BANDWIDTH-NARROWING TECHNIQUES

15.12 The problem of signal-to-noise ratio

Up to this point we have been talking about the various experimental quantities that can be detected, how you might measure them, and what sort of trade-offs you face. As luck would have it, the signals you often want to measure are buried in noise or interference, frequently to the extent that you can't even see them on an oscilloscope. Even when external noise isn't a problem, the statistics of the signal itself may make detection difficult, as, for example, when counting nuclear disintegrations from a weak source, with only a few counts detected per minute. Finally, even when the signal is detectable, you may wish to improve the detected signal strength in order to make a more accurate measurement. In all these cases some tricks are needed to improve the signal/noise ratio; as you will see, they all amount to a narrowing of the detection bandwidth in order to preserve the desired signal while reducing the total amount of (broadband) noise accepted.

The first thing you might be tempted to try when thinking of reducing the bandwidth of a measurement is to hang a simple low-pass filter on the output, in order to average out the noise. There are cases where that therapy will work, but most of the time it will do very little good, for a couple of reasons. First, the signal itself may have some high frequencies in it, or it may be centered at some high frequency. Second, even if the signal is in fact slowly varying or static, you invariably have to contend with the reality that the density of noise signal usually has a $1/f$ character, so as you squeeze the bandwidth down toward dc you gain very little. Electronic and physical systems are twitchy, so to speak.

In practice, there are a few basic techniques of bandwidth narrowing that are in widespread use. They go under names like signal averaging, transient averaging, boxcar integration, multichannel scaling, pulse-height analysis, lock-in detection, and phase-sensitive detection. All these methods assume that you have a repetitive signal; that's no real problem, since there is almost always a way to force the signal to be periodic, assuming it isn't already. Let's see what is going on.

15.13 Signal averaging and multichannel averaging

By forming a cumulative sum of a repetitive signal versus time, you can improve the signal/noise ratio enormously. This usually goes under the heading of "signal averaging," and it is often applied to analog signals. We will consider first what may seem to be an artificial situation, namely a signal consisting of pulses whose rate is proportional to the amplitude of some sought-after waveform versus time. We begin with this example because it makes our calculations easier. In reality, it isn't even an artificial situation, since it is the rule when using pulse-counting electronics such as particle detectors or photomultipliers at low light levels.

Multichannel scalers

We begin with multichannel scaling because it typifies all these techniques and, in addition, is easy to understand and quantify. The multichannel scaler (MCS) is a piece of hardware that contains a set of memory registers (typically 1024 or more), each of which can store a number up to 1 million (20 bits binary or 24 bits BCD) or so. The **MCS** accepts pulses (or continuous voltages, as will be described later) as its input; in addition, it accepts either a channel-advance signal (a pulse) or a parallel multibit channel address. Each time there is an input pulse, the **MCS** increments the count in the memory channel currently being addressed. Additional inputs let you reset the address to 0, clear the memory, etc.

To use an **MCS** you need a signal that repeats itself at some interval. Let's suppose for the time being that the phenomenon you're observing is itself periodic, with period T; although this is not the case most of the time (you usually have to coax the experiment into periodicity), there are good examples in the real world of strictly periodic phenomena, e.g., the light output of a pulsar. Let's suppose that the input consists of pulses, with rate proportional to the signal plus a large background rate of noise pulses, i.e., pulses randomly distributed in time (again, realistic for pulsars, where the actual signal is swamped by light from the night sky). By sending timing pulses to the channel advance and reset inputs, we arrange to sweep the **MCS** repetitively through its 1024 channels once every T seconds, accumulating additional input (signal plus background) counts into the memory channels each sweep. As time goes on, the signal will keep adding counts to the same subgroup of channels, with the background noise adding counts in all channels, because the sweep through the entire set of channels is timed to coincide with the signal's periodicity. Thus the signal keeps adding on top of itself, the accumulated sum getting larger after each repetition.

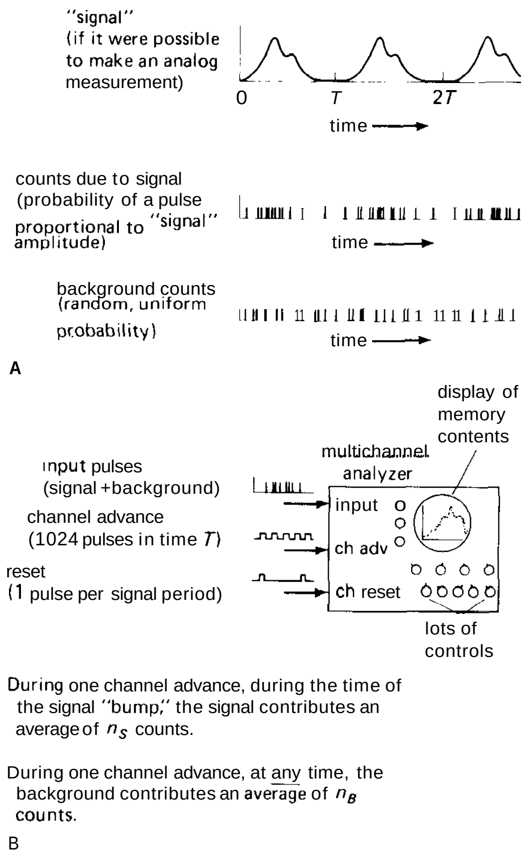

Figure 15.34. Multichannel signal averaging (pulse input).

