

Chapter 2

Theories and Technologies of Tracking and Orbit-Measuring

2.1 General

Tracking and orbit-measuring refers to tracking and measuring the flight trajectory of a flight vehicle by using measurement elements such as range, velocity, angle, and altitude measured by the measuring stations. “Trajectory” is a general term. The unpowered flight trajectory of spacecraft-like satellites refers to orbit, which follows the orbital dynamics. The powered flight trajectory of missiles and rockets refers to ballistic trajectory; the flight trajectory of aircrafts generally refers to track. Trajectory and track both follow the aerodynamics. Thus, orbit-measuring as discussed in this book refers to orbit measurement of spacecraft, trajectory measurement of missiles, and track measurement of aircrafts.

The flight trajectory of center of mass of a spacecraft revolving around the Earth refers to its motion orbit. Under ideal conditions, its motion around the Earth conforms to Kepler’s laws, and its motion trajectory is an established Keplerian ellipse. So the path is called the orbit. In fact, the Earth is not a perfect sphere and its mass is unevenly distributed. Moreover, due to atmospheric drag, attraction of other celestial bodies to the spacecraft, and the acting forces of solar radiation pressure on the spacecraft, the spacecraft orbit becomes a complex curve approximating the Keplerian ellipse. The said acting forces are generally called “perturbing forces,” and the deviation of spacecraft orbit from the Keplerian ellipse is called “perturbation.” Since actual orbit deviates from the theoretical orbit, orbit measurement must be performed to control the spacecraft in order to maintain traveling in originally designed orbit or carry out necessary “orbital transfer.” Methods of orbit determination mainly include:

(1) Geometric orbit determination: Also called kinematics orbit determination, it uses target angle, range, velocity, and other parameters measured by one or more tracking and measuring stations to calculate and obtain instantaneous position and velocity of the spacecraft according to relative geometrical relationship and then obtain its flight trajectory through point fitting. This method is just a geometric approximation to the true orbit and does not reflect actual change law of spacecraft orbit. So it has poor precision.

(2) Dynamics orbit determination: For orbiting spacecraft, first obtain the preliminary model of the spacecraft orbit according to the location, velocity, and direction of the injection point and then use the dynamics model of the spacecraft orbit to determine its orbit (e.g., six Kepler orbit elements) after taking into account various forces on the spacecraft in orbit during its motion. Since spacecraft orbit is determined by orbit dynamics model and its observation values relative to each measuring station can be determined by the model, there is no need of vast measured data to determine spacecraft orbit, and the data measured by the measuring stations are just used to improve the orbit model. That is, use the perturbation theory and statistical approach to solve the theoretical value of the orbit and figure out another orbit value based on the measurement element values obtained from the measuring stations, then take the difference between these two orbit values as the initial value for improved orbit and repeat the iteration to correct parameters and initial conditions of orbit's mathematical model until a high-precision orbit determination is achieved. Due to the application of constraint conditions specific to orbiting, the dynamic method has higher orbit determination precision than geometric orbit determination method. Moreover, it needs fewer measurement elements, so any one measurement element can be used for orbit determination, such as velocity, range, or angle.

For flight vehicles having driving force or more air drag such as missile and aircraft, it is not suitable to use dynamic orbit determination method due to its poor accuracy. Generally, geometric method is applied. Based on precise formula and sound numerical approaches, the data, such as angle, range, and velocity measured by the measuring stations, are used to determine the orbit of the spacecraft, in which the system errors are corrected and the random errors are smoothed to obtain accurate orbit. Since the flight vehicle is in motion, its orbit can be determined only when both its velocity and location are determined, that is, six orbit elements $x, y, z, \dot{x}, \dot{y}, \dot{z}$ should be determined. It can be seen that orbit measurement involves two aspects, namely, velocity measurement and location determination.

2.2 Localization and Orbit-Measuring

2.2.1 Localization

Spatial localization of the flight vehicle is a three-dimensional locating problem. At least three independent location parameters are needed to determine the spatial location of the target. Currently, location parameters which can be measured mainly include range R , azimuth angle A , elevation angle E , range sum S , range difference r , and direction cosine (including l , m , and n), just three of which are enough to determine the spatial location of the flight vehicle. The geometric principle of localization is to make three geometric surfaces by using these three location parameters. The intersection point of these surfaces is the spatial location of the vehicle. This geometric theory localizing the spacecraft is called location geometry. Three location parameters as measured can be used to name the localization system,

for example, (A, E, R) localization system and (R, S, r) localization system. As any three of six location parameters can form a localization system, there are 56 combinations of three that can be drawn from six parameters, namely, 56 localization systems in total. Similarly, there are also 56 systems in velocity measurement, such as $(\dot{R}, \dot{A}, \dot{E})$. At present, there are many definition methods for the locating system. According to the localization geometry theory, the better definition method is using the name of measurement elements.

The geometric surfaces decided by seven major location parameters in engineering application are summarized as follows [1]:

(1) Range R : the distance between orbit-measuring station and the flight vehicle. The flight vehicle may be located at one point on a sphere which takes the measuring station as its center and R as its radius. In the (x, y, z) coordinate system, R meets the following expression:

$$x^2 + y^2 + z^2 = R^2 \quad (2.1)$$

The geometric figure of the sphere is shown in Fig. 2.1.

(2) Elevation angle ε : angle of elevation of the flight vehicle relative to the ground. In the (x, y, z) ground-transmitting coordinate system, elevation angle shall be determined by:

$$x^2 + y^2 = z^2 \cot^2 \varepsilon \quad (2.2)$$

Its geometric figure is a cone, as shown in Fig. 2.2.

As shown in Fig. 2.2, xOy is the horizontal plane. The vehicle may be located at any point on this cone.

(3) Azimuth angle $\varphi(\beta)$: angle of azimuth of the flight vehicle in the (x, y, z) coordinates. The azimuth angle should be determined according to the following expression:

$$y = x \tan \varphi \quad (2.3)$$

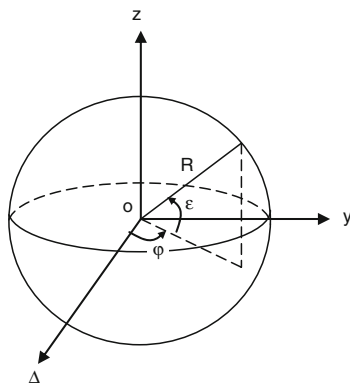


Fig. 2.1 Target location sphere determined by range R

Fig. 2.2 Target location surface determined by elevation angle ε

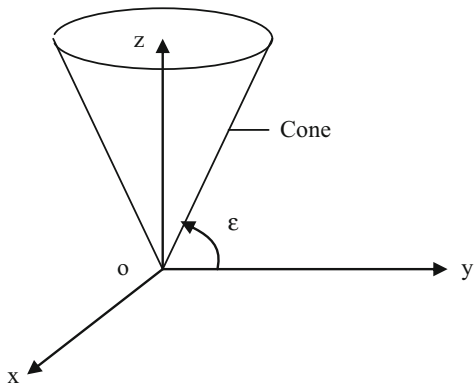
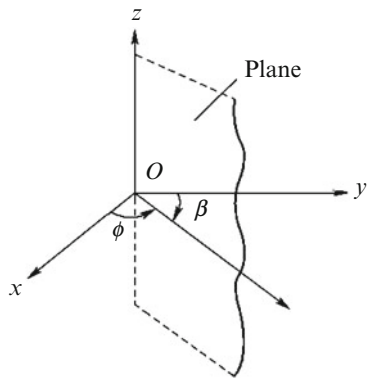


Fig. 2.3 Target location determined by azimuth angle φ



The geometric figure of azimuth angle is a plane perpendicular to xOy horizontal plane, as shown in Fig. 2.3.

The flight vehicle may be located at any point on this plane.

(4) “Range sum” $S = r_i + r_j$, where r_i refers to the distance between the master station and the target and r_j refers to the distance between the slave station and the target. At this point, the target is located on a rotational ellipsoid determined by the following expression:

$$\begin{aligned}
 S &= r_i + r_j \\
 &= \left[(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 \right]^{\frac{1}{2}} \\
 &\quad + \left[(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2 \right]^{\frac{1}{2}}
 \end{aligned} \tag{2.4}$$

where $x_i = [x_i y_i z_i]^T$, $x_j = [x_j y_j z_j]^T$ are the i th and j th station locations, respectively, and $x = [x, y, z]^T$ is spatial location of the target. The rotational ellipsoid determined by the said expression is as shown in Fig. 2.4.

Fig. 2.4 Target location determined by range sum S

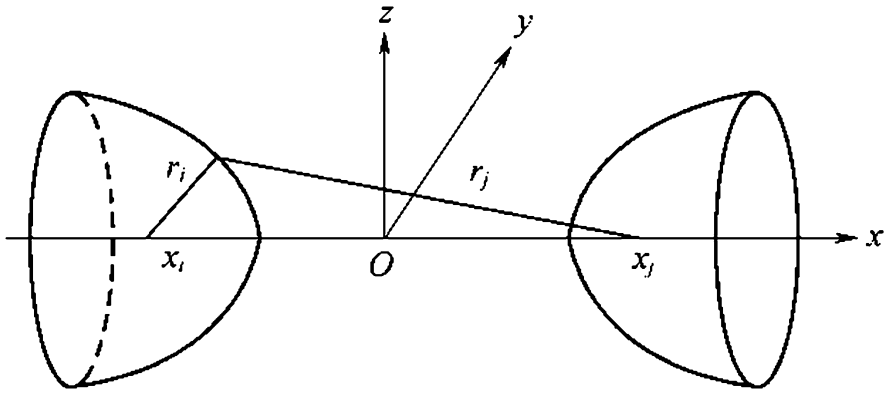
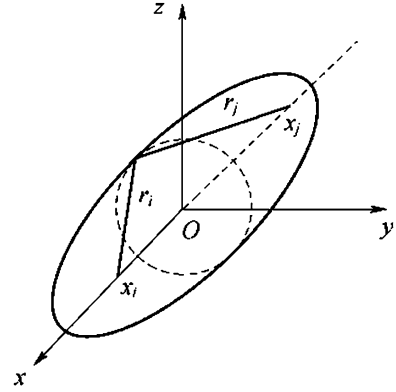


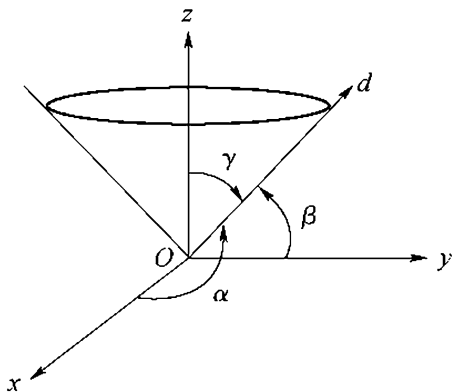
Fig. 2.5 Target location surface determined by range difference r

(5) Range difference $r = r_i - r_j$; at this point, the target is located on a rotational hyperboloid determined by the following expression:

$$\begin{aligned}
 r &= r_i - r_j \\
 &= \left[(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 \right]^{\frac{1}{2}} \\
 &\quad - \left[(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2 \right]^{\frac{1}{2}}
 \end{aligned} \tag{2.5}$$

where all symbols have the same meaning as those in Expression (2.4). The rotational hyperboloid determined by the said expression is as shown in Fig. 2.5.

Fig. 2.6 Target location surface determined by direction cosine



(6) Direction cosine:

$$\begin{aligned} l &= \cos \alpha \\ m &= \cos \beta \\ n &= \cos \gamma = \sqrt{1 - l^2 - m^2} \end{aligned} \quad (2.6)$$

Target location surface determined by direction cosine is shown in Fig. 2.6.

In Fig. 2.6, three direction cosines ($l = \cos \alpha$, $m = \cos \beta$, $n = \cos \gamma$) correspond to three cones which take x -axis, y -axis, and z -axis as cone axis and α , β , and γ as semi-vertex angle, respectively (the figure shows an $n = \cos \gamma$ cone). Any two cones intersect in a line of direction d , i.e., the line of position (LOP).

It can be seen from Fig. 2.6 that target location surfaces determined by both direction cosine and elevation angle are cones (because they are measured angle parameters). The difference is that the geometric surface of target location determined by direction cosine is a cone taking the baseline between two stations as its axis and having an angle α (the angle between target angle direction and the baseline of two stations) while the target location surface determined by elevation angle is a cone taking the normal axis of the antenna as its axis and having an angle ε . The advantage of the method of measuring direction cosine is that higher accuracy of α measurement can be obtained.

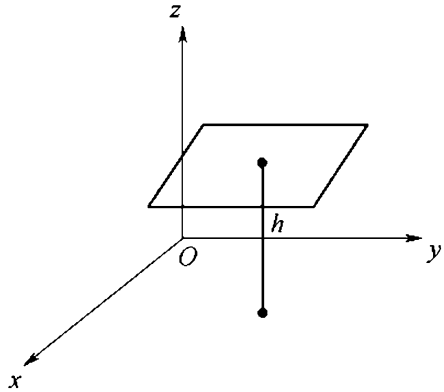
(7) Target altitude h : the altitude is determined by the following expression:

$$z = h \quad (2.7)$$

Expression (2.7) represents a plane with altitude as h , as shown in Fig. 2.7.

According to the foregoing analysis, one location parameter can only determine that the target is located on a surface but can't determine exactly which point it is located on the surface. If there are two location parameters, the two surfaces determined by these two parameters are intersected on a curve. Then, we can only confirm that the target is located in this curve but could not confirm which point it is located on the curve. Only by getting the point of intersection of this curve

Fig. 2.7 Target location surface determined by altitude h



and the surface determined by the third location parameter can finally determine which point the target is located at. R , r , and S are ranging systems, among which one can be converted into another, and each of which is equivalent and can be converted into the pure ranging system.

Among the localization systems, some are suitable for LEO spacecraft, others are suitable for HEO or deep-space spacecraft; some have high accuracy, others have low accuracy; some are high in cost, others are low. So the selection of a localization system is very important in the TT&C system.

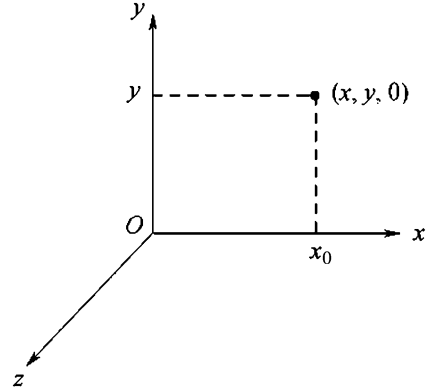
2.2.2 Station Distribution Geometry and Localization Accuracy

Localization methods of spacecraft can be classified into three types: range, azimuth, and elevation (RAE) localization method; angle intersection localization method; and range intersection location method. The point of intersection in space of three geometric surfaces determined by three location parameters is the location of the spacecraft. It is thus clear that the localization accuracy of a spacecraft is related to both measurement accuracy of TT&C equipment and geometry and accuracy of station distribution. Reference [2] gives an in-depth study and provides the relevant results.

Station distribution geometry refers to the relative geometric relationship between orbit-measuring stations and the target, which directly affect the localization accuracy of the target. With the same internal error and external error (or generally same), the different station distribution geometry would cause significant difference in orbit-measuring precision.

“Optimality” metric of station distribution geometry has many criteria. In general, the following minimum criterion is adopted for study of the optimal

Fig. 2.8 Ground non-inertial launch coordinate system



localization geometry (error F expressed in this method is also called statistical module length):

$$F = \sigma_x^2 + \sigma_y^2 + \sigma_z^2$$

Total linear error is

$$\sigma_L = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} = \sqrt{F}$$

where δ_x , δ_y , and δ_z are the location error in the directions of x -axis, y -axis, and z -axis in the ground non-inertial coordinate system, respectively.

It can also refer to velocity error. The study results are applicable to both. A selected coordinate system is shown in Fig. 2.8. The xOy surface and target launch surface coincide with each other; xOz is the ground surface; $(x, y, 0)$ represents the coordinate of the target; and $(x_0, 0, 0)$ refers to the subsatellite point of the target.

2.2.2.1 RAE Localization System

A ground non-inertial launch coordinate system is shown in Fig. 2.8.

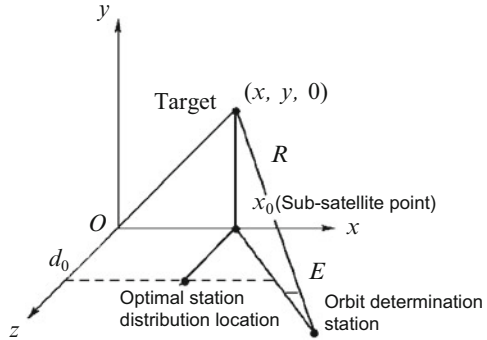
When ranging RMS error is δ_R and angle measurement RMS error is $\delta_A = \delta_E$, the location error expressed by statistical module length F is

$$F = \sigma_R^2 + R^2(1 + \cos^2 E)\sigma_A^2 \quad (2.8)$$

where R is the distance between the measuring station and the target and E is the elevation.

A geometric relationship in single-station orbit determination is shown in Fig. 2.9.

Fig. 2.9 Geometric relationship in single-station orbit determination



When the orbit determination stations are distributed in different locations, F has different values. The station distribution with F being the minimum value is the optimal distribution. According to the analysis of Reference [3], the target is located in the vertical line of Ox axis where $z = 0$, i.e., at subsatellite point $(x_0, 0, 0)$. Then, the location error is

$$F = \sigma_R^2 + (y^2 + 2z_1^2)\sigma_A^2$$

When $z_1 = 0$, then

$$F = \sigma_R^2 + y^2\sigma_A^2 \quad (2.9)$$

where y refers to target altitude.

It is clear that the higher the target, the larger the location error will be.

Considering there is an unsafe area $|z_1| < d_0$ (d_0 is the distance to safety line) in the test range, optimal address of orbit determination station shall be at $(x_0, 0, \pm d_0)$, as shown in Fig. 2.9.

When two sets of RAE orbit-measuring equipment are used, optimal station distribution is located at $(x, 0, d_0)$ and $(x, 0, -d_0)$, respectively. Then, the location error is 1/2 less than that of single-station localization. The optimal station distribution properties of RAE localization are summarized in Table 2.1.

2.2.2.2 “Angle Intersection” Localization

This method adopts two or more high-precision angle measuring systems (such as optical theodolite or interferometer) to find the direction of the target and uses the point of intersection of two or more direction lines to determine the target location.

(1) Dual-station angle intersection localization

Given angle measurement accuracies of two stations are σ_1 and σ_2 , respectively, $\sigma_1 = \sigma_2$. Optimal station address is located at the straight line which is perpendicular

Table 2.1 Optimal station distribution properties of RAE localization

S/N	Item	Single-station localization	Dual-station localization
1	Optimal station distribution location	On the vertical line of x -axis in xOz plane passing through the subsatellite point	On the vertical line of x -axis in xz plane passing through the subsatellite point
2	Vertical distance between orbit-measuring station and subsatellite point	$z_1 = d_0$ (or $z_1 = -d_0$)	$z_1 = d_0$ (or $z_2 = -d_0$)
3	Spatial included angle of the target		$\beta^* = 2^* \arctan(d_0/y)$
4	Elevation between orbit-measuring station and the target	$E = 90^\circ - \alpha^*$	$E = 90^\circ - \alpha^*$
5	Slant range between orbit-measuring station and the target	$R^* = [y^2 + d_0^2]^{\frac{1}{2}}$	$R^* = [y^2 + d_0^2]^{\frac{1}{2}}$
6	Minimum risk value	$F_{\min} = \sigma_R^2 + (2d_0^2 + y^2)\sigma_A^2$	$F_{\min} = 1/2F_{\min}$
7	Optimal estimation accuracy of target location	$\sigma_x = d_0\sigma_A$ $\sigma_y = [(y/R^*)^2\sigma_R^2 + d_0\sigma_A^2]^{\frac{1}{2}}$ $\sigma_z = [(d_0/y)^2\sigma_R^2 + y^2\sigma_A^2]^{\frac{1}{2}}$	$(\sigma_x, \sigma_y, \sigma_z) = \frac{1}{\sqrt{2}}(\sigma_x, \sigma_y, \sigma_z)$

Note: d_0 is the distance to safety line

to Ox axis and passes through the subsatellite point. Two stations are located at both sides of Ox axis, respectively, and the distance between station and Ox axis is:

$$d = 0.67183y \quad (2.10)$$

where y is the target altitude.

Then, optimal localization accuracy is

$$\begin{cases} \sigma_x = 0.4751y\sigma_1, & \sigma_y = 1.5276y\sigma_1 \\ \sigma_z = 1.0263y\sigma_1, & F_{\min} = 3.61235y^2\sigma_1^2 \end{cases} \quad (2.11)$$

Geometry of the station distribution is shown in Fig. 2.10.

(2) Tri-station angle intersection localization

Optimal station distribution is of a regular triangle, and the subsatellite point is located at the center of such regular triangle. The vertex of the regular triangle is located at a station distribution circle. The location error is

$$F = \sigma_a^2 R^4 \frac{(R^4 + y^2 d^2 + 4d^4)}{[3d^2(R^4 + y^2 d^2)]} \quad (2.12)$$

Fig. 2.10 Geometry of optimal station distribution in dual-station angle intersection localization

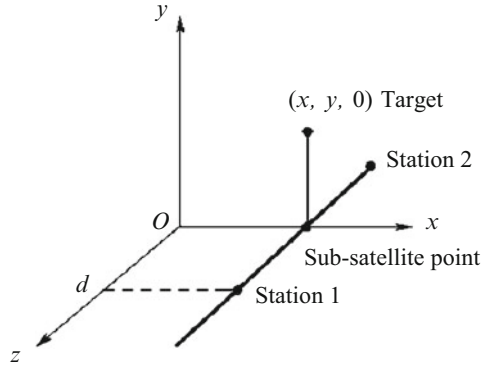
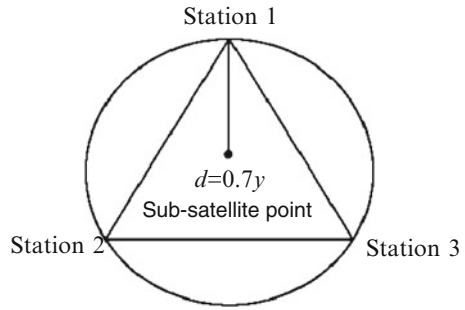


Fig. 2.11 Geometry of optimal station distribution in tri-station angle intersection localization



The optimal radius is

$$d = 0.70387y$$

Geometry of the station distribution is shown in Fig. 2.11.

Then, optimal localization accuracy is

$$\sigma_x = \sigma_z = 0.52y\sigma_1, \quad \sigma_y = 1.23y\sigma_1, \quad F_{\min} = 2.0454y^2\sigma_1^2 \quad (2.13)$$

where y is the target altitude and σ_a is the angle measurement error.

It is thus clear that orbit determination error in angle intersection method also increases with the increase of the target altitude. Station distribution properties of “angle intersection” orbit determination are listed in Table 2.2.

2.2.2.3 “Range Intersection” Localization

This method is to draw multiple spheres taking the stations as the center and the range measured by multiple high-precision ranging systems as the radius. The point

Table 2.2 Optimal station distribution properties of angle intersection localization

S/N	Item	Tri-station intersection localization	Dual-station intersection localization
1	Geometric figure of optimal station distribution (takes the subsatellite point as the center)	Regular triangle	The baseline between two stations passes through the subsatellite point and is perpendicular to x -axis, and the distance between stations and the subsatellite point is equal
2	Radius of station distribution circle	$d = 0.70387y$	$d = 0.67183y$
3	Space intersection angle between two stations and the target	$\beta_{12} = \beta_{13} = \beta_{23} = 59^\circ 48'$	$\beta_{12} = 67^\circ 48'$
4	Elevation between each station and the target	$E_1 = E_2 = E_3 = 20^\circ 18'$	$E_1 = E_2 = 56^\circ 6'$
5	Slant range between each station and the target	$R_1 = R_2 = R_3 = 1.2229y$	$R_1 = R_2 = 1.2047y$
6	Baseline between stations	$D_{12} = D_{13} = D_{23} = 1.21914y$	$D_{12} = 1.34366y$
7	Minimum risk value	$F_{\min} = 2.0454y^2\sigma_a^2$	$F_{\min} = 3.61235y^2\sigma_a^2$
8	Optimal estimation accuracy of target location	$\sigma_x = \sigma_z = 0.52y\sigma_a$ $\sigma_y = 0.123y\sigma_a$	$\sigma_x = 0.4751y\sigma_a$ $\sigma_y = 1.5276y\sigma_a$ $\sigma_z = 1.0263y\sigma_a$
9	Included angle between the radius of station distribution circle and x -axis	$\theta_1 = \theta_1$ (arbitrary, take 30° in practice) $\theta_2 = \theta_1 + 120^\circ$, $\theta_3 = \theta_1 + 240^\circ$	$\theta_1 = 90^\circ$ $\theta_2 = 270^\circ$

of intersection of these spheres is the target location (three or more spheres must be provided).

(1) Tri-station range intersection localization

When three stations are located at the same station distribution circle,

$$F = \left\{ 1 + \left[\frac{2(1 - \cos \beta_{12} \beta_{13} \beta_{23})}{\Delta_1} \right] \right\} \sigma_R^2$$

where

$$\Delta_1 = 1 + 2 \cos \beta_{12} \beta_{13} \beta_{23} - \cos^2 \beta_{12} - \cos^2 \beta_{13} - \cos^2 \beta_{23}$$

where σ_R is ranging error of the measuring station, β_{12} is the space intersection angle of Station 1 and Station 2 at the target, β_{13} is the space intersection angle of Station 1 and Station 3 at the target, and β_{23} is the space intersection angle of Station 2 and Station 3 at the target.

Under the condition of $\beta_{12} = \beta_{13}$, the β_{12} which makes F minimum can be solved as

$$\cos \beta_{12} = \cos \beta_{13} = \cos \beta_{23}$$

Then,

$$\cos \beta_{12} = \cos \beta_{13} = \cos \beta_{23}$$

That is, the geometric figure of tri-station distribution is a regular triangle, and the subsatellite point is located at the center of the regular triangle, and

$$F = \sigma_R^2 + \sigma_R^2 \left\{ \frac{[8(y^2 + d^2) - (2y^2 - d^2)]^3}{27d^4 y^2} \right\} \quad (2.14)$$

It is clear that F also relates to d , i.e., F value changes with d .

In order to compare three localization methods more intuitively, the localization accuracies of RAE method and angle intersection method for geostationary satellite localization are estimated as follows.

For RAE single-station localization method, the expression of total linear error in case of optimal station distribution is

$$\sigma_L = \sqrt{F} = \sqrt{\sigma_R^2 + (2d_0^2 + y^2)\sigma_a^2} \quad (2.15)$$

When ranging error $\sigma_R = 4$ m, angle measurement error $\sigma_a = 0.01^\circ = 0.0001745$ rad, geostationary satellite altitude $y \approx 3.69 \times 10^4$ km, and $d_0 = 0$, through calculation we can get $\delta_L = 6,440$ m.

For tri-station angle intersection localization method, the expression of total linear error in case of optimal station distribution is

$$\sigma_L = \sqrt{F} = \sqrt{2.0454y^2\sigma_a^2} \quad (2.16)$$

When angle measurement error $\sigma_a = 0.01^\circ = 0.0001745$ rad and target altitude $y \approx 3.69 \times 10^4$ km, then we get $\sigma_L = 9,210$ m through calculation.

It can be seen from the above calculations that the tri-station range intersection localization method has the highest accuracy in geostationary satellite localization.

Let $\partial F / \partial d = 0$, the radius of the optimal station distribution circle can be obtained as

$$d = \sqrt{2}y \quad (2.17)$$

and so

$$F_{\min} = 3\sigma_R^2 \quad (2.18)$$

It is thus clear that optimal station distribution geometry in tri-station range intersection localization method can be summarized: three stations are distributed in a circle taking the subsatellite point as its center; the radius of the circle is $d = \sqrt{2}y$; three stations are arranged like a regular triangle.

(2) Four-station range intersection localization

When four stations are distributed on the same circumference, if the four stations constitute a rectangle, compared with other inscribed quadrilateral distribution schemes, it will have a minimum value of F , namely,

$$F = \frac{\sigma_R^2(y^2 + d^2)(4y^2 + d^2)}{4y^2d^2} \quad (2.19)$$

from which it can be seen that F varies with d . By letting $\partial F / \partial d = 0$, the d that makes F minimum can be obtained, i.e.,

$$d = \sqrt{2}y \quad (2.20)$$

and the minimum F is

$$F_{\min} = \left(\frac{9}{4}\right) \sigma_R^2 \quad (2.21)$$

It is clear that $F_{\min}$ is 1.33 times less than that of tri-station localization.

