This work though is meant as the basis for understanding the phenomenon in order to facilitate greater understanding in the use of VCSELs for trace gas detection. To achieve this end an overview of standalone laser noise theory will be presented, followed by feedback noise theory. These then facilitate the knowledge to be used in the qualification of the experimental results.

6.1 Light Injection States

In dealing with feedback it is prudent to properly define the conditions which will be ascribed to the particular experiments undertaken in the course of this research.

There are three different light injection states. In the first of these, the double cavity state, figure 1(a), we assume that the reflected waves maintain a perfect phase correlation between the reintroduced field into the cavity and the principle cavity field present due to stimulated emission. The second state, referred to as the external light injection state of type I does not become the double cavity because we assume that the phase correlation disappears within the feedback roundtrip time, though the reflected waves do form a continuous loop with the cavity field. That is to say both are present within the laser cavity simultaneously as shown in Figure 41 (b-i). The last state is equivalent to an external light injection state but is distinguished from the one above as type II. In this case the laser generates a pulse of sufficiently short duration, indeed shorter than the round trip feedback time, so that only one field source is present within the laser cavity at any one time. Figure 41(b) shows the difference between the two external light injection states of type I and II.
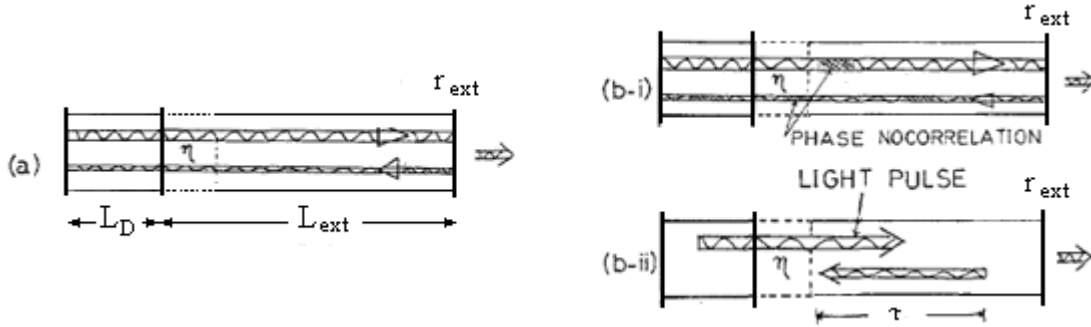
Figure 41. The laser diode cavity is of length L_D, into which a current I is injected. The distance from the emitting laser facet to the feedback reflector rext is Lext with index η; (a) the double cavity state, (b-i) eternal light injection state type I and (b-ii) external light injection state of type II.

As stated above the double cavity state is characterized by the fact that both the pulse length L_P and the coherence length L_C of a laser in the steady state are longer than the loop length $2L_{ext}$ so that the above sates may be expressed by the following relations;

$L_{ext} \leq L_P$ & $L_{ext} \leq L_C$ } double cavity state

$L_{ext} < L_P$ & $L_{ext} > L_C$ } external light injection state type I

$L_{ext} > L_P$ } external light injection state type II

where

$L_P = \tau c/(2\eta),$ & $L_C = c/(2\pi\Delta\upsilon)$

Note that here the coherence and pulse length have been defined with feedback in mind, namely, in the context of the roundtrip distance, which previously has been defined as twice the optical path length from the emitting laser facet to the external reflector.

It must be understood now by the reader that all work that follows belongs specifically to the double cavity state which, as will be seen, is itself comprised of a large breadth of feedback modalities.

6.2 Feedback Effects

Feedback in semiconductor lasers is classified by the effect it has on nominal operation.

By and large feedback of emitted radiation into a semiconductor laser is detrimental to standalone operation. Here the feedback parameters necessary for positive effects such as narrowed linewidth operation will be presented. This section of will also elucidate why, for example, this only applies to vary specific conditions that are difficult to maintain. The feedback region where the linewidth is broadened and an increase in RIN is observed will be focused on since this has the potential of being a principle noise contributor toward the detection of trace gases via tunable diode laser absorption spectroscopy TDLAS.

In [23] Tkach has identified five regimes of operation that are unique and arise due to optical feedback and though his experimental work is done with a DFB laser at 1.5um other lasers too, within a few dB, exhibit the same effects. The important laser parameters affecting feedback sensitivity are the stored energy within the cavity and the coupling of the laser mode to the external field. I apply the notions here for VCSELs too.

Coherence collapse can be brought about quite easily and the negative effects that arise include mode hopping, excess noise, and chaotic behavior. Historically the external reflector is modeled as the reflection off the end of a fiber pigtail as considerations of feedback effects first arose from the telecommunications industry. Nevertheless the developed theory is quite general and does not favor any reflection scheme above another.   Following is an overview of the development of theory that leads to the classification of feedback regimes. It is noted that the theory is developed with weak reflections in mind, a case that wholly applies to VCSELs due to their high emitting facet reflectivity. It is at this point prudent to clearly indicate that the feedback fraction is indicative of the power re-introduced at the laser.

$$f_{ext} = \frac{reflected\ power}{emitted\ power} = R_{ext}$$

Eq. 44

Feedback work is predicated on the cavity roundtrip condition for stand-alone lasers, shown in Figure 42 by the non-shaded region.
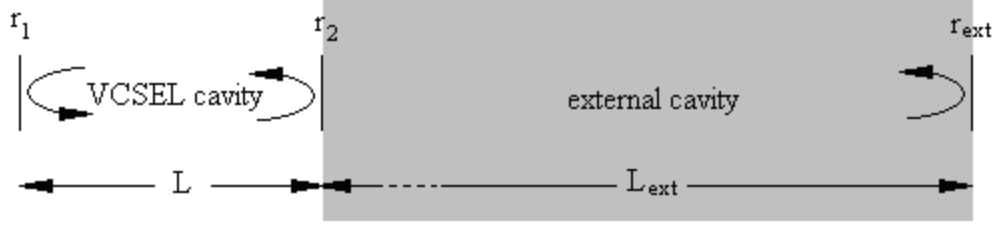
Figure 42. Conceptual depiction of the VCSEL along with an established external cavity.

Simply it is stated as
$$r_1 r_2 \exp\{-2j\beta L + (g - \alpha_s)L\} = 1$$
          Eq. 45
for which the threshold gain condition is
$$g_{th} = \alpha_s + (1/2L)\ln(1/R_1 R_2)$$
          Eq. 46
while the phase condition must satisfy
$$\beta L = m\pi \qquad \{m | m \in Z\}.$$
          Eq. 47
If however feedback is considered, due to the reflected light, the required laser gain changes from that of the standalone laser according to

$$g_c = g_{th} - \frac{\kappa_{ext}}{L}\cos(\phi_{ext})$$
          Eq. 48

Here, g_{th} is the nominal standalone threshold gain, while L and ϕ_{ext} are the cavity length and phase of the reflected light, respectively.   The parameter $\kappa_{ext} = \frac{r_{ext}}{r_2}(1 - |r_2|^2)$ represents the coupling coefficient from the laser to the external cavity [12] with r_{ext} and r_2 representing the effective reflection coefficient and laser facet reflection coefficient, respectively. It is noteworthy to mention here that the reflected polarization is assumed to equal that of the emitted.

Since only slight changes of ϕ_{ext} or the optical frequency v may yield considerable changes in the required threshold gain for lasing according to the above equation a considerable mode hopping may result leading to large mode hopping noise. Normally mode hopping is avoided in single longitudinal mode lasers, like the VCSEL, however the external cavity may introduce additional external cavity modes. This yields possible mode hopping between external cavity modes also for single mode lasers with optical feedback. In deriving the required laser roundtrip phase condition $\Delta\phi_L = m2\pi$, where m is an integer, possible lasing modes are characterized by the threshold gain with feedback equation 46, and taking into account that a change in threshold gain also yields a change in the refractive index via the linewidth enhancement factor alpha

$$\Delta\phi_L = \frac{\tau_L}{\tau_{ext}}[2\pi(\nu - \nu_{th})\tau_{ext} + C\sin(2\pi\nu\tau_{ext} + \arctan\alpha)]$$

Eq. 49

with τ_L indicative of the laser roundtrip time and ν_{th} the lasing frequency, both for the standalone laser, [10]. C is a feedback parameter evidenced in [24] and it characterizes feedback strength, phase, roundtrip delay, and re-penetration into the laser cavity. It is given as follows

$$C = \kappa_{ext}\frac{\tau_{ext}}{\tau_L}\sqrt{1+\alpha^2}$$

Eq. 50

Because C encompasses the gamut of feedback relevant parameters it is by-and-large used as the principal parameter describing feedback. Though equation 49 is derived for Fabry-Perot-type lasers it qualitatively applies for VCSELs too given the simplified model with effective reflectivites and length (section **VCSEL GENERALS**). One must however determine C, which is system dependant. Below, Figure 43 shows the roundtrip phase change $\Delta\phi_L$ as a function of the oscillation frequency ν.
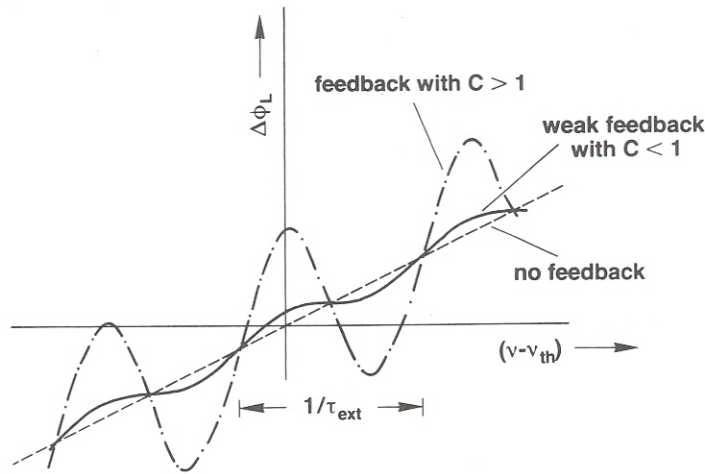
Figure 43. Change in phase versus oscillation frequency. Dashed line is that of no feedback present, solid is with minimal feedback, while the dotted line represents feedback of reasonable strength.

Possible emission frequencies, where the phase condition is satisfied, are characterized by $\Delta\phi_L = 0$. Without feedback, as represented by the dashed curve, $\Delta\phi_L$ varies linearly with ν. Now, as long as C is very small $\Delta\phi_L$ still monotonically increases with ν and the phase condition is still only met for a single frequency, usually only slightly adjusted from the nominal emission frequency. For larger values of C, namely when C>1, the $\Delta\phi_L$ vs. ν profile undergoes strong oscillations whereby multiple $\Delta\phi_L = 0$ phase conditions are met. In this case several modes about ν_{th} may oscillate. Actually, higher values

of C may still yield a single zero for $\Delta\phi_L$ under the right conditions for ϕ_{ext}. Below, Figure 44 shows the $\Delta\phi_L$ vs. $\Delta\nu$ plot with C greater than one.
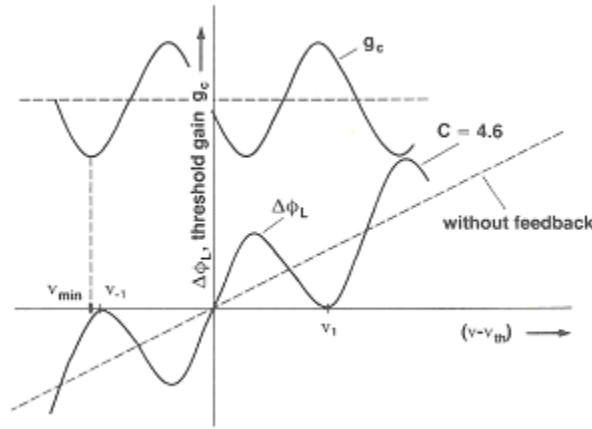
Figure 44. Round trip phase change and threshold gain g_c with the maximum feedback (C=3π/2) for maintaining a single external cavity mode.

In order to allow for a large feedback coefficient C, the feedback phase must be adjusted so that the roundtrip phase has odd symmetry with respect to $\Delta\nu$, as above. From the limits of Figure 44 we can approximate C<3π/2. So, with adjusted phase feedback C<3π/2 can be tolerated without introducing additional external cavity modes. Note that the ν_{-1} and ν_1 modes do not correspond to the minimum threshold gain which occurs for $\phi_{ext} = m2\pi$. Since ν_{min} lies close to ν_{-1}, a zero for $\Delta\phi_L$ for C>3π/2, instabilities may result.

Therefore, single mode conditions may be summed up as follows; for weak feedback, that is C<1, single mode conditions naturally arise while stronger feedback levels, up to C<3π/2, can assume single mode conditions too if the phase is carefully adjusted.

