CHAPTER 6

Limitations of conventional lock-in systems

6.1 Introduction

This short chapter provides an interlude where we can review the main characteristics of conventional lock-in amplifiers before going on to consider more complex system configurations. Of particular interest are the aspects of performance which limit the effectiveness of conventional systems in different applications. If a list of these performance limitations were to be drawn up it would almost certainly include the following items:

(1) Dependence on the availability of a synchronous reference voltage

(ii) Trade-off between key specifications; for example, dynamic-range/linearity, dynamic-reserve/output-stability

(iii) Harmonic responses of the phase-sensitive detector appearing in the overall response of a lock-in amplifier

(iv) Slew-rate limitations in the reference channel

This list is not intended to be comprehensive, nor is it meant to imply that items (i) to (iv) are identified as shortcomings in every application. What we can say is that shortcomings in these areas have prevented conventional lock-in amplifiers from being adopted as general-purpose measurement tools and from performing tasks beyond their traditional role in signal recovery.

The absolute dependence of synchronous detection systems on a reference voltage would appear to be a prime candidate in this respect, yet, in practice, most research workers seem able to devise experiments where a 'local' reference is made available. There are, nevertheless, several examples of experiments where the signal source is remote from the detection system and a local reference must be generated by phase-locking an oscillator to the incoming signal.

An example in this category, the reception of satellite 'beacon' signals, is described in the next chapter. It should not be thought, however, that all phase-lock applications are of this 'remote' type. For example, in Fourier-transform photometry the traditional light 'chopper' is often replaced by a rotating optical grating. This may be so fine-ruled that the usual method of generating a reference by means of an auxiliary light source and phototransistor is impossible to apply. In this case a reference can be generated by phase-locking to the signal in the output of the experiment using the general arrangement illustrated in Fig. 6.1.

The problem of phase-locking to noisy signals merits a fairly extensive discussion in the next chapter where we shall see how phase-locked systems can be bust up using standard lock-in amplifiers and off-the-shelf modules.

Regarding the second item, it must be expected that some sort of trade-off will be encountered when any electronic instrument is operated at its performance limits. The trade-offs referred to in (ii) are inevitable when a conventional lock-in amplifier is operated with a broadband signal channel, but we have seen that the

trade-offs can be improved when filters are used to eliminate unwanted components before detection. For example, the increase in signal-to-noise ratio

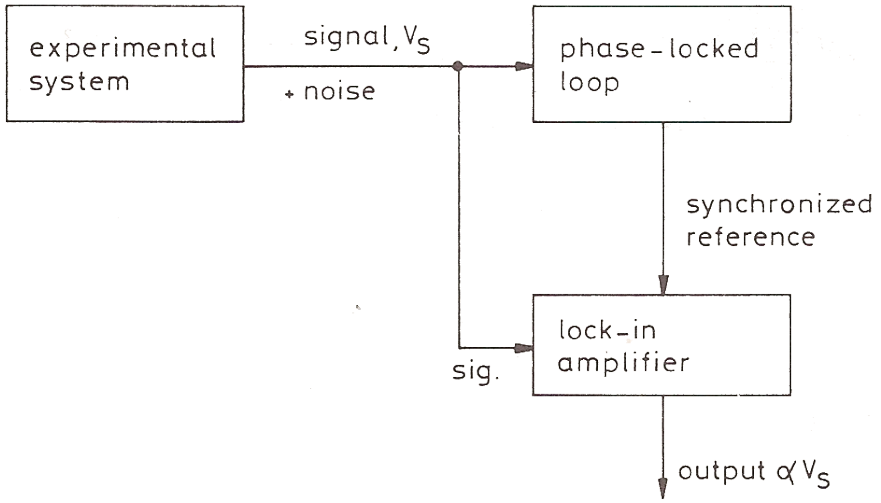

Fig. 6.1 General arrangement for generating a local reference in a lock-in recovery system. The phase-locked loop normally incorporates a second lock-in amplifier or phase-sensitive detector

obtained by filtering can be made equivalent to an increase in dynamic range; the noise-handling capacity of the system is improved without incurring a penalty due to increased offset and drift in the output. Alternatively, the filters could be used to reduce the total level of signal and noise at the input to the phase-sensitive detector with a possible increase in dynamic range sacrificed in favour of an improvement in linearity.

In either case, the improvements attributable to the use of filters are obtained at the expense of wideband capability. Furthermore, the use of filters offers no solution to restoring the loss of dynamic range which is always observed when a phase-sensitive detector is operated towards the upper limit of its recommended frequency range. In general applications, therefore, there may come a point where the trade-offs can only be improved by moving to systems with inherently better performance. This might mean purchasing a conventional lock-in amplifier from a later generation than existing equipment. From the point of view of a designer working at the limits of available technology, it might be more appropriate to investigate techniques which enhance the dynamic range of 'state of the art' phase-sensitive detectors. Methods of achieving this last objective, by means of synchronous heterodyning, will be described in Chapter 8.

The remaining limitations arising from harmonic responses and slew rate can be summarized with reference to specific measurement difficulties that have come to light in earlier chapters. Some solutions will be found in Chapters 8 and 9 which deal with more complex lock-in systems operating on the heterodyne and pulse-width-modulation principles. Unlike the phase-locked systems discussed in the next chapter, heterodyning and p.w.m. lock-in amplifiers are almost invariably supplied and used as self-contained units and cannot, in general, be built up as modular systems. It will be found that some drawbacks of these more advanced systems are highlighted along with their inherent advantages. Certainly no single lock-in amplifier will address all the limitations identified at the beginning of this chapter; it is up to the user to define his own requirements and select the system most suited to his needs.

6.2 Limitations arising from harmonic responses

6.2.1 Susceptibility to interference

The response of a switching phase-sensitive detector is characterized by a set of transmission windows centered on the reference frequency and its odd harmonics. As discussed in Chapter 3, the higher-order windows have negligible effect on the final noise output when the signal is accompanied by broadband white noise, but can have the most serious effect when the signal is accompanied by narrowband noise, or by discrete interference components.

The situation is by far the worst when the lock-in amplifier is synchronized to a low-frequency reference signal at 100 Hz or less. For example, a reference frequency of 100 Hz gives rise to 499 transmission windows between 100 Hz and 100 kHz at, or near to which, the lock-in amplifier can respond to interference voltages. Moreover, at 100 kHz, where the relative magnitude of the harmonic windows is of the order 1/1000, the window frequencies are only 0.2% apart, leaving a very small margin for adjusting the reference frequency if large-scale high-frequency interference components are to be avoided.

The problem is particularly acute when operating photometric equipment incorporating infra-red detectors. Such detectors often exhibit considerable thermal inertia and so limit the usable chopping frequency to a rnaximum of 10 Hz or so. The measurement difficulties are often aggravated by the need for a matching transformer in the signal path to ensure a good noise match to the lock-in amplifier (Appendix 5). This renders the system susceptible to inductive pick-up and can result in relatively high-level interference at mains frequency. Because the chopping frequency is limited to about 10 Hz it is necessary to choose a value of this which does not have an odd harmonic within a few hertz of mains frequency.* Otherwise, measurements will be perturbed by a difference-frequency beat component in the final output. Although the beat can always be reduced by using a larger output time-constant, this may not always be acceptable in view of the increased response time of the measurement system.

The standard technique of suppressing harmonic responses by using a tuned filter in the signal channel of a lock-in amplifier was described in Section 4.4.2. From the discussion given there it is evident that a tuned filter offers only a partial solution to the present problem unless the stability of the chopping system (usually mechanical) is of a very high order. The use of a mains-frequency notch filter would usually provide a better solution to measurement difficulties in this case.

6.2.2 Ambiguity due to the harmonic responses

In a harmonically responding. system the output due to a sinusoidal signal at frequency f_R is indistinguishable from one with three times the amplitude and the correct relative phase at frequency $3f_R$. This ambiguity in the response is not so serious when the requirement is for signal recovery at a fixed frequency with synchronous signal and reference, but can give rise to misleading results elsewhere. For example, the problems with signal distortion in a.c. bridge measurements have been described in Chapter 5. Also in that chapter were examples of operation with asynchronous signals, where the lock-in amplifier is used for wideband spectrum analysis. In this case the harmonic responses. will give rise to spurious lines in the output spectrum which occur whenever an odd harmonic of the reference coincides with a frequency component of the signal

* 9.3 Hz and 11 Hz are popular choices of chopping frequency in infra-red spectroscopy for this reason.

under investigation. Clearly, the use of a tuned filter for harmonic suppression has no.relevance to this type of measurement where the frequency is changing continually.

A further example relates to the discussion on phase-locked loops in Chapter 7. If the phase detector in such a loop is provided by a phase-sensitive detector or lock-in amplifier with harmonic responses it is possible for the loop to lock securely at an odd *sub-harmonic* of the incoming signal frequency. If, in addition, the signal has squarewave rather than sinewave form, there will be a large number of additional frequencies where lock could be acquired. The effect here will be at its worst when the phase-locked loop is required to operate automatically. When the locking signal is very noisy it is recommended that the initial locking conditions are brought under manual control. This would normally be sufficient to ensure that spurious locking to harmonic components was avoided in signal-recovery applications.

6.2.3 Detection of non-sinusoidal signals

The limitation here refers to the problem of obtaining a maximum response to non-sinusoidal signals in the presence of noise. It was shown in Section 3.5 that there is no problem in principle with so-called symmetrical signals. Here, the null-shift procedures can always be applied to produce a maximum response which is first-order independent of errors in the reference-channel phase setting. Otherwise, the response will be less than maximum, resulting in a loss of output signal.to-noise ratio and an increased susceptibility to phase changes in the reference channel and in the applied signal.

The null-shift procedures can be applied with confidence to all types of periodic signal when the lock-in amplifier has fundamental-only response. The detection system is always brought to a condition where the fundamental components of the signal and reference are in phase at the phase-sensitive detector. These systems will also ensure that this condition is reached by the automatic vector tracking system described in Chapter 5 in relation to two-phase lock-in amplifiers, and when the lock-in amplifier is brought under computer control as described in Chapter 10.

In practice, the loss of sensitivity to asymmetrical signals is often marginal in conventional systems and the null-shift procedures offer the ultimate advantage in giving a phase setting which can be reproduced under the noisiest conditions. The problem of defining the phase-shift to give maximum response to an asymmetrical signal remains, however, and serves as another example of the measurement difficulties associated with harmonically responding systems.

6.3 Slew rate limitations

The specification of reference-channel slew rate was dealt with in Chapter 4. The loss of accuracy, or indeed loss of lock, that results from a rapidly changing reference frequency must be considered when setting up any experiment involving a frequency sweep at the reference input.

In most cases, when using time-constant settings of a few hundred milliseconds or greater, the primary limitation on sweep speed comes from the response of the output filter of the lock-in amplifier**.** This aspect was discussed in some detail in Chapter 5 when outlining the behaviour of a two-phase system used as a high-resolution spectrum analyser. When operating at lower resolution, however, say with resolution bandwidths of several hundred hertz, and with a relatively wide frequency sweep (covering, for example, the entire audio frequency range), the

maximum allowable sweep speed will often be determined by the slew rate of the reference channel.

To take an example: suppose we operate with a 12 dB/octave output filter and a time constant setting of 1 ms. This gives a −6 dB frequency resolution of

$$\Delta f = 1/(\pi T_o) \cong 300 \text{ Hz}$$

The maximum allowed sweep-rate to avoid errors greater than 2% in the output filter has been given in Chapter 5. We obtain

$$R_{max} = 1/(8T_o^2) = 125 \text{ kHz/s}$$

If it was required to sweep from 2 kHz to 20 kHz, the system would allow a sweep time

$$T_s = 18.10^3/125.10^3 = 144 \text{ ms}$$

This short sweep time would allow the spectrum-analyser output to be displayed in conventional fashion on an oscilloscope. The problem is, of course, that a sweep time of 144 ms over the range 2 kHz to 20 kHz is equivalent to 72 ms/decade, a figure which is well beyond the slew rate capability of most conventional lock-in amplifiers. The reference processing circuits incorporated in pulse-width-modulated systems offer a solution to the slew rate problem as will be shown in chapter 9.

Phase-locking to noisy signals

7.1 Introduction

Phase-locked loops are used extensively in communications systems in demodulators for phase-and frequency-modulated signals[1-3] . They also form an important component in the reference channels of some conventional lock-in amplifiers and their more sophisticated counterparts to be described in Chapters 8 and 9. The requirement there is for frequency synthesis with the emphasis on precision and wideband capability. In the present context, however, we are specifically interested in the local generation of a reference waveform which is synchronized to a signal which may be obscured by noise. As explained in Chapter 6, this has relevance to all situations where it is not feasible to provide a direct reference connection.

An example of a 'remote' application is in the reception of satellite 'beacon' signals. Microwave beacons are transmitted at fixed power level while received power at an earth station is subject to variations related to atmospheric conditions on the propagation path. Detailed information about earth-space path attenuation can thus be obtained through long-term monitoring of the received signal[4].

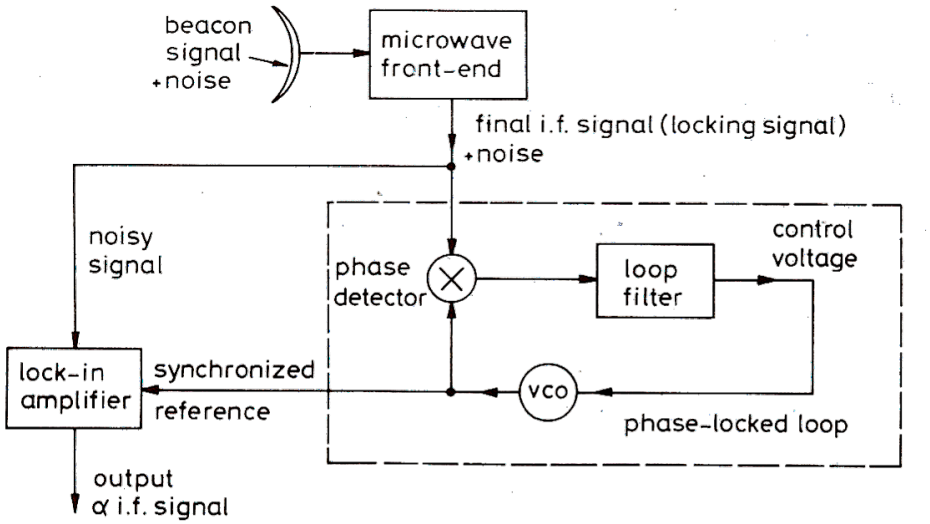

Fig. 7.1 Application of a phase-locked loop for the precision measurement of satellite beacon signals

On reception, the microwave signal is translated to successively lower frequencies and measurement of received signal strength is made at the output of the final i.f. stage. When the signal is subject to deep fading due to attenuation by rain on the propagation path, the output signal-to-noise ratio is seriously degraded and phase-sensitive detection provides the solution to precision measurement over a wide amplitude range. There is no reference directly available, so this must be obtained by phase-locking. In the most simple arrangement, shown in Fig. 7.1, a voltage-controlled oscillator is synchronized to the final i.f. signal. Usually, the

beacon signal is unmodulated, which enables a very narrow bandwidth to be defined for the purpose of signal-to-noise improvement. In experimental systems the final i.f. is often chosen to be less than 1 MHz to enable the use of commercial lock-in amplifiers with wide dynamic range, both for phase detection within the loop and signal measurement.

Fig. 7.1 represents a simple form of coherent receiver in which the phase detector measures the instantaneous phase difference between the voltage-controlled oscillator and the i.f. signal, which – in this context – is referred to as the *locking signal.* The phase detector output constitutes an error signal which is filtered and fed back to the control input of the v.c.o. The noise-rejection properties of the loop are determined by the choice of loop filter which also serves to control the dynamic behaviour of the loop.

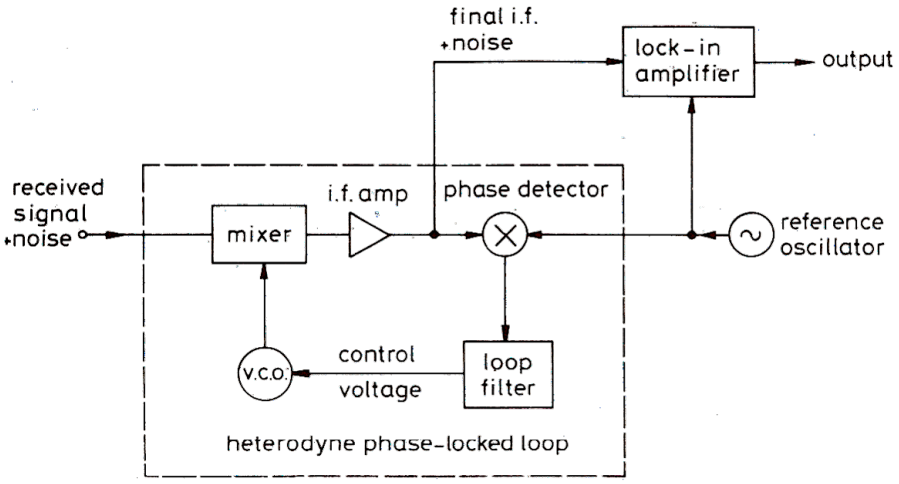

Fig. 7.2 Example of a heterodyne phase-locked loop. The loop is arranged to synchronize the output of the mixer to a stable reference oscillator

Many variations of this basic loop are used in receiver design[2,3]. Very often, the final mixer is incorporated within the loop to enable the final i.f. signal to be synchronized to a stable reference oscillator as shown in Fig. 7.2. This arrangement is widely used in satellite beacon monitoring[5]. Loops involving frequency changing are examples of 'derived' loops and broadly classified as *heterodyne* loops. A further example of a heterodyne loop used for frequency synthesis in a heterodyne lock-in amplifier is discussed in the next chapter.

The purpose of this chapter is to review the procedures whereby the basic loop of Fig. 7.1 can be made to operate successfully when the locking signal is very noisy and has variable amplitude. We shall find it possible to achieve synchronization when the locking signal is as much as 30 dB below the broadband noise level. However, to achieve such extremes of operation it is essential to understand the role of the loop filter in the locking process and to be able to select loop parameters in a systematic way. We shall therefore be looking at the mathematical background to loop operation and selecting some important results which have appeared in the literature in recent years.

In most mathematical treatments of the phase-locked loop the phase detector is modelled as an ideal multiplier. This is entirely suited to our purposes here since it enables us to extend our conclusions to systems using phase-sensitive detectors as phase detectors. In fact, multiplier-detectors are the preferred type when the locking signal is very noisy. Alternative types such as sequential or edge-triggered phase detectors are reserved for applications where the locking signal is

free from noise; for example, in the reference processing systems mentioned earlier.

When describing the properties of the loop we shall maintain a clear separation between the multiplier phase detector and the loop filter, but we must recognize that this separation is not so easily achieved when both functions are incorporated in a fully integrated lock-in amplifier. In fact, many users find the problem of relating the specifications of lock-in equipment to the circuit blocks of an ideal loop model very difficult to overcome. We shall accordingly give due attention to this most important practical aspect and derive optimization procedures on the basis of 'real' equipment.

The literature relating to phase-locked loops is extensive but the general level of treatment is such that the non-specialist reader will probably have considerable difficulty in following the complexities of loop behaviour under conditions of extreme noise. He will also find that signal amplitude variations are often excluded from the analyses entirely. Both of these aspects are of key interest to one whose involvement in phase-locking is motivated by signal recovery, and we shall attempt to take them into account in the course of this chapter.

7.2 'Static' analysis of a phase-locked loop

7.2.1 Phase detector output

We shall assume at the outset that the locking signal is noise-free and unmodulated and that the phase-locked loop has succeeded in pulling the v.c.o. frequency until it coincides exactly with that of the locking signal.

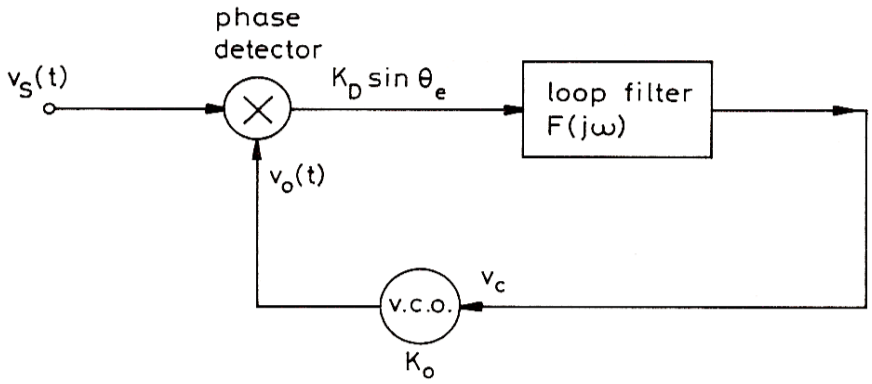

Fig. 7.3 **Static analysis of a phase-locked loop**
$v_s(t) = \sqrt{2}V_s \sin\omega_i t$
$v_o(t) = \sqrt{2}V_o \cos(\omega_i t + \theta_\varepsilon)$

We shall use a multiplier model for the phase detector and refer to the voltages in the loop identified in Fig. 7.3. On examining the loop in the locked condition we find an immediate consequence of using a multiplier-detector, namely that the loop must force a *quadrature* relationship between its two inputs to ensure proper operation (Section 2.4.4). We accordingly write the two phase detector inputs in the form:

$$v_s(t) = \sqrt{2}\, V_s \sin\omega_i t$$

$$v_o(t) = \sqrt{2}\, V_o \cos(\omega_i t + \theta_\varepsilon)$$

where θ_ε represents a small error measured from the ideal quadrature condition. Introducing a constant K, we obtain an output

$$v_D = KV_sV_o \sin\theta_\varepsilon$$

from the phase detector, which, in turn, supplies the feedback essential to loop control.

It is usual to express this result in the form

$$V_D = K_D \sin \theta_\varepsilon$$

Where K_D (volts/radian) measures the sensitivity of the phase detector. Since the amplitude of the v.c.o. is usually fixed we find that the phase detector sensitivity is proportional to the amplitude of the locking signal. In most of the 'standard' analysis which follows it will be assumed that K_D is fixed, but we must eventually consider the effect of variations in signal amplitude.

