

Optical Measurements

My means are sane, my motive and my object mad.

—Captain Ahab, in *Moby Dick* by Herman Melville

10.1 INTRODUCTION

It's easy to be beguiled by reading people's papers into believing that optical instruments are complicated to fit them exquisitely for their purpose, but they aren't.[†] They are complicated in order to work around the limitations of available components, weird sample geometries, the idiosyncrasies and prejudices of experimenters,[‡] and the occasional inconvenient Law of Nature. If we could do stable, flexible, and convenient measurements at the quantum limit all the time with no problems, there would be many fewer types of instrument. Besides component limitations, the main problems come from noisy background signals and environmental effects. That isn't all bad—it provides instrument builders with job security and an unending string of fun problems to solve. There are some enduring principles underlying all this welter, and that's what we'll concentrate on here—background reduction, labeling signal photons to distinguish them from background, and exploiting all the changes that physical interactions can produce in light beams.

10.2 GRASS ON THE EMPIRE STATE BUILDING

As Miller and Friedman engagingly point out in *Photonics Rules of Thumb*, our job is like measuring the height of a tuft of grass on the roof of the Empire State Building, as shown in Figure 10.1; the desired signal is a small delicate thing, riding on a huge ugly background. (It's lucky for New York that the Empire State Building is actually quite nice looking, and doesn't shuck and jive the way our background signals often do.) This

[†]There's a wonderful paper on the tension between frankness and self-promotion, written by Edmond H. Weiss, "Professional Communication" and the "Odor of Mendacity": The Persistent Suspicion that Skillful Writing is Successful Lying'. *IEEE Trans. Professional Commun.* **38**, 3 (September 1995).

[‡]The author's have of course been rehearsed at some length in this book, but other people have them too.

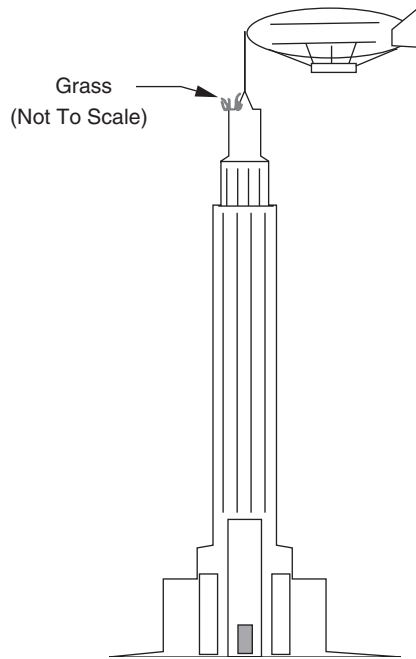


Figure 10.1. Background signals.

whimsical image is a helpful one all through our discussion, so hang on to it. The idea is to make the grass as conspicuous as possible, by fertilizing it to make it grow taller, painting it fluorescent orange to make it stand out, moving it to ground level, climbing the building, or looking at it from a helicopter. In other words, by getting as much signal as possible, labeling signal photons somehow so they can be distinguished from background photons, and getting rid of the background itself, or at least its fluctuations.

10.2.1 Background, Noise, and Spurious Signals

Some background signals are inherent in the measurement, for example, speckle statistics in shearing speckle interferometry or shot noise in fluorescent measurements. Most of the rest come from the immediate environment (e.g., room lights) or the measurement system itself: source noise, stray reflections, and molecular (Rayleigh) scatter.

This junk comes in two flavors: additive and multiplicative. The distinction is based on what we consider to be our signal. In the grass image, the thermal expansion of the building is additive, since it just lifts all the grass the same amount as the building heats and cools, whereas wind is more or less multiplicative, because tall grass waves more than short grass. In spectroscopy, (additive) background shifts the baseline and (multiplicative) gain drift changes the peak heights. Source power variations produce both; the baseline is proportional to the source power, so its fluctuations appear as additive noise, and the peak heights are also proportional to source power, which is multiplicative noise. If we thought of the baseline as signal too, this would be a purely multiplicative noise source. Low frequency multiplicative junk puts noise sidebands on all your measurement signals and carriers, at all frequencies, which are going to stay there

and frustrate your attempts to make measurements at high SNR unless you get rid of them somehow.

10.2.2 Pedestal

A quiet additive background or *pedestal* is like the DC offset of a single-supply circuit. As long as it is really still, it doesn't cause any problem. Background fluctuations are far more likely to be the limiting factor. Unlike a quiet electronic bias level, optical background signals have full shot noise at least, and usually much more of other stuff too.

Occasionally the pedestal is so big that its fixed pattern noise dominates your dynamic range budget. An example is a HgCdTe area array in the thermal IR, where the integration time and signal linearity are limited by the nonuniform sensor seeing a huge 300 K background. The fixed pattern itself is a strong function of intensity, so careful calibration at a variety of scene brightnesses is required, but eventually the fixed pattern noise can be tamed. In most cases, the background shot noise can be made to dominate the fixed-pattern fluctuations and readout noise.

10.2.3 Background Fluctuations

Background fluctuations fall into two classes: photocurrent shot noise and everything else. Shot noise is spectrally white, so it lands on your signal irrespective of where the signal might be found, and has Gaussian amplitude statistics, so its average properties are very easily calculated; the only way to get rid of it is to reduce the background light, using baffles, filters, and so forth.

The other category includes source excess noise, circuit noise, and spurious modulation of the background signal (e.g., 100/120 Hz from room lights). We can usually get round this sort of thing by optical hacks and circuit improvements.

10.2.4 Noise Statistics

As we'll see in Section 13.6.14, shot noise really is Gaussian, to amazing accuracy, and so is Johnson noise, although it's harder to measure. Other types of noise are not so nice, especially $1/f$ noise and mode hopping, which tend to have a strongly popcorn character, that is, lots of fairly sharp step-like jumps, and wildly nonstationary statistics, with long bursts of noise coming at irregular intervals and going away again for awhile. Most measurements are more sensitive to extreme values than to the fuzz level.

10.2.5 Laser Noise

We talked about laser noise in Section 2.13; the two basic kinds are residual intensity noise (RIN) and phase noise. It appears as a purely multiplicative problem in dark-field measurements, and both additive and multiplicative in bright-field ones. It's often quite large, 10^{-3} – 10^{-5} , which will really degrade a measurement if we don't take steps.

Diode lasers have lower RIN power spectral density (at least if you can prevent them from mode hopping), but it extends from DC to a gigahertz and so is hard to avoid by rapid modulation; by contrast, a frequency shift of a few tens of megahertz will get you out of the excess noise region of gas and solid state lasers. If the laser has multiple

longitudinal modes, their beats will produce near-replicas of the baseband noise spaced near multiples of $c/(2n\ell)$, where ℓ is the cavity length and n is the real part of the refractive index at the laser wavelength.

Diode lasers have significantly worse phase noise than other types do, due to the low finesse and intrinsic instability of their resonators. DFB lasers are better than Fabry–Perot types, but still nowhere near the frequency stability of a HeNe. Demodulation of the phase noise and mode partition noise by etalon fringes and multiple scattering is a horrendous problem in bright-field diode laser measurements, as we saw in Section 2.5.3; Section 19.1.1 has a real tear-jerker that hinges on this point.

10.2.6 Lamp Noise

Lamps being thermal devices, one might think their noise would all be down in the very low baseband, but unfortunately that isn't always true; the thermal signal falls off as $1/f^2$ or slower, and plasma noise in arc lamps extends well into the kilohertz. Flashlamps exhibit pulse-to-pulse variations of 1% or so, which puts big noise sidebands on the repetition rate component (see Section 15.9.4).

10.2.7 Media Noise

There are many classes of measurements, for example, photon echo and other spectral hole-burning measurements, optical recording, photographic densitometry, and holographic reconstruction, in which the sample consists of some recording medium in which a pattern has been recorded. The noise budget has to include the noise added in the recording process.

Example 10.1: Media Noise in Spectral Hole Burning. The frequencies and lineshapes of the spectral lines of chromophores (i.e., atoms, molecules, and ions exhibiting spectral lines) depend on their environment. Line broadening is usefully divided into *homogeneous* and *inhomogeneous* broadening. Homogeneous broadening affects all the chromophores the same way, at least on average: collisions in a gas, orbital overlaps in a crystal (which lead to reduced upper state lifetime). Inhomogeneous broadening affects different chromophores differently: Doppler shifts in gases and the different Stark shifts seen by chromophores in different places in a crystal. In a nutshell, in a measurement of a single atom, homogeneous broadening shows up in the width of the transition, and inhomogeneous broadening shows up as a shift in the center frequency—it's the ensemble of many chromophores with different offsets that makes up the inhomogeneous lineshape.

Rare-earth ions are unique because their F-shell electrons are so well shielded by the valence electrons that they are hard to Stark-shift. This means that the homogeneous broadening of such ions when introduced into a suitable crystal can be very small indeed—below 100 Hz for some nominally forbidden transitions at low temperatures (4 K). In an inhomogeneously broadening medium, such as a YAG crystal, this narrow homogeneous linewidth means that there can be 10^3 – 10^6 resolvable frequencies. Since the homogeneous transition is narrow, the upper state lifetime must be long, leading to the possibility of *spectral hole burning*, in which a narrowband laser bleaches the absorption in a narrow bandwidth by pumping all the chromophores in that interval into the upper state. There are a lot of things one can do with this sort

of idea, but one problem is that with that many resolvable frequencies, there aren't that many chromophores per bin—perhaps only 10^6 . Unless the hole is fully saturated (all the chromophores are bleached), this medium is going to have frozen-in counting noise due to the stochastic nature of absorption. Such a medium cannot record any signal with more than 60 dB SNR, no matter what the readout system's photon budget may be—the desired signal isn't identical with the sample property being measured, so the media noise limits the attainable SNR, even though the readout can be made very quiet.

10.2.8 Electrical Interference

This is more of a problem for Chapters 16 and 19, but deserves mention here because sometimes it's easier to avoid localized background effects than to fight them; for example, trying to do sensitive measurements at harmonics of 50 or 60 Hz is a mug's game unless you're running battery powered in the Arctic—and even then, the Earth's magnetospheric activity shows peaks at those frequencies due to power line excitation at lower latitudes that puts 50 and 60 Hz sidebands on the auroras, so you still can't win. Similarly, if your measurement uses a Pockels cell modulated with 1.5 kV RMS of 10 MHz sine wave, it's going to be pretty difficult to make shot noise limited measurements right there. It's certainly possible, but in doing it you'll learn more about grounding and shielding than you probably care to. A slower subsidiary modulation (e.g., chopping the light) will give you signals at, say, $10 \text{ MHz} \pm 2 \text{ kHz}$, which is a lot easier to work with.

