Phase sensitive detection as a means to recover signals buried in noise

1975 J. Phys. E: Sci. Instrum. 8 621

(http://iopscience.iop.org/0022-3735/8/8/001)

Phase sensitive detection as a means to recover signals buried in noise

D P Blair and P H Sydenham
Department of Geophysics, University of New England, New South Wales, Australia

Abstract Phase sensitive detection is formalized in order that its operation can be understood and its practical use optimized. The theory explains how narrow bandwidth amplification is achieved in order to reduce the noise content of a measurand and how the process can be considered simply as a multiplier followed by a low-pass filter. The responses to various excitation and measurand forms are derived for signals ranging from the simple sinusoids to the more complicated signals produced when modulating interference fringes and stabilized laser cavities.

1 Introduction

In practical measurements the measurand is always accompanied by an unwanted proportion of noise energy that limits the ultimate sensitivity that can be attained. When the total noise power approaches or exceeds the signal power special techniques are needed to recover the signal.

This article is concerned with the phase sensitive detection method of limiting bandwidth. General noise reduction of the source and in amplifiers themselves plus studies of correlation methods of signal extraction are the subjects of other articles in this series.

2 The phase sensitive detection concept

When noise is white in character, that is, when it occurs as random variations having constant power per unit bandwidth at all frequencies, the effective noise level can be reduced by limiting the bandwidth of detection to that just necessary to include the range of frequencies occurring in the signal, thus reducing the noise power but not the signal. Practical difficulties restrict the narrowness that can be obtained with conventional filter bandwidths. They include drift, sensitivity to component tolerance changes and sometimes physical size. The problem is accentuated as attempts are made to reduce the bandwidth.

Phase sensitive detection overcomes the practical problems in a spectacular way for it enables extremely narrow bandwidth detection (0·001 Hz is normal) to be achieved with stabilities measured in parts per million. It does, however, require a little more circuitry than simple cascaded filters. It also offers other advantages, one being that the operating frequency of the detection preamplifier can be raised well above the large amplitude low frequency region of flicker noise. (This noise has an amplitude that is inversely proportional to frequency. It is significant in most kinds of pseudo-DC devices, whether they be of electronic, mechanical or optical nature.) Another advantage is the virtually total rejection of discrete frequency noise such as mains pickup in electrical systems and resonant vibrations in mechanical systems. This discussion relates to electrical signals because most measurands are sensed with electrical output transducers. The concepts, however, also apply to the analogous physical systems.

Noise voltage amplitudes rise to millivolt levels. Typical sources might be thermo-electric voltages, thermal resistance noise, photo fluctuations, shot noise and temporal variations due to ambient backgrounds. The signal, however, may be in the nanovolt region. Sophisticated phase sensitive detectors can extract signals that are buried in noise that is 10^7 times the signal magnitude (140 dB below). For example, one commercial unit set to have a 100 s output filter time constant, claims sensitivity to 1 nV.

Costs of implementing phase sensitive detection vary widely. Simple devices assembled from discrete components at a cost of £20 can give useful performance. Modular hybrid specific purpose circuit packages are offered from £60 upward. General purpose instruments range from £400 to £2000 depending upon the degree of sensitivity, stability and versatility.

3 Uses of phase sensitive detection

The principle has numerous applications, and it is surprising, therefore, to find that no satisfactory in depth treatment of the method exists in texts or instrumentation. Papers such as Brower (1968), Coor (1972), Danby (1970), Moore and Chaykowsky (1963), Smith (1972), provide differing forms of explanation but none adequately presents a common theory. Champeney (1973) includes a brief explanation of the phase sensitive detector (PSD) as used in a correlation mode. The subject warrants a complete text on the technique – here we can only lay the foundations leaving it to the reader to expand the mathematics. Phase sensitive detection is a basic tool in the instrument scientist's compendium. It is not new, having been used for close to a century in one form or another. Perhaps the easily 'understood' basic concept coupled with the reliability and power of phase sensitive detection systems has enabled widespread use without the need to resort to formulating a general theoretical approach that covers all situations.

Phase sensitive detection in practice is exemplified in the many uses of the modulated Wheatstone bridge principle (temperature – Bell and Hulley (1966), Larsen (1968), Tempest (1963); strain gauges – Neubert (1968); inductive displacement sensors – Golubtsov and Komarov (1968), Herceg (1972), Noltingk (1960a, b); capacitive displacement sensors – Jones and Richards (1973), Stacey *et al.* (1969); in radiometers – Bach *et al.* (1970), Bevan and Ricketts (1967), Pardy (1969) and in position sensitive photodetectors – Williams (1966)). It is also used in spectrophotometers, nuclear magnetic resonance, electron spin resonance, in optical interferometer fringe position monitors and in the control of gas absorption stabilized lasers.

