New Mode of Operation of a Phase Sensitive Detector*

YARDLEY BEERS
National Bureau of Standards, Boulder, Colorado
(Received 24 August 1964; and in final form, 15 January 1965)

The PS detector plays an important role in the instrumentation for spectroscopy and radio astronomy and in atomic frequency standards. Usually the phase sensitive detector is preceded by a narrow-band-pass amplifier centered on the reference frequency or one of its harmonics. Such a procedure discards useful information contained in the other harmonics. It is suggested that the narrow-band amplifier be replaced by a wide-band one and that the reference waveform be shaped to be identical with that of the desired signal. First order calculations show that this method gives slightly better sensitivity than the conventional one. Also, discrimination against coherent signals of other waveforms can be obtained. If a differentiating circuit is inserted into either the input or the reference circuit, the detector output is an odd function of the phase angle. Therefore, this arrangement can be used as a sensor for a servo system used to lock the phase of the reference oscillator to the signal.

THE type of detector[1,2] which allows the comparison of a signal to a reference voltage is a device used to improve the sensitivity in such diverse fields as microwave spectroscopy and radio astronomy over what is obtainable with a system where no reference voltage is used. It may be used to generate an error signal as part of a servo system for stabilizing the frequency of an oscillator to the frequency of another oscillator. In the conventional mode of operation this detector is used in conjunction with a narrow-band amplifier tuned to the reference frequency. The purpose of this paper is to consider the advantages of another mode of operation where a wide-band amplifier is used and the waveform of the reference voltage is accurately controlled to be identical with that of the desired signal. It is shown, subject to some assumptions to be stated, that this mode in principle provides a slight improvement in signal-to-noise ratio and, in addition, provides suppression of coherent spurious signals such as the background signal in a frequency-modulated microwave spectrometer resulting from impedance mismatches in the rf system. However, in the cases which have come to the author's attention, these advantages are small in practice and may be masked by effects which are neglected in the theory to be presented. The principal apparent advantage of the mode concerns a modification which is appropriate to the problem of frequency stabilization, henceforth called "the phase servo problem." Therefore, the first part of this paper, which contains most of its length, is in practice to be considered as background material for the treatment of the phase servo problem.

The type of detector which allows the comparison of a signal voltage with a reference voltage is referred to by various names: "phase sensitive detector," "synchronous detector," and "lock-in amplifier." Also, less frequently, the terms "coherent detector" and "discriminator" have been used, although the latter term is much more widely applied to circuits of other types. There seems to be a lack of agreement concerning these terms. In this paper "phase sensitive detector" is intended to connote the most general term which includes all applications wherein a signal voltage is compared to a reference signal. In the author's opinion, a "synchronous detector" connotes the particular situation wherein the reference voltage is constant in magnitude but periodically is reversed in polarity (i.e., is a square wave) and the circuit is arranged in such a way that the output is independent of the amplitude of the reference voltage. In most cases, the constancy of the reference voltage in a synchronous detector is obtained by limiting action in the same nonlinear device which gives rise to the detecting action. The author regrets the existence of the term "lock-in amplifier" as being misleading and wishes that its use could be avoided.

The approach of this paper is that of a first order theory. It is supposed that the output current of the nonlinear device which serves as the heart of the detector is expressible by Taylor series as a function of the signal and reference voltages. If these voltages vary in time, they and the output current can be represented as Fourier series. However, the Fourier coefficients of the output current are individually given as functions of the Fourier coefficients of the voltages and of the Taylor coefficients. In the present approach all of the Fourier terms in the output are neglected except those based upon the lowest order nontrivial Taylor coefficient, which paradoxically, in the case of a two terminal nonlinear element, is the coefficient of the second order (quadratic) term. Then the present theory is not valid, except as a rough approximation, for a synchronous detector. The approximation is probably much poorer when the amplitude of the reference voltage is controlled by saturation of the nonlinear element than when it is controlled externally.