6.2.1 Effects on Spectral Linewidth

When feedback is introduced the nominal emission linewidth can change drastically. The spectral linewidth is derived from $\Delta\phi_L$ equation 29 as

$$\Delta\nu \sim [\frac{d}{d\nu}(\Delta\phi_L)]^{-2} \; at \; \Delta\phi_L = 0$$

Eq. 51

resulting in

$$\Delta\nu = \frac{\Delta\nu_0}{[1 + C\cos(\phi_{ext} + \arctan\alpha)]^2}$$

Eq. 52

here $\Delta\nu_0$ is the linewidth of the standalone laser.

From equation 32 above it is evident that the linewidth is bounded by the minimum and maximum linewidths as follows;

56

$$\Delta\nu = \frac{\Delta\nu_0}{[1+C]^2} \qquad \text{Eq. 53(a),} \qquad \Delta\nu = \frac{\Delta\nu_0}{[1-C]^2} \qquad \text{Eq.}$$

53(b)

For C≈1 it appears as though an "infinite" linewidth is predicted. So even for very low feedback levels a substantially large linewidth broadening can arise due to feedback. For C>1 a splitting occurs, as covered previously, and two external cavity modes arise. Here the splitting from the single external cavity mode to the dual mode is accompanied with considerable phase noise and intensity noise. With increasing feedback the modes split further, to a maximum of $1/\tau_{ext}$, upon where the cut-off frequency of the mode hopping noise decreases till one mode only survives. Surprisingly, as elucidated in [24], for measured spectral linewidth for laser diodes with feedback coefficient C>>1 the external cavity mode with the narrowest linewidth is more stable than one for which the threshold gain is lowest. In the coherence collapse regime though, severe instabilities may occur due to the interaction between these modes.

6.3 Regime Classification

The above work enables us to now pool all the above effects and classify feedback types qualitatively. These regimes provide an excellent understanding of feedback from the view of how laser operation is altered and completes the definition of feedback, along with the previously presented states of feedback light.

Regime I: This represents the lowest levels of feedback, feedback parameter C<1, and can result in a narrowing or broadening of the emission line. This effect is dependent upon the phase of the feedback. In his work Tkach finds that a meager power feedback of -80dB can manifest in a 30% effect on the emission linewidth. Figure 45 shows the regimes' linewidth possibilities versus feedback.
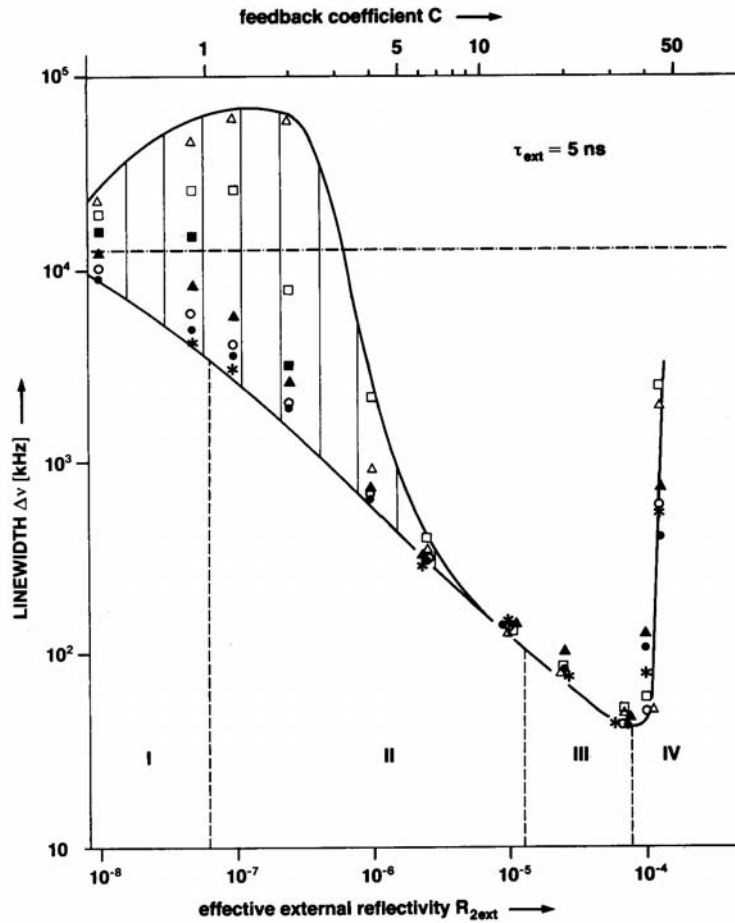
Figure 45. Linewidth versus.feedback for semiconductor lasers and the corresponding feedback regimes I to IV [25].

Regime II: At feedback levels above those of the previous regime the distance to the external reflector as well as the feedback quantity is relevant. Line broadening, which is observed at the lowest levels for out of phase feedback, changes to an apparent line splitting arising from rapid mode hopping. This amount of splitting is directly dependent upon the external reflector proximity.

Regime III: At stronger feedback levels yet, the reflector distance is inconsequential. This regime only occupies a very small range of feedback intensities and the mode hopping from II is suppressed, the laser now operating in a single, narrowed line fashion. Surprisingly, it is the mode from Regime II with the minimum linewidth rather than the mode with minimum gain that becomes the predominant lasing mode [23]. The narrow range of feedback levels which this regime encompasses makes it vary unstable.

Regime IV: Further increases in feedback levels now results in satellite modes, separated from the main lobe by the relaxation oscillation frequency. These grow as the feedback increases and laser line broadens to as much as 50GHz, [23]. It is this regime that is commonly referred to as the coherence collapse regime due to the extreme reduction in the laser's coherence length, to as little as a few

centimeters as reported by some. In figure 6 a drastic increase in linewidth commences at approximately $R_{2ext}=10^{-4}$, coinciding with the onset of coherence collapse. Also, feedback phase is inconsequential in regime IV. Accompanying the increase in linewidth is a substantial increase in relative intensity noise (RIN). Because of the onset of substantial noise this regime will further be explored in following sections.

Regime V: In the extremely high feedback levels of this regime the laser operates as an extended cavity where one can image the long single cavity with a short active region. For the higher than -10dB feedback levels required antireflective coatings are necessary on the coupled laser facet. With VCSELs this regime is not typically observed (to my knowledge).

These five regimes are very well defined and the transitions from one to the other are easily identifiable. Tkach presents experimental measurements of all five regimes in [23] and in Figure 46 below we can observe each regime as a function of feedback power and reflector distance.
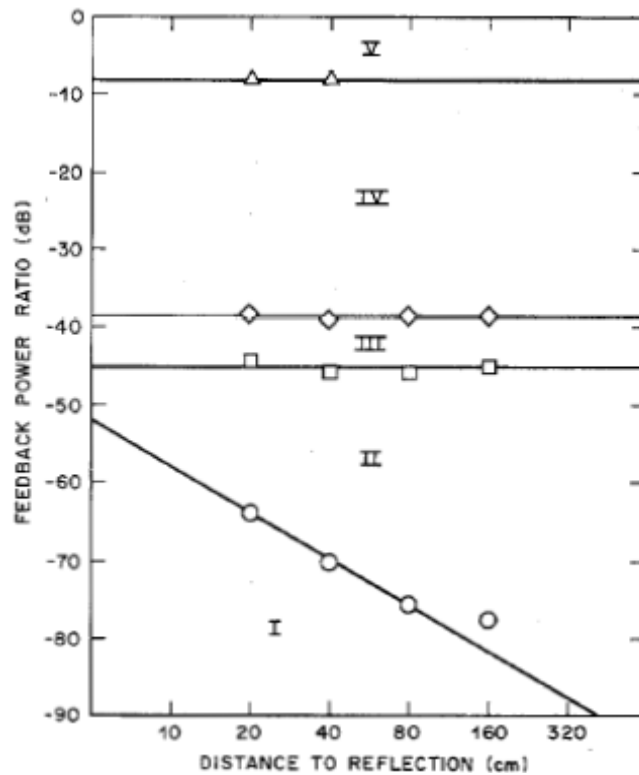
Figure 46. Relationship showing regime delineation as functions of feedback power and distance to reflection.

6.4 Feedback Dynamics

With the background of the previous section the focus now shifts to the dynamic behavior of semiconductor lasers under the influence of feedback. This topic requires analysis of the complex time-dependant field equation within laser diodes. First appearing in [14] the field equation with feedback is given as

$$\frac{d}{dt}E(t) = \left[j(\omega - \omega_{th}) + \frac{1}{2\tau_{ph}}(G-1) \right]E(t) + \frac{1}{\tau_L}\kappa_{ext}E(t-\tau_{ext})e^{-j\omega_{th}\tau_{th}} + E_{sp}(t)$$

Eq. 54

here, ω corresponds to the oscillation frequency, ω_{th} that frequency of oscillation at threshold, and G is the gain ratio g/g_{th}, all of which represent the standalone laser parameters. κ_{ext} as previously, represents the coupling coefficient from the laser diode to the external cavity while $E_{sp}(t)$ as previously represents the complex field amplitude due to spontaneous emission into the lasing mode. As expected the diode field equation with feedback (Eq. 34) collapses to (Eq. 35) when the feedback term is eliminated. Here the slowly varying complex field amplitude as before is written with respect to amplitude and phase as

$$E(t) = \sqrt{S(t)}e^{j\phi(t)}$$

Eq. 55

with $S(t)$ denoting the photon number within the laser diode itself (without the external cavity), yielding for $dS/dt=2Re(E^*(t)dE/dt)$ and $d\phi/dt=Im(E^*(t)dE/dt)/S(t)$. As a result of inserting equation 34 into equation 35 the following relations are given

$$\frac{dS}{dt} = \frac{S(G-1)}{\tau_{ph}} + K_{tot}R_{sp} + \frac{2\kappa_{ext}}{\tau_L}\sqrt{S(t)S(t-\tau_{ext})}\cos\left[\omega_{th}\tau_{ext} + \phi(t) - \phi(t-\tau_{ext})\right] + F_S(t)$$

Eq. 56

$$\frac{d\phi}{dt} = \frac{1}{2}\alpha\upsilon_g\frac{\partial g}{\partial n}(n-n_{th}) + \frac{1}{\tau_L}\kappa_{ext}\frac{\sqrt{S(t-\tau_{ext})}}{\sqrt{S(t)}}\sin\left[\omega_{th}\tau_{ext} + \phi(t) - \phi(t-\tau_{ext})\right] + F_\phi(t)$$

Eq. 57

Those for the intensity and phase are presented with the rate equation for the carrier density

$$\frac{dn}{dt} = \frac{I-I_{th}}{eV} - \frac{1}{\tau_e}(n-n_{th}) - \frac{GS}{V\tau_{ph}} + F_n(t)$$

Eq. 58

These three form the basis of the subsequent analysis for laser noise and dynamic behavior. Note the collapse of (Eq. 56), (Eq. 57), and (Eq. 58) to the standalone laser counterparts (Eq. 20), (Eq. 21), and (Eq. 22) if the feedback effects are eliminated. It can now be stated that it is in fact the numerical solution of (Eq. 36), (Eq. 37), and (Eq. 38) that was used to render Figure 45.

6.5 Intensity Noise with Laser Diode Feedback

The following is a consideration of single-mode like laser diodes with optical feedback. Due to the complexity of these coupled non-linear equations they have been solved numerically or with simplifying assumptions as in [25]..

60
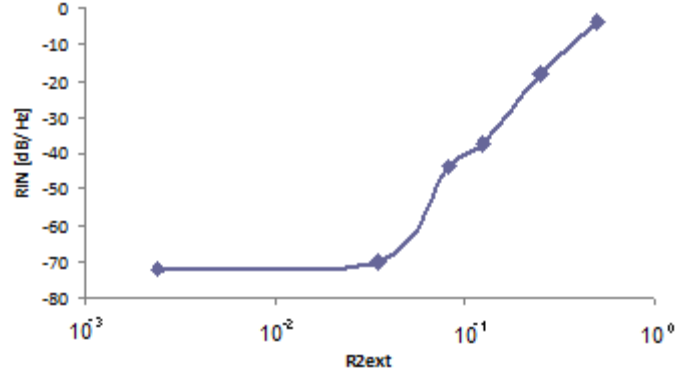
Figure 47. Measured VCSEL RIN. Laboratory measured data.

In Figure 47 above we observe that for $R_{2ext} < 1.5 \cdot 10^{-1}$ the RIN remains relatively low, whereas it increases rapidly afterward. This corresponds to the transition from regime III to IV. From Figure 46 we know that this transition is independent of external cavity length, which applies as long as the feedback coefficient C is large enough. Experiment, therefore, validates this for VCSEL behavior as well. The data from the figure above further strengthens the assumption of VCSEL propriety considering equation 54.

6.6 Feedback Tolerance

From all the analysis provided thus far it is evident that for a single mode laser diode the most critical feedback limit for low noise operation is represented by transition from regime III to regime IV, the coherence-collapse. In designing laser diodes to have low susceptibility to optical feedback therefore it is often desired to shift the coherence-collapse to feedback levels as high as possible.

From the solutions to Eq. 56, Eq. 57, and Eq. 58 several parameters has been shown to have just this effect, the first of these being the laser diode cavity length, L, itself [29].