4 Mathematical treatment

Phase sensitive detection requires the measurand to be modulated at a discrete frequency that is about ten times higher than the highest frequency required from the measurand. After initial amplification, and possibly wide-band filtering, the modulated signal is synchronously detected using the reference signal to form the product in a multiplier circuit. The basic phase sensitive detector can be regarded as a multiplier plus low-pass filter, as shown in figure 1.

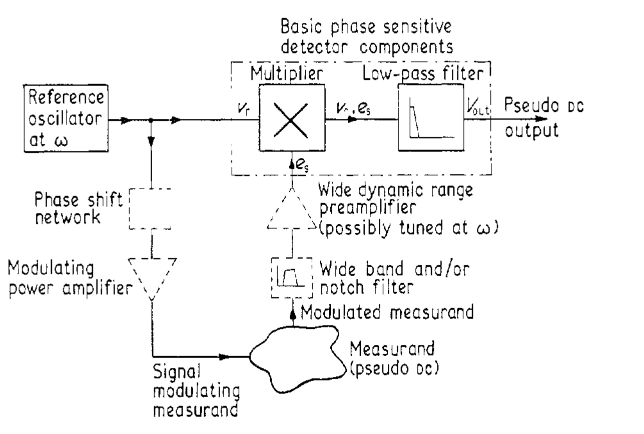

Figure 1 Schematic diagram of phase sensitive detection system. Dotted circuitry are refinements to basic system; they are often incorporated in commercial units

The first step is to form the product of the reference signal and the modulated measurand. The product is then re-expressed as an harmonic series from which we can easily extract the DC contribution. The higher frequencies, being removed by the low-pass filter, are of no consequence.

4.1 *Sinusoidal excitation*

We begin by considering the simplest system to analyse – which occurs when the reference signal and the modulated measurand signal are both sinusoidal waveforms of identical frequency. We will see later that this is not always the case.

Let the signal voltage be

$$e_s = E_s \sin(\omega t + \phi)$$

where ω is the frequency of modulation, t is the time and ϕ is the phase angle, and the reference signal be (this also provides the modulation drive)

$$v_r = V_r \cos \omega t.$$

The output of the multiplier (see figure 1) is given by

$$e_s v_r = \frac{E_s V_r}{2} [\sin(2\omega t + \phi) + \sin \phi].$$

The first term is an AC signal occurring at twice the modulation frequency, the second a DC term. A low-pass filter is used to remove the first, for it has no relevance in the detection process. The remaining term is the output of the phase sensitive detector. Hence

$$V_{out} = \frac{E_s V_r}{2} \sin \phi.$$

This is the basic equation of all PSD systems based on analogue signals – it expresses linearity and sensitivity. Assuming that the amplitude V_r of the reference signal is held constant there are two modes of operation for the system.

Firstly, if $\sin \phi$ remains constant (preferably at unity to maximize the signal amplitude) the detector produces a bipolar linear output related to the amplitude E_s of the modulated measurand. This is the amplitude detection mode; the response is shown in figure 2(*a*). This is useful for recovering small pseudo-DC amplitude effects as found in star photometry, infrared radiation detection, Wheatstone bridge output monitoring and spectrophotometry. (Literal use of the term *phase* sensitive detection can lead to confusion about its operation in this mode!)

Secondly, if the amplitude E_s of the measurand and V_r are held constant during modulation (fringe modulation is an example) the output will be related to the phase difference ϕ between the reference and the measurand signals. This is the phase detection mode (shown as figure 2(*b*)). This is used to

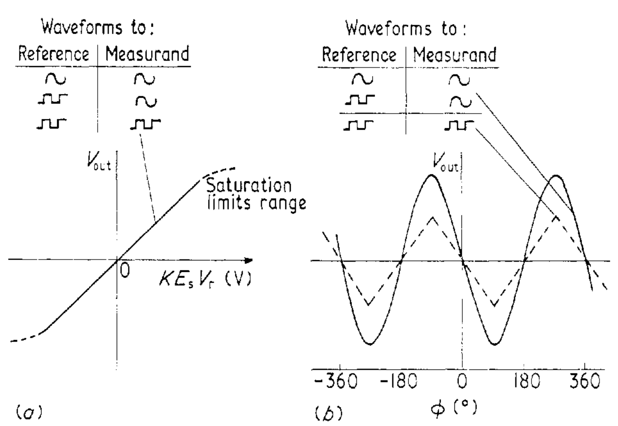

Figure 2 Generalized output responses of phase sensitive detector with various combinations of reference and measurand modulation waveforms. (*a*) Where $\sin \phi$ is held constant – amplitude detection. (*b*) Where KE_sV_r is held constant – phase detection. In general, detector equation is $V_{out} = KE_sV_r \sin \phi$

relate spatial position in photo-electric microscopes and fringe positioning in interferometers. Note that the output in this mode is not linear with input variable change, but sinusoidal. Linearity is reasonable, however, near the zero crossing. The output signal slope changes sign at 180° intervals. These basic factors are most relevant in closed loop mechanisms using the phase detection mode to create the feedback error signal. It will be seen later that the use of square waves overcomes the linearity difficulty but not the others.

