The Ubiquitous Phase Sensitive Detector

Applications and principle of operation

by K. L. Smith, Ph.D.*

One might ask wherein lies the ubiquity of the phase sensitive detector circuit, considering that in many cases a mumble of "I've never used one" might be heard. One reason is that it parades under many a pseudonym and perhaps a list of the more common might ring a mental bell or two—phase sensitive rectifier, homodyne, synchrodyne, product detector, synchronous rectifier, correlation detector, coherent detector, etc. The number and variety of applications in which phase sensitive detectors (p.s.ds) turn up are also pretty widespread. The use of product detectors in single-sideband receivers is well known. Here the carrier is reintroduced to the sideband and audio outputs are obtained by multiplying together this locally-generated carrier and the sideband components. Actually, the precise phase of the reference carrier compared to the original at the transmitter is not as critical as it would have to be with double-sideband suppressed-carrier signals. In the latter case a true p.s.d. would be needed where the phase of the reference would require careful control.

This brings us to the synchrodyne receiving method—a technique not new to *Wireless World* readers.[1] A true p.s.d. is used to demodulate an a.m. r.f. signal. The necessary reference signal with the correct phase is obtained from the carrier by phase locking a local oscillator to it. Bandwidth filtering is obtained after the detection process. (This property of being able to set bandwidths after the detector is a big advantage of the technique.)

There has been a rise into prominence of phase-locked loops in electronic circuitry. An example of this was mentioned above in connection with the synchrodyne. As another example a spot frequency synthesizer can be constructed giving crystal-controlled frequency stability at, say, 30 frequencies in the short-wave band but employing only one crystal. The block diagram of Fig. 1 shows how this is achieved.

The voltage-controlled oscillator is phase locked to each harmonic of the 1-MHz crystal. By applying the two signals to the p.s.d., a direct control voltage is produced proportional to the phase difference between the oscillator and crystal oscillator harmonic, at least for small phase differences. Thus when the oscillator is tuned to lock in at each harmonic, the output is frequency stable—more than that, it is *phase* stable with respect to the crystal.

The full analysis of the operation is complex. The time constant of the d.c. amplifier can be critical; so is the lock-in range. This kind of loop stability problem is common to all negative feedback loops and correct design requires some thought.

In physics laboratories the p.s.d. turns up in scanning spectroscopy. A beam of visible or infra-red radiation being analysed is interrupted by a chopper blade to produce an alternating signal, then amplified and passed to the p.s.d. A reference signal, whose phase can be varied, is obtained from the chopped beam and fed into the other terminals. This method is referred to by physicists[2] as a 'homodyne amplifier' system. The great advantage is that the effective noise bandwidth can be set by filtering after the detection process, a considerably easier job than trying to use selective audio amplifiers before the detector. (Just consider the difficulties with tuned amplifiers of 2-Hz bandwidth if the chopper frequency tends to drift.)

A big improvement in the ability to detect small signals became available for radio astronomical observations when Dicke published his radiometer scheme. This was based on alternately switching the aerial and a standard noise source into the receiver and switching a detector after the video amplifier in step with this, i.e. a p.s.d., to produce a resultant output—Fig. 2. The channel containing the standard noise source produces a known 'temperature' and the other channel can be some cosmic source being picked up by the aerial. If gain drifts slightly, both channels are affected accordingly, but the relative measure remains the same. Again, system bandwidth can be set after the detector.

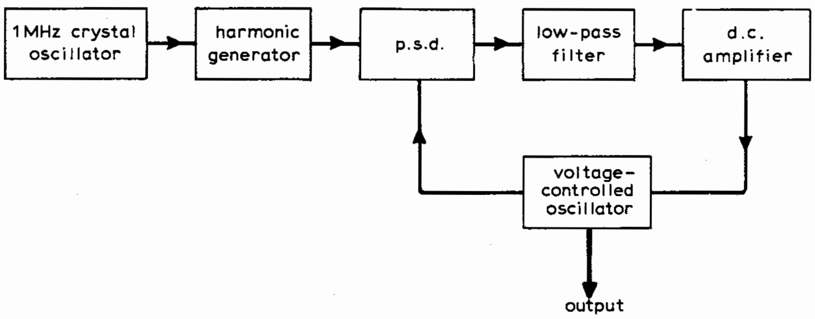

Fig. 1. In some frequency synthesizers, an oscillator is phase locked to a particular harmonic of a crystal oscillator according to the frequency required.