7.2.2 Static phase error

The notion of phase error must arise in a general description of any phase-control system. To calculate its magnitude we must first consider the operation of the v.c.o.

A linear model is to be used in which the frequency offset $\Delta\omega$ from the v.c.o. free-running frequency, ω_o, is proportional to the applied control voltage v_c. This gives

$$\Delta\omega = K_o v_c$$

where K_o is the v.c.o. gain factor with dimensions radians/V-s.

Under the essentially *static* conditions assumed here, when the amplitude and phase of the locking signal are free from variations, the v.c.o. offset frequency will be

$$\Delta\omega = \omega_i - \omega_o$$

which is maintained by a control voltage

$$v_c = K_D F(0) \sin \theta_\varepsilon$$

Here, $F(0)$ is the magnitude of the zero frequency response of the loop filter. The static frequency offset of the v.c.o. from its free-running frequency is therefore

$$\Delta\omega = K_o K_D F(0) \sin \theta_\varepsilon$$

The loop is usually designed to make the static phase error very small. In this case we can make the approximation $\sin \theta_\varepsilon \cong \theta_\varepsilon$ and so obtain

$$\theta_\varepsilon = \frac{\Delta\omega}{K_o K_D F(0)}$$

The static phase error can accordingly be reduced by increasing the d.c. gain of the loop, $K_o K_D F(0)$, and by initially setting the free-running frequency of the v.c.o. as close as possible to the incoming frequency, thus minimizing $\Delta\omega$.

7.2.3 'Hold-in' range

If we now suppose that the incoming frequency is subject to a slow drift, the loop will track and accommodate the frequency change by a change in the v.c.o. offset frequency. However, the results of the last section show that the maximum possible value of the frequency offset is

$$|\Delta\omega|_{max} = K_0 K_D F(0)$$

corresponding to a phase error magnitude of $\pi/2$.

Beyond this point, the loop loses control and synchronization is lost. We accordingly define the *hold-in* range of the phase-locked loop:

$$\omega_H = 2K_O K_D F(0)$$

The hold-in range is proportional to the d.c. gain of the loop and in practice may be far greater than the range over which the static phase error might be considered acceptable.

7.3 Dynamic response

7.3.1 Introduction

Consideration of the static behaviour of the phase-locked loop does not involve any detailed specification of the loop filter. At this point all we can say is that the filter must be effective in suppressing unwanted components at the output of the phase detector and that a high value of gain at zero frequency makes an important contribution to reducing static phase errors and increasing the hold-in range of the loop. These must be major considerations in selecting a filter but the final choice must be consistent with an overall loop response which is stable and predictable. These considerations involve a study of the *dynamic* behaviour of the loop when locked to an incoming signal which carries phase modulation.

7.3.2 The loop equation

We begin with reference to Fig. 7.4 which shows the relevant voltages at different points in the phase-locked loop.

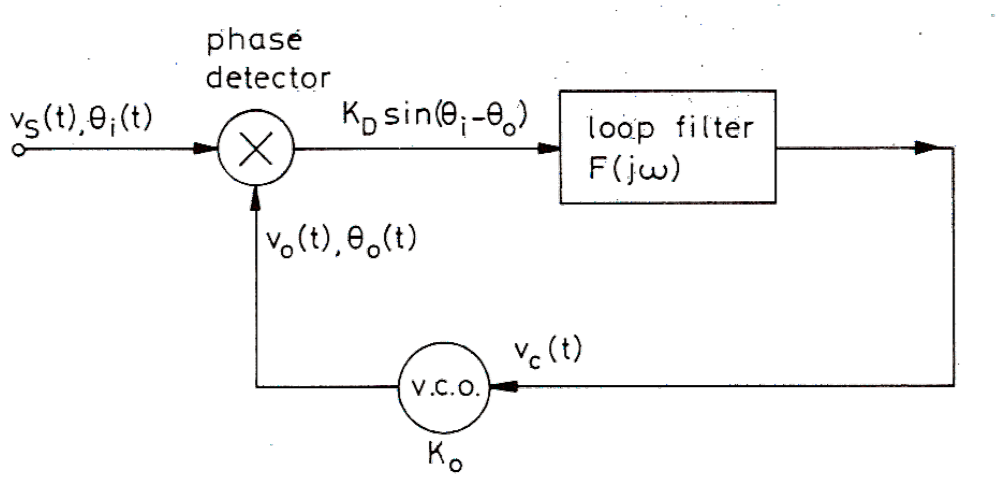

Fig. 7.4 Deriving the loop equation. The input to the loop is taken to be the phase variation of the locking signal

The loop is assumed to be in a locked condition with quadrature voltages at frequency ω_i applied to the phase detector. The v.c.o. phase is modulated by a variation $\theta_o(t)$ in response to the phase-modulation $\theta_i(t)$ carried by the locking signal. It is the nature of this response which is of interest here, in particular: how effective is the loop in 'following' the input phase variation?

The loop operates on the *difference* between $\theta_i(t)$ and $\theta_o(t)$, which produces a phase detector response

$$v_D(t) = K_D \sin [\theta_i(t) - \theta_o(t)]$$

We recall that K_D is strictly proportional to the amplitude of the incoming signal, which is assumed to be fixed throughout the following discussion.

The phase detector response is modified by the loop filter. The resulting output provides the control voltage, $v_c(t)$, to the v.c.o. and can be expressed in terms of the convolution:

$$V_c(t) = K_D \sin [\theta_i(t) - \theta_o(t)] \otimes f_L(t)$$

where $f_L(t)$ is the impulse response of the loop filter.

The resulting frequency deviation of the v.c.o. is proportional to $v_c(t)$, and for our purposes it is convenient to express this deviation in terms of the time derivative of the v.c.o. phase, giving

$$\Delta\omega(t) = d\theta_o(t)/dt = K_o K_D \sin [\theta_i(t) - \theta_o(t)] \otimes f_L(t)$$

This is the general non-linear equation describing the operation of the phase-locked loop. The non-linearity is inherent in the phase detector response, but can be over-come by assuming that the loop is designed to give a very small dynamic phase error. In this case, we write

$$\sin [\theta_i(t) - \theta_o(t)] \cong \theta_i(t) - \theta_o(t)$$

for $|\theta_i(t) - \theta_o(t)| << 1$ radian

and so obtain the *linear* loop equation

$$d\theta_o(t)/dt = K_o K_D [\theta_i(t) - \theta_o(t)] \otimes f_L(t)$$

The conventional approach to this equation is to derive the transfer function relating the 'input' and 'output' phase variations. We accordingly take Laplace transforms of both sides:

$$s\theta_o(s) = K_o K_D [\theta_i(s) - \theta_o(s)] F(s)$$

and rearrange to obtain the transfer function

$$H(s) = \frac{\theta_o(s)}{\theta_i(s)} = \frac{K_0 K_D F(s)}{s + K_0 K_D F(s)}$$

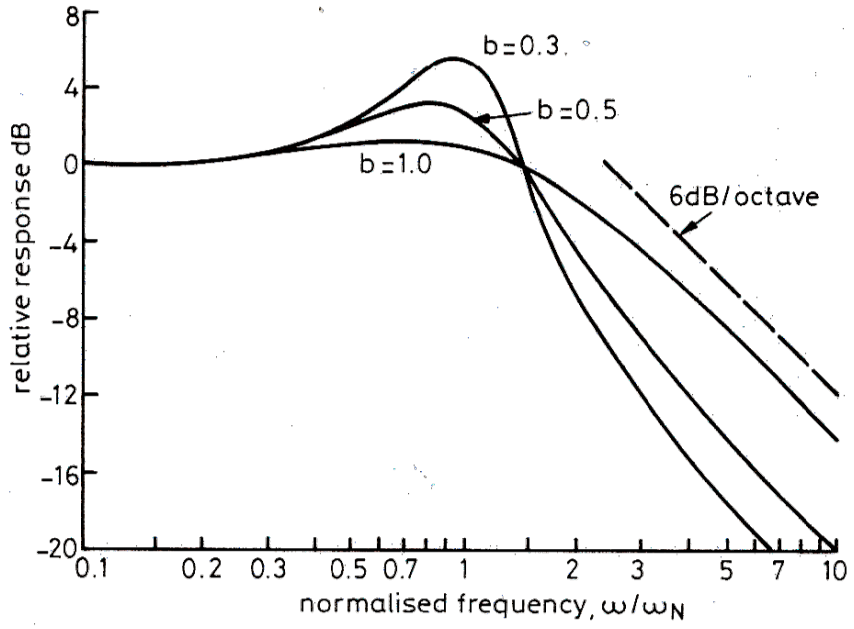

Fig. 7.5 Magnitude of a second-order closed-loop transfer function at various values of the damping ratio, *b*

Given $H(s)$ we can calculate the behaviour of the v.c.o. phase in response to variations in the phase of the locking signal. If we put $s = j\omega$ we obtain the closed-loop frequency-response function $H(j\omega)$ which could – in principle – be of any order, determined by the frequency-response function of the loop filter. In reality phase-locked loops are more often second-order, characterized by two parameters; natural frequency ω_N and damping ratio b. Variation of these parameters gives rise to a family of frequency responses with the magnitudes shown in Fig. 7.5. When the damping is light the loop is resonant and susceptible to oscillatory transients following abrupt changes of phase on the locking signal. As we shall see, it is important to have sufficient design variables to shape the response of the loop. The choice of loop filters in this respect is discussed below.

7.4 The second-order loop

A second-order loop can be obtained by using a loop filter of the simple *RC* low-pass type, shown as an active filter in Fig. 7.6.

The filter has a frequency-response function

$$F(j\omega) = F(0)/(1 + j\omega T_0)$$

$$F(0) = R_0/R_1, \quad T_0 = R_0C$$

resulting in a closed-loop frequency response function:

$$H(j\omega) = \frac{K_0K_DF(0)/T_0}{K_0K_DF(0)/T_0 + j\omega/T_0 - \omega^2}$$

This is the frequency response which would be obtained if a commercial phase-sensitive detector with a standard *RC* filter and 'time constant' control was used in a phase-locked loop. For the purpose of this and following sections we can put the denominator of $H(j\omega)$ in the standard form

$$\omega_N^2 + 2jb\omega\omega_N - \omega^2$$

and then identify the natural frequency and damping factor. In this case, we have

$$\omega_N = (K_oK_DF(0)/T_0)^{½}, \quad b = ½ (K_oK_DF(0)T_0)^{½}$$

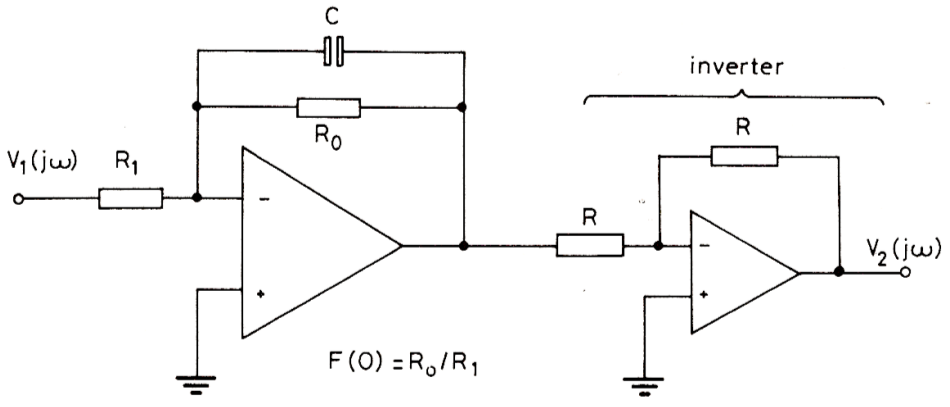

Fig. 7.6 Active RC low-pass filter. The inverter is included to correct for the phase reversal in the first amplifier. $F(j\omega) = V_2(j\omega)/V_1(j\omega)$

The d.c. gain of the loop, $K_0K_DF(0)$ is correctly identified as the loop gain. We find that the natural frequency increases with loop gain while the damping is reduced. In fact, there are insufficient design parameters available to ensure that a particular combination of loop characteristics can be obtained. For example, at a given value of loop gain, the loop bandwidth can only be narrowed by increasing

the time constant of the loop filter, resulting in a loss of damping. The transient response of the loop can thus be seriously degraded when narrow bandwidths are required.

These problems can be overcome by adopting the approach used by control engineers: to use a loop filter with a lag-lead response as in Fig. 7.7. This simple modification is effective because the additional resistor gives control of the damping of the phase-locked loop. Hence, for a fixed value of the loop gain, the natural frequency and damping can be set independently.

A variation on this filter type is shown in Fig. 7.8. We shall refer to this as an 'imperfect integrator'. Both lag-lead filters and imperfect integrators are used in commercial lock-in amplifiers adapted for phase-locking. The extremely high value of low-frequency gain obtainable with the imperfect integrator gives a phase-locked loop with wide tracking capability consistent with a low value of static phase error.

In most systems using lag-lead filters, however, the ratio R_0/R_2 is so high that there is relatively little difference in handling characteristics between the two filter types.

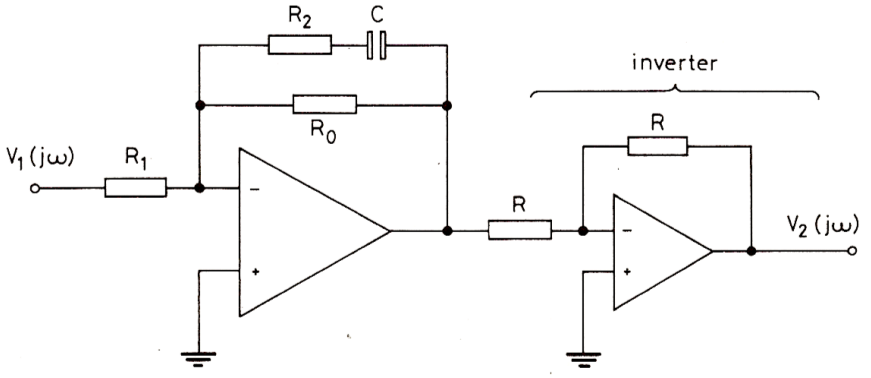

Fig. 7.7 **A lag-lead filter**
$F(\infty) = R_2/R_1,\ R_0 >> R_2$

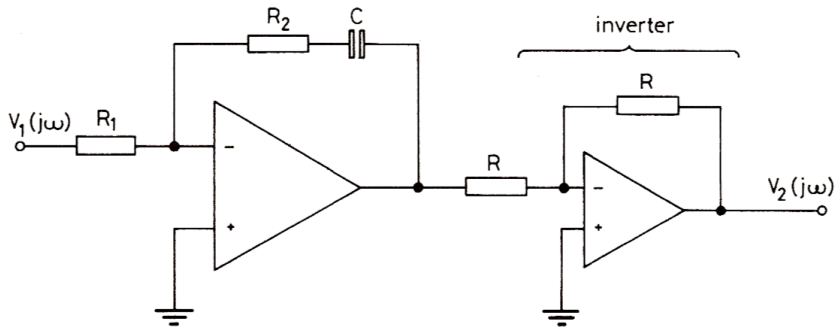

Fig. 7.8 **An imperfect integrator**
$F(\infty) = R_2/R_1$

As we shall see, the recommended procedures for optimizing the loop noise performance are the same in each case. We can adopt a uniform approach that deals correctly with both filter types by expressing the loop filter frequency response functions as follows:

$$\left.\begin{array}{l}\text{lag-lead filter } (R_0/R_2 >> 1)\\ \text{imperfect integrator}\end{array}\right\} F(\mathrm{j}\omega) = F(\infty)(1+\mathrm{j}\omega T_2)/\mathrm{j}\omega T_2,\ T_2 = R_2C$$

where in both cases, the high-frequency gain of the filter, $F(\infty)$ is a real ratio, R_2/R_1.

If the loop-frequency response function is now derived, we find, for both loop filters:

$$\omega_N = (K_0K_DF(\infty)/T_2)^{1/2}$$

and

$$b = \tfrac{1}{2}\,\omega_N T_2$$

The expression for b is exact for the imperfect integrator and an excellent approximation for the lag-lead filter, for all reasonable values of damping ($b^2 >> R_2/R_0$).

Gardner[1] has shown that for this type of filter the loop gain is correctly given by the quantity $K_0K_DF(\infty)$.

The noise bandwidth of the loop becomes very important when we deal with the problems of phase-locking in noise. This can be expressed in terms of ω_N and b as:

$$B_L = \frac{\omega_N}{2}\left[b + \frac{1}{4b}\right]$$

For any value of ω_N the noise bandwidth is minimized when $b = \tfrac{1}{2}$, and is then given by $\omega_N/2$. Note that, following convention, the noise bandwidth is given in hertz, while ω_N is expressed in radians/s.

When using a lag-lead filter or imperfect integrator the natural frequency of the loop and the loop damping ratio exhibit a square-root dependence on loop gain. The practical significance of this becomes evident when the amplitude of the locking signal is allowed to vary in the course of a phase-locked experiment. We saw in section 7.2.1 that the phase-detector constant K_D is strictly proportional to V_s. If all the other loop components are held at a fixed value we obtain the following functional dependence on V_s for ω_N and b:

$$\omega_N = c_1V_s^{1/2}$$

$$b = c_2V_s^{1/2}$$

where c_1 and c_2 are suitably defined constants. The variation of the noise bandwidth now takes the form:

$$B_L = \frac{c_1}{8c_2}\left[1 + 4c_2^2V_s\right]$$

showing that B_L increases in direct proportion to the signal amplitude.

7.5 Noise and phase-locked loops

When the locking signal is accompanied by a random noise disturbance, noise enters the phase-locked loop via the phase detector and gives rise to a random phase error on the v.c.o. output. When the phase error is small it appears as phase

noise or phase jitter on the v.c.o.* relative to the phase of the locking signal. In general, a small amount of jitter should not seriously affect an associated lock-in amplifier when the v.c.o. output is used as a reference voltage for measuring the amplitude of the locking signal.

If the noise input to the loop is allowed to exceed a certain level, the resulting phase error will become sufficiently large to seriously affect loop operation. In the extreme case the loop will drop out of lock completely, but before this point is reached there is a range of noise inputs where the loop is susceptible to a phenomenon known as 'cycle slipping'. In this regime the loop can be kicked temporarily out of lock by a random noise event and then restore its equilibrium an integral number of cycles away from its original condition. The effect can recur repeatedly when the noise levels are sufficiently high.

The analysis of loop operation in this regime is formidable, corresponding as it does to non-linear operation. Fortunately, our main concern is to ensure that the random phase error never reaches a point where there is a significant probability of cycle clipping. Only then can we ensure that the signal recovery system will operate for long periods of time without falling out of lock.

In principle, the random phase error due to external noise can always be reduced by designing for a loop frequency response function with small bandwidth. In this context, the noise bandwidth of the loop introduced in the last section becomes the relevant factor. Its role in evaluating the effect of noise in the input can be clarified as follows. We imaging that the phase-locked loop is a special kind of bandpass filter which accepts a noisy signal at frequency f_s. The output is a clean signal at the same frequency having a residual phase error due to noise. The filter has a frequency response function given by translating the loop frequency-response to the signal frequency f_s. This gives an effective noise bandwidth of $2B_L$ as illustrated in Fig. 7.9.

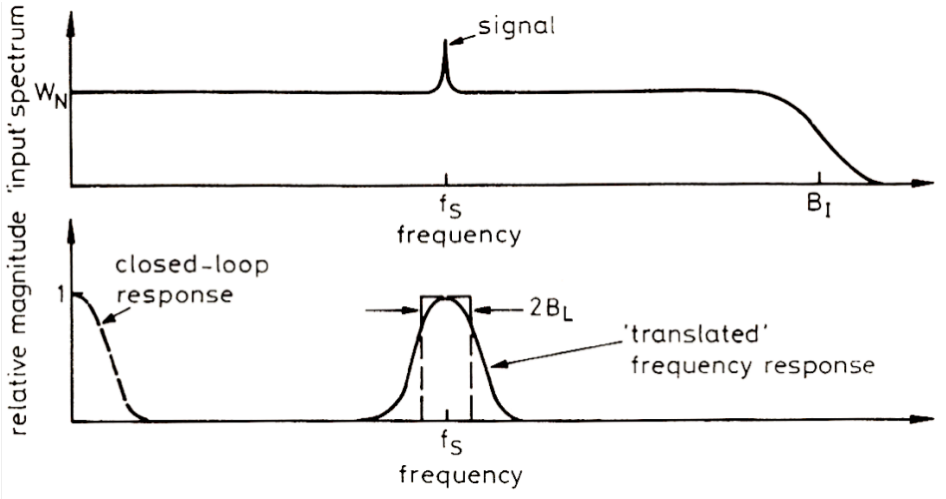

Fig. 7.9 Filter effect of a phase-locked loop
For the purpose of calculation, the loop can be regarded as a bandpass filter with noise bandwidth $2B_L$

The input noise has a bandwidth B_I with uniform spectral density W_N. The input signal-to-noise ratio is therefore

* The effect of phase noise *inherent* in the v.c.o. will be mentioned at a later stage. we shall generally assume that the noise in the loop, giving rise to a fluctuating phase error, is largely of external origin; that is, noise appearing with the locking signal.

$$SNR_I = V_s^2/(2B_I W_N)$$

while the signal-to-noise ratio *measured within the effective bandwidth of the loop* is

$$SNR_L = V_s^2/(4B_L W_N)$$

We shall not attempt to attach any physical significant to SNR_L which is often loosely called the loop signal-to-noise ratio. It can be shown nevertheless that the probability of cycle slipping will be very small provided that this noise measure is greater than about 6 dB (see, for example, the results reviewed by Gardner[1] and Blanchard[2]). We shall err on the side of safety when calculating the required value of B_L and use the criterion:

$$SNR_L = V_s^2/(4B_I W_N) \geq 10$$

In this condition the system is amenable to a small-signal analysis. The results in the literature[1] give a good estimate of the mean-square phase error* on the v.c.o. output, namely:

$$\overline{\theta_N^2} = \frac{1}{2SNR_L}$$

Let us now assume that the signal-to-noise ratio of the locking signal has been estimated and that the input bandwidth is known. We have:

$$SNR_L = SNR_I B_I/(2B_L)$$

and, putting $SNR_L \geq 10$, we obtain the condition:

$$B_L \leq SNR_I B_I/20$$

to ensure that lock is maintained in the presence of noise on the locking signal. With this condition satisfied, the v.c.o. has a mean-square phase error

$$\overline{\theta_N^2} \leq 0.05 \text{ radian}^2$$

The next step is to derive optimization procedures which guarantee that these conditions obtain even when the signal amplitude is allowed to vary over a wide range. This will represent a significant improvement over the 'standard' treatment where the signal – and hence the loop parameters – are fixed, and changes in signal-to-noise ratio are 'arranged' by allowing the noise intensity to vary.