This effect is borne directly by the κ_{ext}/τ_L term within equations equation 56 and equation 57. Likewise, the decrease of the κ_{ext}/τ_L term can be accomplished by the decrease of κ_{ext} directly by increasing the emission facet reflection coefficient. As a note the BTJ-VCSEL has an extremely high emission facet reflectivity ~1 extremely important in offsetting the short cavity length of several microns. This is evidentiary in the high coherence collapse point of the VCSEL seen in Figure 49.

Intuitively another parameter that correlates strongly to the coherence-collapse regime is the optical power which the laser diode is operated at. Increasing the bias emission power shifts the onset of coherence collapse to higher feedback levels [12].

It is by a careful study of underlying phenomena and physical principles that the operational nature of laser diode feedback effects have been presented. The

theoretical background was laid by first the study of standalone laser operation and noises. Serving as a preamble then the idea of optical feedback was applied as an additive phenomenon. The theoretical study then was undertaken whereby this condition was treated as an operational state in its own right. This led first to the study of the deviation of laser linewidth spectral profiles from those predicted by nominal laser diode operation. Next the classification of five regimes, of these the most being the coherence-collapse, the transition between regimes III and IV was presented. In all, this has led to the recognition of operative feedback levels, below those leading to coherence-collapse, as introducing minimally detrimental noises.

6.7 Feedback Experimental Setup

In attempting to study the phenomenon of feedback one has to be careful in orchestrating an effective experimental setup. Several methodologies exist toward this end, the one which was finally settled upon serves as a reasonable compromise between ease of execution, cost, and detection sensitivities. Of practical importance is to feedback appropriate light quantities into the laser diode and coincidentally detect the emitted light so as to ascertain performance variances in lieu of feedback. Below, Figure 48 shows pictorially how this was exacted.
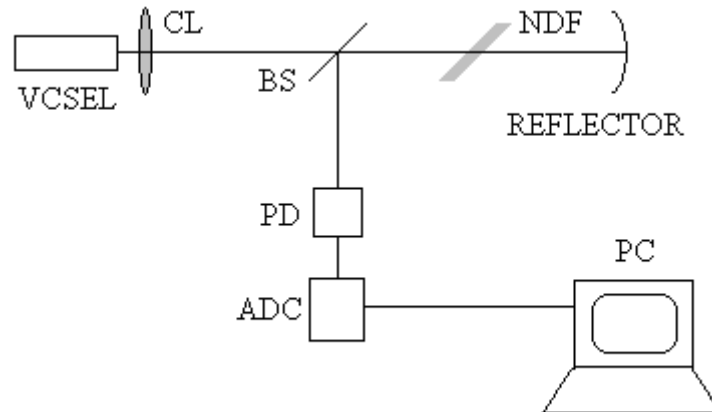
Figure 48. Simplified schematic of the experimental setup. A 50/50 beam splitter was used as a means of coupling some light toward a photodetector. VCSEL-1645.

Varying neutral density filters were placed in the optical path, 1.5m, along with the ability to steer the reflected light by varying the mirror angle were practical means of controlling feedback levels. These combined effects were used to mimic varying reflection levels. With this architecture the practical realization of the theoretical model of laser feedback shown originally in figure 1 has been exacted. The photodetection scheme consisted of a Thorlabs photodiode, model FGA10, in the photoconductive mode shown in Figure 49. The photodetector signal was digitized using the National Instruments PCI-6221, M series DAC.
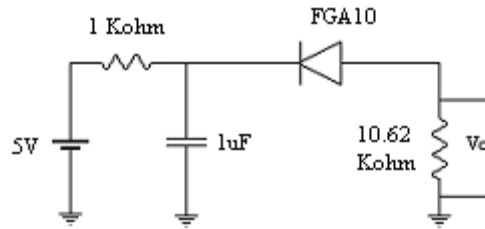
Figure 49. Photodetector Circuit.

In order to put the results we are to analyze here in the right context the free-running laser/detector noise parameters need reviewing. In the section headed "Standalone laser diode operation" the principles of noise were overviewed so the practical specifications are given without much ado.

First shown is the simple association between laser injection current and noise power. In Figure 50 below the general trend shown is for a linear relation between applied current and noise.
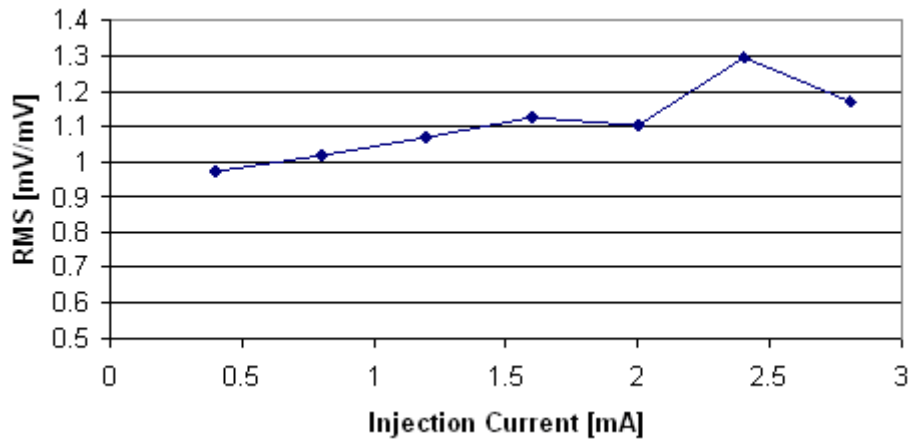
Figure 50. Free-running RMS noise vs. injection current. Operational noise rms normalized relative to the measured dark photodetection noise levels. The time of detection is 50ms. V-1654#1. Laboratory measured data.

One of the most fundamental parameters that can simply be given for feedback systems is the practical measure of noise with said optical feedback. Figure 51 shows this with variances in laser diode injection current.
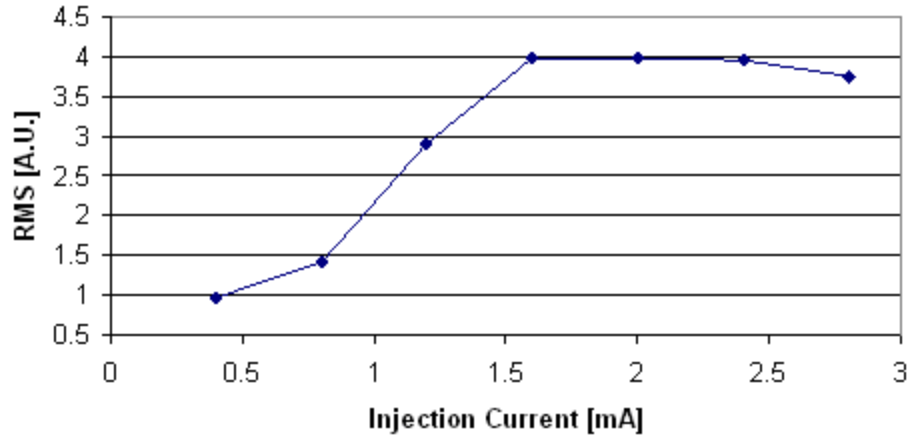
Figure 51. Weak feedback noise RMS vs. injection current. Noise rms detected over 50 ms with "weak" feedback denoting ~10% of maximum attained feedback levels. This does not imply fext as defined by equation (f1) is equal to 0.1. Laboratory measured data.

The figure shows that noise plateaus for the VCSEL at approximately 1.6mA injection current. Note, we can infer that RIN therefore decreases, or conversely SNR increases, for injection currents above this 1.6mA cusp as signal RMS continues increasing while noise does not. Naturally this leads into the study of noise levels at different feedback strengths. To this end Figure 52 below show the noise strength for three different feedback strengths.
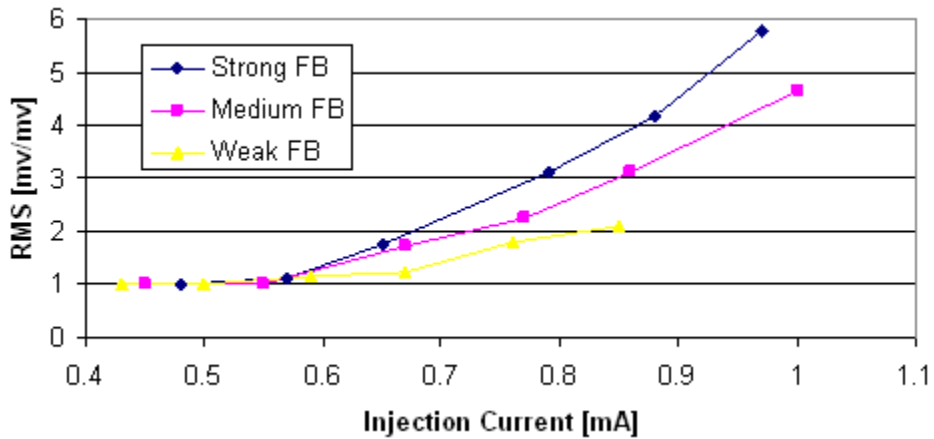
Figure 52. Noise RMS with different feedback levels. Low injection current laser feedback noise at 3 different feedback levels; weak, ~10%, medium, ~50%, and strong amounts, ~100%, relative to the maximum obtainable feedback levels. Laboratory measured data.

From this view we may observe several trends. First is that noise strength increases non-linearly with increasing injection current at all feedback levels. This is in contrast to precisely that for standalone laser diode operation. Second, at these relatively high feedback levels noise increases linearly in strength between feedback strengths' at all current levels. These trends can only be expounded as far as the evidence shows up to a little more than twice the threshold current. Ith~.4mA.

Observing noise power levels though useful do not provide the complete picture, as has been documented within this report. Contextualizing noise levels through relative intensity noise, RIN, is the crux of studying the damaging effects of noises after all.
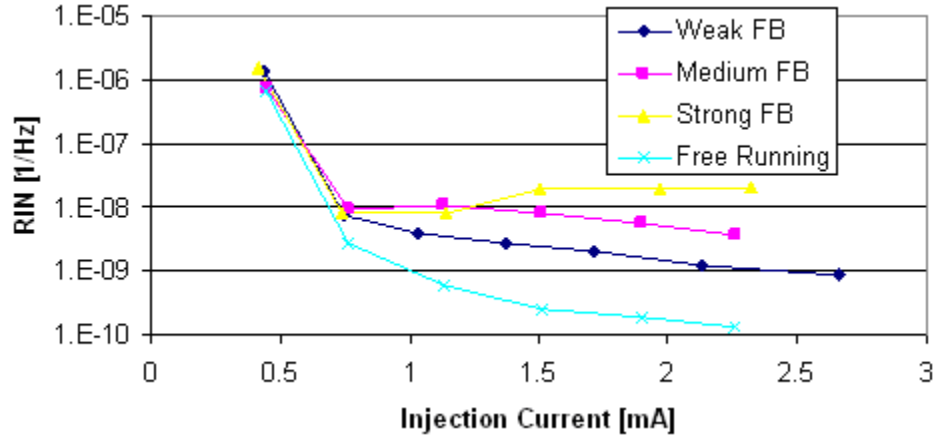
Figure 53. RIN for various feedback levels. Measured RINs with 20 kHz bandwith. Laboratory measured data.

In the above figure we note the relative constancy on RIN levels for all feedback levels beyond a single driving injection current level of approximately 0.7mA. Within Figure 53 the free running laser RIN is shown too, and as theoretically expected these RIN levels fall with greater laser optical power. In fact, for weak feedback levels we can see the both trends marginally, operation in which RIN falls or conversely levels off.

When studying feedback the differences in noise levels compared to those in normal laser diode operation are most evident with high feedback levels. The practical difficulty of measuring the coupled fed back power f_{ext} forces us to rely on experimental verification in gauging whether the VCSEL can be forced into the coherence-collapse regime. Coherence-collapse would catastrophically compromise gas detection. A lack of a drastic increase in RIN therefore implies that even under the strongest feedback levels that I could apply I saw no such evidence. Even so, the subtle increases in RIN noted presented above could be enough to severely limit the detection limit of trace methane. This detriment is now shown with the presentation of 2f acquired signals.

6.8 Spectroscopy with Applied Feedback

Adding traction to the importance of feedback noise in the context of tunable diode laser absorption spectroscopy, several signals with altering feedback levels are presented here. The experimental setup in acquiring data presented for this section is as that shown in Figure 48 with a cell containing methane added between the beam splitter and photodetector, nothing more. The methane gas detection quantity used was 25 ppm·m.