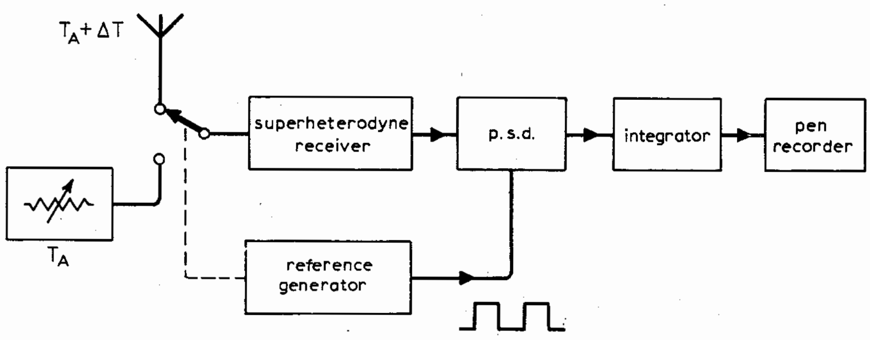

Fig. 2. Although sensitivity is reduced in the Dicke radiometer system, the gain in stability more than compensates.

*Electronics Laboratory, University of Kent.

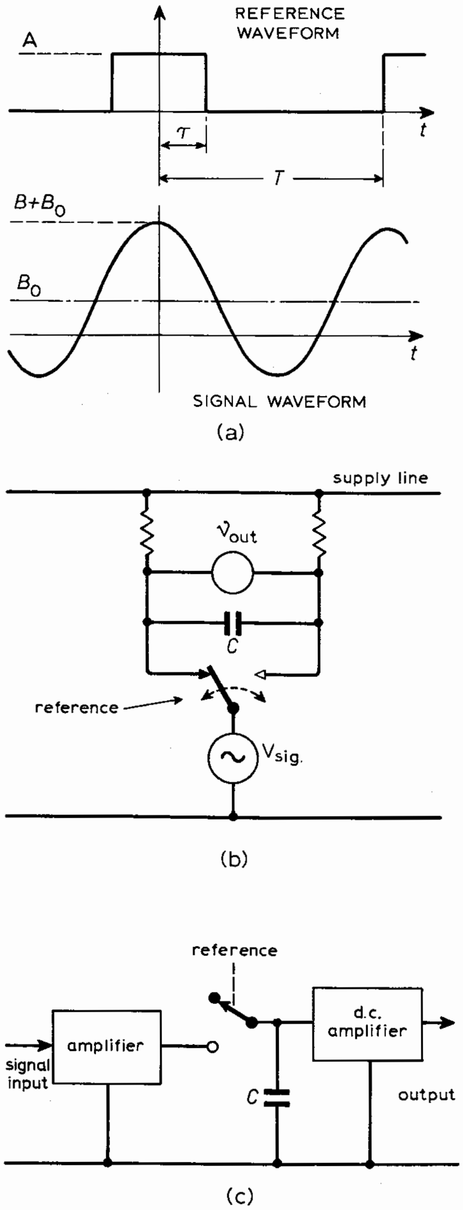

Fig. 3. The phase sensitive detector is typically a synchronous switch. A reference waveform operates the switch and the circuit should not be sensitive to changes in its amplitude.