Signal-to-noise computation

Let's see what happens. To be specific, let the background pulse rate have an average value that contributes n_b pulses per channel each sweep, with the signal contributing an additional n_s pulses into the channel where its peak lies (Fig. 15.34). Let's give ourselves a poor signal/background ratio, i.e., n, $\ll n_b$, meaning that most of the counts added during each sweep through the memory are contributed by background, rather than signal. Now, when the memory contents are graphed, the signal should be recognizable as a bump above the background. You might think the criterion is that the number of signal counts in a channel with signal should be comparable with the number of counts contributed to that channel by the background noise. That would be wrong, since the ***average*** value contributed by noise is quite irrelevant; all that matters is the level of fluctuations of that average value about the mean.

Thus, a poor input signal/noise ratio is actually characterized by $n_s \ll \sqrt{n_b}$, meaning that in one sweep the signal will not be recognizable above the "noise" consisting of an undulating graph of accumulated random background pulses. For purposes of computation, let's let $n_s = 10$ and $n_b = 1000$. Therefore, in one sweep an initially cleared MCS will acquire an average of 1000 counts in each channel, with an additional 10 counts in the channels where the signal peaked. Since the fluctuations in the channel totals equal about 31 (square root of 1000), the actual signal bump is left pretty much buried in the noise after only one sweep. But after 1000 sweeps, say, the average count in any channel is about 1,000,000, with fluctuations of 1000. The channels where the signal peaks have an additional 10,000 counts (1000 sweeps x 10 counts/sweep), for a signal/noise ratio of 10. In other words, the signal has emerged from the background.

Example: Mossbauer resonance

Figure 15.35 shows the results of just such an analysis, in this case a Mossbauer resonance signal consisting of six dips in the transmission of an enriched iron-57 foil to gamma radiation from a cobalt-57 radioactive source. In this case $n_b = 0.4$ and $n_s = 0.1$, approximately, for a situation of poor signal/noise ratio. The Mossbauer signal is totally swamped by noise even after 10 or 100 sweeps; it becomes visible only after 1000 sweeps or so. The results are shown after 1000, 10,000, and 100,000 sweeps, with each graph scaled to keep this signal size the same. Note the rise of the "baseline" caused by the steady background, as well as the nice enhancement of SNR with time.

It is easy to see by what factor the ratio of signal amplitude to background fluctuation ("noise") increases as time goes on. The signal amplitude increases proportional to t; the average background count ("baseline") also increases proportional

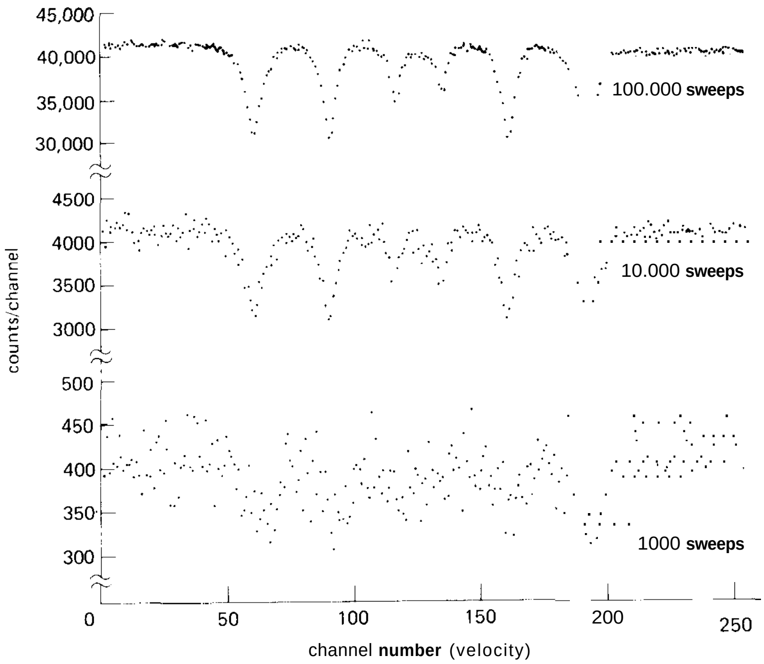

Figure 15.35. Mossbauer absorption spectrum, showing effect of signal averaging.

to t, but thefluctuations in the background count ("noise") rise only proportional to the square root of t. Therefore, the ratio between signal and fluctuations in background increases as t divided by the square root oft. In other words, the signal-to-noise ratio improves in proportion to the square root of time.

Multichannel analysis of analog signals (signal averaging)

You can play the same game with analog signals by simply using a voltage-to-frequency converter at the input. Commercial MCSs often provide the electronics for you, giving you a choice of analog or pulse input modes. In this form you often hear these gadgets called signal averagers or transient averagers. One company (TMC) called theirs a "CAT" (computer of averaged transients), and the name has stuck, in some circles at least.

It is possible to make a completely analog MCS by using a set of integrators to store the accumulated signal. A simpler device, known as a boxcar integrator, is an analog signal averager with a single "sliding channel." With the enormous reductions in digital memory prices that have taken place in the last decade, such analog signal averagers are becoming impractical, except perhaps for specialized applications.