Shown first in Figure 54 below is the detected absorption signal under standard conditions, that is, no feedback.
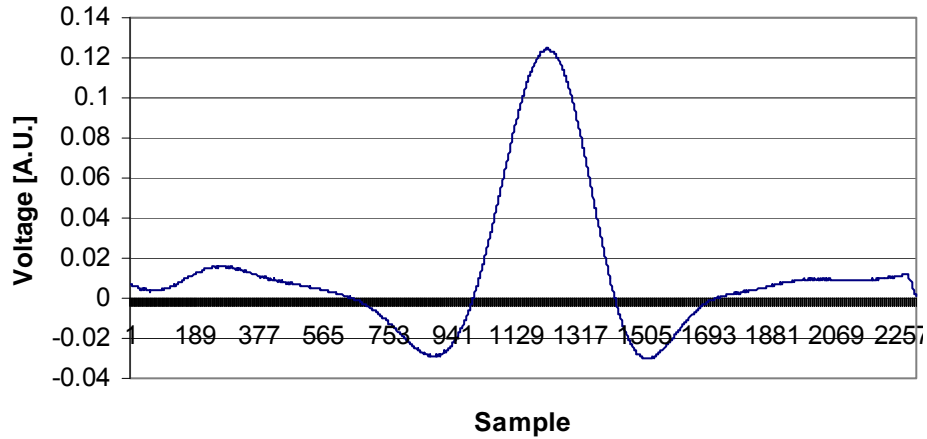
Figure 54. 2*f* absorption signal of 25 ppmm. Nominal 2*f* absorption signal. The signal was obtained with in phase lock-in detection, averaged 50 times. Further details of this technique are expounded in [30]. Laboratory measured data.

From Figure 54 we note the SNR to be approximately 10. The Figure 55 on the next page shows the detection signal again but with feedback applied to the laser diode.
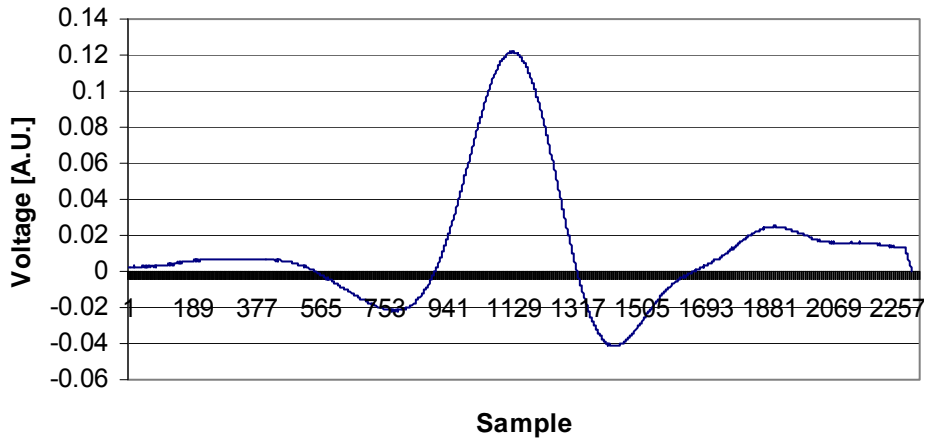
Figure 55. Absorption signal with VCSEL subjected to weak feedback. Laboratory measured data.

In Figure 55 weak feedback denotes precisely 0.0025 of the maximum attainable feedback. The acquired signal in A2 is done under the same conditions as for no feedback, Figure 54. Easily noticeable are the slight degradation of SNR, and the asymmetry of the lower lobes. Tougher to notice is that the entire signal has shifted leftward, a shift of the injection current in fact. This can be correlated to the higher photon count within laser cavities subject to feedback. Finally, Figure 56 shows the 2f signal under high feedback conditions.
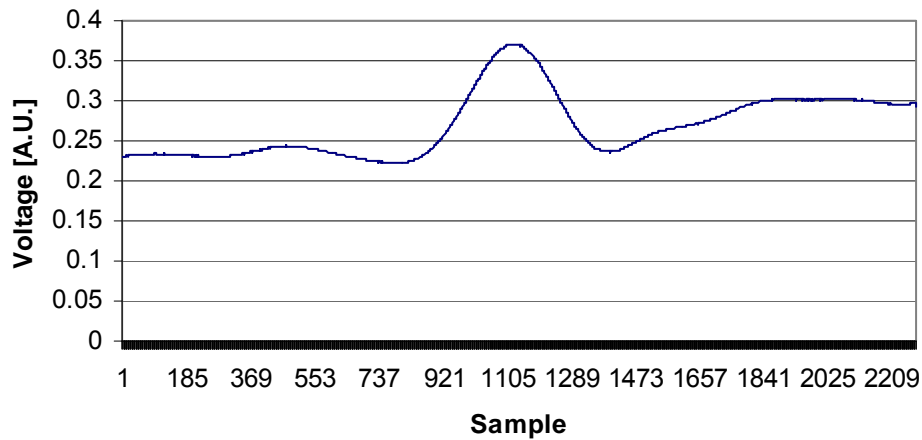
Figure 56. 2*f* absorption with strong feedback. Strong here indicates 0.5 of the maximum feedback strength. Laboratory measured data.

Evidently the SNR has been compromised greatly at this strong feedback level. This is embodied predominantly in the strong shift upward, and hence, diminishing the confidence in any baseline. High noise levels at the 2f frequency may be attributed to this high DC offset.

6.9 Irregularity Highlight

An oscillation of the laser in two different modes, at the highest feedback levels, is apparent as a steady injection current is supplied, Figure 57. This oscillation may certainly be labeled as noise, and if done so would be magnificently larger than noise levels observed in, say, Figure 52 & Figure 53.
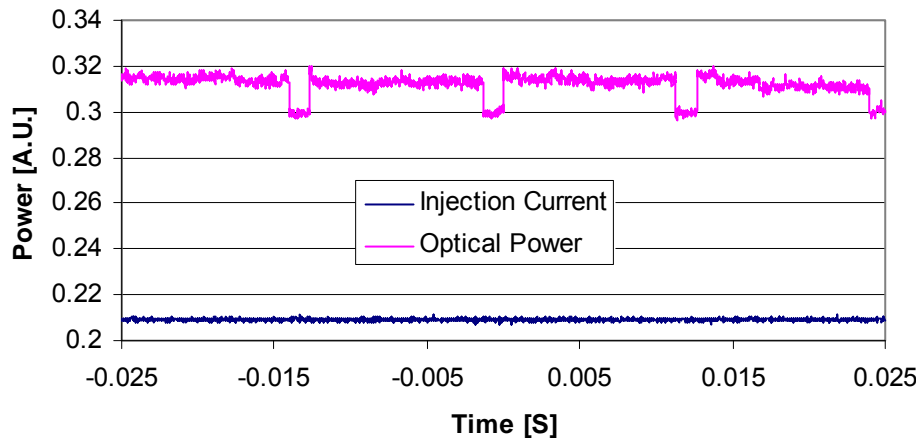
Figure 57. Temporal behavior showing strong noise levels for strong feedback levels. Strong Feedback with 2.09 bias injection current. Indeed this is the feedback level labeled as "strong" in previous figures. Laboratory measured data.

Below, in Figure 58, the same as above is shown but as the title implies the input data has been averaged 16 times to dispel any notion of data irregularity. That is

to say that in some sense the VCSEL is operating deterministically or, "regularly" at this feedback level. No known mechanism for this effect is known to me at this time. The periodicity of the data in Figure 58 is 12.65 mili-seconds with a duty cycle of 10%. The settled levels are approximately 3.122 and 3.012 for top and bottom, respectively, a difference of 0.11. By passing the BTJ-VCSEL's emission through an etalon I was able to correlated this power variance to a frequency span $\Delta\nu \approx 30GHz$. Though the relaxation frequency of the BTJ-VCSEL is not explicitly known to me it would certainly be in the order of gigahertz. This is certainly a point of interest that may be investigated in future studies. Characteristic dynamic laser diode settling can also be noted clearly at both power levels.
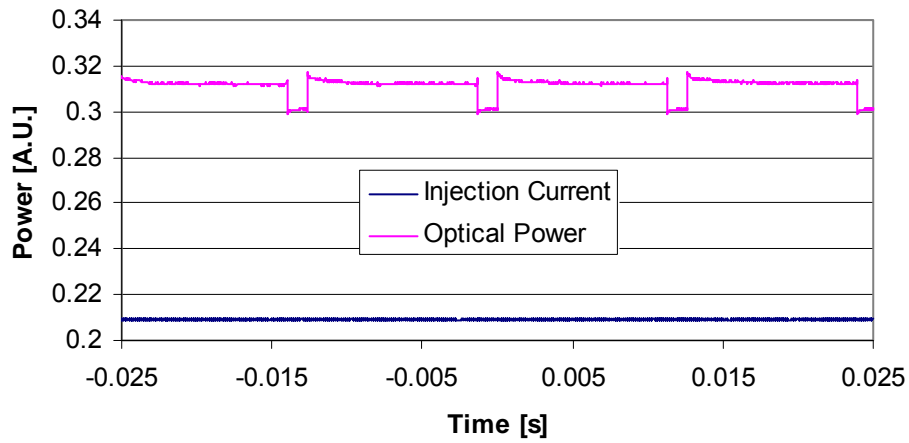
Figure 58. Time averaged signal with strong feedback with input 16x averaged. Similar to figure 18 but time averaged 16 times. Laboratory measured data.

Validation at another bias level was obtained of the same type of behavior and is represented in Figure 59.
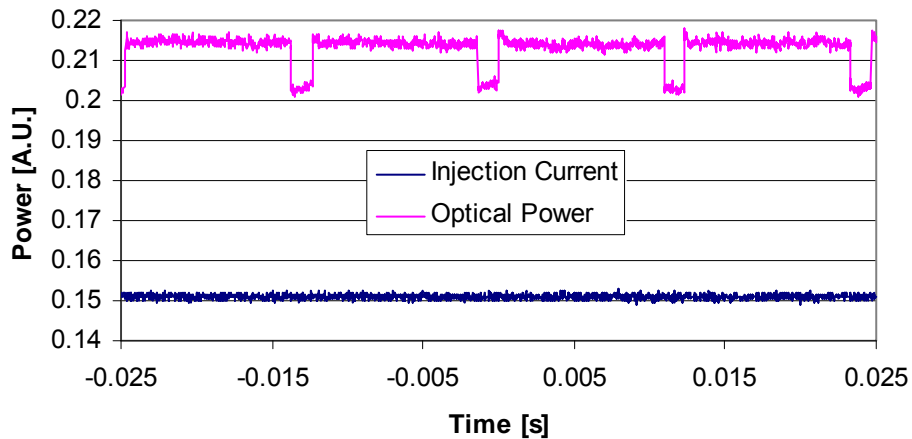
Figure 59. Time average signal with strong Feedback at 1.51 mA bias. Here the repletion rate is 12.31 mili-seconds and the duty cycle is 11%. Laboratory measured data.

At all other feedback levels these effects were not noticed and points toward the poor feedback resolution attainable with the experimental setup used. Furthermore, in order to strengthen the notion that the VCSEL was indeed driven into the coherence-collapse regime at only this strong feedback level optical path distances were varied. This was done to highlight the inconsequence of feedback phase. These types of measurements were made at varying distances and no difference was obvious. This would seem to imply that at the strongest of feedback levels we are only able to drive the VCSEL into regime II, nearer regime III. Noted earlier, there multi-mode operation is expected, with effectively no dependence on distance. From figure 8 we can also note that there, $\sim10^{-5}$, the RIN is in fact maximum, barring the coherence-collapse regime. Though the systematic noise shown above is undocumented at this time, further study an awareness of this phenomenon may even lead to pulsed VCSEL designs.

The experimental data collected shows that undeniably feedback does play a role in increasing VCSEL noise levels. In assessing the detriment of feedback noise several trends have been shown. It is evident that relative intensity noise levels do not fall as one drives VCSELS at levels rising further from threshold. Noise powers in fact rise in tandem with optical power resulting in fairly constant RIN. The span of noise RINs is in fact small though at least one order of magnitude greater than noise levels encountered from nominal laser operation. Coherence-collapse regime operation was attained for the experimental VCSEL but conditions tailored to this effect were highly orchestrated, even by experimental standards including a beam aperture of at least 99%. The alignment sensitivity was so great that a mere tap of the optical table could possibly be enough to misalign the laser/beam splitter/photodetector. Practically, therefore, this is not an easily realizable state. It remains a task to potentially carry out further feedback experimentation with greater sensitivity so that some statistical data may be obtained as well. Requirements for such an endeavor necessitate equipment with greater sensitivity and particularly higher operational speeds.

7. CONCLUSION

The work within this thesis was all conducted with an overview to minimizing the detection limit achievable with an in-situ trace methane detector. An appropriate absorption philosophy was tailored with the VCSEL semiconductor laser diode tabled as an appropriate light source. Performance highlights and limitations of the novel long wavelength VCSEL were examined to show the suitability toward this task. Two primary features make the VCSEL a tangible choice. Firstly, the VCSEL was shown to exhibit excellent fast tuning responsiveness and dynamic operation. Second, the semiconductor VCSEL exhibits a wide single-mode tuning range that is near-linear in frequency output to the driving laser current. This feature enables simplified laser modulation. Additionally, the high cavity quality ensures low power consumption and the inherent VCSEL design enables on wafer testing resulting in the low cost of the laser.