4.2 *Rejection of unwanted frequencies*

It is interesting to see how the system handles signals in the measurand that are not exactly at frequency ω. Consider the measurand signal to be at frequency ω_1, instead of ω, retaining the reference signal at ω. The output before low-pass filtering is then:

$$\text{Output} = V_r \sin \omega t \times E_s \sin \omega_1 t \quad (\phi \text{ taken as } 90°)$$

$$= \frac{V_r E_s}{2} [\cos(\omega - \omega_1)t - \cos(\omega + \omega_1)t].$$

The low-pass filter satisfactorily rejects the second component leaving

$$V_{out} = \frac{V_r E_s}{2} \cos \Delta\omega t$$

where $\Delta\omega = \omega_1 - \omega$. Hence the output of the phase sensitive detector increases in frequency as the frequency of the unwanted signal deviates from ω but retains its amplitude. The low-pass filter response, however, is a gradual fall off with changing frequency so the net result is attenuation of frequencies away from ω giving the effective frequency response shown in figure 3(*a*) for each individual harmonic group present in the excitation signal.

The overall response of the phase sensitive detector is, therefore, equivalent to an extremely stable, very narrow

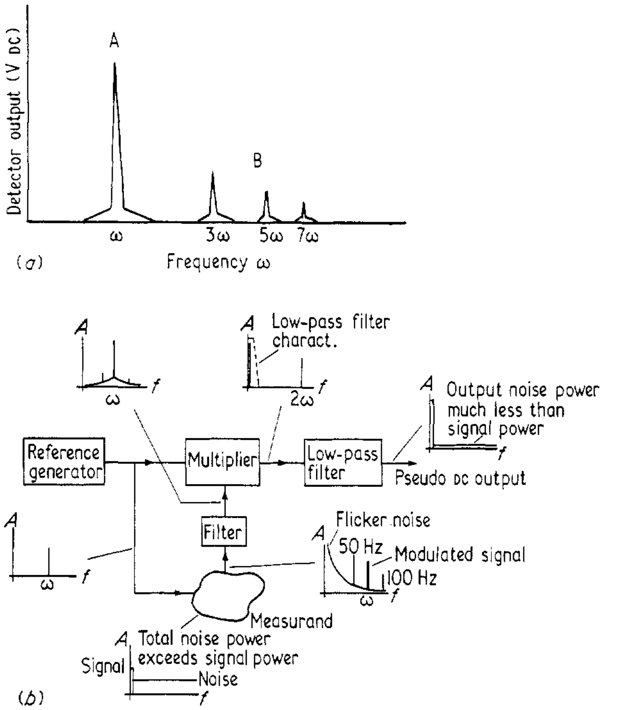

Figure 3 (*a*) Phase sensitive detector output depends upon frequencies present in inputs. A, response broadened by presence of range of frequencies in inputs; B, harmonic contributions provided by distortion in sinusoidal excitation or by use of squarewaves. (*b*) Frequency spectra at stages in phase sensitive detection system

bandwidth filter. For example, consider the simple capacitor–resistor RC low-pass filter that is usually used. This has a half-power bandwidth of

$$f = 1/2\pi RC = 1/2\pi\tau$$

where τ is the time constant. If $\tau = 10$ s the bandwidth from DC to the halfpower frequency is 0·016 Hz. With a modulating reference frequency of 1 kHz (typical values used lie in the range 1 Hz–1 MHz) the 3 dB power loss points will occur at 999·984 Hz and 1000·016 Hz. This represents a filter with a quality factor Q approaching 10^5. The only noise that passes through the detector is that occurring in the 32 mHz band. Provided the signal power exceeds the noise power in this narrow band a useful signal can be extracted. Further, discrete frequency noise such as induced mains supply interference, is effectively rejected provided it is not at the modulating frequency or at a subharmonic of it. Discrete frequency noise can be further reduced (the amplitude that can be eliminated can be gross) by suitably filtering the measurand signal (refer figure 1) before it enters the phase sensitive detector. Figure 3(*b*) shows the frequency spectrum existing at each stage of the basic detector when excited by a sinewave.