10.2.9 Electronic Noise

Electronic noise is dissected minutely in Chapter 18, because it is such a common source of measurement-destroying additive junk. The main effects are Johnson noise in a too-small load resistor, the technical noise of amplifiers, and $1/f$ noise due to not cleaning the circuit board, cracking an IC package by overheating, or just choosing the wrong components (e.g., thick film resistors). No matter how good your detector, a poor choice of front end will make it look bad. A few kinds of electronic noise are multiplicative, such as base resistance voltage noise in an unbiased cascode stage (Section 18.4.4) or denominator noise in an analog divider, and others go as the square root of the photocurrent, for example, the shot noise of a current mirror on a photodiode.

10.2.10 Quantization Noise

A final, and usually more or less additive, noise source is quantization. In Section 13.11.1 we'll see that it has an RMS value of $1/\sqrt{12}$ ADU, and that it will severely limit our dynamic range unless we reduce that range in analog before digitizing.

10.2.11 Baseband Isn't a Great Neighborhood

All the riffraff of the frequency domain collects at baseband. The $1/f$ noise, power supply ripple, arc wander, table vibrations, wind noise, and so on, all happen down in the hertz to kilohertz region. There are two ways to deal with this; either clean up the neighborhood or move away yourself. Both are possible, but cleaning it up prevents it from following you, which it otherwise tends to do, since a lot of it is multiplicative.

Aside: Moving the Grass to Ground Level. The easiest method for measuring the grass is to send someone up to the roof with a trowel to bring it down to ground level, then measure it with a child's ruler. This is where smart measurement ideas are really important: they can make hard things easy by redefining the problem.

10.3 DETECTION ISSUES: WHEN EXACTLY IS BACKGROUND BAD?

Background signals are actually not an unmixed evil, and sometimes are an enormous help in making sensitive measurements. The key is their optical phase correlations with the desired signal. First, let's take a close look at noise in bright- and dark-field measurements.

10.3.1 Dark Field

A dark-field measurement is one where ideally the only light is signal. A dark-field microscope (see Section 9.8.2) is the classical example, but there are lots of others: fluorescence, scintillation, second harmonic generation (SHG). If there is enough light to override the circuit noise, dark-field measurements are terrific.

Unfortunately, that's a pretty stiff requirement, because the detected electrical power is quadratic in the optical power, so the sensitivity of the measurement drops off extremely badly at low light levels; dark-field measurements are almost always severely Johnson noise limited unless you're using electron multiplying detectors (PMTs, APDs, L^3 CCDs, or MCPs). Section 18.2 shows how careful you have to be in order to get to the shot noise limit with a $2\ \mu\text{A}$ photocurrent in a 1 MHz bandwidth; that $2\ \mu\text{A}$ corresponds to a 12 THz average photoelectron count rate. A photon-counting photomultiplier can see single photoelectrons (a good one can at least, providing it's cooled) and will get you up to around 10^6 – 10^7 photoelectrons/s, which is of the order of 1 pW, but that doesn't fill the gaping hole between there and 12 THz. In that empty space, we're stuck with analog mode PMTs and APDs, which are no fun at all when you try to make stable measurements—they're great for on/off, for example, optical fiber communications. Changing to bright field is often a good solution, although that changes the counting statistics, which sometimes matters (see Section 10.3.6).

10.3.2 Bright Field: Amplitude Versus Intensity Sensitivity

We've talked about the advantages of coherent detection in Sections 1.5 and 3.11.7, where we saw that this technique allows shot noise limited measurements at all received power levels. Bright-field measurements have the same benefits, because the background (if it's coherent with the signal light) functions as the local oscillator. One way of looking at this is as dynamic range compression. The interference term's electrical power is linear in the signal beam power, so a 1 pW optical signal winds up being 60 dB (electrical) below a $1\ \mu\text{W}$ signal, rather than 120 dB as in a dark-field measurement. This lifts small optical signals up above the electrical noise most of the time, and this is the key to the advantages of bright-field and dim-field measurements. Since one coherently added photon/s/Hz produces a current change equal to the RMS shot noise, and the shot noise can be made to dominate the circuit noise, the nasty hole goes completely away.

Remember that coherent detectors are highly selective in frequency, position, and angle, which helps reject incoherent background light (the good news) but restricts the optical signal the same way (the bad news).

10.3.3 Coherent Background

So far we've been thinking of additive backgrounds adding in intensity, or equivalently in photocurrent. When the signal and background beams are coherent, though, this is not the only possibility. A pure extinction measurement, where the signal is a small blip on the DC photocurrent from the unattenuated laser, has a shot noise floor of

$$\langle \Delta i^2 \rangle = 2eI_{\text{DC}}. \quad (10.1)$$

There are two ways of looking at this, namely, the *optical theorem* (also known as the *extinction sum rule*) and coherent detection of forward-scattered light. Let's describe the total field as $\psi_{\text{inc}} + \psi_{\text{scat}}$, where ψ_{inc} is the total field in the absence of the absorber.

10.3.4 Optical Theorem

By conservation of energy, the power lost to the beam has to equal the absorption plus the total integrated scatter. We can write the total power as

$$\iint |\psi_{\text{tot}}|^2 d\Omega = \iint |\psi_{\text{inc}} + \psi_{\text{scat}}|^2 d\Omega = \iint (|\psi_{\text{inc}}|^2 + |\psi_{\text{scat}}|^2 + 2\text{Re}\{\psi_{\text{scat}}\psi_{\text{inc}}^*\}) d\Omega. \quad (10.2)$$

For a lossless scatterer, $P_{\text{tot}} = P_{\text{inc}}$, so the last two terms sum to zero:

$$\iint \psi_{\text{scat}}\psi_{\text{scat}}^* d\Omega = -2\text{Re} \left\{ \iint \psi_{\text{scat}}\psi_{\text{inc}}^* d\Omega \right\}. \quad (10.3)$$

In other words, a small detector sensing extinction in the transmit beam receives the same amount of signal as a huge detector covering the whole $4\pi - \epsilon$ sr (excluding the main beam). The bad news is that the shot noise of the beam is also there, which reduces our sensitivity—but it's still a dramatic improvement over a dark-field detector of the same NA.

The other way of looking at it is by the coherent detection Rule of One from Section 1.5.2: *One* coherently added signal photon in *One* second gives an AC measurement with *One* sigma confidence in a *One* hertz bandwidth. Although the beam itself generally has a tiny solid angle Ω_{Tx} , so that the scattered signal power in that solid angle is impossibly small, the coherently added ψ_{inc} lifts it up to the same level as the huge dark-field detector would see. This operation is unable to improve the counting statistics of ψ_{scat} in the small Ω_{Tx} , but due to the Rule of One, it does as well as it can: the measurement sensitivity is limited by the ratio of the power in ψ_{scat} to its *own* shot noise in Ω_{Tx} (not the power in ψ_{scat} divided by the shot noise of ψ_{inc}). (For many years the author had to rederive this fact every time before using it, because it seemed so counterintuitive.)

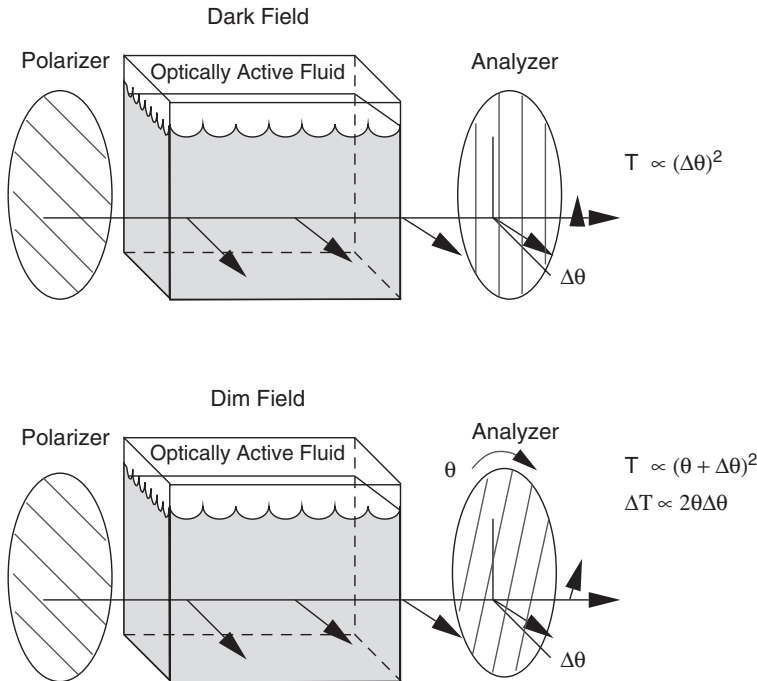


Figure 10.2. Dim-field measurements use a bit of coherent background as an LO signal. Here a crossed-polarizer measurement gets an SNR boost from uncrossing them a bit.

10.3.5 Dim-Field Measurements

Measurements that could be dark field, but where just a bit of coherent background is added are called *dim-field* measurements. An example is the slightly uncrossed crossed-polarizer measurement of Figure 10.2. The coherent background functions as a local oscillator, and so lifts the signal shot noise above the Johnson noise, but not much above; you stop when the multiplicative noise due to LO beam fluctuations becomes the limiting factor, or when you start rejecting your own signal by turning the analyzer too far.

10.3.6 Bright and Dark Fields Are Equivalent

The net of all this is that a bit of coherently added background doesn't hurt anything and may help a lot. As a matter of fact, there is no limit in principle to how much coherent background light you can add. At some point things start deteriorating, due to detector saturation and residual laser noise. We know how to get rid of those pretty well, except when the laser RIN and frequency noise get intermingled, for example, by vignetting or etalon fringes. Another source of trouble is rising electronic noise due, for example, to voltage noise in a cascode stage (Section 18.4.4), but that doesn't happen until we get up into at least hundreds of microamps of photocurrent, and it isn't too hard to get to the shot noise with that. Thus the RMS fractional uncertainty in the photocurrent is the same in bright and dark fields.

They do differ in one moderately important respect: the counting statistics of a bright-field measurement are Gaussian, and those of a photon-counting measurement are Poissonian. This leads to the Bright-Field Rule:

Bright and dark fields are exactly equivalent except for their counting statistics.

This equivalence means that we're free to choose bright or dark field based on other considerations. If you can get a big detection NA, have no background or scattered light problems, and can stand the large dynamic range, you're better off with dark field, whereas in the presence of strong optical or electronic background signals for which the selectivity and gain of bright field helps, or a large signal range, go with bright field or dim field. A noise canceler can get rid of the laser noise, or alternatively (for gas and solid state lasers) a heterodyne interferometer or FM system can avoid the worst of it.

Aside: Counting Statistics: Bright versus Dark Field. In Section 13.6.14 we'll discuss the differences, and how to set detection thresholds intelligently based on the confidence level required. The difference is less striking than it may look, since what's going on is really just the reappportionment of uncertainty; the zero signal state of a Poisson process has zero uncertainty, whereas the variance is fixed in the Gaussian case for all sufficiently weak signal beam levels. What's going on is a redistribution of statistical uncertainty from the finite signal case to zero signal.

