Ph 77 ADVANCED PHYSICS LABORATORY
– ATOMIC AND OPTICAL PHYSICS –

Expt. 72 – Laser Frequency Stabilization

I. BACKGROUND

In many precision optical measurements, it is desirable to have a laser with a well-defined frequency. For example, atomic physics experiments often require lasers with frequencies fixed at or near the frequencies of atomic resonance lines. For tunable lasers, it is therefore necessary to have a means of controlling the laser's operating frequency, stabilizing it and often "locking" it at a desired value. In this lab we will investigate methods for achieving this, and describe one particularly powerful technique known as the Pound-Drever-Hall method (named after R. V. Pound, who first used the technique to stabilize microwave oscillators in the 1940's, and Caltech professor R. W. P. Drever, who extended the ideas to the optical domain in the early 1980's, and John Hall, who added important pieces of RF technology) (Drever *et al.* 1983).

Laser frequency locking is actually quite a complex topic, with a number of interesting subtleties. For example, one can both "frequency lock" and "phase lock" a laser. The former, frequency locking, means that the laser's average frequency is fixed, but its linewidth remains equal to the laser's intrinsic linewidth (for our diode lasers the linewidth is around 1 MHz). With phase-locking techniques, one can take a laser, such as one of our diode lasers, and precisely control the phase ϕ of the electric field, keeping $\phi(t)$ close to $\omega_0 t$, where ω_0 is the laser's desired angular frequency. If the phase is precisely controlled, this serves to *reduce* the laser's linewidth. Laser linewidths have been reduced to sub-Hz levels (so $\Delta\omega/\omega < 10^{-15}$) for periods of several hours using these techniques. In this lab we will not worry about phase locking, and look only at frequency locking. (An introduction to phase locking can be found in Hall and Zhu 1992.)

The topic of laser frequency locking is itself embedded in the much larger field of control theory, which is a branch of engineering that deals with methods for controlling physical systems. You may have heard some of the terminology in control theory, which includes words like "servosystems" and "servomechanisms," "feedback control systems," "feedback loops," and the like. The generic servosystem contains several parts: 1) the *system*, or *plant*, which is something we want to control; 2) an *actuator*, which changes the state of the system; 3) a *sensor*, which detects the state of the system, (usually relative to some *reference)*, and 4) a *controller.* The basic idea is that the controller reads the sensor, then drives the actuator in such a way to send the system to the desired state. If done well, the system will go to the desired state quickly, and then deviate from the desired state as little as possible. For the problem at hand, laser stabilization, the system will be the laser (which has an output frequency we want to control), the actuator is the laser PZT (which changes the laser's frequency), the sensor is some frequency-selective optical element (an atomic resonance cell, or an optical cavity), and the controller is some electronics box that connects the sensor signal with the actuator via a *control loop.*

Side Locking. One of the simplest control techniques is the side-locking method. One starts with some

frequency-selective optical element that produces a voltage signal, $V(\omega)$, as a function of laser frequency. If one wishes to lock the laser frequency at a frequency ω_0, and $dV/d\omega(\omega_0) \neq 0$, then ones subtracts a reference voltage to make an *error signal* $\varepsilon(\omega) = V(\omega) - V(\omega_0)$. This error signal then serves as input to a feedback loop, which adjusts the laser's frequency to make $\varepsilon = 0$. The side-locking method is useful if one wishes, for example, to lock the laser to the side of a peaked resonance feature. It only works, however, if $dV/d\omega(\omega_0) \neq 0$, which means one can lock to the side of a resonance feature, but not to the peak – hence, the name, side-locking.

One often wants to lock the laser to the *peak* of the resonance feature, so what then? The side-locking method would not work in this case, since $dV/d\omega(\omega_0) = 0$; a non-zero error signal $\varepsilon(\omega) = V(\omega) - V(\omega_0)$ would tell you that the laser frequency was not equal to ω_0, but it would be insufficient to say whether the frequency was too high or too low.

Dither Locking. One technique that does work in this circumstance is called dither locking. The idea is to sinusoidally modulate, or dither, the laser frequency at some frequency Ω, producing a voltage signal $V(t) = V(\omega(t)) \simeq V[\omega_{center} + \Delta\omega \cos(\Omega t)]$. If $\beta = \Delta\omega/\Omega \gg 1$, then the voltage signal behaves as if the laser frequency were slowly oscillating back and forth (see the RF modulation discussion in the previous handout).

A lock-in amplifier with reference frequency Ω can then produce an error signal $\varepsilon(\omega)$ that is the Fourier component of $V(t)$ at frequency Ω. Expanding $V(\omega) \approx V_0 - A(\omega - \omega_0)^2$ for ω near the peak frequency ω_0, we have

$$\begin{aligned} V(t) &= V[\omega_{center} + \Delta\omega \cos(\Omega t)] \\ &\simeq V(\omega_{center}) + \frac{dV}{d\omega}(\omega_{center}) \cdot \Delta\omega \cos(\Omega t) + \cdots \end{aligned}$$

and so the Fourier component is

$$\begin{aligned} \varepsilon(\omega) &\sim \frac{dV}{d\omega}(\omega) \cdot \Delta\omega \\ &\sim 2A\,\Delta\omega\,(\omega - \omega_0). \end{aligned}$$

This error signal has the desired properties that $\varepsilon(\omega_0) = 0$ and $d\varepsilon/d\omega(\omega_0) \neq 0$; thus it can be used in a feedback loop to lock the laser frequency at ω_0. (If this isn't clear, draw a picture of $V(\omega)$ near ω_0 and think about it. No matter what the form of $V(\omega)$, selecting the Fourier component of a dithered signal produces an error signal proportional to $dV/d\omega$.)

In the lab, you will set up a dither locking experiment that will lock the laser frequency to a resonance of the confocal cavity, and we'll go over the details of this experiment in the next section.

One problem with the dither-locking method is that the servo bandwidth (how fast the servo can control the system) is limited to a frequency much less than the dither frequency Ω, which in turn must be much less than the frequency scan $\Delta\omega$, and thus much less than the linewidth of the resonance feature on which one wishes to lock. For a very narrow resonance feature, dither locking may not work well. In general, if we want to control something well, we need to control it quickly – in the language of control theory, we need a servo with a large bandwidth.

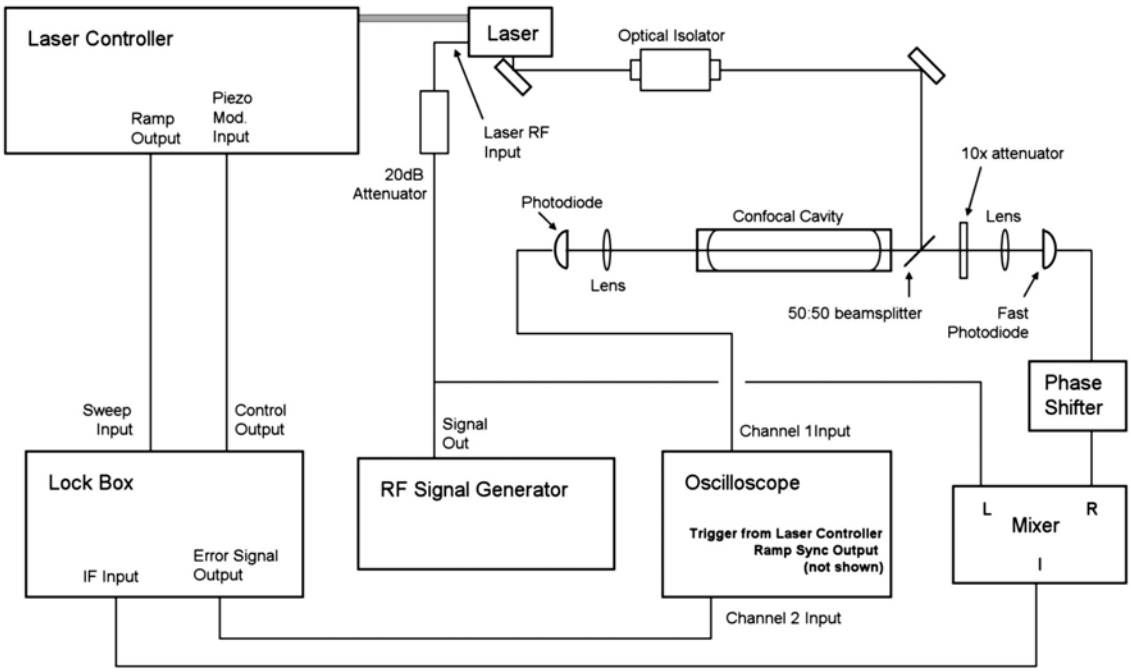

Figure 1. Set up for Pound-Drever-Hall laser locking.

The Pound-Drever-Hall method. We can increase the bandwidth of our dither-locking servo by increasing Ω, but we soon find ourselves in the regime where $\beta \ll 1$. When this happens, we are no longer dithering the frequency, but are modulating the phase of the electric field (again, see the RF modulation discussion in the previous handout). We are adding sidebands to the laser, and our overall picture of what's happening changes, just as it did with RF modulation. The new picture leads us to the Pound-Drever-Hall method.

The P-D-H layout you will use in the lab is shown in Figure 1. It may look complicated at first, but it's fairly simple once you break it down and examine the different pieces:

1. The optics are straightforward, essentially what you set up before. The transmitted signal shows the usual cavity transmission peaks when the laser frequency is scanned. Your goal is to lock the laser to one of the cavity modes, so the transmitted signal stays at the peak value. The transmitted signal is only used as a diagnostic in this lab. When the laser is scanned, it shows that the cavity is properly aligned, and it shows the RF sidebands on the laser. When the laser is locked, you should see a high transmitted signal. This signal is not used as part of the control loop.

2. The RF Signal Generator is used to modulate the laser frequency and thus add sidebands. This is the same as in the previous lab, except the signal is attenuated 100x by a 20dB attenuator. (This is because the mixer needs a large RF amplitude while the laser needs a much smaller RF amplitude.) To set the

notation, we assume two weak sidebands at frequencies $\omega \pm \Omega$ around the laser's central carrier frequency at ω. The modulation frequency Ω is chosen to be slightly greater than the width of the cavity transmission peaks.

3. The Lock Box is some electronics to feed the error signal back to the laser. We'll get to that below, after we look first at how the error signal is produced.

4. The beam reflected off the cavity is sent to a fast photodiode that has a 50 MHz response. For the P-D-H method to work, the bandwidth of the fast photodiode must be higher than the sideband frequency, which will be around 20 MHz. (The other photodiodes only go up to about 1 MHz.) The output from this photodiode will be used to generate the error signal.

5. The Phase Shifter is nothing more than a box of selectable cables of different lengths. Sending the signal through a cable adds a delay, which is the same as adding a phase shift. The longer the cable, the larger the phase shift.