It is probably evident that the circuitry preceding the phase sensitive detection stage, and the stage itself, must possess an adequately wide dynamic range if the small signal is to be recovered without distortion.

4.3 *Use in closed loop applications*

Phase sensitive detectors are sometimes used to provide the corrective error signal in a closed loop feedback controller in the manner shown in figure 4. Examples are fringe-followers used in earth strain meters (Gerard 1969, Sydenham and Blair 1973), and optically controlled ruling machines; image

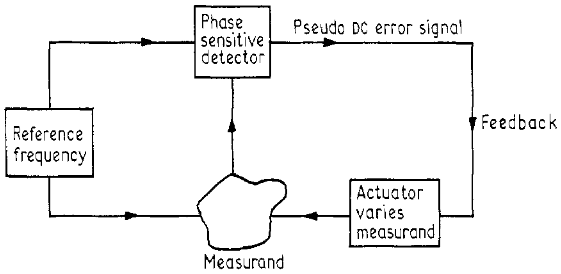

Figure 4 Schematic of feedback controller incorporating phase sensitive detection to provide the error signal

dissector, position sensitive detectors (Williams 1966) and in thermostats (Bell and Hulley 1966, Larsen 1968, Tempest 1963). In these cases it is necessary to be familiar with the limitations inherent in the phase sensitive detection process.

A difficulty is that the bandwidth has been deliberately limited to reduce noise. The response time of a servo mechanism, however, depends on the bandwidth so it is necessary to strike a compromise between tightness and sensitivity. (This is a basic axiom of instrument science – sensitivity and response rate are generally interchangeable parameters.) A servo employing this form of detection must be modulated at a frequency well above the necessary response time of the dominant time constants of elements forming the loop system. Each individual case must be studied on its own merits using servo theory and phase sensitive detection theory combined.

A second difficulty exists in servos employing the phase mode (output of detector proportional to $\sin\phi$). Here the slope of the detector output changes phase at 180° intervals producing positive feedback at certain phase values causing instability. Precautions must be taken to ensure that the phase swings are limited to within less than about $\pm 90°$ or, if unavoidable, that the gain is adequately attenuated as the error signal passes through the unwanted region. For example, in the fringe-follower reported by Sydenham and Blair (1973), their laser interferometer automatic fringe-jumping mechanism avoided this problem by applying an external drive to the servo, the phase sensitive detector output being disconnected as the fringes were forced through potentially unstable zero positions. This difficulty is not present in servos employing the amplitude detection mode; the signal merely saturates away from the null with the correct phase relationship.

4.4 *Response to reference signal harmonics*

In practice the reference signal will not be entirely pure and will have harmonic distortion present to some degree. Harmonics of the fundamental reference signal produce narrow response peaks centred at each harmonic frequency (as shown in figure 3(*a*)). They are, however, easily attenuated by the low-pass filter so the quality of the modulating signal is not important unless the noise reduction requirements are stringent.

4.5 *Square wave reference excitation*

If the reference waveform is a unity mark-space squarewave, i.e. symmetrical about zero in respect of both time and amplitude, and has zero to peak amplitude V_r then the output of the multiplier will be the additive terms of the various Fourier components. This squarewave can be expressed as:

$$v_r = \frac{4V_r}{\pi}\left[\cos\omega_r t - \frac{\cos 3\omega_r t}{3} + \frac{\cos 5\omega_r t}{5} - \ldots\right]$$

and with the measurand modulation as before

$$e_s = E_s \sin(\omega_s t + \phi)$$

the output before filtering will be

$$v_r e_s = \frac{2E_s V_r}{\pi}\left\{\sin[(\omega_s+\omega_r)t+\phi] + \sin[(\omega_s-\omega_r)t+\phi] - \frac{1}{3}\sin[(\omega_s+3\omega_r)t+\phi] - \frac{1}{3}\sin[(\omega_s-3\omega_r)t+\phi] + \frac{1}{5}\ldots\right\}.$$

If $\omega_s = \omega_r$, the only term remaining after low-pass filtering is the DC term giving

$$V_{out} = \frac{2E_s V_r}{\pi}\sin\phi.$$

Thus the squarewave excitation provides the same form of output as sinusoidal signals (shown in figure 2) but is 27% larger. This small advantage is somewhat offset, however, by the increased noise that will enter if the measurand signal has frequencies present at any of the harmonics of the squarewave ($3\omega_s$, $5\omega_s$, $7\omega_s$ etc.) for each of these is a noise acceptance window. The noise degradation is, however, not as great as might be thought. The noise power in each additive term, not being correlated with the signal and being random, adds power only as the square root of its magnitude in each term, whereas the signal gained adds arithmetically. The harmonic noise content can be reduced by passing the measured signal through a filter that rejects $3\omega_s$ and above before the product is formed in the multiplier.