It was noted earlier that we have taken no account of phase noise which is inherent in the v.c.o. itself. In designing phase-locked loops for high precision it is normally arranged that the loop bandwidth is wide enough to accommodate the bulk of these variations, and it can be shown that their effect is reduced in proportion to the loop gain of the system. This requirement is obviously incompatible with choosing a narrow loop bandwidth to combat noise of external origin. Also, it will be shown that having determined the bandwidth of the loop, consistent with a reasonably high damping ratio, the choice of loop gain is restricted. It turns out, however, that when operating in the audio-frequency range with loop components of reasonable stability, any residual jitter due to imperfections on the v.c.o. will be masked by noise on the locking signal, provided that the loop bandwidth is not made unnecessarily small. This is also an important consideration when the loop is expected to track 'slow' frequency variations on the locking signal as will be shown in Section 7.7.

---

* The mean square phase error is calculated on the assumption that the *static* phase error of the loop (see Section 7.2.2) has been trimmed to zero.

7.6 Optimization procedure

The optimization procedure will be derived for a classic signal recovery example, where the locking signal amplitude is varying over a wide range against a noise background with uniform spectral characteristics. It is necessary for us to have an estimate of the minimum value of the locking signal, corresponding to the worst-case signal-to-noise ratio at the input to the loop, $(SNR_I)_{MIN}$.

When the loop incorporates a lag-lead filter or imperfect integrator, the noise bandwidth will be dependent upon signal level as described in Section 7.4. We accordingly design the loop to have its minimum noise bandwidth at minimum signal level. Using the result of Section 7.5, the minimum noise bandwidth is chosen to satisfy:

$$(B_L)_{MIN} \leq (SNR_I)_{MIN} B_I/20$$

The behaviour of the loop as the signal increases from its minimum value can be predicted as follows.

First of all we note that the loop signal-to-noise ratio is given by:

$$SNR_L = V_s^2/(4B_L W_N)$$

Since B_L increases, at most, linearly with signal we find a steady improvement in SNR_L as the locking signal increases, causing a proportional reduction in the mean-square phase error (= $(2SNR_L)^{-1}$) due to external noise. The rise in noise bandwidth also makes the loop more effective in reducing the effect of phase noise inherent in the v.c.o. It is shown in section 7.4 that the loop damping will also rise with signal level. Fortunately, this rise is accompanied by an increase in the natural frequency ω_N. This joint behaviour results in a response which is not too 'sluggish' as would be the case if b, alone, were to increase.

Turning now to the question of loop damping: this would normally be chosen to have a minimum value of about ½ to ensure that the loop transient response is not marred by excessive overshoot and 'ringing'. It also turns out that choosing this particular minimum value greatly simplifies the optimization procedures. The damping ratio falls with locking signal amplitude; we therefore arrange for a minimum damping ratio of ½ at the minimum anticipated signal level.

The general expressions for noise bandwidth and damping ratio are:

$$B_L = \frac{\omega_N}{2}\left[b + \frac{1}{4b}\right]$$

$$b = ½\omega_N T_2$$

Hence, if we put $b = b_{MIN} = ½$ and decide on a minimum value for B_L we find that T_2 is given immediately:

$$T_2 = 1/(2B_L)_{MIN}$$

while the minimum value of ω_N is given by

$$(\omega_N)_{MIN} = 1/T_2 = 2(B_L)_{MIN}$$

Using the value of the phase detector constant appropriate to minimum signal level, together with the required value of T_2, the loop can now be designed to give $\omega_N = (\omega_N)_{MIN}$ at $b = b_{MIN}$. From the results given in Section 7.4 this implies that the minimum value of loop gain is given by

$$K_0(K_D)_{MIN} F(\infty) = 1/T_2$$

T_2 has already been determined, so we cannot improve the minimum loop gain. We have already noted that residual phase noise on the v.c.o. will make a contribution to phase jitter under conditions of low loop bandwidth and low loop gain. The effect of these incidental variations must be checked in a trial run using a noise-free signal. This is, in any case, a useful first step in setting up a phase-locked loop of even moderate complexity.

7.7 Notes on acquisition and tracking

'Acquisition' is a general term which is used to describe the extremely complex processes by which a phase-locked loop picks up an incoming signal and moves towards lock.

When the loop is in an unlocked condition the output from the phase detector will initially be a 'beat' waveform which contains the difference frequency between the locking signal and the instantaneous frequency of the v.c.o. The maximum possible peak-to-peak value of the beat waveform is $\pm K_D$ volts which is subsequently attenuated by the effect of the loop filter. If the resulting voltage swing at the v.c.o. is sufficient to make the v.c.o. and input frequencies coincide, the system moves smoothly into lock without slipping cycles.

In a second-order system using a lag-lead filter the attenuation at high beat frequencies (corresponding to a high initial frequency difference between the locking signal and the v.c.o.) has a constant value $F(\infty)$. In this case the maximum available swing of the v.c.o. frequency is $\pm K_0 K_D F(\infty)$ which defines the *capture range* of the loop.

For initial frequency offsets within the capture range, locking is assured and fast. However, an important consequence of using an integrating loop filter is that the loop will eventually 'pull-in' to a locking signal which is at a frequency far removed from the v.c.o. frequency. The reason is that the 'beat' waveform is highly asymmetrical and contains a d.c. component. This can build up in the integrator and gradually drive the v.c.o. towards the signal. Pull-in can be a cumbersome and time-consuming process which is greatly affected by noise on the signal. It is usually overcome by manual tuning of the v.c.o. to bring the frequency difference within the capture range of the loop. Locking is then, for practical purposes, instantaneous.

It must be assumed that acquisition is always assisted when locking to very noisy signals and that the loop is finally trimmed to minimize the offset frequency of the v.c.o. As indicated in Chapter 6, manual assistance of acquisition serves also to avoid ambiguities in locking when the phase detector has responses to odd harmonics of the v.c.o. frequency.

Finally, let us look briefly at the case where the locking signal frequency is subject to a slow variation or drift. It can be shown that a loop containing an imperfect integrator is capable of tracking a locking signal with a changing frequency, but that the phase of the v.c.o. suffers a 'slip' relative to the phase of the locking signal. This is analogous to the phase slip observed in the reference channel of a lock-in amplifier in response to a changing reference frequency, as discussed in Section 4.5. In the case of the phase-locked loop the rate of change of locking frequency R is related to the phase slip θ by [1,2]:

$$2\pi R = \omega_N^2 \theta$$

where R is expressed in Hz/s. Note that the phase slip is *exclusive* of any transient phase error that may occur following the application of a frequency sweep.

When designing a loop for recovery from noise it has been suggested that the natural frequency has a minimum value

$(\omega_N)_{MIN} = 2(B_L)_{MIN}$

If the phase slip is to have a maximum value θ_{MAX} then we obtain the condition

$R \leq 2(B_L)^2_{MIN}\ \theta_{MAX}/\pi$

For example, a maximum phase slip of 5° gives

$\theta_{MAX} = 5 \times \pi/180$ radians

If $(B_L)_{MIN} = 100$ Hz, then

$R \leq 556$ Hz/s

The consequence of using very low noise bandwidths is a severe reduction in the allowed tracking rate. Thus if $(B_L)_{MIN}$ is only 5 Hz we obtain

$R \leq 1.39$ Hz/s

The same quadratic dependence of frequency sweep-rate on system bandwidth was noted in Section 5.5 in relation to spectrum analysis..

7.8 Using a lock-in amplifier for phase-locking

7.8.1 Introduction

So far, the treatment of phase-locked loops has been fairly general in the sense that a clear distinction has been made between the phase detector and the loop filter. As remarked in the introduction to this chapter, this separation is not so easily achieved when using a fully integrated lock-in amplifier. Let us therefore begin with Fig. 7.10, which shows the internal arrangement of a lock-in amplifier fitted with a 'phase-lock' option.

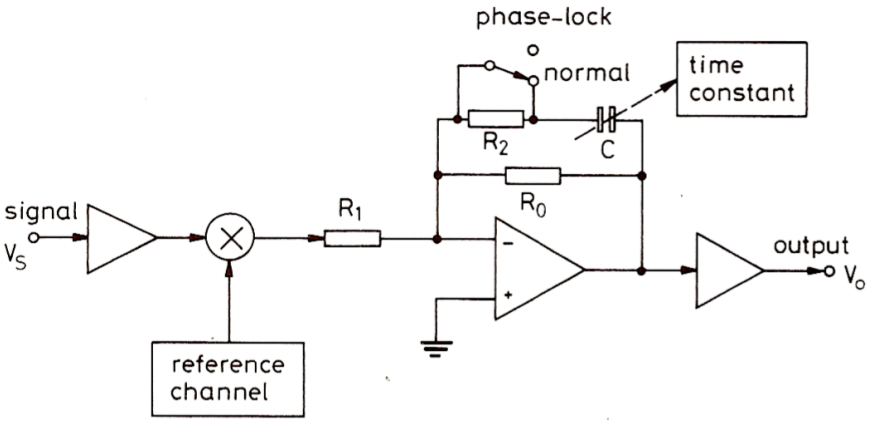

Fig. 7.10 Internal arrangement of a lock-in amplifier with 'phase-lock' option. Time constant $T_0 = CR_0$ in 'normal' mode

The phase-sensitive detector is shown schematically as a multiplier-detector while the output filter now serves a dual function. In 'normal' mode the filter capacitors are selected from the time-constant control, giving a range of values $T_0 = R_0C$. In 'phase-lock' mode the damping resistor R_2 is switched into circuit while R_0, the feedback resistor, may be removed if the loop filter is to be an imperfect integrator. One problem is that R_2 is not usually specified by manufacturers and may not be obtainable except through inquiry or by looking at the circuit diagram. It is certain, however, that the ratio R_0/R_2 will be large. R_0 is usually of the order of 10 MΩ while R_2 is commonly 10 kΩ (for example, in the Brookdeal series of phase-sensitive detectors and lock-in amplifiers).

7.8.2 Identifying the loop constants

When used as a phase detector with a time-constant filter, the observed sensitivity of a lock-in amplifier to static phase errors is, from Section 4.1:

$V_F(V_s/S_F)$ volts radian

This observed phase sensitivity is *inclusive* of the d.c. gain of the time-constant filter, R_0/R_1. In 'phase-lock' operation, the phase detector sensitivity K_D that appears in the loop equations is *exclusive* of the gain provided by the filter block. We thus obtain:

$$K_D = \frac{V_s V_F}{S_F}(R_1/R_0)$$

In practice, the ratio R_0/R_1 accounts for the bulk of the phase-sensitive detector dynamic reserve. The phase detector constant to be used in the loop equations is therefore given approximately by the overall phase sensitivity reduced by a factor equal to the dynamic reserve. For either a lag-lead filter or imperfect integrator we have $F(\infty) = R_2/R_1$. The loop gain is accordingly

$$K_0K_DF(\infty) = K_0\ \frac{V_s\ V_F}{S_F}\left(\frac{R_1}{R_0}\right)\frac{R_2}{R_1}$$

Putting $T_2 = R_2C$, $T_0 = R_0C$ we obtain

$$K_0K_DF(\infty) = K_0\ \frac{V_s\ V_F}{S_F}\left(\frac{T_2}{T_0}\right)$$

Here, T_0 is the time-constant setting on the front panel of the lock-in amplifier. In view of our earlier remarks about assigning a value to R_2, T_2 might have to be identified from a circuit diagram or by inspection of the time constant switch.

Using this value of loop gain, the natural frequency of the loop can now be put in the form

$$\omega_N = \left(\frac{K_0V_sV_F}{S_FT_0}\right)^{1/2}$$

In the usual arrangement, T_0/T_2 appear in a fixed ratio:

$$T_0/T_2 = R_0/R_2 = r$$

The damping ratio is therefore:

$$b = \tfrac{1}{2}\omega_NT_2 = \tfrac{1}{2}\omega_NT_0/r$$

Provided that r can be identified, the loop parameters ω_N and b are now given in terms of V_s and the lock-in amplifier settings; time constant T_0 and sensitivity S_F, for an output voltage swing $\pm V_F$.

7.8.3 Optimization procedures for lock-in amplifiers

The specification for the locking signal gives us a minimum anticipated r.m.s. signal level V_{MIN}, an input noise bandwidth B_I and a worst-case signal-to-noise ratio $(SNR_I)_{MIN}$. This enables us to choose a minimum value of loop noise bandwidth $(B_L)_{MIN}$ commensurate with the input conditions:

$$(B_L)_{MIN} \le (SNR_I)_{MIN}\,B_I/20$$

Using the outline procedure given in Section 7.6 we design for b_{MIN} = ½ and immediately obtain a bound on the required value of T_0 (= rT_2):

$$T_0 \geq \frac{r}{2(B_L)_{MIN}}$$

This leaves us to choose K_0 and S_F to satisfy

$$(\omega_N)_{MIN} = r/T_0 = \left(\frac{K_0 V_{MIN} V_F}{S_F T_0}\right)^{1/2}$$

In summary:

(i) Determine minimum required value of B_L;

(ii) Calculate the time-constant setting T_0 to ensure $(B_L)_{MIN}$ at the smallest anticipated signal level;

(iii) Choose values of K_0 and S_F to satisfy

$K_0 V_{MIN}/S_F = r^2/(V_F T_0)$.

The following examples will help to put the optimization procedures for lock-in amplifiers into perspective.

Example 1

A lock-in amplifier has a maximum time constant of 100 s with the ratio $r = R_0/R_2 = 1000$. What is the smallest value of noise bandwidth that can be achieved consistent with a minimum damping ratio of ½?

The smallest achievable noise bandwidth in phase-locked loop operation is

$$\frac{r}{2(T_0)_{MAX}} = 5 \text{ Hz}$$

Example 2

What is the worst possible signal-to-noise ratio on the locking signal that can be handled by a lock-in amplifier in phase-lock mode?

This is a question which is often asked but to which there is no direct answer. If we calculate the minimum achievable noise bandwidth – as in example 1 – then the best we can do is find a bound on the product $SNR_I B_I$:

$$SNR_I B_I \geq 10r/(T_0)_{MAX}$$

In example 1, the system could cope with a mean-square signal-to-noise ratio of −10 dB (1/10) in an input bandwidth of 1 kHz, or a ratio of −30 db (1/1000) in a 100 kHz bandwidth. We could also be certain that the phase-locked loop could hold lock under more adverse conditions, since a fairly conservative bound on B_L was taken in Section 7.5. It is likely, however, that if the design were carried out with a less stringent bound or with the minimum damping reduced much below ½, then the resulting handling characteristics and phase jitter would verge on the unacceptable.

Example 3

Using the lock-in amplifier specified in example 1, outline the optimization procedures for a signal of 50 kHz appearing at a minimum level of 1 mV with an estimated worst-case signal-to-noise ratio of −20 dB (1/10). The input noise bandwidth is set by signal conditioning filters to a value of 5 kHz and a v.c.o. is available with a sensitivity K_0 of $2\pi 10^4$ radians/V−s at the signal frequency.

First of all, the noise bandwidth of the loop: at minimum signal level this must satisfy

$$(B_L)_{MIN} \leq (SNR_I)_{MIN} B_I/20$$

$$\leq 10^{-1} \times 5 \times 10^3/20 \text{ Hz or } 25 \text{ Hz}$$

The smallest time constant consistent with this value is:

$$T_0 = r/(2 \times 25)$$

We have $r = 1000$; hence

$$T_0 = 20 \text{ s}$$

If a time constant of 20 s is not available, the next largest should be selected. The optimization procedures will ensure that the damping ratio will have a minimum of ½ as required.

Finally:

$$K_0 V_{MIN}/S_F = r^2/(V_F T_0)$$

K_0 has been given. Using $V_F = 10$ V we obtain:

$$V_{MIN}/S_F = 0.079$$

The minimum anticipated signal level is 1 mV r.m.s. The lock-in amplifier should therefore be set to a full-scale sensitivity

$$S_F = V_{MIN} / 0.079 = 12.6 \text{ millivolts}$$

In practice, a full-scale sensitivity of 10 mV will be indistinguishable from the optimum setting. Note that the locking signal may subsequently take larger values which exceed the full-scale sensitivity of the lock-in amplifier without incurring a fault condition, *provided* the total allowable swing on the signal input is not exceeded.*

If an inconveniently low or high ratio V_{MIN}/S_F is predicted, there may be scope to change the value of K_0. It should not be overlooked that many v.c.o.s found in measurement laboratories have overlapping decade ranges. By judicious choice of operating frequency it may be possible to change K_0 by a factor of 10 or even 100, depending on range selection.

When very narrow loop bandwidths are required, the ability to track signals of changing frequency (Section 7.7) must be taken into consideration. Also, a trial run using a noise-free signal should be carried out to assess the residual phase jitter in the loop arising from incidental phase- and frequency-modulation on the v.c.o. Here again, the ability to switch between overlapping ranges might prove useful and enable the v.c.o. to be operated in a region where its self-noise is lower.

7.9 The final measurement

Let us finally return to the measurement system proposed at the beginning of this chapter, where the v.c.o. output is used as a reference voltage for the detection of the locking signal in a second lock-in amplifier.

Two-phase lock-in amplifiers modified for phase-locking are ideal for this type of measurement; ‘quadrature’ channel B is used for phase-locking, leaving the

* This comment is in line with the procedures for increasing the phase sensitivity of lock-in amplifiers that were given in Sections 4.7.5. and 5.2.3.

signal to be measured in ‘in-phase’ channel A. When the dual phase-sensitive detectors are fitted with independently switched ‘expand’ amplifiers it is possible to operate the two channels at sensitivities differing by a factor of 10 or even 100. This is usually sufficient to ensure that the sensitivities of the two phase-sensitive detectors can be separately optimized for phase locking and for signal detection.

When the signal is noisy, the v.c.o. output inevitably exhibits phase jitter relative to the locking signal. Fortunately, as a result of the $\cos\theta$ dependence, this phase jitter has only a second-order effect on the measurement at the second phase-sensitive detector. It turns out in practice that if the phase jitter in the loop is sufficiently small to ensure long-term locked operation without slipping cycles (as we have assumed throughout), then the phase jitter will have minimal effect in the final measurement. In the worst case, at minimum signal level, its effect can be reduced by increasing the output time constant in the second lock-in channel beyond the value normally used at a given signal-to-noise ratio.

7.10 References

1 GARDNER, F.M. (1979): ‘Phase lock Techniques’ (John Wiley, New York)

2 BLANCHARD, A. (1976): 'Phase-locked loops: Application to coherent receiver design’ (John Wiley, New York)

3 LINDSEY, W.C. (1972): ‘Synchronization systems in communication and control’ (Prentice-Hall, N.J.)

4 HOGG, D.C. and CHU, T.S. (1975): ‘The role of rain in satellite communications’, Proc. *IEEE*, 63, pp. 1308-1331

5 BAYLISS, A. (1974): ‘A guide for orbital test satellite experiments’. Proc. European Conf. on Electrotechnics, Eurocon ’74, Amsterdam

CHAPTER 8

Heterodyne lock-in amplifiers

8.1 Introduction

In a heterodyne lock-in amplifier, phase-sensitive detection is carried out at a relatively high, fixed, frequency following a stage of frequency translation of the applied signal. Since the phase-sensitive detector operates at a frequency that bears no harmonic relationship to the applied signal, the harmonic responses that characterize a conventional lock-in system are suppressed. Practical lock-in amplifiers operating on the heterodyne principle thus conform very closely to ideal fundamental-only responding systems. The first of these, offering relative freedom from harmonic responses over a moderate frequency range, was the Ithaco Dynatrac lock-in amplifier. This was introduced in both single- and two-phase versions in the early 1970s.

In the early system the benefits of fundamental-only response were obtained to the detriment of performance in other areas. For example, the true frequency range was only about one decade and coverage of the audio-frequency range required a total of five sets of plug-in circuit cards. Also, the phase accuracy left much to be desired, particularly at the extremes of the individual ranges. Even under the most favourable conditions, this heterodyne system suffered by comparison with the high precision of conventional lock-in amplifiers. The system nevertheless enjoyed considerable success and served to focus attention on the limitations associated with harmonic responses which were reviewed in Chapter 6.

The block diagram of the Ithaco Dynatrac begs comparison with that of a superhet radio, and so the system attracted the "heterodyne" label from the time of its first introduction. This description is now applied more or less indiscriminately to all lock-in amplifiers that incorporate one or more stages of frequency translation. It can be argued that these systems bear only a superficial resemblance to classical heterodyne systems and that the use of expressions such as "intermediate frequency" and "i.f. filter" in relation to lock-in amplifiers is likely to cause confusion, especially with those who have a clear understanding of the conventional usage of these terms. It is therefore necessary to introduce a note of caution about the terminology employed in this chapter which reflects the usage that is now prevalent among lock-in amplifier manufacturers and appears in data sheets and publicity material.

The Ithaco Dynatrac has since been matched by alternative and improved heterodyne lock-in amplifiers from the EG&G companies, PAR and Brookdeal. These lock-in amplifiers have greatly benefited from developments in technology relating to both reference channel and phase-sensitive detector design. In this chapter we shall be taking note of these developments and aiming to highlight areas of specification which are peculiar to heterodyne lock-in amplifiers.

Of particular interest in this respect will be the problem of identifying the spurious responses which occur when the frequency of an asynchronous signal lies close to a "critical" frequency. In a conventional system these critical frequencies correspond to the odd harmonics of the reference frequency. We shall show that *in principle* heterodyne systems can be designed to be inherently free

of these harmonic responses. There are nevertheless a number of additional critical frequencies, each with its related transmission window, which must be taken into account when a heterodyne model is adopted. A major objective of heterodyne system design is therefore to achieve a high degree of suppression of all unwanted responses, consistent with maintaining wideband, wide dynamic range performance.