6. The Mixer (or Double-Balanced Mixer) is another passive element, consisting of a set of transformers and diodes. How it actually works is not trivial, so we're going to skip over that part (you can look it up if you're curious). In a nutshell, the mixer takes two inputs – the Local Oscillator (L) and the RF signal (R) – and essentially multiplies them together. The L and R signals are both RF signals with the same frequency but different phases. The output (I) is then proportional to

$$\cos(\Omega t)\cos(\Omega t + \eta) = \frac{1}{2}\cos(\delta) + \frac{1}{2}\cos(2\Omega t + \eta)$$

which has terms at DC and at 2Ω. The mixer output is sent to the Lock Box, where it is smoothed to remove the high-frequency term, and the smoothed output becomes the error signal.

Putting the pieces together, the error signal can be calculated with the following analysis, adapted from Bjorklund *et al* (1983). The reflection amplitude from a Fabry-Perot cavity is given by (see the previous handout):

$$E_r = \frac{(1 - e^{i\delta})\sqrt{R}}{1 - R\,e^{i\delta}} E_i \equiv A[\delta, R] E_i$$

where R is the mirror reflectivity and $\delta = 2\pi\Delta\omega/\Delta\omega_{FSR}$. You've already seen this in the previous lab, so it should look familiar; a plot of the reflected intensity $|E_r|^2$ shows dips at the cavity resonance (the transmitted light shows peaks, so therefore the reflected light has dips).

For $\beta \ll 1$ we can write the input electric field as a carrier and two weak sidebands

$$E_i = E_0 \left\{ -M e^{i(\omega-\Omega)t} + e^{i\omega t} + M e^{i(\omega+\Omega)t} \right\}$$

which gives the reflected amplitude

$$E_r = E_0 e^{i\omega t} \left\{ -M A_- e^{-i\Omega t} + A_0 + M A_+ e^{i\Omega t} \right\}$$

where

$$\begin{aligned} A_0 &= A[2\pi\Delta\omega/\Delta\omega_{FSR}, R] \\ \text{and } A_\pm &= A[2\pi(\Delta\omega \pm \Omega)/\Delta\omega_{FSR}, R]. \end{aligned}$$

The reflected signal at the photodiode is then $I_{phot} \propto |E_r|^2$, which you can verify is a real function containing DC terms and terms proportional to $\cos(\Omega t)$ and $\sin(\Omega t)$. Because I_{phot} is a real function, the

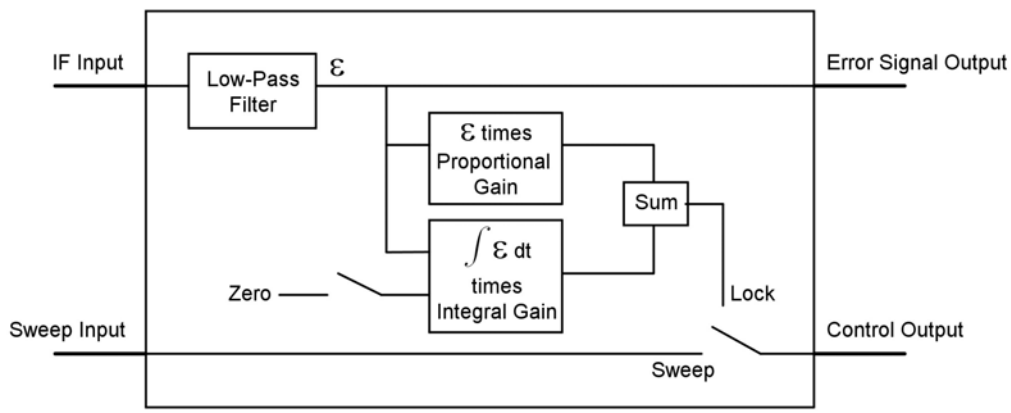

Figure 2. Schematic of the Lock Box.

effect of the mixer is essentially to multiply the photodiode signal by $\exp(i\Omega t + \zeta)$ and take the real part, where ζ is a constant phase factor depending on the relative phase of the photodiode and local oscillator signals at the mixer. Low-pass filtering gives the error signal

$$\varepsilon \propto \mathrm{Re}\left\{e^{i\zeta}(A_0^* A_+ - A_0 A_-^*)\right\}.$$

Problem 1. Calculate and plot $\varepsilon(\Delta\nu)$, where $\Delta\nu = \nu - \nu_0$ in MHz, for a variety of values for the phase factor ζ, assuming a sideband frequency $\Omega/2\pi = 18$ MHz, $\Delta\nu_{FSR} = 375$ MHz, and a cavity reflectivity of $R = 0.95$. (Hint: see the error signal in Figure 4 below.) In the lab, you will generate $\varepsilon(\Delta\nu)$ as the laser frequency is slowly scanned around a cavity resonance and display it on the oscilloscope. If all goes well, the display should match your calculations. You can change ζ using the Phase Shifter, and you should be able to reproduce your calculations at different ζ.

The Lock Box. Figure 2 shows the essential elements of the Lock Box, which is used to feed the error signal back to the laser piezo modulation input, thus controlling the laser frequency and keeping it locked on a cavity resonance. Again, there's quite a bit going on here, so we need to break it down:

1) The IF input is where the error signal enters the Lock Box. Since the mixer output has a substantial 2Ω component, this signal is run through a low-pass filter and an amplifier to clean it up. The result is the smoothed error signal $\varepsilon(t)$. This goes to the Error Signal Output so it can be monitored on the oscilloscope.

2) A ramp signal goes to the Sweep Input (see Figure 1). If the switch on the Lock Box is set to Sweep, then the Sweep Input goes directly to the Control Output. Assuming a triangle-wave input, this simply sweeps the laser frequency back and forth as usual. With the switch set to Sweep, you can view the transmitted light $I_{trans}(\Delta\nu)$ and error signal $\varepsilon(\nu)$ on the oscilloscope to make sure everything looks okay.

3) When the switch is set to Lock, the sweep signal is disconnected. Then the laser frequency is not

sweeping, but the control circuitry will try to lock the laser.

4) Start with the Zero switch closed, which essentially removes the Integral Gain part of the control loop. Then the error signal is simply multiplied by the Proportional Gain and sent to the Control Output; i.e. $C = G_{prop}\varepsilon$, where C is the Control Output signal and G_{prop} is the proportional gain.

If ε is nonzero, then the control signal is nonzero also, and this changes the laser frequency. If the signs are right, then the control will tend to push the laser frequency back toward the cavity resonance. If the gain G_{prop} is very large, then even a small ε will produce a large C, and the laser will stay close to the locked position, as desired. If G_{prop} is not so large (which will be our case in the lab), then the servo will bring the laser closer to the correct frequency, but it will not be able to compensate for large deviations.

One problem with this is that one cannot make G_{prop} very large or the servo will have problems. There are always some phase lags in the electronics and in the mechanical elements, and if the lags are large enough they will result in an instability, causing the servo to oscillate. The theory behind these instabilities can be quite complicated, and we cannot go through it all here. Suffice it to say that problems arise if you increase G_{prop} too much and try to push the servo too hard.

But if G_{prop} is not so large, then the servo will not lock very well. In general, the laser will need a nonzero C to keep the frequency on resonance, and this means ε must be nonzero also, because $C = G_{prop}\varepsilon$. We want to have $\varepsilon = 0$, and in general this cannot happen using proportional gain alone.

5) The next step is to switch in the integral gain, which gives

$$C(t) = G_{prop}\varepsilon(t) + G_{int}\int_{t_0}^{t} \varepsilon(t')dt'$$

where G_{int} is the integral gain and t_0 is the time the zero switch is opened. Now things get interesting, as the servo has some memory. As long as ε is nonzero, the second term in the above will grow (or diminish, if $\varepsilon < 0$). This means C grows also with time. If the signs are right, this will make C grow until ε goes smoothly to zero. And from that point C will remain unchanged as long as ε remains equal to zero. This effective memory means that the Lock Box can produce a nonzero C even when $\varepsilon = 0$. The result, which you will see in the lab, is a better lock.

Finally, as a disclaimer, we note that servo control theory is a very rich field, vastly more than the above quick explanation might suggest. There is a whole branch of engineering devoted to the topic (at Caltech, this is CDS – Control and Dynamical Systems). Control loops are often quite complex, and their analysis is certainly nontrivial. The above discussion only scratches the surface of this field slightly, but it's enough to demonstrate some of the basics in the lab.

II. LABORATORY EXERCISES

The Thermal Expansion of Aluminum. As an amusing warm-up exercise to this lab, use the extreme sensitivity of these optical techniques to measure the thermal expansion coefficient of aluminum.. You can do this by heating the cavity slightly and watching the cavity expand using the optical spectrum analyzer. You can heat the cavity nicely by touching it, and a thermistor is mounted to the cavity tube to

monitor its temperature.

First set up the cavity as shown in Figure 3 below, with the cavity transmission spectrum displayed on the oscilloscope as the laser frequency is swept. Once the system has stabilized, you should not see much drift of the cavity transmission peaks on the 'scope. Then use a digital ohmmeter to measure the resistance of the thermistor on the cavity. The two wires coming from of the body of the cavity are the thermistor leads, and the resistance should be about 2000 Ohms at room temperature. The thermistor resistance changes by about 4.5 percent for each degree of temperature change, with the resistance decreasing as the temperature increases.

The basic idea of this measurement is to count the cavity peaks that drift by on the 'scope as the cavity temperature is changed. One obvious approach to this would be to stabilize the system at one cavity temperature, then change the temperature to some new stable value while counting peaks. The problem with this direct approach is that it takes a long time for the cavity temperature to stabilize.

[Show that the stabilization time scale for the cavity temperature is approximately $\tau \approx CL^2\rho/\kappa$, where C is the specific heat of the aluminum cavity tube, ρ is the density of aluminum, κ is the thermal conductivity, and L is the approximate size scale of the cavity. With $C \approx 900$ J/kg-K, $\kappa \approx 250$ W/m-K, $\rho \approx 2700$ kg/m^3, and $L \approx 0.1$ m, we have $\tau \approx 100$ seconds. And of course one would have to wait several times τ before the temperature is stable.]

A better approach is to simultaneously monitor the time derivatives of the cavity length and the cavity temperature. Starting from a fairly stable temperature, heat the cavity slowly by touching it gently. Try to heat it uniformly along its length (because of the L^2 in the above expression). Record the thermistor resistance while you count cavity peaks going by on the 'scope. This is best done with two people. Make a measurement with the cavity heating, and another with the cavity cooling. From this you can extract the cavity length as a function of cavity temperature, and thus the thermal expansion coefficient. Report your measurement in relative expansion per degree, which should be around 20×10^{-6} K^{-1}.