There is, therefore, a choice to be made between the use of a squarewave or sinewave reference signal. In practice it is usually easiest to use the former: the amplitude V_r can be held more stable and the zero crossings are better defined.

4.6 *Squarewave measurand excitation*

If both inputs to the multiplier are squarewaves at the same frequency, each with amplitudes of E_s, V_r respectively, the DC output, after detection plus filtering, can be shown to be (Smith 1972)

$$V_{out} = \frac{E_s V_r}{\pi^2}\left(1-\frac{2\phi}{\pi}\right) \quad \text{for} \quad 0<\phi<\pi$$

$$= \frac{E_s V_r}{\pi^2}\left(\frac{2\phi}{\pi}-3\right) \quad \text{for} \quad \pi<\phi<2\pi$$

which is a linear waveform, shown dotted in figure 2(*b*). (The derivation used above and taken from Smith (1972) did not incorporate the fixed 90° phase shift and therefore produces an output that is maximum at $\phi=0$, not 90° as shown.)

Hence the use of squarewaves for the reference and the measurand modulation provides a linear output – not sinusoidal, but it still retains the difficulties mentioned in §4.3 when used in servo loops. Square wave modulation is, therefore, desirable with the phase mode of detection since the linear range is greatest. Often it is the most convenient to use, and especially in optical signal detection. It is the easiest to produce in many cases. A chopper disc, for example, modulates both the signal radiation and a reference beam at the same frequency. They are thus generated synchronously. In electrical forms of measurand (bridges etc.) mechanical switches such as vibrators, opto-electronic modulators or solid-state switches are used. Applications do exist, however, where it is not possible to use squarewaves, for the measurand may not respond fast enough to be squarewave modulated with adequate squareness.

Each of the above expressions enables the absolute values of output to be calculated – at least in theory. Current practice of not quoting the internal gain of the multiplier and filter used in commercial units does not, however, enable them to be combined with strict theory. It is necessary to calibrate the unit *in situ*. A statement (by the makers) in terms of V_{out} per $V_r E_s$ product and V_{out} per ϕ (with the product $V_r E_s$ quoted) for each combination of signal inputs would greatly aid system designers using commercial phase sensitive detectors.

5 Advanced application – fringe position measurement

So far here we have considered various cases in which the measurand is modulated to produce the same time function as its source of excitation. Some applications, for example detection of fringe position, involve a more complex situation in that the modulated measurand forms a quite different time function. Fringe-following, first, and then absorption-stabilized laser wavelength control have been chosen to illustrate how to handle this situation. The following treatment of sinusoidal interferometer fringe position monitoring is extracted from the approach by Gerard (1969a,b). Dyson (1970) and Sizgoric and Gundjian (1969) have also published accounts.

Fringes produced in a Michelson interferometer can be described by the intensity function

$$i = i_a + i_0 \cos(4\pi d/\lambda)$$

where i_a is the mean DC level of intensity, i_0 the amplitude, d the distance across the fringe and λ the wavelength (see figure 5). Let the fringe position detector (assumed to be seeing

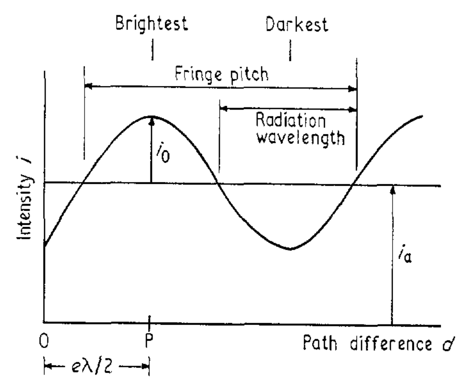

Figure 5 Intensity function for sinusoidal interferometer fringes

a point on the fringe) view the fringe at the relative displacement point P and the path difference OP be given by $1/2e\lambda$ where e is expressed as a fraction of a fringe. In the Michelson interferometer one fringe pitch corresponds to $\lambda/2$.

It is now necessary to modulate the detector position relative to the fringe position in some way if phase sensitive detection is to be used. Gerard (1969) did this by operating on the wavelength of his He–Ne cw laser by modulating the cavity length. The systems described by Dyson (1970), by Sizoric and Gundjian (1969) and Sydenham and Blair (1973) modulate the position of the reflector in the reference arm of the interferometer. Both approaches cause the fringes to vibrate across the input of the fixed detector.