Unfortunately, as we shall see, this last requirement cannot be achieved without sacrificing the total rejection of harmonic responses, which is inherent in an "ideal" heterodyne system. In our discussions we must therefore make a clear distinction between principles and practice and be prepared to examine the trade-offs which are necessary to produce heterodyne systems with good all-round performance.

Also included in this chapter is an appraisal of the synchronous heterodyne technique. Synchronous heterodyning offers a means of improving the dynamic range of phase-sensitive detectors and lock-in amplifiers. The technique can also be used to counteract the loss in dynamic range which occurs when an otherwise conventional phase-sensitive detector is operated towards the upper limit of its frequency range.

This last approach is used to obtain a competitive dynamic range specification in the EG&G Brookdeal heterodyne lock-in amplifiers where phase-sensitive detection is carried out at a fixed high frequency. We shall also be giving consideration to the spurious responses associated with synchronous heterodyning. The treatment given falls short of a full analysis, but it is shown how some of the major contributory factors can be overcome in practical systems.

8.2 Principles of heterodyne operation

The principles of heterodyne lock-in amplifiers can be established in terms of the idealized system shown in Fig. 8.1.

The frequency translator has inputs from the signal channel amplifier, from the applied reference signal and from an internal oscillator. The latter operates at frequency f_I which we shall call the intermediate frequency. The purpose of the frequency translator is to produce an output with magnitude proportional to the signal input but with its frequency shifted from f_s to a new and higher value $f_I + f_R - f_s$. The translated signal is then applied to a phase-sensitive detector which we assume is referenced and phase-shifted at the intermediate frequency.

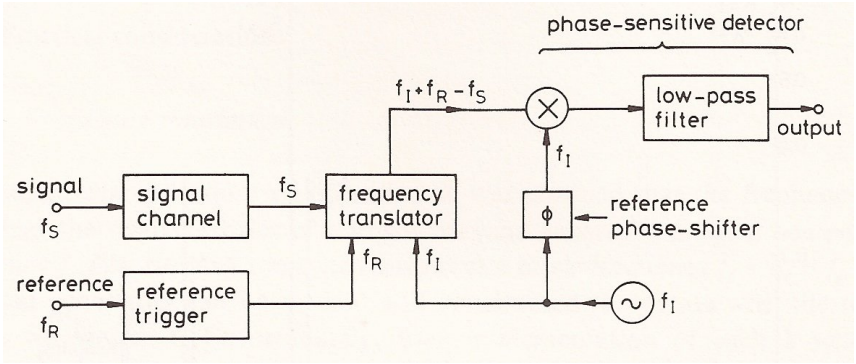

Fig. 8.1 Basic heterodyne lock-in amplifier, showing the frequencies at different points in the system

When the signal and reference are fully synchronous, $f_s = f_R$. The translated output then appears at frequency f_I and yields a classic phase-sensitive response at the phase-sensitive detector.

For signals with frequencies close to f_R the response will be an alternating output at frequency $|f_R - f_s|$, which is attenuated in the usual way by the output low-pass filter. However, in this system there is no scope for unwanted responses when the signal frequency is coincident with an odd harmonic of f_I. The latter are included on the assumption that a switching phase-sensitive detector is used with its associated harmonic transmission windows.

The critical signal frequencies are therefore those which satisfy the relationship:

$$|f_I + f_R - f_s| = Kf_I$$

where K is an odd integer. Solving for f_s, we obtain the critical frequencies:

$$f_s = |(1-K)f_I + f_R|$$

and,

$$f_s = (1+K)f_I + f_R$$

For each value of K there are therefore two frequencies where an interference component will be able to excite a harmonic response in the phase-sensitive detector. The relative sensitivity of the system to inputs at these critical frequencies is denoted by S_K and is given by

$$S_K = 1/K$$

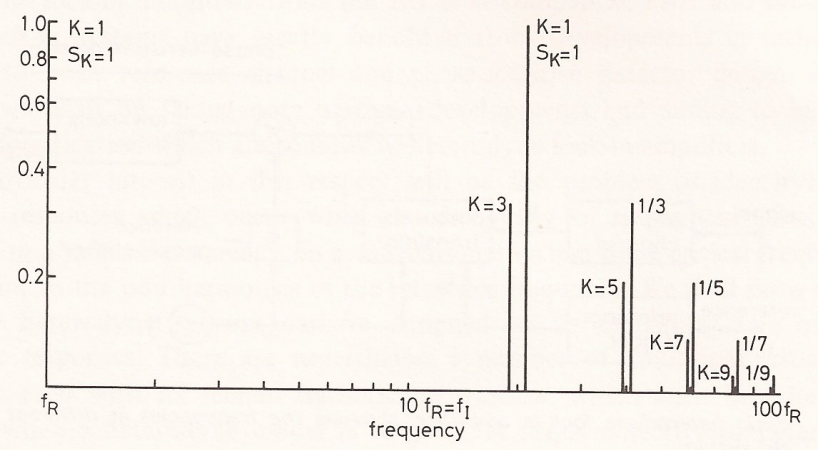

Fig. 8.2 Location and relative magnitude of transmission windows in a heterodyne system, for K = 1, 3, 5, 7 and 9. $f_I = 10f_R$

The transmission windows of the overall detection system are thus derived from the phase-sensitive detector windows and are located at frequencies determined jointly by the intermediate and reference frequencies. The transmission windows corresponding to K in the range 1 to 9 are drawn schematically in Fig 8.2 for the relationship, $f_I = 10 f_R$. The weighting factor assigned to each window is shown in each case and gives the relative sensitivity of the detection system to inputs at the appropriate critical frequency.

Fig. 8.2 reminds us that there must be two transmission windows, corresponding to $K = 1$, where the relative sensitivity is unity. These windows correspond to the "primary" response at $f_s = f_R$, and the "image" response given by signals with frequency $f_s = 2f_I = f_R$. An image response occurs in all systems operating on the heterodyne principle. In this case, a signal for which $f_s \approx 2f_I + f_R$ yields a translator output close to the intermediate frequency and the resulting behaviour of the detection system cannot then be distinguished from the "true" response

when $f_s = f_R$. The only satisfactory way to deal with the image response is to eliminate signal components at the image frequency by filtering. The most convenient arrangement is to use a low-pass filter in the signal channel with a sharp cut-off defined at a frequency below $2f_I + f_R$.

This filter must be introduced *before* the frequency translator and should have a cut-off frequency greater than f_{RMAX}, the maximum anticipated value of the reference frequency. An image filter is an essential component in a heterodyne lock-in amplifier, irrespective of the precise system configuration.

It turns out that in the present, ideal, case a properly designed image suppression filter would be effective in suppressing inputs at all other critical frequencies. This ideal heterodyne system is thus inherently free from responses at harmonics of the reference frequency and can be made relatively immune to the incidence of spurious responses at other, non-related, frequencies.

8.3 Practical considerations

8.3.1 Frequency translation

When discussing principles of operation it was assumed that the frequency translator had the characteristics of a single-sideband generator; a single component at frequency f_s gave rise to a translated output at a *single* frequency $f_I + f_R - f_s$. Single-sideband generators can be devised and constructed to operate with the required degree of precision. Unfortunately, the implementation of such a scheme to operate over a wide frequency range is both complex and expensive. A more cost-effective solution is to use double-sideband generation in which the required translation is achieved with a signal channel mixer and frequency synthesizer as shown in Fig. 8.3.

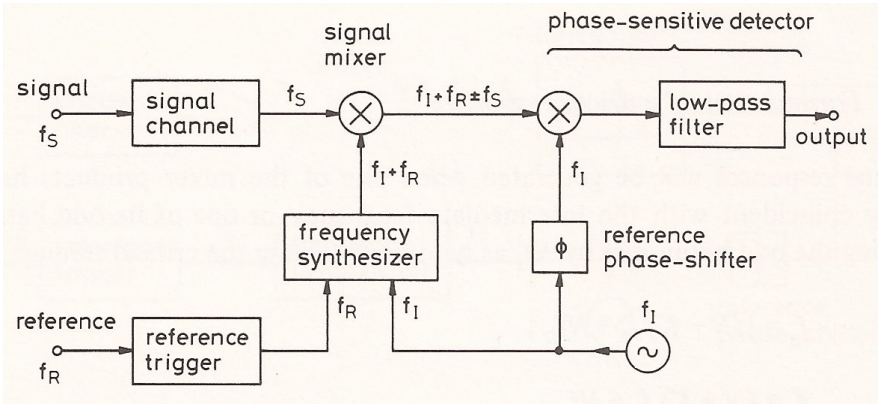

Fig. 8.3 Heterodyne system with double-sideband frequency translation

The frequency synthesizer provides an output at a precisely defined frequency $f_I + f_R$ which yields mixer products at frequencies $f_I + f_R \pm f_s$. When the signal and reference are fully synchronous, the mixer output will consist of two components, one at frequency f_I and the other at $f_I + 2f_R$. The first of these will give rise to a phase-sensitive response as before. The second gives rise to an alternating output at frequency $2f_R$. This is no more serious than the "ripple" component which is expected in conventional phase-sensitive detector operation. As is usually the case, the ripple component is rejected by the low-pass filter in the phase-sensitive detector output.

Turning now to asynchronous signals, we find an important consequence of using a double-sideband generator. This is a reduction in the allowable voltage swing due to asynchronous components at the input to the signal channel. Each of these

gives rise to two mixer products of equal amplitude. The input voltage swing at a given sensitivity is thus limited to half the value which could be sustained when using the phase-sensitive with a single-sideband frequency translator. For a phase-sensitive detector of given characteristics the achievable dynamic reserve is consequently reduced by 6 dB.

Further complications arise in a practical implementation of the frequency translator. The signal channel mixer, like the phase-sensitive detector, is almost invariably a switching multiplier which uses a *squarewave* drive at frequency $f_I + f_R$. This is another example where linearity and dynamic range are obtained through the adoption of a switching operation. The result is that the output of the signal mixer consists of a large number of components at frequencies

$$N(f_I + f_R) \pm f_s, \quad N \text{ odd}$$

The amplitude of each of these components is weighed by a factor $1/N$, reflecting the reduction in the harmonics of the switching waveform with increasing order.

As we shall see, the use of a switching mixer leads to a reappearance of the harmonic responses which were so successfully rejected by the ideal heterodyne scheme.

8.3.2 Formulation of spurious responses

Spurious responses will be generated when one of the mixer products has a frequency coincident with the intermediate frequency or one of its odd harmonics. Denoting the odd harmonics by Kf_I as before, we obtain the critical frequencies:

$$f_s = \left|(N - K)f_I + Nf_R\right|$$

and

$$f_s = (N + K)f_I + Nf_R$$

To calculate the relative sensitivity of the detection system to inputs at these critical frequencies, we must take the following factors into account: First of all, a factor $1/N$ resulting from the use of a squarewave drive to the mixer; secondly a factor $1/K$ to allow for the reduction of the phase-sensitive detector transmission windows with increasing harmonic order. The relative sensitivity at frequencies corresponding to given values of N and K is therefore:

$$S_{N,K} = 1/NK$$

For example, when $N = K = 1$ we obtain the primary response corresponding to $f_s = f_R$ and the image response corresponding to $f_s = 2f_I + f_R$. The relative sensitivity of the system to inputs at these frequencies is unity in both cases.

Let us now take $N = K \neq 1$. In this case, the system will be sensitive to inputs at $f_s = Nf_R$ and $2Nf_I + Nf_R$. The response to the Nth harmonic of f_R is thus reinstated at a relative sensitivity of $1/N^2$, compared to the figure of $1/N$ which would be obtained in a conventional system.

As noted in section 8.2 an image suppression filter would be essential in any practical system. Unfortunately this has no influence on signal components at frequencies Nf_R that fall within the filter bandwidth. The resulting harmonic responses can only be satisfactorily suppressed by using a second, tuned, filter in front of the phase-sensitive detector.

8.3.3 Suppression of the spurious responses

We envisage a system such as that shown in Fig. 8.4 which includes an image suppression filter in the signal channel and a tuned filter following the signal mixer.

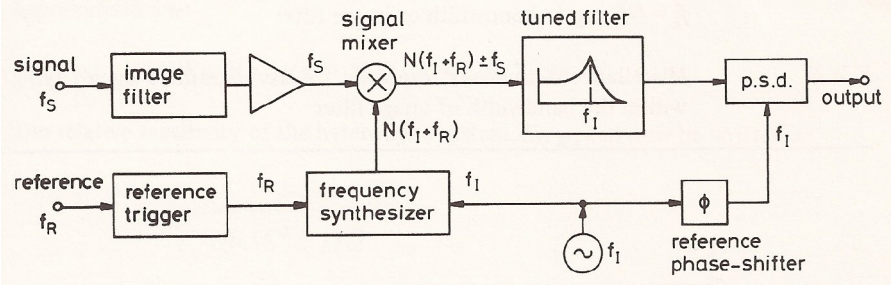

Fig. 8.4 Incorporation of tuned filter in a heterodyne system

The image suppression filter is a low-pass filter which exhibits a "flat" amplitude response for all signal frequencies up to the maximum value of the reference frequency as shown in Fig. 8.5.

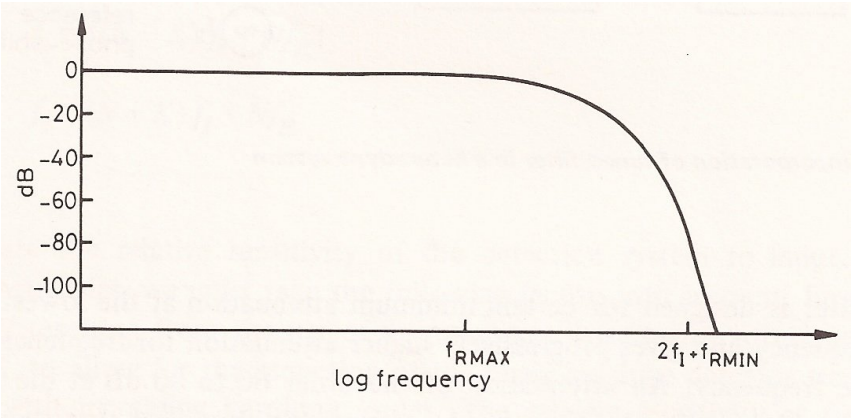

Fig. 8.5 General transmission characteristics of an image-suppression filter

The filter is designed for certain minimum attenuation at the lowest expected image frequency and gives progressively higher attenuation for frequencies beyond the image frequency. An attenuation of the order 60 to 80 dB at the image frequency would be typical of practical systems. We thus find that asynchronous signals characterized by frequencies:

$$f_s = (N+K)f_I + Nf_R; \quad N, K \geq 1$$

all lie beyond the cut-off of the image filter and are effectively eliminated before the signal mixer.

The only asynchronous signals which are likely to yield discernable responses must therefore have frequencies which satisfy the condition:

$$f_s = \left|(N-K)f_I + Nf_R\right|$$

The nature of the responses corresponding to different values of *N* and *K* are summarized in Table 8.1.

Table 8.1 Catalogue of responses for asynchronous signals with frequencies $|(N-K)f_I + Nf_R|$

N, K	Comments
$N = K = 1$	$f_s = f_R$: primary respoonse of system
$N > K$	Asynchronous signals always have frequencies $\geq$ image frequency
$N > K \neq 1$	Spurious responses whenever asynchronous signal has frequency $f_s = Nf_R$ up to bandwidth of image filter
$K > N$	Miscellaneous spurious responses for asynchronous signals lying within the bandwidth of image filter

The first of the categories listed in Table 8.1 corresponds to the "wanted" response of the system while signals in the second category are suppressed by the image filter. In the absence of a tuned filter, the system would display a relative sensitivity $S_{N,K} = 1/NK$ at all critical frequencies in the final two categories. The effect of a tuned filter will be to introduce additional attenuation of these responses giving a relative sensitivity:

$$S_{N,K} = \left|H\left(jK\omega_I\right)\right| / NK$$

Here, $H(j\omega)$ is the frequency response of the tuned filter normalized to a magnitude of unity at the intermediate frequency.

We note that the largest spurious response will be obtained when $N = K = 3$, corresponding to an asynchronous signal with frequency $3f_R$*. The tuned filter should therefore be set up to reduce this response to an acceptably small value.

8.3.4 Tuned filter requirements

In all commercial systems the tuned filter is of the low-pass type described in Appendix 4. The use of a tuned filter for suppressing the harmonic responses of a phase-sensitive detector was discussed in Section 4.5.2. For a low-pass filter tuned to the intermediate frequency f_I, the attenuation at frequency Kf_I in the high-Q approximation is:

$$\left|H\left(jK\omega_I\right)\right| = \frac{1}{(K^2-1)Q}, \quad K = 3,\ 5,\ 7,\ \mathrm{K}$$

The relative sensitivity of the heterodyne system, $S_{N,K}$, can now be written as

$$S_{N,K} = \frac{1}{NK(K^2-1)Q}$$

The value of the Q-factor required for a given level of suppression can thus be calculated.

For example, suppose the system is required to have a relative sensitivity of 10^{-4} (–80 dB) to a signal with frequency close to $3f_R$. From Table 8.1 we put $N = K = 3$ and calculate the required Q-factor of the tuned filter:

---

* Note that the response due to $K = 3$, $N = 1$ corresponds to an asynchronous signal at frequency $2f_I - f_R$.This would normally be heavily attenuated by the image filter.

$$Q = 10^4 / (3 \times 3 \times 8) = 139$$

From the discussion given in Section 4.5.1 it is clear that operation with such a high value of *Q*-factor places severe demands on the system with regard to alignment and maintaining good phase accuracy. This is all the more troublesome in a purpose-built system where the filter is not usually accessible for routine realignment by the user.

An alternative approach which greatly eases alignment problems and which places less demands on filter performance has been used in the EG&G Brookdeal heterodyne system. This is to use two filters of relatively low *Q*-factor in cascade. The attenuation introduced by the filter stage then becomes:

$$\left|H\left(\mathrm{j}K\omega_{\mathrm{I}}\right)\right| = \frac{1}{(K^2-1)^2 Q^2}$$

In this case, a *Q*-factor of 5 is sufficient to give a relative sensitivity of –83 dB to signals with frequency $f_s = 3 f_R$.

Of course, these figures for harmonic rejection are strictly theoretical. There is a considerable technical hurdle to be overcome in order to realize these figures in practice and commercial systems usually specify suppression factors of around –60 dB (×1/1000). In any specific case, it is always worth checking whether the figure given refers solely to third-harmonic rejection or to the maximum level of spurious responses arising from all possible sources.

8.3.5 Phase-shifting

It has been assumed so far that the user-controlled phase-shifter which is essential for phase-sensitive detection is introduced at the intermediate frequency, on the "reference" side of the phase-sensitive detector. This has a distinct advantage in that the phase-shifter can be designed to operate at a fixed frequency rather than over a wide range of frequencies as is usually the case.

In principle, there is no reason why phase-shifting should not be carried out at the original reference frequency or, indeed, on the output of the frequency synthesizer, at frequency $f_I + f_R$. These are all found to be equivalent when system operation is analysed. Phase-shifting at the reference frequency requires similar broadband circuitry to that found in a conventional lock-in amplifier, whereas a phase-shifter placed at the output of the frequency synthesizer, assuming $f_I >> f_{RMAX}$, would require relatively narrowband capability.

8.4 Practical limitations

8.4.1 The frequency synthesizer

The generation of a waveform at a precisely defined frequency $f_I + f_R$ in response to an external signal applied at frequency f_R is the most demanding task facing the designers of heterodyne lock-in systems. The ultimate objective is to produce a system which:

(i) Maximises the range of reference frequencies which can be handled at a particular intermediate frequency.

(ii) Has an acceptably low level of both random and discrete phase-noise over the specified frequency range.

(iii) Minimizes phase errors between the absolute phases of the synthesized waveform and the applied reference waveform.

(iv) Has an acceptable value of reference slew rate and a small acquisition time.

A shortfall in any of these areas would be noticed by any user whose interest in lock-in amplifiers extended beyond the detection of noisy signals at a fixed frequency. The designer's difficulty is to reconcile these requirements and make the correct trade-offs to produce a system with all-round acceptable performance. The problems encountered in reaching an acceptable compromise in synthesizer performance can be highlighted by referring to the system illustrated in Fig. 8.6.

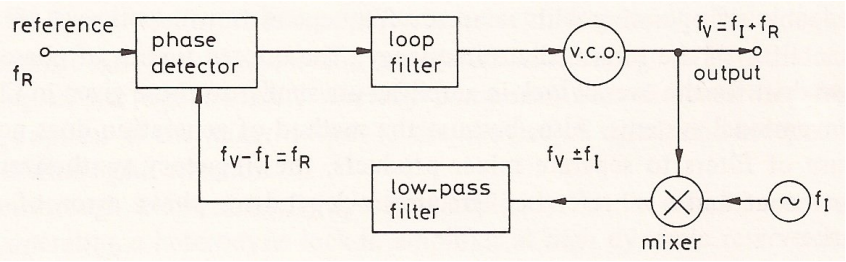

Fig. 8.6 Heterodyning phase-locked loop

This example of a "heterodyning" phase-locked loop received considerable publicity when the Ithaco Dynatrac lock-in amplifiers were first introduced[1].The phase detector compares the phase of its two inputs and its output is amplified and smoothed by the loop filter. The v.c.o. is controlled from the loop filter output and produces an output at frequency f_V, which is mixed with a signal at the intermediate frequency f_I provided from a stable sinewave oscillator. The purpose of the low-pass filter is to transmit only the low-frequency mixer product at frequency $f_V - f_I$ to the phase detector. The difference frequency $f_V - f_I$ is thus phase-locked to the incoming reference signal at frequency f_R, making f_V equal to $f_I + f_R$.

When a strictly conventional approach is taken to designing the phase-locked loop, following the treatment given in Chapter 7, the designer is faced with a difficult decision in choosing the loop characteristics. The loop bandwidth should be small enough to suppress discrete phase modulations resulting from unwanted mixer components at the output of the low-pass filter, and yet wide enough to accommodate the phase noise inherent in the v.c.o. and to give adequate slew rate performance. In addition, severe phase errors might be incurred when the low-pass filter achieves the desired separation of mixer products by virtue of a sharp cut-off. Almost inevitably, the trade-offs are such that acceptable performance can only be achieved over a relatively small range of reference frequencies. Limitations are observed in terms of poor phase tracking and excessive phase noise at the extremes of the operating frequency range.