Laser Locking I - Dither Locking.

The next step is to lock the laser frequency to the confocal cavity using the dither locking technique described in Section I. The optics and electronics set-up is shown in Figure 3. This is quite different from the P-D-H set-up, so let's look at the different pieces:

1. The optics are again familiar, and somewhat simpler than the P-D-H set-up, since you will be using the transmitted signal to generate an error signal. To make life easier, adjust the cavity length and/or the input beam alignment so the cavity has an effective finesse of about 10. The low finesse makes the lock work much better. With a higher finesse, the frequency jitter in the cavity and/or laser makes the cavity output very jumpy. The feedback control won't be fast enough to take out the high-frequency jitter (which comes from acoustic and/or seismic effects), and the resulting noise will quickly throw the servo out of lock. With a low finesse, the jitter doesn't show up much on the cavity output, so the servo only has to compensate for low-frequency drift. To make the servo work with a high cavity finesse, we would probably have to reduce the jitter passively first (with acoustic damping, seismic isolation, etc.).

2. You sweep the laser frequency using the Ramp Generator and Piezo Controller on the Laser Controller,

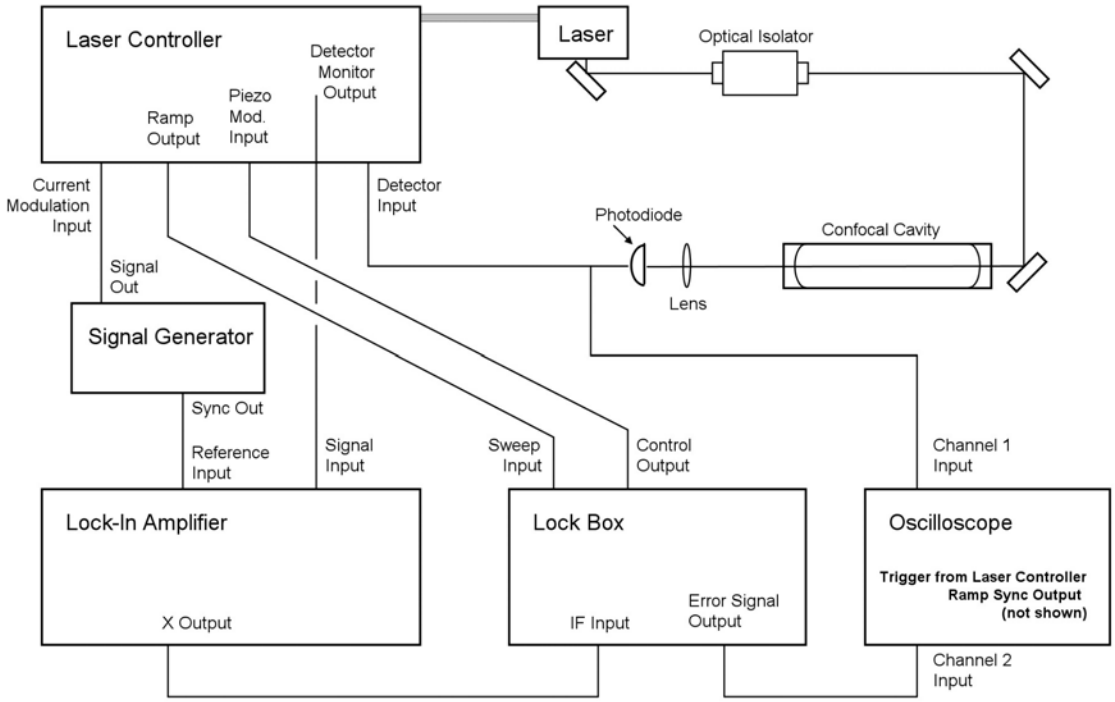

Figure 3. Set-up for dither locking.

basically the same as before. (You should know what this is doing by now. If not, ask your TA to remind you.) The only difference is that the signal is diverted into the Lock Box. When the switch on the Lock Box is set to Sweep, then the Sweep Input goes directly to the Control Output. By setting up these two cables, the laser frequency will scan, as you have done before. Things seem to work best if you turn the Ramp Generator to full amplitude, then turn the Piezo Modulation gain down (both on the Laser Controller) so you are only scanning over a couple of transmission peaks.

3. The photodiode output goes directly to the oscilloscope, so you can monitor the light transmitted by the cavity.

4. Use a BNC tee to also connect the photodiode output to the Detector Input on the Laser Controller. This way you have a separate detector gain control, which will be useful.

5. Configure the Signal Generator to produce a sine wave at 50 kHz; turn the gain down low and make sure the output is set to the 0-2V position. This goes to the Current Modulation Input, which will dither the laser frequency. If you turn the amplitude up, you should be able to see the dither directly on the photodiode output. If this doesn't make sense, or you're not sure what to look for, consult your TA.

6. Now for the error signal. First send the Sync Output from the Signal Generator to the Reference Input of the Lock-In Amplifier. The Lock-In will automatically sense the reference signal and lock onto it. You should see the reference frequency on the Lock-In display. (If not, ask your TA.)

7. Next send the Detector Monitor Output to the Lock-In Signal Input. To begin at least, set the

Lock-In time constant to 100 μsec and the sensitivity to 1 V, and set the input filter to 12dB. With this, the Lock-In will do its thing (ask your TA if you don't now how a lock-in amplifier works) and detect the Fourier component of the input signal at the reference frequency. That is, it multiplies the input signal by $\sin(\omega t)$, integrates for the selected time constant, and sends the result to its X output. (It also multiplies by $\cos(\omega t)$ and sends the result to the Y output, but you won't be using that today.)

8. Then send the X output to the IF input of the Lock Box, and send the Error Signal Output to the Oscilloscope as shown. All the Lock Box does at this point is smooth the signal a bit, so basically you're sending X to the oscilloscope. With this, you should see a sensible error signal on the 'scope. Its shape will be essentially the derivative of the cavity output signal, because that's what the dither-lock set-up does. (Why? See Section I and think about it.) Show your TA the signals before proceeding. It usually takes some tweaking (maybe a lot of tweaking) to get everything looking good before you attempt to lock the laser frequency to the cavity. The goal is to produce a clean, smooth error signal that is about a one volt peak-to-peak.

Check the Phase of the Error Signal. Once the error signal looks good, change the reference phase using the Lock-In controls (ask your TA for details.) Changing the phase from zero means the Lock-In multiplies by $\sin(\omega t + \delta)$ instead of $\sin(\omega t)$, where δ is the Lock-In phase. If you change the phase from 0 to 180 degrees, the error signal should change sign. Check this.

The Lock Box looks at the error signal and feeds a signal back (the Control Output) to the Laser Controller. When you move the switch to Lock, it will try to control the laser frequency and zero the error signal. For this to work, the error signal has to have the correct sign, and you change this using the Lock-In phase. Which sign is the right sign? You just have to try it to see.

Flip the switch on the Lock Box to Lock, while the Integral Gain is off. This means you are only using Proportional Gain. Adjust the Piezo Controller DC Offset by hand to see if you can get the servo to lock right, remembering that the servo has only a small dynamic range (that is, the Control Output signal is fairly small). When you're close to the cavity resonance, you should see the error signal go to zero while the transmitted light signal is near its maximum value. If the sign is right, the error signal will dwell around zero. If the sign is wrong, the error signal will tend to avoid zero. The various nuances of the behavior are hard to explain, so consult your TA at this point.

Try this exercise using both signs of the error signal. You should find that the cavity achieves a weak lock with one sign, but not with the other.

Lock the Laser. Once you know which sign is the right sign, lock the laser again using only Proportional Gain. Keep one hand on the DC offset knob to keep the lock centered so the error signal remains close to zero (while the cavity transmission is near maximum). You are then part of a human servo as you watch the 'scope and feed back to the DC offset knob. If all is well, you can make this work, sort of. Then have your partner place a hand a few centimeters away from the cavity, heating it slightly. (Don't touch the cavity; that's too much heating.) You should see that this induced cavity drift makes it much more difficult for the human servo to keep the laser on resonance.

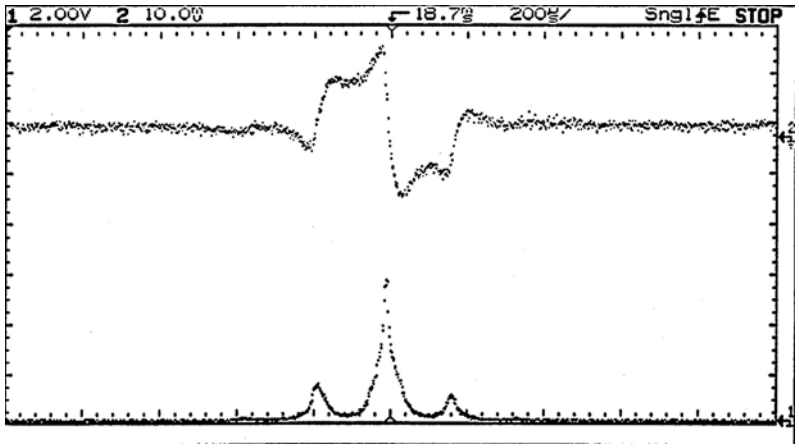

Figure 4. The top trace shows a typical error signal; the bottom trace shows the cavity transmission, taken at the same time.

Take the hand away, get the lock working again, and then switch on the Integral Gain. The cavity should remain locked, and the human servo should no longer be necessary to keep the lock centered. Try heating the cavity again. The lock should stay centered. How does this magic work? Read the explanation in Section I again, and think about it. **Print a copy of the error signal and transmitted signal with the laser locked, and again with the laser scanning.** With the full lock on, clap your hands or tap lightly on the table to test the lock stability. What happens if you change the DC offset on the piezo controller (with the lock on)?

Laser Locking II - The Pound-Drever Hall Method.

Once you are satisfied with your dither lock, you can move on to the Pound-Drever-Hall lock, shown in Figure 1. The basic idea is the same as with dither locking, except the method for generating the error signal is quite different. To begin, use 18 MHz sidebands, the same as you used for your calculations. When you have everything set up, you should see an error signal that looks something like that shown in Figure 4. If not, consult your TA if you need help debugging the set-up. One common problem is saturating the fast photodiode, and you will probably have to attenuate the light hitting the sensor, as shown in Figure 1.

Once you see something reasonable, change the phase using the Phase Shifter, and try changing the RF frequency. **Print out several error signal traces at different phases and frequencies. In particular, try to generate signals that match your calculations.**

When you set out to lock the laser, be sure to trim the DC level of the error signal. Far from resonance, the error signal should be at zero volts. **Print out traces with the laser locked and unlocked.**

III. REFERENCES

Bjorklund, G. C. *et al.* 1983, "Frequency-modulation (FM) Spectroscopy–Theory of Line Shapes and Signal-to-noise Analysis," Appl. Phys. B 31, 145.