In introduction standalone laser diode operation was reviewed. This was particularly done with forethought on noise in this nominal state of operation. The relevance as applies toward VCSELs was maintained by remaining conceptually general in development of the theoretical modeling. Intensity noise was identified as the source of relative intensity noise while frequency noise was shown to cause the finite linewidth of laser emission. As regards optical laser noise, a general approach was applied toward the laser diode models subjected to optical feedback. This led to the clear delineation of five feedback regimes and aided in ascribing tolerable feedback levels. Via experiment noise trends were shown to comply with those predicted by theory and validate the modeling philosophy insofar as the BTJ-VCSEL is concerned. It was shown that though small levels of feedback degraded signal levels, significant feedback quantities pose the greatest concern as the laser is driven to operate in the coherence collapse regime. Given coherence collapse noise levels increase exponentially with increasing levels of feedback. Also demonstrated with high feedback levels was the ability to drive the BTJ-VCSEL into a as of yet explained pulsing optical output.

For $2f$-WMS signal detection the photodetection methods studied showed a simple reverse biased circuit to posses the best combination of desirable characteristics. The simple circuit met the need for linear responsivity over a wide breadth of detected light irradiances, had sufficient response time, enough signal gain, and limited noise. With this a digital system was coupled for $2f$-WMS signal processing. The digital system has advantages over analogue systems due to long term accuracy; analogue systems being dogged by aging components producing performance drift, or thermal susceptibilities. The ease by which parameters, such as filter bandwidths for example, are augmented also makes digital systems time-efficient experimentation tools.

In conclusion, TDLAS is a powerful technique which lends itself to efficient, real time, remote atmospheric sensing, with the continual development of laser

technology driving development of absorption spectroscopy. I have conducted an extensive literature review in order to identify relevant noise sources of laser emission, detection, and interference optics. With the study of noise I was able to gather enough data to yield supportive evidence of the greatest inhibitors of trace gas detection signal SNR. The study of isolated noise phenomena allowed me to devise relevant suppressants of noise and achieve detection sensitivities with low power VCSELs.

Open path applications introduce further noise phenomena that must be considered when employing VCSELs for tunable diode laser absorption spectroscopy. With the study of these additional noise phenomena and with iterative improvements of TDLAS trace gas detection setups I have laid the groundwork for the extension of study to open path systems.

8. BIBLIOGRAPHY

[1] S. Schilt, L. Thevenaz, and P. Robert, "Wavelength modulation spectroscopy: combined frequency and intensity laser modulation," Applied Optics 20, 6728-6738 (2003).

[2] Totschinig G, Lackner M, Shau R, Ortsiefer M, Rosskepf J, Amann M C, Winter F, "High-speed vertical-cavity surface-emitting laser (VCSEL) absorption spectroscopy of ammonia (NH_3) near 1.54 μm," Appl. Phys. B 76, 603-608 (2003).

[3] Zhu X, Cassidy D T, "Modulation spectroscopy with a semiconductor diode laser by injection-current modulation," Optical Society of America B, Vol. 14, No. 8, August 1997.

[4] Wang Jian, Maiorov Mikhail, Jeffries Jay B, Garbuzov Dmitri Z, Connolly John C, Hanson Ronald K, "A potential remote sensor of CO in vehicle exhausts using 2.3 μm diode lasers," Meas. Sci. Technol. 11 1576-1584 (2000).

[5] P. Kluczynski and O. Axner, "Theoretical description based on Fourier analysis of wavelength-modulation spectrometry in terms of analytical and background signals," Applied Optics 20, 5803-5815 (1999).

[6] Silver J A, Wood W R, "Miniature gas sensor for monitoring biological space environments," Proceedings of SPIE, Vol. 4817 (2002).

[8] J. Verdeyen, Laser Electronics (3rd Edition), 1994

[9] A. Yariv, Quantum Electronics (3rd Edition), New York: Wiley, 1989

[10] O. Svelto, Principles of Lasers (4th Edition), New York : Plenum Press, c1998

[11] http://www.rp-photonics.com/relative_intensity_noise.html, see Appendix E

[12] K. Petermann, Laser diode modulation and noise, Kluwer Academic Publishers, Dordrecht 1988.

[13] Silver Joel A, "Frequency-modulation spectroscopy for trace species detection: theory and comparison among experimental methods," Applied Optics 20 Vol. 31, No.6, February 1992.

[14] R. Lang and K. Kobayashi, "External optical feedback effects on semiconductor injection laser properties," IEEE J. Quantum Electron., vol. QE-16, pp. 347–355, Mar. 1980.

[16] D. Babic and S. Corzine, "Analytic Expressions for the Reflection Delay, Penetration Depth, and Absorptance of Quarter-Wave Dielectric Mirrors," IEEE Journal of Quantum Electronics, vol. 28, no. 2, 512-524 (1992)

[17] Ortsiefer M, Baydar S, Windhorn K, Bohm G, Rosskepf J, Shau R, Ronneberg E, Hofmann W, Amann M C, "2.5-mW Single-Mode Operation of 1.55-μm Buried Tunnel Junction VCSELs," IEEE Photonics Technology Letters, Vol. 17, No. 8, August 2005

[18] Ryan N, McDonald D, Lavin A, "Performance improvements of widely tunable lasers for multi gas species detection and analysis," Proceedings of SPIE Vol. 5826 (2005).

[19] Shau R, Ortsiefer M, Rosskepf J, Bohm G, Lauer C, Maute M, Amann M C, "Long-wavelength InP-based VCSELs with Buried Tunnel Junction: Properties and Applications," Proceedings of SPIE Vol. 5364 (2004).

[20] Lytkine A, Jager W, Tulip J, "Frequency tuning of long-wavelength VCSELs," Spectrochimica Acta Part A 2005.

[21] Lytkine A, Jaeger W, Tulip J, "Long-wavelength VCSELs for applications in absorption spectroscopy: tuning rates and modulation performances," Proceedings of SPIE Vol. 5737 (2005).

[22] http://en.wikipedia.org/wiki/Electrical_ballast see Appendix F

[23] R. W. Tkach & A. R. Charaplyvy, "Regimes of Feedback Effects in 1.5-um Distributed Feedback Lasers," J. of Lightwave Technology, vol. LT-4, No. 11, November 1986, pp. 1655-1661.

[24] K. Petermann, "External Optical Feedback Phenomena in Semiconductor Lasers," IEEE Journal of Selected Topics in Quantum Electronics, vol. 1, No. 2, June 1995, pp. 480-488.

[25] N. Schunk and K. Petermann, "Numerical analysis of the feedback regimes for a single-mode semiconductor laser with external feedback," IEEE J. Quantum Electron., vol. 24, pp. 1242–1247, July 1988.

[28] Gaspar J, Chen S F, Gordillo A, Hepp M, Ferreyra P, Marques C, "Digital lock in amplifier: study, design and development with a digital signal processor," Microprocessors and Microsystems 28 157-162 (2004).

[29] M. Johnson, Photodetection and Measurement Maximizing Performance in Optical Systems, McGraw-Hill 2003.

[30] Engelbrecht R, "A compact NIR fiber-optic diode laser spectrometer for CO and CO_2: analysis of observed 2f wavelength modulation spectroscopy line shapes," Sectrochimica Acta Part A 60 3291-3298 (2004).

[31]http://www.physics.utoronto.ca/~astummer/pub/mirror/Projects/Util/Photodio de%20Characteristics%20and%20Applications%20(UDT%20Sensors%20Inc).pd f see Appendix G

[33] Sedra/Smith, Microelectronic Circuits 4ed, Oxford University Press 1998

[35] http://electronicdesign.com/Articles/Index.cfm?AD=1&ArticleID=4346, see Appendix H

[38] Ryan N, McDonald D, Lavin A, "Performance improvements of widely tunable lasers for multi gas species detection and analysis," Proceedings of SPIE Vol. 5826 (2005).

[42] Yu Yanguang, Yau Jianquan, "View for the development of theory on the self-mixing interference and general model of the displacement measurement," Proceedings of SPIE Vol. 4919,2002.

[43] Modugno G, Corsi C, Gabrysch M, Marin F, Inguscio M, "Fundamental noise sources in a high-sensitive two-tone frequency modulation spectrometer and detection of CO_2 at 1.6 μm and 2 μm," Appl. Phys. B 67, 289-296, (1998).

[48] Lackner M, Totschinig G, Winter F, Ortsiefer M, Amann M C, Shau R, Rosskepf J, "Demonstration of methane spectroscopy using a vertical-cavity surface-emitting laser at 1.68 μm with up to 5 MHz repetition rate," Measurement Science and Technology 14 101-106 (2003).

[50] Werle P, and Slemr F, "Signal-to-noise ratio analysis in laser absorption spectrometers using optical multipass cells," Applied Optics Vol. 30, No. 4, 1 February 1991.

[51] Riris H, Carlisle C B, Warren R E, Cooper D E, "Signal-to-noise ratio enhancement in frequency-modulation spectrometers by digital signal processing," Optics Letters Vol. 19, No. 2, January 15 1994.

[52] Riris H, Carlisle C B, Warren R E, Carr L W, Cooper D E, Martinelli R U, Menna R J, "Design of an open path near-infrared diode laser sensor: application to oxygen, water, and carbon dioxide vapor detection," Applied Optics Vol. 33, No. 30, 20 October 1994.

[53] DisselKamp R S, Kelly J F, Sams R L, Anderson G A, "Signal-to-noise enhancement techniques for quantum cascade absorption spectrometers employing optimal filtering and other approaches," Appl. Phys. B 75, 359-366 (2002).

[54] Claps R, Englich F V, Leleux D P, Richter D, Tittel F, and Curl R F, "Ammonia detection by use of near-infrared diode-laser-based overtone spectroscopy," Applied Optics Vol. 40, No. 24, 20 August 2001.

[55] Wainner R T, Green B D, Allen M G, White M A, Stafford-Evans J, Naper R, "Handheld, battery-powered near-IR TDL sensor for stand-off detection of gas and vapor plumes," Appl. Phys, B 75, 249-254 (2002).

[56] Linnerud I, Kaspersen P, Jaeger T, "Gas monitoring in the process industry using diode laser spectroscopy," Laser and Optics, 2 June 1998.

[57] Huang W, Li X, Gao X, Fan H, Li W, Huang T, Pei S, Shao J, Deng L, Zhang W, "Portable Remote Sensor of Methane Leakabe Using Near-IR Diode Laser," Optical Tech, for Atm., Oeean and Enviro. Studies, Proceedings of SPIE Vol. 5832 (2005).

[60] Shao J, Gao X, Zhang W, Yuan Y, Ning L, Yang Y, Pei S, Huang W, ", Highly sensitive diode laser absorption measurements of CO_2 near 1.57 μm at room temperature," Optica Applicata, Vol. XXXV, No. 1, 2005.

[61] Well B v, Murray S, Hodgkinson J, Pride R, Strzoda R, Gibson G, Padgett M, "An open-path, hand-held laser system for the detection of methane gas," J. Opt, A: Appl. Opt, 7 S420-S424 (2005).

[62] Bacsik Z, and Komlosi V, "Comparison of Open Path and Extractive Long-Path FTIR Techniques in Detection of Air Pollutants," Applied Spectroscopy Reviews, 41: 77-97, 2006.

APPENDIX A

Below is provided the exact Matlab code of the m-file that was used to yield the transimpedence amplifier simulation results.

```matlab
%parameters
rsh=38.053e6;  rs=75;
cd=124.731e-12;
rf=1e6*1e0;
cf=10e-12;
in=0.4e-12;
f1=1/(2*pi*rf*cf)
P=5e-6;  Ip=.95*P;

f=[10:10:1e6]';
if f>=100
    en=3e-9;
    else en=3.5e-9;
    end;
w=2*pi*f;
s=j*w;
zcf=1./(s*cf);
zf=rf*zcf./(rf+zcf);
zcsh=1./(s*cd);
zsh=rsh*zcsh./(rsh+zcsh);

G=-rf/rs
G_c=abs(G./(1+w/(f1*2*pi)));

nsh=0.57*sqrt(Ip/1e-6)*1e-12*rf;    G_nsh=abs(nsh./(1+w/(f1*2*pi)));

nth=4*sqrt(rf/1e3)*1e-9;            G_nth=abs(nth./(1+w/(f1*2*pi)));

G_nin=abs(in*rf./(1+w/(f1*2*pi)));

ven=abs(en*(1+zf./zsh));
ntot=sqrt(G_nsh.^2+G_nth.^2+G_nin.^2+ven.^2);
figure(1);
loglog(f,G_nsh,':',f,G_nth,'.',f,G_nin,'-.',f,ven,'--',f,ntot,'-');
legend('shot','thermal','amp cur','amp vol','total');
title('Noise profile for a PD transimpedance amplifier');
ylabel('Output voltage (V)');
xlabel('Frequency (Hz)');

G_Ip=abs(Ip*rf./(1+w/(f1*2*pi)));
SNR=G_Ip./ntot;

figure(2);
loglog(f,SNR);
return;
```

APPENDIX B

Below are provided the exact Matlab codes of the m-files that were used to yield the wavelength modulation spectroscopy simulation results.