The changing path difference produced by modulation (in Gerard's case) is given by

$$d=\frac{e\lambda}{2}+\frac{m\lambda}{2}\sin\omega t.$$

This produces a photodetector output of the form

$$i=i_a+i_0\cos\left[2\pi(m\sin\omega t+e)\right]$$

which drives the measurand input of the phase sensitive detector. We have previously established above that the phase sensitive detector performs an harmonic analysis of the measurand at the reference signal frequency. It is, therefore, appropriate in order to understand what is happening in the PSD output to extract the relevant harmonic components of the signal i by expressing it in a harmonic series form. This can be achieved using Bessel functions of the first kind.

Hence

$$I=i_a+i_0\{[J_0(2\pi m)+2(J_2(2\pi m)\cos 2\omega t+\ldots)]\cos 2\pi e$$
$$-2[J_1(2\pi m)\sin\omega t+J_3(2\pi m)\sin 3\omega t+\ldots]\times\sin 2\pi e+\ldots\}.$$

Extracting the coefficient of the sin ωt term (which is the role of the phase sensitive detector unit) we see that the output of the multiplier after filtering – the output of the PSD unit – is proportional to

$$-2i_0J_1(2\pi m)\sin 2\pi e.$$

In closed loop applications, such as those in which the fringe is held stationary by a displacement mechanism that counteracts the interferometer measuring arm changes in length, the path difference OP is held small giving $\sin 2\pi e\simeq 2\pi e$ and

$$V_{\text{out}}=-2\pi eJ_1(2\pi m).$$

This expression enables the optimum practical conditions to be achieved. They are quite unexpected.

Firstly, if e happens to be zero the output of the phase sensitive detector will be zero. This condition occurs when the photodetector sees the modulation across the maxima or minima of the intensity functions. Modulation produces a detector waveform at double the source frequency. Note that each half of the cyclic signal will be of equal amplitude if centred, and uneven if not. This feature can be used to set the fringe position (to within $10^{-3}\lambda$) without the need to use a phase sensitive detector – the waveform produced by the photodetector is simply viewed on an oscilloscope and adjacent amplitudes balanced to equality. For further explanation see Dyson (1970).

Secondly, it is essential to know the optimum modulation width that should be used, for a wrong choice can provide no output whatsoever. This is decided by studying the nature of the Bessel function $J_1(2\pi m)$. This looks like a damped sinusoid having alternate maxima and minima at $m=1{\cdot}8/2\pi$, $5{\cdot}3/2\pi$, $8{\cdot}5/2\pi$ etc. (0·286, 0·85, 1·36 etc. of a fringe in each case). The relative outputs of each are 1·00: 0·58: 0·48 showing that the optimum is a scan width of 0·386 fringe. Experience, however, has shown that 0·85 fringe modulation width also operates satisfactorily (Gerard 1969b). Certain scan widths ($m=0$, 0·61, 1·1 . . . of a fringe) produce no output for these are the zero crossings of the Bessel function.

In the above analysis the phase shift between the modulating and reference signals was assumed to be zero. If a phase shift ϕ does occur such as may happen across the modulating power amplifier stage, the phase sensitive detector output will be modified by a sin ϕ term. Hence, if this provides a shift of 90°, the output will be zero. A system should therefore have the phase shift adjusted upon assembly to produce a maximum signal.

6 Dealing with intensity distribution functions in a general manner

In §5 the method was developed to handle measurands with sinusoidal intensity variation. Many processes needing phase sensitive detection, however, have other functional relationships that do not enable such a specific mathematical approach to be used. One example is Fabry–Perot interferometer fringes (Dyson 1970). Another is the intensity characteristic for a laser cavity – in this section we look at a generalized procedure that handles all cases of intensity function.

The same fundamental approach can be used as with earlier discussions – form the product of the reference signal and the modulated measurand, extract an harmonic series representation and identify the DC components.

To illustrate the generalized approach for cases where the intensity function is more complicated than sinusoidal, we now study an example in which the length of a CW laser cavity, incorporating a gas absorption cell, is servo-locked on to very small amplitude intensity changes. These changes occur at certain highly specific wavelengths and therefore form the basis of an improved length standard.