Drever, R. W. P. *et al.* 1983, "Laser Phase and Frequency Stabilization Using an Optical Resonator," Appl. Phys. B 31, 97.

Hall, J. L. and Zhu, M. 1992, "An Introduction to Phase-Stable Optical Sources," *Proceedings of the International School of Physics – Enrico Fermi*, edited by E. Arimondo, W. D. Phillips, and F. Strumia (North-Holland).

Yariv, A. 1991, *Optical Electronics*, Saunders College Publishing, 4th ed.

An introduction to Pound–Drever–Hall laser frequency stabilization

Eric D. Black
LIGO Project, California Institute of Technology, Mail Code 264-33, Pasadena, California 91125

(Received 3 January 2000; accepted 4 April 2000)

This paper is an introduction to an elegant and powerful technique in modern optics: Pound–Drever–Hall laser frequency stabilization. This introduction is primarily meant to be conceptual, but it includes enough quantitative detail to allow the reader to immediately design a real setup, suitable for research or industrial application. The intended audience is both the researcher learning the technique for the first time and the teacher who wants to cover modern laser locking in an upper-level physics or electrical engineering course. © *2001 American Association of Physics Teachers.*
[DOI: 10.1119/1.1286663]

I. INTRODUCTION

Pound–Drever–Hall laser frequency stabilization is a powerful technique for improving an existing laser's frequency stability,[1,2] and it is an essential part of the technology of interferometric gravitational-wave detectors.[3] The technique has been used to demonstrate, using a commercial laser, a frequency standard as relatively stable as a pulsar.[4,5]

The physical basis of the Pound–Drever–Hall technique has a broad range of applications in addition to gravitational-wave detection. A closely related technique is employed in atomic physics, where it goes by the name frequency-modulation (fm) spectroscopy and is used for probing optical resonances. (See, for example, Refs. 6–8. Both techniques are similar to an older method used in microwave applications, invented in the forties by R. V. Pound.[9]) The conceptual foundations of fm spectroscopy and Pound–Drever–Hall laser locking are quite similar. If you can understand one, you will have a good handle on the other.

The idea behind the Pound–Drever–Hall method is simple in principle: A laser's frequency is measured with a Fabry–Perot cavity, and this measurement is fed back to the laser to suppress frequency fluctuations. The measurement is made using a form of nulled lock-in detection, which decouples the frequency measurement from the laser's intensity. An additional benefit of this method is that the system is not limited by the response time of the Fabry–Perot cavity. You can measure, and suppress, frequency fluctuations that occur faster than the cavity can respond.

The technique is both simple and powerful; it can be taught in an advanced undergraduate laboratory course.[10] It is my hope that this paper will provide a clear conceptual introduction to the Pound–Drever–Hall method. I am going to try and demonstrate both the physical basis of the technique and its fundamental limitations. I also hope that a more widespread understanding of the technique will stimulate further development of laser frequency stabilization (and perhaps fm spectroscopy) in general.

In this paper I am going to focus on the frequency measurement, also called the *error signal*. That is really the heart of the technique, and it is often the point of maximum confusion when one first encounters it. The frequency measurement is also an essential part of fm spectroscopy, and a good understanding of it will get the reader off to a good start in that field as well.

In this paper I will assume that the reader is already familiar with Fabry–Perot cavities as they would be covered in a good introductory optics course. (See, for example, Refs. 11 and 12.) For some very good comprehensive introductory materials on both control theory and Fabry–Perot cavities, see Refs. 13–17. An excellent introduction to interferometric gravitational-wave detectors is Ref. 18.

II. A CONCEPTUAL MODEL

Suppose we have a laser that we want to use for some experiment, but we need better frequency stability than the laser provides ''out of the box.'' Many modern lasers are tuneable: They come with some input port into which you can feed an electrical signal and adjust the output frequency. If we have an accurate way to measure the laser's frequency, then we can feed this measurement into the tuning port, with appropriate amplification and filtering, to hold the frequency (roughly) constant.

One good way to measure the frequency of a laser's beam is to send it into a Fabry–Perot cavity and look at what gets transmitted (or reflected). Recall that light can only pass through a Fabry–Perot cavity if twice the length of the cavity is equal to an integer number of wavelengths of the light. Another way to say this is that the frequency of the light's electromagnetic wave must be an integer number times the cavity's free spectral range $\Delta\nu_{\mathrm{fsr}} \equiv c/2L$, where L is the length of the cavity and c is the speed of light. The cavity acts as a filter, with transmission lines, or resonances, spaced evenly in frequency every free spectral range. Figure 1 shows a plot of the fraction of light transmitted through a Fabry–Perot cavity versus the frequency of the light.

If we were to operate just to one side of one of these resonances, but near enough that some light gets transmitted (say, half the maximum transmitted power), then a small change in laser frequency would produce a proportional change in the transmitted intensity. We could then measure the transmitted intensity of the light and feed this signal back to the laser to hold this intensity (and hence the laser frequency) constant.

This was often how laser locking was done before the development of the Pound–Drever–Hall method, and it suffers from a few flaws, one of which is that the system cannot distinguish between fluctuations in the laser's frequency, which changes the intensity transmitted through the cavity, and fluctuations in the intensity of the laser itself.

We could build a separate system to stabilize the laser's intensity, which was done with some success in the early seventies,[19] but a better method would be to measure the reflected intensity and hold that at zero, which would decouple intensity and frequency noise. The only problem with this scheme is that the intensity of the reflected beam is symmetric about resonance. If the laser frequency drifts out of

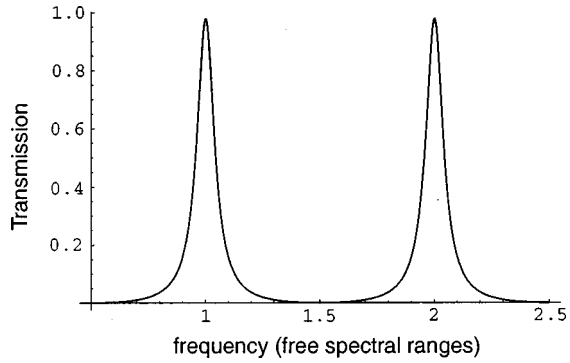

Fig. 1. Transmission of a Fabry–Perot cavity vs frequency of the incident light. This cavity has a fairly low finesse, about 12, to make the structure of the transmission lines easy to see.

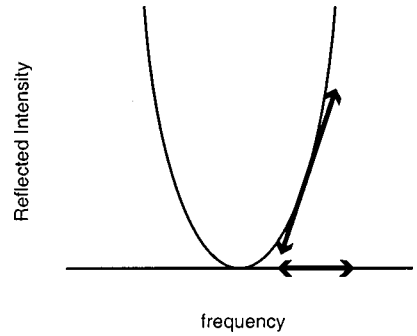

Fig. 2. The reflected light intensity from a Fabry–Perot cavity as a function of laser frequency, near resonance. If you modulate the laser frequency, you can tell which side of resonance you are on by how the reflected power changes.

resonance with the cavity, you can't tell just by looking at the reflected intensity whether the frequency needs to be increased or decreased to bring it back onto resonance. The derivative of the reflected intensity, however, is antisymmetric about resonance. If we were to measure this derivative, we would have an error signal that we can use to lock the laser. Fortunately, this is not too hard to do: We can just vary the frequency a little bit and see how the reflected beam responds.

Above resonance, the derivative of the reflected intensity with respect to laser frequency is positive. If we vary the laser's frequency sinusoidally over a small range, then the reflected intensity will also vary sinusoidally, in phase with the variation in frequency. (See Fig. 2.)

Below resonance, this derivative is negative. Here the reflected intensity will vary 180° out of phase from the frequency. On resonance the reflected intensity is at a minimum, and a small frequency variation will produce no change in the reflected intensity.

By comparing the variation in the reflected intensity with the frequency variation we can tell which side of resonance we are on. Once we have a measure of the derivative of the reflected intensity with respect to frequency, we can feed this measurement back to the laser to hold it on resonance. The purpose of the Pound–Drever–Hall method is to do just this. Figure 3 shows a basic setup. Here the frequency is modulated with a Pockels cell,[20] driven by some local oscillator. The reflected beam is picked off with an optical isolator (a polarizing beamsplitter and a quarter-wave plate makes a good isolator) and sent into a photodetector, whose output is compared with the local oscillator's signal via a mixer. We can think of a mixer as a device whose output is the product of its inputs, so this output will contain signals at both dc (or very low frequency) and twice the modulation frequency. It is the low frequency signal that we are interested in, since that is what will tell us the derivative of the reflected intensity. A low-pass filter on the output of the mixer isolates this low frequency signal, which then goes through a servo amplifier and into the tuning port on the laser, locking the laser to the cavity.

The Faraday isolator shown in Fig. 3 keeps the reflected beam from getting back into the laser and destabilizing it. This isolator is not necessary for understanding the technique, but it is essential in a real system. In practice, the small amount of reflected beam that gets through the optical isolator is usually enough to destabilize the laser. Similarly, the phase shifter is not essential in an ideal system but is useful in practice to compensate for unequal delays in the two signal paths. (In our example, it could just as easily go between the local oscillator and the Pockels cell.)

This conceptual model is really only valid if you are dithering the laser frequency slowly. If you dither the frequency too fast, the light resonating inside the cavity won't have time to completely build up or settle down, and the output will not follow the curve shown in Fig. 2. However, the technique still works at higher modulation frequencies, and both the noise performance and bandwidth of the servo are typically improved. Before we address a conceptual picture that does apply to the high-frequency regime, we must establish a quantitative model.

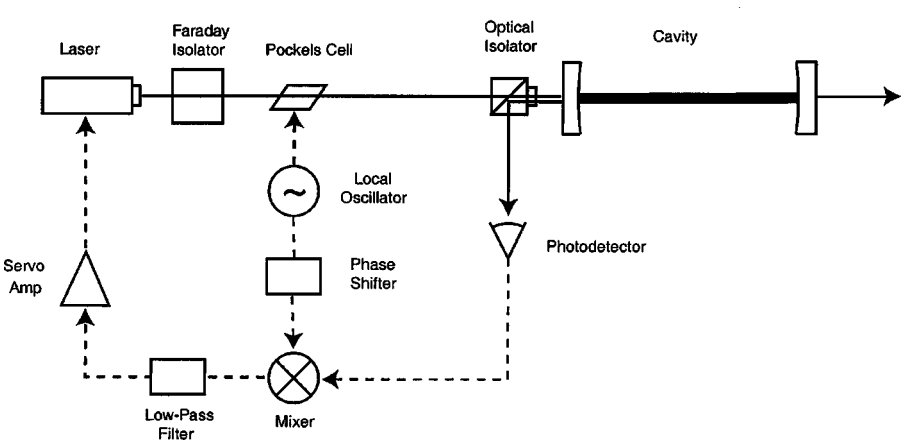

Fig. 3. The basic layout for locking a cavity to a laser. Solid lines are optical paths and dashed lines are signal paths. The signal going to the laser controls its frequency.

