A New Principle of Linear Phase Discrimination - Irregular Phase Discrimination

Zhou Wei
Shaanxi Institute for Measurement and Testing
30, West Xianning Road
XI'AN, CHINA

Abstract: The principle of irregular phase discrimination for getting the relative phase value consists of applying two signals which are of different frequencies to a linear phase discriminator. When the frequencies of these two signals are f_A=A*C, f_B=B*C, where C can be any frequency, A and B are positive integers without any common divisor. The equivalent frequency of the phase discrimination obtained in this way is f=A*B*C. The advantages of the principle of phase discrimination are: low demand on the performance of the equipment, high precision of the comparison and the flexibility in the frequency used for phase discrimination. The principle will generally be used in the design of frequency standards, for frequency measurement and for general and special phase (time interval) comparisons.

The two phase compared signals must have identical averages in the conventional phase detection method. Not only does this method limit the frequency from going higher, but it also has some nonlinearity in the phase detection at HF, and has a dead zone problem.

Phase comparison with better precision is usually required in frequency and time measurements. It is not required that the absolute frequency or phase be measured, but only the relative values. A lot of analyses and experiments have shown that it is not necessary to require that the two phase comparing signals have the identical frequency for obtaining the relative phase value, as long as their frequencies agree with some precision. A typical phase comparison device (with JK flip-flop characteristics) shown in Fig. 1 may be used. The frequencies of the two inputs are different. The two frequencies are:

F_A=A*C (1)
F_B=B*C (2)

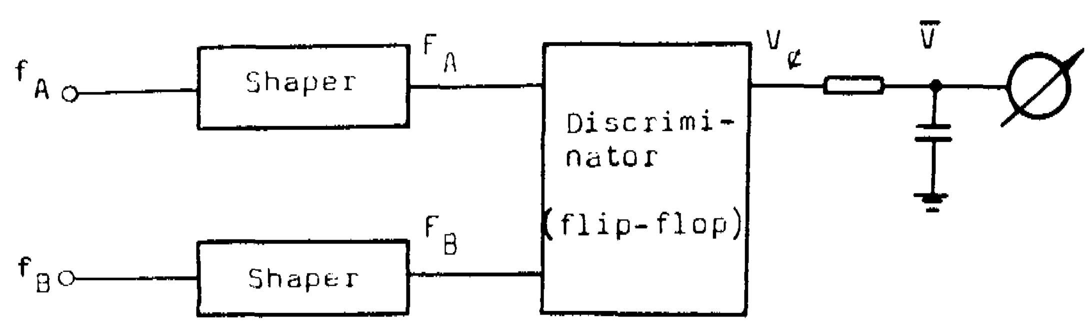

Fig. 1 Typical Phase Comparison Block Diagram

where C can Be any frequency; A and B are integral numbers without any common divisor. The periods of the two signals are, respectively,

$$T_A=1/f_A=1/A*C \qquad (3)$$
$$T_B=1/f_B=1/B*C \qquad (4).$$

Thus, in 1/C seconds, f_A completes A periods, while F_B completes B periods. With every 1/C seconds, there are always A periods of f_A compared with B periods of f_B in the switching phase detector. The pulse widths of the square wave output from the switching phase detector differ from each other in every 1/C seconds since the periods T_A and T_B are rather different. However, the output square waves repeat every 1/C seconds periodically. The pulse width of each square wave is the same as the one 1/C seconds later) except for the phase shifts caused by the frequency differences between the two signals and for their average values).

The switching phase detector is activated by A triggering signals of f_A and B triggering signals of f_B every 1/C seconds (see Fig.1 and Fig. 2). Suppose that A>B, then there are B square pulses output from the phase detector in 1/C seconds. The widths of the B pulses differ from each other in a group, and all the pulse widths of B strings, with an interval of 1/C seconds, change with the phase variations of the two compared input signals.

Equations (1) and (2) represent the situation where f_A and f_B are coherent. If they are independent of each other, f_B is supposed to be the reference signal, $f_A=A*C+ f=(A+ f/C)*C$. That is, $T_B=1/B*C$, $T_A=1/(A+ f/C)*C$.