10.3.7 Heterodyne Interferometry

The heterodyne interferometer is the queen of laser-based optical instruments (Figure 10.3). It is demanding and occasionally temperamental, but unique in its ability to function in extreme situations, for example, very strong backscatter that would cause interference and possibly mode hopping. The backscattered light is frequency shifted away from the cavity transmission peak, so it bounces harmlessly off the output coupler. Heterodyne measurements are normally 3 dB worse than shot noise limited homodyne or baseband measurements due to the image frequency noise, just as in radios (see Section 13.8.7). With gas or solid state lasers, an 80–200 MHz AOD can get us pretty well out of the laser noise and into the shot noise region, at least as far as additive effects go.

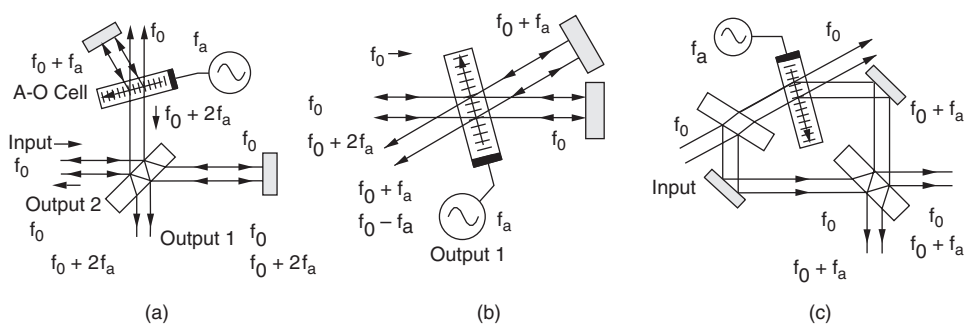


Figure 10.3. Heterodyne interferometers: (a) Michelson with separate beamsplitter, (b) Michelson with Bragg cell as beamsplitter, and (c) Mach-Zehnder.

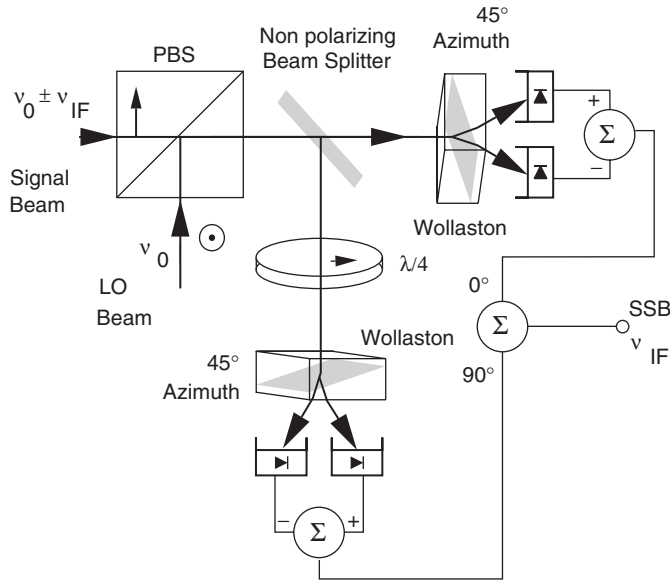


Figure 10.4. Optical single-sideband (SSB) mixer.

10.3.8 SSB Interferometers

In a heterodyne interferometer, the frequency mixing occurs at optical frequency, with a fractional offset of maybe 1 part in 10^6 if we're lucky, which makes the usual radio trick of filtering out the image frequency next to impossible. This costs us 3 dB SNR, and if that doesn't bother you by now, you haven't been paying attention. Two ways round this are to use homodyne detection (IF at DC), which makes the signal and image frequencies coincide, so that only one bandwidth's worth of noise gets through, or to build an optical SSB detector[†].

Optical SSB mixers work just like the ones in Section 13.8.7. Combine two beams in orthogonal polarizations (Figure 10.4). Split the combination beam without polarizing it, for example, with a patterned metal film (you may need a bit of spatial filtering to take out the pupil function weirdness this causes). Put one-half through a quarter-wave plate aligned to one of the polarization axes, detect separately, using Wollastons with their outputs subtracted, and add or subtract the result. You'll get a zero-background output that is purely USB or LSB. Whether all this foofaraw is worth it depends on how valuable your photons are.

10.3.9 Shot Noise Limited Measurements at Baseband

You really can get to the shot noise limit in a bright-field measurement, even at baseband, and even if the laser is an ugly mass of RIN. The trick is to use a laser noise canceler, which will get you 50–70 dB rejection of additive laser noise at low frequency, and

[†]Bo F. Jørgensen, Benny Mikkelsen, and Cathal J. Mahon, Analysis of optical amplifier noise in coherent optical communication systems with optical image rejection receivers. *J. Lightwave Technol.*, **10**(5), 660–671 (1992).

40 dB out to 10 MHz or so. Using the ratiometric output of the canceler will also get rid of the multiplicative noise to the tune of 70 dB from DC to some kilohertz, where you need it most. It's a pretty ugly laser indeed that needs more than that. See Sections 10.8.6 and 18.6.3 for the details.

10.4 MEASURE THE RIGHT THING

The best way to get better data is to measure something closer to what you need to know. A photoacoustic measurement, where a chopped laser heats a surface and a laser interferometer detects the surface motion, could probably be improved by using a focused ultrasonic transducer to detect the thermal signal instead. There are lots of other examples, for example, using a Nomarski microscope instead of a dark-field microscope to look at weak phase objects such as unstained animal cells.

Example 10.2: Nomarski DIC Microscope. The Nomarski differential interference contrast microscope is a white-light shearing interference microscope with zero path difference between beams (Figure 10.5). The shear is very small, less than one spot diameter, and is produced by a modified Wollaston prism (see Section 6.7.4) in the pupil of the microscope objective. It produces an image that is roughly

$$I(x, y) = |h(\mathbf{x}) * E_x(x, y, z(x, y)) + e^{i\phi} h(\mathbf{x}) * \delta(x - \xi) * E_y(x, y, z(x, y))|^2$$

$$\approx |h_2(\mathbf{x}) * E(x, y, z(x, y))|^2 + 2 \cos\left(\phi + \xi \frac{\partial \theta}{\partial x}\right) \cdot |(h_2(\mathbf{x}) * E(x, y, z(x, y)))|^2, \quad (10.4)$$

where θ is the phase of E , so that an amplitude and a phase gradient image are superimposed. The phase shift ϕ between the images can be varied by moving the prisms laterally to get the best contrast.

With thermal light, it is important that the illumination passes through a matching prism; two passes through the one prism works in reflection, but you need a second one in transmission. (Why?)

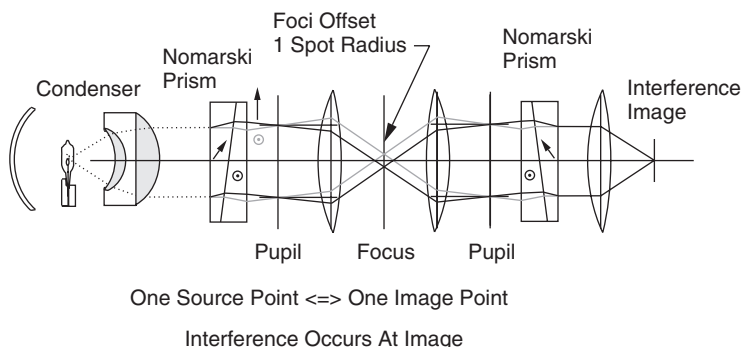


Figure 10.5. The Nomarski differential interference contrast microscope is a shearing microscope with a very small shear and zero path difference.

10.4.1 Phase Measurements

Always consider measuring phase, and think hard. Optical phase, modulation phase, any phase you can think of usually holds useful information. If your current measurement is phase insensitive, what phase would get you? And how hard would it be to get? Section 15.5 has lots on how to measure the phase of an RF signal accurately, so if you're doing a single-point AC measurement, it isn't a difficult matter.

Example 10.3: Attractive Mode Atomic Force Microscope (AFM). The attractive mode AFM of Figure 10.6 uses a vibrating cantilever to sense the force gradient between a very sharp tip and a sample, with a heterodyne interferometer to turn the tip vibration into phase modulation of the 80 MHz carrier. The force gradient acts like a change in the spring constant, and so changes the resonant frequency of the cantilever a bit. The usual way of detecting this is to drive the cantilever piezo with some frequency just above its resonance, and servo the tip-sample distance by forcing the amplitude to remain constant. Too low an amplitude means the tip is too close.

The tip has a tendency to stick to the surface because of the surface tension of the adsorbed water layer. Sticking of course stops the vibration entirely, causing a gigantic error signal, so the servo controller pulls the tip back and back until it lets go. Unfortunately, that plucks the cantilever very hard, making it ring strongly at its free resonant frequency. The servo thinks this means that the tip is way, way too far away, so it sends it crashing right into the sample again, and the cycle repeats until the tip is bashed into oblivion.

The solution is to measure the phase of the modulation instead of its depth—the phase of the phase, if you will. The modulation phase contains tuning information but is completely immune to this problem, because it can be tuned so that the free resonance is outside the passband, and the error signal is not enormously larger when the tip is stuck to the surface; if there's zero oscillation, there's zero output from the phase detector, which isn't too far from where it expects to be. You do need some amplitude sensitivity, because otherwise the tip will happily drag along the surface at zero oscillation amplitude, but

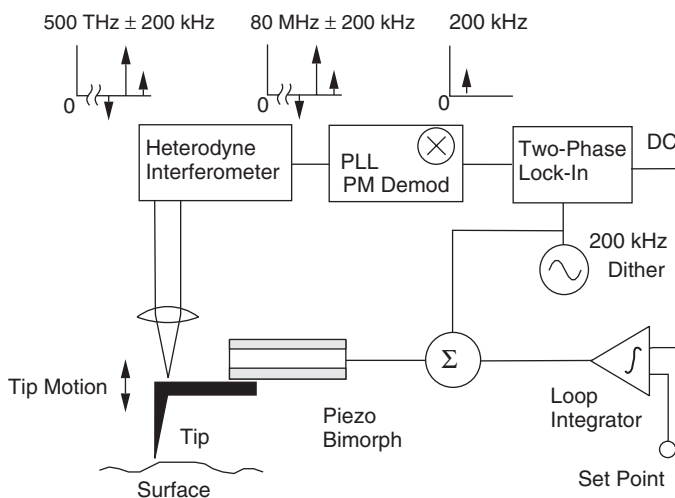


Figure 10.6. Attractive mode AFM.

that's just a matter of servoing slightly off the null of the phase detector. The increased resistance to tapping allows good measurements with the tip hovering closer to the sample, which translates to higher spatial resolution and sensitivity.

10.4.2 Multiple-Scale Measurements Extend Dynamic Range

Measure with a micrometer, mark with chalk, cut with an axe.