Signal.m

```matlab
function signal;
%
% This function is designed to apply the theoretical
% analysis to an IM-FM signal

global t ph pom pOM ph m w delvl vl Io IOM;

%t=[0:.000004:.999996]';
t=[0:.00004:.99996]';   %The time interval in insufficient for f=11kHz

degph=10;  ph=degph*pi/180;
m=2.00;   f=10e3;   w=2*pi*f;   a=2;
Io=3;   pom=-2.64; pOM=-1.05;   delvl=.07327;

x=12*t-m*cos(w*t)-6;
%x = x+1*rand(size(t))-1/2;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% This should be used if a representation with
% the ramp frequency is explicitly desired.
%x=114*t-m*cos(w*t)-6;
%per=length(t)/9.5;
%for i=2:10,
% if i~=10
%  x(1+round((i-1)*per):round(i*per))=x(1+round((i-1)*per):round(i*per))-12*(i-1);
% else
%  x(1+round((i-1)*per):end)=x(1+round((i-1)*per):end)-(x(1+round((i-1)*per))-
x(1));
%end;end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

IOM=pOM*delvl*x+1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Schilt's eqn 12 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
harmonic=2;              % desired harmonic
detph=harmonic*ph+pi/2;      % the detection phase (maximum sig at harmonic*phase)
sIMFM=Io*(sphase(harmonic,x)*cos(detph)+squad(harmonic,x)*sin(detph));
[sx,i]=sort(x);
%correlate(sIMFM(i));
% The y axis is normalized
figure(1);   plot(sx,(sIMFM(i)/max(sIMFM))); grid;
title(['IM-FM stats: Harmonic ',num2str(harmonic),'; Mod Index ',num2str(m),...
'; s mag max ',num2str(max(sIMFM)),'; s ampl max ',num2str(range(sIMFM))]);

pom=0; pOM=0;   IOM=1;
sFM=Io*(sphase(harmonic,x)*cos(detph)+squad(harmonic,x)*sin(detph));
% The y axis is normalized
hold on;      plot(sx,(sFM(i)/max(sFM)),'r--');   legend('IM-FM','FM');   hold
off;

%plot(x,att);
return;

%%%%%%%%%%%%%%%%%%%%%%% using equation 8; method 2 %%%%%%%%%%%%%%%%%%%%%%%%%%%%
% There is some ambiguity to it.  The n=0 term begets an e(-1), and I've
% assumed this to be 2.
sIMFM1=zeros(size(x));
for n=0:9,
 sp=sphase(n,x);
 sq=squad(n,x);
 sIMFM1=sIMFM1+sp.*cos(n*w*t)-sq.*sin(n*w*t);
end;
me2=Io*sIMFM1;
%figure(2);
```

```
%subplot(1,2,1);plot(me2);
%subplot(1,2,2);plot(abs(fft(me2)));

filter=zeros(size(x));
filter(end-11151+1:end-10850+1)=1;
filter(10850:11151)=1;
ME2=fft(me2).*filter;
m2=ifft(ME2);
maco1=angle(m2)>pi/2;maco2=angle(m2)<-pi/2;
maco=maco1+maco2;
maco=maco*-2+1; mac=abs(m2).*maco;
figure(2);
plot(mac);
```

En.m
```
function eterm=en(n);

switch n
  case 0
    eterm=1;
  otherwise
    eterm=2;
end;
```

S.m
```
function s=s(n,x);

global t m Io;

a=10;
sub=1-i*x;

fir=(-i)^n*en(n)*(sqrt(sub.^2+m^2)-sub).^n./(m^n*sqrt(sub.^2+m^2));
s=Io*(besselj(n,0)-a*(fir+conj(fir))/2);

return;
```

Correlate.m
```
function correlate(x,xnoisy);

global t;% xnoisy;

%nlvl = 10;
%xnoisy = x+nlvl*rand(size(t))-nlvl/2;

mx=mean(x);
my=mean(xnoisy);
denom = sqrt(sum((x-mx).^2)*sum((xnoisy-my).^2));
numer = sum((x-mx).*(xnoisy-my));
r = numer/denom
return;
```

Sphase.m
```
function sp=sphase(n,x);

global IOM ph pOM delvl m pom;

sp=IOM*cos(n*ph).*s(n,x)-pOM*delvl*m/2*(en(n)/2-en(n)+1)...
    *cos(n*ph)*s(n+1,x)-pom*delvl*m/2*(2/en(n-1)*cos((n-1)*ph)...
    *s(n-1,x)+(en(n)-1)*cos((n+1)*ph)*s(n+1,x));
return;
```
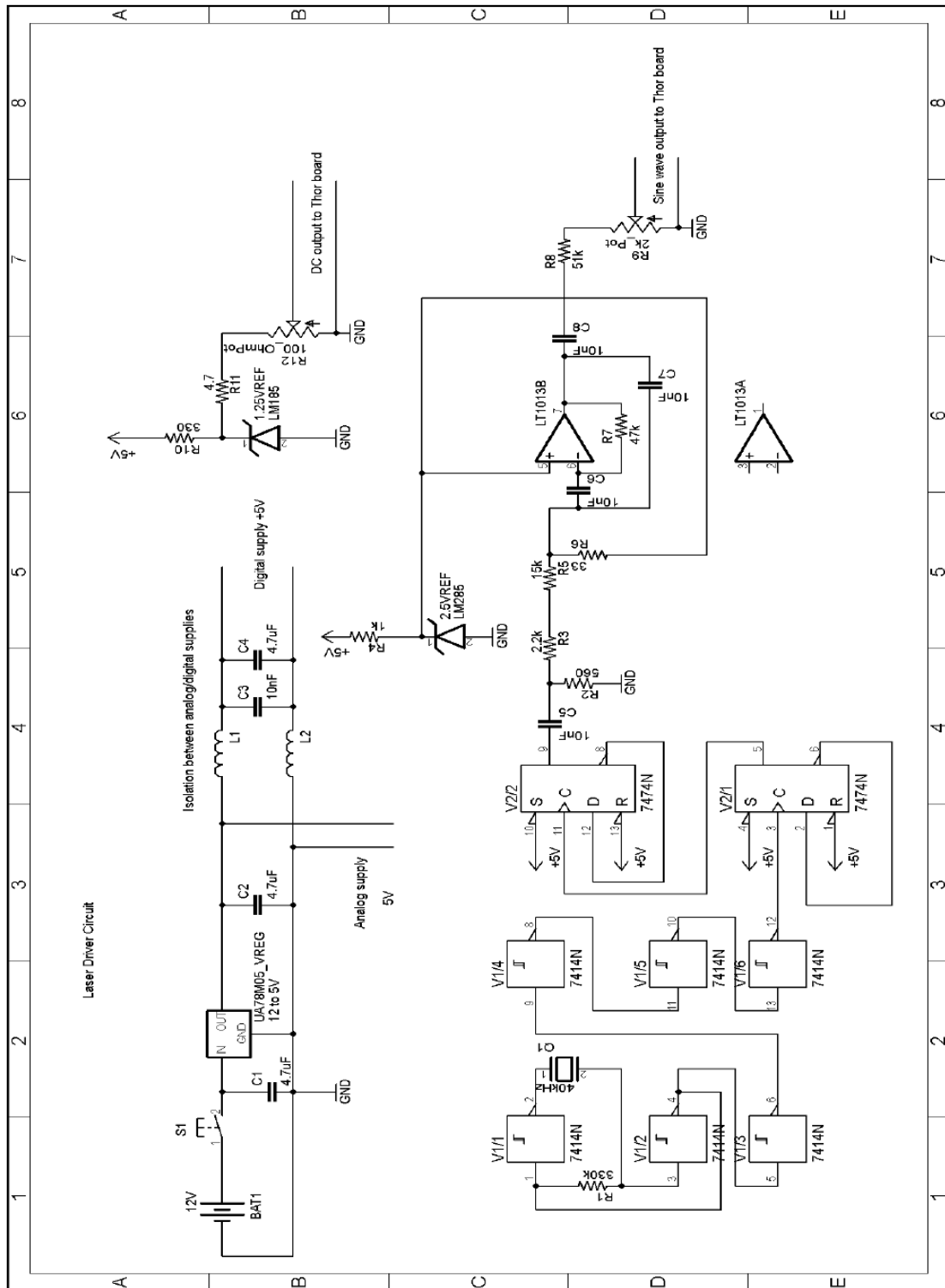
Squad.m
```
function sq=squad(n,x);
```

```
global IOM ph pOM delvl m pom;

sq=IOM*sin(n*ph).*s(n,x)-pOM*delvl*m/2*(en(n)/2-en(n)+1)...
    *sin(n*ph)*s(n+1,x)-pom*delvl*m/2*(2/en(n-1)*sin((n-1)*ph)...
    *s(n-1,x)+(en(n)-1)*sin((n+1)*ph)*s(n+1,x));
return;
```

APPENDIX C

Laser modulation circuit

APPENDIX D

Supplemental terms

Dark Current (I$_{DRK}$)

Dark current is leakage current which flows when no light is irradiated upon the photodiode but the diode circuit is on, that is to say, the appropriate biases are applied. With bias, voltage is applied across the junction. This voltage may be as low as 10 mV or as high as 50 V and the dark currents may vary from pAs to µAs depending upon the junction area and the manufacturing process used. On datasheets, dark current is always specified at a particular value of reverse applied voltage.

The dark current is temperature dependent. The rule of thumb is that the dark current will approximately double for every 10°C increase in ambient temperature. However, specific diode types can vary considerably from this relationship.

Shunt Resistance (R$_{SH}$)

Values for the shunt resistance can vary from 100k to 100G ohms. The noise performance and the linearity of the short circuit photocurrent are directly related to the value of the shunt resistance. This parameter is voltage dependent but still is quite useful in calculating the offset gain ($G_{OS} = 1 + R_F / R_{SH}$) in transimpedance amplifier circuits. The shunt resistance is dependent on the active area of the diode chip and on the type of processing used in constructing the diode. It is also temperature dependent, decreasing with increasing temperature.

Junction Capacitance (C$_J$)

A capacitance is associated with the depletion region which exists at the P-N junction of the photodiode. The response time of a photodiode is dependent to a large extent upon the product of the junction capacitance and the external load resistor. Recalling that capacitance may be defined by $C = \varepsilon A/d$, it is easy to see that junction capacitance increases with increasing junction area of the photodiode chip. It is also a function of any reverse voltage applied across the photodiode. When a reverse voltage is applied the depletion layer will expand and the junction capacitance will decrease. The capacitance will continue to decrease with increasing reverse applied voltage until the depletion region expands to the back surface of the photodiode chip at which point the diode's capacitance becomes nearly constant. Also, PD response time to a sudden change in light level is commonly expressed in terms of the rise time (t_R) or the fall time (t_F) where:

t_R = The time required for the output to rise from 10% to 90% of its final value.
t_F = The time required for the output to fall from 90% to 10% of its on state value.

The response time of a photodiode depends upon many factors, including the wavelength of the light, the value of the applied voltage across the diode (since this has a major effect of the junction capacitance), and the load resistance.

Characteristic curves given with each process show that the nonsaturated response time is largely dependent upon the product of the junction capacitance and load resistance. However, when this product is small, other effects become significant and limit the response time.

Noise Current (I_N)

A photodiode will act as a source for electrical noise and generate a noise current (I_N). The noise current will limit the usefulness of the photodiode at very low light levels where the magnitude of the noise approached that of the signal photocurrent. The amount of noise generated is dependent upon the characteristics of the photodiode and the operating conditions. There are three main components which go into making up the total noise generated by the photodiode; thermal noise, shot noise, and excess noise.

Thermal (or Johnson noise) is inversely related to the value of the shunt resistance (R_{SH}) of the photodiode. Thermal noise tends to be the dominant noise component when the diode is operated under zero applied reverse bias conditions.

Shot noise is dependent upon the leakage or dark current (I_{DRK}) of the photodiode. It tends to dominate when the photodiode is used in the photoconductive mode where an external reverse bias is applied across the device.

Excess noise, also known as $1/f$ noise, is unlike thermal or shot noise in that it possesses a $1/f$ spectral density. Excess noise may dominate when the bandwidth of interest contains frequencies less than 1 kHz. As noted elsewhere it is one of the detrimental factors leading to the adoption of the $2f$-WMS mode for concentration detection of trace methane.