Then all the pulse widths of the B train of square waves vary with the period of 1/C seconds, and the period of every square wave is T_B. However, the width of the output square pulse can only vary over the range 0 $-T_A$ due to f. Along with this variation of phase, the pulse widths of a group of square waves in the train of B pulses, repeated each 1/C seconds, will either continually become smaller until it reaches zero, or larger until it reaches T_A. Then the pulse width jumps; it alternates between 0 and T_A (the jump is between T_B and T_B and T_A at the opposite output of the phase detector). The sum of all the widths of the B pulses, which repeat every 1/C seconds, jumps when the pulse width of any group jumps. If f is constant, all the jumps in the pulse widths of the train of B pulses happen sequentially in a certain direction. Between every two jumps the variation of the sums of the pulse widths are identical and equal to the jumping value. When A>B, the pulse width jumping group in the B train of pulses will always jump by an absolute value of 1/A*C seconds. That is, certain pulse groups, whose period is 1/C seconds jump because their phase has drifted 1/A*C seconds, since f is not zero. The output square waves from the phase detector have been modified from each other by the B train of pulses, which have a period of 1/C seconds. When f is a constant, the pulse widths of the B train of pulses always either increase or decrease, and jump sequentially in a certain direction. Between every two jumps of a pulse group, the two comparing signals have their relative phase shifted 1/A*C seconds because $f\neq0$, and all other B-1 square wave groups injected into other neighborhoods have also had a pulse width jump. The phase shift, when the pulse width of each square wave group jumps, is identical, as well as the opposite phase shift between every two jumps. Then the equivalent frequency of the phase comparison is:

$$f=A*B*C \qquad (5).$$

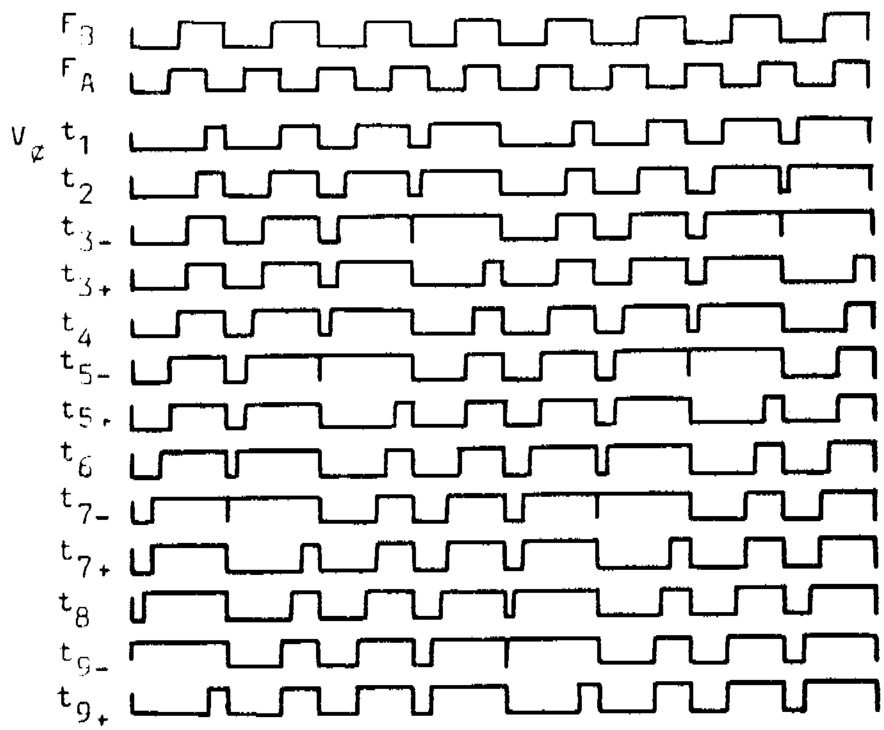

Fig.2 Waveforms of irregular phase detection

The waveforms of the aforementioned operational procedure are shown in Fig.2. The curve f_B is a reference signal and V_ϕ is the output from the phase detector on each row, showing the phase detecting output during a short time interval at a given instant. Some typical phase detecting output waveforms as functions of time are shown in the different rows.