Anonymous

Remember the vernier dial? You still see them on inexpensive calipers, but that's about it nowadays. A vernier allows us to get three-figure accuracy from a two-figure eyeball. It works by spreading a small residual error out over many divisions, where it's easy to spot. We can do the same thing in optical measurements. The idea of this quotation actually works if you do it backwards; for example, good wide range position measurements can be made with an interferometer by counting fringes, or by subranging with a tunable source. If you have a 20 dB SNR, a straight measurement will get you an RMS accuracy of 7% in amplitude and 70 mrad in phase (Section 13.6.9), which is no great shakes. One solution is to use a two-wavelength interferometer, whose output is periodic at the beat length λ_B , where

$$\frac{1}{\lambda_B} = \frac{1}{\lambda_1} - \frac{1}{\lambda_2}. \quad (10.5)$$

Stepping λ_B by factors of 3 or so and measuring the phase residual at each step will give you the absolute distance to any desired accuracy, if you know the laser frequencies accurately enough. If you count fringes instead, that 20 dB is enough that you don't lose count under reasonable circumstances, so by counting fringes throughout the motion, you can measure long distances to an accuracy approaching $\lambda/10$ —once again, if you really know the laser frequency.

10.4.3 Fringes

Fringes and speckle corrupt all sorts of measurements, and especially bright-field laser measurements. With some work, these nuisances can be turned to good account; adding a piezoelectric mirror can make a phase-shifting interferometer, and a TV camera and video capture card can turn a HeNe laser into an electronic speckle pattern interferometry (ESPI) system that can measure surface deformations both in-plane and out-of-plane.

10.5 GETTING MORE SIGNAL PHOTONS

A bit of water and fertilizer can make the grass taller; an extra inch or two of height can make the difference between having a measurement and not having one. Be careful to keep the fertilizer out of your apparatus and especially your published data.

10.5.1 Don't Throw Photons Away

Thrift is a virtue in optics as in life. You can get more out of your photons by steering each of them to its appropriate bucket, rather than just throwing away the ones you're

not instantly interested in. In a flood-illuminated situation, where the light source is large in area or wavelength spread, you can't get all the photons into one resolution element. Use a staring detector that measures all the grass at once, rather than scanning one blade at a time.

It's okay to look at only one (band, pixel) combination at a time if your light source is doing the same (e.g., laser spectroscopy). There are lots of suggestions starting in Section 10.2 on how to reduce the background and get more photons.

10.5.2 Optimize the Geometry

The optimal measurement geometry depends on the interaction strength. Optimal sensitivity is reached at an absorption of order 1 (i.e., a transmission of order $1/e$). Some samples are so absorbing that we have to use attenuated TIR to look at their absorption spectra; we bounce our light off the hypotenuse of a prism, so that it is totally reflected from the interface, making the effective path length in the absorber very small.

Others are very weak, for example, the O_2 overtone absorption line at 760 nm, which is useful for measuring oxygen concentration with a diode laser. Those need a multiple-pass cell or a resonant cavity to provide enough interaction length.

10.5.3 Use TDI

Time delay integration (TDI) is a sometimes useful scheme for making 1D CCD detectors less bad at imaging areas. The idea is to make a narrow 2D array, maybe 20 pixels wide and 4096 long, and shift it in the frame direction at just the right speed that the motion of the image and of the photoelectrons is synchronized. Ideally, that gets us $20\times$ more photons per pixel with no other worries except a longer integration time. Unfortunately, it isn't easy to make it work so neatly; the tolerances on the scan speed, clock rate, and the geometrical distortion of the lens are very tight, so the SNR improvement is less and the image degradation more than you might think. Altogether, TDI arrays are for people with big budgets and lots of patience.

10.5.4 Consider OMA Spectroscopy

The light from a grating is spread out linearly with wavelength, and so can be sampled with a 1D multielement detector such as a CCD or photodiode array, an arrangement called an optical multichannel analyzer (OMA). These have much better efficiency than monochromator types, but as there is no exit slit or second grating for protection against stray light, they are much poorer when the dynamic range gets large. Their typical 1024 element detector provides about 5 \AA resolution across the visible; since silicon detectors are insensitive beyond $1\text{ }\mu\text{m}$ and IR arrays are expensive, insensitive, or both, nearly all OMAs are visible-light devices, with a detection range of 400–800 nm. Most OMAs operate at room temperature, which raises the dark current of the detector by orders of magnitude, and tend to have 12 bit ADCs, so that their dynamic range wouldn't improve much even if they were cooled. High performance OMA spectrometers with cooled detectors can provide shot noise limited spectral measurements, at least in dim light. Since we're still limited by the étendue of the entrance slit, that still isn't that great, but it's fine for lots of things.

10.5.5 Consider Slitless Spectroscopy

If we're doing spatially resolved spectral measurements (e.g., spectral reflectance to get film thickness), we can get more photons by removing the output slit of the monochromator and shining the dispersed light directly on the sample; provided that it's in focus, every point on the sample will be illuminated with a definite wavelength. Scanning the grating makes the pattern move across the sample. A camera looking at the sample will therefore get lots more photons, and the measurements can be figured out afterwards in software.

10.5.6 Consider Fourier Transform Infrared (FTIR) Spectroscopy

We saw in Section 1.3.8 that the power spectrum of a signal is the Fourier transform of its autocorrelation. A two-beam interferometer with nonzero path length produces a DC term, plus an AC interference term whose time average is the autocorrelation of the input radiation. Accordingly, measuring the autocorrelation at appropriately chosen path differences and taking the Fourier transform numerically gives us the power spectrum of the radiation.

This approach has two important benefits, which in spectroscopy are customarily known as the *Felgett advantage* (frequency multiplexing) and the *Jacquinot advantage* (étendue). The Felgett advantage comes from all wavelengths being detected at a duty cycle of 100%, which improves the shot noise—we don't block any photons with input and output slits of a spectrometer. The Jacquinot advantage comes from the fact that aperture angle enters only quadratically into the optical path length of the interferometer, and hence into the frequency resolution, whereas there is a linear trade-off between resolution and slit width in a monochromator system; thus the FTIR's detection NA can be widened enormously. It's as if the slit width were more like slit length; instead of a long narrow slit, you can use an aperture as wide as it is tall, and you can imagine how much brighter that is. The two produce an efficiency advantage on the order of 10^5 in a typical FTIR versus a Czerny–Turner, and 200 compared with an OMA type. There is also no grating order overlap to worry about, though you do have to move the interferometer mirrors very precisely, scanning is relatively slow, and the shot noise of bright components can corrupt the measurements of weak ones.

You get all this with a single-element detector, out there in the IR where even poor detectors cost a lot. It's a remarkable measurement idea altogether.

Aside: FTIR Errors. The simple account of FTIR operation given above ignores the infrared self-luminosity of the detector, mirrors, and beamsplitter. In the mid- and far-IR, FTIRs have to take account of these effects, which require careful calibration and a significant amount of thought.

10.5.7 Use Laser Scanning Measurements

With all that has gone before about widening our apertures and using staring sensors, it may seem contradictory to recommend a flying-spot laser scanner as a good way to get more signal, but nevertheless, it is. The reason is that although each point on the sample sees the beam only once per frame, a duty cycle of perhaps 10^{-6} , the beam is very very bright during that time, so its total dose per unit area is the same, and hence the counting statistics are too, for constant collection NA.

The other benefit is that the detector sees a very brightly illuminated sample point at all times, and the incoherent optical background signals get detected with that 10^{-6} duty cycle too, assuming of course that the baffles on the receive side are that selective, or that there is some heterodyne gain lifting the signal up. The detector side can have a high NA but a lowish étendue, because it is being descanned.

The very short dwell time per pixel also eliminates speckle (because only one sample point is being illuminated, it has nothing to interfere with to make speckles). There is no blur due to finite integration time, because the dwell time per pixel is 10^{-6} frame times. (Grass has to be blowing around pretty fast to cause much blur in a 50 ns integration time.) Furthermore, flying-spot measurements are naturally suited to signal averaging.

Very high contrast samples are hard to work with in a full field system, because of scattered light and possibly saturation, and things like Raman spectroscopy have such an enormous dynamic range that they are not possible without rejecting the unshifted beam with a sharp interference filter. Due to angle tuning, filtering is a lot easier in the collimated light you get after descanning. Selectivity is far better with scanning because the interfering area is not illuminated—no glints, no speckle—you name it, it's gone.

10.5.8 Modify the Sample

Sometimes you can modify the sample to make it easier to measure, either by preparing it differently, changing its size or concentration, or, as in this example, controlling its motions.

Example 10.4: Sub-Doppler Spectroscopy. Example 1.1 describes two-photon Doppler-free spectroscopy. It's an optical hack for eliminating Doppler broadening, which is limited to two-photon transitions. Other methods usually concentrate on making the molecules' radial velocities very small. This is done by supersonic cooling, which makes the random velocities small, or by the atomic fountain, where very cold gas rises and falls under gravity, with the measurement occurring at the peak of the motion. These are cryogenic, high vacuum techniques. Another technique is trapping, where molecules are held in a small volume for a long time. Ions can be trapped electromagnetically, and neutral species with *optical molasses*. In the optical molasses approach, the molecules are held at the intersection of three sets of counterpropagating beams, tuned just slightly above a strong resonance of the molecule. A molecule that is moving will be Doppler shifted closer to resonance and will absorb photons from the beams opposing its motion. It will reradiate into 4π steradians, so that the expectation value of the momentum change is just the momentum of the laser photon. This slows the particle down. There's a certain irreducible residual motion, since the random character of the photoemission means that the resulting impulse never quite cancels the particle's momentum.

Optical molasses is not a true trap, because there's no potential well, just optical friction. Particles can diffuse out of it, but it's a good way of making very slow-moving molecules.

10.5.9 Corral Those Photons

There are often geometric limits to how many photons we can easily recover. If we're using a small light source on a specular object 10 meters away, we'd need a truly huge

collecting system to achieve decent photon efficiency because most of the light is reflected away. A partial answer is retroreflective tape, which collects all the photons that hit it and bounces them more or less back into their source; the difference in signal power can easily be 60 or 70 dB (see Sections 7.8 and 11.8).

Aside: Video Techniques. Talking about staring measurements inevitably brings up video. Properly designed CCD detectors are excellent devices, but video cameras are another matter. Their automatic gain and black level controls corrupt our attempts at calibration, which makes intensity measurements difficult. (See Sections 3.9.1 and 3.9.14.)

Besides mapping intensity, cameras are used a lot in full field interferometric techniques (phase shifting, moiré, and holographic), and with *structured light* measurements. For example, consider a fan-shaped beam projected on a curved surface and looked at with a camera placed out of the plane of the fan beam; wiggles in the image of the line correspond to variations in height. With a camera as detector, even quite large variations in the surface (e.g., tubes of different diameters) can be accommodated without changing the setup, which is a very valuable attribute, but the attainable precision is poor.

In these situations, the efficiency and repeatability of the camera as a photon detector are of secondary concern, but the geometric distortion caused by the camera lens is often not properly accounted for. This distortion is a few percent for most lenses, and as much as 10% in a wide-angle, which makes a precise geometric measurement using a camera lens a nontrivial exercise. This is particularly so if the focus is being adjusted, or if the surface is going in and out of focus due to large curvature. If the camera is seeing the surface at an angle, the magnification will be a function of position; this is responsible for the familiar “keystone” effect with projectors.