According to the foregoing, an optimal station distribution geometry in a four-station range localization method can be summarized: four stations are distributed in a circle taking the subsatellite point as its center; the optimum radius of the circle is $d = \sqrt{2}y$; four stations are arranged like a square.

Geometric graph of optimal station distribution in “range intersection” localization is shown in Fig. 2.12.

Optimal station distribution properties of range intersection localization are listed in Table 2.3.

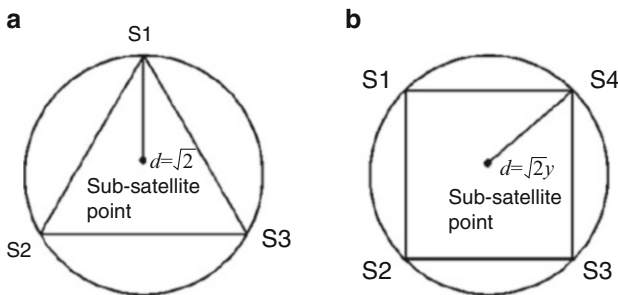


Fig. 2.12 Geometric graph of optimal station distribution in range intersection localization. (a) Tri-station intersection and (b) four-station intersection

Table 2.3 Optimal station distribution properties of range intersection localization

S/N	Item	Four-station intersection localization	Tri-station intersection localization
1	Geometric figure of optimal station distribution (takes the subsatellite point as the center)	Square	Regular triangle
2	Radius of station distribution circle	$d = \sqrt{2}y$	$d = \sqrt{2}y$
3	Space intersection angle between two stations and the target	$\beta_{12} = \beta_{14} = \beta_{23} = \beta_{34} = 70^\circ 30'$ $\beta_{13} = \beta_{24} = 109^\circ 30'$	$\beta_{12} = \beta_{13} = \beta_{23} = 90^\circ$
4	Elevation between each station and the target	$E_i = 35^\circ 16' (i = 1, 2, 3, 4)$	$E_i = 35^\circ 16' (i = 1, 2, 3)$
5	Slant range between each station and the target	$R_i = \sqrt{3}y (i = 1, 2, 3, 4)$	$R_i = \sqrt{3}y (i = 1, 2, 3)$
6	Baseline between stations	$D_{12} = D_{14} = D_{23} = D_{34} = 2y$ $D_{13} = D_{24} = 2\sqrt{2}y$	$D_{12} = D_{13} = D_{23} = \sqrt{6}y$
7	Minimum risk value	$F_{\min} = (9/4)\sigma_a^2$	$F_{\min} = 3\sigma_a^2$
8	Optimal estimation accuracy of target location	$\sigma_x = \sigma_z = \sigma_y = (\sqrt{3}/2)\sigma_R$	$\sigma_x = \sigma_y = \sigma_z = \sigma_R$
9	Included angle θ_1 between the radius of station distribution circle and x -axis	$\theta_1 = \theta_1$ (arbitrary, take 45° in practice) $\theta_2 = \theta_1 + \pi/2$, $\theta_3 = \theta_1 + \pi$ $\theta_4 = \theta_1 + \frac{3}{2}\pi$	$\theta_1 = \theta_1$ (arbitrary, take 30° in practice) $\theta_2 = \theta_1 + 120^\circ$ $\theta_3 = \theta_1 + 240^\circ$

According to the foregoing, it can be summarized as follows:

- (1) Localization accuracy of RAE method and angle intersection method decreases with the increase of spacecraft altitude y . So both methods are more applicable to LEO spacecraft, instead of HEO spacecraft.
- (2) Localization accuracy of range intersection method is independent of the spacecraft altitude and is only related to ranging accuracy σ_R in optimal station distribution. So it is most favorable to the localization of HEO spacecraft.
- (3) The optimal station distribution in range intersection localization is of a circle with the diameter equal to $\sqrt{2}y$. When $y = 4 \times 10^4$ km, the diameter is 5.6×10^4 km. It is clear that this could not be realized on the surface of the Earth. So it is impossible to realize optimal station distribution on the Earth. What can only be done is to expand the distance between orbit-measuring stations as much as possible. Then, specific station distribution geometry requires specific localization accuracy calculation.

2.2.3 Trajectory Measurement System

In order to measure the trajectory of the vehicle, at least six measurement elements must be measured. Several common systems composed of these measurement elements and their relevant characteristics are summarized as follows [4].

2.2.3.1 RAE System

The RAE system is also called single-station system or ranging and angle measurement system. Measurement elements R , A , and E are used to determine the location of the flight vehicle. Through differential smoothing, $\dot{R}$, $\dot{A}$, and $\dot{E}$ can also be used to determine the velocity of the flight vehicle. If higher velocity-measuring accuracy is required, the carrier Doppler velocity measurement can be used. Continuous wave radar, pulse radar, and optical theodolite with a laser ranging device can constitute this system. Due to limited angle measurement accuracy, this system has poor measurement accuracy for long-range targets. So it is a mid- and low-accuracy measurement system. It is a main system in satellite TT&C systems.

2.2.3.2 Multi- RR System

The multi- RR system is also called the range intersection system, in which each station independently measures R and $\dot{R}$ (the number of stations ≥ 3) and baseline is generally several hundred kilometers to several thousand kilometers. The transponder operating in this multi-receiver and multi-transmitter system must make a multichannel response in frequency division or code division method. So the multichannel transponder is one of the key issues. This system is applicable to intersection measurement using multiple pulse radars and continuous wave radars. Without need of baseline transmission, each station is relatively independent. So the station is distributed in a flexible way. According to the accuracy of the measuring station, a high- and mid-accuracy measurement system can be constituted. Currently, it has been widely used in missile high-accuracy measurement system, satellite TT&C system, and satellite navigation system and is a promising high-accuracy measurement system.

2.2.3.3 Rlm System

The Rlm system is also called short-baseline interferometer system. It consists of one master station and two slave stations. In general, the baseline is very short, only tens of meters to several kilometers. The range between the master station and the

target is R ; direction cosines of range difference between master station and slave station are l and m . Measurement elements of this system are $R, l, m, \dot{R}, \dot{l}$, and $\dot{m}$. It uses carrier frequency signals to measure range difference and has very high accuracy. However, due to a short carrier frequency period, the range ambiguity is very serious. So it is required to place multiple antennas on the baseline to solve the ambiguity. Under short-baseline conditions, ambiguity resolving is easy. But if the baseline is too long, the number of antennas will increase, the system will be complex, and the cost will be very expensive. Under short baseline conditions, by dividing the range difference by the baseline length, one can obtain direction cosines l and m . Signals between master station and slave station can be transmitted through phase-compensated cables or optical fibers. Due to lower measurement accuracy of direction cosines, this system is applicable to mid-accuracy measurement.

2.2.3.4 Rr_i System

The Rr_i system is also called medium- and long-baseline interferometer system. Measurement elements of medium- and long-baseline interferometer system are $R, r_i, \dot{R}$, and $\dot{r}_i$ ($i = 1, 2, \dots, n, n \geq 2$, r is the range difference between master station and slave station). Baseline length is generally dozens of kilometers, i.e., LOS. Signals can be transmitted through microwave communication or optical fibers. As the baseline length increases, range ambiguity when measuring range difference by the carrier will be more serious. So we have to use modulation signals to measure the range difference. The accuracy of measuring range difference in this system is lower than short-baseline system. But because the baseline is longer, equivalent angle measurement accuracy is higher than short-baseline system. When signals are transmitted via a microwave link between master and slave stations, it is required to build a baseline transmission tower. Standard frequency of master station should be sent to slave stations as their frequency reference; slave stations should convert and transmit received signals to the master station. The master station will extract, record, and transmit in real time information about velocity-measuring and ranging of master and slave stations to the data processing center, while the slave station will only receive and retransmit the measurement information, without need of timing and data recording devices.

2.2.3.5 Multi-AE System

The multi-AE system is also called angle intersection system. It can be constituted by multiple interferometers or by a combination of a cinetheodolite, electrooptic theodolite, and monopulse radar. Its measurement elements are multiple groups of A and E . Radar and optical observations are independent. There is no relation

between devices. Associated instrumentation refers to combined utilization in data. In principle, arbitrary combination is available. At present, with technical development, the baseline uses optical fiber transmission and connected element interferometer (CEI), and very long baseline interferometer with intercontinental baseline appeared, thus significantly improving measurement and localization accuracy. Since only a beacon is installed on board to perform broadcast transmission and one-way measurement is performed on ground to implement passive tracking, equipment is simplified. Moreover, there is no rang system error introduced by the transponder in ranging, and radio star can be used for angle calibration, so another idea is provided for improving localization accuracy.

2.2.3.6 Multi- $\dot{SS}$ System

The multi- $\dot{SS}$ is also called $k\dot{SS}$ system. This system requires measuring the range sum and its rate ($k \geq 3$) only. Generally, it consists of one transmitting station and multiple receiving stations. It can be classified into medium-baseline system and long-baseline system. There are two methods in measuring range sum, namely, modulation signal ranging method and integral ranging method.

(1) Medium-baseline $k\dot{SS}$ system

Its system structure is very similar to that of a medium-baseline interferometer. The difference is that information transmission on baseline is in opposite direction. The transmitting station transmits time and frequency standards to the receiving station, and the receiving station performs $\dot{SS}$ measurement. The $k\dot{SS}$ system and interferometer have no essential distinction. They both use $\dot{SS}$ measured by the stations to calculate $r\dot{r}$.

(2) Long-baseline $k\dot{SS}$ system

This system features a single-carrier frequency signal and simple equipment. However, due to long baseline, if integral ranging method is used, other measures provided by external equipment or the internal system are required to provide the start point of integral, and no contingency interrupt is allowed for signals. All stations should provide timing, frequency reference, communication, and data recording devices to independently measure $\dot{S}$. No signal is transmitted between stations. Due to long baseline, this system is applicable to high-accuracy measurement.

The above-mentioned systems are quite distinct from each other. In selecting a system, one should take an overall consideration in terms of system engineering according to measurement accuracy requirements, operating requirements, and cost.

2.3 Velocity-Measuring Theory and Technology: Two-Way, Three-Way, and Single-Way Velocity-Measuring Technologies

2.3.1 CW Velocity-Measuring

The principle of measuring the target velocity is that, based on the Doppler effect in physics, when CW radar wave illuminates the target, the target will move and produce Doppler frequency which is in direct proportion to the velocity of the target relative to radar movement and in inverse proportion to wavelength λ . When the target flies to the radar station, Doppler frequency will be positive and the frequency of receiving signals is higher than that of transmitting signals. When the target departs from the radar, Doppler frequency will be negative and the frequency of receiving signal is lower than that of transmitting signals.

The velocity can be measured by measuring the Doppler frequency shift. The most basic method is the two-way carrier Doppler velocity-measuring. The two-way carrier Doppler velocity-measuring system is mainly composed of a ground transmitter, flight vehicle transponder, and ground receiver. Generally, a phase-lock coherent transponder is used. A noncoherent transponder with independent local oscillator is also available. "Coherent" refers to a certain phase relation between output and input signals; otherwise, it is "noncoherent." Velocity measurement corresponding to the former is called coherent Doppler velocity-measuring. Velocity measurement corresponding to the latter is called noncoherent Doppler velocity-measuring.

(1) Two-way Doppler velocity-measuring with coherent transponder

For coherent transponders, because after an uplink one-way Doppler frequency is retransmitted by a coherent transponder with a coherent turnaround ratio $1/\rho = f_{\text{uplink}}/f_{\text{downlink}}$, Doppler frequency is transformed $1/\rho$ times. The combined Doppler frequency of uplink and downlink is almost twice of downlink one-way Doppler frequency. It is well known that the radial velocity of the target is:

$$v = -\frac{f_d}{2f_r + f_d}c, \quad (2.22)$$

where f_r is the downlink receiving frequency on ground.

The coherent transponder is generally a phase-lock transponder. Figure 2.13 shows a common solution.

In Fig. 2.13, the receiving frequency of the transponder is

$$f_{r\text{Trans}} = \left(N_1 + N_2 + \frac{1}{2}\right)f_v$$

where f_v is the VCO frequency.

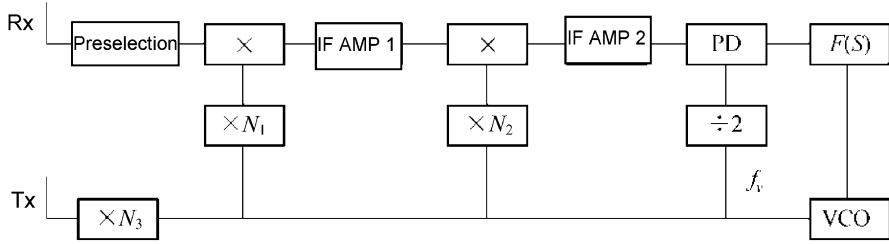


Fig. 2.13 Principle block diagram of phase-lock transponder

Transmitting frequency is

$$f_{bTrans} = N_3 f_v$$

Phase-lock transponder turnaround ratio is

$$\rho = \frac{f_{bTrans}}{f_{Trans}} = \frac{N_3}{N_1 + N_2 + 1/2} = \frac{f_{downlink}}{f_{uplink}} \quad (2.23)$$

(2) Two-way Doppler velocity-measuring with noncoherent transponder

A noncoherent transponder can also be used for a two-way Doppler velocity measurement system. But in order to improve velocity measurement accuracy, reasonable measures should be taken to eliminate noncoherent components from transponder noncoherent LO. According to different elimination schemes, the method can be classified as “ground station canceling method,” “transponder canceling method,” and “transponder measuring and ground processing method.” These three methods will be introduced in detail below. Noncoherent transponder scheme is often used.

1) Mixed Repeating Noncoherent Transponder. As shown in Fig. 2.14, this transponder has an independent crystal oscillator with frequency of f_L . The input signal will be down-converted, amplified, filtered, and up-converted and then retransmitted via a power amplifier. Down-conversion LO is $(m-1)f_L$, up-conversion LO is mf_L , and the receiving frequency of the transponder is the sum of uplink transmitting frequency f_i and uplink Doppler frequency f_{id} , namely, $f_{rTrans} = f_i + f_{id}$.

Being converted twice, it becomes

$$f_{bTrans} = f_i + f_{id} + f_L$$

Transmitting frequency contains independent LO component f_L . In order to eliminate the incoherence, mf_L , the m -time multiplier of f_L , should be transmitted to the ground station to eliminate f_L incoherence in the process of extracting Doppler frequency. Figure 2.15 is the block diagram showing the ground station extracts of Doppler frequency.

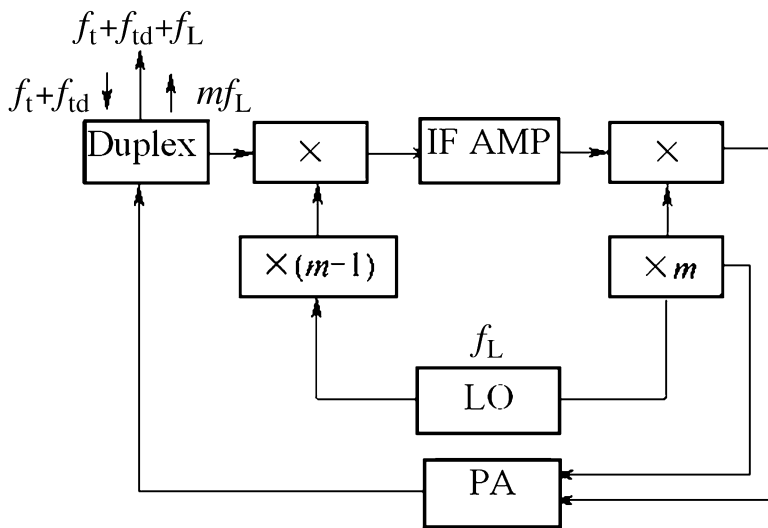


Fig. 2.14 Mixed repeating noncoherent transponder

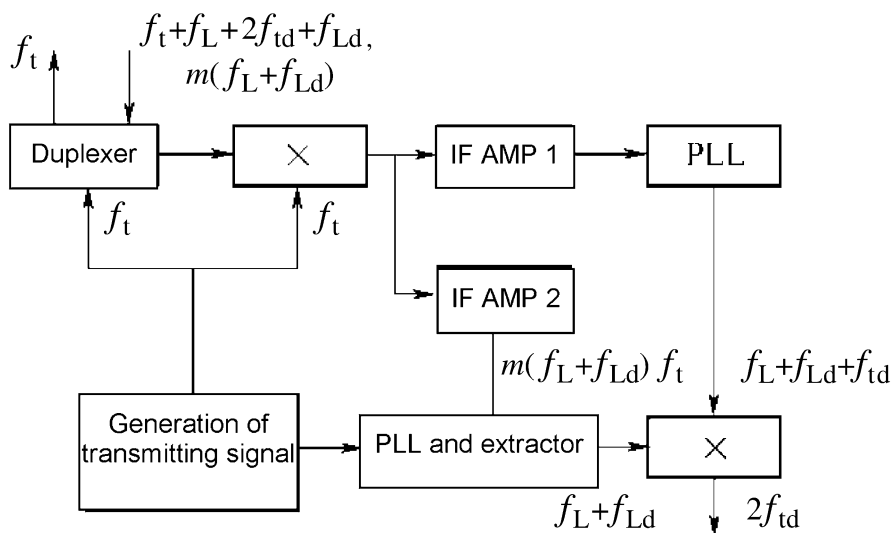


Fig. 2.15 Block diagram of Doppler frequency extraction in ground stations

Downlink signal returned from the transponder enters into the receiver, which contains two signal frequencies: $f_t + f_L + 2f_{td} + f_{Ld}$ and $m(f_L + f_{Ld})$, where f_{Ld} and f_{td} refer to one-way Doppler frequencies of the independent master oscillator of the transponder and of ground-transmitting frequency, respectively. Two signal frequencies are mixed with transmitting frequency f_t and filtered by two IF amplifiers at different frequencies and then extracted with the phase-locked loop. IF amplifier

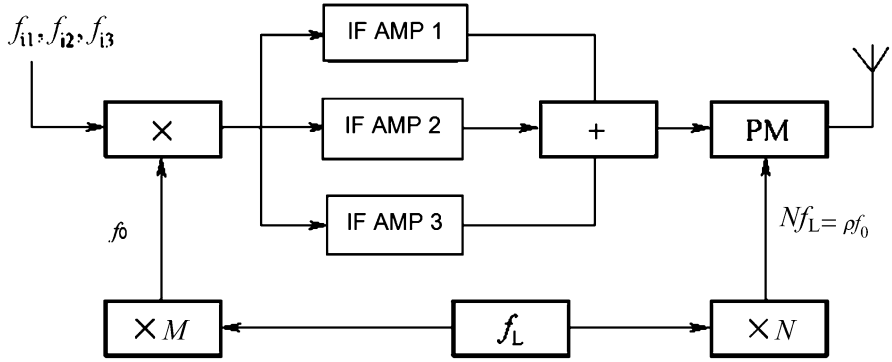


Fig. 2.16 IF modulation repeating noncoherent transponder

1 extracts and outputs $f_L + f_{Ld} + 2f_{id}$; the branch of IF amplifier 2, PLL and extractor, outputs $f_L + f_{Ld}$. After mixing and subtracting the two outputs, pure Doppler frequency $2f_{id}$ will be outputted, thus eliminating the incoherence caused by the independent crystal oscillator of the transponder.

2) Modulation Repeating Noncoherent Transponder. As shown in Fig. 2.16, the transmitting signal of the transponder is a constant envelope. So the power amplifier can operate in a saturated mode, thus avoiding the effect of intermodulation interference and amplitude/phase conversion on measurement accuracy. A transponder is used to repeat signals from three ground stations. The carrier frequency difference among three ground stations is small, several megahertz in general. The transponder receives transmitting frequencies from three ground stations, converting them into IF f_{i1} , f_{i2} , and f_{i3} via same mixer. Noncoherent LO frequency f_0 is M -time frequency multiplication of f_L . Three IF frequencies are filtered and amplified by three IF amplifiers in parallel and summed to modulate the phase of transmitting frequency of the transponder Nf_L and then transmitted to the ground stations. Since three phase modulation subcarriers (i.e., three intermediate frequencies) and the transmitting frequency of the transponder contain noncoherent components of master oscillator f_L , after phase-locking and extracting the main carrier and subcarrier in ground stations and proper frequency adding and subtracting, noncoherent components of master oscillator of the transponder f_L can be eliminated to obtain pure Doppler frequency f_D (Fig. 2.17).

2.3.2 Doppler Frequency Measurement Method

Currently, the digital phase-locked loop (PLL) velocity measurement method has been widely used. Previous methods such as “determination of integer ambiguity by fixed time,” “time determination by fixed integer ambiguity,” and “determination of integer ambiguity by basic fixed time” are used in certain special circumstances

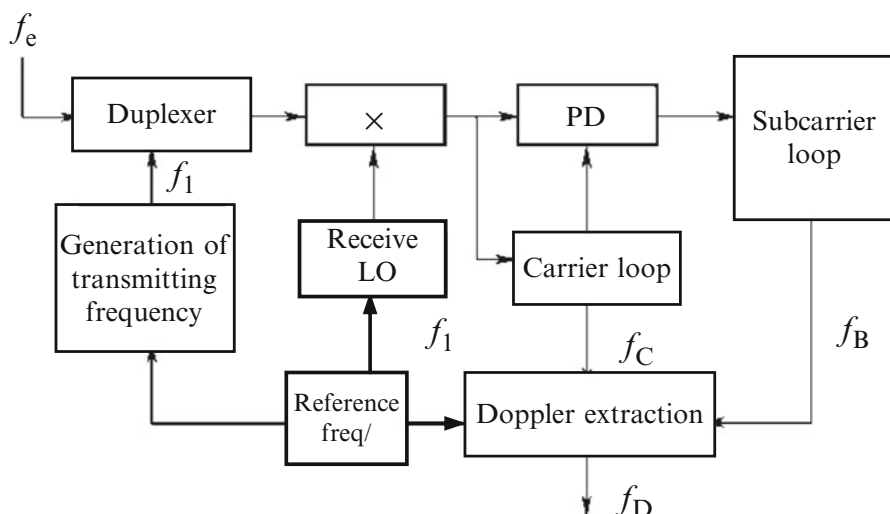


Fig. 2.17 Doppler extraction in ground station for IF modulation transponder

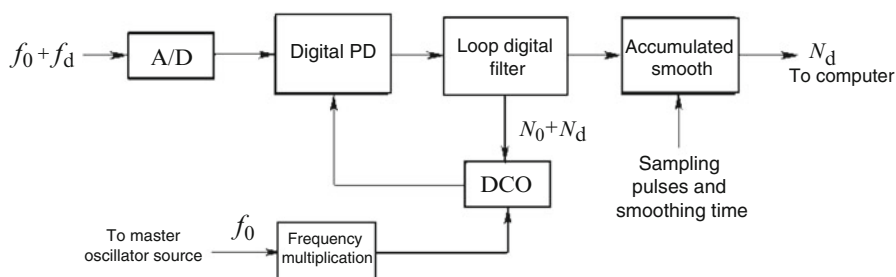


Fig. 2.18 The extraction and measuring of Doppler frequency in IF digital phase-locked loop

only. Now, the digital PLL of the phase-locked receiver captures and tracks the downlink signals and extracts and measures carrier Doppler frequency. This not only realizes software technology and multiple functions but also greatly simplifies velocity-measuring terminal equipment. Moreover, equipment simplification reduces additional phase noise and increases velocity measurement accuracy.

As shown in Fig. 2.18, IF digital PLL is mainly composed of digital circuits, which eliminate the slow drift in simulating parameters of phase discriminator, active filter, and VCO in the PLL with the change of time, temperature, and level. The phase discriminator of the digital PLL consists of a digital multiplier and an accumulator. The loop filter is a digital filter, generally a digital signal processor (DSP). The digital-controlled oscillator (DCO) is composed of a digital synthesizer driven by externally high-frequency clock. Output frequency and phase are under control by output of the loop filter. After loop-locking, the DCO frequency is equal to the frequency of input signals. The variation of DCO frequency control code

fully corresponds to the frequency change of input signals, with a correct mathematical corresponding relation.

Multiplying the variation of frequency control code by determined conversion coefficient equals Doppler frequency shift of input signals. Then, output frequency control code finishes extraction and measurement of Doppler frequency, and by continuously outputting frequency control code (at an interval of 5-loop filter operation repetition period), the velocity value can be continuously outputted.

Certainly, frequency code will be affected and changed by input thermal noise. In order to eliminate the effect of thermal noise, frequency code should be smoothed and filtered. In general, accumulative smooth is used. Accumulative start time is under the control of sampling pulse, and smooth time is also controlled by external commands. After smooth, frequency code will be sent to the computer for conversion into velocity data.

Given the frequency control coefficient of the DCO is k_i , input frequency is $f_{io} + f_d$, center frequency f_{io} , and loop filter outputs control code $N_0 + N_d$, then $N_0 = f_{io}/k$, $N_d = f_d/k_i$. The average value of Doppler frequency code of the accumulated smoothing outputs is N_d , and average value of Doppler frequency $f_D = k_i N_d$.

In order to eliminate noncoherent changes of DCO-driven clock frequency and avoid noncoherent velocity-measuring error, high-frequency clock of DCO shall be generated by multiplication of master oscillator frequency f_0 of the transmitter to make transmitter frequency source and receiver Doppler extraction circuit form a complete coherent system. Moreover, master oscillator f_0 is of good short-term stability, thus decreasing velocity-measuring error caused by introduction of Doppler frequency measurement reference.

The digital method often has quantization error. In design, efforts should be made to reduce the quantization error to an extent much smaller than the error caused by inputting thermal noise. When the DSP is used as the loop filter arithmetic unit, a 16-bit operation or double-precision 32-bit operation is often used. 16-bit operation can be used in loop filter operation. Loop self-adjustment can be used to eliminate the effect of operation quantization error on DCO mean frequency. The resolution of frequency control code outputted after operation will be guaranteed jointly by loop filter output word length and DCO frequency adjustment accuracy. For design, first determine the shortest word length of f_d code according to velocity range and velocity resolution. Given maximum velocity of the flight vehicle 12 km/s, velocity resolution 0.1 cm/s, minimum capacity of f_d code $12 \text{ km}/0.1 \text{ cm} = 1.2 \times 10^7$, and $2^{24} = 1.68 \times 10^7$, the shortest word length of f_d code is a 24-bit binary code. Then, derive DCO frequency control resolution from the velocity resolution.

Given DCO frequency adjustment step size ΔF and velocity resolution Δv , for a two-way Doppler, there is

$$\Delta F = \frac{2\Delta V}{c} f_r \quad (2.24)$$

When $f_r = 2 \text{ GHz}$, $\Delta V = 0.1 \text{ cm/s}$, we can obtain a maximum DCO frequency adjustment step size of 0.013 Hz.

DCO can be implemented by the DDS which consists of binary phase accumulator and digital sine (cosine) table. Frequency control resolution depends on bits of phase accumulator and clock frequency, namely,

$$\Delta F = \frac{f_{\text{op}}}{2^n} \quad (2.25)$$

The above discussion about the quantization is a sufficient condition, but not a necessary condition. When the DCO frequency adjustment step size is longer or loop filter operation word length is shorter, DCO mean frequency is always equal to input signal frequency due to the loop self-adjustment. During locking process, DCO control code fluctuates around the control value of the mean frequency, and time average of the control code will be a mixed decimal. Thus, when DCO or loop filter operation resolution is not high enough, output resolution can be improved through time accumulation of frequency control code. Certainly, instantaneous or short-time high resolution could not be obtained. Time accumulation of frequency control code can both eliminate the effect of thermal noise and improve velocity resolution in times equal to the number of smooth.

2.3.3 *Theoretical Analysis of Two-Way Coherent Doppler Velocity-Measuring Error*

In the radio velocity measurement technology field, continuous wave coherent Doppler velocity measurement is a velocity measurement system with highest accuracy. When a higher accuracy is required, short-term stability of velocity measurement signals is an important factor affecting the velocity measurement accuracy. This is a special issue in continuous wave velocity measurement technology. This section will first give an introduction to the issues on signal short-term stability.