Generally, the signal voltage has a complex waveform: that is, has more than one finite Fourier amplitude. With the conventional mode, which uses a narrow-band filter,

* A portion of this work was performed while the author was at New York University and working under contract with the U. S. Office of Naval Research.
[1] C. H. Walter, Z. Tech. Physik 13, 363 (1932).
[2] C. R. Cosens, Proc. Phys. Soc. (London) 46, 818 (1934).

ideally only one of the signal components is accepted, and thus useful information contained in the signal is discarded. The new mode uses all of the information contained in the signal. To do this, however, it is necessary that the reference voltage contain all of the Fourier components that the signal does. On the other hand, if the reference voltage contains Fourier components not contained in the signal voltage it needlessly sensitizes the detector to produce output for signal frequencies which are of no interest and thus makes the detector unnecessarily vulnerable to interference and noise. With the conventional mode, this interference ideally is eliminated by the narrow-band filter, but with actual, nonideal filters and very strong interfering signals, trouble could result from including unneeded Fourier coefficients in the reference voltage. Accordingly, the common practice of using a narrow-band filter with a synchronous detector seems obviously illogical and undesirable. If a narrow-band filter is used, the reference voltage should have a sinusoidal waveform. The treatment which follows strongly implies, but does not rigorously prove for all cases, that optimum performance is obtained when the waveforms of the signal and reference voltages are identical, regardless of whether the conventional or the new mode is used. In the particular case of the Dicke radiometer, the advantage of matching waveforms has been demonstrated by the calculations of Johnson.[3]

THE DETECTION PROBLEM USING THE CONVENTIONAL MODE

The conventional mode is illustrated by Fig. 1. The word "phenomenon" refers to a spectral line, a radio star, or something else which is to be observed. The phenomenon is modulated at an angular frequency ω. In microwave spectroscopy the phenomenon itself is modulated by applying a Stark or Zeeman field to the sample giving rise to the line. In radio astronomy, however, the primary effect is not directly accessible for modulation, but modulation is accomplished in the radio receiver which serves as the first detector by switching the frequency from that of the source to some other one or by switching the input from the antenna to a resistor. In these applications to radio astronomy, the principal purpose of the phase sensitive detector is to eliminate fluctuations in gain rather than to improve the signal-to-noise ratio.

The filter selects the k'th harmonic (most often $k=1$) so that the input to the detector is equal to $A_k \cos(k\omega t-\varphi_k)$. Generally the reference voltage is represented by a Fourier series, but if the signal input is confined to the k'th harmonic, only the k'th harmonic of the reference signal contributes to the dc output of the detector. The average output current then is, according to the first order theory,

$$i=gA_kB_k\cos(\varphi_k-\varphi_k'), \tag{1}$$

[3] W. A. Johnson, Proc. IEEE 52, 1242 (1964).

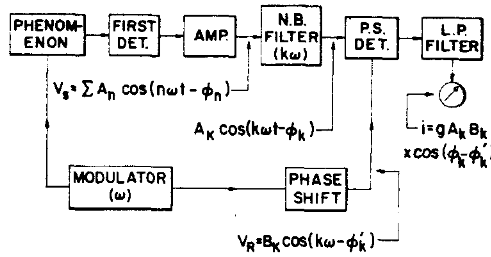

FIG. 1. Block diagram of conventional mode.

where B_k and φ_k' are, respectively, the amplitude and phase angle of the k'th harmonic of the reference voltage and g is a constant. Usually the phase shifter is adjusted to make the cosine factor unity when the phenomenon is at its peak.

In principle, the output consists not only of a steady current as given by Eq. (1) but also of fluctuations arising from noise produced in the first detector or in its input. The proper fluctuation output can be considered as being due to noise voltage components at the input of the phase sensitive detector beating against the reference voltage. However, the phase sensitive detector employs a nonlinear element, which also can act as a rectifier. Therefore, there is also a fluctuating output resulting from noise components beating against one another. According to first order theory, increasing the reference voltage increases the signal output and the proper noise output in the same ratio while leaving the noise output due to rectification unchanged. Therefore, proper operation requires that the reference voltage be large enough that the proper noise output be large compared to the rectified noise output, but this requirement is usually satisfied automatically whenever B_k is large compared to A_k. Usually this condition can be satisfied easily. There is, of course, an upper limit to the magnitude of the reference signal, which results in saturation effects and the failure of first order theory. However, under normal conditions, the band width of the low pass filter at the output is considerably smaller than that of the narrow band filter at the input, the output signal-to-noise ratio is independent of the width of the narrow band filter.