In the loop used by EG&G Brookdeal, switching waveforms derived from the applied reference and from the intermediate-frequency oscillator are combined in a multiplexer. The output is a switching waveform in which the positive transitions occur an average rate $(f_I + f_R)/4$. This entire waveform is then used to phase-lock a v.c.o. in a phase-locked loop which incorporates a ÷4 counter. In a conventional phase-locked loop, the output from the v.c.o would be a squarewave at frequency $f_I + f_R$ subjected to a high level of discrete phase-modulation which could only be suppressed by using a loop of small bandwidth. In the EG&G Brookdeal system the unwanted modulation is well defined and can be suppressed to a high order by adding a compensation signal to the output of the phase detector. Provided the compensation signal is accurately generated and controlled, the constraints on loop bandwidth are considerably relaxed. The loop characteristics can then be optimized with respect to a low incidence of phase noise and obtaining acceptable handling characteristics over a wide range of reference frequencies. The system, in fact, proves capable of operating with

reference frequencies from less than 1 Hz up to about one fifth of the intermediate frequency. The quoted figures of phase noise and phase drift for the overall lock-in amplifier are similar to those given in Chapter 4 for conventional systems. Also, because the method of generation does not rely on the use of filters to separate mixer products, the frequency synthesizer itself does not contribute a reference frequency-dependent phase error of major significance.

8.4.2 The image filter

It has been noted that the image filter should have a flat response up to signal frequencies corresponding to f_{RMAX}. The rate of cut-off beyond this point must then be extremely large in order to introduce adequate attenuation at the image frequency. In most commercial systems the ratio f_I / f_{RMAX} is less than 10, so that a filter of high order, 4 to 6 pole, is required to achieve the necessary roll-off.

Like all filters used for signal conditioning, the image filter introduces phase errors into the measurement system. When the filter is of high order, the phase error can be in excess of 100° at about one half of the cut-off frequency. A fortunate consequence of using high-order filters is that the amplitude response can be made uniformly flat up to frequencies very close to cut off. In addition, the phase shift within this range can be made proportional to frequency. When the filter approximates to such a linear-phase model, the phase-shift of the signal channel can be compensated by introducing a suitable time delay in the reference channel. In the Ithaco Dynatrac, the image-filter phase characteristic was actually compensated by the characteristic of the low-pass filter in the synthesizer phase-locked loop. In other systems, such as that from EG&G Brookdeal, the synthesizer has inherently low phase error and the compensating time delay need be no more complicated than a monostable circuit operating at a fixed pulse width.

When we take matching constraints into account and add uncompensated phase errors accrued in the intermediate-frequency filter and in the synthesizer, we conclude that the overall phase precision of heterodyne systems must fall short of that obtainable in a conventional lock-in amplifier operating in the same reference frequency range.

8.4.3 The signal mixer

The most serious performance limitations associated with the signal mixer are due to non-linearity and "feedthrough". It has been stressed elsewhere that the synchronous demodulation process should be supported by linearity in all preceding stages; hence the linearity of the signal mixer should be at least of the same order as the linearity of the phase-sensitive detector. This requirement is eased in practice because the two components, being similar in concept, draw upon similar technologies.

"Feedthrough" is a phenomenon associated with voltage offsets and capacitive coupling in the signal mixer, whereby components at the switching frequency, $f_I + f_R$, appear at the mixer output in the absence of signal, and independently of signal channel gain selection.

Feedthrough is not normally specified explicitly but its effect becomes evident when operating a heterodyne lock-in amplifier at high dynamic reserve. It appears as an alternating component at frequency f_R in the output of the system and appears at its worst when operating at low reference frequencies with a relatively short time constant selected.

A level of feedthrough of 10^{-5} (–100dB) represents a reasonable target for a signal channel mixer operating with an intermediate frequency in excess of 100 kHz. When used in conjunction with a phase-sensitive detector having a

dynamic reserve of $\times 10^4$ (80 dB) the output a.c. variation at frequency f_R could have a maximum value of 1/10 (–20 dB) of full scale. Unless the intermediate-frequency tuned filter has unusually high *Q*-factor - and correspondingly narrow bandwidth - it is unlikely to have much influence on the level of feedthrough reaching the phase-sensitive detector.

8.4.4 The phase-sensitive detector

It was noted in Chapter 3 that the dynamic range of switching phase-sensitive detectors is reduced at high reference frequencies owing to the appearance of the so-called "h.f. offset". The mechanisms at work here are very similar to those which give rise to feedthrough in switching mixers. It is claimed, quite reasonably, that in heterodyne lock-in amplifiers operating at a fixed value of intermediate frequency, the dynamic range is constant over the entire range of signal and reference frequencies. It should also be said, however, that the phase-sensitive detector is fated always to operate at a relatively high frequency where its dynamic range is less than optimum. In the Ithaco system, operation in a given frequency range involved selecting plug-in circuit cards to provide a suitably high value of intermediate frequency. Not surprisingly, the figures for dynamic range showed a marked deterioration as the system was configured to operate at higher reference frequencies and correspondingly higher values of intermediate frequency.

In the EG&G Bookdeal heterodyne lock-in amplifier the maximum intermediate frequency is 1 MHz, which allows operation with reference frequencies over the entire audio-frequency range up to 200 kHz. In this case, the dynamic range of the phase-sensitive detector is maintained at a competitively high value (>120 dB) by a technique known as *synchronous heterodyning*. It is shown in Section 8.8 that this approach can be used to improve the dynamic range of phase-sensitive detectors operating in either a conventional lock-in amplifier or in a heterodyne system. Synchronous heterodyning is not an alternative to the heterodyne mode of operation described so far; rather it is a supplementary technique applied with the objective of improving dynamic range.

8.5 Overload capability of heterodyne systems

It was the advent of the Ithaco Dynatrac that focused popular attention on the specification of overload capability in lock-in amplifiers. Clearly, a simple statement of phase-sensitive detector dynamic reserve is not sufficient when a lock-in system is so heavily supported by filters.

From the point of view of overload capability, the image suppression filter in any heterodyne lock-in amplifier can be regarded as a low-pass signal conditioning filter. The rate of cut-off of the image filter is extremely high and this leads to a dramatic improvement in overload capability for asynchronous signals having frequencies greater than about one half the intermediate frequency.

The selectivity of the tuned intermediate-frequency filter is chosen solely on the grounds of suppressing harmonic responses. Claims that the tuned filter is responsible for a significant increase in overload capability should therefore be treated with caution, particularly when the reference frequency is very low compared with the intermediate frequency. In a heterodyne system the overload characteristics are similar to those of a conventional lock-in amplifier using a tuned filter at the reference frequency with a bandwidth equal to that of the intermediate frequency filter, f_I / Q. The *effective Q*-factor is thus dependent on the reference frequency and is given by $Q_{\text{eff}} = Q \times f_R / f_I$. As a result, the filter is likely to have a significant effect on overload capability only when the reference frequency approaches its maximum value; even then the maximum reference

frequency f_{RMAX} should represent a substantial fraction of f_I. Moreover, the *Q*-factor of the tuned filter itself should also be high; if sharp cut-off is obtained by cascading low-*Q* sections as described in Section 8.3.4. the overload characteristics will be broadly independent of frequency within the bandwidth of the image-suppression filter.

The interference rejection characteristics plotted in Fig. 8.7 serve to emphasize these points. The graphs are drawn for a system in which $f_I = 5 f_{RMAX}$ and the intermediate-frequency tuned filter has a *Q*-factor of 20. These conditions approximate to those it the Ithaco Dynatrac lock-in amplifier. In a system such as

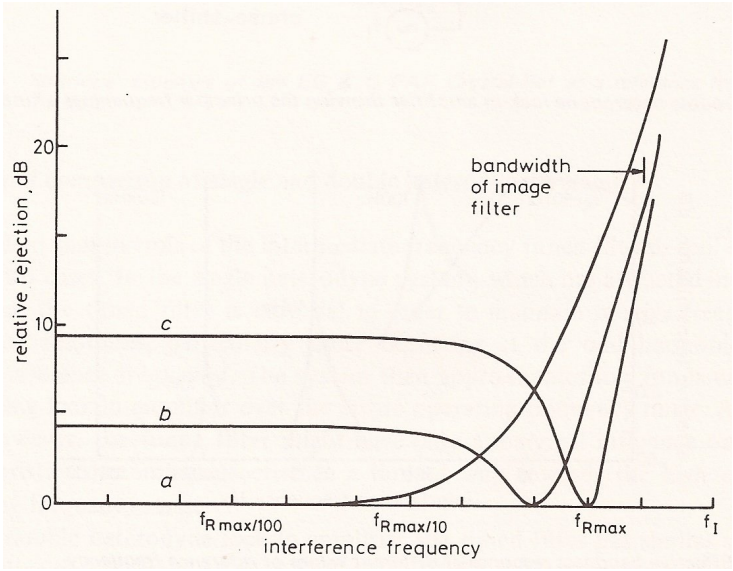

Fig. 8.7 Interference rejection in a heterodyne lock-in amplifier (a) $f_R = f_{RMAX}/10$; (b) $f_R = f_{RMAX}/2$; (c) $f_R = f_{RMAX}$

the EG&G Brookdeal heterodyne, the reference frequency is allowed to take values up to 6 decades below the intermediate frequency. Rejection characteristics of type (a) in Fig. 8.7 are thus applicable over most of the operating frequency range.

8.6 Double heterodyne lock-in amplifiers

In the heterodyne system described so far, the final detection has been carried out at the intermediate frequency. In a double heterodyne detection is carried out at the original reference frequency as shown in Fig. 8.8.

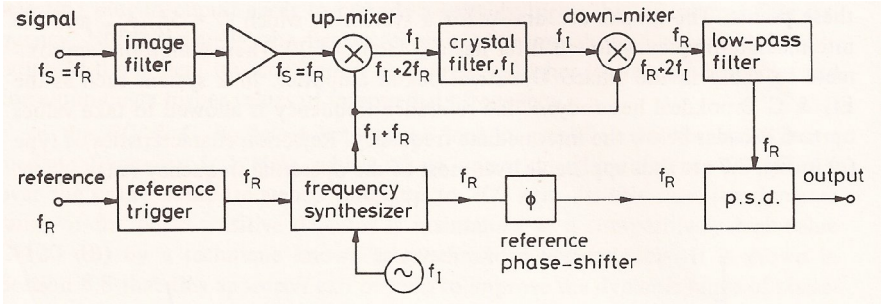

Fig. 8.8 Double-heterodyne lock-in amplifier showing the principal frequencies of interest

The intermediate frequency stage incorporates a filter of extremely high selectivity. In the case of the EG&G PAR Crystal-Het lock-in amplifier a two-section crystal filter is used to define a *Q*-factor of 50 000 at the intermediate frequency of 1MHz. The resulting 20 Hz bandwidth is significantly smaller than the reference frequency over much of the 100kHz frequency range, and thus has a profound effect in removing harmonically related components and noise before detection.

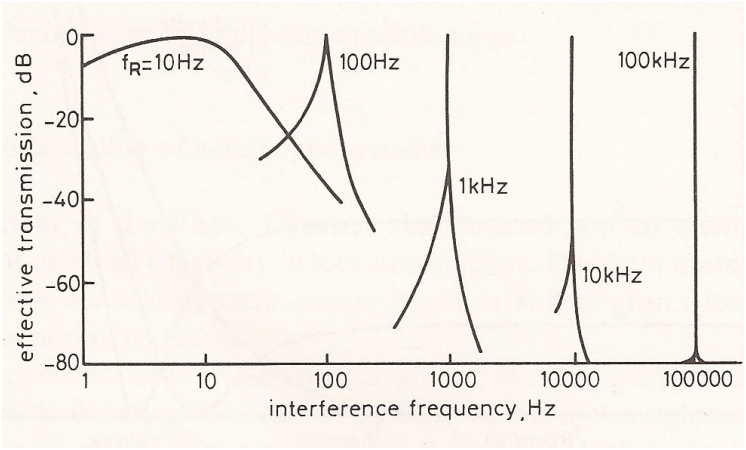

Fig 8.9 Effective bandpass response at different values of reference frequency

The system can be thought of in terms of a tracking filter with a fixed bandwidth of 20 Hz, giving an effective *Q*-factor, $Q_{\text{eff}} = 50\,000 \times f_R / f_I$. The effectiveness of the system in rejecting interference components is illustrated in Fig. 8.9 which is drawn for different reference frequencies.

As might be expected, the phase-tracking of such a narrowband system is greatly inferior to that of a conventional lock-in amplifier. The level of spurious responses is nevertheless very low at midband. This is evident from the results given in Fig. 8.10 for operation at a reference frequency of 1kHz. The largest spurious response in this case occurs at a level of – 70dB relative to the "primary" response at 1kHz.

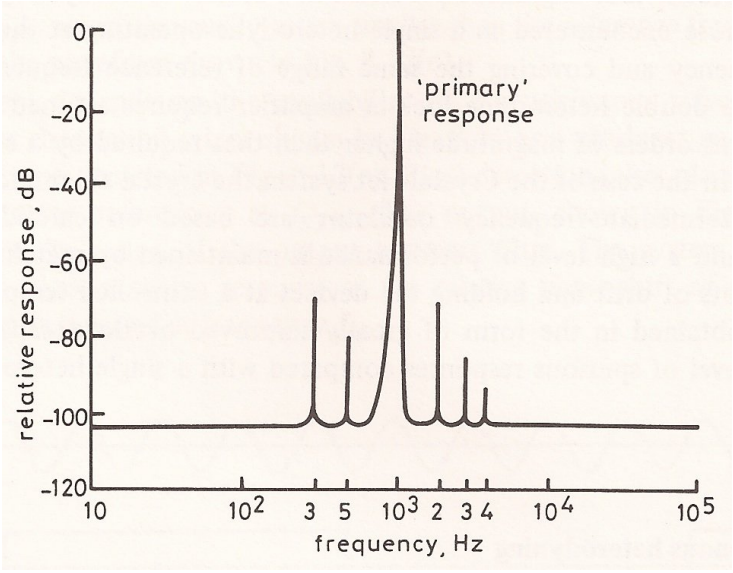

Fig. 8.10 Spurious responses of the EG&G PAR Crystal-Het at a reference frequency of 1kHz

8.7 Brief comparison of single and double heterodyne systems

It is evident that the role of the intermediate frequency tuned filter is quite different in the two cases. In the single heterodyne system, which has attracted most of our attention, the tuned filter is essential in order to maintain relative freedom from unwanted responses, particularly those occurring at the odd harmonics of the applied reference frequency. The system then approximates to a fundamental-only responding lock-in amplifier over the entire operating frequency range. As we have seen, however, the tuned filter might have only a marginal influence on overload characteristics, becoming effective in a limited way towards the high end of the operating frequency range.

In a double heterodyne lock-in amplifier the tuned signal has similar status to a tuned filter incorporated in the signal channel of a conventional lock-in amplifier. The considerations of Section 4.4 thus apply provided that allowance is made for the variation of the effective *Q*-factor with operating frequency. The essential difference between the two heterodyne schemes is that, in the double heterodyne system, harmonic rejection depends on the *effective Q*-factor; in a single hetrodyne, harmonic rejection depends only on the actual *Q*-factor of the tuned filter measured at the intermediate frequency. We can conclude from this that the double heterodyne approximates to a fundamental-only responding system only for reference frequencies greater than a specified value. In the case of the EG&G PAR Crystal-Het system referred to earlier, this minimum value of reference frequency must certainly be greater than 20 Hz, which limits the system's effectiveness in the critical low-frequency regime identified in Chapter 6.

It is also evident that alignment problems in a double heterodyne system are far worse than those encountered in a single heterodyne operating at the same intermediate frequency and covering the same range of reference frequencies. As we have seen, the double heterodyne lock-in amplifier requires a tuned filter with a *Q*-factor several orders of magnitude higher than that required by a single heterodyne system. In the case of the Crystal-Het system the crystal filter and the crystal-controlled intermediate-frequency oscillator are based on carefully matched components and a high level of performance is maintained by matching temperature coefficients of drift and holding the devices at a controlled temperature. The benefits are obtained in the form of greatly improved overload capability and a much lower level of spurious responses compared with a single heterodyne lock-in amplifier.

8.8 Synchronous heterodyning

8.8.1 Introduction

In the mid 1970s EG&G PAR introduced a lock-in amplifier known as the Syncro-Het which offered a considerable improvement in dynamic range over conventional lock-in amplifiers. The system enjoyed success in its own right, but attracted little in the way of direct competition. It appears that no attempt was made to refine the basic technique or to develop further commercial instruments operating on the synchronous heterodyne principle. This was the situation until recently when synchronous heterodyning was revived in the context of high-frequency lock-in systems. The objective was stated in Section 8.4.4; namely to improve the dynamic range of phase-sensitive detectors operating at a high frequency.

It was remarked in Section 3.6.5 that the d.c. response to a synchronous signal can be separated from offset voltages at the output of a phase-sensitive detector by introducing a phase reversal of 180° in the reference channel. This approach to

signal detection forms the basis of the synchronous heterodyne system illustrated in Fig. 8.11. Here, the phase reversal is introduced by systematically switching the signal with a squarewave taking values +1 and –1. The switching is carried out at a frequency f_{SYN}. Final detection takes place in the phase-sensitive detector referenced at f_{SYN}. The output, shown in Fig. 8.12(f), is smoothed by the action of the output low-pass filter. The system thus yields a conventional phase-sensitive response as the relative phase-shift of the signal and reference inputs is varied at frequency f_R.

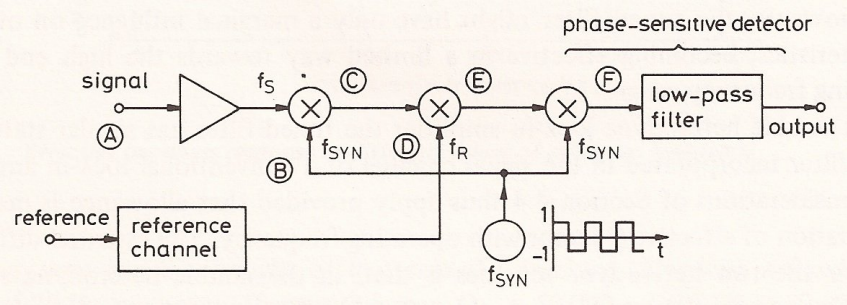

Fig. 8.11 Principles of synchronous heterodyning

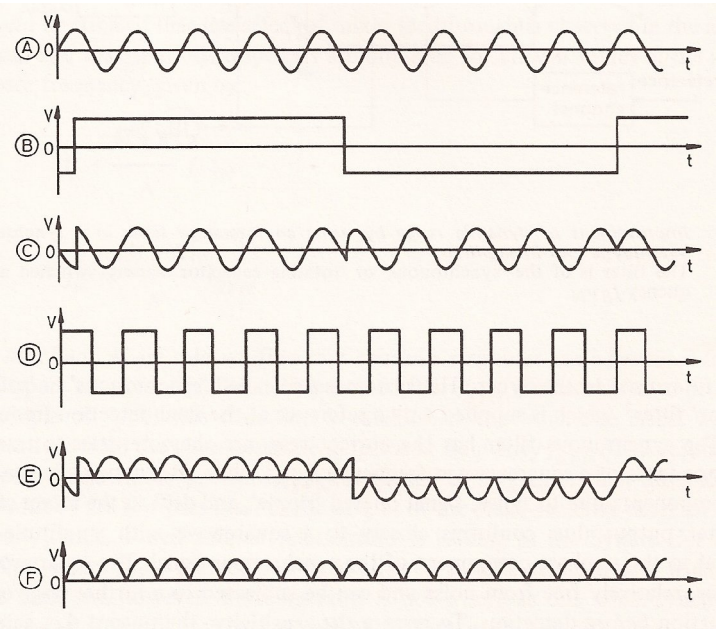

Fig. 8.12 Waveforms in the system of Fig. 8.11

In this mode of operation we find that the phase-sensitive detector can be constrained to operate in a frequency range well below the actual value of the reference frequency f_R. For example, phase-sensitive detection of a signal at 100 kHz could be achieved with a "syncrohet" frequency, f_{SYN}, of 100 Hz. The overall dynamic range would then be equivalent to that obtainable when operating the phase-sensitive detector in a conventional fashion at 100 Hz rather than the higher value of 100 kHz. It is this aspect of operation that has made the synchronous heterodyne approach so attractive in heterodyne systems where the phase-sensitive detector is referenced at a high value of intermediate frequency.

8.8.2 Dynamic range improvement

The synchronous heterodyne system described so far offers no inherent improvement in dynamic range at low reference frequencies. This was obtained

in the EG&G PAR Syncro-Het system by imposing a filter between the final mixer and the phase-sensitive detector as shown in Fig. 8.13.

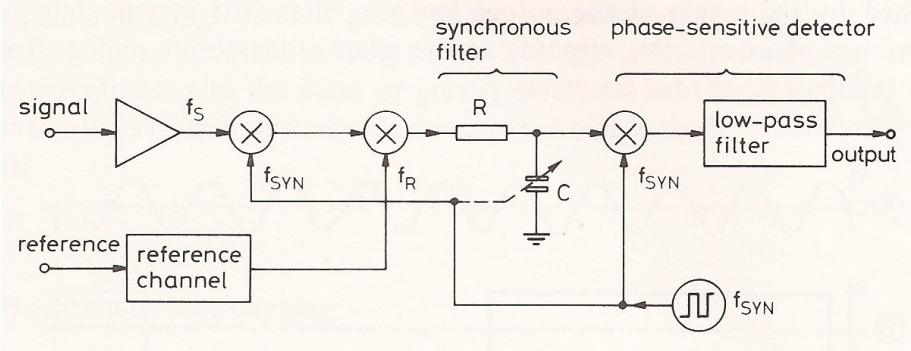

Fig 8.13 Improvement of dynamic range by using an interstage filter in a synchronous heterodyne lock-in amplifier. The filter is of the "synchronous" or "rotating capacitor" variety switched at frequency f_{SYN}

The filter used in the Syncro-Het system is a so-called "synchronous" or "rotating capacitor" filter[2] which is supplied with a reference at the final detection frequency f_{SYN}. The synchronous filter has the correct response characteristics to transmit the components of a squarewave at frequency f_{SYN}, while attenuating all asynchronous components due to noise, signal related "ripple", and drift in the mixer stages. The filter output thus conforms closely to a squarewave with a amplitude proportional to the in-phase component of the synchronous signal. The squarewave is, moreover, relatively free from noise and can be subjected to a further stage of a.c. amplification *before* detection. To restore the sensitivity, the output d.c. gain can then be reduced in proportion to the extra a.c. gain supplied in the signal path. The result is a system which offers improved output stability at a given level of dynamic reserve. In the light of the definitions given in Chapter 3, this is equivalent to a system with improved dynamic range.