2.3.3.1 Phase Noise Power Spectrum Density and Allan Variance [4, 6]

Phase noise is often characterized by several spectrum density functions:

- (1) $S_{\Phi}(f)$: baseband spectrum of phase noise $\Phi(t)$
- (2) $P_{\text{RF}}(f)$: RF signal amplitude spectrum observed on the spectrum analyzer, by which P_s/N_0 can be obtained
- (3) $\mathcal{E}(\Delta f)$: phase noise spectrum value displayed on the phase noise tester, also called normalized spectrum
- (4) $S_y(f)$: baseband spectrum of frequency fluctuation noise

The said spectrums are common one-sided spectrums in engineering. $S_{\Phi}(f)$ is a low-pass one-sided spectrum of phase noise modulation signal $\Phi(t)$, used for

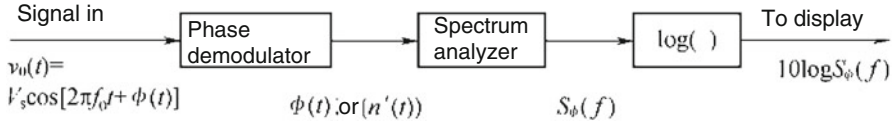


Fig. 2.19 Block diagram of $S_{\Phi}(f)$ phase noise analyzer

measuring $S_{\Phi}(f)$, as shown in Fig. 2.19. The measuring instrument consists of a $\Phi(t)$ phase demodulator, a low-spectrum analyzer receiving $S_{\Phi}(f)$ signals, and a logarithmic converter used for display. The logarithmic scale is usually used to display the relation between $10 \log S_{\Phi}(f)$ and frequency. $S_{\Phi}(f)$ is expressed in rad^2/Hz , which should be explained to be dB relative to $1 \text{ rad}^2/\text{Hz}$, with respect to the dB scale.

In engineering practice, although $S_{\Phi}(f)$ is really measured, $10 \log \mathcal{L}(\Delta f)$ is always displayed in the ordinate. This value is considered to be a phase noise within 1 Hz bandwidth in a sideband at the frequency deviation. If phase noise amplitude is small enough, $\mathcal{L}(\Delta f) \approx S_{\Phi}(f)/2$. When it is expressed in logarithm, by increasing 3 dB in low-pass spectrum of $10 \log [\mathcal{L}(\Delta f)]$, one can obtain a correct $10 \log [S_{\Phi}(f)]$.

The above-mentioned spectrum density functions maintain a relationship, namely,

$$S_{\Phi}(f) = S_y(f) \left(\frac{f_0}{f} \right)^2 \quad (2.26)$$

$$\mathcal{L}(\Delta f) \approx \frac{S_{\Phi}(f)}{2} \quad (2.27)$$

When white Gaussian noise is added, phase noise spectrum density is

$$N_{\Phi}(f) = W'_n(f) = \frac{2N_0}{V_s^2} = \frac{N_0}{P_s} \quad (2.28)$$

where P_s is signal power, N_0 is noise power spectrum density, $N_{\Phi}(f)$ is the spectrum density of random variable $n'(t)$ at output end of phase demodulator, and V_s is signal amplitude.

Another characterization method of phase noise spectrum density is Allan variance. The Allan variance $\sigma_y^2(\tau)$ and phase noise power spectrum density are two features both used to describe the short-term frequency stability of one signal. $S_{\Phi}(f)$ is used for frequency domain and $\sigma_y^2(\tau)$ is for time domain. However, with respect to their physical essence, they are indexes used to describe the same short-term stability. They are related to each other through Fourier transform.

The relationship between phase noise power density $S_{\Phi}(\omega)$ and Allan variance $\sigma_y^2(\tau)$ will be represented in three methods.

(1) Integral representation method

A lot of documents have given the relational expression of $\sigma_y^2(\tau)$ and $S_y(\omega)$ [4]:

$$\sigma_y^2(\tau) = \int_0^\infty S_y(f) \left[\frac{2 \sin^4(\pi f \tau)}{(\pi f \tau)^2} \right] df = \int_0^\infty S_\Phi(f) \frac{f^2}{f_0^2} \left[\frac{2 \sin^4(\pi f \tau)}{(\pi f \tau)^2} \right] df \quad (2.29)$$

where τ is sampling interval (s) of measuring Allan variance; f is frequency interval (Hz) deviating from carrier frequency f_0 ; $S_y(f)$ is power spectrum density of relative frequency fluctuation $\Delta f/f_0$, short for relative frequency fluctuation spectrum density; and $S_\Phi(f) = S_y(f)(f_0/f)^2$ is phase fluctuation spectrum density.

Expression (2.29) shows that the solution of Allan variance in time domain is equivalent to $H_y(f) = (2 \sin^4(\pi f \tau)/(\pi f \tau)^2)$ characteristic filtering in frequency domain $S_\phi(\omega)$, where $H_y(f)$ is the transfer function of $S_y(f)$ Allan variance and $H_\Phi(f) = \left[\frac{2 \sin^4(\pi f \tau)}{(\pi f \tau)^2} \left(\frac{f}{f_0} \right)^2 \right]$ is the transfer function of $S_\phi(f)$ Allan variance.

(2) Analytic representation method

In theoretical analysis, $S_\phi(f)$ is often of a power law spectrum model. It is an approximate assumption. The model is

$$S_\phi(f) = h_{-2} f_0^2 f^{-4} + h_{-1} f_0^2 f^{-3} + h_0 f_0^2 f^{-2} + h_{+1} f_0^2 f^{-1} + h_{+2} f_0^2 \quad (2.30)$$

$$S_y(f) = h_{-2} f^{-2} + h_{-1} f^{-1} + h_0 f^0 + h_{+1} f^{+1} + h_{+2} f^{+2} \quad (2.31)$$

where $S_\phi(f)$ is expressed in rad^2/Hz and $S_y(f)$ is expressed in $(\text{rad/s})^2/\text{Hz}$.

Items in the above expression are in direct proportion to powers of f , so it is called power law spectrum. The relationship between these five noise types and spectrum density is shown in Table 2.4.

In Table 2.4, α is each power component of power law spectrum expression in the frequency domain, representing the type of frequency fluctuation noise. Similar to the frequency domain, in the time domain, power law expression of Allan variance is $\sigma_y^2(\tau) = \sum_{\alpha=-2}^2 D_\alpha \tau^\mu$, where μ is the power, $\mu = -\alpha - 1$ (if $\alpha > 1$, $\mu = -2$), representing the type of frequency fluctuation noise in the time domain and D_α and h_α represent the intensity of corresponding noise. A different oscillator has a different noise type. Quartz crystal oscillator mainly has three types of noise, namely, PM white noise, FM white noise, and FM flicker noise. For atomic clocks, the phase noise spectrum is greatly narrowed.

Table 2.4 Spectrum densities of five types of noise

Noise type	α	$S_y(f)$	$S_\phi(f)$	μ
PM white noise	2	$h_2 f^2$	$h_2 f_0^2$	-2
PM flicker noise	1	$h_1 f$	$h_1 f_0^2 f^{-1}$	-2
FM white noise	0	h_0	$h_0 f_0^2 f^{-2}$	-1
FM flicker noise	-1	$h_{-1} f^{-1}$	$h_{-1} f_0^2 f^{-3}$	0
Frequency random walk	-2	$h_{-2} f^{-2}$	$h_{-2} f_0^2 f^{-4}$	1

$S_\phi(f)$ is expressed in a logarithmic way:

$$[S_\phi(f)]_{-2} = 10\lg h_{-2}f_0^2 - 40\lg f, \quad (f_1 \sim f_2 \text{ interval, frequency random walk noise}) \quad (2.32)$$

$$[S_\phi(f)]_{-1} = 10\lg h_{-1}f_0^2 - 30\lg f, \quad (f_2 \sim f_3 \text{ interval, FM flicker noise}) \quad (2.33)$$

$$[S_\phi(f)]_0 = 10\lg h_0f_0^2 - 20\lg f, \quad (f_3 \sim f_4 \text{ interval, FM white noise}) \quad (2.34)$$

$$[S_\phi(f)]_{+1} = 10\lg h_{+1}f_0^2 - 10\lg f, \quad (f_4 \sim f_5 \text{ interval, PM flicker noise}) \quad (2.35)$$

$$[S_\phi(f)]_{+2} = 10\lg h_{+2}f_0^2, \quad (f_5 \sim f_6 \text{ interval, PM white noise}) \quad (2.36)$$

where $S_\phi(f)$, expressed in dBc/Hz, is the relative spectrum density relative to the carrier.

The physical essence of the above expression is to use broken lines to approximate actual phase noise curves (represented by logarithmic coordinates). Each broken line segment has 40, 30, 20, and 0 dB/10 octave slope, representing one phase noise type respectively. h_{-2} , h_{-1} , h_0 , h_{+1} , and h_{+2} are the heights of each broken line, respectively, representing phase noise amplitude, called intensity coefficient, of each type of phase noise.

The curve graph approximated by phase noise power law spectrum broken line is shown in Fig. 2.20 (approximation of three segments $[S_\phi(f)]_{-1}$, $[S_\phi(f)]_0$, and $[S_\phi(f)]_{+2}$).

The above logarithmic expressions are common approximate representations of phase noise index in engineering. If the above curves of signals are measured, when $\lg f = 0$, one can solve h_{-2} , h_{-1} , h_0 , h_{+1} , and h_{+2} in Expressions (2.32), (2.33), (2.34), (2.35), and (2.36), namely,

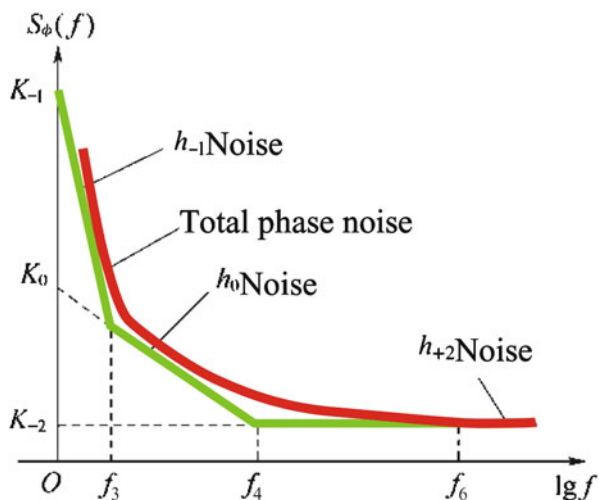


Fig. 2.20 Phase noise power law spectrum density curve

$$h_{-2} = f_0^{-2} \lg^{-1}(0.1K_{-2}) \quad (2.37)$$

$$h_{-1} = f_0^{-2} \lg^{-1}(0.1K_{-1}) \quad (2.38)$$

$$h_0 = f_0^{-2} \lg^{-1}(0.1K_0) \quad (2.39)$$

$$h_{+1} = f_0^{-2} \lg^{-1}(0.1K_{+1}) \quad (2.40)$$

$$h_{+2} = f_0^{-2} \lg^{-1}(0.1K_{+2}) \quad (2.41)$$

where K_{-2} , K_{-1} , K_0 , K_{+1} , and K_{+2} refer to the values of the intersection points of each approximate broken line of h_{-2} , h_{-1} , h_0 , h_{+1} , and h_{+2} with the vertical axis at $\lg f = 0$, respectively, which can be solved by the phase noise spectrum density curve. They can either be given by the analytic expression of Expressions (2.32), (2.33), (2.34), (2.35), and (2.36) or directly solved by Expressions (2.37), (2.38), (2.39), (2.40), and (2.41). Furthermore, it can be seen from the said expression that when f_0 increases N times, dB in the first item of the expression will increase $20 \lg N$. Therefore, when the frequency reference source is the same (e.g., 10 MHz frequency standard), due to its h_{-2} , h_{-1} , h_0 , h_{+1} , and h_{+2} are fixed, its phase noise index will be different after N times of ideal frequency multiplication to generate f_0 , i.e., will increase by $20 \lg N$.

By integral solving through Expression (2.29), one can obtain Allan variance corresponding to each item of $S_\phi(f)$. For example, for h_{+2} item (PM white noise), $S_\Phi = h_{+2}f_0^2$.

$$\begin{aligned} [\sigma_y(\tau)]_{+2}^2 &= \frac{2h_{+2}}{\pi^2\tau^2} \int_{f_5}^{f_6} \sin^4(\pi\tau f) df = \frac{2h_{+2}}{\pi^2\tau^2} \left\langle \frac{3(f_6 - f_5)}{8} \right. \\ &\quad \left. - \frac{1}{4\pi\tau} \cdot \left\{ \sin^3[\pi\tau(f_6 - f_5)] \cos[\pi\tau(f_6 - f_5)] + \frac{3}{4} \sin[2\pi\tau(f_6 - f_5)] \right\} \right\rangle \end{aligned} \quad (2.42)$$

where f_6 is cutoff frequency of integral operation in calculating short-term stability, corresponding to effective bandwidth the phase noise may give an effect on.

For example, in a phase-locked receiver, it refers to a single-side bandwidth of the phase-locked loop. This relationship shows that short-term stability relates to the phase-locked receiver bandwidth. Since short-term stability affects the velocity measurement accuracy, velocity measurement accuracy relates to the phase-locked loop bandwidth. When $f_6 - f_5 \gg 0.2/\tau$, PM white noise is

$$[\sigma_y(\tau)]_{+2}^2 \approx \frac{3h_{+2}(f_6 - f_5)}{4\pi^2\tau^2} \quad (2.43)$$

Similarly, the following items can be solved by integral [5], that is, PM flicker noise:

$$[\sigma_y(\tau)]_{+1}^2 \approx \frac{1.038 + 3[\ln 2\pi\tau(f_5 - f_4)]}{4\pi^2\tau^2} h_{+1}, \quad \text{if } 2\pi\tau(f_5 - f_4) \gg 1 \quad (2.44)$$

FM white noise:

$$[\sigma_y(\tau)]_0^2 \approx \frac{h_0}{2\tau} \quad (2.45)$$

FM flicker noise:

$$[\sigma_y(\tau)]_{-1}^2 = 2h_{-1}\ln 2 \quad (2.46)$$

Frequency random walk:

$$[\sigma_y(\tau)]_{-2}^2 \approx \frac{2\pi^2\tau h_{-2}}{3} \quad (2.47)$$

After solving the five items of Allan variance, total Allan variance is

$$[\sigma_y(\tau)]_{\Sigma}^2 = [\sigma_y(\tau)]_{-2}^2 + [\sigma_y(\tau)]_{-1}^2 + [\sigma_y(\tau)]_0^2 + [\sigma_y(\tau)]_{+1}^2 + [\sigma_y(\tau)]_{+2}^2 \quad (2.48)$$

The above six expressions have the following meanings in the engineering application:

1) For each type of phase noise, when $\sigma_y(\tau_1)$ of a τ_1 value is given, one can solve its corresponding h_{-2} , h_{-1} , h_0 , h_{+1} , and h_{+2} and then obtain $\sigma_y(\tau_x)$ of any other τ_x value.

2) By measuring $\sigma_y(\tau_1)$, $\sigma_y(\tau_2)$... at different τ_1 , τ_2 ..., one can obtain the $\sigma_y^2(\tau)$ to τ curve and then approximately obtain the distribution frequency interval (corresponding to $f=1/2\tau$) of each type of phase noise from the above expression and thus determine the type of phase noise. For example, when $\sigma_y^2(\tau)$ is in direct proportion to $1/\tau$, the noise is h_0 phase noise (if the phase noise could not be determined due to too large interval of τ_1 , τ_2 ..., reduce $\Delta\tau$ until the phase noise can be determined), and then solve h_0 . Similarly, h_{+2} , h_{+1} , and h_{-2} can be solved, and thus phase noise characteristic $S_\phi(f)$ will be obtained.

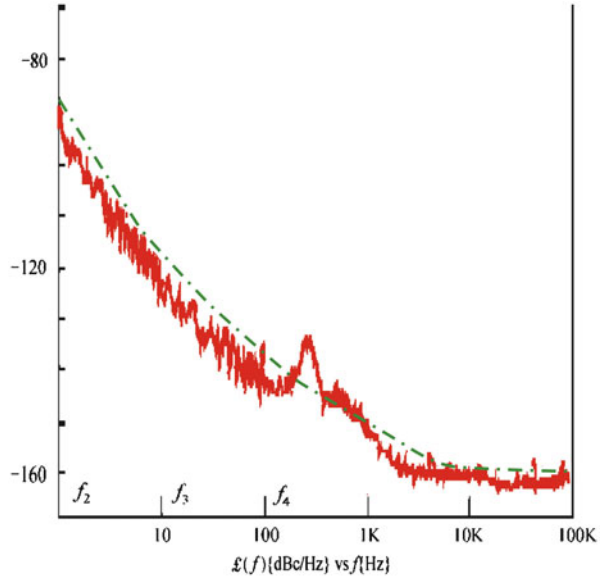
3) If the power law spectrum characteristic curve is given, one can obtain Allan variance $\sigma_y(\tau)$.

The following is an example by using measured phase noise curve of a crystal oscillator to solve corresponding $\sigma_y(\tau)$.

As shown in Fig. 2.21, spectrum density of measured phase noise is $\mathcal{L}(f)$. As $S_\phi(f) = 2\mathcal{L}(f)$ has been described in previous sections, use a power law spectrum broken line 3 dB more than $\mathcal{L}(f)$ in the figure to obtain an approximate $S_\phi(f)$ (refer to Fig. 2.20, shown in a dashed line). Then, the values of intersection points of the extended line of each broken line segment with $\lg(f) = 0$ axis are

$$\begin{aligned} K_{-1} &= -88 \text{ dBc/Hz}, & K_0 &= -98 \text{ dBc/Hz} \\ K_{+1} &= -118 \text{ dBc/Hz}, & K_{+2} &= -158 \text{ dBc/Hz} \end{aligned}$$

Fig. 2.21 Measured phase noise and short-term stability of a 10 MHz crystal oscillator



and the corner frequencies of each broken line are

$$f_2 = 1 \text{ Hz}, f_3 = 10 \text{ Hz}, f_4 = 100 \text{ Hz}, f_5 = 4 \text{ kHz}$$

From Fig. 2.21, one can get $f_6 = 6.3 \times 10^3 \text{ Hz}$ since $\sigma_y(\tau)$ is measured at a cutoff frequency of $6.3 \times 10^3 \text{ Hz}$.

According to the above known parameters, one can use Expression (2.39) to obtain

$$h_0 = f_0^{-2} \lg^{-1}(0.1K_0) = (10 \times 10^6)^{-2} \lg[0.1 \times (-98)] = 1.58 \times 10^{-24} (\text{rad})^2 / \text{Hz}$$

Similarly,

$$\begin{aligned} h_{-1} &= 1.58 \times 10^{-23} (\text{rad})^2 \\ h_{+1} &= 1.58 \times 10^{-26} (\text{rad})^2 / \text{Hz}^2, \\ h_{+2} &= 1.58 \times 10^{-30} (\text{rad})^2 / \text{Hz}^3 \end{aligned}$$

Then, by Expression (2.45) one can obtain

$$\begin{aligned} [\sigma_y(20 \text{ ms})]_0^2 &= \frac{h_0}{2\tau} = 39 \times 10^{-24} (\text{rad})^2 \\ [\sigma_y(20 \text{ ms})]_0 &= 6.24 \times 10^{-12} (\text{rad}) \end{aligned}$$

Similarly,

$$\begin{aligned} [\sigma_y(20 \text{ ms})]_{+1} &= 4.2 \times 10^{-12}(\text{rad}) \\ [\sigma_y(20 \text{ ms})]_{+2} &= 0.85 \times 10^{-12}(\text{rad}) \\ [\sigma_y(20 \text{ ms})]_{-1} &= 4.7 \times 10^{-12}(\text{rad}) \end{aligned}$$

By Expression (2.48), one can obtain

$$[\sigma_y(20 \text{ ms})]_{\Sigma} = 8.8 \times 10^{-12}(\text{rad})$$

The measured value is 6.625×10^{-12} , rather consistent with the calculation result. The larger result is caused by bigger $S_{\Phi}(f)$ approximate broken line.

Similarly, the short-term stability can be calculated according to phase noise indexes given in the specifications. For example, for the following phase noise spectrum density $S_{\Phi}(f)$ indexes:

$$\begin{aligned} -27 - 30\lg f(\text{dBc/Hz}), & \quad 10 \text{ Hz} \leq f \leq 100 \text{ Hz} \\ -32 - 20\lg f(\text{dBc/Hz}), & \quad 100 \text{ Hz} \leq f \leq 1 \text{ kHz} \\ -52 - 10\lg f(\text{dBc/Hz}), & \quad 1 \text{ kHz} \leq f \leq 100 \text{ kHz} \end{aligned}$$

Solving steps of short-term stability are:

1) If $f_0 = 2.1 \text{ GHz}$, by Expressions (2.37), (2.38), (2.39), (2.40), and (2.41), one can obtain

$$\begin{aligned} K_{-1} &= -27 \text{ dBc/Hz} \\ K_0 &= -32 \text{ dBc/Hz} \\ K_{+1} &= -52 \text{ dBc/Hz}, \end{aligned}$$

Then, solve the short-term stability of noise through Expressions (2.43), (2.44), (2.45), (2.46), (2.47), and (2.48).

2) When the unilateral loop of phase-locked receiver is 1 kHz, there are only h_{-1} and h_0 noise.

$$\begin{aligned} [\sigma_y(20 \text{ ms})]_{-1}^2 &= 2h_{-1}\ln 2 = 6.3 \times 10^{-22}, \quad [\sigma_y(20 \text{ ms})]_{-1} = 2.52 \times 10^{-11} \\ [\sigma_y(20 \text{ ms})]_0^2 &= \frac{h_0}{2\tau} = 3.6 \times 10^{-22}, \quad [\sigma_y(20 \text{ ms})]_0 = 1.9 \times 10^{-11} \\ [\sigma_y(20 \text{ ms})]_{\Sigma} &= 3.1 \times 10^{-11} \end{aligned}$$

3) When cutoff frequency is 6.3 kHz, h_{+1} noise exists, $f_5 = 6.3 \text{ kHz}$, $f_4 = 1 \text{ kHz}$.

$$\begin{aligned}
 [\sigma_y(20 \text{ ms})]_{+1}^2 &= \frac{1.038 + 3\ln[(2\pi\tau(f_5 - f_4))]}{4\pi^2\tau^2} h_{+1} = 18.8 \times 10^{-22} \\
 [\sigma_y(20 \text{ ms})]_{+1} &= 4.3 \times 10^{-11} \\
 [\sigma_y(20 \text{ ms})]_{\Sigma} &= 5.37 \times 10^{-11}
 \end{aligned}$$

(3) Feature point correspondence method

Analytic function representation method can be used for design calculation and index allocation but is still complex for adjustment, testing, and fault isolation, and its physical conception is less intuitive. Therefore, the “feature point” correspondence method is proposed. It refers to measured data correspondence method. “Feature point” refers to the point imposing the largest impact.

It is obvious from Expression (2.29) that $\sigma_y^2(\tau)$ is an output of $S_\phi(\omega)$ through a transfer function of $H_\phi(f) = \sin^4(\pi\tau f)$. The filtering characteristic curve of $\sin^4(\pi\tau f)$ is shown in Fig. 2.22.

In Fig. 2.22, $S_\phi(f)$ is phase fluctuation spectrum density, $|H_\phi(f)|^2$ is the equivalent transfer function of Allan variance when inputting $S_\phi(f)$, $S_y(f)$ is phase-frequency fluctuation spectrum density, and $|H_y(f)|^2$ is the equivalent transfer function of Allan variance when inputting $S_y(f)$.

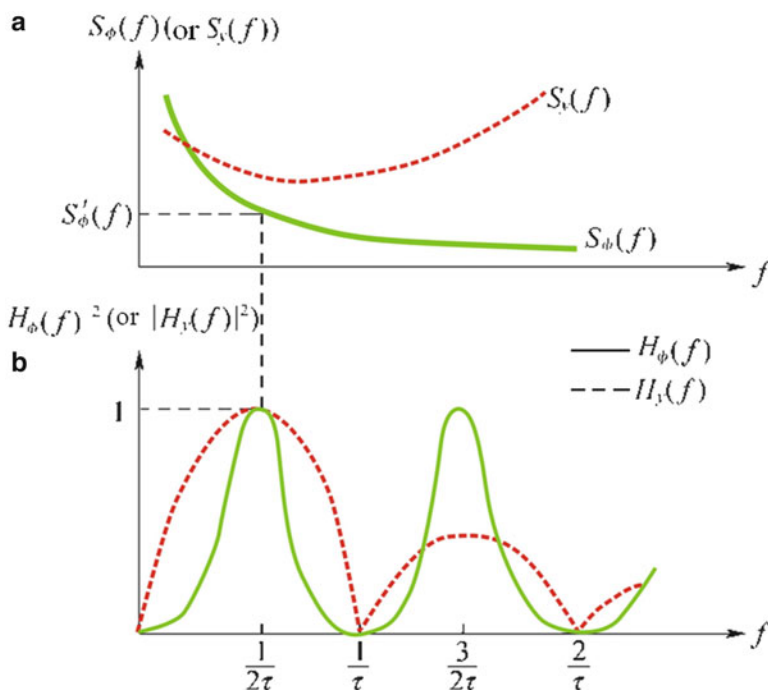


Fig. 2.22 Filtering characteristic of Allan variance $\sigma_y(\tau)$

It can be seen from Fig. 2.22 that filtering characteristic has a peak value at $f = 1/2\tau$. $S_\phi(\omega)$ sharply decreases with f increase, so phase noise power spectrum density at $f = 1/2\tau$ will impose greatest impact on the output (i.e., the frequency window at $1/2\tau$ corresponds to the time window at τ). According to this feature point, one can roughly solve the relationship between output Allan variance and phase noise at $f = 1/2\tau$ frequency point, that is to say, measured $S_\phi(1/2\tau)$ can obtain an approximate $\sigma_y(\tau)$ value. For example, if measured data shown in Fig. 2.21 are given, measured $\sigma_y(20 \text{ ms})$ is 6.63×10^{-12} and $\mathcal{L}(f)$ corresponding to $1/2\tau = 25 \text{ Hz}$ is -130 dBc/Hz (relevant $S_\phi(f)$ increases 3 dB, namely, -127 dBc/Hz), only the phase noise is to be measured in future test to obtain approximate Allan variance. That is, only phase noise at $f = 25 \text{ Hz}$ is lower than -130 dBc/Hz and $\sigma_y(20 \text{ ms})$ will be better than 6.63×10^{-12} . This result is approximate but is very convenient in application during development, and its physical concept is also clear and definite. If you want to represent it more accurately, additional feature points such as $3\tau/2$ and $5\tau/2$ can be used.

2.3.3.2 Theoretical Analysis of Two-Way Coherent Doppler Velocity-Measuring Error

Time Domain Analysis Method for Velocity Measurement Accuracy

(1) Fundamental principle of time domain analysis method

Before fundamental principle description, it is firstly assumed that the additional frequency fluctuation is not induced by either the uplink and downlink channels or the transponder for a two-way coherent Doppler velocity measurement and the receiving/transmitting frequency fluctuations are only induced by master oscillator source, with a spatial time delay of T between the two frequency fluctuations. The velocity measurement model is shown in Fig. 2.23.

It can be seen from Fig. 2.23 that Doppler frequency is equal to the difference between output frequencies of a transmitter in different times. The integration time is τ which is approximately equal to sampling time. The average smooth output value is

$$\overline{f_d(t)} = \overline{f_1(t+T) - f_1(t)}$$

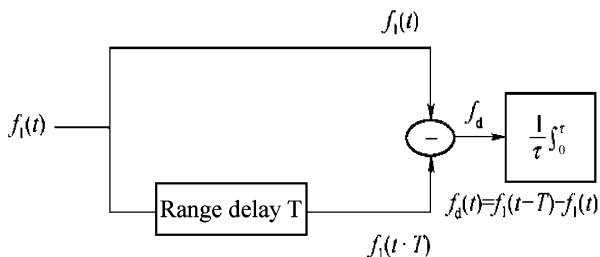


Fig. 2.23 Two-way coherent Doppler velocity measurement model

According to the theorem of mathematical expectation, the average of differences is equal to difference between average values,

$$\overline{f_d(t)} = \overline{f_1(t+T)} - \overline{f_1(t)} \quad (2.49)$$

where the first item and second item in the right side of the equation are two measured values of transmitting frequencies with T time difference. Corresponding time relationship is shown in Fig. 2.24.

In Fig. 2.24, if measuring at t_1 , delay time of R/T signal is T . It is obvious that measurement of R/T signal as shown in Fig. 2.24 is equivalent to measurement of transmitting signal as shown in Fig. 2.25.