Another consequence of the nonlinearity is the generation of harmonics of both signal and reference voltages. For simplicity, in this paper the effects of these harmonics are neglected. Many detectors employ balanced detectors. With these, the effects of rectification noise and harmonics are eliminated or reduced.

One function of the narrow band filter, then, is to choose the harmonic of the input voltage which is to give rise to the output. As far as this purpose is concerned, the narrow band filter could just as well have been placed in the reference circuit as in the signal circuit, and if either the signal or reference signal or both should consist essentially only of the k'th harmonic, this filter could be omitted. However,

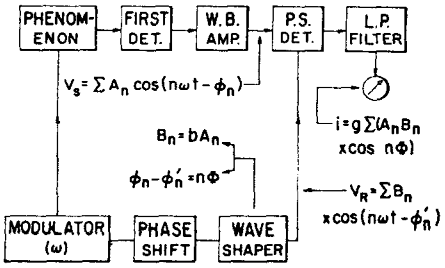

FIG. 2. Block diagram of new mode. The narrow-band filter has been removed, and a wave-shaping circuit has been introduced into the reference line to make its waveform identical with that of the signal.

there are some secondary advantages to having this filter. In the first place, it is useful in discriminating against spurious signals, especially if reference voltage harmonics other than the k'th have appreciable amplitudes.

Secondly, by limiting the predetection bandwidth, the filter reduces the number of noise components which beat together to give noise output by rectifier action. Often when the device is used with marginal signals, there is danger of overloading the amplifier by noise. In Fig. 1, for simplicity, the filter is shown at the output of the amplifier. In practical design it usually is incorporated in one of the interstage coupling networks. Then it performs the very important function of reducing the danger of overload in the stages which follow.

To satisfy these purposes, it is merely necessary that this filter have sufficient selectivity to reject adjacent harmonics. In particular, it is undesirable to attempt to achieve an unnecessarily small predetection bandwidth by cascading a large number of filter sections since a small fluctuation in the frequency ω causes a large shift in phase and results in difficulty in keeping the phase shifter set to its optimum position.

THE DETECTION PROBLEM USING THE NEW MODE

The previous discussion points out the fact that the narrow-band filter is not fundamental to the operation of the detector but greatly assists it. It also implies that useful information contained in the other harmonics of the signal is being discarded. (It should be remarked in passing, however, that Shimoda,[4] in a circuit for stabilizing a laser, has used several detectors, each with a filter tuned to different harmonic, to utilize the available information more fully.) Accordingly, in the new mode, the narrow-band filter is omitted; and the reference voltage waveform is controlled to be the same as that of the signal, for then the contributions of all the harmonics add constructively at the detector output. To show this, it is assumed for the moment that they are of different waveforms and that the voltages are given, respectively, by

$$V_S = \sum A_n \cos(n\omega t - \varphi_n) \tag{2}$$

and

$$V_R = \sum B_n \cos(n\omega t - \varphi_n'), \tag{3}$$

where n ranges from one to infinity. (It is assumed that the dc components are blocked out.)

Each harmonic contributes to the average output current in accordance with Eq. (1), and the total average output is given by the sum of such terms,

$$i = g\sum A_n B_n \cos(\varphi_n - \varphi_n'). \tag{4}$$

If the signal and reference voltages have arbitrary waveforms, generally some of the cosine factors are positive and some are negative, and there is a tendency for the terms to annul one another. However, if the waveforms are identical,

$$B_n = bA_n, \tag{5}$$

and

$$\varphi_n - \varphi_n' = n\Phi, \tag{6}$$

where b and Φ are independent of n. Then, if the phase shifter is adjusted to make Φ equal to zero, the cosine factors all become plus one and the terms add constructively to give

$$i = bg\sum A_n^2. \tag{7}$$

Suppose that there is present a coherent spurious signal with the same fundamental frequency but with a different waveform. Such a situation exists in a frequency-modulated microwave spectrometer. Ideally, in the absence of a sample, the output is frequency independent, and no modulated output should be present. In practice, however, this condition is never fulfilled, and there is an output which can mask the effect of the sample, if the effect of the sample is small. This voltage may be written in the form of Eq. (2) as