The best way of using a camera for surface measurements is to keep the axis of the camera lens normal to the surface, to eliminate keystone and minimize defocus, and carefully calibrate the geometric distortion of the lens at different positions and different states of focus. A telecentric lens makes this dramatically easier, since to leading order the magnification is independent of focus; you get the equivalent of an isometric projection—see Section 9.3.7. Failing that, a careful 3D calibration is needed, which is sufficiently painful that you won’t want to do it twice.

10.6 REDUCING THE BACKGROUND FLUCTUATIONS

It really isn’t necessary to get rid of the background altogether, as long as its fluctuations aren’t limiting the measurement unnecessarily. Getting bright-field measurements to do as well as ideal dark-field ones with the same NA requires getting the background fluctuations down to the shot noise. Fortunately, that’s perfectly possible a lot of the time.

10.6.1 Beam Pointing Stabilization

Active-optics beam pointing stabilization systems are based on shining a sample of a beam into a quad cell and tipping a mirror or tuning AODs to keep it still. In a limited bandwidth, they can do a very good job. One limitation is the interaction of angular and position wobbles; focusing the beam on the quad cell with a lens will make the system

ignore wobbles in position but correct pointing errors. Operating near the null point makes the beam pointing insensitive to RIN, so that with enough control bandwidth, only shot noise and beam profile changes are left to cause errors.

10.6.2 Beam Intensity Stabilization

You can get feedback controlled attenuators that adjust the intensity of the laser beam to keep it constant, e.g., Noise Eaters. They're great when the beam intensity is what matters, for example, in dark-field measurements, laser heating, distribution of beams to several places, or pumping other processes. They can do a great job of stabilizing dark-field systems, because the shot noise limited SNR is much lower than that of the full beam (so it's easier), and because they get rid of the low frequency multiplicative noise, which is usually the main culprit in dark field.

Unfortunately, they are not able to get the noise down to the shot noise of the beam, which is why they're not a complete solution. There are two reasons why. First, they're feedback devices that rely on the loop gain to suppress noise (unlike noise cancelers and differential measurements, which rely on matching instead), and they don't have enough gain and bandwidth. The second reason is more fundamental and affects noise cancelers and ratiometric measurements as well; if we're using beam *A* as a reference to stabilize beam *B*, the shot noise of the two beams is uncorrelated. Thus by a standard propagation-of-error calculation (you remember—take partial derivatives and multiply by the corresponding rms noise, then take the RMS sum of all contributions), assuming we're well inside the bandwidth, the shot noise limited CNR_{out} of the stabilized beam obeys

$$\frac{1}{\text{CNR}_{\text{out}}} = \frac{1}{\text{CNR}_A} + \frac{1}{\text{CNR}_B}, \quad (10.6)$$

where CNR_A and CNR_B are the shot noise limited CNRs of the two beams. There is no way to stabilize a beam down close to the shot noise this way without wasting a whole lot of it; a 50% loss will get you to 3 dB above the shot noise.

10.6.3 Photocurrent Stabilization

For bright-field measurements, we have to cope with the shot noise of the beam anyway, so there's no need to stabilize the strength of the *beam*, just the photocurrent. Stabilizing the beam is one way to do that, but it's the hard way. An alternative is to use a differential or ratiometric measurement of the photocurrent, where a comparison current derived from a sample of the source beam is subtracted or divided out. Since both beams will see the same fractional noise, this will leave a quiet result. We'll leave subtraction for Section 10.8 and the detailed discussion of the principle to Section 18.6.3. This is a very powerful technique that should be considered in any laser measurement.

10.6.4 Ratiometric Measurements and Phase Errors

The ratiometric approach promises to correct for source intensity variations, and hence eliminate both additive and multiplicative source noise at a stroke. Examples of ratiometric techniques are analog division, deriving the ADC reference voltage from the source

power rather than from a voltage reference, or using the log ratio output of a noise canceler. We'll see in Sections 15.11 and 18.6.2 that divider ICs have very limited dynamic range and are slow and nonlinear, so they aren't as attractive as they look, though they are adequate for some things, especially slow measurements with quiet sources. They work OK for taking out 120 Hz from tungsten bulbs, for example.

For slow things like that, the trick of deriving the ADC reference voltage from the comparison photocurrent is also competitive; you do have to worry about matching the phase shifts in all the signal paths, though, or you may wind up making things worse rather than better—the more circuitry there is in the way, the harder the matching is to get. Since the phases have to be matched to 0.01 radian (0.6°) to get 40 dB noise suppression, and op amps have phase shifts of about f/f_c radians (where the -3 dB frequency f_c is not very well controlled), this gets hard very rapidly for signals above 1 kHz if each current has its own amplifier. Dividers have separate sections for numerator and denominator, so they suffer from this even if you wire the two in parallel for test purposes. It's possible to put in a phase tweak, but since it doesn't usually have the right frequency dependence, you can't get that far with it and it isn't especially stable with time and conditions; we're usually talking about subpicofarad strays, and those will change with time, temperature, and humidity.

The logarithmic output of the laser noise canceler doesn't have these problems, for two reasons: (1) the photodiodes are basically wired in parallel for AC, so the strays and phase shifts have little opportunity to differ; and (2) the balanced subtraction gets rid of the additive noise at all frequencies, irrespective of the feedback loop bandwidth, so all the loop has to cope with is the multiplicative noise, which is usually some orders of magnitude smaller and concentrated at baseband where the loop has an easy time. See Sections 10.8.6 and 18.6.3 for the details.

10.6.5 Changing the Physics

Sometimes we can solve a problem by changing the measurement physics. For example, consider a spectrometer for measuring impurity concentrations in ultrapure water. To get decent sensitivity, we'll need a long interaction length, which typically means a multipass cell, with mirrors bouncing the beam back and forth many times through different parts of the sample cell. However, even ultrapure water has a strong scattering background from Rayleigh scatter and bubbles, which depends on the temperature and pressure histories, dissolved gas concentration, hydrodynamic vibration, and other things that we don't regard as data. Since the scattered light misses the detector, it shows up as a spurious absorption signal and is difficult to correct for.

Now consider the same sample cell inside an integrating sphere, with source and detector shielded with baffles. The white walls of the sphere cause the light to make many passes through the sample before being absorbed, so the extra sensitivity of the multipass arrangement is preserved. However, the same walls homogenize all the light into π steradians every 15 cm or so, so the scattered photons are no more likely to miss the detector than the unscattered ones, and the scatter signal goes away almost entirely. The photon efficiency of such a measurement is of course much lower than in the multipass cell, where almost the entire beam goes into the detector, but a few inexpensive photons are a small price to pay for a stable and sensitive measurement.[†]

[†]Private communication from Mark Johnson.

10.7 OPTICALLY ZERO BACKGROUND MEASUREMENTS

An optically zero background measurement is one where the unaltered illumination beam doesn't get to the detector—in other words, we're talking about dark-field measurements of one sort or another. These are great if you're either in the photon-counting regime or the $>1\ \mu\text{A}$ regime, but not so nice in between. The analogy is climbing the building to measure the grass; you probably get a better measurement, but your footing may be a bit precarious.

10.7.1 Dark Field

Ordinary dark-field microscopes aren't actually all that dark because of scatter, but they work pretty well visually. A *doubly-dark-field* measurement is one that uses dark-field illumination plus crossed polarizers, looking for specific kinds of problems such as diagonal scratches or particles that are diffuse scatterers.

10.7.2 Fringe-Based Devices

Two laser beams producing a fringe pattern within the sample space can do some very useful things. The simplest one is time-modulated scattering, as in *laser Doppler velocimetry* (LDV), which sends aerosol particles along $\Delta\mathbf{k} = \mathbf{k}_2 - \mathbf{k}_1$ and measures the resulting amplitude-modulated light pulses with one or two photomultipliers in a dark-field geometry; the frequency of the pulses gives the velocity, and (interestingly) the phase shift between different scattering directions tells something about the particle size. Using variable-pitch fringes to encode measurement data is also possible; if the LDV fringes were made from two spherical waves instead, it would be possible to work out where the particle went through the beam by the change of frequency with time.

A more complicated one is the *photochemical grating*, where the fringes pump some process in the sample, making a grating structure that can diffract a probe laser beam, with the diffracted beam being the signal. This is a bit like a thermal lens measurement, in that the grating can be made in the IR and the probe in the visible. The biggest advantages are that the signal beam winds up going off away from the other beams and that its amplitude is zero when there's nothing to detect. That makes it easy to detect very small signals. Fringe-based measurements are a spatial version of modulation mixing (see Sections 10.9.4 and 13.10.6).

10.7.3 Monochromator-Based Measurements

The classical way to do spectroscopy is a scanning Czerny–Turner spectrometer with a single-element detector, looking at the emitted or transmitted light from some sample. This is an enormous waste of good photons. Nobody in his right mind would do that today except with very bright light, or in very special circumstances, for example, Raman spectroscopy where the close proximity of the enormous unshifted line makes double and triple Czerny–Turners very helpful.

10.7.4 Fluorescence and Photon Counting

Fluorescence solves the source shot noise problem and ideally has the same sensitivity as a dark-field or coherent detection measurement with the same detection NA.

The improvements in practice come from being able to reject the incident light via highly effective optical filters, and being able to distinguish one sort of molecule from another—the ones we care about label themselves for us. Thus fluorescence measurements can be highly sensitive. Especially with the new development of negative electron affinity (NEA) photocathodes (see Section 3.6.1), a fluorescence detection instrument can detect more than 25% of all photons emitted by the fluor.

The bad news is that the quantum yield of the fluorescence process—the ratio of all fluorescent photons emitted to photons absorbed—is often very low, in the 10^{-5} range, although some highly efficient fluors such as rhodamine laser dyes are a lot better, a few percent to tens of percent. On top of this, real samples are usually optically thin at the pump wavelength, so only a small percentage of incident photons are absorbed. This means that photon-counting detection using PMTs must usually be used.

Laser dyes and other organic fluorophores tend to die at around 10^6 emitted photons per molecule, because the upper state of the fluorescence occasionally decays by electron transfer (i.e., chemical reaction).

Fluorescence methods are best used where the desired species can be made to fluoresce efficiently, usually by being labeled with a specially chosen marker such as fluorescein. Thus they are used extensively in biochemistry. Other uses are atomic fluorescence measurements, where an atom or molecule in the gas phase is pumped with one wavelength and detected at another. By chopping the pump beam, the desired signal can be made to appear at AC, with the background left at DC. Of course, this assumes that the pump has no detectable energy at the probe wavelength, and that there is no residual fluorescence of other components in the system at the probe wavelength. Colored glass filters and plastic or glass components exposed to pump radiation must be scrutinized especially carefully.

Putting filters in the wrong order can cause enough filter fluorescence to get through that the measurement becomes impossible—see Section 5.5.1.