It can be seen in Fig. 2.25 measurement of frequency difference Δf between R/T signals with delay time T is equivalent to measurement of frequency difference ($f_2 - f_1$) of the transmitting signal between frequency f_2 at $(t_1 - T)$ and frequency f_1 at t_1 . For this measurement, two times in a group and multiple group method are typically employed. The frequency fluctuation variance is calculated and then averaged to obtain the Doppler frequency output variance of $\langle (\overline{f_2} - \overline{f_1})^2_m \rangle$,

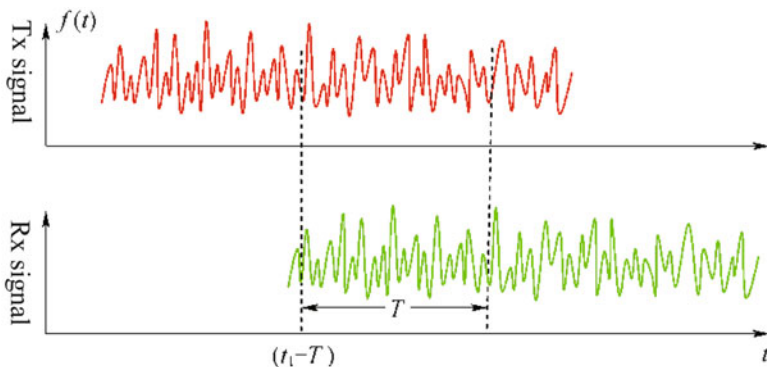


Fig. 2.24 Phase and frequency relationships between R/T signals

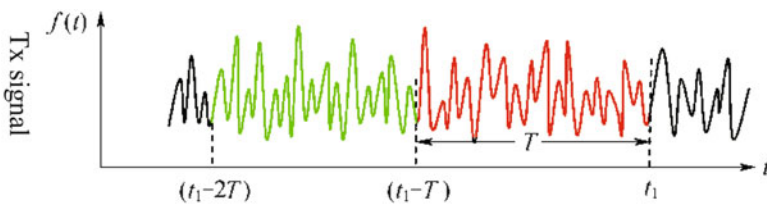


Fig. 2.25 Equivalent measurement of transmitting signal

where, $\langle \dots \rangle$ represents an average value of m groups of measurements. According to the definition of gapped Allan variance $\langle \sigma_y^2(2, T, \tau) \rangle$:

$$2\langle \sigma_y^2(2, T, \tau) \rangle = \langle (\bar{y}_2 - \bar{y}_1) \rangle = \frac{\langle (\bar{f}_2 - \bar{f}_1)_m^2 \rangle}{f_0^2} \quad (2.50)$$

where $y = \Delta f/f_0$ is the relative frequency instability; y_2 and y_1 are the value of two measurements, respectively; T is measurement time interval; and τ is integration time. Expression (2.50) is the expression of gapped Allan variance of short-term stability. So the measurement of Doppler frequency variance is equivalent to the measurement of short-term stability of transmitter signals, i.e., if the $\langle \sigma_y^2(2, t_0, T) \rangle$ is measured, then the velocity can be measured.

Short-term instability of signal frequency is typically expressed by Allan variance $\langle \sigma_y^2(2, \tau, \tau) \rangle = \sigma_y^2(\tau)$. If the relation between $\langle \sigma_y^2(2, t_0, T) \rangle$ and Allan variance is solved, velocity-measuring error can be calculated by Allan variance. Reference [6] gives this relation, namely, Banes second bias function

$$B_2(\gamma, \mu) = \frac{\langle \sigma_y^2(2, T, \tau) \rangle}{\sigma_y^2(\tau)} \quad (2.51)$$

where $\gamma = T/\tau$ and μ is the noise type index. So the velocity-measuring error caused by short-term stability is:

$$\begin{aligned} \sigma_R^2 &= \left(\frac{c}{2}\right)^2 \frac{\langle (\bar{f}_2 - \bar{f}_1)_m^2 \rangle}{f_0^2} = \left(\frac{c}{2}\right)^2 \times 2\langle \sigma_y^2(2, t_0, T) \rangle \\ &= \frac{1}{2} C^2 B_2(\gamma, \mu) \sigma_y^2(\tau) \\ \sigma_R &= \frac{\sqrt{B_2(\gamma, \mu)}}{\sqrt{2}} C \sigma_y(\tau) \end{aligned} \quad (2.52)$$

References [6] and [7] have given B_2 of various types of noise:

1) For $\mu = 0$, i.e., FM flicker noise with $\alpha = -1$,

$$B_2 = \begin{cases} 0 & \text{when } \gamma = 0 \\ \frac{-2\gamma^2 \ln \gamma + (\gamma + 1)^2 \ln(\gamma + 1) + (\gamma - 1)^2 \ln(\gamma - 1)}{4 \ln 2} & \text{when } \gamma > 0 \end{cases} \quad (2.53)$$

2) For $\mu = -1$, i.e., FM white noise with $\alpha = 0$,

$$B_2 = \begin{cases} \gamma, & \text{when } 0 \leq \gamma \leq 1 \\ 1, & \text{when } \gamma > 1 \end{cases} \quad (2.54)$$

Table 2.5 $B_2(\gamma, \mu)$ function table

$\gamma \backslash \mu$	0	0.1	0.8	1	1.1	2	4	16	64	256	1024
0	0	0.0274	0.7667	1.000	1.089	1.566	2.078	3.082	4.082	5.082	6.082
-1	0	0.1	0.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-2	0	0.6667	0.6667	1.000	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667
+1	0			1.000	1.15	2.5	5.5	23.5			

3) For $\mu = -2$, i.e., PM white noise with $\alpha = +2$ and PM flicker noise with $\alpha = +1$,

$$B_2 = \begin{cases} 0, & \gamma = 0 \\ 1, & \gamma = 1 \\ 2/3, & \gamma \text{ others} \end{cases} \quad (2.55)$$

$B_2(\gamma, \mu)$ function table is listed in Table 2.5.

In order to use Table 2.5 for lookup $B_2(\gamma, \mu)$, one should firstly find out γ and μ , where γ can be solved through $\gamma = T/\tau$ for a different target range time delay T and sampling time τ . It can be seen in the above table that, for $\mu = 0$ noise type, $B_2(\gamma_1, 0)$ will gradually increase as T increases. This results in increasing velocity-measuring error $\sigma \dot{R}$. The physical reason is that the noise is an FM flicker noise component with $\mu = 0$ which is a slowly changed random fluctuation and also a non-stable random fluctuation. Within an infinitely long time, there must be infinitely great fluctuations, that is, the frequency source exits slow drift. The gapped Allan variance could not neutralize this slow drift as T increases, so the $\sigma_y^2(2, T, \tau)$ increases. It is worth noticing that Allan variance will not increase because the $T = \tau$ is not high. Based on the velocity-measuring principles, when reference frequency is drifted, it will be introduced to the Doppler value from measurement. As mentioned above, in order to minimize velocity-measuring error, the first is to reduce T/τ value (e.g., increase τ for constant T) and the second is to improve the stability of the frequency source (e.g., the Allan variance with given average sampling time τ). For that reason, a high-stability frequency source with higher τ and lower Allan variance should be used when T is high (e.g., deep-space TT&C). For the noise of $\mu = -1$ and $\mu = -2$, both are stable random fluctuations and their $B_2(\gamma, \mu)$ does not relate to T .

The following is an example to explain velocity-measuring error caused by phase noise (or short-term stability) of reference frequency source. Taking the phase noise of a 10 MHz reference source in Fig. 2.21 as an example, at target range 2,500 km and sampling time 20 ms the velocity-measuring error can be calculated by the following steps:

1) Solve the power law spectrum coefficient. According to Fig. 2.21, there are four types of phase noise: h_{-1} , h_0 , h_{+1} , and h_{+2} , which have been solved in Sect. 2.3.3.1,

$$\begin{aligned}
h_{-1} &= 1.58 \times 10^{-23} (\text{rad})^2 \\
h_0 &= 1.58 \times 10^{-24} (\text{rad})^2 / \text{Hz} \\
h_{+1} &= 1.58 \times 10^{-26} (\text{rad})^2 / \text{Hz}^2 \\
h_{+2} &= 1.58 \times 10^{-30} (\text{rad})^2 / \text{Hz}^3
\end{aligned}$$

2) Solve the short-term stability of each type of phase noise at sampling time 20 ms, which has been solved in Sect. 2.3.3.1, namely, $[\sigma_y(20 \text{ ms})]_\alpha$ (here, noise type is represented by α).

3) Solve Banes second bias function $B_2(\gamma, \mu)$ of each type of phase noise. Given $T = 16.66 \text{ ms}$ for the range of 2,500 km, $\tau = 20 \text{ ms}$, $\gamma = T/\tau = 0.8$. It can be solved by Expressions (2.53), (2.54), and (2.55) or searched in Table 2.5 (here, noise type is represented by μ . The correspondence between μ and α is listed in Table 2.4).

$$\begin{aligned}
B_2(0.8, 0) &= 0.7667 \text{ corresponds to } h_{-1} \text{ noise (see Table 2.4)} \\
B_2(0.8, -1) &= 0.8 \text{ corresponds to } h_0 \text{ noise (see Table 2.4)} \\
B_2(0.8, -2) &= 0.667 \text{ corresponds to } h_{+1} \text{ noise (see Table 2.4)} \\
B_2(0.8, -2) &= 0.667 \text{ corresponds to } h_{+2} \text{ noise (see Table 2.4)}
\end{aligned}$$

4) Solve velocity-measuring error caused by various types of phase noise. By Expression (2.52), one can obtain

$$\begin{aligned}
[\sigma_{\dot{R}}]_{-1}^2 &= 75.9 \times 10^{-8} (\text{m/s})^2 \\
[\sigma_{\dot{R}}]_0^2 &= 140.5 \times 10^{-8} (\text{m/s})^2 \\
[\sigma_{\dot{R}}]_{+1}^2 &= 56.16 \times 10^{-8} (\text{m/s})^2 \\
[\sigma_{\dot{R}}]_{+2}^2 &= 2.1 \times 10^{-8} (\text{m/s})^2
\end{aligned}$$

Total velocity-measuring error is

$$\begin{aligned}
\sigma_{\dot{R}}^2 &= [\sigma_{\dot{R}}]_{-1}^2 + [\sigma_{\dot{R}}]_0^2 + [\sigma_{\dot{R}}]_{+1}^2 + [\sigma_{\dot{R}}]_{+2}^2 = 274.66 \times 10^{-8} (\text{m/s})^2 \\
\sigma_{\dot{R}} &= 16.57 \times 10^{-4} \text{ m/s} = 1.6 \text{ mm/s}
\end{aligned}$$

If the vehicle continues to fly to the Mars, the range is $401.3 \times 10^6 \text{ km}$, corresponding to $T = 2,675.3 \text{ s}$, $\gamma = 133,765$. In this case, $B_2(133,765, 0)$ would be very big, thus causing larger velocity-measuring error. This is because the correlation of R/T signal frequency fluctuation at receiving time is reduced to decrease the counteract effect of frequency fluctuation noise. In deep-space TT&C, in order to reduce the velocity-measuring error, integration time τ should be increased ($\tau = 60 \text{ s}$ for some stations) to reduce γ and h_{-1} (improve short-term stability).

If the total short-term stability of each noise type is provided instead of short-term stabilities of individual noise type, the following method may be used for approximate calculation. It can be seen in Fig. 2.22, $|H_y(f)|^2$ at $1/2\tau$ frequency is maximum value. Typically, frequency fluctuation $\sigma_y(\tau)$ output at $1/2\tau$ is the main component of total fluctuation, i.e., $S_y(f)$ noise type corresponding to $1/2\tau$ is main noise type of output $\sigma_y(\tau)$. Hence, approximate calculation may be performed based on this noise type and total short-term stability. With the phase noise characteristics of a crystal oscillator shown in Fig. 2.21, it is FM white noise (corresponding $\tau = 1/2f = 50 - 5$ ms) for $f = 10-100$ Hz, PM white noise ($\tau < 0.125$ ms) for $f > 4$ kHz, and FM flicker noise ($\tau > 50$ ms) for $f < 10$ Hz.

Accordingly, velocity-measuring error at different integration time τ can be calculated approximately:

① For $\tau = 50-5$ ms, FM white noise mainly occurs. In this case, the corresponding $\mu = -1$ and $B_2 = 1$ with $\gamma \geq 1$ as lookup for Table 2.5. So the following expression can be obtained according to Expression (2.52):

$$\sigma_R^* = \frac{C}{\sqrt{2}} \sigma_y(\tau) \quad (2.56)$$

② For $\tau = 0.5-10$ s, FM flicker noise mainly occurs. In this case, corresponding $\mu = 0$ and B_2 will vary with T/τ as shown in Table 2.5 and the specific value can be obtained by lookup table.

③ For $\tau \leq 0.5$ ms, the errors are PM white noise and PM flicker noise components. In this case, corresponding $\mu = -2$ and as shown in Table 2.5:

If $\gamma = 1$, then $B_2 = 1$, and

$$\sigma_R^* = \frac{C}{\sqrt{2}} \sigma_y(\tau) \quad (2.57)$$

When If $\gamma \neq 1$, then $B_2 = 2/3$, and

$$\sigma_R^* = \frac{C}{\sqrt{3}} \sigma_y(\tau) \quad (2.58)$$

When $\gamma = 0$, $B_2 = 0$, and

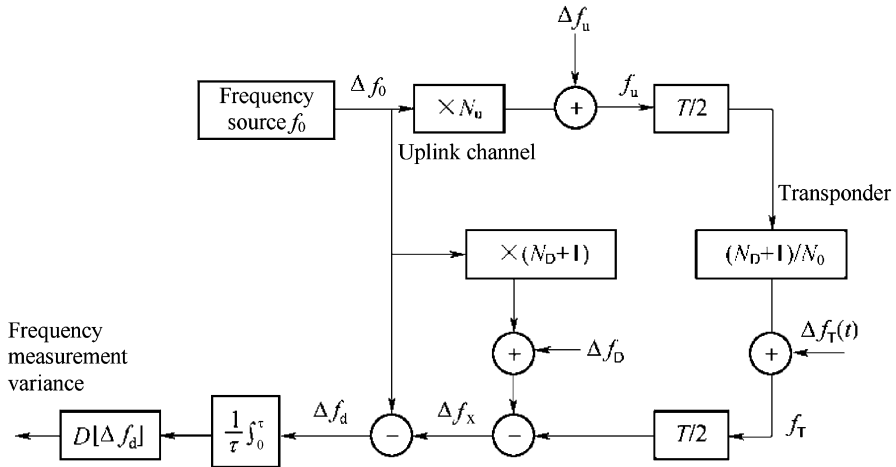
$$\sigma_R^* = 0 \quad (2.59)$$

The above calculation is performed for phase noise characteristics of the crystal oscillator shown in Fig. 2.21. This calculation also can be performed using the short-term stability specifications, such as short-term stability specifications of a hydrogen atomic clock as listed in Table 2.6.

In the range of $f = 0.5-50$ Hz, $\sigma_y(\tau)$ is in proportion to $1/\tau$. Thus according to Expressions (2.43), (2.44), (2.45), (2.46), and (2.47), the phase noise is h_0 noise (FM white noise). If $f > 50$ Hz, $\sigma_y(\tau)$ is almost independent of τ and the phase noise is h_{-1} noise (FM flicker noise). After the noise type is determined, corresponding

Table 2.6 Short-term stability specifications of hydrogen atomic clock

τ/s	1	10	100	1,000	1,000 above
$\sigma_y(\tau)$	1.5×10^{-13}	2×10^{-14}	2×10^{-15}	2×10^{-15}	2×10^{-15}
$f = \frac{1}{2\tau}/Hz$	0.5	5	50	500	500 above

**Fig. 2.26** Simplified frequency flow of a two-way coherent Doppler velocity measuring

B_2 can be looked up. Then, the velocity-measuring error can be calculated according to an appropriate expression.

According to the comparison of the above two types of oscillator phase noise, spectrum width of phase noise for atomic clock significantly reduces, and spectrum components of every phase noise move to low frequencies.

(2) Effect of additional phase noise [8]

The above analysis assumes that R/T frequency fluctuations are caused by the frequency reference source. However, the actual TT&C system includes multiple oscillators, such as frequency reference source (5 or 10 MHz source), local oscillator of transmitter (may be multiple), receiver local oscillator (may be multiple), and local oscillator of transponder. The above sections only give an analysis to effects of short-term stability of frequency reference source on Doppler velocity measurement accuracy but not to the effect of the thermal noise from other oscillators and the receiver. In practice, short-term stability of various coherent oscillators is different from that of the frequency reference source, which includes additional noncoherent phase noise. For example, many phase-locked frequency synthesizers have higher additional phase noise which result in its short-term stability lower than that of the frequency reference source. Moreover, these additional phase noises could not be offset in coherent velocity measurement. Thus, such additional phase noise would increase the velocity-measuring error. Specific analysis follows.

Simplified frequency flow of a two-way coherent Doppler velocity measuring is as shown in Fig. 2.26.

In Fig. 2.26, Δf_0 is frequency fluctuation noise caused by short-term instability of the frequency source (short-term stability is $\Delta f_0/f_0$); Δf_u , Δf_T , and Δf_D are frequency fluctuation noises added by the uplink channel, transponder, and downlink channel, respectively; τ is integration time; T is R/T delay time; f_u is uplink RF frequency; f_T is downlink RF frequency; and Δf_d is received two-way Doppler frequency shift. Due to frequency fluctuation noise, Δf_d is a random variable and $D[\Delta f_d]$ is its variance.

Velocity measurement is

$$\dot{R} = C \left(\frac{\Delta f_d}{f_u + f_T} \right) \quad (2.60)$$

Total frequency fluctuation noise at the output end of the receiver is

$$\begin{aligned} \Delta f_\Sigma(t) = & (N_D + 1)\Delta f_0(t - T) + \left(\frac{N_D + 1}{N_u} \right) \Delta f_u(t - T) \\ & + \Delta f_T \left(t - \frac{T}{2} \right) - (N_D)\Delta f_0(t) - \Delta f_D(t) \end{aligned} \quad (2.61)$$

Frequency fluctuation noise at the output end of a Doppler frequency extractor is

$$\begin{aligned} \Delta f_d(t) = & \Delta f_\Sigma(t) - \Delta f_0(t) \\ = & (N_D + 1)[\Delta f_0(t - T) - \Delta f_0(t)] + \left(\frac{N_D + 1}{N_u} \right) \Delta f_u(t - T) \\ & + \Delta f_T \left(t - \frac{T}{2} \right) - \Delta f_D(t) \end{aligned} \quad (2.62)$$

Since random variables $\Delta f_0(t)$, $\Delta f_u(t)$, and $\Delta f_T(t)$ are independent from each another, according to digital characteristics theorem of random variables,

$$\begin{aligned} D[\Delta f_d(t)] = & (N_D + 1)D \left[\Delta f_0 \left(t - \frac{T}{2} \right) - \Delta f_0(t) \right] + \left(\frac{N_D + 1}{N_u} \right) D[\Delta f_u(t - T)] \\ & + D \left[\Delta f_T \left(t - \frac{T}{2} \right) \right] - D[\Delta f_D(t)] \end{aligned} \quad (2.63)$$

In order to simplify the expression, the substitution is made:

$$\begin{aligned}
D[\Delta f_d(t)] &= D[\Delta f_d], & D[\Delta f_u(t-T)] &= D[\Delta f_u], \\
D\left[\Delta f_T\left(t - \frac{T}{2}\right)\right] &= D[\Delta f_T], & D[\Delta f_D(t)] &= D[\Delta f_D],
\end{aligned} \tag{2.64}$$

Since Δf_u , Δf_T , and Δf_D are independent of one another and are all stationary random processes, such substitution will not lose its general sense. Expression (2.63) can be expressed as

$$D[\Delta f_d] = (N_D + 1)D[\Delta f_0(t)] + \left(\frac{N_D + 1}{N_u}\right)D[\Delta f_u] + D[\Delta f_T] - D[\Delta f_D] \tag{2.65}$$

According to Expression (2.60), velocity-measuring error caused by frequency fluctuation noise is:

$$\begin{aligned}
\sigma_R^* &= C \left(\frac{\sqrt{D[\Delta f_d]}}{f_u + f_T} \right) \\
&= \frac{C}{f_u + f_T} \sqrt{(N_D + 1)D[\Delta f_0(t-T) - \Delta f_0(t)] + \left(\frac{N_D + 1}{N_u}\right)D[\Delta f_u] + D[\Delta f_T] - D[\Delta f_D]}
\end{aligned} \tag{2.66}$$

When $\Delta f_u = 0$, $\Delta f_T = 0$, and $\Delta f_D = 0$ (consider only the effect of frequency fluctuation noise of frequency source f_0),

$$D[\Delta f_d] = (N_D + 1)D[\Delta f_0(t-T) - \Delta f_0(t)] \tag{2.67}$$

This is the short-term stability of the frequency source effect on the velocity measurement accuracy discussed in other references.

When $(N_D + 1)D[\Delta f_0(t-T) - \Delta f_0(t)] \ll ((N_D + 1)/N_u)D[\Delta f_u] + D[\Delta f_T] - D[\Delta f_D]$, i.e., when additional frequency fluctuation noise is higher than frequency fluctuation noise of the frequency source f_0 ,

$$D[\Delta f_d] = \frac{(N_D + 1)}{N_u}D[\Delta f_u] + D[\Delta f_T] - D[\Delta f_D] \tag{2.68}$$

$\Delta f_0/f_0$ could not be improved by increasing short-term stability of the frequency source. These two conditions would occur in practical engineering and will be discussed as follows.

1) Effect of short-term stability of frequency source. When we only consider the effect of Δf_0 but don't consider the effect of Δf_u , Δf_T , and Δf_D , Expression (2.62) can be simplified as

$$\Delta f_{d0}(t) = (N_D + 1)[\Delta f_0(t-T)] - (N_D + 1)[\Delta f_0(t)] \tag{2.69}$$

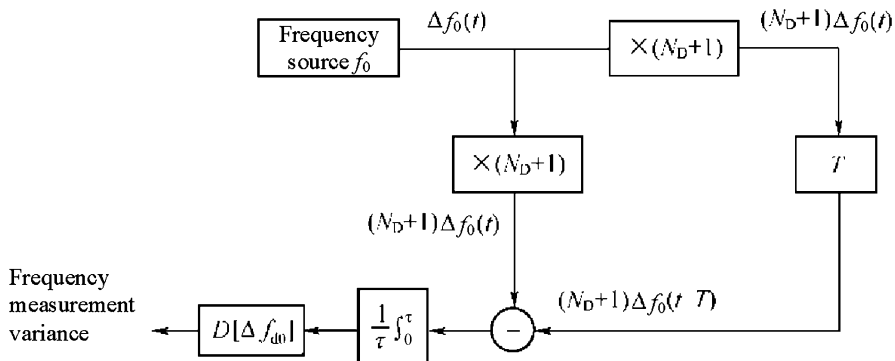


Fig. 2.27 Model of a two-way Doppler velocity-measuring error caused by short-term stability of frequency source

where $\Delta f_{d0}(t)$ is the output frequency fluctuation noise when only short-term stability of the frequency source f_0 is considered.

Expression (2.69) can be expressed as the form of Fig. 2.27.

In Fig. 2.27, $D[\Delta f_{d0}]$ refers to frequency measurement variance caused by short-term stability of frequency source f_0 . Corresponding time domain process is shown in Fig. 2.28.

Figure 2.28a shows frequency fluctuation noise of transmitting RF signal. Figure 2.28b shows frequency fluctuation noise of receiving RF signal, and T is time delay of R/T signals. Figure 2.28c shows frequency difference $\Delta f_{d0}(t)$ obtained from subtraction between transmitting RF and receiving RF. Transmitting/receiving frequency noise has certain correlation effects, so $\Delta f_{d0}(t)$ after subtraction decreases. With T increasing, related effects become weak to make $\Delta f_{d0}(t)$ gradually increasing. When T increases to the level in which the T is not correlated with R/T Δf , $\Delta f_{d0}(t)$ will be the mean square sum of R/T $\Delta f(t)$ and is about twice of transmitting signal $\Delta f(t)$ variance. $\Delta f_{d0}(t)$ is the average of sampled values within τ time. The frequency measurement variance can be obtained by continuously sampling rN times and then calculating the variance. Figure 2.28d shows sampling time τ . It can be seen in the figure that averaging $\Delta f_{d0}(t)$ within τ time is equivalent to subtract mean value $(N_D+1)[\Delta f_0(t)]$ averaging $(N_D+1)[\Delta f_0(t)]$ within τ'' from mean value $(N_D+1)[\Delta f_0(t-T)]$ averaging $(N_D+1)[\Delta f_0(t)]$ within τ' with sampling time $\tau' = \tau'' = \tau$ and sampling interval of T . This is the gapped sampling variance of transmitting RF signal, which can be used to calculate velocity-measuring error.

Figures 2.27 and 2.28 show the physical processes of velocity measuring, which can be represented by mathematical expressions as follows.

Doppler frequency difference is:

$$\Delta f_{d0}(t) = [\Delta f_0(t-T) - \Delta f_0(t)](N_D + 1) \quad (2.70)$$

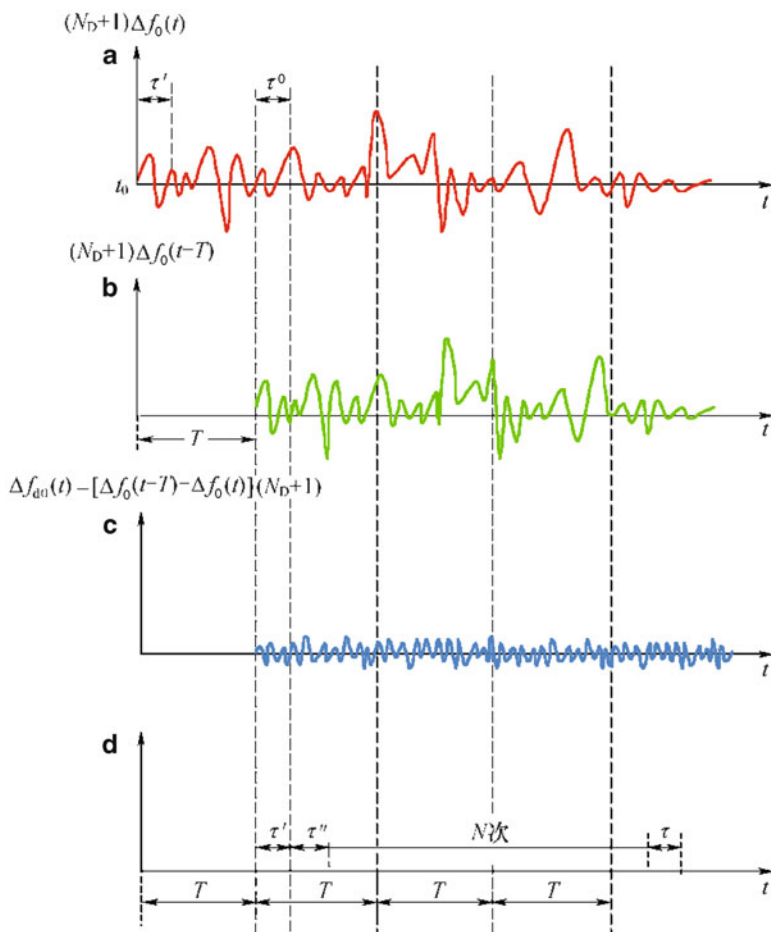


Fig. 2.28 Time domain relationship of Doppler velocity measuring

The integration over the integral time τ can be achieved through the averaging within time τ , i.e.,

$$\begin{aligned} \Delta \bar{f}_{d0}(t) &= [\Delta \bar{f}_0(t-T) - \Delta \bar{f}_0(t)](N_D + 1) \\ D[\Delta f_{d0}] &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \left[(\Delta f_{d0})_n - \frac{1}{N} \sum_{n=1}^N (\Delta f_{d0})_n \right]^2 \end{aligned} \quad (2.71)$$

According to the mathematical expectation theorem of random variable,

$$\begin{aligned} \Delta \bar{f}_{d0}(t) &= [\Delta \bar{f}_0(t-T) - \Delta \bar{f}_0(t)](N_D + 1) \\ D[\Delta f_{d0}] &= \left[\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N (\Delta f_{d0})_n \right]^2 \end{aligned} \quad (2.72)$$

where $(\Delta f_{d0})_n$ is n th sampling value of Δf_{d0} (similarly, n th sampling value of any random variable X is expressed as $(X)_n$).