APPENDIX E

Excerpt from http://www.rp-photonics.com/relative_intensity_noise.html [11]

Relative Intensity Noise

Acronym: RIN

Definition: noise of the optical intensity (or actually power), normalized to its average value

In the context of intensity noise (optical power fluctuations) of a laser, it is common to specify the *relative intensity noise* (RIN), which is the power noise normalized to the average power level. The optical power of the laser can be considered to be

$$P(t) = \bar{P} + \delta P(t)$$

with an average value and a fluctuating quantity δP with zero mean value. The relative intensity noise is then that of δP divided by the average power; in the following, this quantity is called I. The relative intensity noise can then be statistically described with a power spectral density (PSD):

$$S_I(f) = \frac{2}{\bar{P}^2} \int_{-\infty}^{+\infty} \langle \delta P(t)\, \delta P(t + \tau) \rangle \exp\left(i 2\pi f \tau\right) d\tau$$

which depends on the noise frequency f. It can be calculated as the Fourier transform of the autocorrelation function of the normalized power fluctuations (see the equation), or measured e.g. with a photodiode and an electronic spectrum analyzer. (The factor of 2 in the formula above leads to a one-sided PSD as usually used in the engineering disciplines.) The units of this RIN PSD are Hz^{-1}, but it is common to specify 10 times the logarithm (to base 10) of that quantity in dBc/Hz (see also: decibel). The PSD may also be integrated over an interval $[f_1,\ f_2]$ of noise frequencies to obtain a root mean square (r.m.s.) value of relative intensity noise

$$\left.\frac{\delta P}{\bar{P}}\right|_{rms} = \sqrt{\int_{f_1}^{f_2} S_I(f)\, df}$$

which is then often specified in percent.

Note that it is not sensible (although common) to specify relative intensity noise in percent (e.g. as ±0.5%) without clarifying whether this means an r.m.s. value or something else. See the article on noise specifications for more such details.

APPENDIX F

Electrical ballast

From Wikipedia, the free encyclopedia

Jump to: navigation, search

"Choke ballast" (inductor) used in older lighting. This example is from a tanning bed. Requires a lamp starter (below) and capacitor.

Lamp starter, required with some inductor type ballasts. Connects both ends of the lamp together to "preheat" the lamp ends for 1 second before lighting.

An **electrical ballast** (sometimes called **control gear**) is a device intended to limit the amount of current in an electric circuit.

Ballasts vary greatly in complexity. They can be as simple as a series resistor as commonly used with small neon lamps or light-emitting diodes (LEDs). For higher-power installations, too much energy would be wasted in a resistive ballast, so alternatives are used that depend upon the reactance of inductors, capacitors, or both. Finally, ballasts can be as complex as the computerized, remote-controlled electronic ballasts now often used with fluorescent lamps.

Contents

[hide]

Current Limiting

Ballasts are used where an electrical load cannot effectively regulate its current use. These are most often used when an electrical circuit or device presents a negative (differential) resistance to the supply. If such a device were connected to a constant-voltage power supply, it would draw an increasing amount of current until it was destroyed or caused the power supply to fail. To prevent this, a ballast provides a positive resistance or reactance that limits the ultimate current to an appropriate level. In this way, the ballast provides for the proper operation of the negative-resistance device by appearing to be a legitimate, stable resistance in the circuit.

Examples of such negative-resistance devices are gas-discharge lamps.

Ballasts can also be used simply to deliberately reduce the current in an ordinary, positive-resistance circuit.

Prior to the advent of solid-state ignition, automobile ignition systems commonly included a ballast resistor to regulate the voltage applied to the ignition system.

Although LEDs are positive resistance devices, they have insufficient resistance to regulate their current consumption when operated from a voltage controlled source, so ballasts are used to control the current through the LED. Because the power dissipation is minuscule, simple resistor ballasts are normally used.

Resistors

The term **ballast resistor** primarily refers to a resistor which compensates for normal or incidental changes in the physical state of a system. It may be a fixed or variable resistor.

Fixed resistors

For simple, low-powered loads such as a neon lamp or LED, a fixed resistor is commonly used. Because the resistance of the ballast resistor is large it dominates the current in the circuit, even in the face of negative resistance introduced by the neon lamp.

The term also refers to an automobile engine component that lowers the supply voltage to the ignition system after the engine has been started. Because cranking the engine causes a very heavy load on the battery, the system voltage can drop quite low during cranking. To allow the engine to start, the ignition system must be designed to operate on this lower voltage. But once cranking is completed, the normal operating voltage is regained; this voltage would overload the ignition system. To avoid this problem, a ballast resistor is inserted in series with the supply voltage feeding the ignition system. Occasionally, this ballast resistor will fail and the classic symptom of this failure is that the engine runs while being cranked (while the resistor is bypassed) but stalls immediately when cranking ceases (and the resistor is re-connected in the circuit).

Modern electronic ignition systems do not require a ballast resistor as they are flexible enough to operate on the low cranking voltage or the ordinary operating voltage.

In some old AC/DC receivers (universal sets), the vacuum tube heaters are connected in series. Since the voltage drop across all the filaments in series is sometimes less than the full mains voltage, it was often necessary to get rid of the excess voltage. A ballast resistor was often used for this purpose, as it was cheap and worked with both AC and DC.

Self-variable resistors

Some ballast resistors have the property of increasing in resistance as current through them increases, and decreasing in resistance as current decreases. Physically, some such devices are often built quite like incandescent lamps. Like the tungsten filament of an ordinary incandescent lamp, if current increases, the ballast resistor gets hotter, its resistance goes up, and its voltage drop increases. If current decreases, the ballast resistor gets colder, its resistance drops, and the voltage drop decreases. Therefore the ballast resistor reduces variations in current, despite variations in applied voltage or changes in the rest of an electric circuit. These devices are sometimes termed barretters.

This property can lead to more precise current control than merely choosing an appropriate fixed resistor. The power lost in the resistive ballast is also reduced because a smaller portion of the overall power is dropped in the ballast compared to what might be required with a fixed resistor.

In times past, household clothes dryers sometimes incorporated a germicidal lamp in series with an ordinary incandescent lamp; the incandescent lamp operated as the ballast for the germicidal lamp. A commonly used light in the home in the 1960s in 220-240V countries was a circleline tube ballasted by an under-run regular mains filament lamp. Self ballasted mercury-vapor lamps incorporate ordinary tungsten filaments within the overall envelope of the lamp to act as the ballast, and it supplements the otherwise lacking red area of the light spectrum produced.

Reactive ballasts

Several typical magnetic ballasts for fluorescent lamps. The top is a high-power factor, lead-lag ballast for two 30-40W lamps. The middle is a low power factor ballast for a single 30-40W lamp while the bottom ballast is a simple inductor used with a 15W preheat lamp.

Because of the power that would be lost, resistors are not used as ballasts for lamps of more than about two watts. Instead, a reactance is used. Losses in the ballast due to its resistance and losses in its magnetic core may be significant, on the order of 5 to 25% of the lamp input wattage. Practical lighting design calculations must allow for ballast loss in estimating the running cost of a lighting installation.

An inductor is very common in line-frequency ballasts to provide the proper starting and operating electrical condition to power a fluorescent lamp, neon lamp, or high intensity discharge (HID) lamp. (Because of the use of the inductor, such ballasts are usually called *magnetic ballasts*.) The inductor has two benefits:

1. Its reactance limits the power available to the lamp with only minimal power losses in the inductor
2. The voltage spike produced when current through the inductor is rapidly interrupted is used in some circuits to first strike the arc in the lamp.

A disadvantage of the inductor is that current is shifted out of phase with the voltage, producing a poor power factor. In more expensive ballasts, a capacitor is often paired with the inductor to correct the power factor. In ballasts that control two or more lamps, line-frequency ballasts commonly use different phase relationships between the multiple lamps. This not only mitigates the flicker of the individual lamps, it also helps maintain a high power factor. These ballasts are often called *lead-lag* ballasts because the current in one lamp leads the mains phase and the current in the other lamp lags the mains phase.

For large lamps, line voltage may not be sufficient to start the lamp, so an autotransformer winding is included in the ballast to step up the voltage. The autotransformer is designed with enough leakage inductance so that the current is appropriately limited.

Because of the large inductors and capacitors that must be used, reactive ballasts operated at line frequency tend to be large and heavy. They commonly also produce acoustic noise (line-frequency hum).

Prior to 1980 in the United States, PCB-based oils were used as an insulating oil in many ballasts to provide cooling and electrical isolation (see transformer oil)[1][2].

Electronic ballasts
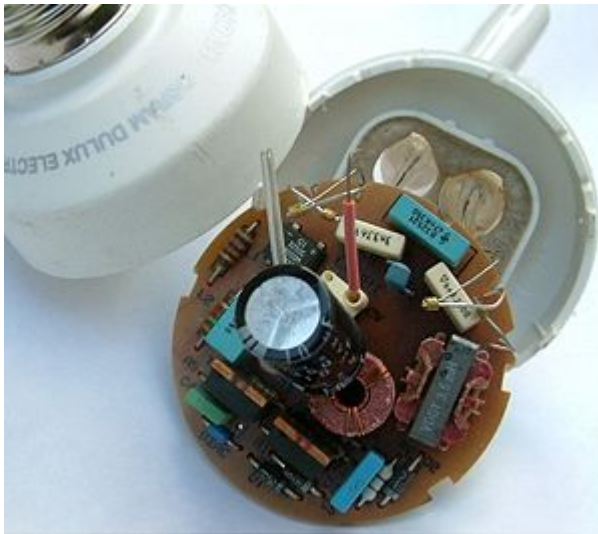
Electronic ballast of a compact fluorescent lamp

An **electronic lamp ballast** uses solid state electronic circuitry to provide the proper starting and operating electrical condition to power one or more fluorescent lamps and more recently HID lamps. Electronic ballasts usually change the frequency of the power from the standard mains (e.g., 60 Hz in U.S.) frequency to 20,000 Hz or higher, substantially eliminating the stroboscopic effect of flicker (a product of the line frequency) associated with fluorescent lighting

(see photosensitive epilepsy). In addition, because more gas remains ionized in the arc stream, the lamps actually operate at about 9% higher efficacy above approximately 10 kHz. Lamp efficacy increases sharply at about 10 kHz and continues to improve until approximately 20 kHz.[3] Because of the higher efficiency of the ballast itself and the improvement of lamp efficacy by operating at a higher frequency electronic ballasts offer higher system efficacy. In addition, the higher operating frequency means that it is often practical to use a capacitor as the current-limiting reactance rather than the inductor required at line frequencies. Capacitors tend to be much lower in loss than inductors, allowing them to more closely approach an "ideal reactance".

Electronic ballasts are often based on the SMPS topology, first rectifying the input power and then chopping it at a high frequency. Advanced electronic ballasts may allow dimming via pulse-width modulation and remote control and monitoring via networks such as LonWorks, DALI, DMX-512, DSI or simple analog control using a 0-10V DC brightness control signal.

Fluorescent lamp ballasts

Instant start

An instant start ballast starts lamps without heating the cathodes at all by using a high voltage (around 600 V). It is the most energy efficient type, but gives the least number of starts from a lamp as emissive oxides are blasted from the cold cathode surfaces each time the lamp is started. This is the best type for installations where lamps are not turned on and off very often.

Rapid start

A rapid start ballast applies voltage and heats the cathodes simultaneously. Provides superior lamp life and more cycle life, but uses slightly more energy as the cathodes in each end of the lamp continue to consume heating power as the lamp operates. A dimming circuit can be used with a dimming ballast, which maintains the heating current while allowing lamp current to be controlled.

Programmed start

A programmed-start ballast is a more advanced version of rapid start. This ballast applies power to the filaments first, then after a short delay to allow the cathodes to preheat, applies voltage to the lamps to strike an arc. This ballast gives the best life and most starts from lamps, and so is preferred for applications with very frequent power cycling such as vision examination rooms and restrooms with a motion detector switch.

References

1. **^** [Hazardous Fluorescent Lamp Ballasts](#)
2. **^** [Recycling Fluorescent Lights](#)
3. **^** IES Lighting Handbook 1984

APPENDIX G

Excerpts from
http://www.physics.utoronto.ca/~astummer/pub/mirror/Projects/Util/Photodiode%20Characteristics%20and%20Applications%20(UDT%20Sensors%20Inc).pdf
[31]

■ ELECTRICAL CHARACTERISTICS

A silicon photodiode can be represented by a current source in parallel with an ideal diode (Figure. 3). The current source represents the current generated by the incident radiation, and the diode represents the p-n junction. In addition, a *junction capacitance* (C_j) and a *shunt resistance* (R_{SH}) are in parallel with the other components. *Series resistance* (R_S) is connected in series with all components in this model.
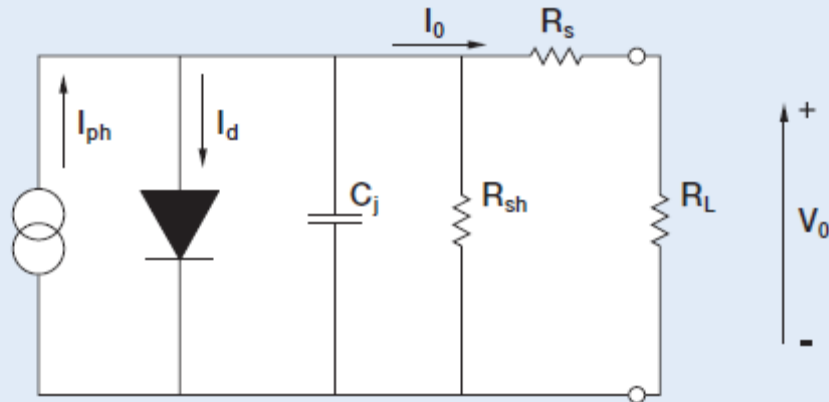

Figure 3. Equivalent Circuit for the silicon photodiode

APPENDIX H

What's All This Transimpedance Amplifier Stuff, Anyhow? (Part 1)

Bob Pease
ED Online ID #4346

One of the first things you learn about operational amplifiers (op amps) is that the op amp's gain is very high. Now, let's connect a feedback resistor across it, from the output to the −input. When you put some input current into the −input (also known as the *summing point*), the gain is so high that all of the current must go through the feedback resistor. So, the output will be $V_{OUT} = -(I_{IN} \times R_F)$. That's neat. While we used to call this a "current-to-voltage converter," which it is indeed, it's also sometimes referred to as a "transimpedance amplifier," where the "gain" or "transimpedance" is equal to R_F.