We must first understand the practical background of the technique. When the length of the cavity of a laser incorporating an absorption cell is varied from one extreme to the other the output radiation intensity varies in a manner depicted in figure 6. The depression in the Doppler profile curve is the

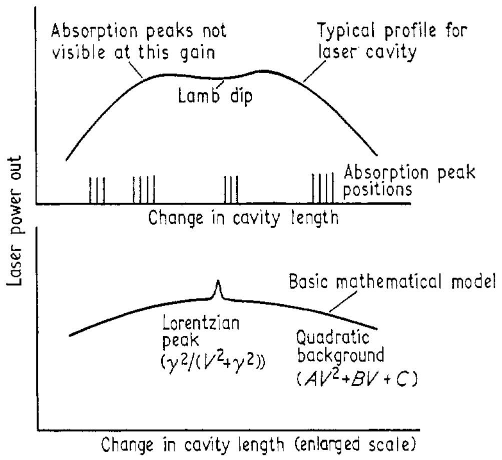

Figure 6 Output power variation of a CW laser cavity (that contains iodine vapour absorption cell) as the cavity length is altered

Lamb dip, which results as the influence of the whole cavity system. Small peaks (in reality barely visible on the scale shown) occur on this profile as the result of the selective absorption phenomenon of the gas in the absorption cell.

Early laser wavelength controllers locked on to the bottom of the Lamb dip profile (see Gerard 1969), but developments now enable us to lock the wavelength on to one of the absorption peaks thus increasing the wavelength stability by many orders of magnitude over Lamb dip methods of frequency stabilization.

To investigate the feasibility of servo-locking on to such fine peaks when they have such small amplitude and a nonconstant background, it is necessary to produce an equation describing the intensity characteristic. If it is assumed that lock is achieved the curve can be regarded as a background quadratic to which is added a sole absorption peak. It is known that the absorption peak characteristic is of lorentzian profile.

The net result is an intensity function of the form (see Wallard 1972)

$$F(V)=\gamma^2/(V^2+\gamma^2)+AV^2+BV+C,$$

where V is the output voltage produced by a photodetector for a cavity scan with zero modulation; γ is the full width at half maximum, FWHM, of the lorentzian curve. In the presence of modulation the photodiode output becomes $F(V+m\sin\omega t)$, where m is modulation depth.

To obtain suitable servo-error signals the phase sensitive detector output must cross zero at the maximum of the lorentzian profile.

A normal phase sensitive detection process (reference at the same frequency as the impressed modulation on the measurand) could not provide a lock on to the maximum of the absorption peak, for the latter can lie virtually anywhere on the background curve depending upon manufacturing parameters. It would attempt instead to home to the dip of the Lamb dip.

If, however, the phase sensitive detection is performed at a frequency of $n\omega$ then the detector output is essentially the nth derivative of the $F(V)$ (see Wallard 1972). Hence for phase sensitive detection at the modulation frequency the background contribution is $2AV+B$, which does not, in general, provide a phase sensitive detector output passing through zero at the maximum of the lorentzian profile. If the peak lies on the top of the general cavity profile, first derivative locking is possible as has been demonstrated by Levine and Hall (1972) in their earth strain meter. Going one step further, double differentiation provides no central polarity change at all and also has the undesirable DC offset of $2A$. Differentiating again (triple differentiation), however, provides a suitable control signal that has a slope passing through zero; the background quadratic contribution is completely eliminated, leaving only the harmonic content of the lorentzian signal as shown in figure 7.

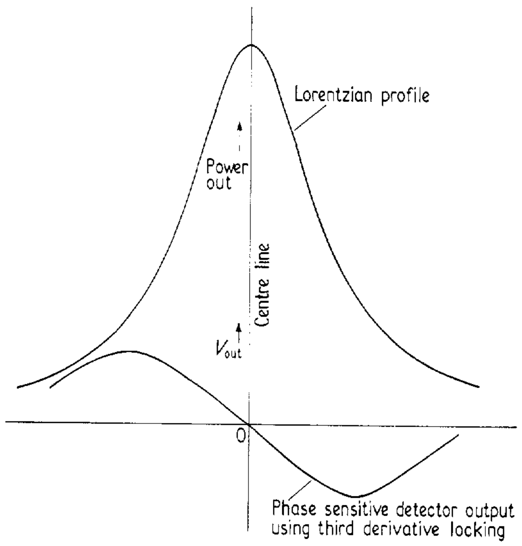

Figure 7 Phase sensitive detector output produced by third derivative locking on to modulated absorption-stabilized laser cavity

Triple differentiation is easily performed in practice by using phase sensitive detection which has a reference signal at 3ω and cavity modulation at ω (provided the latter contains negligible distortion). This recently introduced technique is now commonly known as 'third derivative locking' and is used by all groups developing iodine-stabilized lasers.