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N (\Delta f_{d0})_n &= \frac{(N_D + 1)^2}{N} \sum_{n=1}^N [\overline{\Delta f_0(t - T)} - \Delta f_0(t)]_n \\ &= (N_D + 1)^2 \left\{ \frac{1}{N} \sum_{n=1}^N [\Delta \bar{f}_0(t - T)]_n - \frac{1}{N} \sum_{n=1}^N [\Delta \bar{f}_0]_n \right\} \end{aligned} \quad (2.73)$$

where both two items are mean values of frequency drift of frequency source f_0 . Since the stability of frequency source f_0 is very high, both items are approximate to 0. For example, when $\tau = 20$ ms and $N = 1,000$ (mean value within 20 s),

$$\frac{1}{N} \sum_{n=1}^N (\Delta f_{d0})_n \approx 0.$$

$$\begin{aligned} D[\Delta f_{d0}] &= \lim_{N \rightarrow \infty} \frac{(N_D + 1)^2}{N} \sum_{n=1}^N [(\Delta f_{d0})_n]^2 \\ &= \lim_{N \rightarrow \infty} \frac{(N_D + 1)^2}{N} \sum_{n=1}^N [\Delta \bar{f}_0(t - T) - \Delta \bar{f}_0(t)]_n^2 \end{aligned} \quad (2.74)$$

According to the definition of gapped Allan variance,

$$[\Delta \bar{f}_0(t - T) - \Delta \bar{f}_0(t)]_n^2 = 2f_0^2 \sigma_y(2, T, \tau) \quad (2.75)$$

where $\sigma_y(2, T, \tau)$ is gapped sampling Allan variance of relative value $\Delta f_0(t)/f_0$ of frequency fluctuation noise $\Delta f_0(t)$ of frequency source f_0 , where sampling interval is T , number of sampling is 2, and average time is τ .

In the above expression, T is the target time delay in actual velocity measuring and τ is integration time of velocity-measuring data processing. So

$$\begin{aligned} D[\Delta f_{d0}] &= 2 \lim_{N \rightarrow \infty} \frac{(N_D + 1)^2}{N} f_0^2 \sum_{n=1}^N [\sigma_y(2, T, \tau)]_n \\ &= 2(N_D + 1)^2 f_0^2 \langle \sigma_y(2, T, \tau) \rangle \end{aligned} \quad (2.76)$$

where $\langle \dots \rangle$ represents mean value.

Then, substitute B_2 into Expression (2.51) to obtain

$$D[\Delta f_{d0}] = 2(N_D + 1)^2 f_0^2 B_2 \sigma_y^2(\tau), \quad D[\Delta f_0] = 2f_0^2 B_2 \sigma_y^2(\tau) \quad (2.77)$$

The velocity-measuring error can be obtained by Expression (2.66):

$$\sigma_{\dot{R}} = \frac{C}{2} \frac{\sqrt{D[\Delta f_{d0}]}}{(N_D + 1)f_0} = \frac{C}{2} \frac{\sqrt{2B_2(N_D + 1)^2 f_0^2}}{(N_D + 1)f_0} \sqrt{D[\Delta f_{d0}]} = \frac{\sqrt{B_2}}{\sqrt{2}} C \sigma_y(\tau) \quad (2.78)$$

According to Expression (2.78), velocity-measuring error $\sigma_{\dot{R}}$ can be solved by B_2 obtained from Allan variance and lookup table. The result is the same as the result of Expression (2.52).

2) Effect of additional noise. When additional frequency fluctuation noise is a primary contributing factor, velocity-measuring error is expressed as follows according to Expressions (2.66) and (2.68):

$$\sigma_{\dot{R}} = \frac{C}{f_u + f_T} \sqrt{\frac{(N_D + 1)}{N_u} D[\Delta f_u] + D[\Delta f_T] - D[\Delta f_D]} \quad (2.79)$$

According to Expressions (2.63), (2.64), (2.65), (2.66), (2.67), (2.68), (2.69), (2.70), (2.71), (2.72), (2.73), (2.74), (2.75), (2.76), and (2.77) and Reference [8]:

$$\begin{aligned} D[\Delta f_u] &= 2B_2 f_u^2 \sigma_{yu}^2(\tau), \quad D[\Delta f_T] = 2B_2 f_T^2 \sigma_{yT}^2(\tau), \quad D[\Delta f_D] \\ &= 2B_2 f_D^2 \sigma_{yD}^2(\tau) \end{aligned} \quad (2.80)$$

where $\sigma_{yu}^2(\tau)$, $\sigma_{yT}^2(\tau)$, and $\sigma_{yD}^2(\tau)$ are short-term stabilities of each unit.

Substituting it into Expression (2.79) then obtains:

$$\sigma_{\dot{R}} = C \sqrt{2B_2 \left[\frac{(N_D + 1)}{N_u} \left(\frac{f_u}{f_u + f_T} \right)^2 \sigma_{yu}^2(\tau) + \left(\frac{f_T}{f_u + f_T} \right)^2 \sigma_{yT}^2(\tau) + \left(\frac{f_D}{f_u + f_T} \right)^2 \sigma_{yD}^2(\tau) \right]} \quad (2.81)$$

According to Expression (2.81), the lower the f_u, f_T, f_D , the smaller the effects on $\sigma_{\dot{R}}$ and thus lower short-term stability requirements. In practical engineering, “weighted allotment” or “equal contribution” allotment can be used to determine the short-term stability requirements of each unit. If “equal contribution” allotment is used,

$$\frac{(N_D + 1)}{N_u} \left(\frac{f_u}{f_u + f_T} \right)^2 \sigma_{yu}^2 = \left(\frac{f_T}{f_u + f_T} \right)^2 \sigma_{yT}^2(\tau) = \left(\frac{f_D}{f_u + f_T} \right)^2 \sigma_{yD}^2(\tau) \quad (2.82)$$

According to Expression (2.82), we can obtain

$$\frac{\sigma_{yT}(\tau)}{\sigma_{yD}(\tau)} = \frac{f_D}{f_T} \quad (2.83)$$

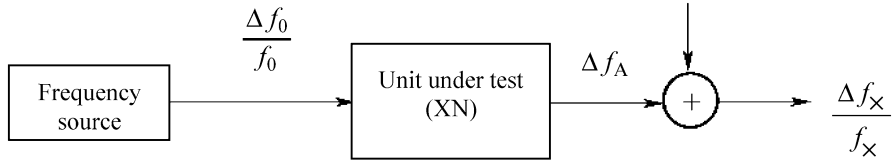


Fig. 2.29 Short-term stability contribution of additional phase noise

$$\frac{\sigma_{yT}(\tau)}{\sigma_{yD}(\tau)} = \sqrt{\frac{(N_D + 1)f_u}{N_u f_T}} \quad (2.84)$$

Expressions (2.83) and (2.84) can be used for performance allotment to each unit under “equal contribution criteria.” The contribution of short-term stability of additional phase noise of each unit to the total short-term stability is shown in Fig. 2.29.

Given the short-term stability of frequency source f_0 is $\sigma_{y0}(\tau)$, the frequency after passing through a unit is converted into Nf_0 , additional frequency fluctuation noise during conversion is Δf_A , its short-term stability is $\sigma_{yA}(\tau)$, and total short-term stability is $\sigma_{yN}(\tau)$.

Because

$$\Delta f_N = \sqrt{[N(\Delta f_0)]^2 + [\Delta f_0]^2}, \quad f_N = Nf_0$$

we have

$$\left(\frac{\Delta f_N}{f_N}\right)^2 = \left(\frac{\Delta f_0}{f_0}\right)^2 + \left(\frac{\Delta f_A}{f_N}\right)^2, \quad \left(\frac{\Delta f_A}{f_N}\right)^2 = \left(\frac{\Delta f_N}{f_N}\right)^2 - \left(\frac{\Delta f_0}{f_0}\right)^2 \quad (2.85)$$

Total short-term stability after adding frequency fluctuation noise Δf_A into this unit is

$$\sigma_{yN}(\tau) = \sqrt{\sigma_{yA}^2(\tau) - \sigma_{y0}^2(\tau)} \quad (2.86)$$

When $\sigma_{yN} \gg \sigma_{y0}$, $\sigma_{yN}(\tau) \approx \sigma_{yA}(\tau)$, i.e., when additional frequency fluctuation noise is higher, the short-term stability of output signal of this unit will depend on additional noise and can't be improved by increasing short-term stability $\sigma_{y0}(\tau)$ of frequency source. This is a phenomenon found in system ground joint test. In addition, in the ground joint test, $T \rightarrow 0$ results in $\gamma \rightarrow 0$ and thus $B_2 \rightarrow 0$ according to Table 2.5. According to Expression (2.78), velocity-measuring error caused by short-term stability shall be 0, but the actual measured result can't be zero. This is the result of additional frequency fluctuation noise.

Short-term stability of frequency source and effects of additional frequency fluctuation noise of each unit in the link have been discussed in the above paragraphs. Because they aren't correlated with each other, the comprehensive effects of them should be considered by using mean square sum to solve total frequency fluctuation noise caused by them and Allan variance. Since additional noise is PM white noise and hence $B_2 = 2/3$ is feasible, velocity-measuring error component caused by additional noise can be obtained using the Expression (2.78) as follows:

$$\sigma_{RA}^* = \frac{C}{\sqrt{3}} \sigma_{yA}(\tau) \quad (2.87)$$

because total velocity-measuring error is expressed as

$$\sigma_R^* = \sqrt{\sigma_{Ro}^2 + \sigma_{RA}^2} \quad (2.88)$$

Therefore, for $\mu = -2$ noise (PM white noise and PM flicker noise), it is expressed as

$$\sigma_R^* = \frac{C}{\sqrt{3}} \sqrt{\sigma_y^2(\tau) + \sigma_{yA}^2(\tau)} \quad (2.89)$$

For $\mu = 0, \mu = -1$ noise (FM white noise and FM flicker noise), it is expressed as

$$\sigma_R^* = C \sqrt{\frac{\sigma_{yA}^2(\tau)}{3} + \frac{B_2}{2} \sigma_y^2(\tau)} \quad (2.90)$$

where B_2 is obtained by lookup Table 2.5 with $\mu = -1$ and $\gamma = T/\tau$.

The above time domain analysis method can be used for performance allotment and accuracy calculation using short-term stability. However, the measurement of short-term stability during test and adjustment is very difficult and requires specific and expensive measuring instrument for short-term stability. When the “frequency domain analysis method” is used, the spectrum analyzer can be used to directly measure the phase noise. This “frequency domain analysis method” will be discussed in the following paragraph.

Frequency Domain Analysis Method of Velocity Measurement Accuracy [3]

(1) Relationship between velocity measurement error and spectral density of phase noise

In section above, the relationship between Allan variance and the velocity measurement error obtained by time domain analysis method is:

$$\sigma_R^2(\tau) = \frac{c^2}{2} \times B_2(\gamma, \mu) \sigma_y^2(\tau) \quad (2.91)$$

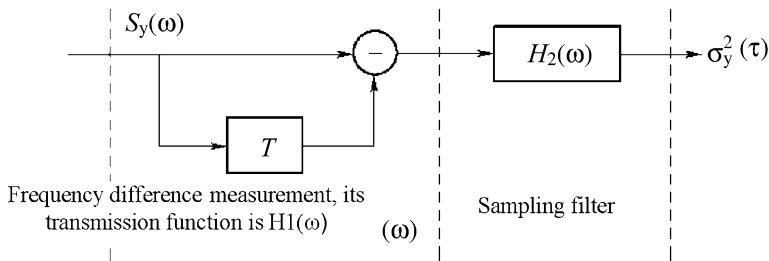


Fig. 2.30 The model of velocity measurement error caused by the phase noise of transmission signal

The power law spectrum model is hypothetical and is not fully consistent with actual phase noise curve; in conjunction with the $B_2(\gamma, \mu)$ obtained on the basis of some supposed conditions, Expression (2.91) is approximate. This is the disadvantage of time domain analysis method. Frequency domain analysis method can be used to overcome these disadvantages, in which the relationship between power spectral density of phase noise and velocity measurement error is derived as follows:

The model of velocity measurement error caused by the phase noise of transmission signal is shown in Fig. 2.30.

In Fig. 2.30, $H_1(\omega)$ is delay T cancellation, its power transmission function feature is expressed as:

$$|H_1(\omega)|^2 = 4 \sin^2 \frac{\omega T}{2} \quad (2.92)$$

Its characteristic curve is as shown as $|H_1(f)|$ in Fig. 2.31.

$H_2(\omega)$ is sampling smoothing filter within sampling time τ , and its power transmission function feature is [4]:

$$|H_2(\omega)|^2 = \sin^2 \frac{(\omega\tau/2)}{(\omega\tau/2)^2} \quad (2.93)$$

Its feature curve is as shown as $|H_2(f)|$ in Fig. 2.31.

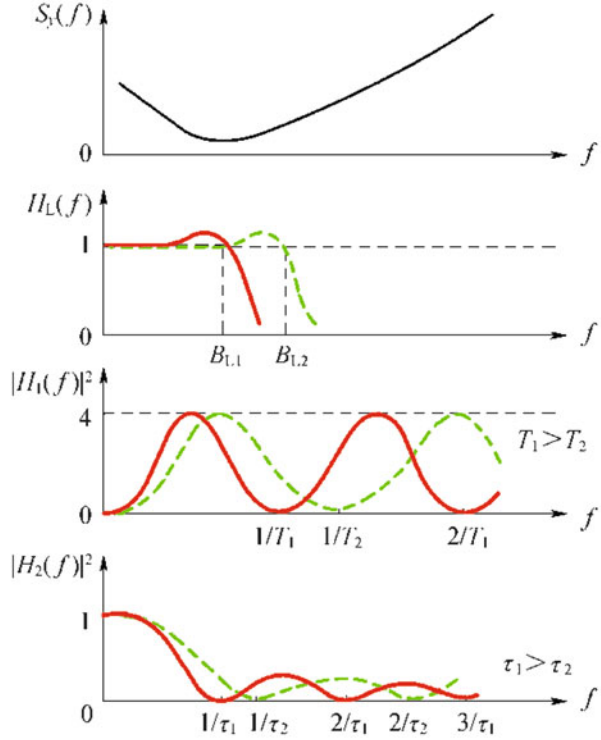
The sampling variance value of relative frequency fluctuating noise at the output end is

$$\sigma_y^2(\tau, T) = \frac{1}{2\pi} \int_0^\infty S_y(\omega) |H_1(\omega)|^2 |H_2(\omega)|^2 d\omega \quad (2.94)$$

Create simultaneous expression out of combined Expression (2.92) with Expression (2.94) to yield:

$$\sigma_y^2(\tau, T) = \frac{1}{2\pi} \int_0^\infty S_y(\omega) \left[4 \sin^2 \left(\frac{\omega T}{2} \right) \right] \left[\sin^2 \frac{(\omega T/2)}{(\omega T/2)^2} \right] d\omega \quad (2.95)$$

Fig. 2.31 Illustration of frequency domain analysis of velocity measurement error



where ω_0^2, T, τ are known values. As $S_y(\omega) = S_\Phi(\Phi)(\omega^2/\omega_0^2), \sigma_R^2(\tau, T)$ can be obtained via computer using measured $S_\Phi(\omega)$ (or using power law spectrum model). The multiplier ω^2 denotes that the spectral density of frequency fluctuation noise will increase as ω increases. In practical proposal, before the Doppler frequency difference is measured, the phase-locked loop typically is used to perform tracking filter to $S_\Phi(\omega)$. It supposes the single-side bandwidth of the phase-locked loop is $\omega_1(\omega_1 = 2\pi B_L)$; then, Expression (2.95) becomes:

$$\sigma_y^2(\tau, T, B_L) = \frac{1}{\pi} \int_0^{\omega_L} S_y(\omega) \left[\sin^2 \left(\frac{\omega T}{2} \right) \right] \left[\sin^2 \left(\frac{\omega \tau}{2} \right) / \left(\frac{\omega \tau}{2} \right)^2 \right] d\omega \quad (2.96)$$

Expression (2.96) shows that $\sigma_y^2(\tau, T, B_L)$ is correlated with τ, T , and B_L . The smaller τ is (i.e., the wider the $H_2(\omega)$ is) and the bigger B_L is, and so the bigger $\sigma_R^2(\tau, T, B_L)$ is.

The relationship between the gapped sampling variance and spectral density is as follows [4]:

$$\sigma_y^2(N, T, \tau) = \frac{N}{N-1} \int_0^\infty S_y(f) \left(\frac{\sin^2 \pi \tau f}{\pi \tau f} \right)^2 \left[1 - \left(\frac{\sin N \pi \tau f}{N \sin \pi \tau f} \right)^2 \right] df$$

where $\gamma = T$ and $\omega = 2$. Substituting $N = 2$ into above expression, the gapped Allan variance within the bandwidth can be obtained:

$$\sigma_y^2(2, T, \tau, B_L) = \frac{1}{\pi} \int_0^{\omega L} S_y(\omega) \left[\sin^2 \left(\frac{\omega T}{2} \right) \right] \left[\sin^2 \frac{(\omega \tau / 2)}{(\omega \tau / 2)^2} \right] d\omega$$

Therefore,

$$\sigma_y^2(\tau, T, B_L) = 2\sigma_y^2(2, T, \tau, B_L)$$

and the velocity measurement error is:

$$\begin{aligned} \sigma_R^2 &= \left(\frac{c}{2} \right)^2 \sigma_y^2(\tau, T, B_L) = \left(\frac{c}{2} \right)^2 2\sigma_y^2(2, T, \tau, B_L) = \left(\frac{c}{2} \right)^2 2B(\gamma) \sigma_y^2(\tau) \\ \sigma_R &= \frac{c}{\sqrt{2}} \sqrt{B_2} \sigma_y(\tau) \end{aligned} \quad (2.97)$$

Expression (2.97) is in consistency with the velocity measurement error Expression (2.51) obtained by the time domain analysis method. But the frequency domain method is derived by filtering phase noise frequency spectrum, and its $\sigma_y^2(\tau)$ is the result after the signal is subject to phase-locked filtering (the bandwidth is B_L). The physical concept of such method is clear and applicable to solve actual problems in projects.

The physical implication is shown from Fig. 2.31: Suppose that input is frequency modulation white noise and its power spectrum density is $S_y(f)$.

It can be seen in Fig. 2.31 that for the given distance (corresponding to radio transmission delay T_1 in the figure), the smaller τ_2 is, the bigger $1/\tau_2$ is and the wider the corresponding $|H_2(f)|$ frequency window. Therefore, the bigger the frequency fluctuating noise is, the bigger the velocity measurement error becomes.

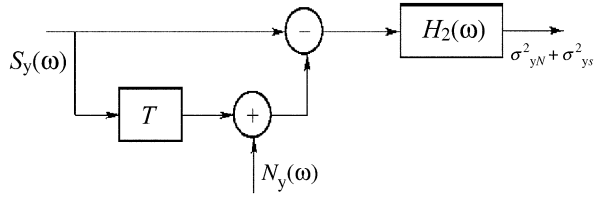
For a given sampling time (such as t_1 in Fig. 2.31), the smaller the T is, the value of $|H_1(f)|$ corresponding to frequency window of $1/\tau_1$ will be. Therefore, the smaller the frequency fluctuating noise is, the velocity measurement error will become.

For given τ and T , the wider B_L is, the greater the areas of $|H_1(f)|$ and $|H_2(f)|$ (equivalent noise bandwidths) will be. Therefore, the bigger the output frequency fluctuating noise is, the bigger the velocity measurement error will become.

n/τ , $n = 1, 2, 3 \dots$ positive integers are dumb points of the sampling filter. Therefore, τ value can be selected to eliminate interference. For example, for 50 Hz power supply jamming, $n/\tau = 50$ Hz can be selected, namely, $\tau = n \times 20$ ms (at this time, τ may be 20 ms, 40 ms, 80 ms, 120 ms. . .).

If noncoherent velocity measurement is used, there is no filtering of $H_1(f)$ and the velocity measurement error will increase. Especially, there is no null point of

Fig. 2.32 Velocity-measurement error model caused by noise of the receiving system



$H_1(f)$ in the vicinity of $f=0$, which strengthens the impact of frequency random walk and frequency modulation flash noise.

(2) Relationship between thermal noise of a receiver and velocity-measuring error

Velocity-measuring error is not only correlated with short-term frequency stability but also relevant to the thermal noise of the receiving system. The model of velocity measurement error caused by noise of receiving system is shown in Fig. 2.32.

In Fig. 2.32, $S_y(\omega)$ is the one-sided power spectrum density of relative frequency fluctuation noise (short-term frequency stability) of signals and $N_y(\omega)$ is generated by noise of the receiving system (thermal noise of the receiver is generally considered as additive normal white noise). Section 2.3.3.1 has provided the formula $N_y(\omega) = N_\phi(\omega)(\omega^2/\omega_0^2) = (N_0/P_s)(\omega^2/\omega_0^2)$ (P_s refers to the power of receiving signals and N_0 refers to power spectrum density of amplitude noise of the receiving system). Since $N_y(f)$ is non-correlated with $S_y(f)$, $N_y(f)$ and $S_y(f)$ power is added when the difference is solved. The output after $|H_2(f)|$ filtering includes two parts: σ_{yN}^2 (velocity variance generated by noise of receiving system) and σ_{ys}^2 (velocity variance generated by short-term frequency stability of transmitting signals). The total output is the mean square sum. The solution of σ_{ys}^2 has been provided above, and σ_{yN}^2 can be derived from Fig. 2.32, namely,

$$\sigma_{yN}^2(\tau, B_L) = \frac{1}{2\pi} \int_0^{\omega_L} N_y(\omega) |H_2(\omega)|^2 d\omega \quad (2.98)$$

where $|H_2(\omega)|^2$ has been given in Expression (2.93), substituting it into Expression (2.98) to obtain:

$$\begin{aligned} \sigma_{yN}^2(\tau, B_L) &= \frac{1}{2\pi} \int_0^{\omega_L} \frac{N_0}{P_s} \frac{\omega^2}{\omega_0^2} \left[\frac{\sin^2(\omega\tau/2)}{(\omega\tau/2)^2} \right] d\omega \\ &= \frac{1}{\pi \omega_0^2 \tau^3} \frac{N_0}{P_s} (\omega_L \tau - \sin \omega_L \tau) \end{aligned} \quad (2.99)$$

When $\omega_L \tau \gg 1$, we can get

$$\sigma_{yN}^2(\tau, B_L) = \frac{1}{\sqrt{2}\pi f_0 \tau} \sqrt{\frac{N_0}{P_s}} B_L \quad (2.100)$$

The velocity-measuring error introduced by thermal noises is:

$$\sigma_{\dot{R}N}(\tau, B_L) = \frac{C}{2\sqrt{2}\pi f_0 \tau} \sqrt{\frac{N_0}{P_s}} B_L = \frac{C}{2\sqrt{2}\pi f_0 \tau} \sqrt{\frac{1}{\rho_L}} \quad (2.101)$$

where $\rho_1 = P_s/B_L N_0$ is SNR of the downlink carrier loop; $N_0 = KT$, K is Boltzmann constant and T refers to the noise temperature; and B_L is the loop sideband equivalent noise bandwidth.

According to Expression (2.101), by reducing B_L , the velocity measurement error introduced by thermal noise can be reduced accordingly. However, decrease in B_L will enlarge dynamic error. To meet dynamic error requirements, higher-type higher-order narrowband loop shall be employed (“type” refers to the number of integrators in a loop). According to the classification method of the phase-locked loop types in Reference [9], the transfer functions and the calculation formula of its bandwidth of several phase-locked loops are introduced in the section below:

Type II order II:

$$H(s) = \frac{2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (2.102)$$

Type II order III:

$$H(s) = \frac{K_{\tau_2}(s\tau_2 + 1)}{s^3\tau_2^3/b + s^2\tau_2^2 + Ks\tau_2^2 + K\tau_2} \quad (2.103)$$

Type III order III:

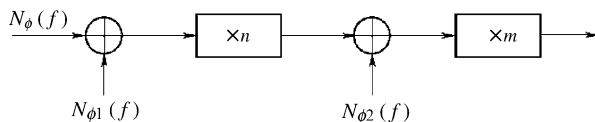
$$H(s) = \frac{K\left(s^2 + \frac{K_2}{K_1}s + \frac{K_3}{K_1}\right)}{s^3 + K\left(s^2 + \frac{K_2}{K_1}s + \frac{K_3}{K_1}\right)} \quad (2.104)$$

B_L of these loops can be calculated by the formula in Expression (2.7).

(3) Features of frequency domain method

According to the analysis above, the features of frequency domain analysis method are concluded that such method is more practical for equipment R&D because the power spectrum density of phase noise can be obtained through measuring with spectrum analyzer and the physical concept is more clear and intuitive and more favorable for the researcher to find and solve problems.

Fig. 2.33 Equivalent circuit of a frequency multiplier



Frequency domain method enjoys higher accuracy. The power spectrum density of phase noise measured by spectrum analyzer can work out the velocity measurement error, or the measured power spectrum density can be input into computers for accurate calculation or simulation.

“Phase domain transfer” and “amplitude domain transfer” are similar in many aspects, including:

1) Frequency doubling of “phase domain” is equivalent to “amplification” of “amplitude domain.” Namely, n times of doubling of signal frequency equals to n times of amplification of phase noise. The signal frequency is equivalent to signal amplitude. The power spectrum density of output phase noise is n^2 times of input phase noise spectrum density, that is, dB is $20 \lg n$. When the frequency multiplier generates internal phase noise $N_\phi(f)$, the equivalent circuit is as shown in Fig. 2.33. Similar to low-noise amplifier, phase noise is mainly influenced at the first stage of a frequency multiplier; therefore, the frequency multiplier with low-phase noise shall be designed at the first stage. The subsequent order phase noise $N_{\phi 2}(f)$ influence can be reduced by increasing n . Analysis of frequency divider is similar to that of frequency multiplier.

2) The “subtractor” in “phase domain” is equivalent to “down frequency mixer” in “amplitude domain.” For phase, the down frequency mixer equals to the “subtractor.” Similarly, for up-frequency mixing, it equals to “adder.” When phase noise of ϕ_1 and ϕ_2 is not correlated mutually, both “subtractor” and “adder” implement mean square addition to phase noise.

3) Frequency multiplier and phase-locked frequency synthesizer: the power spectrum density of phase noise output from n -time frequency multiplier increases by n^2 times in theory, but in fact, partial internal noise will be added at the front stage of the frequency multiplier. If frequency multiplier with low-phase noise is designed, such additional internal noise is relatively small. When phase-locked loop frequency synthesizer is used to implement similar frequency doubling functions, the following phase noise will be generated in the phase-locked frequency synthesizer:

- ① Phase noise of frequency divider
- ② Phase discriminator of locked-phase loop and noise of its post amplifier, which will cause phase jitter of VCO
- ③ Phase noise of VCO, influencing phase noise outside of the phase-locked loop bandwidth

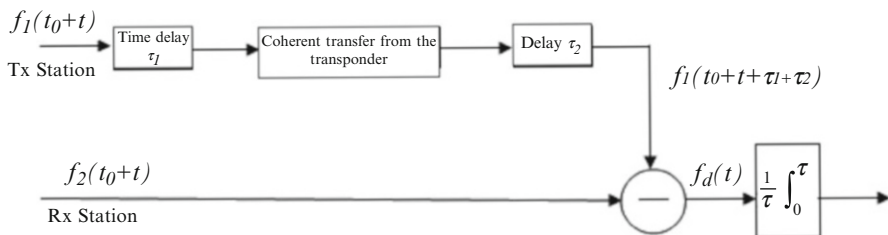


Fig. 2.34 The model of a three-way incoherent Doppler velocity measuring

2.3.4 Three-Way Noncoherent Doppler Velocity Measuring and $\dot{S}$ Measuring

“Three-way” is relative to “two-way.” In “two-way” measuring, only one ground station is needed; the forward and backward signals of which are used to realize two-way measuring. Three-way measuring happens in the condition that another ground station is used to realize receiving in a third way; in this case, frequency sources of the transmitting and receiving stations are incoherent, that is the reason it is called noncoherent velocity measuring. The model of a three-way noncoherent Doppler velocity measuring is shown in Fig. 2.34.