$$V_D = \sum D_n \cos(n\omega t - \varphi_n''), \tag{8}$$

and it gives an output of

$$i_D = bg\sum A_n D_n \cos(\varphi_n - \varphi_n''). \tag{9}$$

The suppression ratio S may be defined as i/i_D. Then

$$S = \sum A_n^2 / \sum A_n D_n \cos(\varphi_n - \varphi_n''). \tag{10}$$

One possible advantage of the new mode can be seen by examination of Eq. (10). By proper selection of the type of modulation it may be possible to cause some of the cosine factors to be positive and some negative so that the terms of the denominator may combine destructively while, as has been pointed out, those in the numerator add constructively. Thus the suppression ratio may exceed the value it would have with the conventional method, where there would be only one term in numerator and denominator.

[4] K. Shimoda, Buteri 19, 120 (1964). Also 14th General Assembly of URSI, Tokyo, September (1964).

This advantage is of little interest to potential users unless it can be shown that the signal-to-noise ratio is at least as good as with the conventional mode. Therefore, this ratio will be computed now.

For the moment, it will be supposed that there has been inserted into the reference circuit a filter which passes only the k'th harmonic, and it is assumed that the system is operated in such a way that noise contributions to the output due to rectification can be neglected. Then the noise bandwidth f of the system is determined by the low-pass filter at the output. (Actually f is twice the noise bandwidth of the low-pass filter since a detector with such a filter of width $f/2$ accepts noise from a band of width f centered at $(k\omega)/(2\pi)$ after beating with the reference voltage.) The rms equivalent noise voltage at the input of the first detector is given by

$$v_N = (4F\mathrm{k}TRf)^{\frac{1}{2}}, \tag{11}$$

where F is the noise figure of the system, k is Boltzmann's constant (not to be confused with the index k) and T is the reference temperature. R is the resistance associated with the first detector input. The noise voltage at the input of the phase sensitive detector is found by multiplying this by the gain between the two detectors, G, which is assumed to be frequency independent. The rms output noise current i_{Nk} is found by replacing A_k in Eq. (1) by this voltage and by replacing the cosine factor by its rms value $(\frac{1}{2})^{\frac{1}{2}}$. With the use of Eq. (5) this becomes

$$i_{Nk} = bgGA_k(2F\mathrm{k}TRf)^{\frac{1}{2}}. \tag{12}$$

If the filter in the reference circuit is removed, each harmonic of the reference voltage makes a contribution given by Eq. (12). These contributions are independent. Therefore the total noise current i_N is found by taking the square root of the sum of these squares of the individual contributions. The output current signal-to-noise ratio N is found by dividing Eq. (7) by i_N. In making the computation it is convenient to define a quantity, which has the dimensions of power

$$P = (\sum A_n^2)/(RG^2), \tag{13}$$

in which the index k has been replaced by n. If this calculation is made, after some cancellation, it is found that

$$N = (P/2F\mathrm{k}Tf)^{\frac{1}{2}}. \tag{14}$$

This result, of course, is based upon the assumption, usually valid, that noise generated in the reference circuit can be neglected. The derivation of the signal-to-noise ratio for the conventional mode is identical except the summations reduce to a single term. Therefore Eq. (14) applies to both modes.

The quantity P can be recognized as the total power dissipated in the input circuit of the first detector associated with the *modulation of* the primary phenomenon. Therefore, the signal-to-noise ratio depends upon how completely the phenomenon can be modulated. With the conventional mode, to obtain the maximum sensitivity, it is necessary to completely modulate the phenomenon and yet throw all of the power into a single harmonic. With the new mode it is merely necessary to completely modulate the phenomenon, but it is unimportant how the power is distributed among the harmonics. Generally it is impossible to obtain a high modulation index and confine the power entirely to a single harmonic. Therefore, it can be expected that the new mode is capable of giving at least as good a signal-to-noise ratio and probably a somewhat better one than the conventional mode, although with most waveforms, any such improvement would be small.