III. A QUANTITATIVE MODEL

A. Reflection of a monochromatic beam from a Fabry–Perot cavity

To describe the behavior of the reflected beam quantitatively, we can pick a point outside the cavity and measure the electric field over time. The magnitude of the electric field of the incident beam can be written

$$E_{\text{inc}}=E_0 e^{i\omega t}.$$

The electric field of the reflected beam (measured at the same point) is

$$E_{\text{ref}}=E_1 e^{i\omega t}.$$

We account for the relative phase between the two waves by letting E_0 and E_1 be complex. The *reflection coefficient* $F(\omega)$ is the ratio of E_{ref} and E_{inc}, and for a symmetric cavity with no losses it is given by

$$F(\omega)=E_{\text{ref}}/E_{\text{inc}}=\frac{r\left(\exp\left(i\dfrac{\omega}{\Delta\nu_{\text{fsr}}}\right)-1\right)}{1-r^2\exp\left(i\dfrac{\omega}{\Delta\nu_{\text{fsr}}}\right)}, \tag{3.1}$$

where r is the amplitude reflection coefficient of each mirror, and $\Delta\nu_{\text{fsr}}=c/2L$ is the free spectral range of the cavity of length L.

The beam that reflects from a Fabry–Perot cavity is actually the coherent sum of two different beams: the *promptly reflected beam*, which bounces off the first mirror and never enters the cavity; and a *leakage beam*, which is the small part of the standing wave inside the cavity that leaks back through the first mirror, which is never perfectly reflecting. These two beams have the same frequency, and near resonance (for our lossless, symmetric cavity) their intensities are almost the same as well. Their relative phase, however, depends strongly on the frequency of the laser beam.

If the cavity is resonating perfectly, i.e., the laser's frequency is exactly an integer multiple of the cavity's free spectral range, then the promptly reflected beam and the leakage beam have the same amplitude and are exactly 180° out of phase. In this case the two beams interfere destructively, and the total reflected beam vanishes.

If the cavity is not quite perfectly resonant, that is, the laser's frequency is not exactly an integer multiple of the free spectral range but close enough to build up a standing wave, then the phase difference between the two beams will not be exactly 180°, and they will not completely cancel each other out. (Their intensities will still be about the same.) Some light gets reflected off the cavity, and its phase tells you which side of resonance your laser is on. Figure 4 shows a plot of the intensity and phase of the reflection coefficient around resonance.

We will find it useful to look at the properties of $F(\omega)$ in the complex plane. (See Fig. 5.) It is not too hard to show (see Appendix A) that the value of F always lies on a circle in the complex plane, centered on the real axis, with ω being the parameter that determines where on this circle F will be. $|F(\omega)|^2$ gives the intensity of the reflected beam, and it is given by the familiar Airy function. F is symmetric around resonance, but its phase is different depending on whether the laser's frequency is above or below the cavity's resonance. As ω increases, F advances counterclockwise around the circle. For the symmetric, lossless cavity we are considering, this circle intersects the origin, with $F=0$ on resonance. Very near resonance, F is nearly on the imaginary axis, being in the lower half plane below resonance and in the upper half plane above resonance.

We will use this graphical representation of F in the complex plane when we try to understand the results of our quantitative model.

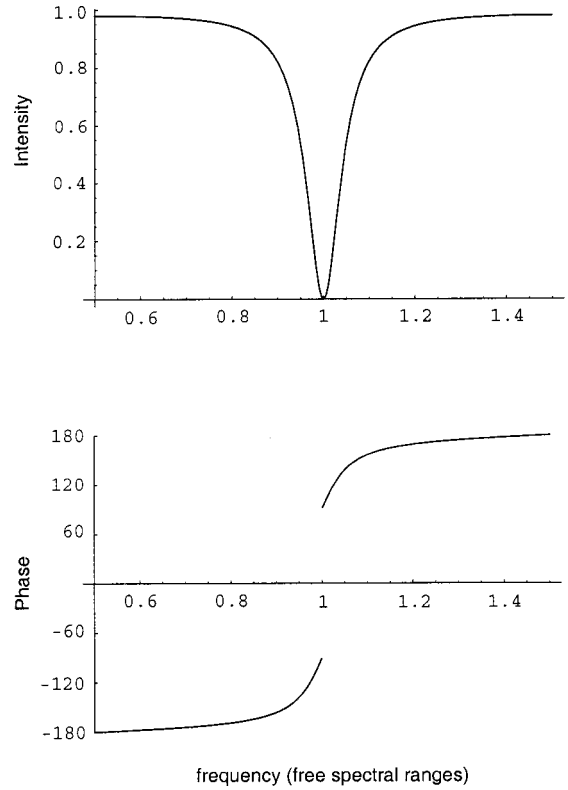

Fig. 4. Magnitude and phase of the reflection coefficient for a Fabry–Perot cavity. As in Fig. 1, the finesse is about 12. Note the discontinuity in phase, caused by the reflected power vanishing at resonance.

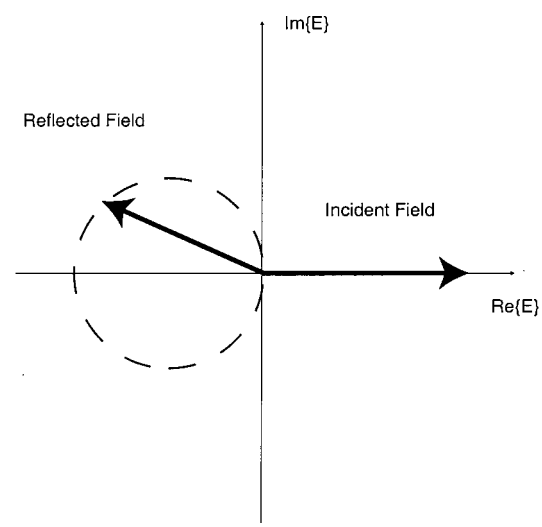

Fig. 5. The reflection coefficient in the complex plane. As the laser frequency (or equivalently, the cavity length) increases, $F(\omega)$ traces out a circle (counterclockwise). Most of the time, F is near the real axis at the left edge of the circle. Only near resonance does the imaginary part of F become appreciable. Exactly on resonance, F is zero.

B. Measuring the phase of the reflected beam

To tell whether the laser's frequency is above or below the cavity resonance, we need to measure the phase of the reflected beam. We do not, as of this writing, know how to build electronics that can directly measure the electric field (and hence the phase) of a light wave, but the Pound–Drever–Hall method (and fm spectroscopy) provides us with a way of indirectly measuring the phase.

Our conceptual model suggests that if we dither the frequency of the laser, that will give us enough information to tell which side of resonance we are on. A more quantitative way of thinking about this frequency dither is this: Modulating the laser's frequency (or phase) will generate sidebands with a definite phase relationship to the incident and reflected beams. These sidebands will not be at the same frequency as the incident and reflected beams, but a definite phase relation will be there nonetheless. If we interfere these sidebands with the reflected beam, the sum will display a beat pattern at the modulation frequency, and we can measure the phase of this beat pattern. The phase of this beat pattern will tell us the phase of the reflected beam. The sidebands effectively set a phase standard with which we can measure the phase of the reflected beam.

C. Modulating the beam: Sidebands

I talked about varying the frequency of this beam in the qualitative model, but in practice it is easier to modulate the phase. The results are essentially the same, but the math that describes phase modulation is simpler than the math for frequency modulation. Phase modulation is also easy to implement with a Pockels cell, as shown in Fig. 3. After the beam has passed through the Pockels cell, its electric field has its phase modulated and becomes

$$E_{\text{inc}}=E_0 e^{i(\omega t+\beta \sin \Omega t)}.$$

We can expand this expression, using Bessel functions, to[21]

$$\begin{aligned} E_{\text{inc}} &\approx [J_0(\beta)+2iJ_1(\beta)\sin \Omega t]e^{i\omega t} \\ &= E_0[J_0(\beta)e^{i\omega t}+J_1(\beta)e^{i(\omega+\Omega)t}-J_1(\beta)e^{i(\omega-\Omega)t}]. \end{aligned} \tag{3.2}$$

I have written it in this form to show that there are actually three different beams incident on the cavity: a carrier, with (angular) frequency ω, and two sidebands with frequencies $\omega\pm\Omega$. Here, Ω is the phase modulation frequency and β is known as the modulation depth. If $P_0\equiv|E_0|^2$ is the total power in the incident beam, then the power in the carrier is (neglecting interference effects for now)

$$P_c=J_0^2(\beta)P_0,$$

and the power in each first-order sideband is

$$P_s=J_1^2(\beta)P_0.$$

When the modulation depth is small ($\beta<1$), almost all of the power is in the carrier and the first-order sidebands,

$$P_c+2P_s\approx P_0.$$

D. Reflection of a modulated beam: The error signal

To calculate the reflected beam's field when there are several incident beams, we can treat each beam independently and multiply each one by the reflection coefficient at the appropriate frequency. In the Pound–Drever–Hall setup, where we have a carrier and two sidebands, the total reflected beam is

$$\begin{aligned} E_{\text{ref}}=E_0[&F(\omega)J_0(\beta)e^{i\omega t}+F(\omega+\Omega)J_1(\beta)e^{i(\omega+\Omega)t} \\ &-F(\omega-\Omega)J_1(\beta)e^{i(\omega-\Omega)t}]. \end{aligned}$$

What we really want is the power in the reflected beam, since that is what we measure with the photodetector. This is just $P_{\text{ref}}=|E_{\text{ref}}|^2$, or after some algebra

$$\begin{aligned} P_{\text{ref}}=&P_c|F(\omega)|^2+P_s\{|F(\omega+\Omega)|^2+|F(\omega-\Omega)|^2\} \\ &+2\sqrt{P_cP_s}\{\text{Re}[F(\omega)F^*(\omega+\Omega) \\ &-F^*(\omega)F(\omega-\Omega)]\cos \Omega t+\text{Im}[F(\omega)F^*(\omega+\Omega) \\ &-F^*(\omega)F(\omega-\Omega)]\sin \Omega t\}+(2\Omega \text{ terms}). \end{aligned} \tag{3.3}$$

We have added three waves of different frequencies, the carrier, at ω, and the upper and lower sidebands at $\omega\pm\Omega$. The result is a wave with a nominal frequency of ω, but with an envelope displaying a beat pattern with two frequencies. The Ω terms arise from the interference between the carrier and the sidebands, and the 2Ω terms come from the sidebands interfering with each other.[22]

We are interested in the two terms that are oscillating at the modulation frequency Ω because they sample the phase of the reflected carrier. There are two terms in this expression: a sine term and a cosine term. Usually, only one of them will be important. The other will vanish. Which one vanishes and which one survives depends on the modulation frequency. In the next section we will show that at low modulation frequencies (slow enough for the internal field of the cavity to have time to respond, or $\Omega\ll\Delta\nu_{\text{fsr}}/\mathcal{F}$), $F(\omega)F^*(\omega+\Omega)-F^*(\omega)F(\omega-\Omega)$ is purely real, and only the cosine term survives. At high $\Omega(\Omega\gg\Delta\nu_{\text{fsr}}/\mathcal{F})$ near resonance it is purely imaginary, and only the sine term is important.