In Fig. 2.34, $f_1(t)$ is the uplink frequency of a ground-transmitting station and $f_2(t)$ means the reference frequency of a ground receiving station. Using $f_2(t)$ to measure $f_d(t)$, we find that $f_1(t)$ and $f_2(t)$ are incoherent. t_0 is the starting time to measure frequency drifts of $f_1(t)$ and $f_2(t)$, $(\tau_1 + \tau_2)$ is the time delay from the transmitting station to the receiving station, and t represents the continuous variable starting from 0. Compared with two-way coherent velocity measuring, the three-way incoherent measuring introduces the following velocity-measuring errors:

(1) The error caused by frequency drift $v(T)$

The $v(T)$ is defined as the change of frequency appears after time interval of T , which starts from t_0 ; and $(t_0 + T)$ here means the time of Doppler frequency measuring. Assume that frequency drifts of $f_1(t)$ and $f_2(t)$ are $v_1(T)$ and $v_2(T)$, respectively; then, at the time of $(t_0 + T)$, they will introduce an error of Doppler frequency measuring:

$$F_1(T) = \sqrt{v_1^2(T) + v_2^2(T)} \quad (2.105)$$

The reason why the above expression is in the form of mean square addition is because $f_1(t)$ and $f_2(t)$ are independent. Take f_2 as a reference, the velocity-measuring error resulted from F_1 is

$$\Delta_1 = \frac{C}{2} \frac{\sqrt{v_1^2(T) + v_2^2(T)}}{f_2} \quad (2.106)$$

(2) The error caused by frequency accuracy $f_1(t_0)$ and $f_2(t_0)$ are respectively the actual values of $f_1(t)$ and $f_2(t)$ at the time of t_0 , which have an error to the given nominal value. As velocity calculation is based on the nominal value, this error will directly bring in a Doppler frequency error. Assume that frequency accuracies of $f_1(t)$ and $f_2(t)$ are $A_1 = \Delta f_1/f_1$ and $A_2 = \Delta f_2/f_2$, respectively; then, the Doppler frequency error introduced by them is

$$F_2(t_0) = \sqrt{A_1^2 f_1^2 + A_2^2 f_2^2} \quad (2.107)$$

The velocity-measuring error caused therefrom is:

$$\Delta_2 = \frac{C}{2} \frac{\sqrt{A_1^2 f_1^2 + A_2^2 f_2^2}}{f_2} \quad (2.108)$$

(3) The velocity-measuring error caused by short-term instability
According to Expression (2.49), the mean value of $f_d(T)$ is:

$$\begin{aligned} \overline{f_d(T)} &= \overline{f_1(t_0 + T + \tau_1 + \tau_2)} - \overline{f_1(t_0 + T)} \\ &= \left[\overline{f_1(t_0 + T + \tau_1 + \tau_2)} - \overline{f_1(t_0)} \right] - \left[\overline{f_2(t_0 + T)} - \overline{f_2(t_0)} \right] \\ &\quad + \left[\overline{f_2(t_0)} - \overline{f_1(t_0)} \right] \end{aligned} \quad (2.109)$$

Expression (2.109) includes three components:

1) Item $\left[\overline{f_2(t_0)} - \overline{f_1(t_0)} \right]$

This is the system error caused by frequency accuracy, which can be calculated according to Expression (2.108).

2) Item $\left[\overline{f_2(t_0 + T)} - \overline{f_2(t_0)} \right]$

This is the random error resulted from $f_2(t)$ short-term instability. According to Expression (2.50), there is:

$$\left\langle \left[\overline{f_2(t_0 + t)} - \overline{f_2(t_0)} \right]_{m^2} \right\rangle = 2f_2^2 \left\langle \sigma_{y_2}^2(2, T, \tau) \right\rangle$$

where $\sigma_{y_2}^2(2, T, \tau)$ is the gapped Allan variance of $f_2(t)$ when the sampling interval is T and the integration time is τ . This error contains FM white noise, PM white noise, FM flicker noise, and frequency random walk noise. As T lasts too long (e.g., several hours or days) during a three-way velocity measuring, the last two items are contained in frequency drift. The other two errors can be calculated according to the following formula:

① FM white noise ($\mu = -1$ noise):

According to Expression (2.52), the variance of the velocity-measuring error caused by $f_2(t)$ short-term instability is:

$$\left[\sigma_{\dot{R}_2}^2 \right]_{-1} = \frac{C^2}{2} B_2(\gamma, \mu) \sigma_{y_2}^2(\tau)$$

As the value of T is quite large, $\gamma = T/\tau$ is also huge; we look up $B_2(\gamma, \mu) = 1$ from Table (2.5); then, the velocity-measuring error caused by PM white noise is

$$\left[\sigma_{\dot{R}_2} \right]_{-1} = \frac{C}{\sqrt{2}} \sigma_{y_1}(\tau) \quad (2.110)$$

where $\sigma_{y_2}^2(\tau)$ is the Allan variance of $f_2(t)$ at the sampling time τ .

② PM white noise (noise for $\mu = -2$)

In the same way the error of $\left[\sigma_{\dot{R}_2} \right]_{-2}$ is obtained, we get $B_2(\gamma, \mu) = 2/3$ from Fig. 2.5, thus the velocity-measuring error caused by PM white noise is

$$\left[\sigma_{\dot{R}_2} \right]_{-2} = \frac{C}{\sqrt{3}} \sigma_{y_1}(\tau) \quad (2.111)$$

3) Item $\left[\overline{f_1(t_0 + T + \tau_1 + \tau_2)} - \overline{f_1(t_0)} \right]$

Through analysis the same with that for $\left[\overline{f_2(t_0 + T)} - \overline{f_2(t_0)} \right]$, we get:

① FM white noise

$$\left[\sigma_{\dot{R}_1} \right]_{-1} = \frac{C}{\sqrt{2}} \sigma_{y_1}(\tau) \quad (2.112)$$

where $\sigma_{y_1}(\tau)$ is the Allan variance of $f_1(t)$ at the sampling time τ .

② PM white noise

$$\left[\sigma_{\dot{R}_2} \right]_{-2} = \frac{C}{\sqrt{3}} \sigma_{y_1}(\tau) \quad (2.113)$$

As the errors above are independent, the total velocity-measuring variance $\sigma_{\dot{R}}^2$ obtained in Fig. 2.34 is the mean square sum of the above errors.

$$\sigma_{\dot{R}}^2 = \Delta_1^2 + \Delta_2^2 + \left[\sigma_{\dot{R}_2} \right]_{-1}^2 + \left[\sigma_{\dot{R}_2} \right]_{-2}^2 + \left[\sigma_{\dot{R}_1} \right]_{-1}^2 + \left[\sigma_{\dot{R}_1} \right]_{-2}^2 \quad (2.114)$$

By substituting (2.106), (2.108), (2.110), (2.111), (2.112), and (2.113) into (2.114), we get:

$$\sigma_{\dot{R}} = \frac{C}{\sqrt{2}} \left[\frac{V_1^2(T) + V_2^2(T) + A_1^2 f_1^2 + A_2^2 f_2^2}{2f_2^2} + \frac{5}{3} \sigma_{y_2}^2(\tau) + \frac{5}{3} \sigma_{y_1}^2(\tau) \right]^{\frac{1}{2}} \quad (2.115)$$

Expression (2.114) indicates that the accuracy of a three-way incoherent velocity measuring is not as satisfactory as that of the two-way coherent velocity measuring, the reason of which is (1) errors caused by “frequency drift” and “accuracy” of frequency sources and (2) among error terms resulted from the short-term stability; the mean square of each error term is added in three-way measuring, while in two-way measuring, the mean squares can be partly counteracted. Though there are weak points as mentioned above, three-way incoherent velocity measuring is still applied in some special areas, such as:

(a) Deep-space TT&C: that deep-space probes are far away from the Earth results into long delay of TT&C signals from a probe to the ground station or vice versa. This means when the echo arrives at the Earth, the original transmitting ground station has lost its sight to the echo because it is moving with the Earth’s rotation (namely, the station’s location is below the horizon). In this case, another ground station to receive the echo is necessary. Although the probe is far away from the ground station, the visual angles of the transmitting and the receiving stations for the vehicle are basically the same, so do the forward and the backward velocity. Thus calculation can be done according to Expression (2.115).

(b) Multiple $\dot{S}$ measuring at a range: in this case, multiple slave ground stations for only receiving are needed to measure the velocity and ($\dot{S}$) from ground main station $\rightarrow$ transponder $\rightarrow$ slave ground station. It is also a kind of incoherent three-way Doppler velocity measuring. In this case, if the visual angle difference of the transmitting station and the receiving station for the vehicle is not very small, then the forward and the backward velocities are different; and the measured is the sum of the forward/backward Doppler frequencies, namely, the measured is $\dot{S}$, the expression of which is:

$$\sigma_{\dot{S}} = \sqrt{2}C \left[\frac{V_1^2(T) + V_2^2(T) + A_1^2 f_1^2 + A_2^2 f_2^2}{2f_2^2} + \frac{5}{3} \sigma_{y_2}^2(\tau) + \frac{5}{3} \sigma_{y_1}^2(\tau) \right]^{\frac{1}{2}} \quad (2.116)$$

2.3.5 Two-Way Noncoherent Doppler Velocity Measuring

The expression “two-way noncoherent” means a ground station is used to transmit and receive forward and backward signals, but a transponder is used for noncoherent responding. Compared with coherent responding, it will introduce a noncoherent system error, so the coherent Doppler velocity measurement provides highest accuracy. To reduce the velocity-measuring system error caused by noncoherent components, measures are taken to counteract the noncoherent components.

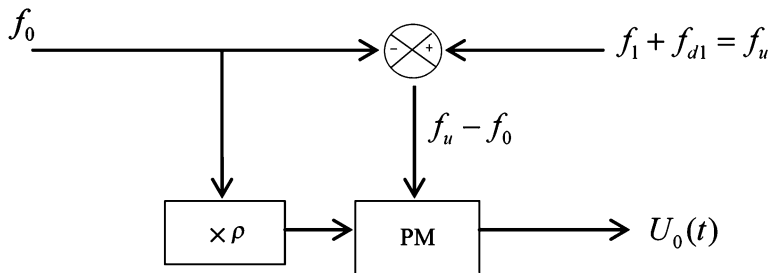


Fig. 2.35 Principle of transponder modulation retransmission

2.3.5.1 Ground Counteracting

These options include “mixing responding” and “modulation responding.” The latter is comparatively better, the block diagrams of which are shown in Figs. 2.16 and 2.17. For the convenience of introduction of its counteracting principle of noncoherent components, Fig. 2.16 is simplified to the condition that the transponder responds to only one uplink signal, as shown in Fig. 2.35.

In Fig. 2.35, $\rho = N/M$ is the turnaround ratio, f_1 is the uplink frequency transmitted by the ground station, which is the reference in measuring of Doppler frequency drift. f_0 is the local oscillator (LO) frequency introduced by the transponder, which is noncoherent to f_1 . When an onboard transponder moves at a speed of V , the received frequency is $f_1 + f_{d1}$, and f_{d1} is an uplink Doppler frequency. According to the Doppler effect, there is

$$\begin{aligned} f_1 + f_{d1} &= \frac{C - V}{C} f_1 = K_1 f_1 \\ K_1 &= \frac{C - V}{C} \end{aligned} \quad (2.117)$$

where K_1 is the coefficient of the uplink Doppler frequency drift.

$(f_1 + f_{d1})$ and f_0 is mixed to form subcarrier frequency $[(f_1 + f_{d1}) - f_0]$ which is used to conduct phase modulation (PM) for the noncoherent frequency (ρf_0) , and the generated PM wave is $U_1(t)$:

$$\begin{aligned} U_1(t) &= J_0(m) \sin \rho \omega_0 t + J_1(m) [\rho \omega_0 + (K_1 \omega_1 - \omega_0)] t \\ &\quad - J_1(m) \sin [\rho \omega_0 - (K_1 \omega_1 - \omega_0)] t + \dots \end{aligned} \quad (2.118)$$

where m is a PM index and $J_0(m), J_1(m)$ are zero-order and first-order Bessel functions, respectively. The following only discuss $J_0(m), J_1(m)$. $U_1(t)$, when

reaching the ground station and changing into $U_2(t)$, and under the influence of the Doppler effect, its frequency f_2 changes into:

$$\begin{aligned} f_2 + f_{d2} &= f_2 \frac{C - V}{C} = K_2 f_2 \\ K_2 &= \frac{C}{C + V} \end{aligned} \quad (2.119)$$

where K_2 is the coefficient of the downlink Doppler frequency drift. With this coefficient, we get the following from Expression (2.118):

$$\begin{aligned} U_2(t) &= J_0(m) \sin K_2 \rho \omega_0 t + J_1(m) \sin K_2 [\rho \omega_0 + (K_1 \omega_1 - \omega_0)] t \\ &\quad - J_1(m) \sin K_2 [\rho \omega_0 - (K_1 \omega_1 - \omega_0)] t \end{aligned}$$

Use a phase-locked loop to draw out residual carrier item $c(t)$:

$$c(t) = J_0(m) \sin K_2 \rho \omega_0 t$$

where $\omega_c = K_2 \rho \omega_0$ is the residual carrier frequency; use $c(t) \times U_2(t)$ to demodulate subcarrier item $B(t)$.

$$B(t) = J_1(m) \sin [K_2 (K_1 \omega_1 - \omega_0)] t = J_1(m) \sin \omega_B t$$

where $\omega_B = K_2 (K_1 \omega_1 - \omega_0)$ is a subcarrier frequency.

Measure out ω_c , ω_B and ω_1 at the ground station and then get the two-way carrier Doppler frequency ω_D by the following calculation:

$$\omega_D = \omega_B + \frac{\omega_c}{\rho} - \omega_1 = K_1 K_2 \omega_1 - \omega_1 \quad (2.120)$$

Substitute Expressions (2.117) and (2.119) into Expression (2.120) to obtain:

$$\begin{aligned} \omega_D &= \left[\left(\frac{C - V}{C} \right) \left[\frac{C}{C + V} \right] - 1 \right] \omega_1 = \frac{-2V}{C + V} \omega_1 \\ V &= \frac{\omega_D}{2\omega_1 + \omega_D} C = \frac{f_D}{2f_1 + f_D} C \end{aligned} \quad (2.121)$$

By comparing Expression (2.121) with Expression (2.22) in Sect. 2.3.1, we find that they are completely the same. Thus we know that the application of this noncoherent measuring of “modulation responding” realizes coherent velocity measuring and their measuring accuracy analyses are also the same, but two additional errors will be introduced.

(1) The influence of additional non-coherent phase noise: In Expression (2.120), suppose ω_c/ρ of $c(t)$ and $K_2\omega_0$ of $B(t)$ are equal, so their difference is zero. This result is based on frequency accuracy, drift, short-term stability, etc. of f_0 generated coherent components in $c(t)$ and $B(t)$. However, if there are some noncoherent components added into $c(t)$ and $B(t)$, then their mean squares will be added rather than they are counteracted; these additional noncoherent phase noises include: ① phase noises introduced by different parts of transmission channels of $c(t)$ and $B(t)$; ② measurement error and phase noises introduced respectively while ω_c and ω_B are being measured; ③ after f_0 passes through $c(t)$ and $B(t)$, the cross-correlation weakens due to different time delays of the transmission channels of $c(t)$ and $B(t)$; and ④ phase noises added when frequency turnaround ratio ρ is generated, namely, in Fig. 2.16, the noncoherent phase noises added respectively when M-order and N-order frequency doubling are generated; this condition is equal to make random changes to ρ . In calculation of Expression (2.120), ρ is assumed to be a designed constant, so the calculation result will have an error. Above-mentioned factors in the development should be reduced as much as possible; it has been proved in practice that their influence can be reduced to a tiny extent, only occupying a very small part of velocity-measuring error.

(2) The influence of thermal noises: Phase noises caused by thermal noises in $c(t)$ and $B(t)$ are noncoherent. Assume that the carrier-to-noise ratio of $c(t)$ is J_0/N_0 ; then, the power spectral density of the relevant phase noise is N_0/J_0 . The two are noncoherent, the power spectral density of the phase noise of their sum Φ_D is:

$$\Phi_D = \frac{N_0}{J_1} + \frac{N_0}{J_0} = \left(\frac{1}{J_1} + \frac{1}{J_0} \right) N_0 = \left(1 + \frac{J_1}{J_0} \right) \frac{N_0}{J_1} \quad (2.122)$$

It is clear from here that J_0/N_0 of the residual carrier will produce additional velocity-measuring error; in the expression, $(1 + J_1/J_0)$ is an additional factor, namely, the measuring error caused by thermal noises will deteriorate $\sqrt{1 + J_1/J_0}$ times.

One of the distinguishing good points of noncoherent “modulation responding” is that multiple subcarriers can be adopted to realize multi-station intersected measuring with frequency division multiple access (FDMA). So long as each subcarrier frequency is carefully designed to prevent their combination interferences from falling into useful bandwidth, it will achieve a high measuring accuracy. This technique has been applied in multiple RR high-accuracy measuring systems.

2.3.5.2 Measuring in the Air and Processing on the Ground

This method uses a transponder to measure the data of uplink Doppler frequency and transmits such data to the ground through a data channel, then jointly processes the above-mentioned data with the data of downlink Doppler frequency measured

on the ground, and finally, a two-way Doppler is obtained. The principle is shown as follows:

For the same point on a flight locus, from simultaneous expressions of (2.117) and (2.119), we have:

$$v = \frac{[f_1 f_2 - (f_1 + f_{d1})(f_2 + f_{d2})]}{[f_1 f_2 + (f_1 + f_{d1})(f_2 + f_{d2})]} c \quad (2.123)$$

where f_1 and f_2 are uplink and downlink frequencies, respectively, and $(f_1 + f_{d1})$ and $(f_2 + f_{d2})$ are frequency values to be measured; when the TT&C system is in velocity measuring, f_1 will be regarded as the reference and f_2 is noncoherent with f_1 . One of the conditions that make Expression (2.123) tenable is that the four frequency values in the expression are defined according to the same reference, generally, the ground clock is used as the reference; otherwise, Expressions (2.117) and (2.119) cannot be made simultaneous, and each piece of measured information cannot be integrated either.

In this case, the transponder and the ground station use different frequency measurement clocks. Commonly, $(f_1 + f_{d1})$ is measured out from f_1 , f_2 , and $(f_1 + f_{d2})$ with the ground F_g as the reference and $(f_1 + f_1)_s$ is measured with the transponder clock F_s as the reference; their relationship is:

$$(f_1 + f_{d1}) = \frac{F_s}{F_g} (f_1 + f_{d1})_s \quad (2.124)$$

Substitute Expression (2.124) into Expression (2.123) to obtain:

$$v = \frac{[f_1 f_2 - \frac{F_s}{F_g} (f_1 + f_{d1})_s (f_2 + f_{d2})]}{[f_1 f_2 + \frac{F_s}{F_g} (f_1 + f_{d1})_s (f_2 + f_{d2})]} c \quad (2.125)$$

As noncoherent F_s components are already contained when transponder F_s is taken as the reference to measure $(f_1 + f_{d2})_s$, and transponder downlink transmitting frequency f_2 is obtained with F_s as its reference, thus if they are jointly processed, a part of noncoherent components can be eliminated, but there are still some of them remaining. It is obvious from Expression (2.125) that there are six error factors that can affect v measuring accuracy, which are $(f_1 + f_{d1})_s$, $(f_2 + f_{d2})_s$, F_s , F_g , f_1 and f_2 . Measuring error, accuracy, or changes of these factors will induce velocity-measuring error. Apply total differential to Expression (2.125) to acquire each velocity-measuring error.

Expression (2.125) can also be expressed as follows:

$$(f_1 + f_{d1}) = \frac{F_s(1 + \sigma_s)}{F_g(1 + \sigma_g)} (f_1 + f_{d1})_s$$

where σ_s and σ_g are the short-term stability of the transponder clock and the ground station frequency clock, respectively, and F_s and F_g are the sums of their frequency drifts and frequency accuracy, respectively.

From the above expression, it is clear that σ_s , σ_g and changes of F_s and F_g and will cause velocity-measuring error; accuracy of F_s and F_g will influence velocity-measuring accuracy; as $(f_1 + f_{d1})_s$ is measured onboard (on the transponder), its error is comparatively large (e.g., it is difficult to use an extreme narrowband phase-locked loop). Besides, Expression (2.125) is obtained from simultaneous expressions of (2.117) and (2.119) under the same orbit point condition. Thus, above conditions will not be satisfied and a velocity-measuring error will be introduced if there is relative dislocation or different smoothing time lengths during measuring frequencies of $(f_1 + f_{d1})_s$ and $(f_2 + f_{d2})_s$.

From comparison between Expressions (2.125) and (2.22), it is clear that there are only two error factors f_2 and f_1 in coherent Doppler velocity measuring, while there are six error factors in this noncoherent measuring ($f_2, f_{d2}, f_1, f_{d1}, F_s$ and F_g). Thus, velocity-measuring error in this plan is more than that in coherent measuring; the good point of which is that no additional noncoherent f_0 beacon is needed, but uplink frequency needs to be measured on the transponder, which complicates aerial equipment. Another advantage is that technology of code division multiple access (CDMA) can be used in both up- and downlink to realize modes like multi-station measuring single satellite or single-station measuring multi-satellite and downlink can take broadcast transmission to save power, but CDMA interference to measuring accuracy should be taken into account.

2.3.5.3 Onboard Counteracting

The block diagram of this kind of transponder [10] is shown as follows:

In Fig. 2.36, f_L is an introduced noncoherent frequency. Generally, it makes f_2 noncoherent to f_1 , but if an intermediate frequency $F_2 = \rho F'_1 = \frac{N}{M} F_1$ is generated in a digital signal processor, the output of such transponder will be

$$f_2 = Nf_L + \frac{N}{M} F_1 = Nf_L + \frac{N}{M} (f_1 - Mf_L) = \frac{N}{M} f_1 \quad (2.126)$$

In this scheme, f_2 and f_1 are coherent, and (M/N) is the turnaround ratio. Its physical meaning is that signal frequency changes caused by F_L at terminals A and B of a transmitting mixer are completely coherent, which are entirely counteracted after frequency mixing. However, in an actual circuit, they may only be coherent in part, introducing velocity-measuring error. Reasons for noncoherence include:

(1) $\rho F'_1 \neq \frac{N}{M} F_1$. In the expression, M and N are designed deterministic values, and F_1 is a continuous variable. But this plan aims to use digital signal processing circuits to generate $\rho F'_1$ which is disperse, having quantizing error. The error Δf produced by it is

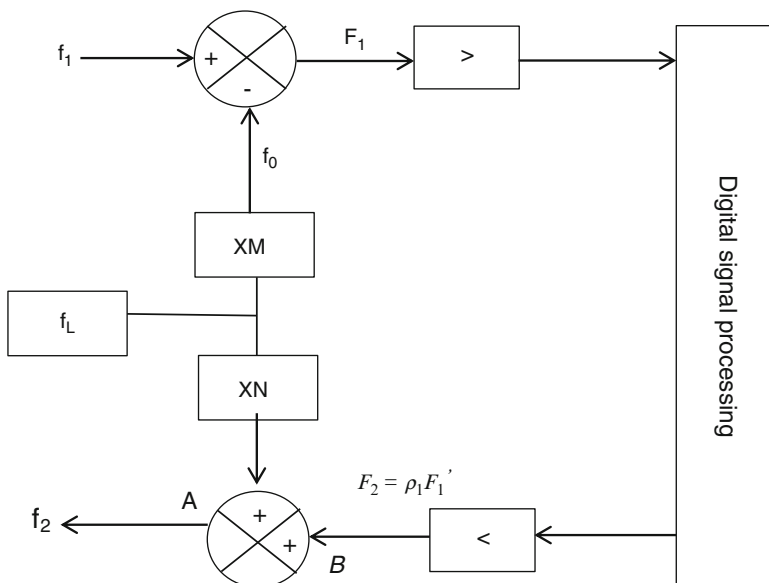


Fig. 2.36 Block diagram of the counteracting principle of noncoherent components in a transponder

$$\Delta f = \frac{N}{M} F_1 - \rho F_1'$$

But Δf will produce velocity-measuring error Δv :

$$\Delta v = \frac{1}{2} \frac{\Delta f}{f_2} C \quad (2.127)$$

A common solution is to use DDS₁ of a digital carrier loop to measure F_1 , getting a proximate frequency code F_1' . A binary number ρ is used to approximate M/N and is multiplied by the frequency code. Their product is filtered digitally and then used to control another DDS₂ to produce f_2 . In this procedure, there exist three parts of error:

- 1) The error in measuring of F_1 with the digital carrier loop: its calculation is determined by phase accumulation digits of DDS₁. Please refer to Expressions (2.24) and (2.25) in this book.
- 2) The error between ρ and the digital filter: it is determined by its word length. Refer to Expression (2.121) for its calculation. As in computer the word length of binary numbers can be long, thus this is not a main component.
- 3) The quantizing error of DDS₂: it is mainly determined by multiples of the phase accumulator of DDS₂. Please refer to Expressions (2.24) and (2.25) in this book for its calculation.

(2) Additional phase noises: In circuit between Points A and B in Fig. 2.36, except f_L , additional phase noises produced by other parts (including those caused by onboard vibration) are noncoherent. They cannot be subtracted but their mean squares are added in the transmitting mixer, causing velocity-measuring error. The analysis about this can be found in Sect. 2.3.3.2 in this book; and refer to Expression (2.87) for its calculation.

(3) Influence from circuit delay time: Namely, make $\rho F' = \frac{N}{M} F_1$, but delay time from f_L to Point A and that from f_L to Point B is different, causing a bad cross-correlation of f_L with Point A and Point B. Thus, the influence from f_L cannot be totally counteracted. Its physical mechanism is the same with that of the influence of transmission signal short-term instability to velocity-measuring accuracy. Please refer to Sect. 2.3.3.2 in this book for its analysis and to Expression (2.52) for its calculation. In the expression $V = T/\tau$, if T is made very small, the error will be small too.

This scheme simplifies and digitalizes the transponder, bringing a series of good points of digitalization and software using. Its weak point is it can only “approximate coherence” rather than “achieve a complete one,” hence introducing a noncoherent velocity-measuring error.

2.3.6 One-Way Noncoherent Doppler Velocity Measuring

The simplified model of a one-way noncoherent Doppler velocity measuring is shown in Fig. 2.37.

In Fig. 2.37, $f_1(t)$ is the frequency of velocity-measuring beacon onboard, $f_2(t)$ is the reference frequency used by the ground station to measure Doppler frequency f_d , t_0 is the starting time to measure frequency drifts of $f_1(t)$ and $f_2(t)$, delay time T is the time difference between t_0 and measuring time ($t_0 + T$), and τ is the integration time of velocity measuring.

(1) The velocity-measuring error introduced by thermal noises

Thermal noises only add phase noises in $f_1(t)$, the analysis method of which is the same with that of the two-way coherent measuring system (see Sect. 2.3.3.2).

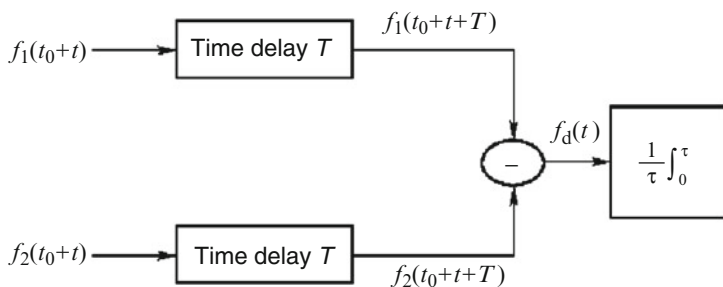


Fig. 2.37 The model of a one-way noncoherent Doppler velocity measuring

One-way Doppler frequency offset is 1 time smaller than that of the two-way one; thus, the error is 1 time larger, hence the expression:

When $\omega_L \tau \gg 1$, there is

$$\sigma_{\dot{R}_1}(\tau, B_L) = 2\sigma_{\dot{R}_N}(\tau, B_L) = \frac{c}{\sqrt{2}\pi f_0 \tau} \sqrt{\frac{N_0 B_L}{P_s}} = \frac{c}{\sqrt{2}\pi f_0 \tau} \sqrt{\frac{1}{\rho_L}} \quad (2.128)$$

where $\rho_1 = P_s/B_L N_0$ is the signal-to-noise ratio of a downlink carrier loop.