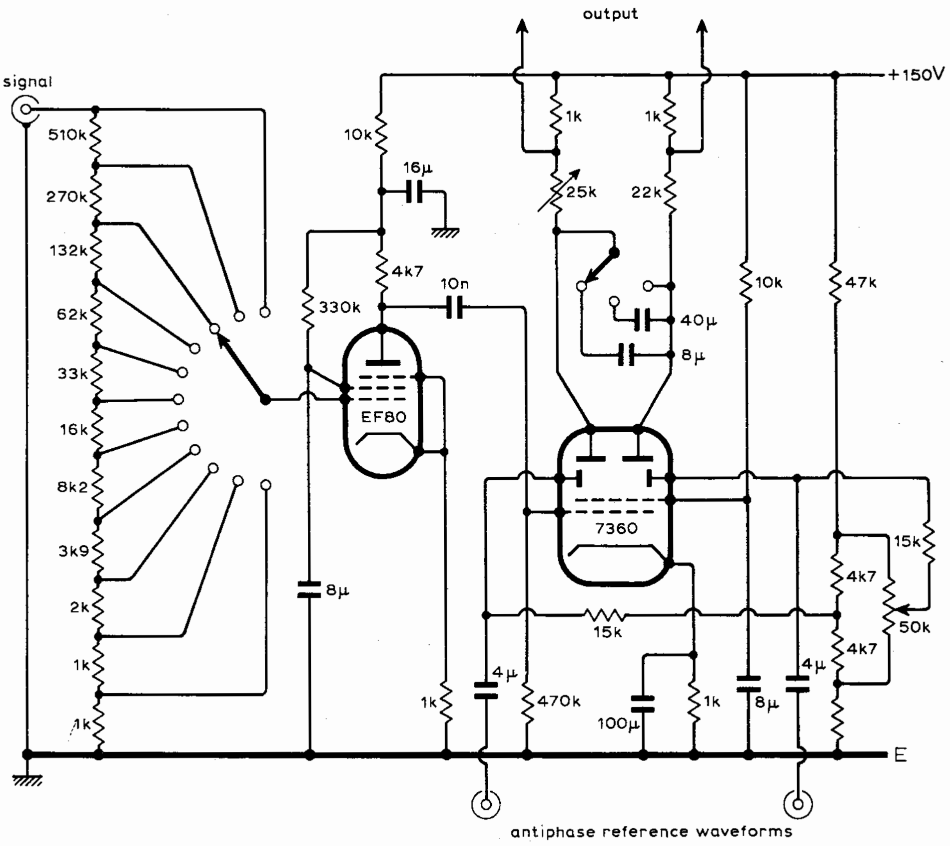

Fig. 4. Practical phase detector using beam-switching valve.

Because radio astronomical signals are wideband noise, the minimum change in source noise temperature observable, ΔT, is improved by having a large bandwidth, B, before the detector and a small bandwidth after it. The latter requirement is the same thing as saying that the detector is followed by a filter with a long integrating time, τ. In fact, by considering the statistics of the wideband noise signals, we find

$$\Delta T = \frac{2T}{\sqrt{B\tau}}\ *$$

Filtering after phase sensitive detection to obtain a large τ is relatively simple, as we shall see later.

It is a reminder to look back at the list at this point, and to realize that the old synchronous vibrators were a typical p.s.d. system, with the secondary of the transformer switched in phase across the smoothing filter. This switched or synchronous rectifier idea is often found in the armoury of the control engineer with his servo systems. The error signal is often in the form of a phase shift in one waveform relative to another and a p.s.d. gives a d.c. output proportional to this. (Actually it is proportional to the cosine of the phase angle in the case of sine wave signals, see appendix.)

Perhaps there has been a television receiver in your workshop with something a bit odd happening in the flywheel sync circuit. (You might remember the swearing involved!) Right there was the p.s.d. of course, which should have been keeping the timebase oscillator phase locked to the sync pulses. Quite unabashed we find p.s.ds all over colour television sets. The colour burst during the period of one of the porches on the video waveform is used to phase lock a crystal oscillator in the receiver. The local oscillation is used to demodulate the chrominance signals in a pair of p.s.ds.[3]

Working near the limits set by random noise, a periodic signal might be completely buried. Any signal variation which at first sight looks completely random (on a cathode ray tube, for instance), may have a periodic component. If then we phase detect with a reference signal whose frequency is chosen to equal the periodic component we know or suspect is in the noisy signal any output, after a suitably long integration time to reduce the random noise, indicates the presence of such a component. A remarkably small signal in a huge amount of noise has been detected in this way. In fact the Dicke radiometer mentioned earlier operates this way.

The p.s.d. used to dig out a weak signal in noise is working as a correlation detector, and the process is known as cross-correlating the signal and reference. Of course, we could have taken the original noisy signal and correlated it with a time delayed version of itself (auto-correlation). If there is no periodic signal component, then there will be one large output when the delay is zero and all the variations are in phase. Nothing further is obtained as output whatever the delay, because no phase coherence is obtained again. But if there is a periodic component, every time the delay is one period, there is an in-phase condition at the detector and therefore an output. The output is a series of spikes, separated by the periodic time. (A very informative article appeared in *Wireless World* on this topic in March 1955. In it, James Franklin discussed the recovery of signals by correlation methods and gave some very interesting applications, including one where a tape recorder was used as a correlation analogue computer.)