Multichannel analysis as bandwidth narrowing

We suggested at the beginning of this discussion that there was an equivalence between the magical SNR-reduction methods and a reduction in effective measurement bandwidth. It is not hard to see how that goes in this case. Imagine another (interfering) signal added into the input, but with periodicity T' slightly different from the desired signal of period T. After just a few sweeps, its signal will also begin to accumulate, causing trouble. But wait – as time goes on, its "bump" will gradually drift along through the channels, successively contributing counts through all the channels. It will have drifted all the way around through all the channels once after a time.

$$t = 1/\Delta f$$

where Af is the frequency difference $1/T - 1/T'$ between the desired signal and the interfering signal.

EXERCISE 15.1
Derive this result.

In other words, by accumulating data for a time t (as given in the preceding equation), the interfering signal has been spread equally through all the channels. Another way to say the same thing is that the measurement's bandwidth is reduced roughly to

$$\mathbf{A}f = 1/t$$

after accumulating data for time t. By running for a long time, you reduce the bandwidth and exclude nearby interfering signals! In fact, you also exclude most of the noise, since it is spread evenly in frequency. Viewed in this light, the effect of multichannel analysis is to narrow the accepted bandwidth, thereby accepting the signal power but squeezing down the amount of noise power.

Let's see how the calculation goes. After time t, the bandwidth is narrowed to $\mathbf{A}f$ =llt. If the noise power density is p_n watts per hertz, and the signal power P_s stays within the measurement bandwidth, then the SNR after time t is

$$\text{SNR} = 10\log(P_s t/p_n)$$

The signal amplitude improves proportionally to the square root of t (3dB for each doubling of t), just as we found in the analysis we did earlier by considering the number of counts per channel and its fluctuations.

15.14 Making a signal periodic

We mentioned initially that all signal-averaging schemes require a signal that repeats many times in order to realize significant reduction in signal/noise ratio. Since most measurements don't involve intrinsically periodic quantities, it is usually necessary to force the signal to repeat. There are many ways to do this, depending on the particular measurement. It is probably easiest to give a few examples, rather than attempt to set down rules.

A measurable quantity that depends on some external parameter can easily be made periodic – just vary the external parameter. In NMR (nuclear magnetic resonance) the resonance frequency varies linearly with the applied field, so it is standard to modulate the current in a small additional magnet winding. In Mossbauer studies you vary the source velocity. In quadrupole resonance you can sweep the oscillator.

In other cases an effect may have its own well-defined transient, but allow external triggering. A classic example is the pulse of depolarization in a nerve fiber. In order to generate a clean graph of the waveform of such a pulse, you can simply trigger the nerve with an externally applied voltage pulse, starting the MCS sweep at the same time (or even "anticipating" the trigger by starting the sweep, then triggering the nerve with a delayed pulse); in this case you would pick a repetition period long enough so that the nerve has fully recovered before the next pulse. This last case illustrates graphically the importance of a repeatable phenomenon as fodder for signal averaging; if the frog whose leg is twitching chances to expire, your experiment is over, whatever the signal/noise ratio!

It should be pointed out that cases where the phenomenon you're measuring has its own well-defined periodicity may in fact be the most difficult to work with, since you have to know the periodicity precisely. The graph of the "light curve" (brightness versus time) in Figure 15.36 is an example. We made this curve by using an MCS on the output of a photomultiplier stationed at the focus of a 60 inch telescope, run

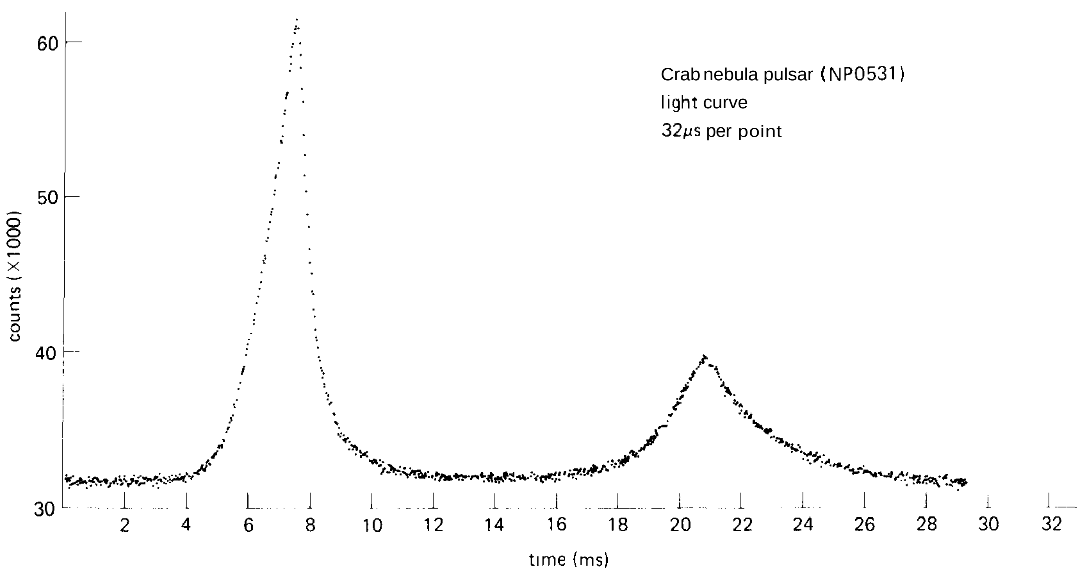

Figure 15.36. Crab nebula pulsar brightness versus time (light curve).

exactly in synchronism with the pulsar's rotation. Even with that size telescope it required an average of approximately 5 million sweeps to generate such a clean curve, since the average number of detected photons for each entire pulsar pulse was about 1. With such a short period, that puts enormous accuracy requirements on the MCS channel-advance circuitry, in this case requiring clocks of part-per-billion stability and frequent adjustment of the clock rate to compensate for the earth's motion.