The Syncro-Het lock-in amplifier operated with f_{SYN} = 11 Hz and proved capable of ×3000 (70 dB) dynamic reserve, consistent with an output stability of 10 p.p.m. This is equivalent to an input dynamic range of a massive 3.10^8 (170 dB). Unfortunately, the overall performance was marred by a series of spurious low-frequency outputs which occurred for critical combinations of the applied reference frequency and the internal "syncrohet" frequency f_{SYN}.

Mixer feedthrough, giving rise to a residual mixer output in the absence of signal, is a major cause of these low-frequency "beat" products. In general, the residual output from the second mixer will comprise components at combinations of f_R, f_{SYN} and their odd harmonics, characterized by frequencies:

$$\left| Mf_{SYN} \pm Nf_R \right| ; \quad M, N \text{ odd}$$

The phase-sensitive detector is referenced at frequency f_{SYN} and is responsive to inputs at frequencies $K f_{SYN}$ where K is an odd integer. Spurious responses will occur whenever f_R takes values such that:

$$\left| Mf_{SYN} \pm Nf_R \right| = Kf_{SYN}$$

It should be stressed that the effect of mixer feedthrough is observed in the absence of signal and that spurious responses are obtained for critical values of the applied *reference* frequency, given by:

$$f_R = \frac{M+K}{N} f_{SYN}$$

and

$$f_R = \frac{|M-K|}{N} f_{SYN}$$

Since *M, N* and *K* are odd we find that spurious responses can be avoided (at least with respect to mixer feedthrough) provided that f_{SYN} is constrained to be an odd submultiple of f_R. If this condition is established, the worse-case effect will be a beat product from the phase-sensitive detector at frequency f_{SYN}. This response will not be discernible provided f_{SYN} is chosen to be greater than the maximum bandwidth of the low-pass filter in the output of the phase-sensitive detector.

Clearly, such a well-defined relationship is lacking in the Syncro-Het lock-in amplifier, where f_{SYN} is fixed and f_R is allowed to take values over a wide range. As we shall see, however, the constraint is not a serious one when synchronous heterodyning is applied to a heterodyne lock-in amplifier operating at a fixed intermediate frequency.

8.8.3 Application to heterodyne lock-in amplifiers

The implementation of synchronous heterodyning in the EG&G Brookdeal heterodyne lock-in amplifiers follows the block diagram shown in Fig. 8.14(a). The phase sensitive detector operates at a frequency f_{SYN} which is well below the highest intermediate frequency of 1MHz. The complexities of interstage filtering are avoided, and further reduction in complexity is achieved by using only one mixer as opposed to the double mixer stage assumed earlier.

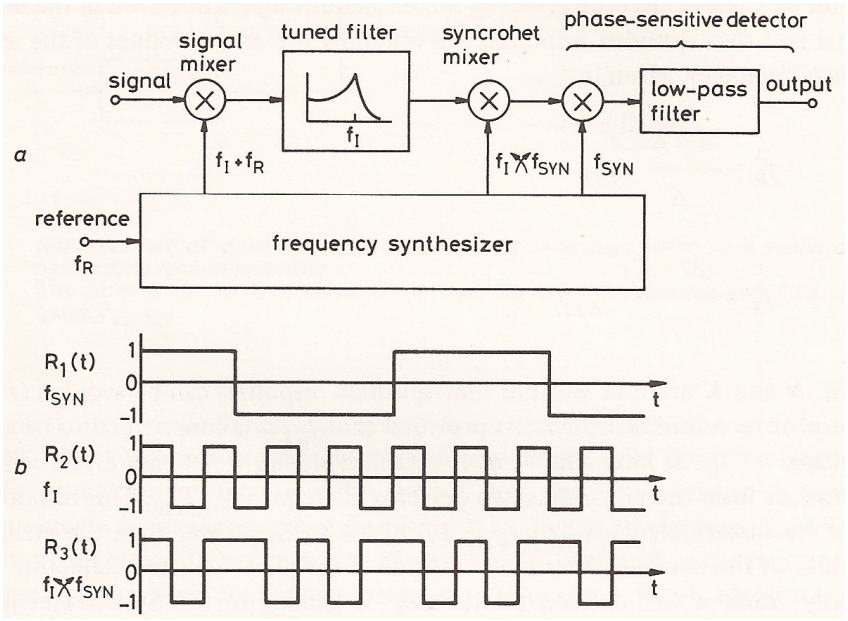

Fig. 8.14 **(a) Incorporation of synchronous heterodyning in a heterodyne lock-in amplifier. Reference processing and provision for phase-shifting are omitted for clarity**
(b) Generation of the switching input to the syncrohet mixer f_{SYN} is obtained by odd-integer frequency division from f_I to minimize spurious responses. In practice, f_{SYN} will be several orders of magnitude less than the highest value of intermediate frequency

In our original description of synchronous heterodyning, the signal was mixed successively with signals at frequencies f_{SYN} and f_R. For a signal $s(t)$, the output from the second mixer has the form:

$$s(t)\, R_1(t)\, R_2(t)$$

where $R_1(t)$ and $R_2(t)$ are squarewaves at f_{SYN} and f_R.

The alternative approach, using only one mixer, is to multiply the signal directly by a two-state waveform:

$$R_3(t) = R_1(t)\, R_2(t)$$

In practice, $R_3(t)$ can be formed by combining the individual switching waveforms $R_1(t)$ and $R_2(t)$ in an exclusive-NOR circuit. The resulting waveforms are derived in Fig. 8.14(b). In the EG&G Brookdeal heterodyne lock-in amplifier, f_{SYN} is chosen to be about 1 kHz, to lie beyond the bandwidth of the output filter, and is generated by odd-integer frequency division from the intermediate frequency oscillator. As explained in the last section this minimises the incidence of spurious outputs resulting from feedthrough in the syncrohet mixer. The method of generation avoids the coincidence of transitions in the switching waveforms $R_1(t)$ and $R_3(t)$ which is, in itself, a source of spurious outputs in systems where f_{SYN} takes arbitrary values.

8.9 Conclusions

This chapter has shown how a simple idea, that of frequency-shifting to avoid harmonic responses, has been progressively modified to allow for practical limitations in the performance of the various subsystems which comprise a heterodyne lock-in amplifier. The result is an instrument operating at a level of complexity greatly in excess of a conventional broadband lock-in amplifier and which falls short of conventional systems in several important specification areas, notably dynamic range and phase accuracy.

These drawbacks, to judge from the popularity of heterodyne systems, are more than offset by the advantages in operating with fundamental-only response allied to a low level of spurious responses. In terms of all-round performance, therefore, heterodyne systems confirm most closely with the characteristics demanded of a "general purpose" measurement tool capable of making unambiguous measurements with a variety of signal types over a wide range of operating frequencies. The availability of these instruments with a comprehensive facility for digital control adds to their appeal in a wide range of applications. Some of the characteristics of these systems are reviewed in Chapter 10, while a comparison with p.w.m. systems, which offer an alternative approach to achieving fundamental-only response, is given at the end of Chapter 9.

8.10 References

1 MUNROE, D.M. (1973): "The heterodyning lock-in amplifier". Technical Bulletin, Ithaco Corp., Ithaca, NY.

2 FRANKS, L.E. and SANDBERG, I.W. (1960): "An alternative approach to the realization of network transfer functions: The N-path filter", *Bell Syst. Tech. J.*, 39, pp. 1321-1350.

CHAPTER 9

P.W.M systems

9.1 Introduction

Single and double heterodyne systems were developed to overcome the problem of harmonic responses in lock-in detection while retaining the ability to track signals over a wide frequency range. In this chapter, we shall be looking at an alternative method of suppressing harmonic responses, by using pulse-width modulation in the reference channel.

It will become apparent that heterodyne and pulse-width modulation systems have rather different status. Whereas heterodyne lock-in amplifiers have a system configuration which is far removed from the "conventional" arrangement, the pulse-width modulation approach requires only a relatively simple modification to an otherwise basic system. As a result, the pulse-width modulation, or p.w.m., circuitry can be supplied as an option to a conventional lock-in amplifier, leaving the user to select either conventional response or fundamental-only response as required. As in the case of heterodyne systems, the suppression of harmonic responses is achieved at the expense of spurious responses at apparently arbitrary frequencies and is attended by a loss of dynamic range. We shall therefore be examining some of the trade-offs which must be made to achieve a detection system with overall acceptable characteristics.

P.W.M. lock-in amplifiers are characterized by phase accuracy of a very high order and represent a solution to suppressing harmonic responses which is both cost-effective and flexible. Other benefits, such as the potential to operate with greatly improved slew rate and the ability to operate the phase sensitive detector as an "ideal" multiplier, will also be noted.

A brief comparison of p.w.m. systems with heterodyne lock-in amplifiers is given at the end of this chapter.

9.2 Principles of operation

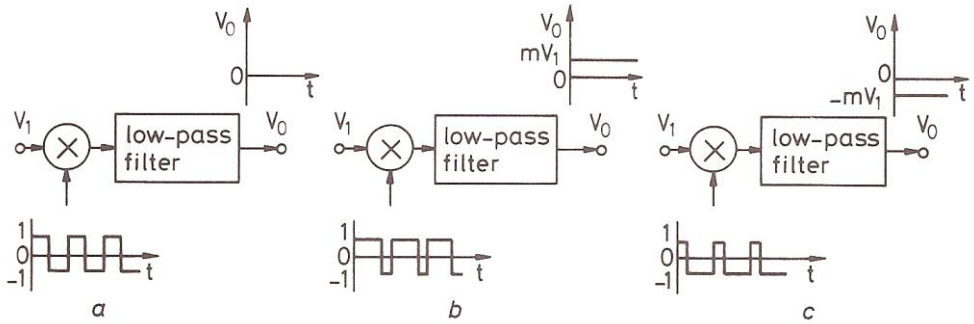

Fig. 9.1 Effect of reference waveform symmetry on the average gain of a phase-sensitive detector shown for the signal at a fixed level, V_1 volts. The average gain is (a) zero; (b) +*m*; (c) –*m*

The harmonic responses of a phase-sensitive detector result from the abrupt change of gain between +1 and –1 which occurs whenever the reference

switching waveform changes polarity. If we were to measure the *average* gain of the phase-sensitive detector over many consecutive reference cycles the result would of course be zero, provided that the reference switching waveform spent an equal amount of time in its two states. As shown in Fig. 9.1(a), (b) and (c) the effect of changing the symmetry of the switching waveform is to change the average gain of the phase-sensitive detector to a value somewhere between the extremes of +1 and -1.

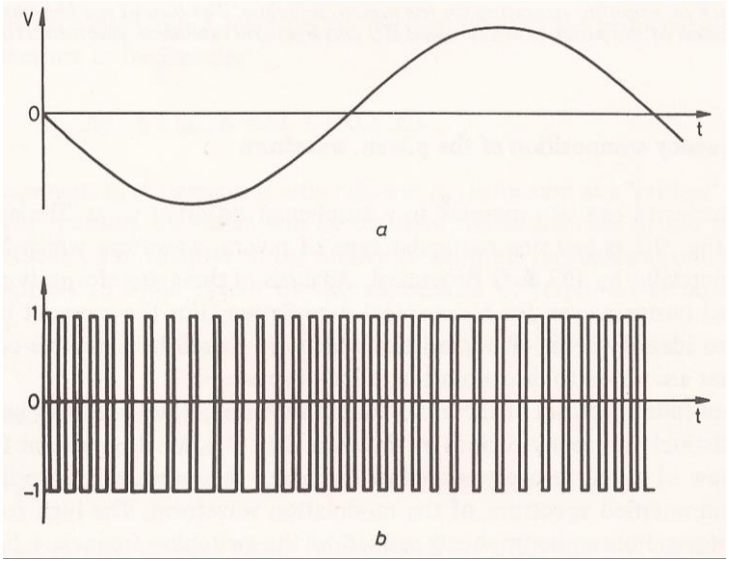

Fig. 9.2 **(a) Modulation voltage; (b) pulse-width-modulated switching waveform**

Suppose now that the waveform symmetry is subject to a long-term variation imposed by modulating the switching waveform as shown in Fig.9.2. The average gain of the phase-sensitive detector (Measured over a time which is long compared with a switching cycle but short compared with a modulation period) is now subject to a continuous variation which is free from discontinuities. The variation in gain, moreover, reproduces the modulation waveform exactly.

This approach, the pulse-width modulated reference channel, provides a means of achieving a sinusoidal gain variation in the signal path while retaining the dynamic range benefits of a switching phase-sensitive detector. A sinusoidal variation has been chosen because this clearly brings us a step closer to a system with fundamental-only response. To take a broader view: the system described approximates to an ideal multiplier model in which the reference input could be of any waveform, supplied in the form of a modulation voltage. Fig. 9.3 shows this configuration of a lock-in system working on this principle.

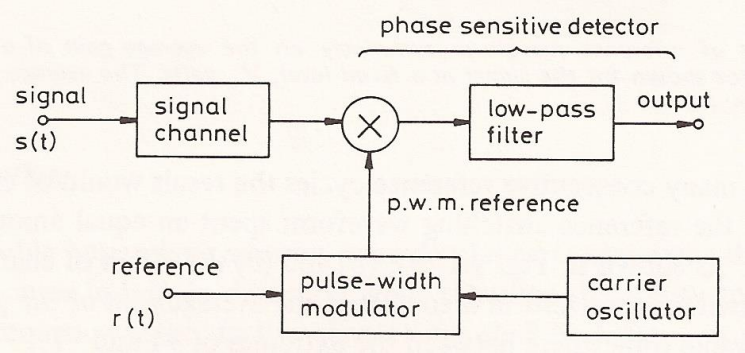

Fig. 9.3 **Lock-in amplifier operating on the p.w.m. principle. The output is a low-pass filtered version of the product of the signal *s*(*t*) and a general reference waveform *r*(*t*)**

9.3 Frequency composition of the p.w.m. waveform

P.W.M. waveforms can be generated in a number of different ways. The waveform shown in Fig. 9.2 is just one particular type of p.w.m. waveform which has been used commercially by EG&G Brookdeal. Analysis of these waveforms is generally a protracted business even for a sinusoidal modulation. For the moment it will be sufficient to identify some characteristics which are shared by a number of p.w.m. schemes that are suited to incorporation in lock-in systems.

The amplitude spectrum of these modulated switching waveforms can be divided into two distinct frequency regions as shown in Fig 9.4. Most important from the point of view of signal recovery is the low-frequency region which incorporates the complete unmodified spectrum of the modulation waveform. The high frequency region contains Fourier components related to the switching frequency f_0 and its harmonics, each of which carries sidebands derived from the Fourier components of the modulation.

The switching waveform thus combines the Fourier components of the modulation voltage with other, non-harmonically related, components at higher frequencies. The latter account for the switching characteristics of the waveform and each is associated with a transmission window where the detection system is susceptible to interference components.

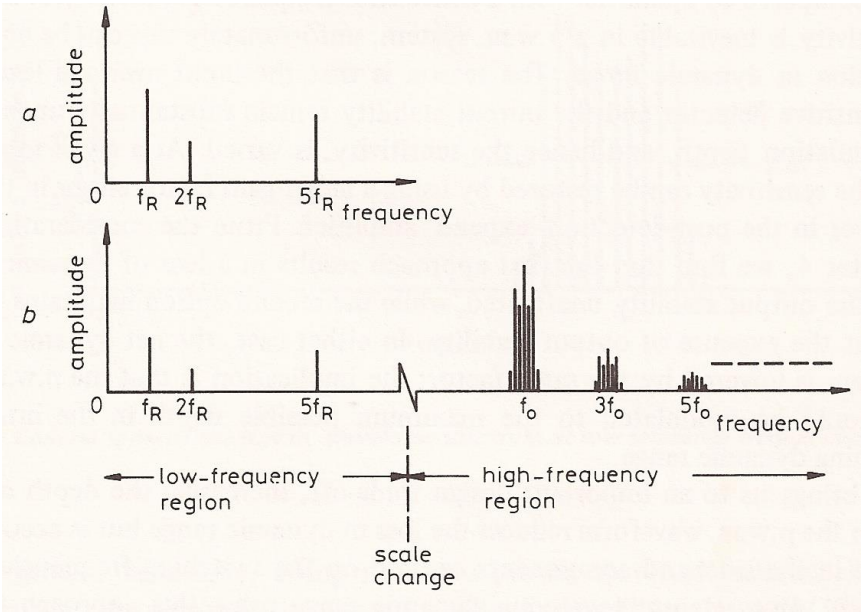

Fig 9.4 Amplitude spectra of: (a) modulation voltage; (b) a typical p.w.m. switching waveform

For fundamental-only response, the modulation will be sinusoidal at frequency f_R which is usually much less than f_0. We then identify the principal components of the spectrum at frequencies:

$$f_R,\ Kf_0 \pm Lf_R;\quad K \text{ odd},\ L = 0,\ 1,\ 2,\ 3\ \ldots$$

Each component in the spectrum, other than at f_R, is located at a "critical" frequency near which spurious responses will be obtained. As in the case of the heterodyne systems discussed in Chapter 8, the ability to suppress responses at odd harmonics of f_R is offset to some extent by the appearance of responses at unrelated frequencies.

9.4 Basic design considerations

9.4.1 Dynamic range

In practical p.w.m. systems giving fundamental-only response the magnitude of the component at frequency f_R is found to vary linearly with the amplitude of the modulation voltage which, in turn, determines the depth of modulation in the p.w.m. waveform. The component at f_R can take reasonably large values - of the same order of magnitude as the dominant high-frequency components - without causing over-modulation. This turns out to be very important in phase-sensitive detection where the "primary" response at $f_s = f_R$ should be as large as possible. Failure to achieve this means that the overall detection system will suffer a significant loss of sensitivity compared to operation with a conventional squarewave reference. Some loss of sensitivity is inevitable in a p.w.m. system: unfortunately this can be equated to a reduction in dynamic range. The reason is that the input overload level to the phase-sensitive detector and the output stability remain substantially unchanged as the modulation depth, and hence the sensitivity, is varied. At a given modulation depth, the sensitivity can be restored by using a larger gain factor either in the signal channel or in the post-detection "expand" amplifier. From the considerations given in Chapter 4, we find that the first approach results in a loss of dynamic reserve, leaving the output stability unaffected, while the second option maintains dynamic reserve at the expense of output stability. In either case, the net dynamic range of the system is lowered by the same factor; the implication is that the p.w.m. waveform should be modulated to the maximum possible depth in the interests of maintaining dynamic range.

This brings us to an important design trade-off. Increasing the depth of modulation on the p.w.m. waveform reduces the loss in dynamic range but is accompanied by a rise in the sideband components centred on the switching frequencies and its harmonics. Any attempt to recoup dynamic range using this approach is consequently matched by an increase in the general level of spurious responses. In addition, high-order sidebands which had negligible magnitude will now introduce transmission windows in a frequency range closer to the reference frequency, and so assume greater practical importance.

9.4.2 Spurious responses

In general, the magnitude and extent of the sideband arrays centred on f_0 and its harmonics vary non-linearly with modulation depth and can only be determined for a given modulation scheme by exact analysis. The scheme exploited by EG&G Brookdeal in their Sinetrac systems is particularly difficult to analyse because, in addition to modulating the mark/space ratio of the switching waveform, the modulating signal also causes a shift of the carrier frequency. We shall therefore restrict ourselves to a general review of system behaviour.

In the Sinetrac system, the sinewave modulation voltage at frequency f_R is specified at a standard level of 1 V r.m.s. The modulation depth corresponding to this level of modulation signal gives a loss of 10 dB in the sensitivity of the phase-detector. The sensitivity is regained by introducing an extra gain stage of 10 dB in the signal channel, with a consequent loss of 10 dB in the system dynamic reserve.

When the modulation is at a reference frequency f_R which is much lower than f_0, the sidebands centred on the carrier frequency are closely spaced at frequencies $f_0 \pm Lf_R$ but the *extent* of the sideband array is strictly limited as shown

schematically* in Fig. 9.5. The total width, $2\Delta f$, is typically 20 kHz and substantially independent of modulation frequency when f_R takes sufficiently low values, giving a spectrum similar to that of a frequency-modulated carrier under conditions of large-index modulation.

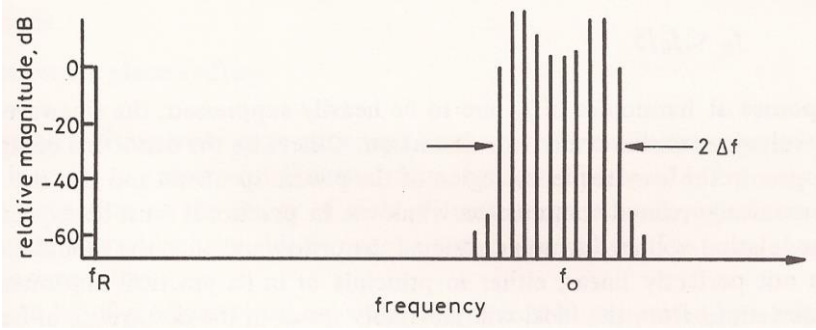

Fig. 9.5 General form of the p.w.m. waveform spectrum at low reference frequencies

This general behaviour, giving rise to a p.w.m. spectrum with sensibly fixed characteristics, corresponds to operation with f_R less than about $f_0/30$. For reference frequencies in this range, a clear separation is maintained between the low- and high-frequency regions of the spectrum. The lowest critical frequency, where a spurious response of significant magnitude could be obtained, is now in the region of $f_0 - \Delta f$, which is well removed from f_R.

When f_R is increased, keeping the depth of modulation constant, the separation of the sidebands becomes correspondingly larger. In order to predict the critical frequencies and the magnitude of their associated transmission windows, it now becomes necessary to give individual attention to each of the sidebands. In practice it will be the sidebands below the carrier frequency which prove to be the most troublesome, located at frequencies $f_0 - Lf_R$.