(2) The velocity-measuring error introduced under non-coherent conditions

From Fig. 2.37, it is obvious that:

$$\begin{aligned} f_d(t) &= [f_1(t_0 + t + T) - f_2(t_0 + t + T)] \\ &= [f_1(t_0 + t + T) - f_1(t_0)] - [f_2(t_0 + t + T) - f_2(t_0)] + [f_1(t_0) - f_2(t_0)] \end{aligned} \quad (2.129)$$

As $f_2(t)$ is on the ground, its short-term instability can be made better, and there are no phase noises added by its transmission channel; moreover, it has no Doppler changes introduced dynamically by targets, thus it can apply comparatively large time constant filtering. Therefore, generally speaking, the short-term instability of $f_2(t)$ is far better than that of $f_1(t)$, namely,

$$f_d(t) = [f_1(t_0 + t + T) - f_1(t_0 + t + T)] + \Delta F(t_0) \quad (2.130)$$

$$\Delta F(t_0) = [f_1(t_0) - f_2(t_0)] \quad (2.131)$$

From Expressions (2.129) and (2.130), one-way noncoherent Doppler velocity measuring is a special example of the three-way one. The first item of Expression (2.130) can be analyzed with the short-term stability of $f_1(t)$. It is a random error, the analysis method of which is the same with that of a three-way noncoherent Doppler velocity measuring. From Expression (2.115), there is:

$$\sigma_{\dot{R}} = \frac{2c}{\sqrt{2}} \left[\frac{V_1^2(T) + A_1^2 f_1^2}{2f_2^2} + \frac{5}{3} \sigma_{y1}^2(t) \right]^{\frac{1}{2}} \quad (2.132)$$

From Expression (2.132), it is clear that the error of one-way noncoherent Doppler velocity measuring is larger than that of two-way coherent Doppler measuring. The reasons are:

1) For a spacecraft not far from the ground, because T is comparatively large (in two-way coherent measuring, T is just the delay time of signal receiving/transmitting), $r = T/\tau$ increases. From Table 2.5, it is clear that the velocity-measuring error caused by $\mu = +1$ noise (frequency random walk noise) and $\mu = 0$ noise (frequency modulation flicker noise) becomes larger. But in deep-space exploration, as time delay of signal receiving/transmitting is large, the two measuring methods share comparatively small difference. Thus good points of one-way Doppler measuring appear more outstanding.

2) As the Doppler frequency offset of one-way measuring is one time smaller than that of the two-way one, there is one time larger error.

3) System error $\Delta F(t_0)$ brought by frequency drift and accuracy.

The short-term stability of velocity-measuring signals exerts comparatively large influence on error of one-way noncoherent velocity measuring, but if measuring error caused by short-term stability can be made smaller than that of thermal noises in the system design phase, then the influence on the system's total velocity-measuring accuracy is comparatively small.

From the analysis above, it is clear that two-way coherent Doppler velocity measuring can acquire Doppler data with the highest accuracy so far, because it takes the same ground-based frequency standard as the reference signal for Doppler frequency measuring of up- and downlink signals. However, as spacecraft oscillator stability improves, one-way Doppler velocity measuring can match the two-way one, especially in deep-space TT&C. One-way Doppler measuring enjoys simplified equipment and provides better signal-to-noise ratio (SNR) for receiving of spacecraft telemetry data. Reasons for SNR improvement: First, one-way transmission is above spacecraft receiver thermal noises; second, the ground station uses one-way receiving mode during one-way velocity measuring; however, the two-way measuring needs duplex mode which will raise the equivalent noise temperature of the ground receiving system; third, in deep-space TT&C, short-term stability of two-way transmission will be affected by the influence of solar plasma on uplink signals; and fourth, there is another far-reaching good point about one-way measuring, namely, one antenna is enough to receive at the same time one-way Doppler and telemetry information of multiple aircrafts and landers within the same beam, with no need for multiple uplink signals. Therefore, such configuration will economize ground-based resources more effectively and improve estimation accuracy of orbit solution and lander location by carrying out differential measuring.

2.3.7 Theoretical Calculation of Velocity-Measuring Accuracy

Use the following expression to calculate coherent Doppler velocity-measuring error:

$$V = -\frac{1}{2} \frac{f_d}{f_0 + f_d} \quad (2.133)$$

Use the following expression to calculate one-way noncoherent Doppler velocity-measuring error:

$$V = -\frac{c}{f_0} f_d \quad (2.134)$$

2.3.7.1 Random Error

(1) The velocity-measuring error caused by system short-term instability

During coherent Doppler velocity measuring, its expression has already been obtained from Expression (2.52), namely,

$$\sigma_{\dot{R}_1} = \frac{\sqrt{B_2}}{\sqrt{2}} C \sigma_y(\tau) \quad (2.135)$$

where $\sigma_y(\tau)$ is Allan variance and B_2 means Barnas second offset function; refer to Sect. 2.3.3.2 for value choosing of τ and B_2 .

During single-way noncoherent Doppler velocity measuring, there is:

$$\sigma_{\dot{R}} = 1.8c\sigma_y(\tau) \quad (2.136)$$

(2) The velocity-measuring error introduced by thermal noises

1) Coherent Doppler velocity measuring. Expression (2.101) has already given its expression as follows:

$$\sigma_{\dot{R}_2} = \frac{C}{2\sqrt{2}\pi f_0 \tau} \sqrt{\frac{N_0 B_L}{P_s}} = \frac{C}{2\sqrt{2}\pi f_0 \tau} \sqrt{\frac{1}{\rho_L}} \quad (2.137)$$

where f_0 is the carrier frequency; τ means integration time; B_L represents carrier loop single-side bandwidth (refer to Table 2.7 for B_L calculation expression of different types of phase-locked loops); P_s is receiving power; N_0 stands for noise power spectral density; and ρ_L symbolizes the SNR of downlink carrier loop when there still exist residual carriers.

The value of ρ_L changes along with different carrier modulation modes. The expressions are as follows:

(a) Carriers are modulated with non-return-to-zero (NRZ) signals rather than subcarriers, but there still exist residual carriers. The carrier loop SNR will produce additional loss which arises from increased effective noise level of carrier

Table 2.7 Noise bandwidth of common PLL

Types of PLL	Noise bandwidth, B_L/Hz
Type I order I	$K/4$
Type I order II (lagging filter)	$K/4$
Type II order II	$\frac{\omega_n}{2} \left(\zeta + \frac{1}{4\zeta} \right)$
Type II order III	$\frac{K}{4} \left(\frac{1+1/K_{r2}}{1-1/b} \right)$
Type III order III	$\frac{K}{4} \left(\frac{1+\frac{K_2}{KK_1}+\frac{K_1K_3}{KK_2^2}}{1-\frac{K_1K_3}{KK_2^2}} \right)$

synchronization due to overlapping of residual carriers in frequency domain by data sides. In this case, ρ_L shall be calculated according to the expression below:

$$\rho_L = \frac{P_C}{N_0} \Big|_{D/L} \cdot \frac{1}{B_L} \cdot \frac{1}{1 + 2E_S/N_0} \quad (2.138)$$

where E_S/N_0 means the ratio between energy and noise spectral density of each symbol.

(b) In the case of suppressed carrier tracking, the carrier loop SNR is:

$$\rho_L = \frac{P_T}{N_0} \Big|_{D/L} \cdot \frac{S_L}{B_L} \quad (2.139)$$

where $P_T/N_0|_{D/L}$ is the ratio between total signal power and noise spectral density of the downlink; S_L means the square loss of Costas loop.

$$S_L = \frac{2(E_S/N_0)}{1 + 2(E_S/N_0)} \quad (2.140)$$

(c) In the case of QPSK, carrier loop SNR is:

$$\rho_L = \frac{P_T}{N_0} \Big|_{D/L} \cdot \frac{S_{LQ}}{B_L} \quad (2.141)$$

where S_{LQ} is the square loss of the QPSK Costas loop.

$$S_{LQ} = \frac{1}{1 + \frac{9}{2E_{SQ}/N_0} + \frac{6}{(E_{SQ}/N_0)^2} + \frac{3}{2(E_{SQ}/N_0)^3}} \quad (2.142)$$

where E_S/N_0 means ratio between energy and noise spectral density of each quaternary channel symbol.

When telemetry data in non-return-to-zero (NRZ) form is used to directly modulate carriers (i.e., without subcarriers), and there exists imbalance in data (i.e., the number of logic 1 is not equal to that of 0), then the residual carrier loop will jitter. This jitter is a kind of error source for Doppler measuring. The influence of this error depends on the statistical distribution of telemetry data.

2) One-way noncoherent Doppler velocity measuring

$$\sigma_{\dot{R}_2} = \frac{C}{\sqrt{2\pi f_0 \tau}} \sqrt{\frac{N_0 B_L}{P_s}} = \frac{C}{\sqrt{2\pi f_0 \tau}} \sqrt{\frac{1}{\rho_L}} \quad (2.143)$$

3) Thermal noises introduced by the transponder

$$\sigma_{\dot{R}_3} = \frac{c}{2\sqrt{2\pi f_c T}} \sqrt{\frac{1}{\rho_L} + \frac{G^2 B_L}{P_C/N_0|_{U/L}}} \quad (2.144)$$

where T is the measured integration time (S), f_c is downlink carrier frequency (Hz), c is light speed in vacuum (mm/s), G is retransmission ration, B_L is single-side equivalent noise bandwidth of the downlink carrier loop, and $P_C/N_0|_{U/L}$ is the ratio between uplink carrier power and noise spectral density.

The fact that the carrier loop bandwidth of the transponder (uplink) is larger than that of the downlink is supposed in Expression (2.144).

(3) The velocity-measuring error introduced by clutters

Clutters of the transmitter power supply cause spurious phase modulation to the phase modulator and power amplifier; the introduced velocity-measuring error is

$$\sigma_{\dot{R}_4} = \frac{\lambda_i}{2} \frac{\sigma_\varphi}{2\pi T} \quad (2.145)$$

where λ_i is the uplink carrier wavelength, σ_φ is phase jitter caused by spurious phase modulation, and T is sampling time.

(4) The velocity-measuring error introduced by VCO short-term instability

In the receiver phase-locked loop, the measuring error introduced by loop output phase jitter caused by VCO short-term instability is:

$$\sigma_{\dot{R}_5} = \frac{\lambda_i}{2\pi T} \sigma_{\theta V} \quad (2.146)$$

where $\sigma_{\theta V}$ is the loop output phase jitter caused by VCO short-term instability and λ_i is downlink carrier wavelength.

(5) Quantization error in carrier loop velocity measuring

Directly extract Doppler frequency from the receiver digital carrier loop, smooth it, and then work out velocity-measuring error with the computer. In this case, there are no $\sigma_{\dot{R}_4}$ and $\sigma_{\dot{R}_5}$ but quantization error $\sigma_{\dot{R}_6}$:

$$\sigma_{\dot{R}_6} = \frac{1}{\sqrt{6}} \frac{\dot{R}}{N} \quad (2.147)$$

where $\dot{R}$ is flying target velocity and N is digital filter word length.

(6) The velocity-measuring error introduced by VCO short-term frequency instability

The velocity-measuring error caused by VCO short-term frequency instability in the coherent transponder is:

$$\sigma_{\dot{R}_7} = \frac{\lambda_i \sigma_v}{4\pi\tau} \quad (2.148)$$

where σ_v is phase jitter caused by VCO short-term instability.

Total velocity-measuring error is:

$$\sigma_R = \sqrt{\sigma_{R_1}^2 + \sigma_{R_2}^2 + \sigma_{R_3}^2 + \sigma_{R_4}^2 + \sigma_{R_5}^2 + \sigma_{R_6}^2 + \sigma_{R_7}^2 + \sigma_{R_8}^2} \quad (2.149)$$

2.3.7.2 Velocity-Measuring System Error

(1) The error introduced by long-term frequency instability of the master oscillator

Find the differential of velocity-measuring formula with respect to f_i to obtain the caused velocity-measuring error:

$$\Delta V = \frac{Cf_d}{2f_i^2} \Delta f_i = \frac{Cf_d}{2f_i} \frac{\Delta f_i}{f_i} = V \times \frac{\Delta f_i}{f_i} \quad (2.150)$$

The error is equal to the product of the measured velocity value and the relative long-term instability of the master oscillator; long-term instability includes drift and accuracy error. When $V = 10$ km/s and long-term instability $= 1 \times 10^{-9}$, the caused velocity-measuring error is just 10^{-5} m/S. It is obvious that velocity-measuring system error complies to relevant requirement easily.

(2) Dynamic error

Velocity-measuring dynamic error (frequency error, unit rad/s) of different types of phase-locked loops can be calculated with Table 2.8 [9].

(3) The velocity-measuring error introduced by light speed ambiguity

The velocity-measuring error introduced by light speed ambiguity is

$$\sigma_{R_8} = 3.33 \times 10^{-7} \dot{R} \quad (2.151)$$

Table 2.8 Dynamic error of carrier Doppler velocity measuring

Loop	Constant range rate	Constant range rate derivative	Second derivative of constant range rate
	Constant Doppler frequency drift	Constant Doppler rate	Constant Doppler acceleration
Type II loop with standard underdamping	0	0	$\left(\frac{9\pi\beta}{16B_L^2}\right)$
Type II loop with super-critical underdamping	0	0	$\left(\frac{25\pi\beta}{32B_L^2}\right)$
Type III loop with standard underdamping	0	0	0
Type III loop with super-critical underdamping	0	0	0

Note: B_L is equivalent loop bandwidth of carrier loop (single-side) noises; when continuous Doppler acceleration appears, Type II loop periodically carries out cycle skip; β is Doppler acceleration (Hz/s²)

2.4 Ranging Techniques Theory and Technology: Two-Way, Three-Way, and Single-Way Ranging Technologies

2.4.1 Continuous Wave Two-Way Ranging Methods

Continuous wave ranging means certain forms of ranging signals are modulated on a continuous carrier to acquire the range information by comparing time delay between transmitted and received ranging signals. The above-mentioned certain forms of ranging signals refer to signals with particular time marks (such as the zero crossing point (ZCP) of sine waves) or phase marks. The characteristics of these continuous wave ranging signals are that there exists phase ambiguity, so the range de-ambiguity is the most special problem for continuous wave ranging. At present, there are mainly three types of continuous wave ranging signals: pure tone signal (single-frequency sine signal or square signal), pseudorandom code signal (for short, pseudo code signal, abbrev: PN code), and tone and code-mixed signal (the mixed signal of pure tone signal with pseudorandom code signal). Besides, based on the similar principles, there exist other ranging methods such as spread spectrum ranging, carrier ranging, and Doppler accumulative ranging. This section mainly introduces the first three ranging methods. Spread spectrum ranging will be described in the section of spread spectrum TT&C, Doppler accumulative ranging will be explained along with range acquisition, and carrier ranging has been applied in GPS, for which readers can refer to relevant materials.

2.4.1.1 Tone Ranging System

Tone means pure sine wave signal modulated on a carrier. It is also called side tones since its frequency spectrum is located at both sides of the carrier. The relationship between range R and tone time delay T and its phase shift Φ can be expressed as follows:

$$R = \frac{cT}{2} = \frac{c\varphi}{2\omega} = \frac{\Phi c}{720F} \quad (2.152)$$

where ω is the angular frequency of the ranging signal (rad/s), F is frequency of the ranging signal (Hz), Φ is phase shift in unit of degree ($^\circ$), φ is phase shift in unit of radian (rad), and c is light speed (m/s).

Therefore, time delay T can be obtained by measuring the phase shift. This method is called phase measuring method. Another method is to directly measure ZCP time delay of the phase between the transmitted and received tones. This is called time measuring method. It is clear from Expression (2.152) that if a certain phase measuring error $\Delta\Phi$ exists, the higher ω is, the smaller the time-delay error will be. Thus the ranging accuracy can be improved by using high frequency

ranging signal. As the phase of a sine periodic signal circulates by 2π , the real phase shift = the measured sine wave phase shift + $2n\pi$, where $n = 0, 1, 2, \dots$. This leads to multi-valuedness of the real phase of sine waves. This phenomenon is called phase ambiguity, the corresponding range of which produces range ambiguity. One solution to this problem is to constitute a set of tone signals with multiple pure sine wave signals in different frequencies. The set of tone signals are used to measure the same range simultaneously. Use the signal $U_2(t)$ in a lower frequency F_2 to solve the phase ambiguity produced by signal $U_1(t)$ in a higher frequency F_1 and make F_1 be K_1 times as large as F_2 , namely, $F_1 = K_1 F_2$, where K_1 is called the frequency ratio of F_1 to F_2 , also called matching ratio.

The most prominent feature of continuous wave ranging is the existence of range ambiguity; therefore, it is comparatively complicated to be realized. At present there are many ways for solving ambiguity, the main of which are listed as follows:

(1) Simultaneous multitone method

Its principle is shown as in Fig. 2.38 (where a uniform motion target is taken as an example).

In Fig. 2.38, the ordinate indicates the phase difference Φ between received/transmitted tones; the abscissa indicates the corresponding time delay t of target in range R ; the solid line Φ_1 represents the phase change curve of F_1 ; the dotted line Φ_2 is the phase change curve of F_2 ; F_1 is the higher tone, and F_2 is the lower tone. In this example, $K_1 = (F_1/F_2) = 4$, it is obvious from the figure that when the target starts to move from zero range, the echo delay time t increases gradually, so does Φ_1 of F_1 ; when $t = 1/F_1$, Φ reaches to 360° . Because of sine wave periodicity, at the next moment, Φ_1 jumps to 0° from 360° . If this 0° is used for ranging, its corresponding range will be 0. Apparently, it is wrong. This phenomenon is called

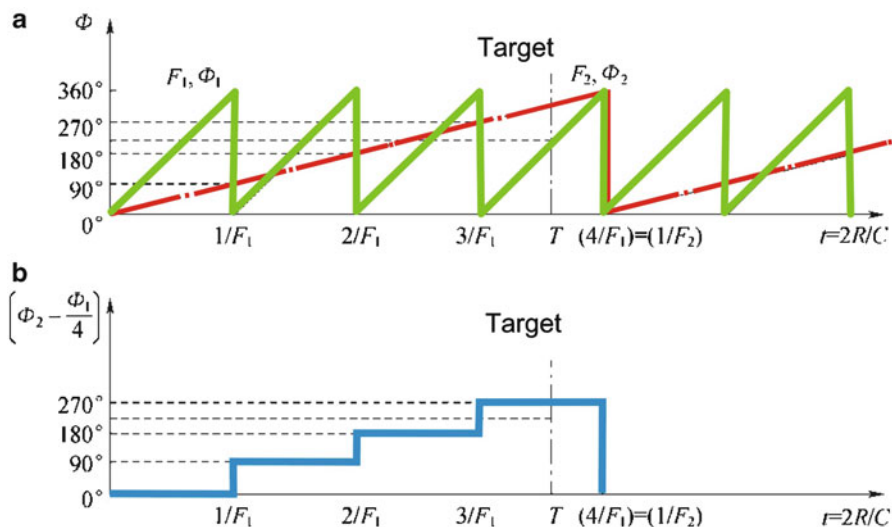


Fig. 2.38 Principle of multitone de-ambiguity

the appearance of range ambiguity. When the target continues to move, at time $t = 2/F_1$, the second step from 360° to 0° occurs. Its ambiguity number of times $n = 2$; when the target moves to time $t = 3/F_3$, a third ambiguity with $n = 3$ will happen. . . . When the target echo is at time T , the measured phase of F_1 tone is Φ_1 . However, as it has already generated three times of ambiguity ($3 \times 360^\circ + \Phi_1$), and in the foresaid example $n = 3$, thus when continuous wave ranging is underway, apart from “odd item” Φ_1 , n is also needed to be worked out, which is the number of range space corresponding to $\Delta R = c/2F_1$ and “to solve n ” is called de-ambiguity.

For de-ambiguity, a secondary tone F_2 , which frequency is lower than that of high tone F_1 , is needed. When F_1 has produced three times of ambiguity, the phase Φ_2 of F_2 remains in the m -th period at a range of 0° – 360° . Thus, the following criteria can be used for de-ambiguity:

$$\begin{cases} \text{when } 0^\circ < \Phi_2 < 90^\circ, & n = 0 \\ \text{when } 90^\circ < \Phi_2 < 180^\circ, & n = 1 \\ \text{when } 180^\circ < \Phi_2 < 270^\circ, & n = 2 \\ \text{when } 270^\circ < \Phi_2 < 306^\circ, & n = 3 \end{cases} \quad (2.153)$$

In this way, the phase value of the secondary tone can be used to solve the phase ambiguity of the high tone, which is called “using F_2 to solve R ambiguity.” At this time, the total phase shift $\Phi_\Sigma = n \times 360^\circ + \Phi_1$. This is the basic principle of range de-ambiguity. When the target flies further, making F_2 produce new phase ambiguity, it is feasible to use tone F_3 whose frequency is lower than that of F_2 to solve F_2 ambiguity. In the same manner, multiple secondary tones can be used to solve further range ambiguity. A priori information of the orbit can be used to reduce numbers of secondary tones, so that de-ambiguity is to be conducted only on a section of orbit, simplifying the equipment.

The foresaid basic principle of de-ambiguity is theoretically tenable, but in engineering implementation, at the boundaries (e.g., 0° , 90° , 180° . and 270°) of above-mentioned criteria, random error (e.g., phase jitter) and system error (e.g., group delay time error) of Φ_2 may cause erroneous judgment. For example, assuming actual $\Phi_2 = 179^\circ$ without above-mentioned error, then, according to the criteria, the corresponding $n = 1$. But if the error exists, making $\Phi_2 = 181^\circ$, then the judgment will be $n = 2$, producing range jump of $\Delta R = C/2F_1$, which is called “mismatch error.” The following criteria can be adopted to reduce such influence:

$$\begin{cases} \text{when } -45^\circ < (\Phi_2 - \Phi_1/4) < +45^\circ, & n = 0 \\ \text{when } 45^\circ < (\Phi_2 - \Phi_1/4) < 135^\circ, & n = 1 \\ \text{when } 135^\circ < (\Phi_2 - \Phi_1/4) < 225^\circ, & n = 2 \\ \text{when } 225^\circ < (\Phi_2 - \Phi_1/4) < 360^\circ, & n = 3 \end{cases} \quad (2.154)$$

The $(\Phi_2 - \Phi_1/4)$ curve of these criteria is shown as in Fig. 2.38b, the judgment of n mainly bases on this step trait. It is clear from this figure that corresponding to a period of Φ_1 without ambiguity, judgment value $(\Phi_2 - \Phi_1/4)$ is a constant value with an error tolerance of $\pm 45^\circ$. Therefore, if each kind of phase error (mainly

caused by inconsistent group delay of each tone and by time-delay variance of each tone after demodulation) does not exceed $\pm 45^\circ$, the ranging value will not be affected by mismatch error. Thus the tolerance becomes much looser. To make plus-minus error share the same margin to maximize error tolerance, in equipment commissioning, mean value of $(\Phi_2 - \Phi_1/4)$ shall be adjusted to 0° , 90° , 180° , and 270° . This procedure is called “phase matching” adjustment of Φ_1 and Φ_2 , also named as “range matching” adjustment among ranging tones (in broad terms, “phase matching” means the phase of each tone shares certain interrelation, e.g., align their zero phase points or align the last tone phase to the tolerance middle point of the next tone phase). In order to simplify engineering implementation, change the criteria of Expression (2.154) as follows: use 90° (judgment tolerance range $\Delta\Phi$) to divide the above criterion expression to obtain:

① When $-0.5 < ((\Phi_2 - \Phi_1/4)/\Delta\Phi) < +0.5$, $n = 0$. Round the calculated value of this criterion expression to obtain the value of n (here $n = 0$), and the tolerance zone middle point is 0.0.

② When $0.5 < ((\Phi_2 - \Phi_1/4)/\Delta\Phi) < 1.5$, $n = 1$. Round the calculated value of this criterion expression to obtain the value of n (here $n = 1$), and the tolerance zone middle point is 1.0.

③ When $1.5 < ((\Phi_2 - \Phi_1/4)/\Delta\Phi) < 2.5$, $n = 2$. Round the calculated value of this criterion expression to obtain the value of n (here $n = 2$), and the tolerance zone middle point is 2.0.

④ When $2.5 < ((\Phi_2 - \Phi_1/4)/\Delta\Phi) < 3.5$, $n = 3$. Round the calculated value of this criterion expression to obtain the value of n (here $n = 3$), and the tolerance zone middle point is 3.0.

According to the above expression, the range matching adjustment procedure is that under static state and in the case of strong S or large Φ , ① measure the values of Φ_2 and Φ_1 ; ② work out the value of $(\Phi_2 - \Phi_1/4)/\Delta\Phi$ and adjust the phase Φ_2 of the secondary tone F_2 (or Φ_1 of the main tone F_1) to make the digits of the mean value after the decimal point get close to “0,” namely, “XX.00...”; and ③ round the value gained in the last step to obtain the “large number” n (or add “0.5” to the value of $(\Phi_2 - \Phi_1/4)/\Delta\Phi$ and then “round” it).

The above example is for the case of $n = 3$, namely, there are three times of ambiguity; in the same manner, the general condition where $n = N$ can be obtained. Under this condition, there is:

$$F_2 = \frac{F_1}{(N+1)} = \frac{F_1}{K_1} \quad (2.155)$$

where K_1 is matching ratio and $K_1 = N + 1$.

$$\Delta\Phi = \frac{360^\circ}{(N+1)} = \frac{360^\circ}{K_1} \quad (2.156)$$

The general expression of the judgment expression is:

$$\frac{(\Phi_2 - \Phi_1/K_1)}{360^\circ/K_1} = \frac{K_1\Phi_2 - \Phi_1}{360^\circ} \quad (2.157)$$

After the above-mentioned phase matching, $(K_1\Phi_2 - \Phi_1)/360^\circ$ shall be at the middle point of tolerance range $(360^\circ/K_1)$, and its margin at plus and minus directions is symmetrical. In this case, the probability of matching error caused by a normal noise is:

$$P(\Phi_n > \pi/K_1) = 1 - \operatorname{erf}\left(\frac{\pi}{K_1} \sqrt{\left(\frac{S}{N}\right)_L}\right) \quad (2.158)$$

where Φ_n is tone phase error (rad) generally determined by Φ_2 , $(S/N)_L$ is tone loop SNR, and erf is error function.

Expression (2.158) also determines the probability of “mismatch error.”

From Expression (2.158), generally speaking, the calculation expression of the large number n is:

$$n = 1NT \left[\frac{K_1\Phi_2 - \Phi_1}{360^\circ} + 0.5 \right] \quad (2.159)$$

where $1NT$ is “rounding” operation and K_1 is matching ratio.

The total phase shift is:

$$\Phi = n \times 360^\circ + \Phi_1 \quad (2.160)$$

The corresponding range value is:

$$R = \frac{\Phi C}{720F} \quad (2.161)$$

If there are multiple secondary tones used to solve range ambiguity, the method can be employed in like manner. Total phase of F_1 is:

$$\Phi = (n_m K_{m-1} K_{m-2} \cdots K_2 + n_{m-1} K_{m-2} \cdots K_2 + \cdots + n_2 K_2 + n_1) 360^\circ + \Phi_1 \quad (2.162)$$

where $K_2 = F_3/F_2$, $K_m = F_{m+1}/F_m$ is matching ratio; $F_1, F_2, \dots$ are secondary tones arranged in order from high frequency to low frequency; generally, F_1 is the main tone which determines range accuracy; $n_1, n_2, \dots, n_m$ are the numbers of complete cycles contained by each secondary tones, among which, n_1 is the number of

complete cycles of F_1 contained by F_2 , and n_m is the number of complete cycles of F_m contained by F_{m+1} .

The selection of the highest fine measuring frequency F_1 depends on the requirement for ranging accuracy. The stricter the requirement is, the higher the frequency shall be:

$$F_1 \geq \frac{C}{\left(4\pi\sigma_R\sqrt{P/N}\right)} \quad (2.163)$$

where c is light speed, σ_R is required ranging accuracy, and S/N is signal SNR.

The selection of the lowest coarse measuring frequency depends on the requirement for the largest nonambiguous range; namely, the further the range is, the lower the coarse measuring frequency F_L shall be:

$$F_L \leq \frac{c}{R_{\max}} \quad (2.164)$$

where c is light speed and $R_{\max}$ is the largest nonambiguous range (the sum of a two-way range).

When the nonambiguous range is required to be comparatively far, F_L is pretty low. The tone spectrum after carrier which is directly modulated by this low frequency tone will get very close to carrier spectrum. This will not only affect the normal tracking for carrier but also make it difficult to demodulate the tone signals with such a low frequency. Therefore, a folded tone scheme is commonly adopted for solving this problem, that is, to conduct amplitude modulation to a certain secondary tone, then use its upper sideband-folded spectrum to modulate a carrier. Taking a unified S-band TT&C system as an example, besides an 100 kHz main tone and a 20 kHz secondary tone, 4 and 20 kHz are folded to generate 16 kHz folded tone, and then the 16 kHz tone is folded with an 800 Hz tone in turn to generate 16.8, 16.16, 16.032, and 16.08 kHz folded tones, which are used together with the main tone to collectively perform phase modulation to the carrier.

The foresaid basic principle for de-ambiguity is based on simultaneously transmitting a group of tones and measuring n and fine measured phase Φ_1 (corresponding to n and Φ_1 at the same range) at the same sampling time. So this method is called “simultaneous multitone method.”