10.7.5 Nonlinear Measurements

A nonlinear generation measurement produces no output signal with no sample, which makes it a dark-field measurement. In fact, a lot of the time they produce no output even with a sample, but you can make them if you try hard. Examples are second harmonic generation (SHG) and Raman spectroscopy. Raman is a special case, since separating out the unshifted light is a big job—it's very bright, and very nearby.

10.7.6 Nonoptical Detection

Some kinds of samples are kind enough to work as detectors for us; examples are photoionization spectroscopy, photoacoustic measurements, and some ultrafast ones such as picosecond electro-optical sampling, where the output is a short pulse on a transmission line. These count as zero background only because we think of the whole output as the signal.

10.7.7 Active Fringe Surfing

You can get a more-or-less dark-field measurement by actively stabilizing an interferometer on a dark fringe. You can do this by dithering one arm slightly with a liquid crystal

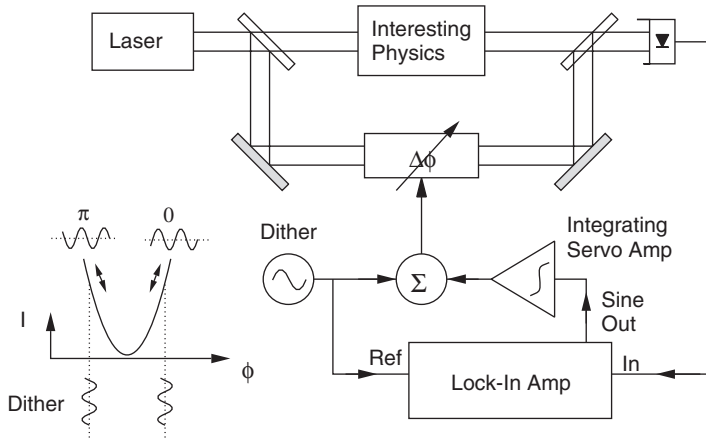


Figure 10.7. Fringe surfing: the phase of the fundamental signal changes 180° going through the null, so a lock-in can keep us right on the null.

phase shifter and using a lock-in amplifier to sense when the fundamental signal goes through 0, as shown in Figure 10.7. The fundamental signal goes as the first derivative of the fringe slope, so it goes through zero and changes sign at the null; thus a lock-in produces a properly bipolar error signal, and a simple integrator will stabilize the operating point. You wire it up, close the loop, and *presto*, you're sitting on a dark fringe. Using the control voltage as the output, signals well inside the feedback loop bandwidth will be detected, and those far outside will be rejected. Using the lock-in output, it's the other way round. Either way, additive laser noise (including shot noise) will be greatly reduced, because most of the light is going somewhere else.

Everyone loves fringe surfing (probably because most of us have invented it independently at some point, see Section 8.16.3). It's deceptively easy; you see, the surfing part is well behaved—it's the dark fringe itself that is noisy and poorly defined. When you think about it, it's really pretty silly to servo on a dark fringe, because the SNR there is zero (not 0 dB, *zero*). It's also a tough thing to do in a real instrument, because it's mechanically flaky and you'll lose lock if the phase shifter goes out of range, which it's going to, probably at an inconvenient time. Phase shifters are nonlinear, so even if it comes into lock somewhere else, it isn't simple to stitch together data from the different operating points (see Section 8.16.3).

The other things to remember are that the coherent background you get from working off the null actually improves the SNR a great deal, as we saw in Section 10.3.5, and that the laser intensity noise can be got rid of with a laser noise canceler. Thus if you must surf fringes, it's much better to do it in bright field (for more details, see Sections 10.8.3 and 10.8.6).

10.7.8 Polarization Tricks

High school and undergraduate optics labs are full of crossed-polarizer experiments, where some sample such as cellophane tape or a plastic model of some architectural structure is placed between crossed polarizers, and the colors of the output are photographed. This is a famous old trick for finding stress concentrations in 2D models.

Dithering the polarization with a liquid crystal retarder or a Pockels cell allows the same sorts of tricks as in the previous section, and it's more useful since a lot of the time the background light isn't coherent with the signal, especially if we're using a thermal source with appreciable NA.

10.7.9 Optical Time Gating

In a pulsed measurement, the background can often be eliminated by time-gating the detection side appropriately, for example, an OTDR, where the huge pulses from the ends of the fiber and from the connectors will saturate the detector and so mask nearby signals. An APD or MCP detector can be turned on and off very rapidly, so this isn't hard to do.

10.8 ELECTRONICALLY ZERO BACKGROUND MEASUREMENTS

There are lots of methods for reducing the influence of noise in measurements, many of which are explained in Chapters 13 and 17, and most of which boil down to filtering in one form or another. Here we'll concentrate on how to keep that gunk out of the data in the first place. These methods get rid of the excess noise of the background but are powerless to reduce the shot noise. They are thus quite competitive with optical background reduction approaches for coherent bright field, but distinctly inferior for dark field or when there's an incoherent background. The analogy here is looking from a neighboring building or a helicopter; if the two buildings are swaying in the wind, the measurement isn't as stable, although much better than from ground level, because the angular size of the grass tuft is much bigger.

10.8.1 Polarization Flopping

In polarizing measurements such as Nomarski, a ferroelectric liquid crystal half-wave plate between the Nomarski prisms will invert the image contrast when its state changes (see Example 7.2). Using a CCD detector or instrument-grade camera (*not* an auto-gain, auto-black-level video camera) and flopping the phase on alternate frames allows separating the amplitude and phase information; summing alternate frames gives amplitude, and subtracting them and dividing by their sum gives $\cos \phi$. You need more phases than that for a real phase measuring system (at least three phases spaced around the unit circle), but just separating out amplitude and phase is often good enough.

10.8.2 Electronic Time Gating

If the pulse interference can't be gated out optically, there are other things we can do, for example, an IF noise blanker, which looks for noise pulses above some threshold, and cuts off the IF gain when one arrives; this is how some car radios deal with ignition noise, for example.

If the signal pulses have a low duty cycle, then turning off the receiver in between reduces the noise power we detect by a factor of the duty cycle; of course, the bandwidth has to be wide enough to reproduce the pulse, but we're normally stuck with that anyway. The ping-pong gated integrator idea from Section 15.5.6 is one example; even though

the noise is being integrated 100% of the time, if there's no signal pulse, we just throw the result away, so it doesn't reduce the measurement SNR.

10.8.3 Nulling Measurements

Fringe surfing doesn't have to be around a dark fringe; since an interferometer has two output ports, we can servo the phase of one arm to keep the two output powers exactly equal. Although the shot noise of the two arms of course adds, this isn't a serious limitation because of the Bright Field Rule of Section 10.3.6.

It is also possible to use this idea to null out other effects. The long-distance particle counter of Figure 10.8 is based on extinction (Figure 10.9) and uses an AOD at a pupil to make a laser beam wiggle back and forth sinusoidally. Because the AOD is at the pupil (where the beam comes to a focus in this case), when the beam is recollimated by the next lens, its motion becomes a pure translation, with no angular scan (it is intended for use in a 30 m long by 15 cm high belt oven full of hydrogen at 1000 °C). A particle will be in the beam part of the time, and out of it part of the time; thus it will cause a differential extinction signal at the scan frequency. Dust on the optics, minor misalignment, and varying Fresnel reflections cause the output to have a large background component at the fundamental, which will mask the small (10^{-6}) transient extinction signals of the particles. It can be nulled out with a lock-in sensing the fundamental, connected to an integrating servo amplifier that controls the center frequency of the scan. If there's

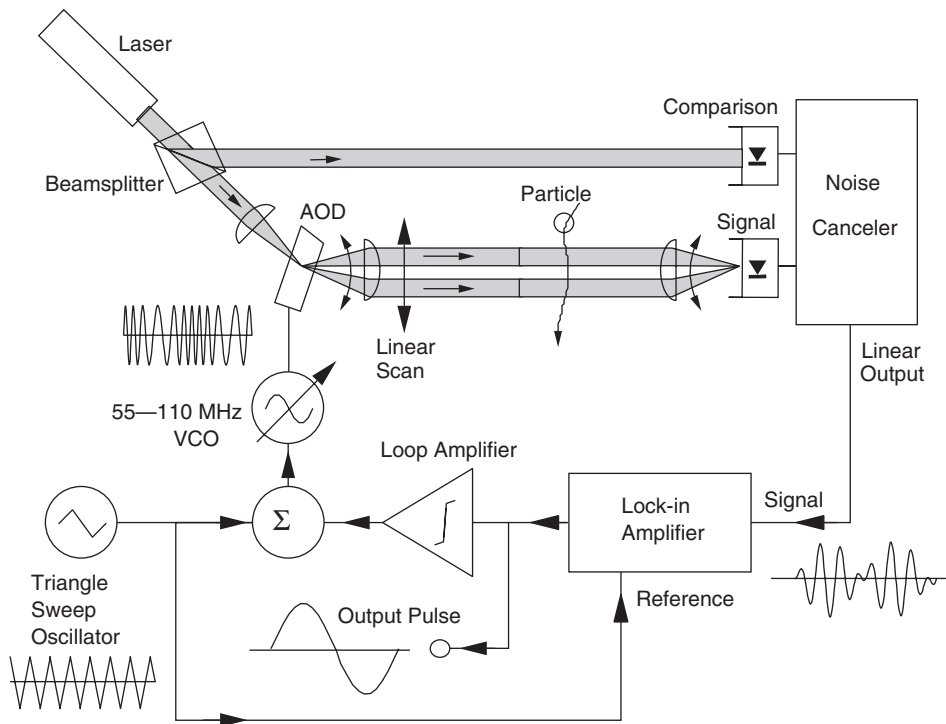


Figure 10.8. This acousto-optically scanned long-distance particle counter uses the parabolic shape of the diffraction efficiency curve of the AOD to null out any static background at the fundamental (e.g., due to vignetting or dust).

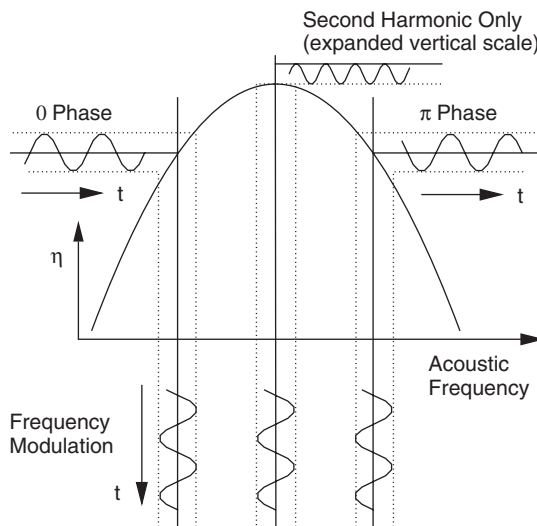


Figure 10.9. Operating principle of the extinction system of Figure 10.8: the sign of the fundamental depends on which side of the diffraction peak it's on, so nulling gets rid of it.

more extinction on the positive half-cycle, the center frequency will slide slightly down the low frequency skirt of the diffraction efficiency peak, causing just enough variation to cancel the background. The lock-in's output is a highpass filtered version of the extinction signal from the particles. This is dramatically better than surfing a dark fringe, because the SNR is so much higher. As usual, analog techniques are much superior in nulling measurements—simpler, faster, quieter, and more accurate. Null first and digitize second.