It is worth saying again that the essence of signal averaging is a reduction in bandwidth, gained by running an experiment for a long period of time. The bottom line here is the total length of the experiment; the particular rate of scanning, or modulation, is usually not important, as long as it takes you far enough from the $1/f$ noise present near dc. You can think of the modulation as simply shifting the signal you wish to measure from dc up to the modulating frequency. The effect of the long data accumulation is then to center an effective bandwidth $\Delta f = 1/T$ at f_{mod}, rather than at dc.

15.15 Lock-in detection

This is a method of considerable subtlety. In order to understand the method, it is necessary to take a short detour into the phase detector, a subject we first took up in Section 9.27.

Phase detectors

In Section 9.27 we described phase detectors that produce an output voltage proportional to the phase difference between two digital (logic-level) signals. For purposes of lock-in detection, you need to know about linear phase detectors, since you are nearly always dealing with analog voltage levels.

The basic circuit is shown in Figure 15.37. An analog signal passes through a linear amplifier whose gain is reversed by a square-wave "reference" signal controlling a FET switch. The output signal passes through a low-pass filter, RC. That's all there is to it. Let's see what you can do with it.

□ ***Phase-detector output.*** To analyze the phase-detector operation, let's assume we apply a signal

$$E_s \cos(\omega t + \phi)$$

to such a phase detector, whose reference signal is a square wave with transitions at the zeros of sinwt, i.e., at $t = 0, \pi/\omega, 2\pi/\omega$, etc. Let us further assume that we' average the output, V_{out}, by passing it through a low-pass filter whose time constant is longer than one period:

$$\tau = RC \gg T = 2\pi/\omega$$

Then the low-pass filter output is

$$\langle E_s \cos(\omega t + \phi)\rangle\big|_0^{\pi/\omega} - \langle E_s \cos(\omega t + \phi)\rangle\big|_{\pi/\omega}^{2\pi/\omega}$$

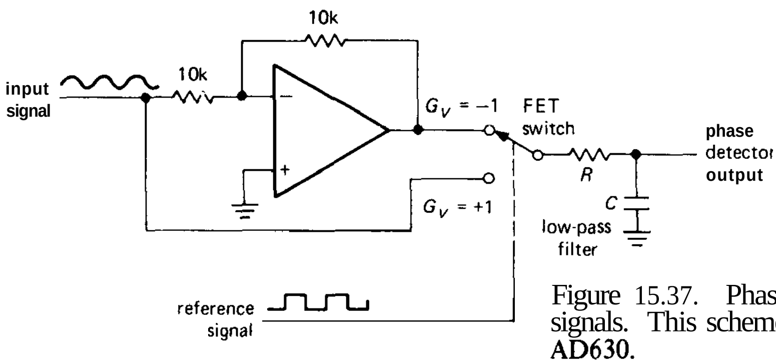

Figure 15.37. Phase detector for linear input signals. This scheme is used in the monolithic AD630.

where the brackets represent averages, and the minus sign comes from the gain reversal over alternate half cycles of V_{ref}. As an exercise, you can show that

$$\langle V_{\text{out}} \rangle = -(2E_s/\pi)\sin\phi$$

EXERCISE 15.2

Perform the indicated averages by explicit integration to obtain the preceding result for unity gain.

Our result shows that the averaged output, for *an input signal of the same frequency as the reference signal,* is proportional to the amplitude of V_s and sinusoidal in the relative phase.

We need one more result before going on: What is the output voltage for an input signal whose frequency is close to (but not equal to) the reference signal? This is easy, since in the preceding equations the quantity ϕ now varies slowly, at the difference frequency:

$$\cos(\omega + \Delta\omega)t = \cos(\omega t + \phi)$$
$$\text{with}\, \phi = t\Delta\omega$$

giving an output signal that is a slow sinusoid:

$$V_{\text{out}} = (2E_s/\pi)\sin(\Delta\omega)t$$

which will pass through the low-pass filter relatively unscathed if Aw $< 1/\tau = 1/RC$ and will be heavily attenuated if $\Delta\omega > 1/\tau$.

The lock-in method

Now the so-called lock-in (or phase-sensitive) amplifier should make sense. First you make a weak signal periodic, as we've discussed, typically at a frequency in the neighborhood of 100Hz. The weak signal, contaminated by noise, is amplified and phase-detected relative to the modulating signal. Look at Figure 15.38. You need an experiment with two "knobs" on it, one for fast modulation in order to do phase detection and one for a slow sweep through the interesting features of the signal (in NMR, for example, the fast modulation might be a small 100Hz modulation of the magnetic field, and the slow modulation might be a frequency sweep 10 minutes in duration through the resonance). The phase shifter is adjusted to give maximum output signal, and the low-pass filter is set for a time constant long enough to give good signal/noise ratio. The low-pass-filter rolloff sets the bandwidth, so a 1Hz rolloff, for example, gives you sensitivity to spurious signals and noise only within 1Hz of the desired signal. The bandwidth also determines how fast you can adjust the "slow modulation," since now you must not sweep through any features of the signal faster than the filter can respond. People use time constants of fractions of a second up to tens of seconds and often do the slow modulation with a geared-down