In either case (high or low Ω) we will measure $F(\omega)F^*(\omega+\Omega)-F^*(\omega)F(\omega-\Omega)$ and determine the laser frequency from that.

E. Measuring the error signal

We measure the reflected power given in Eq. (3.3) with a high-frequency photodetector, as shown in Fig. 3. The output of this photodetector includes all terms in Eq. (3.3), but we are only interested in the $\sin(\Omega t)$ or $\cos(\Omega t)$ part, which we isolate using a mixer and a low-pass filter. Recall that a mixer forms the product of its inputs, and that the product of two sine waves is

$$\sin(\Omega t)\sin(\Omega' t)=\tfrac{1}{2}\{\cos[(\Omega-\Omega')t]-\cos[(\Omega+\Omega')t]\}.$$

If we feed the modulation signal (at Ω) into one input of the mixer and some other signal (at Ω') into the other input, the output will contain signals at both the sum $(\Omega+\Omega')$ and difference $(\Omega-\Omega')$ frequencies. If Ω' is equal to Ω, as is the case for the part of the signal we are interested in, then the $\cos[(\Omega-\Omega')t]$ term will be a dc signal, which we can isolate with a low-pass filter, as shown in Fig. 3.

Note that if we mix a sine and a cosine signal, rather than two sines, we get

$$\sin(\Omega t)\cos(\Omega' t)=\tfrac{1}{2}\{\sin[(\Omega-\Omega')t]-\sin[(\Omega+\Omega')t]\}.$$

In this case, if $\Omega=\Omega'$ our dc signal vanishes! If we want to measure the error signal when the modulation frequency is low we must match the phases of the two signals going into the mixer. Turning a sine into a cosine is a simple matter of introducing a 90° phase shift, which we can do with a phase shifter (or delay line), as shown in Fig. 3.

In practice, you need a phase shifter even when the modulation frequency is high. There are almost always unequal delays in the two signal paths that need to be compensated for to produce two pure sine terms at the inputs of the mixer. The output of the mixer when the phases of its two inputs are not matched can produce some odd-looking error signals (see Bjorklund[7]), and when setting up a Pound–Drever–Hall lock you usually scan the laser frequency and empirically adjust the phase in one signal path until you get an error signal that looks like Fig. 7.

IV. UNDERSTANDING THE QUANTITATIVE MODEL

A. Slow modulation: Quantifying the conceptual model

Let's see how the quantitative model compares with our conceptual model, where we slowly dithered the laser frequency and looked at the reflected power. For our phase modulated beam, the instantaneous frequency is

$$\omega(t)=\frac{d}{dt}(\omega t+\beta\sin\Omega t)=\omega+\Omega\beta\cos\Omega t.$$

The reflected power is just $P_{\mathrm{ref}}=P_0|F(\omega)|^2$, and we might expect it to vary over time as

$$P_{\mathrm{ref}}(\omega+\Omega\beta\cos\Omega t)\approx P_{\mathrm{ref}}(\omega)+\frac{dP_{\mathrm{ref}}}{d\omega}\Omega\beta\cos\Omega t$$
$$\approx P_{\mathrm{ref}}(\omega)+P_0\frac{d|F|^2}{d\omega}\Omega\beta\cos\Omega t.$$

In the conceptual model, we dithered the frequency of the laser adiabatically, slowly enough that the standing wave inside the cavity was always in equilibrium with the incident beam. We can express this in the quantitative model by making Ω very small. In this regime the expression

$$F(\omega)F^*(\omega+\Omega)-F^*(\omega)F(\omega-\Omega)$$
$$\approx 2\,\mathrm{Re}\left\{F(\omega)\frac{d}{d\omega}F^*(\omega)\right\}\Omega\approx\frac{d|F|^2}{d\omega}\Omega,$$

which is purely real. Of the Ω terms, only the cosine term in Eq. (3.3) survives.

If we approximate $\sqrt{P_cP_s}\approx P_0\beta/2$, the reflected power from Eq. (3.3) becomes

$$P_{\mathrm{ref}}\approx(\text{constant terms})+P_0\frac{d|F|^2}{d\omega}\Omega\beta\cos\Omega t$$
$$+(2\Omega\ \text{terms}),$$

in agreement with our expectation from the conceptual model.

The mixer will filter out everything but the term that varies as $\cos\Omega t$. (We may have to adjust the phase of the signal before we feed it into the mixer.) The Pound–Drever–Hall error signal is then

$$\epsilon=P_0\frac{d|F|^2}{d\omega}\Omega\beta\approx 2\sqrt{P_cP_s}\,\frac{d|F|^2}{d\omega}\Omega.$$

Figure 6 shows a plot of this error signal.

B. Fast modulation near resonance: Pound–Drever–Hall in practice

When the carrier is near resonance and the modulation frequency is high enough that the sidebands are not, we can assume that the sidebands are totally reflected, $F(\omega\pm\Omega)\approx-1$. Then the expression

$$F(\omega)F^*(\omega+\Omega)-F^*(\omega)F(\omega-\Omega)\approx -i2\,\mathrm{Im}\{F(\omega)\},\tag{4.1}$$

is purely imaginary. In this regime, the cosine term in Eq. (3.3) is negligible, and our error signal becomes

$$\epsilon=-2\sqrt{P_cP_s}\,\mathrm{Im}\{F(\omega)F^*(\omega+\Omega)-F^*(\omega)F(\omega-\Omega)\}.$$

Figure 7 shows a plot of this error signal.

Near resonance the reflected power essentially vanishes, since $|F(\omega)|^2\approx 0$. We do want to retain terms to first order in $F(\omega)$, however, to approximate the error signal,

$$P_{\mathrm{ref}}\approx 2P_s-4\sqrt{P_cP_s}\,\mathrm{Im}\{F(\omega)\}\sin\Omega t+(2\Omega\ \text{terms}).$$

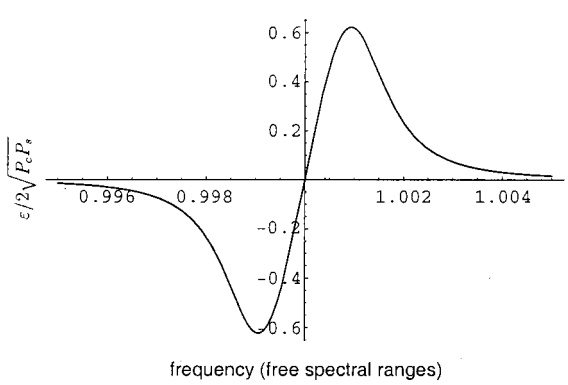

Fig. 6. The Pound–Drever–Hall error signal, $\epsilon/2\sqrt{P_cP_s}$ vs $\omega/\Delta\nu_{\mathrm{fsr}}$, when the modulation frequency is low. The modulation frequency is about half a linewidth: about 10^{-3} of a free spectral range, with a cavity finesse of 500.

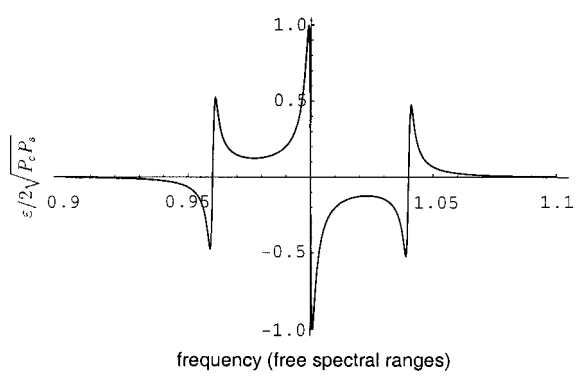

Fig. 7. The Pound–Drever–Hall error signal, $\epsilon/2\sqrt{P_cP_s}$ vs $\omega/\Delta\nu_{\mathrm{fsr}}$, when the modulation frequency is high. Here, the modulation frequency is about 20 linewidths: roughly 4% of a free spectral range, with a cavity finesse of 500.

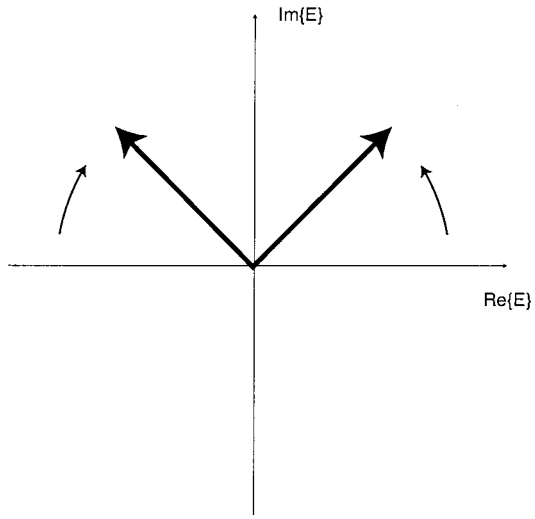

Fig. 8. Sidebands at $\omega \pm \Omega$.

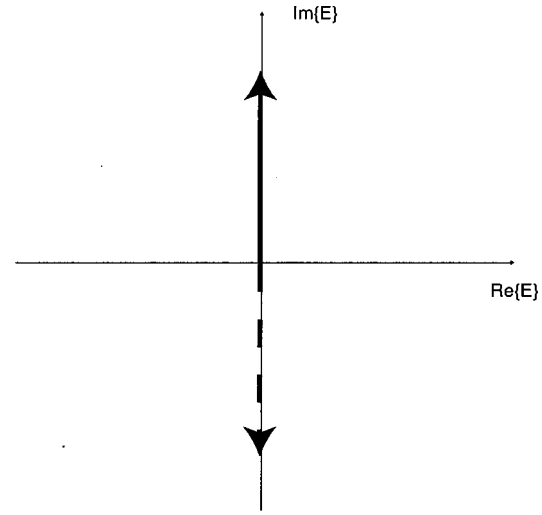

Fig. 9. The sum of the two sidebands shown in Fig. 8. This is the actual electric field produced when the two sidebands interfere with each other. Note that the intensity oscillates at 2Ω.