There's a whole class of applications in which this configuration is quite useful and important. An important case is when you need an op amp to amplify the signal from a sensor, such as a photodiode. Photodiodes put out current at high impedance (high at dc), but often they have a lot of capacitance. If you just let the photo diode dump its current out into a resistor, there are two problems. If the sense resistor is large, then the gain can be fairly large, but the response will be slow and the time-constant will be large: $\tau = R_L \times C_S$. But if you choose a small sense resistor to get a small τ, the gain will be low. The signal-to-noise ratio (SNR) may also be unacceptable. How can you avoid poor gain and/or poor response? Kay garney? (That's Nepali for "What to do?")

To avoid this terrible compromise, it's a good idea to feed the photodiode's output current directly into the summing point of a transimpedance amplifier. Here, the response time is not $R_F \times C_S$, but considerably faster. Plus, the gain can be considerably larger, because now you can use a larger R_F. This helps improve the signal-to-noise ratio too!

When you connect up the diode like this, the first thing you realize is that the darned thing is oscillating! Why? Well, it's well known that the input capacitance of an op amp (and its circuitry) can cause instability when the op amp is used with a feedback resistor. You *usually* need to add a feedback capacitor across R_F to make it stable. In the old days, it was stated that:

$C_F \times R_F = C_{IN} \times R_{IN}$

So if you have a unity-gain inverter with $R_{IN} = R_F = 1$ MΩ, and the input capacitance of the op amp is 10 pF, then you're supposed to install a feedback capacitor of 10 pF. That's what people said for *years*. The LF156 data sheet stated this, and it still does. But that's not exactly true. A complete explanation is a bit beyond the scope of this column, but in practice you can usually get away with a much smaller feedback capacitor. In many cases, you can get a response that's improved by a factor of five or 10, and still not get excessive (more than 5% or 10%) overshoot. In practice, you have to *tweak* and optimize the feedback capacitance as you observe the response.

The formula for the optimized amount of C_F is, if:

$$\left(\frac{R_F}{R_{IN}} + 1\right) \geq 2\sqrt{GBW \times R_F \times C_S}$$

then:

$$C_F = \frac{C_S}{2\left(\frac{R_F}{R_{IN}} + 1\right)}$$

but if:

$$\left(\frac{R_F}{R_{IN}} + 1\right) < 2\sqrt{GBW \times R_F \times C_S}$$

the feedback capacitor C_F should be:

$$C_F = \sqrt{\frac{C_S}{GBW \times R_F}}$$

Now, whenever you have an op amp with a large C_S, a large R_F, and a small C_F, the noise gain will rise at moderate frequencies. The definition of noise gain is the reciprocal of the attenuation from the output back to the −input. In other words, if the attenuation is $Z_{IN}/(Z_{IN} + Z_F)$, then the noise gain is $1 + Z_F/Z_{IN}$.

At moderate frequencies, the Z_F is determined by R_F, and Z_{IN} is established by C_S. So, the noise gain will rise until the frequency where the impedance of C_F becomes equal to R_F. Then the noise gain flattens out, typically at a large number, such as 20, 40, or 80. We do this because if the noise gain kept rising at 6 dB/octave while the op amp's gain is rolling off at 6 dB/octave, the loop is going to be unstable, and it will oscillate. The reason that we choose a small value of C_F is to make the noise gain flatten out, make the loop stable, and stop the oscillation and ringing.

If you make $C_F = C_{IN}$, you can get the noise-gain curve to stay flat as in line A-E. It will be very stable but have a very slow response. If you add no feedback capacitor, the noise gain will tend to rise as per line A-B-C. This will cause instability. Selecting a suitable small value for C_F can get the smooth results shown by line A-B-D. Yeah, it's as easy as ABD to get fast, stable response by picking a small C_F. So, we have made the feedback capacitance big enough to stop the oscillation and minimize the overshoot. Now what?

There's a pretty good book by Jerald Graeme (ex-Burr-Brown) on the topic of the transconductance amplifier: *Photodiode Amplifiers—Op Amp Solutions*. Jerry and I have definitely come to the same basic conclusion. When you want to optimize a transimpedance amplifier, everything interacts. Therefore, every time you compute the response and the noise, and change any factor, the computations may change considerably. There's no simple or obvious way to compute or optimize the performance. The performance, in terms of response or bandwidth, in terms of peaking or overshoot, and in terms of noise or SNR, is an extremely complicated, nonlinear, and highly interacting function of:

- the feedback resistor
- the source capacitance
- the feedback capacitance
- the desired bandwidth
- the desired gain factor (which does predict the full-scale output voltage)
- the voltage noise of the op amp

- the current noise of the op amp
- the input capacitances of the op amp
- and the gain-bandwidth product of the op amp.

Jerry and I certainly agree on that. Jerry's book is well written, and for just $55, it's pretty much a bargain. I recommend it: ISBN = 0-07-024237-X *(www.amazon.com/exec/obidos/ASIN/007024247X/o/qid=968200121/sr=8-1/ref=aps_sr_b_1_3/002-6674439-7948805).*

But I also have worked on this general problem many times over the years and have several suggestions that go be-yond Jerry's book. More on this later.

There are several basic rules of thumb that Jerry and I agree upon:

(A) You want to avoid an op amp with high voltage noise ($nV/\sqrt{Hz}$).

(B) You want to avoid an op amp with high current noise ($pA/\sqrt{Hz}$). (Most bipolar op amps have much higher current noise than FETs.) It's a rare case when an op amp with bipolar input transistors is better, except when R_S is very low or resistive (or in cases where the input is capacitive but the bandwidth is narrow).

(C) You usually want to avoid an op amp with large input capacitance. Unfortunately, most data sheets don't properly specify the op amp's input capacitances, neither differential-mode nor common-mode. But it's fair to assume that most "low-noise" op amps have a larger input capacitance than ordinary op amps. You may want to ask the manufacturer, or you might just decide to measure it yourself.

(D) Much of the noise of such a transimpedance amplifier is proportional to $\sqrt{BW}$ &215; $C_{SOURCE} \times V_N$ of the op amp. So if you want to get low noise, you must optimize very carefully. Specifically, begin by computing the im-pedance Z_S of your sensor at the maximum frequency of interest:

$Z_S = 1/2\pi FC_S$

For a good amplifier, the voltage noise and the current noise times Z_S should both be as small as you can get. If one of these noises is much larger than the other, then you're probably far off optimum.

(E) If you have any choice of what sensor you employ, try to find a lower-capacitance sensor. Furthermore, make a low-capacitance layout between the sensor and the op amp.

If you want to get fast response, low noise, or wide bandwidth, Jerry's book offers some pretty good advice. More on that later.

But Jerry didn't include a list of good op amps that have low voltage noise, and/or low current noise, and/or low input capacitance. Because some are better than others, I bet you can use Paul Grohe's selector guide to find some low-noise op amps. See *www.national.com/selguide* for free "Selguide" software that can run on your PC to help you select a good, low-noise, inexpensive op amp.

Also, Jerry neglected to mention that you can design your own op amp with better, lower voltage noise and better bandwidth. I mean, op amps that you can buy off the shelf cover a wide array of cases where they are optimized for low V_{NOISE} and low I_{NOISE}, wide bandwidth, low power drain, and so on. But you can "roll your own" surprisingly easily and accomplish even better performance for a specified application! I'm not proposing that you design a complete op amp, but it's simple to just add a new low-noise front end ahead of a suitable op amp.

The basic idea is to add a couple of good low-noise FETs in front of an existing op amp. Most op amps don't operate the front-end transistors as rich as the output. Yet in a case like this, there's no reason at all not to run more current through the front end than in the rest of the op amp. My first pick is the 2N5486, which has less than 1 pf of C_{RSS}, but has a lot of g_{ms} (4 millimhos) and low voltage noise (at I_S = 3 mA). So for my first design, I'll just put a matched pair* of 2N5486s in front of a decent wideband op amp, such as the LM6171. What's the voltage noise of this amplifier? We may be able to get an average of 3 nV/√Hz, out to 10 kHz.

When you're designing an op amp, remember this: adding gain is one of the cheapest things you can add. You on-ly need to be careful about how to give that gain away—to roll it off.

In this case, it's easy. The R1-C1 network in Figure 5 just rolls off the gain for a fairly smooth frequency response. To achieve 2 MHz of bandwidth and a fairly good, smooth 6-dB/octave rolloff, I suggest R1 = 75 Ω, and C1 = 100 pF as a good place to start your design.

But now, look at the refinements in Figure 6. We can roll off the amplifier's gain simply in two swoops. The low-frequency gain is rolled off by R_X and C_X. Then after the gain rolls off flatly, we roll it off some more by R_Y and C_Y. When we are finished, it should look something like curve X in Figure 7. This isn't exactly rocket science. We just want to make it a practical design. But this is a whole system design. You can't very well design and optimize the op amp alone. It's the op amp, the feedback system, the noise filters, and the post-amplifiers that have to be considered and optimized all together. My first-hack proposals for these damping/stabilization components are:

R_X = 5.1 kΩ, C_X = 50 pF

R_Y = 330 Ω, C_Y = 7.5 pF

The whole point behind making your own op amp is that you do not have to just build an op amp with a smooth 6-dB/octave rolloff, all the way out to a few megahertz. You can roll off the gain at a 6 dB/octave out to some intermediate frequency, and then flatten out the gain. Then, at a higher frequency, let it roll off some more in some vaguely controlled way. This would make a lousy general-purpose op amp, but it might be ideal for a case where the noise gain is rising, such as in a transimpedance amplifier. (Look at the old LM709. When you choose the correct damping networks, it can provide a gain of 1000 out to some high frequency like 1 MHz.)

Also note that I added a second pair of 2N5486s to improve the voltage noise. Yes, this will approximately *double* the input capacitance. But if your C_S is already large, this may easily improve the signal-to-noise ratio. If it's good to have two, will three be better? I'll let you figure that out! But, yes, four or five may provide definite improvements... or that might not be the case.

I won't recommend that you design your own op amp if you can buy one that does the job. But if the best one you can buy isn't good enough, then there's some hope here. Designing your own composite op amp is not that hard, and not that expensive, even if you are going to build one or 10 or 1000. The post-amplifier can be inexpensive. Of course, all of the basic designs will be somewhat different if you are running on ±5-V supplies, or ±15-V supplies.

Either way, it's not that difficult, but the design compromises are slightly different. Here, I just showed a couple of ±15-V applications. (The ±5-V designs differ mostly by using a low-voltage, rail-to-rail-output op amp.)

In future columns on this topic, I will comment on other aspects of design and optimization for transimpedance am-plifiers.

Meanwhile, try to avoid Tee networks in the feedback network. They often cause poor signal-to-noise ratios. Next time, I'll explain that completely. Yes, a Tee network might help you avoid buying 1000-MΩ resistors, but that's only okay when you have proven that the noise is okay.

All for now. / Comments invited!
RAP / Robert A. Pease / Engineer
rap@galaxy.nsc.com—or:

Mail Stop D2597A
National Semiconductor
P.O. Box 58090
Santa Clara, CA 95052-8090

*For this case, grade a good number of 2N5486s into 20-mV bins of V_S, with V_{GD} = 7 V, and I_S = 3.8 mA. Take units out of the same bin for good matched pairs.

P.S. If you design in an op amp, try to avoid relying on nonguaranteed characteristics, such as noise, which is rarely guaranteed.

P.P.S. I neglected to mention that any resistor may have a built-in capacitance of 0.3 to 0.8 pF. If you add that to any imperfect layout, the capacitance could be so big that you wish it were smaller. Good layout and good engineering can easily cut the C to less than 0.2 pf. For example, make the feedback resistance out of three or four resistors in series, and install a shield land between the ends of the resistor. More later. /rap