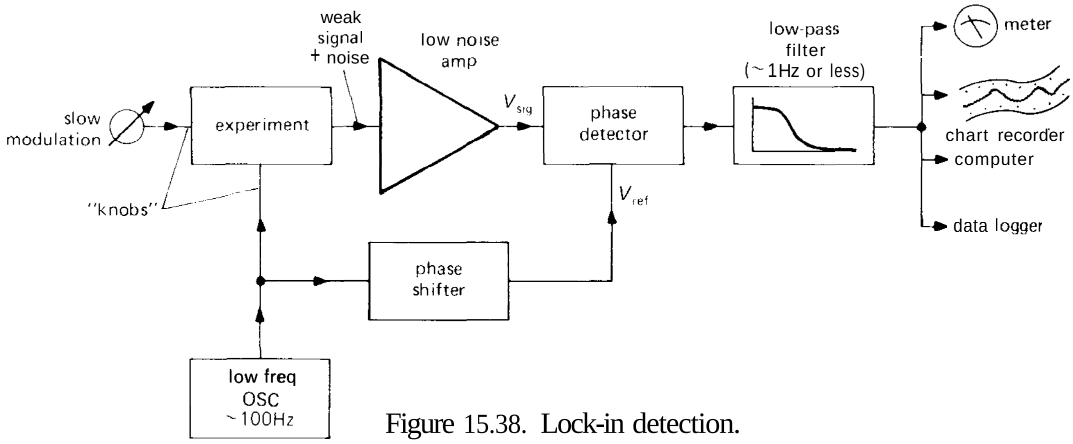

Figure 15.38. Lock-in detection.

clock motor turning an actual knob on something!

Note that lock-in detection amounts to bandwidth narrowing again, with the bandwidth set by the post-detection low-pass filter. As with signal averaging, the effect of the modulation is to center the signal at the fast modulation frequency, rather than at dc, in order to get away from $1/f$ noise (flicker noise, drifts, and the like).

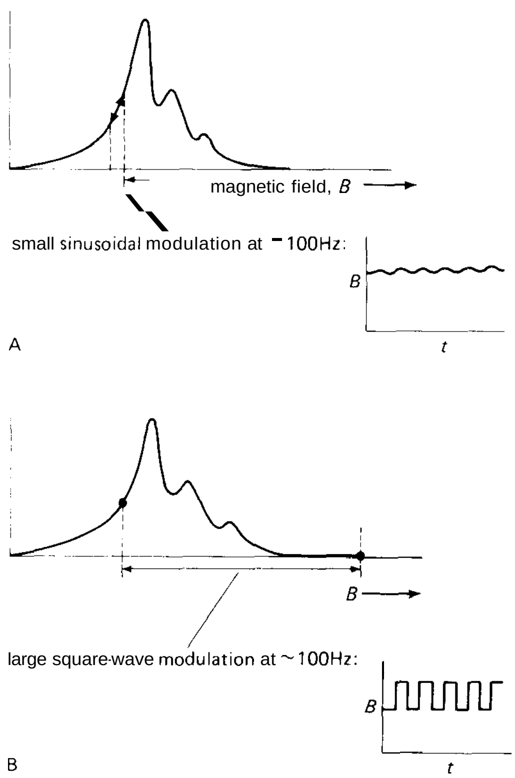

Figure 15.39. Lock-in modulation methods.
A. Small sinusoid.
B. Large square wave.

Two methods of "fast modulation"

There are some ways to do the fast modulation: The modulation waveform can be either a very small sine wave or a very large square wave compared with the features of the sought-after signal (line shape versus magnetic field, for example, in NMR), as sketched in Figure 15.39. In the first case the output signal from the phase-sensitive detector is proportional to the slope of the line shape (**i.e.**, its derivative), whereas in the second case it is proportional to the line shape itself (providing there aren't any other lines out at the other endpoint of the modulation waveform). This is the reason all those simple NMR resonance lines come out looking like dispersion curves (Fig. 15.40).

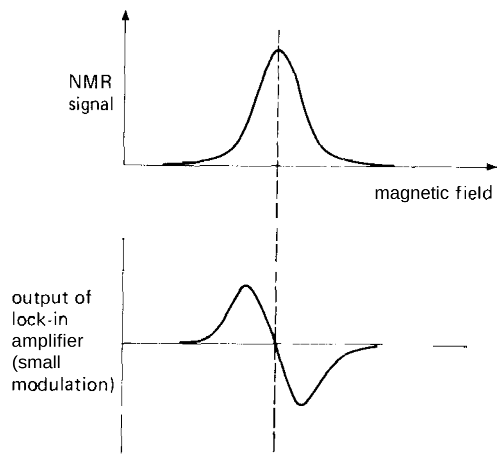

Figure 15.40. Line shape differentiation resulting from lock-in detection.

For large-shift square-wave modulation there's a clever method for suppressing modulation feedthrough, in cases where that is a problem. Figure 15.41 shows the modulation waveform. The offsets above and below the central value kill the signal, causing an **on/off** modulation of the signal at ***twice*** the fundamental of the modulating waveform. This is a method for use in special cases only; don't get carried away by the beauty of it all!

Large-amplitude square-wave modulation is a favorite with those dealing in infrared astronomy, where the telescope secondary mirrors are rocked to switch the image back and forth on an infrared source.

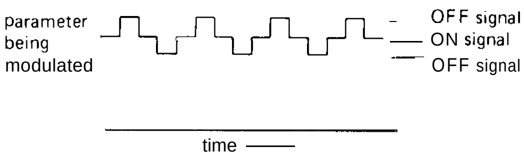

Figure 15.41. Modulation scheme for suppressing modulation feedthrough.