A theoretical derivation of the phase sensitive detector output in this case has been developed. Wallard (1972), for instance, expanded the intensity function in terms of a Taylor series in order to extract magnitudes of the harmonics but this is applicable for small modulation widths only. Since then a more general method (Blair 1974) has been reported. After forming the product of the intensity function with the sinusoidal reference signal it is expanded, for first derivative locking (neglecting the quadratic background contribution), as

$$V_1(x, m)=-2\Sigma_n A_n \sin(nx)\mathrm{J}_1(nm).$$

For third derivative locking the method gives

$$V_3(x, m)=2\Sigma_n A_n \sin(nx)\mathrm{J}_3(nm)$$

where A_n are the Fourier coefficients of the unmodulated intensity function (in the lorentzian case $A_0=0{\cdot}402$, $A_1=0{\cdot}389$, $A_2=0{\cdot}128$ etc.) and $n=0$, 1, 2, 3, etc., m is the fraction of modulation width of a FWHM$=1{\cdot}0$.

Using the above approach, it is possible to predict the phase sensitive detector output in the presence of harmonic distortion in the modulation. The theory can also be used to ascertain the output under the conditions when two or more lorentzian peaks exist close together.

Closer definition of this problem is enhancing the reproducibility of the laser wavelength showing that it is sophisticated phase sensitive detection that has made the new, greatly improved, length standard a reality. To the authors' knowledge nth derivative locking has not yet been used elsewhere in other applications.

7 Summary

We have seen that the phase sensitive detection process consists basically of a multiplier followed by a low-pass filter. This fundamental approach enables the behaviour of any phase sensitive detection system to be studied theoretically. Explanatory articles have appeared that are based on the use of waveform explanations but whilst these help the practitioner to some degree they do not enable the system to be optimized. Complicated situations cannot be modelled in terms of graphical explanations.

Space has permitted only an overview – no attempt has been made here to describe practical circuitry. For practical realization see Clayton (1973), Faulkner and Harding (1966), Jones and Richards (1973), Linsley (1970), Smith (1972) and Williams (1965).

References

Bach J H, Street P J and Twamley C S 1970 *J. Phys.* E: *Sci. Instrum.* **3** 281–6

Bell E C and Hulley L N 1966 *Proc. IEE* **113** 1671–7

Bevan H and Ricketts R E 1967 *J. Sci. Instrum.* **44** 1048–50

Blair D P 1974 *Appl. Phys. Lett.* **25** 71–3

Brower R 1968 *Electronics* **40** 80–90

Champeney D C 1973 *Fourier Transforms and their Physical Applications* (London: Academic)

Clayton G B 1973 *Wireless World* **79** 355–6

Coor T 1972 *Electron. Today Int.* (Aust. Edn) **1** 30–8 and 115

Danby F C G 1970 *Electron. Engng* **42** 36–41

Dyson J 1970 *Interferometry as a Measuring Tool* (Brighton: Machinery Publishing Co)

Faulkner E A and Harding D W 1966 *J. Sci. Instrum.* **43** 97–9

Gerard V B 1969a *J. Phys.* E: *Sci. Instrum.* **2** 749–50

Gerard V B 1969b *J. Phys.* E: *Sci. Instrum.* **2** 933–5

Golubtsov A I and Komarov V V 1968 *Geod. Aerophotogr.* **1** 53–6

Herceg E E 1972 *Handbook of Measurement and Control* (Pennsauken, New Jersey: Schaevitz Engineering)

Jones R V and Richards J C S 1973 *J. Phys.* E: *Sci. Instrum.* **6** 589–600

Larsen N T 1968 *Rev. Sci. Instrum.* **39** 1–12

Levine J and Hall J L 1972 *J. Geophys. Res.* **77** 2595–609

Linsley Hood J L 1970 *Electron. Engng* **42** 62–3

Moore R D and Chaykowsky O C 1963 *Princeton Applied Res. Corp. Tech. Bull.* 109

Neubert H K P 1968 *Strain Gauges – Kinds and Uses* (London: Macmillan)

Noltingk B E 1960a *Instrum. Pract.* **14** 749–50

Noltingk B E 1960b *Br. Commun. Electron.* **7** 430–5

Pardy D 1969 *J. Phys.* E: *Sci. Instrum.* **2** 526–30

Sizgoric S and Gundjian A A 1969 *Proc. IEEE* **57** 1313–4

Smith K L 1972 *Wireless World* **78** 367–70

Stacey F D, Rynn J M W, Little E C and Croskell C 1969 *J. Phys.* E: *Sci. Instrum.* **2** 945–9

Sydenham P H and Blair D P 1973 *Proc. Laser Applications Conf.* (Melbourne: Ian Clunies Ross Memorial Foundation)

Tempest W 1963 *Electron. Engng* **35** 814–6

Wallard A J 1972 *J. Phys.* E: *Sci. Instrum.* **5** 926–30

Williams P 1965 *J. Sci. Instrum.* **42** 474–6

Williams T L 1966 *Instrum. Rev.* **13** 88–90, 150–2, 189–92