Rather closely connected with correlation, at least it would seem from the mathematical form, is an operation known as convolution. There are subtle differences however. Two functions can be convolved and in effect one is smoothed by the other. If one function has a number of fine details, these are lost if the other is a broad one. Convolution is in fact a scanning of one function by another and multiplying. So was correlation, but there one looked for similarities in the two functions. In convolution one function is reversed, then scanned across the other. An aerial beam of known distribution, looking at a source sweeping through would be a convolution process;

*The number '2' is a particular value; other slightly different constants appear according to the circuit used.

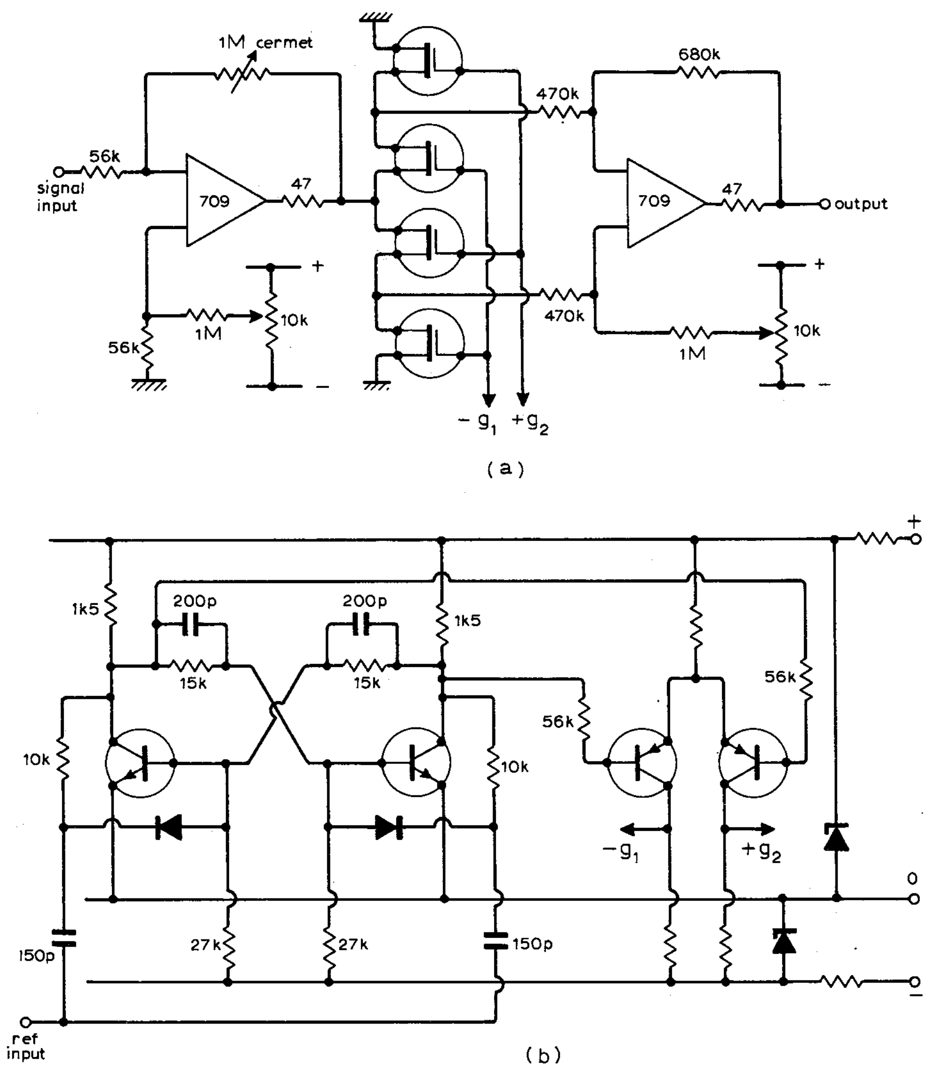

Fig. 5. This circuit—used in an astronomical radiometer—has a full-wave synchronous bridge detector. Output amplifier effectively produces a single-ended output from its balanced input signal. (Circuit of (a) connects to that of (b) at points g_1 and g_2).