In practice, it may be difficult or impossible to achieve this improvement in signal-to-noise ratio. It should be remembered that this is a first order theory which neglects overload effects. If the amplifier has sufficient bandwidth to pass several harmonics, these effects are likely to be severe, while, as has been mentioned previously, with the conventional mode the inclusion of the filter greatly reduces the danger of overload.

Whether the new mode is to be preferred to the conventional one can be determined only by consideration of details in individual cases. The new mode is unlikely to be preferred in situations where there are no coherent spurious signals. The new mode is more likely to be useful in the phase servo problem, which is discussed herewith, than with the detection problem.

THE PHASE SERVO PROBLEM

In the previous discussion it has been assumed that the objective has been to detect the presence or absence of some weak phenomenon, and the use of a phase sensitive detector greatly increases the sensitivity of the process. There is another type of application in which the phase sensitive detector is very useful. This is when it is desired to synchronize the frequency of an oscillator to a spectral line, as in the case of a cesium beam frequency standard, or to synchronize the phase of an oscillator with some received signal of high stability, as might be the case with the reception of LF or VLF standard frequency and time radio signals.

In the following discussion it will be supposed that the oscillator is connected to a wave shaping circuit which generates the same waveform as that of the signal from the source of stabilization. Equations (4) and (6) suggest that, if the oscillator is connected to the reference input and the stabilization signal is connected to the signal input of the detector, a maximum output is obtained when the two are in phase. However, in a servo system for correcting the phase of the oscillator, it is necessary to generate an error voltage which is zero when the phase angle Φ is zero and which is an odd function of Φ. It will now be shown that such a voltage can be obtained if there is inserted in either the signal or reference line a circuit whose output is proportional to the time derivative of its input.

For the sake of definiteness it is assumed that this is placed in the signal input. Then, by the use of Eq. (2), the signal input becomes

$$V_s' = -\sum n\omega A_n \sin(n\omega t - \Phi_n) = \sum n\omega A_n \cos(n\omega t - \Phi_n + \pi/2). \tag{15}$$

Then the output current i' is given by an expression which is analogous to Eq. (7),

$$i' = -bg\sum(n\omega A_n^2 \sin n\Phi). \tag{16}$$

Equation (16) shows that i' has the desired property of being an odd function of Φ. In contrast to Eq. (7) it is to be noted that the terms that are being summed in Eq. (16) contain a factor of n. Thus the higher harmonics play a much more important role in the phase servo problem than in the detection problem.

It can usually be assumed that noise in the oscillator circuit can be neglected. A differentiating circuit is one whose amplitude response increases with frequency. Therefore, the gain G is no longer frequency independent. As an approximation, it may be assumed that the gain at the n'th harmonic is given by

$$G_n = G_0 n. \tag{17}$$

Then by analogy to Eq. (12) and by reference to the paragraph which follows it, the noise output current is

$$i_N' = bG_0(2FkTRf)^{\frac{1}{2}}(\sum n^2 A_n^2)^{\frac{1}{2}}. \tag{18}$$

Again the importance of the higher harmonics is to be noted.

The factors b and G_0 are arbitrary as far as the present discussion is concerned, but it is to be recognized that there exist practical limits to their values. A discussion of these limits is beyond the scope of this paper. If the differentiating circuit is placed in the oscillator line, it can be shown easily that forms identical with Eqs. (16) and (18) are obtained except that the gain in the signal channel G appears as an arbitrary factor. Thus, as far as these first order considerations are concerned, it is immaterial where the differentiating circuit is located. However, when practical limits upon the values of these arbitrary factors are considered, a preference may be indicated. Also, if a coherent spurious signal is present, a greater suppression ratio may be obtainable with one choice than with the other.

The signal-to-noise ratio may be computed by dividing Eq. (16) by Eq. (18). In this case, algebraic cancellation does not simplify the expression to allow the substitution of the power P. Therefore, the explicit relation will not be given here.

Noise can be considered to be equivalent to a fluctuation in phase angle. The rms value may be determined by substituting numerical values of the coefficients in Eqs. (16) and (18), equating i' and i_n', and solving the resulting transcendental equation. By doing this for several alternative types or degrees of modulation, the optimum choice can be determined.

ACKNOWLEDGMENT

The writer wishes to thank Neil T. Larsen for helpful discussions.