Because the frequencies for phase detection are different and the output pulse widths of the square wave from the phase detector seem irregular at first glance, it has been called "irregular phase detection". The phase variation of the compared signals is very precisely analogized by the level variation from a low pass filter on the phase detector output.

Some characteristics of the irregular phase detection are as follows:

1. Equations (1), (2) and (5) show that the bigger the values of A and B (while C is smaller by some degree), the the higher the equivalent phase detecting frequency.

For two high frequency signals with identical effective values, the linearity and the precision of the phase detection may be improved by frequency division before phase detection. There can be no common factor between the divisors of the two signals; otherwise the equivalent frequency and the precision of the phase detection will be lowered proportional to the common divisor.

2. Comparing the full periodical outputs, the voltage amplitude from the irregular phase detection is smaller than that from the identical frequency phase detection.

The output pulse width does not change totally periodically from the phase detector in irregular phase detection. Furthermore, only one of the B trains jumps during a full phase detecting period, and the pulse widths of all the other trains are meanwhile gradually drifting along with the phase drift of the comparing signals. Thus, from the phase detector made of switching components, the output equivalent phase detecting voltage in a full period of irregular phase detection is:

$$V=V_m*f_B/f_A$$

$$=V_m*(B*C)/(A*B*C)$$

$$=V_m/A \qquad (A>B) \qquad (6)$$

where f is the equivalent phase detection frequency; V_m is the output phase detecting voltage in a full period of the identical frequency phase detection, which is about the voltage of the power supply if the saturation voltage of the switching phase detector and the attenuation of the filter are ignored.

3. The linearity of irregular phase detection is much better than that of the identical frequency phase detection. As shown in Fig. 2, since $f_A \neq f_B$ in the case of irregular phase detection, only one of the two extreme situations of the duty factor (1 and 0) can exist in the output of the phase detector. Among the injected pulse groups of the B trains, when the duty factor of one train is 1 or 0, all the other B-1 trains are situated between the possible maximum and minimum of the duty factor. Nonlinearity will not be caused by the limited switching speed of the phase detector. The phase detecting output, after filtering, is equal to the average D.C. voltage of the square waves. Nonlinearity has been reduced considerably since a possible nonlinear case has been averaged by B-1 normally linear cases. By increasing the values of B and A, the phase detecting nonlinearity can be improved while the equivalent frequency of phase detection stays constant. The two comparing signals can be divided by large numbers (without common divisors) before phase detection to improve the linearity; however, it will reduce the output voltage of the phase detection in a full period.

For identical frequency phase detection at 5 MHz, the dead and serious nonlinear zones are about 15% of a phase detecting period. For irregular phase detection at an equivalent frequency of 50 MHz, using a similar phase detector, the abnormal zone is smaller than 1% if A and B are properly chosen.

4. In irregular phase detection, the period of the output square wave from the phase detector is equal to that of the signal, which has a lower frequency, and the range of the pulse width variation is equal to the period of the signal which has a higher frequency. The period of the output square wave groups is equal to the reciprocal of the common factor C in $f_A=A*C$ and $f_B=B*C$; therefore, 1/C corresponds to the upper limit of the responding speed of the phase detector.

The phase detector designed for the measurement of atomic frequency standards using irregular phase detection may get a similar result to that of the identical frequency phase detector at 100 MHz, but it works at a lower frequency. The precision of irregular phase detection can be as high as $3.5*10^{-15}$/day and $1*10^{-13}$/hour, yet the equipment is kept simple.

It is a common situation in radio electronics that two signals with different frequencies need to be phase compared. Usually there is great difficulty turning them into identical frequencies, but irregular phase detection can simplify the complicated circuits in this situation and avoid some additional noise.

This new principle of phase comparison may be widely used in phase locking techniques, in the design of new frequency standards, frequency measurement, automatic control systems, and for general and special phase comparison. Higher precision may be achieved using simplified systems.