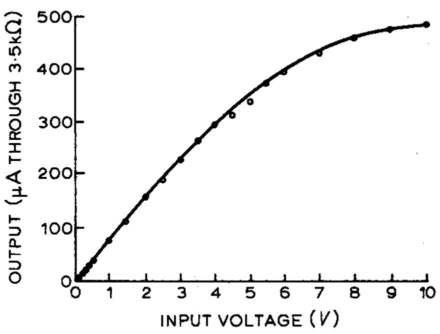

Fig. 6. The 7360 circuit of Fig. 4 is linear up to 3 or 4V input to the grid of the valve to accommodate large noise amplitudes.

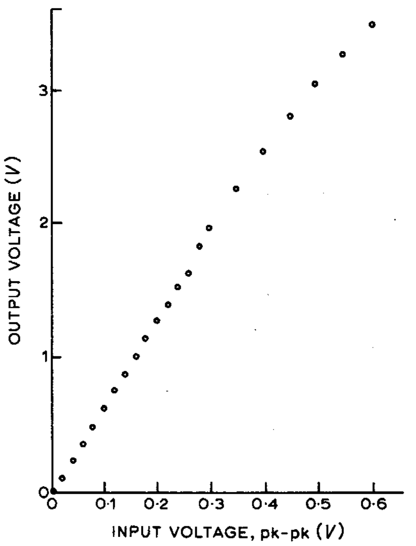

Fig. 7. Linearity of the solid-state circuit measured for one gain setting of the driver amplifier. Graph indicates order of performance one would expect in a reasonable phase detector.

so would a spectral line scanning across a slit at the output of a spectrograph.

The p.s.d. is well suited to demonstrate correlation and convolution of periodic signals. If one waveform is fed to the reference input and the other to the signal input and a slight difference in frequency exists, then as the waveforms slowly sweep 'past' each other (as seen on a double-beam oscilloscope for instance), the output of the p.s.d. will represent the correlation function of the waveforms, or the convolution if one wave shape is imagined as folded back on itself, i.e. is the reverse of the function being studied. So one can use the p.s.d. to demonstrate (and perhaps to use seriously in a system) the convolution and correlation of periodic functions, two very important ideas in modern communication theory.

Practical detectors

The basis for a phase sensitive detector system is a one-way (half-wave) or two-way (full-wave) switching circuit. Mechanical switches or commutators can be and have been used, but electronic switching is much more precise and convenient. The principle is illustrated in Fig. 3. If the switch is operated by the reference waveform in synchronism with the signal, then one half cycle only will pass to charge capacitor C in the case of half-wave switching, Fig. 3(b). In the case of full-wave switching, C is charged by both half cycles; all the positives going one way, all the negatives the other.†

Some time ago a very useful valve appeared (designed for colour television purposes) known as a beam deflection tube (R.C.A. 7360). In this device two deflector plates, rather like those in a cathode ray tube, switch the beam of electrons alternately from one anode to another. If the reference waveform is applied to the plates and a signal to the control grid, synchronous rectification takes place at the anodes. The circuit of a balanced full-wave phase-sensitive detector using this valve is shown in Fig. 4. The detector was actually developed and used with success in an infrared spectrographic system.

Of course there is some advantage in using semiconductor devices. A number of switching f.e.ts are available now and can be very successfully used in p.s.ds. The second practical example given here is working at the moment in a radiometer for solar and atmospheric observations. The arms of the bridge of f.e.ts—Fig. 5—are alternately switched by the reference waveforms in synchronism with the signal waveform at the output of the first 709 operational amplifier. In effect this means that the second operational amplifier input is alternately switched in phase across the output terminals of the first one. Filter C_T, R_T forms the integrator.

A p.s.d. worth its salt should have a fairly wide linear working region so that large noise amplitudes on small signals can be

†So far it appears that just a switching on and off is involved. In fact the signals in the two channels are multiplied together. While it is easy to see that a switch can have zero or unity gain, and therefore that the output will be the signal multiplied by a reference signal of either zero or unity, it may not always be obvious that in detectors with a sinusoidal or otherwise non-rectangular reference that the output is the product of the inputs. Looking on the p.s.d. as an analogue multiplier removes this problem. While this is not the place to discuss techniques of analogue multiplication, one particular realisation may be helpful. Imagine a signal attenuated in a potentiometer, the output being the product of tapping position—expressed as a fraction of the whole—and input signal amplitude. Multiplication by a second, reference signal is then achieved by applying it to the potentiometer wiper electromechanically. In the special case where the reference is rectangular, one is operating a switch that is either on or off, the product being either zero or finite (unity where there is no gain), depending on whether the switch has zero or unity gain. One can visualize this as multiplication by thinking of the two extremes of potentiometer tap as representing zero and unity gain.

accommodated without running the circuit into overloading. The circuits I have described produced the input-output linearity curves shown in Figs 6 and 7.