It is also popular in radioastronomy, where it's called a Dicke switch.

Commercial lock-in amplifiers have a variable-frequency modulating source and tracking filter, a switchable time-constant post-detection filter, a good low-noise wide-dynamic-range amplifier (you wouldn't be using lock-in detection if you weren't having noise problems), and a nice linear phase detector. They also let you use an external source of modulation. There's a knob that adjusts the phase shift, so you can maximize the detected signal. The whole item comes packed in a handsome cabinet, with a meter to read output signal. Typically these things cost a few thousand dollars and are manufactured by companies like EG&G Princeton Applied Research, Ithaco, and Stanford Research Systems. Board-level components are made by Evans Electronics, among others.

In order to illustrate the power of lock-in detection, we usually set up a small demonstration for our students. We use a lock-in to modulate a small LED of the kind used for panel indicators, with a modulation rate of a kilohertz or so. The current is very low, and you can hardly see the LED glowing in normal room light. Six feet away a phototransistor looks in the general direction of the LED, with its output fed to the lock-in. With the room lights out, there's a tiny signal from the phototransistor at the modulating frequency (mixed with plenty of noise), and the lock-in easily detects it, using a time constant of a few seconds. Then we turn the room lights on (fluorescent), at which point the signal from the phototransistor becomes just a huge messy 120Hz waveform, jumping in amplitude by 50dB or more. The situation looks hopeless on the oscilloscope, but the lock-in just sits there, unperturbed, calmly detecting the same LED signal at the same level. You can check that it's really working by sticking your hand in between the LED and the detector. It's darned impressive.

15.16 Pulse-height analysis

A pulse-height analyzer (PHA) is a simple extension of the multichannel scaler principle, and it is a very important instrument in nuclear and radiation physics. The idea is simplicity itself: Pulses with a range of amplitudes are input to a peak-detector/ADC circuit that converts the relative pulse height to a channel address. A multichannel scaler then increments the contents of the selected address. The result is a graph that is a histogram of pulse heights. That's all there is to it.

The enormous utility of pulse-height analyzers stems from the fact that many detectors of charged particles, x rays, and gamma rays have output pulse sizes proportional to the energy of the radiation detected (e.g., proportional counters, solid-state detectors, surface-barrier detectors, and scintillators, as we discussed in Section 15.07). Thus a pulse-height analyzer converts the detector's output to an energy spectrum.

Pulse-height analyzers used to be designed as dedicated hardware devices, with buckets of ICs and discrete components. Nowadays the standard method is to use an off-the-shelf microcomputer, preceded by a fast pulse-input ADC. That way you can build in all sorts of useful computational routines, e.g., background subtraction, energy calibration and line identification, disk and tape storage, and on-line control of the experiment. We have an apparatus that scans a proton microbeam over a specimen in a two-dimensional raster pattern, detects the emitted x rays, sorts them by chemical element, and stores a picture of the distribution of each element in the sample, all the while letting you view the x-ray spectrum and images as the picture accumulates. The whole operation is handled by a pulse-height analyzer that doesn't realize that it's really a computer.

There is an interesting subtlety involving the ADC front end of a pulse-height analyzer. It turns out that you can't use something like a successive-approximation A/D converter, in spite of its superior speed, because you wouldn't get exact equality of channel widths, with the disastrous effect of producing a lumpy baseline from a smooth continuum of input radiation. All PHAs use a so-called Wilkinson converter, a variation on single-slope conversion whereby an input pulse charges a capacitor, which is then discharged by a constant current while a fast counter (200MHz is typical) counts up the address. This has the disadvantage of giving an analyzer "dead time" that depends on the height of the last pulse, but it gives absolute equality of channel widths.

Most pulse-height analyzers provide inputs so that you can use them as multichannel scalers. Why shouldn't they? All the electronics are already there. Some big names in pulse-height analyzers are Canberra, EG&G, Nuclear Data, and Tracor-Northern.

15.17 Time-to-amplitude converters

In nuclear physics it is often important to know the distribution of decay times of some short-lived particle. This turns out to be easy to measure, by simply hooking a time-to-amplitude converter (TAC) in front of a pulse-height analyzer. The TAC starts a ramp when it receives a pulse at one input and stops it when it receives a pulse at a second input, discharging the ramp and generating an output pulse proportional to the time interval between pulses. It is possible to build these things with resolution down in the picoseconds. Figure 15.42 shows a measurement of the muon lifetime made by a student by timing the delay between the capture of a cosmic-ray muon in a scintillator and its subsequent decay. Each event creates a flash of light, and a TAC is used to convert the intervals into pulses. A cosmic-ray muon decayed in this student's apparatus once a minute on the average, so he accumulated data for 18 days to determine a lifetime of 2.198 ± 0.02μs (accepted value is 2.197134 ± 0.00008μs). Note the use of log-lin axes to plot data that should be an exponential, and the systematic shift of $n^{1/2}$ (counting) error bars. The line plotted is the decay according to the accepted value, $n(t) = n_0 \exp(-t/\tau)$.