It turns out that for sufficiently high reference frequencies, the major contribution to spurious responses comes from the low-order sidebands for which $L \leq 4$. This is illustrated by Fig. 9.6, drawn for the specific case where $f_R = f_0/10$. Here, the separation between the low- and high-frequency regions of the p.w.m. spectrum is much less well defined. Also, it is evident that, if f_R exceeds a certain value, one of the sidebands will enter the frequency range *below* f_R.

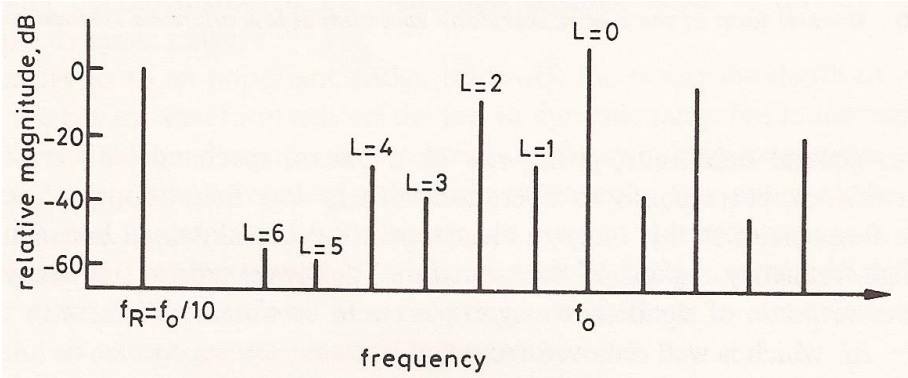

Fig. 9.6 Major sideband components for $f_R = f_0/10$, shown relative to the "primary" component at f_R

* Note that, in the Sinetrac system, application of the modulation voltage shifts the mean carrier frequency by about 10% from its "free-running" value. We shall ignore this effect in the following discussion.

To avoid the possibility of a transmission window appearing at an apparently arbitrary frequency below the reference, we must place an upper bound on the reference frequency. Assuming that sidebands corresponding to $L > 4$ have negligible magnitude, this gives us the condition

$$f_0 - 4f_R > f_R$$

which limits the reference frequency to the range

$$f_R < f_0 / 5$$

If responses at harmonics of f_R are to be heavily suppressed, the sinewave modulation voltage must have very low distortion. Otherwise the distortion components will appear in the low-frequency region of the p.w.m. spectrum and give rise to a set of harmonically-related transmission windows. In practice it must be expected that the modulation voltage has some residual distortion and that the modulation process is not perfectly linear, either in principle or in its practical implementation. Such deviations from the ideal will inevitably result in the occurrence of harmonic responses, albeit at low level.

9.4.3 Choice of switching frequency

For maximum separation between the low- and high-frequency regions of the p.w.m. spectrum, the switching frequency f_0 should be chosen to be much higher than the maximum anticipated reference frequency. The limitation on f_0 is decided ultimately, by the dynamic range of the phase-sensitive detector which deteriorates at high switching frequencies. This deterioration is compounded by the additional loss in dynamic range inherent in the p.w.m. approach. It has been observed that a p.w.m. reference channel can be configured as an option to an otherwise conventional lock-in amplifier. If this is the case, the phase-sensitive detector will be optimized over a range of frequencies rather than at a fixed high frequency. The p.w.m. switching frequency must then be chosen to be comparable with the highest frequency envisaged in conventional operation. Inevitably, the highest permitted value of f_R in p.w.m. operation must then be significantly less than this value. The EG&G Brookdeal Sinetrac lock-in amplifiers are subject to such a constraint; here the maximum recommended reference frequency in p.w.m. operation is 25 kHz in a system which can operate to frequencies above 100 kHz in conventional mode.

9.5 Reference phase-shifting

As described so far, p.w.m. systems depend on the provision of a sinewave reference voltage in order to achieve fundamental-only response. This would obviously place a severe restriction on the utility of such systems compared with, say, heterodyne lock-up amplifiers which are able to operate with a wide range of externally applied reference waveforms.

Also, to be of practical value, a p.w.m. system must be supported by a reference phase-shift network to enable the phase of the sinewave modulation voltage to be adjusted relative to a synchronous signal.

Clearly, these two drawbacks can only be overcome by adding to the complexity of the reference channel. In deciding on a suitable processing system, the following factors must be taken into account:

(i) Although we have identified an upper limit on f_R, to avoid low-frequency spurious responses, there is no fundamental limit on the lowest value of f_R which might be used. This implies that to exploit the p.w.m. technique to the full, the reference processing circuits should be capable of operating over a wide range of frequencies, amounting to several decades.

(ii) The sensitivity of the phase-sensitive detector, and hence the calibration of the overall system, depends directly on the amplitude of the sinusoidal modulation voltage, which should therefore have constant value over the entire operating range.

(iii) The sinewave applied to the pulse-width modulator should have a very low level of distortion if the detection system is to reject responses at harmonics of the reference frequency.

These requirements are fully met by the reference channel used in the EG&G Brookdeal Sinetrac series of lock-in amplifiers. The configuration is shown in Fig. 9.7. The reference input stage and the broadband phase-shift network are those of a strictly conventional lock-in amplifier, capable of operating with high precision over a frequency range in excess of five decades. The phase-shifted output of the reference channel is a closely controlled squarewave which is subsequently converted, first to a triangle and then to a sinewave. The circuits used for squarewave-to-triangle conversion have been described by Carter and Faulkner[1]; the final conversion to sinewave form is given by a piecewise linear network adjusted for a low level of distortion over the full frequency range.

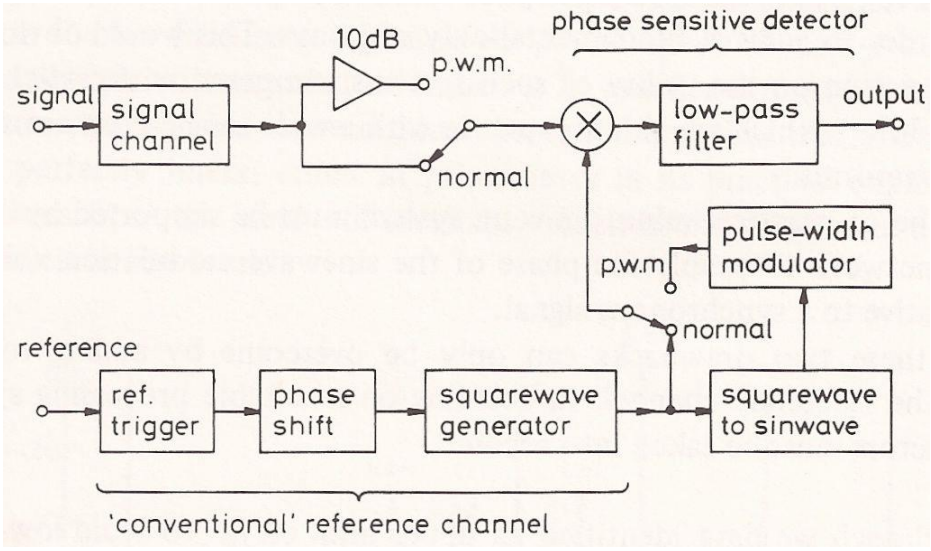

Fig. 9.7 Broadband lock-in amplifier with facility for fundamental-only operation in p.w.m. mode

Fig. 9.7 also shows the relatively simple arrangement of switches required to convert a harmonically-responding, conventional, lock-in amplifier to a system with fundamental-only response. An additional signal-channel gain stage of 10 dB is used to maintain the overall sensitivity of the system in p.w.m. mode. The diagram serves to emphasise that the p.w.m. configuration does not rely on the use of filters, and is consequently free from phase and amplitude errors due to filter misalignment. The phase accuracy and the residual phase noise are therefore comparable with those of the basic conventional system and significantly better than that of a heterodyne lock-in amplifier operating in the same reference frequency range.

9.6 Two-phase systems

The extension to two-phase systems requires an additional squarewave-to-sinewave convertor operating on the squarewave output of the quadrature reference channel. A second pulse-width modulator, operating independently of the first, is used to supply the reference input to the quadrature phase-sensitive detector.

It should be noted that p.w.m. systems do not depend on a carefully adjusted carrier frequency. In practice, the carrier frequencies of the two pulse-width modulators used in a two-phase lock-in amplifier need to be no more than nominally equal.

9.7 Analogue correlation

"Analogue correlation" is a term used in the context of p.w.m. lock-in amplifiers to cover the various modes of operation made possible when an external reference waveform is applied directly to the pulse-width modulator. This opens up numerous possibilities some of which are reviewed in the following sections.

9.7.1 Matched detection

The idea of a "matched" detector was mentioned in Section 3.5.5 in relation to the measurement of a squarewave signal using a conventional squarewave reference. If we now suppose that a periodic but non-sinusoidal waveform was used as the modulation input to a p.w.m. lock-in amplifier, the resulting detection system would be characterized by a set of transmission windows which could be exactly matched in amplitude, frequency and phase to the Fourier components of the signal. Such a system proves capable of yielding the best possible output signal-to-noise ratio for signals obscured by white noise. Unfortunately, signal recovery problems are usually associated with noise spectra far more complicated than this, so that the benefits of matched detection (which are, in many cases, marginal) are difficult to realize in practice. It usually turns out that the ability to operate in a fundamental only response mode with relative freedom from transmission windows close to the reference frequency gives a far greater advantage when measuring non-sinusoidal signals in noise. As noted in Section 3.5.5 and in Chapter 6, the null-shift procedures can be applied when the detection system has fundamental-only response, to quickly bring the reference phase-shift to an optimum setting under very noisy conditions.

9.7.2 Two-frequency lock-in analysis

In a two-phase lock-in amplifier operating on the p.w.m. principle, the pulse-width modulators associated with each phase-sensitive detector operate independently and have separate inputs. These can be supplied with external reference waveforms having different fundamental frequencies and different waveforms if so required. The ability to use independent reference inputs means that different spectral components of the signal can be separately measured on each of the phase-sensitive detectors.

Since direct connection of external references to the modulator inputs bypasses the reference phase-shifting networks, this facility is likely to be most useful in such applications as optical spectroscopy where there is minimal phase-shift between the reference and signal. An example of spectrometer operating with two chopper frequencies is shown in Fig. 9.8. The light paths from the two samples are combined at the cathode of single photomultiplier. A two-phase p.w.m. lock-in amplifier is subsequently used to measure the two chopped signals separately and simultaneously, using "analogue" reference inputs derived from the drives to the optical choppers. A ratiometer can be a useful accessory in this type of measurement; however, a digital interface of the type described in Chapter 10 provides a more flexible approach to processing the outputs from the twin channels.

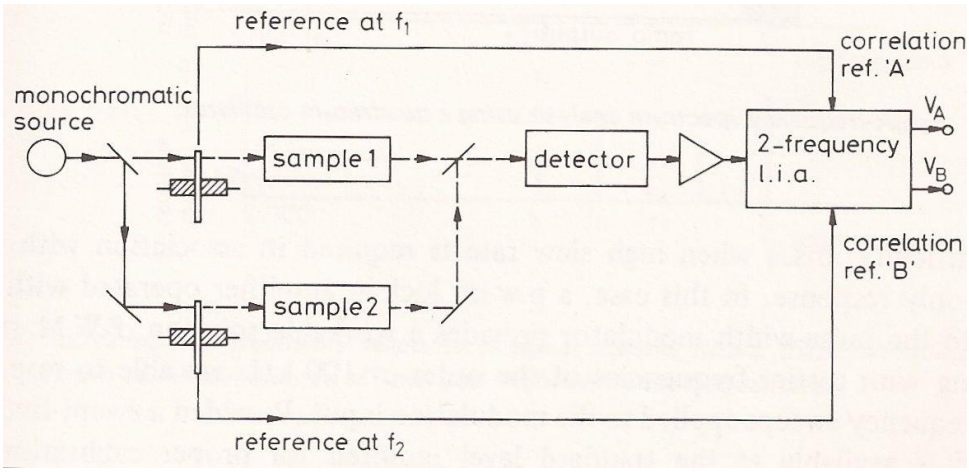

Fig. 9.8 **A two-frequency lock-in amplifier application. The outputs V_A and V_B are proportional to the signals transmitted by samples 1 and 2 respectively.**

9.7.3 High slew rate applications

In principle, the limited slew rate capability of conventional lock-in amplifiers could be overcome by bypassing the reference phase-shifting networks and applying a swept-frequency switching waveform directly to the phase-sensitive detector. Such a procedure is valid when the signal and reference remain sensibly in phase over the desired frequency sweep. In practice, the procedure can be applied to conventional lock-in systems only when the phase-sensitive detector is accessible in the form of a modular unit.

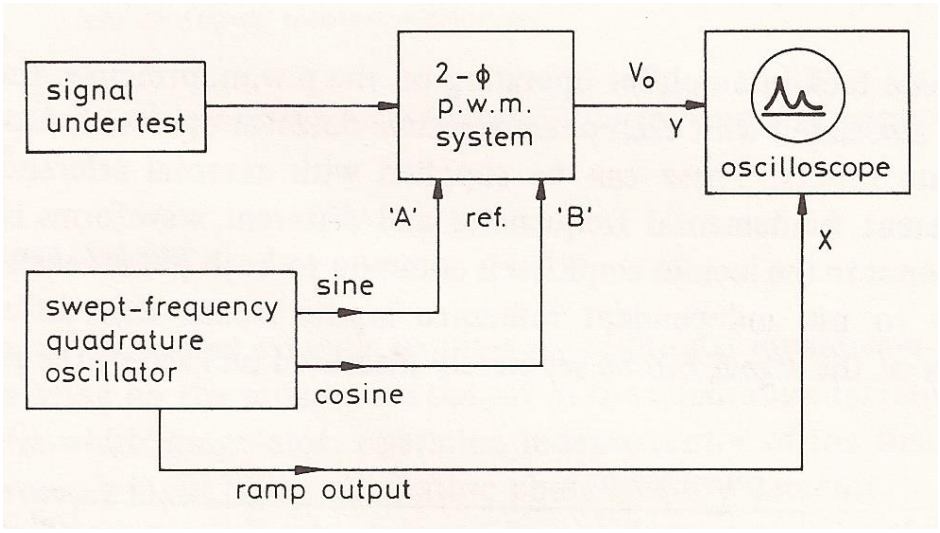

Fig. 9.9 **Swept-frequency spectrum analysis using a quadrature oscillator**

A difficulty arises when high slew rate is required in association with fundamental-only response. In this case, a p.w.m. lock-in amplifier operated with direct access to the pulse-width modulator provides a workable solution. P.W.M. systems operating with carrier frequencies of the order of 100 kHz are able to respond to rapid frequency sweeps applied to the modulation input. Provided a swept-frequency sinusoid is available at the standard level required for proper calibration, it is possible to obtain fundamental-only response consistent with slew rates far in excess of most practical requirements.

Wide-band swept spectrum analysis was cited in Chapter 6 as a lock-in amplifier application where fundamental-only response and high slew rate were essential joint requirements. In order to exploit the characteristics of a two-phase p.w.m.

system in this application, it is necessary that the two pulse-width modulators are operated in strict quadrature at the swept reference frequency. This highlights the need for a swept-frequency oscillator providing quadrature sinewave outputs at a standard level. Such oscillators have been made available as accessories to two-phase p.w.m. lock-in amplifiers for use in the configuration shown in Fig. 9.9. In wideband applications using high frequency resolution, the sweep-rate limitation in this type of system lies with the output filters as explained in Section 5.5.3.

9.8 Interference rejection filters

In an earlier version of the Sinetrac lock-in amplifier, the signal channel was fitted with a 2-pole low-pass filter cutting off above the maximum reference frequency of 25 kHz but well below the 100 kHz switching frequency. In addition, a notch filter was used to enhance the suppression of signal components in the region of 100 kHz. The object was to overcome the major spurious responses associated with p.w.m. operation.

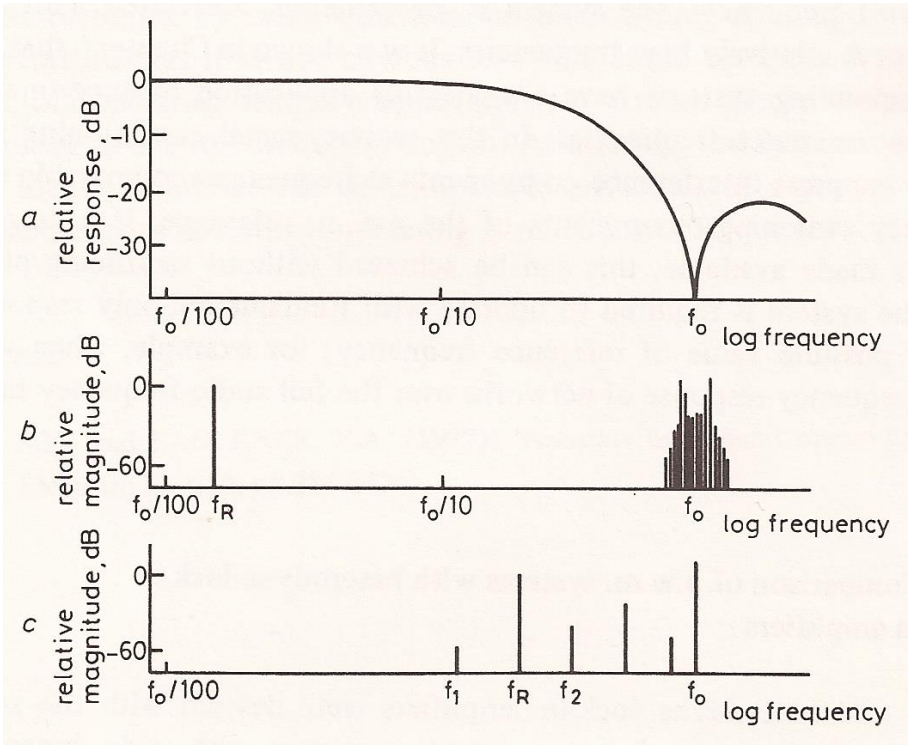

Fig 9.10 (a) Combined frequency response of signal channel filters. (b), (c) Amplitude spectrum of PWM reference at "low" and "high" reference frequencies

In practice, such a combination of filters is likely to prove desirable only at comparatively low reference frequencies for the following reasons. First of all, we have seen that at low reference frequencies the transmission windows are concentrated in the spectral regions close to the switching frequency and its odd harmonics. Spurious responses are therefore associated with high-frequency interference and the filters are most effective in suppressing these in advance of detection. Secondly, at low frequencies the phase-shift introduced by the filters will be relatively small, enabling the overall system to operate with good phase precision.

Conversely, at high reference frequencies, the signal channel filters will introduce large phase errors into the measurement system while the transmission windows move to much lower frequencies by reason of the wider sideband separation. The contrasting situations are illustrated in Fig. 9.10. We find that at higher reference

frequencies there will be a number of critical frequencies, for example f_1 and f_2 which fall within the bandwidth of the signal channel filter. Interference components close to these frequencies would suffer minimal attenuation in the filter and be able to excite spurious responses. In some cases, therefore, the inclusion of the filter acts more to the detriment of the system, introducing large phase errors while having only partial success in overcoming the problem of spurious responses.

The decision to eliminate interference rejection filters with fixed characteristics in later versions of the Sinetrac system resulted in a lock-in amplifier capable of fundamental-only response, consistent with excellent phase accuracy, over many decades of frequency. The system is, nevertheless, associated with large spurious responses at relatively high frequencies. It was shown in Chapter 6 that fundamental-only responding systems have considerable application in experiments operating with low reference frequencies. In this regime, signal conditioning filters can be used to suppress interference components at frequencies comparable with the high-frequency switching components of the p.w.m. reference. If a range of optional filters is made available, this can be achieved without sacrificing phase accuracy when the system is required to operate with fundamental only response up to the highest possible value of reference frequency: for example, when measuring the swept-frequency response of networks over the full audio-frequency range.

9.9 Comparison of p.w.m. systems with heterodyne lock-in amplifiers

P.W.M. and heterodyne lock-in amplifiers were devised with the same objective in mind; to give a synchronous detection system with wide dynamic range and relative freedom from harmonic responses over a wide range of reference frequencies. As we have seen, the two approaches to this problem lead to vastly different solutions, both of which involve system designers in a number of trade-offs and compromises.

On balance, modern heterodyne lock-in amplifiers appear to offer the widest frequency range consistent with the lowest level of spurious responses, whereas commercial versions of the p.w.m. system operate up to a maximum frequency of about 25 kHz and have a number of large transmission windows accessible in the frequency range immediately beyond this value.

The overall phase accuracy of p.w.m. systems is superior to that of heterodyne lock-in amplifiers, which are susceptible to alignment errors in a number of sub-systems and generally require a far more complex configuration. The difference in complexity is reflected in system cost, since p.w.m. lock-in amplifiers usually offer a cheaper means of obtaining fundamental-only response than their heterodyne counterparts. Also p.w.m. systems can usually be converted by pushbutton selection to operate as conventional lock-in amplifiers, giving an extension of the frequency range and allowing the fundamental-only response to be traded for greater dynamic range. This flexibility in choosing the response of the system extends to choosing an arbitrary response given by the Fourier components of the applied reference waveform.

A feature that both heterodyne and p.w.m lock-in amplifiers have in common is that fundamental-only response at low frequencies is obtained by operating the phase-sensitive detector at a relatively high frequency. The result in both cases is a system with a dynamic range independent of reference frequency but less than that which might be achieved if the phase-sensitive detector was operated in conventional fashion. It was shown in Chapter 8 that the dynamic range of

heterodyne systems can be recouped by using synchronous heterodyning as a supplementary technique. Unfortunately, it is not obvious how this could be applied to phase-sensitive detectors operating with a p.w.m. reference without incurring additional spurious responses in the low-frequency region. As a result, the dynamic range of a heterodyne lock-in amplifier can be comparable with that of a p.w.m. system where the phase-sensitive detector is operating at much lower frequency.

9.10 Reference

CARTER, S.F., and FAULKNER, E.A. (1977): "Accurate broadband square-to-triangle converter", Electron, Lett., 3, pp. 381-382.