Since we are near resonance we can write

$$\frac{\omega}{\Delta \nu_{\mathrm{fsr}}} = 2\pi N + \frac{\delta\omega}{\Delta \nu_{\mathrm{fsr}}},$$

where N is an integer and $\delta\omega$ is the deviation of the laser frequency from resonance. It is useful at this point to make the approximation that the cavity has a high finesse $\mathcal{F} \approx \pi/(1-r^2)$. The reflection coefficient is then

$$F \approx \frac{i}{\pi}\frac{\delta\omega}{\delta\nu},$$

where $\delta\nu \equiv \Delta \nu_{\mathrm{fsr}}/\mathcal{F}$ is the cavity's linewidth. The error signal is then proportional to $\delta\omega$, and this approximation is good as long as $\delta\omega \ll \delta\nu$,

$$\epsilon \approx -\frac{4}{\pi}\sqrt{P_c P_s}\,\frac{\delta\omega}{\delta\nu}.$$

That the error signal is linear near resonance allows us to use the standard tools of control theory to suppress frequency noise. We will use this linear behavior later on to examine some fundamental noise limits. It will be useful for us to write the error signal in terms of the regular frequency $f = \omega/2\pi$, instead of ω, and define the proportionality constant between ϵ and δf,

$$\epsilon = D\,\delta f,$$

where the proportionality constant,

$$D \equiv -\frac{8\sqrt{P_c P_s}}{\delta\nu}, \tag{4.2}$$

is called the frequency discriminant.

C. A conceptual model good for high modulation frequency

When the modulation frequency was low, we could picture the reflected power in the time domain and compare it with the modulation of the laser. At high modulation frequencies, we can still conceptualize the technique, but we will have to be a bit more subtle.

We can represent the electric fields of each beam by time-varying vectors in a complex plane that rotates along with the carrier at frequency ω. We can choose this ''moving reference frame'' such that the incident carrier's electric field component always lies along the real axis. The part of the carrier that gets reflected from the Fabry–Perot cavity is also represented by a vector in this plane, as shown in Fig. 5, and near resonance it is given by

$$E_{\mathrm{carrier}} \approx i\sqrt{P_c}\,\frac{\delta\omega}{\pi\delta\nu}.$$

The sidebands have different frequencies than the carrier, so they are represented by vectors that spin around in this reference frame. The upper $(\omega+\Omega)$ sideband has a higher frequency than the carrier, so its vector rotates counterclockwise in the plane with angular frequency Ω. The lower side-

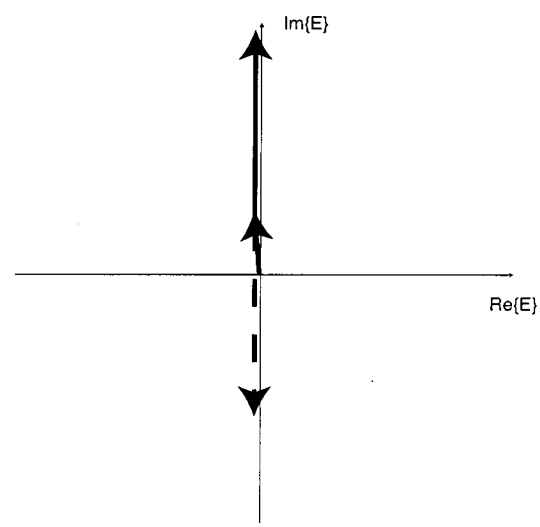

Fig. 10. Sidebands and the reflected carrier near resonance. The actual reflected beam is the coherent sum of these two fields. The small reflected carrier introduces an asymmetry in the intensity over its 2Ω period, which produces a component that varies at Ω. This Ω component is the beat signal between the carrier and the sidebands, and its sign tells you whether you are above resonance or below it.

band has a lower frequency and rotates clockwise at $-\Omega$. (See Fig. 8.) The sum of the two sidebands, when they are both completely reflected off the cavity, is a single vector that oscillates up and down along the imaginary axis. (See Fig. 9.) This field is given by [see Eq. (3.2)]

$$E_{\text{sidebands}} = -i2\sqrt{P_s}\sin\Omega t.$$

The total field reflected off the cavity is the vector sum of the reflected carrier and the two sidebands. (See Fig. 10.) We measure the intensity of this field with the photodetector, and that is just the magnitude (squared) of the total field,

$$\begin{aligned} P_{\text{ref}} &= |E_{\text{carrier}} + E_{\text{sidebands}}|^2 \\ &\approx P_c\left(\frac{\delta\omega}{\pi\delta\nu}\right)^2 + 2P_s - 4\sqrt{P_c P_s}\,\frac{\delta\omega}{\pi\delta\nu}\sin\Omega t \\ &\quad - 2P_s\cos 2\Omega t. \end{aligned}$$

The cross term proportional to $\sin\Omega t$ represents the beating of the sidebands with the reflected carrier, and its sign tells you which side of resonance you are on. The 2Ω term is the result of the two sidebands beating together.

Now we are in a position to understand why the error signal is not limited by the bandwidth of the cavity. Whenever there is a phase mismatch between the promptly reflected field and the leakage field, we get an error signal. For very fast changes in the frequency of the incident (and promptly reflected) beam, the leakage beam acts as a stable reference, averaging both the frequency and the phase of the laser over the storage time of the cavity.[2] If the promptly reflected beam (which provides an effectively instantaneous measure of the incident beam) ''hiccups,'' i.e., jumps away from this average, the error signal will immediately register this jump, and the feedback loop can compensate for it. We are effectively locking the laser to a time average of itself over the storage time of the cavity.

V. NOISE AND FUNDAMENTAL LIMITS: HOW WELL CAN YOU DO?

A. Noise in various parameters

I have only talked about laser frequency so far, but it is a straightforward exercise to extend this analysis in terms of both frequency and cavity length. A little algebra shows that the laser frequency and the cavity length are on equal footings near resonance. For high modulation frequencies,

$$\epsilon = -8\sqrt{P_c P_s}\,\frac{2L\mathcal{F}}{\lambda}\left\{\frac{\delta f}{f} + \frac{\delta L}{L}\right\},$$

where δL is the deviation of the cavity length from resonance, analogous to δf. All along I have been talking about measuring the frequency noise of the laser and locking it to a quiet cavity, but we could have just as easily measured the length noise in the cavity (provided the laser was relatively quiet) and locked the cavity to the laser.

Note that it is not possible to distinguish laser frequency noise from cavity noise just by looking at the error signal. One naturally wonders what other noise sources contribute to the error signal. It is a straightforward exercise to show that none of the following contribute to the error signal to first order: variation in the laser power, response of the photodiode used to measure the reflected signal, the modulation depth β, the relative phase of the two signals going into the mixer, and the modulation frequency Ω. The system is insensitive to each of these because we are locking on resonance, where the reflected carrier vanishes. This causes all of these first-order terms to vanish in a Taylor expansion of the error signal about resonance. (A good treatment of optically related noise sources in a gravitational-wave detector can be found in Ref. 23.)

The system is first-order sensitive to fluctuations in the sideband power at the modulation frequency Ω. Most noise sources fall off as frequency increases, so we can usually reduce them as much as we want by going to a high enough modulation frequency. There is one noise source, however, that does not trail off at high frequencies, and that is the shot noise in the reflected sidebands. Shot noise has a flat spectrum, and for high enough modulation frequencies it is the dominant noise source.

B. Shot noise limited resolution

Any noise in the error signal itself is indistinguishable from noise in the laser's frequency. There is a fundamental limit to how quiet the error signal can be, due to the quantum nature of light.[24]

On resonance, the reflected carrier will vanish, and only the sidebands will reflect off the cavity and fall on the photodetector. These sidebands will produce a signal that oscillates at harmonics of the modulation frequency. Calculating the shot noise in such a cyclostationary signal is fairly subtle,[25] but for our purposes we may estimate it by replacing this cyclostationary signal with an averaged, dc signal. The average power falling on the photodiode is approximately $P_{\text{ref}} = 2P_s$. The shot noise in this signal has a flat spectrum with spectral density of

$$S_e = \sqrt{2\frac{hc}{\lambda}(2P_s)}.$$

Dividing the error signal spectrum by D gives us the apparent frequency noise,

$$S_f = \frac{\sqrt{hc^3}}{8}\,\frac{1}{\mathcal{F}L\sqrt{\lambda P_c}}.$$

Since you can't resolve the frequency any better than this, you can't get it any more stable than this by using feedback to control the laser. Note that the shot noise limit does not explicitly depend on the power in the sidebands, as you might expect. It only depends on the power in the carrier.[26]

It's worth putting in some numbers to get a feel for these limits. For this example we will use a cavity that is 20 cm long and has a finesse of 10^4, and a laser that operates at 500 mW with a wavelength of 1064 nm. If the cavity had no length noise and we locked the laser to it, the best frequency stability we could get would be

$$S_f = \left(1.2\times 10^{-5}\frac{\text{Hz}}{\sqrt{\text{Hz}}}\right)\frac{10^4}{\mathcal{F}}\,\frac{20\ \text{cm}}{L}\sqrt{\frac{1064\ \text{nm}}{\lambda}\,\frac{500\ \text{mW}}{P_c}}.$$

The same shot noise would limit your sensitivity to cavity length if you were locking the cavity to the laser. In this case, the apparent length noise would be

$$S_L = \frac{L}{f}S_f = \frac{\sqrt{hc}}{8}\,\frac{\sqrt{\lambda}}{\mathcal{F}\sqrt{P_c}}. \tag{5.1}$$

For the example cavity and laser we used above, this would be

$$S_L = \left(8.1\times 10^{-21}\frac{m}{\sqrt{\mathrm{Hz}}}\right)\frac{10^4}{\mathcal{F}}\sqrt{\frac{\lambda}{1064\ \mathrm{nm}}\frac{500\ \mathrm{mW}}{P_c}}.$$

ACKNOWLEDGMENTS

Many thanks go to Ken Libbrecht for patiently supporting me while I worked on this paper, and to Stefan Seel for helping me implement my first Pound–Drever–Hall lock. I also thank Ron Drever for wonderful insight and some very interesting discussions and Stan Whitcomb and David Shoemaker for carefully reading and reviewing this manuscript. This work was supported by the National Science Foundation as part of the LIGO project, Grant No. PHY98-01158.