10.8.4 Differential Measurements

Interferometers have differential outputs, as we've seen, and there are lots of other cases, too. Dual-beam spectrometers, where one beam passes through the sample and one doesn't, are another example. Subtracting two carefully balanced photocurrents ideally cancels the source excess noise exactly, at least if the noise in the two beams hasn't been decorrelated by vignetting or etalon fringes (which turn FM noise into AM). In the Nomarski case, the normalized difference signal corresponds to the phase, and this is true of other interferometers as well, when they're working near null, and the total phase excursion is small. As we've seen elsewhere, the problem is how to keep them in balance.

10.8.5 Other Linear Combinations

Mere differences don't exhaust the usefulness of linear combinations. An interferometer has output beams, $a(1 + \cos \phi)/2$ and $a(1 - \cos \phi)/2$, so that their sum is a and their normalized difference is $\cos \phi$. Operating near $\phi = \pm\pi/2$ gives an electronically zero background phase measurement, even if we don't divide out the amplitude, and is linear over a range of $\pm 5^\circ$ or so, which is frequently lots, especially in common-path

interferometers where the phase drift is small. The beam intensities will need a nulling adjustment.

10.8.6 Laser Noise Cancelers

As we've already noted in Section 10.6, the laser noise canceler is capable of making bright-field laser measurements equivalent to ideal dark-field ones in practice, as they are in theory. A good example of this is tunable diode laser absorption spectroscopy, as shown in Figure 10.10.[†] The diode laser is tunable over a wave number or two by changing its bias current, but in the process its output power changes by a factor of 2, which makes gigantic additive and multiplicative backgrounds. An extremely simple optical system, consisting of a Wollaston prism, one polarizer, a sample cell, and a noise canceler, can make absorption measurements down to the 10^{-6} range even in the face of huge variations like this.

The circuit details are in Section 18.6.3, but the main point is that the automatically balanced subtraction gets rid of the additive noise, and the log ratio output takes this additive-noise-free signal and gets rid of the low frequency multiplicative noise as well, so that the peak heights are independent of the laser power, and does it without the serious noise and accuracy limitations of analog dividers, or the bandwidth, phase shift, and dynamic range problems of digital methods based on numerically dividing the two detected signals. Etalon fringes can be subtracted out by admitting air into the sample cell, which pressure-broadens the real absorption into oblivion, leaving only the etalon effects, which can then be subtracted out. The before and after pictures are shown in Figure 10.11. The two just-discernible bumps on a huge sloping background have turned into a well-resolved spectrum with ppm sensitivity and a flat baseline.

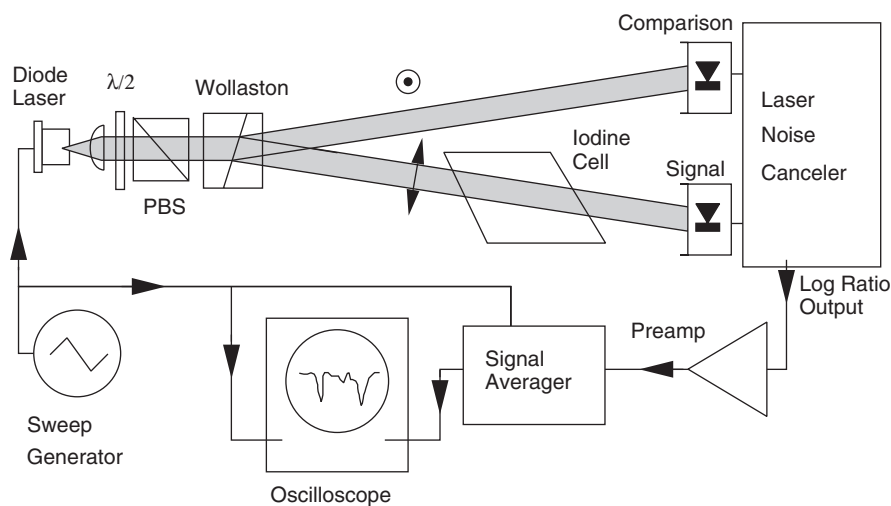


Figure 10.10. Tunable diode laser spectroscopy with a laser noise canceler.

[†]Kurt L. Haller and Philip C. D. Hobbs, Double beam laser absorption spectroscopy: shot-noise limited performance at baseband with a novel electronic noise canceler. *Proc. SPIE* **1435**, 298–309 (1991). (Available at <http://electrooptical.net/www/canceller/iodine.pdf>.)

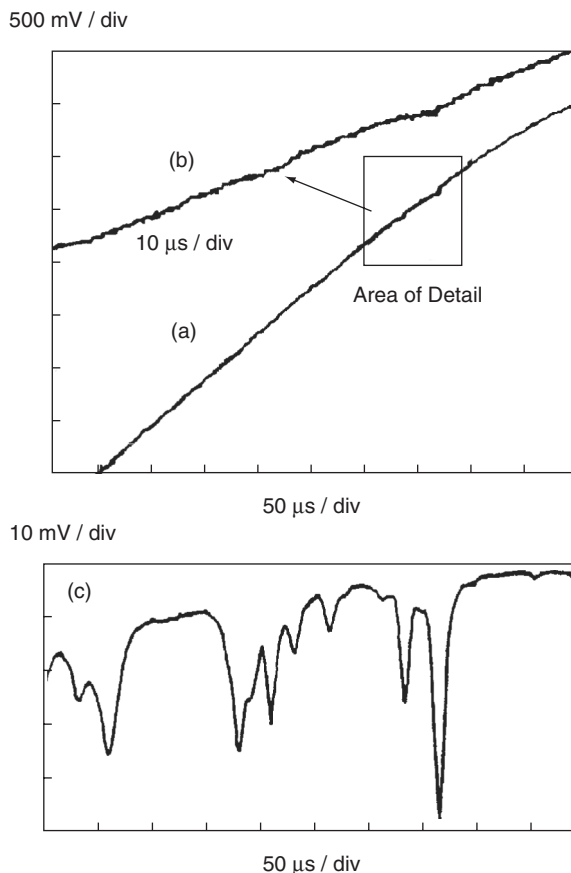


Figure 10.11. Before and after pictures of the signal from the tunable diode laser spectrometer: (a) signal-averaged photocurrent, showing the huge background, (b) detail of (a), and (c) the same spectral region as (b), seen by the log ratio output of the noise canceler; the largest peak has about 1% absorption. Other peaks were observed at 12 ppm absorption with good SNR.

Unfortunately, noise cancelers don't work as well with fiber systems, because of fiber's nasty way of producing huge amounts of coherence fluctuation noise through etalon fringes and double Rayleigh scatter (see Sections 2.5.3 and 8.5.13). Different fibers, or even different polarizations in the same fiber, have different fringe phases, so the noise they produce is uncorrelated and hence uncancelable. Once again, this is a beam problem, not a circuit problem—the noise canceler gives you a really great measurement of the uncorrelated part of the noise. This probably isn't what you want, but it's as much as any technique can do with such badly corrupted beams.

10.9 LABELING SIGNAL PHOTONS

Painting the grass fluorescent orange is useful in two ways; we can wait until night time, shine a black light at it, and look for its fluorescence through an orange filter (a dark-field technique), or we can do it in daylight and do some sort of spectral differencing

measurement, where grey things like concrete cancel out but orange things don't (a photon-labeling technique).

Labeling good photons may let us subtract off the background photocurrent, but not of course its shot noise—it's a lot like the electronic zero background approaches, which isn't too surprising when you think about it. Like them, photon labeling is second best unless we're doing a shot noise limited bright-field measurement. There are lots of ways to label photons, so if you can, try to use ones that preserve your SNR by not wasting most of the power in the carrier.

10.9.1 Chopping

The naïve approach to laser-based AC measurements is to chop the laser beam and use a lock-in to look at the amplitude of the chopped signal. It does avoid electronic drift and $1/f$ noise, but that is very rarely the problem. It doesn't help optical stability one bit, because all the low frequency laser junk just gets translated up to the chopping frequency; AC measurements help only when they treat the signal and noise differently. If the reason this doesn't work is unclear to you, you may wish to read Chapter 13 carefully before proceeding.

If you can chop the right thing, though, this can be a good technique. It's rarely the best available, because the frequency shifts you can get are modest, the photon efficiency poor, and mechanical jitter high. Acousto-optic and photoelastic modulators make better choppers than choppers do, at least with lasers.

Chopping the right thing means chopping the sample. One example is a pump/probe measurement, where a chopped beam pumps the sample and another beam measures the changes so produced. Another would be scanning on and off the sample, or moving the sample very rapidly in and out of the beam, e.g. with kilohertz microfluidics. Chopping isn't just for laser experiments; infrared astronomy would be almost impossible without it.

Example 10.5: Chopping Secondary IR Telescopes. Infrared astronomers have a tough job. From a 270 K planet with a 100 K sky temperature, using lousy detectors, they have to see the IR emission from a protostellar nebula a kiloparsec away or a primeval galaxy on the rim of the universe. These signals are far below the sky brightness level, and that sky brightness changes with time and atmospheric conditions (the situation is redressed in part by being able to do it during the day as well as at night—the λ^{-4} dependence of scattered intensity chops the Rayleigh scatter down pretty far in the IR). The solution they adopt is the *chopping secondary*. The telescope's secondary mirror flips back and forth very slightly at a rep rate of a hertz or so. The detector accordingly sees the source alternating with a patch of sky a few arc seconds away, whose background level is very nearly the same. The sky brightness signal is thus DC, and the source signal is AC. Using two detectors works even better, since no signal photons are lost, and the background gradient can be removed too.

One really important caveat is to make sure that you have enough actuator and detector bandwidth that the chopped signal has a chance to settle down on each cycle. Otherwise you'll find yourself measuring the slew artifacts rather than the signal.

10.9.2 Scanning

Some classes of measurement rely on making the sample look different at different times, so that the signal alone is modulated in a distinctive way. The simplest one is scanning (see Section 7.9). Scanning can be done in x or λ , making the signal periodic at the scan frequency; signal averaging will select the signal and leave the low frequency noise and instability behind (Section 13.10.4). If it's possible to keep the optical system constant and scan the sample, either by moving it on an accurate translation stage or changing some other sample property, the background is more nearly constant, and hence easier to get rid of. Any scanning artifacts (e.g., changing etalon fringes, vignetting, or laser power variations) will usually be indistinguishable from signal, so we have to go further.