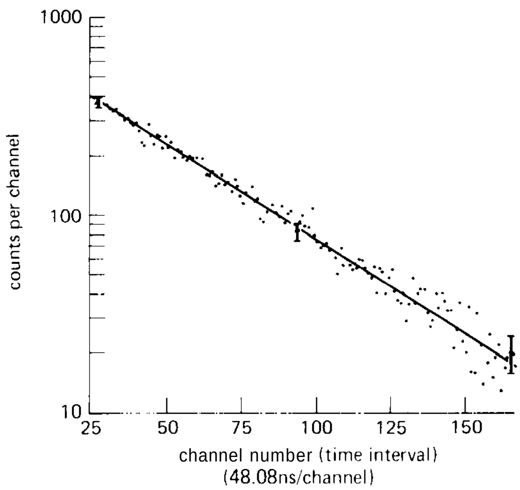

Figure 15.42. Muon lifetime measurement from time-interval spectrum (TAC + PHA).

SPECTRUM ANALYSIS AND FOURIER TRANSFORMS

15.18 Spectrum analyzers

An instrument of considerable utility, particularly in radiofrequency work, is the spectrum analyzer. These devices generate an xy oscilloscope display, with y representing signal strength (usually logarithmic, i.e., in decibels), but with x representing frequency. In other words, a spectrum analyzer lets you look in the *frequency domain,* plotting the amount of input signal versus its frequency. You can think of it as a Fourier decomposition of the input waveform (if you know about such things), or as the response you would get as you tuned the dial of a broadrange high-performance (wide dynamic range, stable, sensitive) receiver through its frequency range. This ability can be very handy when analyzing modulated signals, looking for intermodulation products or distortion, analyzing noise and drift, trying to make accurate frequency measurements on weak signals in the presence of stronger signals, and making a host of other measurements.

Spectrum analyzers come in two basic varieties: swept-tuned and real-time. Swept analyzers are the most common variety, and they work as shown in Figure 15.43. What you have is basically a superheterodyne receiver (see Section 13.16), with a local oscillator (LO) that can be swept by an internally generated ramp waveform. As the LO is swept through its range of frequencies, different input frequencies are successively mixed to pass through the IF amplifier and filter. For example, suppose you have a spectrum analyzer with an IF of 200MHz and an LO that can sweep from 200MHz to 300MHz. When the LO is at 210MHz, input signals at 10MHz (± the IF filter bandwidth) pass through to the detector and produce vertical deflection on the scope. Signals at 410MHz (an "image" frequency) would also pass through, which is the reason for the low-pass filter at the input. At any given time, input frequencies 200MHz lower than the LO are detected.

Real spectrum analyzers allow lots of flexibility as to sweep range, center frequency, filter bandwidth, display scales, etc. Typical input frequency ranges go from hertz to gigahertz, with selectable bandwidths ranging from hertz to megahertz. A range of 10MHz to 22GHz is popular, with resolution bandwidths of 10Hz to 3MHz. In addition, sophisticated spectrum analyzers have convenience features such as absolute amplitude calibration, storage of spectra to prevent flicker during sweeping, additional storage for comparison and normalization, and display of digital information on the screen. Fancy spectrum analyzers let you analyze phase versus frequency, generate frequency markers, program the operation via the IEEE-488 bus, include tracking oscillators (for increased dynamic range), make precise frequency measurements of features in the spectrum, generate tracking noise voltages for system stimulus, and even do signal averaging (particularly useful for noisy signals).

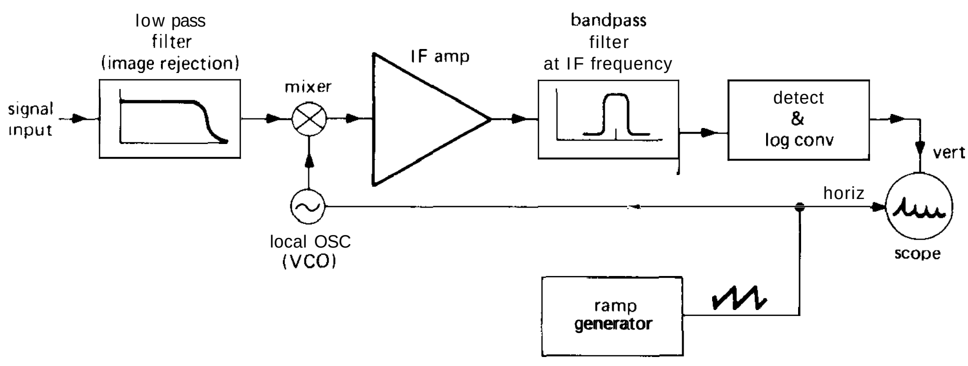

Figure 15.43. Swept-LO spectrum analyzer.

Note that this sort of swept spectrum analyzer looks at only one frequency at a time and generates a complete spectrum by sweeping in time. This can be a real disadvantage, since you can't look at transient events. In addition, when scanning with narrow bandwidth, the sweep rate must be kept slow. Finally, only a small portion of the input signal is being used at any one time.

These disadvantages of swept spectrum analysis are remedied in real-time spectrum analyzers. Again, there are several approaches. The clumsy method employs a set of narrow filters to look at a range of frequencies simultaneously. More recently, sophisticated analyzers based on digital Fourier analysis (in particular, the famous Cooley-Tukey fast Fourier transform, FFT for short) are becoming popular. These instruments convert the analog input signal (after mixing, etc.) to numbers, using a fast analog/digital converter. Then a special-purpose computer turns the crank on the FFT, generating a digital frequency spectrum. Since this method looks at all frequencies simultaneously, it has excellent sensitivity and speed, and it can be used for analysis of transients. It is particularly good for low-frequency signals, where swept analyzers are too slow. In addition, it can perform correlations between signals. Since the data comes out in digital form, it is natural to apply the full power of signal averaging, a feature available in some commercial instruments.