CHAPTER 10

Computer-controlled lock-in amplifiers

10.1 Introduction

The advent of the microprocessor and the increasing availabity of desk-top computing power have provided a challenge to both designers and users of electronic measuring equipment. Instrument designers are faced with a demand for 'intelligent' instruments capable of performing programmed tasks or able to communicate with other instruments via a computer controller. As for the instrument user; he is concerned with using these new instruments to the best effect and with devising measurement procedures that take advantage of the latest developments in instrument technology.

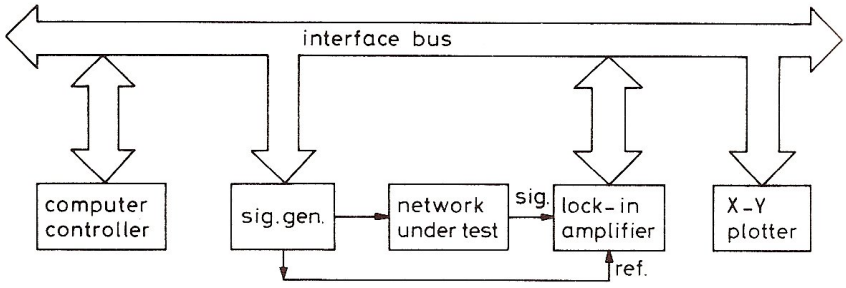

Fig. 10.1 Bus-controlled measurement system

This general situation is reflected in the specific case of instrumentation for signal recovery where increasing emphasis is being placed on computer control in its widest sense. Developments in this area have been greatly influenced by the widespread adoption of the IEEE-488 bus system for providing two-way digital communication between measuring instruments and a computer-controller. This has resulted in the availability of a large range of compatible instruments that can be, literally, plugged together to produce a computer-controlled measurement system. Many of the applications listed in Appendix 1 could benefit from such an approach; for example, Fig. 10.1 shows a bus-compatible lock-in amplifier operated in conjunction with a number of other controllable devices to provide an automatic system for frequency-response measurement.

The IEEE-488 bus protocol is rigidly defined; thus, at any time, only one device is permitted to 'talk', that is send data or commands over the bus, while several devices may 'listen' in order to receive data or commands. In the example shown, the *X-Y* plotter and signal generator would probably operate as 'listeners' while the lock-in amplifier would both 'talk' and 'listen', to transmit data to the computer and receive commands. Overall control comes from the computer, which is programmed to change the frequency of the signal generator in discrete steps, to manage the take-up of data from the lock-in amplifier, and to process data for presentation on the *X-Y* plotter.

In the computer-controlled lock-in amplifiers to be discussed in this chapter, the phase-sensitive detector remains intact at the heart of the system, supported by signal and reference channels having characteristics similar to those described in

earlier chapters. The incorporation of digital control lines to switch the sensitivity and the internal configuration of the lock-in amplifier can be achieved without compromising key specifications such as input dynamic range and operating frequency range. In practice, therefore, the only serious limitation incurred in operating with a computer interface is with regard to output dynamic range. This is now limited by the use of an analogue-to-digital convertor on the phase-sensitive detector output. An attempt to match the 100 dB dynamic range of a typical analogue output would require a 17-bit conversion and would be difficult to justify on grounds of cost in a general-purpose measurement system. The usual provision is for a $3^1/_2$ digit conversion, giving a resolution of 10 mV in a 10 V output with 100% over-range. If this is inadequate for a particular application, the analogue output is available on its usual socket and can be separately converted to high precision if so required.*

The handling characteristics of the lock-in amplifier being relatively unchanged, the main problem in digital control is to create programs which reproduce the measurement routines and setting-up routines that are associated with the detection of noisy signals. In giving consideration to these routines it will be convenient to distinguish between the two main types of controllable lock-in amplifier in general use. These are 'programmable' lock-in amplifiers where the software control routines are resident in an external computer controller, and microprocessor-based systems – so called 'intelligent' lock-in amplifiers. The latter feature a number of stored software routines that can be initiated by front-panel switch selection or by a command transmitted on the interface bus.

10.2 Programmable lock-in amplifiers

In early lock-in amplifiers, the pushbutton and switch selectors controlling the overall system configuration were heavily interlocked and interlinked and required front panel assemblies that were both complex and labour-intensive in production. At a later stage, f.e.t. switches, controlled by the application of standard logic levels, became widely used for both gain selection and mode selection and there was a move to transfer hard-wired switching logic to integrated circuits mounted on the printed circuit board. This change leads to a dramatic simplification in switch design. For example, the sensitivity switch of a typical lock-in amplifier is reduced from a multi-wafer assembly to a single-pole selector as illustrated in Fig. 10.2.

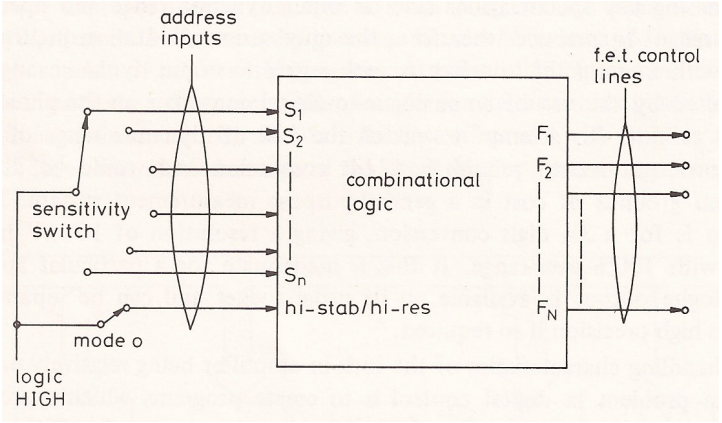

Fig. 10.2 Simplification of switching operations by using a combinational logic circuit

* Note that to exploit the full output dynamic range of these instruments generally places severe demands on peripheral equipment, both analogue and digital.

The sensitivity lines S_1 to S_n provide an address input to a combinational logic block which is tabled to produce the appropriate output combination on the f.e.t. control lines F_1 to F_N. These are then used to switch the gain of the amplifier stages in the lock-in amplifier signal channel. In a comprehensive detection system, giving a choice of 'high stability' or 'high reserve' operation, the address lines might be augmented by an additional input which is either HIGH or LOW depending on mode selection. The combinational logic circuit is then arranged to control the f.e.t. switches for any combination of mode and sensitivity, without adding to the complexity of the mechanical switching assembly.

The transition from a 'standard' instrument designed along these lines, to one where the internal switches can be controlled by the application of logic levels from an external source is relatively straightforward. In a fully 'programmable' lock-in amplifier, all the functions that are normally switched from the front panel, such as sensitivity, time constant, phase quadrants and 'expand', can be controlled from logic levels applied to a 'digital' input. In lock-in amplifiers having a voltage-controlled phase-shifter, the provision of a digital-to-analogue convertor enables the reference phase to be added to the list of controllable parameters. In its simplest form, the digital input might be a multi-way socket connected to a set of remotely operated switches. In modern instruments, however, the digital input is more likely to be a port having access to a standard bus system, such as the IEEE-488 bus referred to earlier. When the lock-in amplifier output is provided with an analogue-to-digital convertor, a properly defined bus system enables data to be transferred to the computer controller and to other instruments connected to, and controlled from, the bus.

In order to exploit this type of system effectively, the computer-controller must be provided with programs sufficiently powerful to undertake the management of the lock-in amplifier under a wide range of signal and noise conditions. We thus envisage a control program that defines an overall measurement procedure and contains a number of subroutines for sensitivity and phase selection. The specification of these subroutines requires a certain familiarity on the part of the user with the handling characteristics of lock-in amplifiers and would normally involve several stages of refinement before an acceptable solution was found. Some essential features of these routines are identified and discussed in Sections 10.4 and 10.5.

10.3 Microprocessor-based systems

The incorporation of a microprocessor to monitor and supervise the switching functions of a lock-in amplifier represents a significant advance in system concept and design. The result is a self-contained lock-in amplifier with the ability to undertake *sequential* switching operations controlled by software associated with the microprocessor. The control system takes additional data from

(i) front panel switch arrays

(ii) a digital interface to an external keyboard or controller

(iii) the outputs of the phase-sensitive detectors.

A lock-in system with this overall capability greatly cases demands on the user who no longer requires such detailed familiarity either with lock-in techniques in particular or signal recovery in general. For example, the resident software routines could enable the lock-in amplifier to adjust sensitivity and phase automatically to maximize the output for a given signal. As far as the user is concerned, the lock-in amplifier now operates like a special type of a.c. microvoltmeter which can read the amplitude and phase of a signal in response to

a single key stroke or bus command. This reduced level of operational complexity is reflected in the amount of programming effort required to control the lock-in system when it forms a component part of a larger bus-controlled system.

There are other aspects of operation which benefit non-specialist users that apply to almost any type of microprocessor-based instrument. For example, the digital output can be scaled to reflect the input signal level, taking all factors such as sensitivity multipliers and amplifier gains into account. The system is then much less prone to operator error than a mechanically switched system fitted with a pointer scale where the danger of overlooking a scaling factor is always present. Also, since all switching operations from the front panel are supervised by the microprocessor, the system is able to inhibit or give warning of undesirable or unorthodox combinations of front-panel controls, supported by a display or print-out of the appropriate error message.

In an instrument such as a lock-in amplifier which is subject to frequency-dependent errors in the signal and reference channel circuits, there is ample scope for using the microprocessor in an automatic calibration routine. This would measure and store calibration errors over the frequency range of the instrument with the object of providing corrected results in the final measurement. In principle, the calibration routine could be extended to correct phase-sensitive detector offsets and to compensate the amplitude and phase characteristics of signal conditioning filters introduced into the signal channel.

Clearly, the incorporation of a microprocessor has progressively greater impact as the complexity of the lock-in system is increased and should, ideally, enable a greater number of facilities to be offered without sacrificing case and clarity of operation. This objective generally requires a fresh approach to front panel design. For example, the familiar phase dial of a lock-in amplifier might be replaced by a counter that can be incremented or decremented using a pushbutton switch. The reference-channel phase-shifter is then controlled from a digital-to-analogue convertor taking its input from the microprocessor data bus. The problem of monitoring the status of the instrument is overcome by displaying the phase setting on a digital panel meter which also serves to display error codes and fault conditions when the system is operated.

As an additional constraint on the system designers, experienced users would normally require that the system is able to revert to full manual control where various combinations of front-panel settings could be tried without being restricted to operate from a ‘menu’ of stored routines. This constraint would certainly apply to any microprocessor-based lock-in amplifier that was offered as a general purpose measurement tool rather than as a special-purpose instrument, rigidly programmed to perform a specific range of tasks.

The selection of routines available on commercial instruments is limited but carefully chosen to enhance the handling characteristics of the lock-in amplifier in a wide range of applications. In addition to the software routines for sensitivity and phase selection referred to earlier, there is usually the possibility to offset data by a fixed amount and to normalize data, providing an output expressed as a fraction of percentage of some predetermined level. Routines of this type are therefore applied after detection and serve as a first stage of output processing. If more complex processing is required, this would normally be carried out by a computer interfaced to the lock-in amplifier, programmed to suit the needs of a specific experiment.

Management of this interface by a microprocessor resident in the lock-in amplifier offers several advantages over a ‘hardware only’ design. Thus, transmitted data can be presented in an easily understood format, only relevant

data need be transmitted, and received commands can be less complex and more meaningful. The impact of the microprocessor on the design of an IEEE-488 compatible lock-in amplifier is brought out in Table 10.1. Comparison is made with a notional design based on classical techniques, and improvements are attributed to specific characteristics of the microprocessor.

Regarding the routines for sensitivity and phase adjustment; it is essential here that the criteria for range and phase switching are clearly stated if the lock-in amplifier is to behave predictably under the worst conditions of signal and noise. The following sections give further discussion on these routines and apply equally to a detection system under software control from a microprocessor or from an external controller linked by a data bus.

Table 10.1

Instrument characteristic	Classical design techniques	Improvement in microprocessor-based design	Microprocessor characteristic leading to improvement
Digital display of output	Output in range 0 to ± 10V scaled by reference to gain setting and ×1, ×2, ×5 multipliers	Direct scaling reflecting input level	Multiplication program
Digital display of phase	0 to 99.9° Quadrant information on +90°and +180° switches	0 to 359.9° Direct indication	Addition capability
Autorange mode	Hardware design requiring physical links with sensitivity and time-constant switches	Software-only design: no additional hardware required to implement mode	Digital comparison Data manipulation Program storage
Zero offset	Manual operation	Manual-automatic operation	Program storage Data manipulation Subtraction capability
Initial Set-up	None	Automatic operation	Program storage Data manipulation Digital comparison Mathematical capability
Normalize	Manual operation	Manual/automatic operation	Program storage Data manipulation mathematical capability

Table 10.1 (continued)

Variable phase and zero-offset hardware	3-digit d.a.c. or 10-bit binary d.a.c. plus b.c.d. to binary conversion hardware	10-bit d.a.c.	B.C.D. to binary program storage
Output conversion	$\pm 3^1/_2$ digit a.d.c.	12-bit binary a.d.c.	Binary to b.c.d. program storage
IEEE 488 Transmitted data	Fixed format. Interpretation required for phase and output data	Flexible format No interpretation required No redundant data need be transmitted	Data manipulation Storage of format styles Mathematical capability Read/write storage capability
IEEE 488 Received commands	Fixed format: hardware determined	Flexible format Meaningful commands	Data manipulation Data storage Storage of format styles
Front panel control	Rotary, pushbutton and toggle switches. Hardware determined	Pushbutton switches only. Ergonomic improvements. Group of controls not constrained by internal design of instrument	Data manipulation Program storage

10.4 Automatic sensitivity selection

When a lock-in amplifier is operated manually, the response to a synchronous signal is usually adjusted by switching sensitivity to obtain an output as close as possible to a full-scale reading.

In a single-phase lock-in amplifier, the response will not necessarily be maximum and may even take negative values unless the phase of the reference channel has been correctly adjusted. When using a two-phase system, the sensitivity is usually adjusted to maximize the output of the phase-sensitive detector giving the largest response. Alternatively, when the use of a vector computer is appropriate, the sensitivity can be adjusted by observing the ‘magnitude’ output of the vector computer.

If the residual noise output of the lock-in amplifier is sufficiently large, it will be necessary to ensure that fluctuations in the output do not carry the indication due to the signal beyond full-scale. This would normally require an observation time amounting to several time constants and may result in the system being switched to lower sensitivity. Most systems offer a 1:2:5 or 1:3:10 switching sequence**,** so this final step can usually be achieved without a significant loss of output voltage.

The object of a sensitivity routine or autorange routine is to bring this sequence of operations under automatic control by comparing the *magnitude* of the lock-in amplifier output with predetermined ‘threshold’ levels. Some of the difficulties encountered when autoranging with a noisy signal are demonstrated by the following examples.

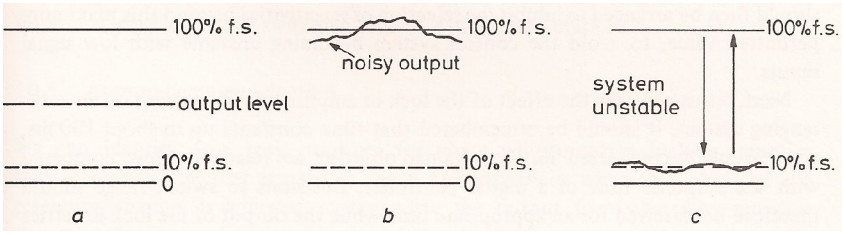

Fig. 10.3 Definition of switching thresholds in a decade autoranging system.

Suppose we have a lock-in amplifier where the sensitivity control is divided into a series of decade ranges, and that the response to a synchronous signal has been brought to a level corresponding to about half-scale output as shown in Fig. 10.3(*a*).

The lock-in amplifier controller is programmed to switch to a lower sensitivity (range-down) when the signal increases to a level corresponding to full scale output, and to switch to higher sensitivity (range-up) when the signal falls to 10% of full-scale. To be realistic we must allow for a small amount of residual noise appearing with the signal as shown in Fig. 10.3(*b*); hence the first 'down' transition will occur when the total output due to signal plus noise exceeds the 100% level. Fig. 10.3(*c*) shows the new situation which applies after the sensitivity has been switched. Signal and noise appear in the same relation as before, so it is only a matter of time before the total output falls below the 10% threshold level, causing the system to range-up to its original sensitivity. The situation illustrated in Fig. 10.3(*b*) is thus restored. If the signal remains at a constant value the system will attempt to switch gain alternately 'up' and 'down' with the result that a stable condition is never attained. To overcome this difficulty it is necessary to redefine the switching thresholds, for example by increasing the upper threshold to greater than 100%, or by reducing the lower threshold to less than 10%. In this way, the system can be made to tolerate residual noise on the output, at least up to a certain peak-to-peak level, and will be able to switch sensitivity to reach a well-defined condition.

Let us now look at the behaviour of the system under conditions of very low signal. If, at some point, the output signal-to-noise ratio falls drastically, or if the signal is removed, leaving only noise in the output, the sensitivity controller will attempt to switch gain to the maximum achievable value. The system will then remain in a stable condition at maximum sensitivity, *provided* the output noise peaks do not exceed the upper threshold level. If this level is exceeded, the controller will switch the sensitivity to a lower value. Unfortunately, the noise is bipolar and so repeatedly takes values close to zero voltage. The controller will thus restore the sensitivity to its maximum value at the first opportunity and subsequently make random transitions between the two most sensitive range positions.

These considerations suggest that the maximum usable sensitivity in autorange operation is where the peak output noise is just less than the full-scale output. This maximum sensitivity can be determined by experiment; the autorange program should then be arranged to inhibit the selection of sensitivities beyond this maximum permitted value, to avoid the control system becoming unstable with low signal inputs.

Next, let us consider the effect of the lock-in amplifier time constant on an auto-ranging system. It should be remembered that time constants up to about 100 ms, which would be considered 'fast' by a human observer, are relatively 'slow'

compared with the response time of a digital controller. Decisions to switch range should therefore be deferred for an appropriate time while the output of the lock-in settles to a new value following a switching operation. In commercial systems supplied with an autorange facility, a settling time of four or five time constants is usually allowed between successive switching operations. In a fully integrated system, the micro-processor will be provided with the time-constant setting as a matter of course. In a bus-controlled system, the time-constant setting may have to be 'read' by the controller via the bus interface and entered into the autorange subroutine. The subroutine would usually feature a number of WAIT instructions to ensure that the program runs at an appropriate rate for the particular time-constant selection.

The final point to be taken into consideration concerns the role of 'expand' selection in determining the sensitivity in autorange operation. It is shown in Chapter 4 that, in some lock-in amplifiers, a given sensitivity can he obtained for two or more combinations of a.c. and d.c. gain in the system and that the choice of combination influences the dynamic performance of the lock-in amplifier. It follows that, if the lock-in amplifier is required to autorange in a 'high reserve' mode, the switching program should be arranged to give a gain combination that uses the maximum possible value of expand gain. Conversely, a 'high stability' switching program would be biased in favour of using high a.c. gain in order to achieve the best possible output stability for precision measurements.

In a microprocessor-based system, these factors could be taken into account automatically, depending on the mode of operation selected by the user. Other facilities that would normally be made available include a procedure for entering the maximum autorange sensitivity (in the interests of system stability as described above) and a procedure for entering the threshold switching points, usually expressed as a percentage of full-scale deflection. In some cases, the upper threshold is fixed at 110% of full-scale output while the autorange routine covers the 1:2:5 range sequence of the lock-in amplifier. A system with these characteristics would switch range until the output indication lay somewhere between 40% and 110% of full-scale with the overall switching time determined by the time-constant selected on the lock-in amplifier.

It should be acknowledged that autorange switching routines are, at best, systematic and, at worst, cumbersome. At a time constant of 1 second, a typical autorange routine would take about 1 minute to switch from minimum to maximum sensitivity in a 1:2:5 sequence. When operating with a very wide range of signal levels, using a programmable system with a choice of programs, there is a possibility to include a 'trial' routine confined to decade switching. The idea is to obtain an order-of-magnitude estimate of signal level; this estimate can then be improved using the 1:2:5 switching sequence in a final iteration.

10.5 Automatic phase selection

We can identify two basic routines for phase adjustment in lock-in amplifier measurements. The first is used in signal recovery work where the phase of the reference channel is adjusted to maximize the output from the phase-sensitive detector. The second is associated with precision phase measurement where the phase adjustment is made with the objective of nulling the output of the phase-sensitive detector.

Phase measurements might be made with either a single- or two-phase lock-in amplifier. If the latter is used, it is normally the output of the quadrature phase-sensitive detector which is to be brought to a null condition. Since phase measurements are usually made with noise-free signals, the accuracy to which the null can be set will depend on the ability of the analogue-to-digital convertor

associated with the phase-sensitive detector to resolve small output changes. The resolution of the phase-shift control will be similarly limited in a digitally controlled system. In commercial systems the phase can usually be advanced in increments as small as 0.1°, which is comparable with the resolution of a conventional phase dial giving a continuous adjustment.

The phase-null routine can be initiated by subtracting increments of about 30° until the output changes sign. Smaller increments, say 5°, are then progressively added to the set phase until the output changes sign yet again. The procedure is repeated with successively smaller increments until the null is achieved to within the resolution capability of the system, or to within some specified limit.

Regarding signal recovery applications using a single-phase lock-in amplifier, an alternative approach to setting the phase is defined as follows, starting from an arbitrary initial phase condition:

(i) 'read' the in-phase value of the signal, V_A

(ii) add 90° to the set phase of the reference channel

(iii) 'read' the quadrature value of the signal, V_B.

(iv) compute $\phi = \tan^{-1} V_B/V_A$: reduce set phase by 90°

(v) add ϕ to the set phase.

This routine could be accomplished in a time equivalent to about 10 lock-in amplifier time-constants. When the lock-in amplifier has fundamental-only response, the resulting response will always be maximized and first-order independent of errors accrued in the measurement and in the computation of the signal phase. This procedure is perfectly adequate for use in general signal recovery applications and, more importantly, can be used to extend the usefulness of single-phase lock-in amplifiers in tasks which are normally reserved for two-phase systems.

Of course, if a two-phase system is available, the problem of setting phase need not arise in signal recovery work. All that is required is an autoranging control system to bring the vector magnitude to a suitable 'on-scale' value. In the case of a single-phase lock-in amplifier, an autoranging routine would normally be executed prior to setting the phase, followed by a final autorange routine to bring the maximized response within range.