APPENDIX A: A PROOF OF THE CIRCLE THEOREM

For the general case of a Fabry–Perot cavity with lossy mirrors, the reflection coefficient is

$$F = \frac{-r_1 + r_2(r_1^2+t_1^2)\exp\left(i\dfrac{\omega}{\Delta\nu_{\mathrm{fsr}}}\right)}{1-r_1 r_2 \exp\left(i\dfrac{\omega}{\Delta\nu_{\mathrm{fsr}}}\right)}.$$

Here, r_1 and t_1 are the amplitude reflection and transmission coefficients of the input mirror, and r_2 is the amplitude reflection coefficient of the end mirror. (Note that the reflection coefficient does not depend on the transmission of the end mirror!) It is straightforward algebra to show that F satisfies the equation of a circle:

$$|F(\omega)-Z_0|^2 = R^2$$

for all ω. Z_0 and R are both real and are given by

$$Z_0 = -\frac{r_1}{1-r_1^2 r_2^2}[1-r_2^2(r_1^2+t_1^2)]$$

and

$$R = \frac{t_1^2 r_2}{1-r_1^2 r_2^2}.$$

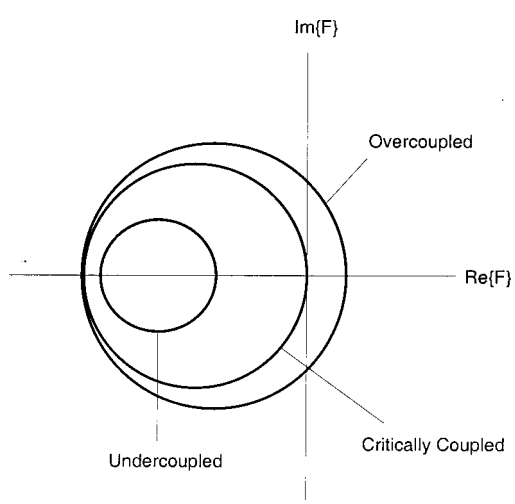

Fig. 11. Plots of F in the complex plane for various couplings. Only for the impedance matched case does the reflected intensity vanish on resonance.

If this circle intersects the origin (i.e., the reflected intensity vanishes on resonance) then the cavity is said to be *critically coupled*. The requirement for critical coupling is that $R = |Z_0|$, or

$$r_1[1-r_2^2(r_1^2+t_1^2)] = t_1^2 r_2. \tag{A1}$$

Our lossless, symmetric cavity had $r_1 = r_2 \equiv r$, $t_1 = t_2 \equiv t$, $r^2+t^2=1$, and satisfied the conditions for critical coupling. A lossy, asymmetric cavity can also be critically coupled, provided its mirror parameters satisfy Eq. (A1), which can be rewritten as

$$r_2 = \frac{t_1^2 + \sqrt{t_1^4 + 4r_1^2(r_1^2+t_1^2)}}{2r_1(r_1^2+t_1^2)}.$$

Other cases, known as undercoupling and overcoupling, are illustrated along with critical coupling in Fig. 11. Overcoupling plays a central role in interferometric gravitational-wave detectors and can be achieved by making the end mirror much more reflective than the near mirror ($t_2 \ll t_1$).

APPENDIX B: OPTIMUM MODULATION DEPTH

It is sometimes useful to maximize the slope of the error signal D [recall Eq. (4.2)]. This slope is a measure of the sensitivity of the error signal to fluctuations in the laser frequency (or cavity length). One example of when you might need a high sensitivity in this discriminant is if you need a large gain in the feedback loop.

The discriminant D depends on the cavity finesse, the laser wavelength, and the power in the sidebands and the carrier. Experimental details usually restrict your finesse and wavelength choices, but you often have quite a bit of freedom in adjusting the sideband power. The question I want to address in this section is this: How does D depend on the sideband power?

D is proportional to the square root of the product of the sideband and carrier power. [See Eq. (4.2).] This has a very simple form when $P_c + 2P_s \approx P_0$, i.e., when negligible power goes into the higher order sidebands,

$$D \propto \sqrt{P_c P_s} \approx \sqrt{\frac{P_0}{2}}\sqrt{\left(1-\frac{P_c}{P_0}\right)\frac{P_c}{P_0}}.$$

A plot of D in this approximation against P_c/P_0 traces out the top half of a circle, with a maximum at $P_c/P_0 = 1/2$. (See Fig. 12.) It is useful to express the power in the sidebands

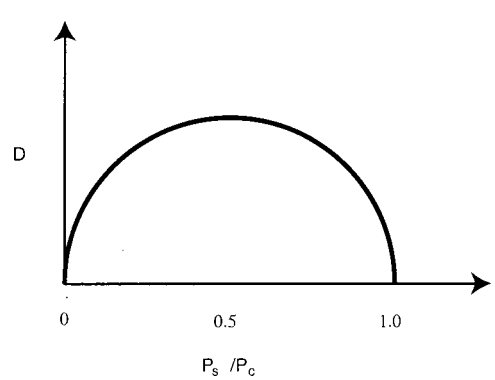

Fig. 12. An approximate plot of D, the slope of the error signal near resonance, vs P_s/P_c. The optimum value is at $P_s/P_c = 1/2$, and the maximum is very broad.

relative to the power in the carrier, and this gives

$$\frac{P_s}{P_c}=\frac{1}{2}.$$

D is maximized when the power in each sideband is half the power in the carrier, and this maximum is fairly broad.

If you want do a more careful analysis, write D in terms of the Bessel functions of the modulation depth and find its maximum. You'll find the optimum modulation depth to be $\beta=1.08$, and you'll come up with essentially the same answer as with the simple estimate:

$$\frac{P_s}{P_c}=0.42.$$

[1]R. W. P. Drever *et al.*, ''A Gravity-Wave Detector Using Optical Cavity Sensing,'' in *Proceedings of the Ninth International Conference on General Relativity and Gravitation, Jena, July 1980*, edited by E. Schmutzer (Cambridge U.P., Cambridge, 1983), pp. 265–267.

[2]R. W. P. Drever *et al.*, ''Laser phase and frequency stabilization using an optical resonator,'' Appl. Phys. B: Photophys. Laser Chem. **31**, 97–105 (1983).

[3]A. Abromovici *et al.*, ''LIGO: The laser interferometer gravitational-wave observatory,'' Science **256**, 325–333 (1992).

[4]S. Seel, R. Storz, G. Ruoso, J. Mlynek, and S. Schiller, ''Cryogenic optical resonators: A new tool for laser frequency stabilization at the 1 Hz level,'' Phys. Rev. Lett. **78** (25), 4741–4744 (1997).

[5]L. A. Rawley *et al.*, ''Millisecond Pulsar PSR 1937+21: A Highly Stable Clock,'' Science **238**, 761–765 (1987).

[6]Axel Schenzle, Ralph G. DeVoe, and Richard G. Brewer, ''Phase-modulation laser spectroscopy,'' Phys. Rev. A **25**, 2606–2621 (1982).

[7]G. C. Bjorklund *et al.*, ''Frequency Modulation (FM) Spectroscopy: Theory of Line shapes and Signal-to-Noise Analysis,'' Appl. Phys. B: Photophys. Laser Chem. **32**, 145–152 (1983).

[8]Gary C. Bjorklund, ''Frequency-modulation spectroscopy: A new method for measuring weak absorptions and dispersions,'' Opt. Lett. **5**, 15–17 (1980).

[9]R. V. Pound, ''Electronic Frequency Stabilization of Microwave Oscillators,'' Rev. Sci. Instrum. **17**, 490–505 (1946).

[10]R. A. Boyd, J. L. Bliss, and K. G. Libbrecht, ''Teaching physics with 670-nm diode lasers—experiments with Fabry–Perot cavities,'' Am. J. Phys. **64**, 1109–1116 (1996).

[11]Eugene Hecht, *Optics* (Addison–Wesley, Reading, MA, 1998).

[12]Grant R. Fowles, *Introduction to Modern Optics* (Dover, New York, 1975).

[13]G. F. Franklin, J. D. Powell, and A. Emami-Naeni, *Feedback Control of Dynamic Systems* (Addison–Wesley, Reading, MA, 1987).

[14]B. Friedland, *Control System Design: An Introduction to State Space Methods* (McGraw–Hill, New York, 1986).

[15]H. Kogelnik and T. Li, ''Laser beams and resonators,'' Appl. Opt. **5**, 1550–1567 (1966).

[16]A. E. Siegman. *Lasers* (University Science Books, Mill Valley, CA, 1986).

[17]Amnon Yariv, *Optical Electronics in Modern Communications* (Oxford U.P., New York, 1997).

[18]Peter R. Saulson, *Fundamentals of Interferometric Gravitational Wave Detectors* (World Scientific, Singapore, 1994).

[19]R. L. Barger, M. S. Sorem, and J. L. Hall, ''Frequency stabilization of a cw dye laser,'' Appl. Phys. Lett. **22**, 573–575 (1973).

[20]The Pockels cell actually modulates the laser's phase, but the distinction between phase and frequency modulation is irrelevant for the conceptual model.

[21]The reader who doesn't like Bessel functions will find that the small angle expansion, $E_{\rm inc}\approx E_0[1+i\beta\sin\Omega t]e^{i\omega t}=E_0[1+(\beta/2)(e^{i\Omega t}-e^{-i\Omega t})]e^{i\omega t}$ works just about as well.

[22]There may also be some contribution from higher order terms that we neglected when we expanded $e^{i(\omega t+\beta\sin\Omega t)}$ in terms of Bessel functions. These may make a significant contribution to the 2Ω term, but we do not need to consider them in this tutorial.

[23]J. B. Camp, H. Yamamoto, S. E. Whitcomb, and D. E. McClelland, ''Analysis of Light Noise Sources in a Recycled Michelson Interferometer with Fabry–Perot Arms,'' J. Opt. Soc. Am. A (to be published).

[24]You can get around this quantum limit to some extent by squeezing the light, but that is beyond the scope of this article.

[25]T. T. Lyons, M. W. Regehr, and F. J. Raab. ''Shot Noise in Gravitational-Wave Detectors with Fabry–Perot Arms'' (unpublished).

[26]The shot noise limit does depend implicitly on the power in the sidebands, since $P_c=P_0-P_s$, but this is a relatively minor effect.

GRAND CANYON BOATWOMAN

Lorna [Corson] rows gracefully and likes being in control. Her least-favorite rapid is unpredictable Granite. ''It's sloppy, there's no finesse.'' Lorna's favorite rapid is Deubendorff at low water, when you have to make an exact entry or risk smashing into black fang rocks disguised as foam at the bottom. ''I love reading the water, estimating what you think it will do, getting your angles right, and finding out how close your calculations were—just the physics of water.''

Louise Teal, *Breaking into the Current: Boatwoman of the Grand Canyon* (University of Arizona Press, Tucson, 1994), p. 130.