Perhaps this outline discussion of the phase sensitive detector begins to show some of the ubiquity. There is much more that could be said and a suggestion for further reading is given at the end of the references. The appendix gives a few sums for those interested, but the arguments are not necessary for the main ideas of my text.

Appendix

The two instantaneous values of the reference and signal waveforms occurring at every instant, are multiplied together by a phase sensitive detector, hence the other name—product detector. If the reference signal is $f_1(t)$ and the main signal is $f_2(t)$, the output of the p.s.d is $f_1(t)f_2(t)$ in the absence of overloading and before smoothing or integration. The area under the product curve of $f_1(t)$, $f_2(t)$ is

$$\int_{-\infty}^{\infty} f_1(t)f_2(t)\,dt$$

It is somewhat arbitrary to go ahead and talk about the average value of $f_1(t)f_2(t)$ if there is just one occurrence (for instance, the average value of the above integral over an infinite time is zero), but if the functions are periodic with period T the average value is

$$\frac{1}{T}\int_{-T/2}^{T/2} f_1(t)f_2(t)\,dt$$

and this is the value that would be produced across, say, a large smoothing capacitor at the output of the p.s.d.

Suppose $f_2(t)$ is a periodic reference signal of symmetrical square waveform. Fourier tells us

$$f_2(t) = A_2\left(\frac{1}{2}+\frac{2}{\pi}\cos\omega t-\frac{2}{3\pi}\cos 3\omega t+\frac{2}{5\pi}\cos 5\omega t-\ldots\right)$$

where A_2 is the amplitude, the first term a d.c. component and the remainder the fundamental and odd harmonics. If $f_1(t)$ is a sine wave signal with amplitude A_1 and phase angle ϕ, then

$$f_1(t) = A_0+A_1\cos(\omega t+\phi)$$

Multiplying $f_1(t)$ and $f_2(t)$ we get a series whose terms are all zero on average, except for the first two

$$\frac{A_0A_2}{2}+\frac{A_1A_2}{\pi}\cos\phi$$

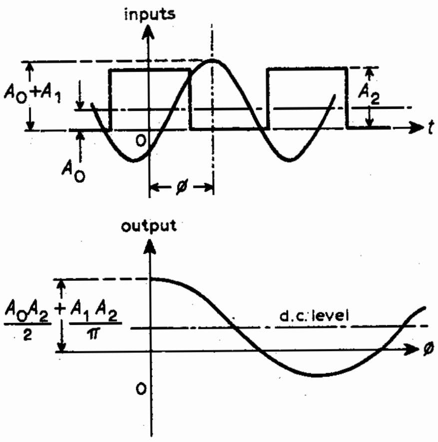

Fig. 8. The p.s.d. as a phase meter—with sinusoidal input the output is proportional to the cosine of ϕ (about any d.c. level).

This shows that the output varies with the cosine of the phase angle and has maximum and minimum values of $+A_1A_2/\pi$ (assuming from now on that the d.c. components of both signals are zero) as Fig. 8 shows.

Suppose that the input signal $f_1(t)$ is also a symmetrical square wave, then the output will be the average value of

$$f_1(t)f_2(t) = \left[\frac{2A_1}{\pi}\cos\omega t-\frac{2A_1}{3\pi}\cos 3\omega t+\ldots\right]\cdot\left[\frac{2A_2}{\pi}\cos(\omega t-\phi)-\frac{2A_2}{3\pi}\cos 3(\omega t-\phi)\ldots\right]$$

Now more of the terms multiply up to give a contribution to the averaged output

$$\frac{4A_1A_2}{\pi^2}\cos\phi+\frac{4A_1A_2}{9\pi^2}\cos 3\phi+\frac{4A_1A_2}{25\pi^2}\cos 5\phi+\ldots$$

and by trigonometry, the sum of this series can be shown to be

$$\frac{4A_1A_2}{\pi^2}\left(1-\frac{2\phi}{\pi}\right)$$

for values of ϕ between 0 and π radians, and

$$\frac{4A_1A_2}{\pi^2}\left(\frac{2\phi}{\pi}-3\right)$$

for values of ϕ between π and 2π radians. In words, the output is a triangular wave as ϕ varies, as shown in Fig. 9.