Note that these digital spectrum analyzers, being limited by computational speed, have ***much*** less bandwidth than the radiofrequency analog types (swept-LO or filter-bank). For example, the popular HP 3561A goes from 125μHz to 100kHz. You can, of course, use it to look at a 100kHz band centered at some higher frequency, by translating that band down in frequency with heterodyne techniques.

A clever real-time spectrum analyzer can also be constructed using the so-called chirp/Z transform. In this method a dispersive filter (delay time proportional to frequency) replaces the IF bandpass filter in the swept-LO analyzer (Fig. 15.43). By matching the LO sweep rate to the filter's dispersion, you get an output that superficially resembles the swept analyzer output, namely a linear scan of frequency versus time during each sweep. However, in contrast to the swept-LO analyzer, this scheme gathers signals from the entire band of frequencies continuously. Another interesting technique for real-time spectral analysis is the Bragg cell (or "acousto-optic spectrometer"), in which the IF signal is used to generate acoustic waves in a transparent crystal. These deformations diffract a laser beam, generating a real-time display of the frequency spectrum as light intensity versus position. An array of photodetectors completes the analyzer output. Bragg-cell spectrometers are used in radio astronomy. A typical unit has 2GHz instantaneous bandwidth, analyzed into 16,000 channels of 125kHz bandwidth each. When choosing a spectrum analyzer type, be sure to consider trade-offs among bandwidth, resolution, linearity, and dynamic range.

Figure 15.44 shows the sort of radiofrequency spectra that endear spectrum analyzers to people who earn their living above 1MHz. The first four spectra show oscillators: A is just a pure sine-wave oscillator, B is distorted (as indicated by its harmonics), *C* has noise sidebands, and D has some frequency instability (drifting or residual FM). You can measure amplifier intermodulation products, as in E, where second-, third-, and fourth-order intermodulation frequencies are visible in the output of an amplifier driven by a "two-tone" test signal consisting of pure sine waves at frequencies f_1 and f_2. Finally, in F you can see the uncouth behavior of a double-balanced mixer; there is feedthrough of both the LO and input signal, as well as distortion terms ($f_{LO} \pm 2f_{sig}, f_{LO} \pm 3f_{sig}$). This last spectrum may actually indicate quite respectable mixer performance, depending on the vertical scale shown. Spectrum analyzers are designed with enormous dynamic range (internally generated distortion products are typically down by 70dB or more; with a "tracking preselector" they're down by 100dB) so that you can see the failings of even a very good circuit.

The last graph G in Figure 15.44 shows what happens when you sweep the LO too fast in a swept analyzer. If the sweep causes a signal to pass through the filter bandwidth Δf in a time shorter than $\Delta t \approx 1/\Delta f$, it will be broadened, roughly to $\Delta f' \approx 1/\Delta t$.

15.19 Off-line spectrum analysis

The fast Fourier transform applied to digitized data from an experiment provides a very powerful method of signal analysis, particularly the recognition of weak signals of well-defined periodicity buried in interfering signals or noise, or the recognition of vibrations or oscillatory modes. For instance, we have used the FFT to search for pulsars, perform audio analysis, enhance the resolution of astronomical images (speckle imaging), and look for signals from intelligent life in space (SETI). In the last experiment, a GaAs FET amplifier connected to a receiving dish 84 feet in diameter drives a heterodyne receiver, with 400kHz of bandwidth analyzed (in real time) into 8 million simultaneous 0.05Hz channels. Our digital spectrum analyzer has 20,000 ICs and a half million solder joints (all done by hand!) and can detect narrowband signals 60dB below receiver noise in a 20 second integration. This corresponds to a radio flux of less than 1 nanowatt total over the entire earth's disk!

SELF-EXPLANATORY CIRCUITS

15.20 Circuit ideas

In Figure 15.45 we've collected some circuits that are useful in measurement and control applications.

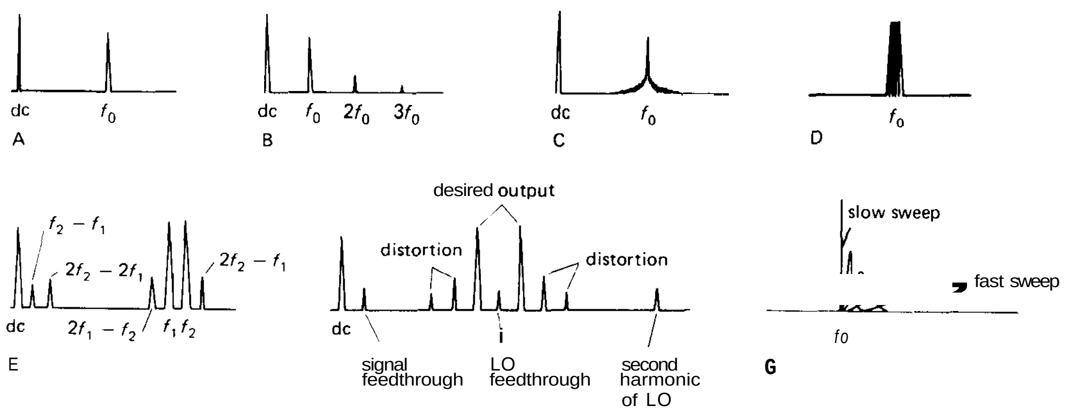

Figure 15.44. Spectrum analyzer displays.