Continuous scanning produces an ordinary sampled measurement (see Section 17.4.3), but a stepped scan requires that the electrical bandwidth is wide enough so that the signal settles accurately between samples. A triangle-wave scan is better than a sawtooth, because settling artifacts show up clearly when the trace and retrace don't line up with each other. A combination of scanning and chopping can work well too, especially if you have a rapidly moving baseline; if you have random access to your sample points, instead of scanning $p_0, p_0 + \Delta, p_0 + 2\Delta, \dots, p_0 + (N-1)\Delta$, you can go $p_0, p_0 + \Delta, p_0 + 2\Delta, p_0 + 3\Delta, p_0, p_0 + 3\Delta, p_0 + 4\Delta, p_0 + 5\Delta, p_0, \dots$. That way, every frequency is measured twice at different times in the sampling window, and there are a whole bunch of baseline correction points; discrepancies due to settling funnies become obvious, and the baseline error can often be fixed very accurately. Do keep the bandwidth wide enough for all that jumping around, though.

10.9.3 AC Measurements

The most common way of labeling photons is by making them arrive at some modulation frequency sufficiently far from DC that the worst junk can be filtered out. Heterodyne measurements are great at this, but modulation techniques are widely used too, for example, FM derivative spectroscopy, in which a small, very fast dither i_{dither} is superimposed on the diode laser bias current I_{bias} . (See Section 13.1 for more on AC versus DC.)

Example 10.6: FM Derivative Spectroscopy. By Taylor's theorem, if the tuning sensitivity is K and the total photon efficiency of the system (including sample absorption) is $A(\omega)$, the detected photocurrent is

$$I_{\text{photo}} = BPA \approx B \left(\sum_{j=0}^{\infty} \frac{(i_{\text{dither}})^j}{j!} \frac{d^j P}{di^j} \right) \left(\sum_{j=0}^{\infty} \frac{(i_{\text{dither}})^j}{j!} K^j \frac{d^j A}{d\omega^j} \right). \quad (10.7)$$

The double sum is approximate because the laser power and tuning are not uniquely defined functions of I_{bias} , but depend on temperature and history as well.

Detecting the second harmonic gives terms proportional to $d^2 A/d\omega^2$, $d^2 P/dI^2$ (the second derivative of laser power vs. bias current), and the cross term of their first derivatives. Since the laser power is reasonably linear in I_{bias} above threshold, its second derivative is small, so its additive noise is strongly suppressed. Differentiating sharp spectral features makes them grow, so the cross term is suppressed as well.

The lineshapes you get from this sort of measurement aren't what you'd measure in an ordinary spectrometer, which is a drawback. For measurements of the concentrations of known species, it works pretty well, and the optical apparatus is about as simple as it gets: a collimated diode laser, sample cell, and detector.

FM spectroscopy is an AC measurement done by modulating the beam, but it is very different from naïve chopping. In a chopping measurement, the beam switches from full on to full off regardless, but in the FM spectroscopy case there's almost no second derivative signal when there's no sample absorption peak.

10.9.4 Modulation Mixing

The FM spectroscopy example leads to another possibility, namely two-tone spectroscopy. Modulating the laser with two frequencies at once will cause mixing products that depend only on the cross terms in an expanded version of (10.7), and so are more nearly zero background. A high-frequency modulation and a lower-frequency one can do measurements far from the baseband noise and scan artifacts, with a nearly zero background characteristic. The downside of this is poorer shot-noise-limited SNR, because we're starting with two relatively weak sidebands rather than the main beam, and of course lots of spurious signals due to all the possible sidebands that can mix together.

The Doppler-free two-photon spectroscopy experiment of Example 1.1 is another example; the sum and difference frequencies come only from molecules that have absorbed one photon from each beam, and hence exhibit no Doppler broadening. Another one is the attractive mode AFM of Example 10.3, where the two modulations are the cantilever vibrations and the heterodyne interferometer's carrier frequency, and the resulting mixing appears as PM sidebands. Photoacoustic measurements with optical readout can be done this way too. There's more theoretical detail on modulation mixing in Section 13.10.6.

10.9.5 AC interference

One good way to make good fringes different from bad ones is to make them move in a rapid and predictable fashion, by using two beams of slightly different frequency. The frequency difference can be produced with an acousto-optic (Bragg) cell, a two-frequency laser such as a Zeeman-split HeNe, or by rapid frequency modulation of a diode laser, together with a path delay in one arm of the interferometer. Slower fringe motion can be obtained using mechanical means, such as a corner reflector spinning about its three-fold symmetry axis.

There is a trade-off of optical versus electronic complexity. Some people like to use multiple Bragg cells, in order to get a fixed IF directly from the photodiode. This is OK if you have a drawer full of Bragg cells and are just using it in the lab, but in an instrument it wastes photons, costs a lot, and isn't really needed by anybody who can build a radio. It's generally better to use more analog electronic signal processing and less optics; this preserves dynamic range while reducing complexity and cost.

10.9.6 Labeling Modulation Phase

If your measurement physics has an intrinsically slow response time (e.g., thermal effects), you can phase shift the signal with respect to the background by chopping the

excitation light. Choose a modulation frequency where the real signal is phase-shifted by $20\text{--}45^\circ$ from the chopped background (above 45° the amplitude falloff starts hurting your SNR). Null the background by tweaking the lock-in's phase. A nice stable setup will give you 40 dB (electrical) background rejection this way.

10.9.7 Labeling Arrival Time

There are other time-sensitive measurements besides OTDR. For example, you can make good range measurements by putting a linear current ramp on a diode laser in a Michelson interferometer and looking at the frequency of the output signal. If the tuning rate is $\dot{\omega}$ and the round-trip delay is Δt , the output frequency is $\dot{\omega} \Delta t$. You can do a good job with this, if you can make your diode stable enough.

10.9.8 Labeling Time Dependence

Pulsed lasers are rotten spectroscopic sources, because of their pulse-to-pulse variation in power, which changes irregularly over time. One way of dealing with it is the *cavity ring-down* approach, a kind of optical ju-jitsu; let the pulse have whatever energy it wants, and watch it decay as it rattles around in a resonator containing the sample. Regardless of what the pulse energy is, the attenuation per pass gives the sample absorption.

You use a laser whose pulse width is shorter than the cavity round-trip time, so that the pulses don't interfere with themselves, and look at the leakage out the end mirrors. A logarithmic detector (usually a DLVA) will produce an output whose slope is proportional to the fractional loss on each pass.

10.9.9 Labeling Wavelength

Spectroscopic measurements are of course labeled with wavelength, but as a background rejection device we need something like a two-wavelength colorimetric chemical measurement. There are a number of reagents whose color changes in the presence of some other chemical, for example, copper sulfate and ammonium hydroxide. Measuring the absorption of the mixture at two wavelengths can normalize out minor concentration errors and source brightness, leaving the ratio of the absorbances at the two wavelengths.

10.9.10 Labeling Coherence

Light whose coherence function is narrow (e.g., LED light or thermal light) can be used in interferometers for absolute distance measurements, because the fringe visibility goes through such a big peak near zero path difference. A sinusoidally modulated diode laser can do the same thing; its fringe visibility goes to 0 at the nulls of the Bessel function for its phase modulation (see Section 13.3). The modulation index m is the peak *relative* phase deviation in radians, which depends on the transit time as well as the FM spectrum itself.

10.9.11 Labeling Coincidence

If you're down in the mud scrabbling for photons, and the dark current of your photomultiplier is a serious problem, coincidence detection is a lifesaver. We usually think of it

in nuclear measurements, e.g. two-gamma transitions in nuclei, or scintillation counting. However, it is equally applicable to other measurements, such as surface scatter from a particle adhering to a silicon wafer. Multiple detectors looking at a large solid angle are one way to get all the photons, and if you run a coincidence or majority-rules scheme to weed out noise pulses, you can avoid the limitations of dark pulses in PMTs. (If you're using Geiger-mode APDs, remember they emit light when they avalanche—see Section 3.6.4.)

10.9.12 Labeling Position

Even in a complicated, multiple-scattering sort of sample, it is possible to restrict the measurement to a small region. While this may waste photons, those weren't going to help anyway. An example is an ordinary bright-field microscope with a translucent sample, where out-of-focus planes scatter light into the detector that makes the image fuzzy and ambiguous. Good baffles and spatial filters are a good way to do this, for example, the confocal microscope, where light has to make it back through a pinhole before reaching the detector; the light that would have blurred the image misses the pinhole.

10.9.13 Labeling Polarization

Sometimes we can manoeuvre the background signal into a polarization we can do without, and then get rid of it with an analyzer. If you have a sample that you can modulate, e.g. a magneto-optic film that responds strongly to an applied AC field, this allows the same sort of electronic zero background measurement as sitting on the lock-in's null.

Example 10.7: Patterned Thin Film Thickness Measurements. Let's say you have a dielectric stack with patterned metal at the bottom—it could be a silicon wafer or a magnetic recording head—and you're doing a broad-area thickness measurement based on interference between layers as a function of λ . The patterned metal doesn't obey the nice plane-parallel film assumptions we usually make in film thickness measurements, so the a priori approach won't work. If you care about the interference between the wafer surface and the top of the metal, but not about the thickness of the dielectric, you can use p polarization at Brewster's angle, which will pretty well get rid of it.

It's more interesting if you want the dielectric thickness: in that case, you might try flopping between s and p with a liquid crystal variable retarder. The fringes in the p -polarized light will contain mainly the metal-to-substrate fringes, and the s will have all three pairs. By knowing the metal–substrate distance, sorting out the dielectric thickness from the three sets of fringes isn't too hard.

10.10 CLOSURE

There is a broad class of measurements based on *closure*, first used as a way of correcting for atmospheric phase shifts in very long baseline interferometry (VLBI) astronomical measurements. The basic idea is that if you have N sources, you can get $N(N - 1)/2$ independent pairwise comparisons, which gives an overdetermined system when $N > 3$. Overdetermined systems can be solved with least squares by using singular value decomposition (see Section 17.9).

If you need to measure the short-term stability of a laser, the classical method is to build two sources, lock them loosely together, and look at the beat note between them. That gives you a signal you can measure on a spectrum analyzer, and it's reasonable to expect their short-term instabilities to be uncorrelated, provided you account for microphonics, temperature gradients, power supply variations, and stray pickup. If you're more concerned with drift, temperature stability, and so forth, you can lock one laser to an atomic transition (e.g., an iodine line) and beat it against your unknown.

The closure approach would be to build three lasers and look at the pairwise correlations. From measurements of pairwise phase deviations $\Delta\phi_{12}(t)$, $\Delta\phi_{23}(t)$, and $\Delta\phi_{31}(t)$, you can solve for $\Delta\phi_1 - \Delta\phi_3$, the individual deviations from the mean phase. (You can't get the mean phase itself—you're never going to be able to measure the carrier frequency from the relative phase fluctuations, for instance.) The nice thing about this is that you can assign an individual phase deviation to each source. In the laser case, that means you don't have to assume that the lasers have the same noise, as you do in the two-channel measurement. Similarly, if the measurements themselves have noise (which they will), adding more sources and using SVD will improve the estimates further. The two-channel correlation noise measurement of Section 17.11.6 uses a similar idea to get $N(N-1)/2$ separate measurements from N voltmeters, allowing noise measurements $\sim 20 \log N - 6$ dB below the noise floor; with closure, you can measure signal as well as noise.