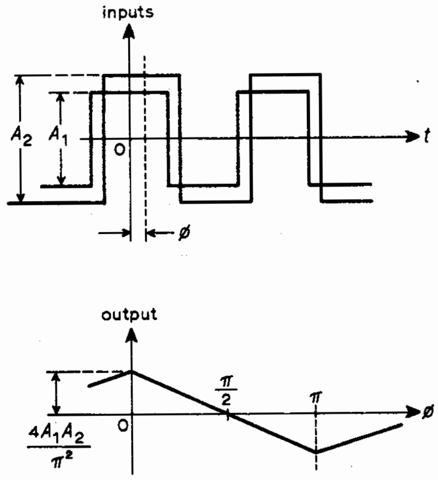

Fig. 9. In the case of square waves, the p.s.d. output is directly proportional to ϕ.

Rewriting the phase changes in terms of time variations $\cos(\omega t-\phi)$ can be written $\cos\omega(t-\phi/\omega)$ or writing ϕ/ω as τ, and applying this idea to any function, we can finally write $f_1(t-\tau)$ for the signal and interpret it as function $f_1(t)$ sweeping along the axis as τ varies. This is true even if the function is no longer periodic. This idea of a whole function sweeping along the axis might be a little difficult to visualize, but think of a wave motion, for instance, and it should become clearer. Fig. 10 shows this for $f_1(t-\tau)$. If τ is added to t, that is $f_1(t+\tau)$, then the function moves along in the other direction, as indicated in Fig. 11.

It is now possible to see what happens when one function is scanned across either a replica of itself or a different function; an action obtained in practice, for example, by using a variable delay network in one channel. The mathematics of this is fairly easy to see now, so consider the following average

$$C_{11}(\tau) = \frac{1}{T}\int_0^T f_1(t)f_1(t+\tau)\,dt$$

where T is the period.

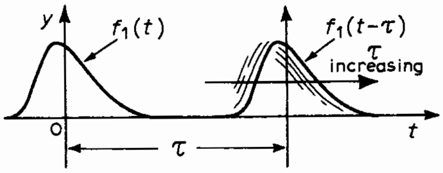

Fig. 10. As τ varies, the whole function $f_1(t-\tau)$ moves along the t axis.

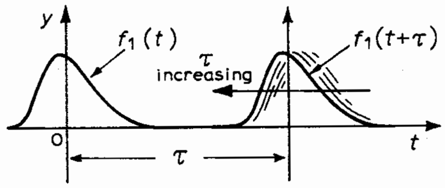

Fig. 11. With the opposite sign, the direction of movement of the function with τ is reversed.

Function $C_{11}(\tau)$ is known as the autocorrelation function; in this case, for periodic functions. Two different functions multiplied and averaged in this way gives the cross-correlation function

$$C_{12}(\tau) = \frac{1}{T}\int_0^T f_1(t)f_2(t+\tau)\,dt$$

These operations search out any similarities in the functions, hence give a measure of the correlation between them. They are very significant in statistical communication theory and crop up in such applications as digging out small signals in large amounts of noise.

Mathematically speaking, convolution has a lot in common with the ideas mentioned above. There is one important difference however. One of the functions is reversed—hence the name of the operation, to convolve—and then it is scanned to produce the product which is finally integrated and averaged

$$\text{convolution, } C_{1,-2} = \frac{1}{T}\int_0^T f_1(t)f_2(\tau-t)\,dt$$

Typical convolutions include the case of a spectral line scanning past a slit, or a distant radio source distribution being scanned by an aerial beam. Of course, convolution gives an output not primarily dependent on any similarity between the functions (contrast correlation), but is a kind of 'smoothing' of one function by the other.

The phase sensitive detector shows all these operations for periodic functions.

References

1. 'Cathode Ray' The Synchrodyne, *Wireless World*, vol. 53, 1948, pp. 277–81.
2. R. A. Smith, F. E. Jones and R. P. Chasmar, The Detection and Measurement of Infra-red Radiation, Oxford, 1957.
3. S. C. Ryder-Smith, The P.A.L. Colour TV System, *Wireless World*, vol. 73, 1968, pp. 628–33.
A good general reference to the kind of material discussed in the appendix and which I have found very readable is chapter 7 in Information Transmission Modulation and Noise by Mischa Schwartz (McGraw Hill).