(2) Large Number Accumulation Method

From Fig. 2.38, it is also clear that each time when F_1 produces an ambiguity, F_1 's phase Φ_1 generates a $360^\circ \rightarrow 0^\circ$ step, making relative n be added by “1.” Therefore, phase change of Φ_1 can be detected by making the large number n added by “1” (in reverse motion direction, subtracted by “1”) each time when phase step occurs. This is called the “large number accumulation method.” It is mainly used under a condition that range has been captured and initial range R_0 has been obtained, for which the large number n in subsequent range changes can be acquired in this method. As detection of Φ_1 is conducted between two sampling points,

assuming sampling time is τ and target radial motion velocity is $v(t)$, the moving distance in time τ is $\tau v(t)$ corresponding phase change is $2\tau v(t) \times 360^\circ/c$, the resulting number of phase steps producing phase ambiguity is decreased to $[360^\circ - 2\tau v F_1 \times 360^\circ/c]$, and the phase difference $\Delta\Phi$ between two sampling points shall be within the following scope:

$$\frac{(2\tau v F_1 \times 360^\circ)}{c} < \Delta\Phi < \left[360^\circ - \frac{(2\tau v F_1 \times 360^\circ)}{c} \right] \quad (2.165)$$

so as to make a judgment that a phase ambiguity has occurred. But a proper threshold value Φ_p should be chosen within this scope. Considering that $\Delta\Phi$ may be with plus and minus error, if the error influence makes $\Delta\Phi$ go beyond the foresaid boundary scope values, wrong n added by “1” (or n decreased by 1) will happen, leading to mismatch error of the ranging data. Therefore, it is optimal to place threshold Φ_p in the middle of abovesaid scope, making both sides have the same error margin. Thus, there is

$$\Phi_p = \frac{[360^\circ - (2\tau v F_1 \times 360^\circ)/C] + [(2\tau v F_1 \times 360^\circ)/C]}{2} = 180^\circ \quad (2.166)$$

Therefore, when $|\Delta\Phi| > 180^\circ$, a time of phase ambiguity can be judged to have occurred and the large number n shall be added by “1” (or subtracted by 1).

However, if Expression (2.165) cannot be satisfied, the “large number accumulation method” shall not be applied. For instance, if the frequency of F_1 is too high (or target velocity is too fast, or sampling time τ is too long), it is possible that Expression (2.165) may become false. In this case, it may make $(2\tau v F_1 \times 360^\circ) > 180^\circ$ which will lead to a wrong judgment. Moreover, this method also has following disadvantages:

1) As large numbers n and phase Φ_1 are not measured every time when range values are read and the “large number is accumulated,” the happening of ranging mismatch error may not get aware of, accumulation may be going on, which cannot be found and corrected in a timely manner. But the above-mentioned “simultaneous multitone method” is free from this disadvantage.

2) As “large number accumulation” is a sort of open-loop measuring technique, it does not have automatic adjustment and correction functions as the closed-loop mode does.

3) When this method is adopted, the range capture scheme for acquiring initial range is usually different from the subsequent scheme for range tracking. Connection between both schemes is required, which is apt to introduce ranging error (e.g., time is not aligned), resulting in different ranging values to appear at each capture.

4) Wrong $\Delta\Phi$ measurement will bring about ranging mismatch error, and the random mistake in $\Delta\Phi$ measurement may result from several influences, such as dropped frames of ranging data, time difference caused by inaccurate timing, cycle slip of F_1 phase-locked loop, interference, and strong power supply fluctuation.

However, this scheme also has its advantages, which include simple equipment and less power consumption due to the fact that only fine tone is transmitted in ranging process.

(3) Time-sharing multitone method

In the above concurrent multitone method, it requires measuring the receiving/transmission phase shift Φ at the same range (i.e., at the same time) to obtain corresponding n and calculate the true range. Such multitone at the same time may occupy some power, and each tone needs simultaneous phase measurement, which causes complexity to equipment. Time-sharing multitone method is employed in engineering, and only one tone is transmitted for each time. The following two schemes are usually adopted.

1) Signal replication method

Ranging by signal replication method is to replicate the received tone signal by frequency division chain under high-tone driving, and then conduct non-ambiguity ranging with the lowest tone. Its working principle is as follows.

Under frequency scale synchronization, a tone generator produces tone frequency and transmits it from high to low in a time-sharing mode. However, the high tone is constantly transmitted and initial phase of low tone can be controlled and regulated. IF phase-modulated signals and reference signals from a receiver shall be firstly subject to coherent demodulation so as to obtain basic low tones in a time-sharing mode. High tone is purified by phase-locked loop and low tone is replicated successively by a frequency demultiplier. The tone generator compares the phase difference of each received and replicated secondary tone, regulates state of frequency division chain at each level to make each pair of tones in the same phase. In this way, the lowest secondary tone output by the frequency division chain is the replication of the received lowest tone and its time delay is equal to transmission delay of radio wave. When a range counter is formed through filling of a high-stability frequency clock and is enabled from the lowest transmitted tone and disabled at the lowest tone of frequency division chain, its output data will be the target delay of frontier moment for transmitting lowest tone. Sinusoidal function generator is employed but not frequency division chain method for transmission of primary tone, which can greatly increase the bandwidth of primary tone filter, improve delay stability and introduce no ranging drift error. Due to effect of dispersion from channel transmission, different tones create different delays. Initial phase of the transmitted tones shall be regulated to make the received tones aligned at the zero point so as to avoid frequency division chain capturing high error. Such adjustment is accomplished in system joint test.

Because the frequency division chain replicates and records all tones and range change is also recorded through driving of received primary tone to the frequency division chain, ambiguity solution for each tone at the same range point can be realized even though tones are transmitted in a time-sharing mode when the aforesaid method is adopted.

The nearest frontier alignment method is usually employed for phase matching zero method of system tones, which regulates transmitted tone phase on the basis of

received frequency division chain to make positive direction of received tones aligned crossing the zero point. The specific process is as follows.

Firstly set the initial phase of transmitted tones as zero; frequency division chain is under random state after primary tone is Loop-locked. The correlator measures the phase difference $\phi_i - \phi_{0i}$ between echo tone and frequency division chain, which will be taken as neighboring frequency ratio after $(360^\circ/a)$ modulus and remainder $\Delta\phi_i, a$, then

$$\Delta\phi_i = (\phi - \phi_{0i}) \bmod [360^\circ/a] (^\circ) \quad (2.167)$$

After regulating the transmitted tone phase with remainder $\Delta\phi_i$, received tone zero point can be aligned. Such zero method is to align the lower tone to the nearest positive direction zero point, i.e., near matching. As the equipment does not eliminate the additional phase shift of high tone, the system may have high zero-range set constant (e.g., several kilometers). Maybe it is strange in theory, the second step of matching zero method is to eliminate high zero-range set constant. Its method is to calculate and determine which low tones will be given high phase shift from the system according to current zero-range set constant, change the initial phase of several transmitted tones through keyboard (change value is one or several phase interval $(360^\circ/a)$), and move forward the received low tone across the zero point to decrease zero-range set constant.

Feature of such signal replication method: use received primary tone to promote frequency divider and make phase matching by correlational method to replicate lowest tone signals. These lowest tone signals follow primary tone signals for successive tracking and possess the accuracy of primary tone phase as well as period of lowest tones, thereby realizing non-ambiguity and high-accuracy ranging at a long distance.

Its defect: Constant transmission of primary tone requires transmission of two tones at each time. Although capture time may be shorter, transmission energy is bigger. One possible improvement method is to correlate transmission tone with transmission carrier. If correlation ratio of primary tone and carrier is m_1 , carrier correlation transmission ratio of transponder is m_2 , then the correlation ratio of primary tone and received carrier is $m_1 \times m_2$, conduct $m_1 \times m_2$ time of frequency division of received carrier to make signal replication and ranging according to above method after received primary tone. But phase ambiguity from $m_1 \times m_2$ frequency division can be solved by real-time measurement on the phase difference of frequency division primary tone and received primary tone. Apply it into range calculation formula to obtain no-fuzzy range and the accuracy of high tone is determined by carrier C/N_0 . The phase difference of high tone ambiguity solution can be decreased by increasing integral time to obtain high ranging accuracy. Such method can make time-sharing transmission of sequence tone, but memory and driving of tone is replaced by carrier frequency division.

2) Phase measurement calculation method

Such method makes time-sharing measurement of receiving/transmission phase shift of each tone and calculates ambiguity solution to obtain range value.

For example, transmitted tones are folded tones of 100, 20, 16, 16.8, 16.16, 16.032 Hz, and 16.008 kHz. No need to restore 16.008 kHz to 8 Hz at receiving end. We can directly measure 16.008 kHz phase shift, minus 16 kHz phase shift to obtain 8 Hz phase shift:

Sum frequency phase shift: $\Phi_{\Sigma n} = (\omega_B + \omega_n)\tau = \omega_B\tau + \omega_n\tau$

Base frequency phase shift: $\Phi_B = \omega_B\tau$

Secondary tone f_n phase shift: $\Phi_n = \Phi_{\Sigma n} - \Phi_B = \omega_n\tau$

It can directly measure phase shift at sum frequency, which is one advantage of tone phase calculation method.

Range finder transmits one tone for each time under time-sharing mode. After lowest tone of phase loop is locked, it transmits from low to high and finally accomplishes locking of primary tone. In case of digital signal processing, phase measurement can be done through digital phase-locked loop. Tone generator produces primary tone and various secondary tone waveforms, the order of which are sent to uplink modulator for carrier modulation.

Tone digital signal from receiver is sent to tone extracting loop for tone phase extraction and measurement of receiving/transmission phase difference. Tone frequency is produced and extracted from low to high and the receiving/transmission phase shift measurement of such tone is done after indication of tone extraction loop-lock is stable, phase ambiguity resolution is done in order. Receiving/transmission phase difference is sent to data processing computer for calculation of tone phase ambiguity resolution and calculation of target range. Time scale of ranging data is the time of tone receiving end.

Capture time of ranging depends on the locking time of tone loop. Wideband loop capture and narrowband loop tracking is employed to shorten capture time and ensure ranging accuracy. Tone loop employs Doppler compensation, which guides rapid locking of tone loop by use of Doppler information from R&C unit.

Range capture procedures are as follows:

① Process of tone transmission. In case of range capture, tone transmission is done in order and there is only one tone at any time. For transmission of one tone, wait for loop-locking of received tone and delay for some time; make R/T phase shift measurement after loop DCO phase is stable.

Firstly transmit the reference frequency of combined tone $F_1 = 16$ kHz, then transmit tone frequency from low to high one by one, so as to calculate the secondary tone phase shift from combined tone phase shift and provide phase ambiguity resolution in order.

② Phase calibration. Calculate and store the R/T phase zero of R/T equipment channel to tones, i.e.,

$$\theta_{0i} = \theta_{\text{Trans}0i} - \theta_{\text{Reciev}0i}, \quad i = 1 \sim 7 \quad (2.168)$$

where $\theta_{\text{Trans}0i}$ is the sample value of tone phase of transmitted tone generator in case of phase calibration; $\theta_{\text{Reciev}0i}$ is the sample value of tone phase of received tone loop DCO in case of phase calibration.

In case of range capture on target, tone R/T phase shift caused by target range delay is:

$$\varphi_i = \theta_i - \theta_{0i}, \quad i = 1 \sim 7$$

where

$$\theta_i = \theta_{\text{Transi}} - \theta_{\text{Recievi}}, \quad i = 1 \sim 7$$

③ Transmit 7 tones in order and record R/T phase shift at corresponding moments of $t_1 \sim t_7$. For ambiguity solution at the same range point, conduct range matching calculation after obtaining phase shift value of 7 tones at t_7 . Phaseshift increment $\Delta\varphi_{di}$ from target movement in time periods of t_1-t_7 shall be calculated.

$$\Delta\phi_{di} = \int_{t_i}^{t_7} f_d(t)dt \quad (2.169)$$

where $f_d(t)$ refers to Doppler frequency.

Expression (2.146) is called Doppler integration, which calculates $\Delta\phi_{di}$ and corresponding range variation value. Specific implementation methods include Doppler frequency single- point sampling method, multi-point average method, fragment accumulation method, in which fragment accumulation method has higher accuracy.

R/T phase shift of combined tone at the moment of 6 is:

$$\Phi_{\Sigma i} = \phi_i + \Delta\phi_{di}, \quad i = 1 \sim 7$$

④ Calculate the R/T phase shifts of basic tones.

R/T phase shift is $\Phi_1 = \Phi_{\Sigma 1} - \phi_1$ for $f_1 = 8$ Hz; R/T phase shift is $\Phi_2 = \Phi_{\Sigma 3} - \phi_1$ for $f_2 = 32$ Hz; R/T phase shift is $\Phi_3 = \Phi_{\Sigma 3} - \phi_1$ for $f_3 = 160$ Hz; R/T phase shift is $\Phi_4 = \Phi_{\Sigma 4} - \phi_1$ for $f_4 = 800$ Hz; R/T phase shift is $\Phi_5 = \Phi_{\Sigma 5} - \phi_1$ for $f_5 = 4$ kHz; R/T phase shift is $\Phi_6 = \phi_6$ for $f_6 = 20$ kHz; R/T phase shift is $\Phi_7 = \phi_7$ for $f_7 = 100$ kHz.

Feature of phase measurement calculation method: discrete capturing of phase and Doppler data of each tone in order. As ambiguity solution calculation shall be done at the same time, Doppler integration method is employed to calculate the phase of each tone to arrival time of final tone, and then conduct ambiguity solution calculation, which causes following problems.

① Doppler integration is impacted by measurement accuracy, sampling interval, integration time and integration algorithm, there is error for such phase calculation mode. When time is longer, speed and acceleration is higher, its error is bigger, which may cause matching mistake and mismatch error for range capture. There is serious defect for ranging of long range of deep space, etc. and high-speed target.

② Its phase measurement is accomplished by phase-locked loop. As there is such special problem as capture time, capture bandwidth, cycle slip for

phase-locked loop, it may cause phase measurement error and matching error when design is improper (capture time is not sufficient).

③ Time difference, interference, etc. from frame-falling, inaccurate time scale may cause capture mistake. Its advantage is to transmit one tone at each time.

Doppler integration ranging in Expression (2.169) is also one range determination method, which has two problems: firstly, starting value of range must be known; secondly, Doppler integration error, the longer the integration time lasts, the bigger the accumulation error becomes. It can calculate range increment of high accuracy and high resolution to short time period by use of velocity integration with known original value of range and be applied to following cases.

① When ranging time and data sampling time is different and there is difference of AT time interval, make Doppler integration in AT to obtain the range value at sampling time.

② When the time interval of ranging is long (e.g., PN code period is long) and there is high requirement for data sampling rate, Doppler integration ranging can be used for high resolution compensation between two ranging moments.

③ It can modify the propagation delay error of ionosphere and plasma.

2.4.1.2 Pseudorandom Code (PN Code) Ranging System

Tone ranging method needs to use multiple tone signals and it has a disadvantage that the phase de-ambiguity will get very complicated. There is another approach called PN code ranging which uses only one signal rather than multiple signals. It makes use of long period of the PN code to solve ambiguity problem and takes code clock as accurate measurement signal similar to high tone. This is the original intention when the PN code ranging was conceived. With the development of technology, now it has also been endowed with other functions such as anti-interception, anti-interference, spread spectrum and code division multiple access (CDMA). For ranging, phase of transmitted carrier is under 0 or π phase modulated by binary PN code and target reflected echo is received by the receiver, in which there is also a same PN code generator served for reproducing the received PN code. The local pseudo code is multiplied with the modulated IF signal of the received pseudo code in a multiplier, which equivalently performs a modulation of 0 or π phase shift on the input received signal. If local pseudo code is completely consistent with input modulation pseudo code in time, the occurrence order of 0 and 1 will be exactly the same (reproduced) and the original 0 or π modulation will get removed by such counter modulation. So the receiving multiplier will eliminate the pseudo code modulation on input carrier to make the input signal become a carrier, which then can be extracted by a narrow band-pass filter. If the local pseudo code is not aligned to the modulated PN code of input signal in time, the multiplier cannot remove the modulation on PN code and will add a new modulation, so signal carrier cannot be restored and the band-pass filter will have no carrier signal output. The multiplier and band-pass filter constitute the IF correlator of PN code modulation. Output signal of the band-pass filter depends on the time relationship between local

PN code and input PN code. Analysis shows that when the time difference of 2 PN codes is more than 1 code element, output is close to zero ($1/p$ of the maximum output, where p is the number of code elements in a pseudo code period, called code length); when time difference is zero, i.e., when two sequences are completely aligned, output is maximum; when time difference is within plus/minus 1 code element, correlation value changes with time difference. In one sequence period, correlation value has obvious output only in one code element width, having a form like a triangle. Other time output is zero.

Figure 2.39 is the diagram of autocorrelation function of an m -sequence pseudo code.

The shape of the correlation peak is like a triangle. Its hypotenuse shows that in one code element width the correlation value is linearly related with time, which can be used to constitute the time discriminator for PN code tracking loop. Such discrimination feature is as shown in Fig. 2.40. In Fig. 2.40, where Δ is code element width, the voltage level change caused by time change is used to control local PN code phase to make it synchronous to input PN code.

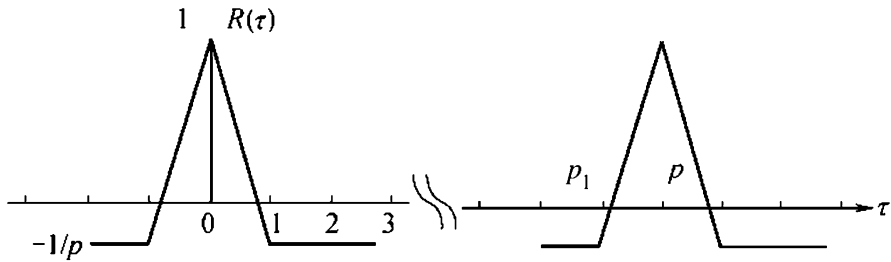


Fig. 2.39 Autocorrelation function of an m -sequence pseudo code

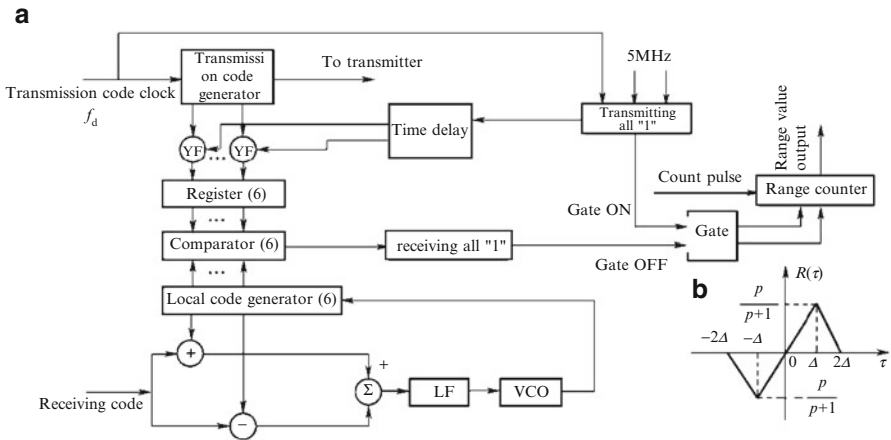


Fig. 2.40 Principle of pseudo code ranging

As the band-pass filter bandwidth of correlator is narrow and bandwidth of PN code loop is also narrow, it can filter much input noise. PN code tracking loop composed of correlator can work under low SNR environment and accomplish ranging task.

Its work process can be divided into following four steps.

(1) Use transmission end PN code to modulate carrier and transmit it to target. After sampling pulse for ranging arrives, its rising edge is used to put 0 and 1 states of all levels of shift register in transmission end PN code generator into the latch. Shift register states represent the phase of PN code sequence. At the same time, sampling pulse rising edge is used as transmitting “1” pulse to make range counter counts code clock.

(2) Produce a local code identical to transmitting code but with a certain delay to it at reception end, and make frequency and code sequence of the local code clock synchronous to receiving code, thus the local code can be served as the reproducing signal of receiving code. Track the delay change of input signal by PN code tracking loop. As both the initial states of local code and receiving code are random, a capture circuit is needed to roughly align local code and receiving code within one code element. Discriminator will have error output after the time difference is within the linear area of discriminator’s characteristic curve. After filtering, the error output is used to control VCO, making local code align with receiving code and so realizing synchronous tracking, i.e., composing a delay locked loop.

(3) After synchronization, the states of each level of shift register in local PN code generator, which represent the phase of local PN code sequence, are sent to a multi-digit comparator for comparing with latched states of the shift register in transmission end. With the time passing, when the multi-digit states of local code match with latched transmitting states, the comparator will output a matching pulse as receiving “1” pulse to make range counter stop counting. The time interval between two pulses of transmitting “1” and receiving “1” is the time delay of local code relative to transmission end, which represents transmission delay of target echo.

(4) The counting pulse number N of code clock of range register in one Gate ON – OFF period is coarse-range and fine range is determined by code clock phase Φ_0 . The range can be expressed as:

$$R = \frac{(N \cdot 360^\circ + \Phi)c}{720F} \quad (2.170)$$

where F is code clock frequency.

The pseudo code ranging can use either single code or compound code.

Compound code is composed of several single codes (also called subcode) of different code length in some logic combination. The code lengths of subcodes are better to be coprime (relatively prime) and closer with each other for the most. For example, in US “Apollo” TT&C system, ranging code was a compound code composed of five subcodes, whose code lengths were: $CL = 2$, $A = 11$, $B = 31$, $C = 63$, and $D = 127$.

As code lengths of subcodes are relatively prime, the code length of the compound code is the least common multiple of subcode code lengths (product of each subcode's code length), i.e.,

$$P = \prod_{i=1}^n p_i \quad (2.171)$$

For the “Apollo” example, $P = 2 \times 11 \times 31 \times 63 \times 127 = 5,456,682$

As can be seen, the code length of such compound code is long enough so that the corresponding maximum unambiguous range is farther than the distance to the Moon.

Capture time of compound code is the sum of individual capture time of each subcode, i.e., if each subcode captures, the whole compound code captures.

$$T_{\text{compound}} = \left(\sum_{i=1}^n p_i \right) \tau_0 \quad (2.172)$$

where p_i is code length of subcode; n is the number of subcodes composing the compound code; τ_0 is the capture time of each code element. The capture time of a single pseudo code with the same code length as the compound code is:

$$T_{\text{single}} = P\tau_0 = \left(\prod_{i=1}^n p_i \right) \tau_0 \quad (2.173)$$

As $T_{\text{single}} \geq T_{\text{compound}}$, in general pseudo code ranging system the signal form of single pseudo code is rarely used. The ranging accuracy of pseudo code depends on code element width (corresponding with code clock). The smaller the code element width is, the smaller the ranging error will be.

2.4.1.3 Hybrid Ranging System of Code and Tone

Pure tone ranging and pseudo code ranging have their respective advantages and disadvantages. Main advantages of pure tone ranging are that the high-accuracy ranging can be achieved by raising tone frequency, and its occupied bandwidth is rather narrow and capture is fast; but its phase de-ambiguity is complicated. Main advantage of pseudo code ranging is the long non-ambiguity range; but code element width must be decreased to raise accuracy, which increases occupied bandwidth, makes code capture more complicated, and takes longer time. Hybrid system of pseudo code and tone can make them have complementary advantages. Its basic idea is to use high tone for accuracy, use pseudo code for range de-ambiguity, and use pseudo code signal for PSK demodulation on high tone.

There are following three schemes for hybrid system of code and tone.

(1) Pseudo code + high tone

It is a common scheme. It makes the pseudo code be PSK modulated on high tone and selects a higher frequency for high tone to improve ranging accuracy. Its basic principle is the same as above-mentioned pseudo code modulated on carrier, and so it is not necessary for more details here.

(2) Spread spectrum code + high tone

For composing ranging signal, PN code is firstly used to spread the spectrum of the low rate de-ambiguity code (Barker code). The spread-spectrum signal is used to conduct BPSK modulation on 1 MHz high tone and then phase modulation on carrier. Carrier modulation sideband is used for ranging and residual carrier is used for velocity measurement. Spread-spectrum ranging signal has strong anti-interference capability and has CDMA capability in the case of multi-station working at the same time.

The ratio of PN code clock frequency to high tone frequency can be selected to be rather small, so it is easy to implement complete cycle de-ambiguity of the high tone by code clock phase and decrease the accuracy requirements on PN code clock phase.

De-ambiguity code for maximum range is a 13-bit Barker code, which is used for PN code phase de-ambiguity. If the non-ambiguity range is required to be farther, the 16-bit de-ambiguity code can be selected. De-ambiguity code shall have sharp self-correlation peak and low side-lobe.

By contrast, in some TT&C systems which adopt hybrid system of tone and code, when pseudo code uses multiple code, signal power will be occupied, causing signal power dispersion, reducing ranging accuracy and de-ambiguity capability, and capture time of multiple code structure is hard to be shortened.

In such scheme, Barker code is spectrum-spread by pseudo code, reception end accomplishes code tracking and recovery, high tone is directly modulated by pseudo code, which not only improves anti-interference capability, but also makes full use of signal energy. High tone, pseudo code, and Barker code do not separately occupy signal energy. Correlation receiver of de-ambiguity code uses the signal energy in 13 pseudo code cycles, which makes correlation peak for range de-ambiguity have higher output SNR to ensure no mismatch error occurs during ranging.

Meanwhile, single code modulation makes code capture process simple and code capture time can be shortened by semi-parallel processing of multiple correlators. The capture time of N correlators in parallel can be N times shorter than the capture time of N correlator in serial.

After PN code is captured, the arrival time of correlation peaks can be detected by de-ambiguity hardware of matching filters. Ranging process is simple and capture time is composed by two parts of PN code capture and tone loop-locking.

Its composition block diagram is as shown in Fig. 2.41. Ranging terminal is composed of ranging signal generator, ranging signal extractor, range delay measurement unit, ranging computer, etc. Ranging signal extractor is composed of PN code parallel detector, PN code tracking loop, high tone loop, and Barker code matching filter.

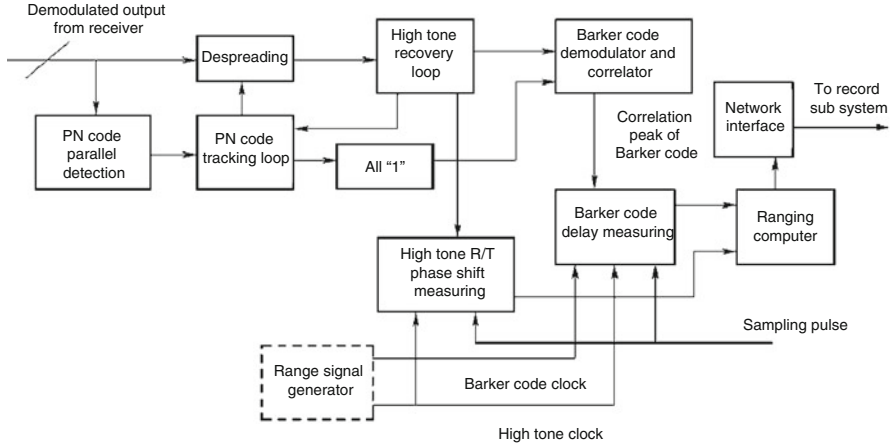


Fig. 2.41 Composition block diagram of ranging terminal

The uplink carrier modulated by ranging signal is returned to ground receiver by transponder. After demodulation there is much noise overlapped in the ranging signal and so filtering is needed to recover the pure ranging signal for measuring the R/T delay difference, which is sent to ranging computer to calculate target radial range and then sent to data record subsystem through network interface.

In measuring range through delay, both Barker code delay measurement and high tone R/T phase shift measurement are employed for facilitating measuring and obtaining high resolution. Barker code delay is used to measure coarse range (quantization unit is the cycle number of high tone) and R/T phase shift of high tone is used to measure fine range. PN code clock is used for phase de-ambiguity of high tone.

(3) Sequence code + high tone

The uplink ranging signal of such scheme includes high frequency tone f_r and sequence codes $r_n(t)$ from its subharmonic waves, which are phase-modulated on the tone. The resulting ranging signal can be expressed as:

$$s_{ur}(t) = \cos[2\pi f_r t + \Phi_r r_n(t)] \quad (2.174)$$

1) Ranging code

The de-ambiguity signal $r_n(t)$ is produced by subharmonic wave of the ranging tone, its period is:

$$T_n = \frac{2^n}{f_r} \quad 0 \leq n \leq 20 \quad (2.175)$$

where 2^n is code length and f_r is ranging signal frequency.

Code $r_n(t)$ is a bipolar (± 1) periodic rectangular function:

$$r_n(t) = Q_1 \oplus Q_2 \oplus Q_3 \oplus \cdots \oplus Q_n \quad (2.176)$$

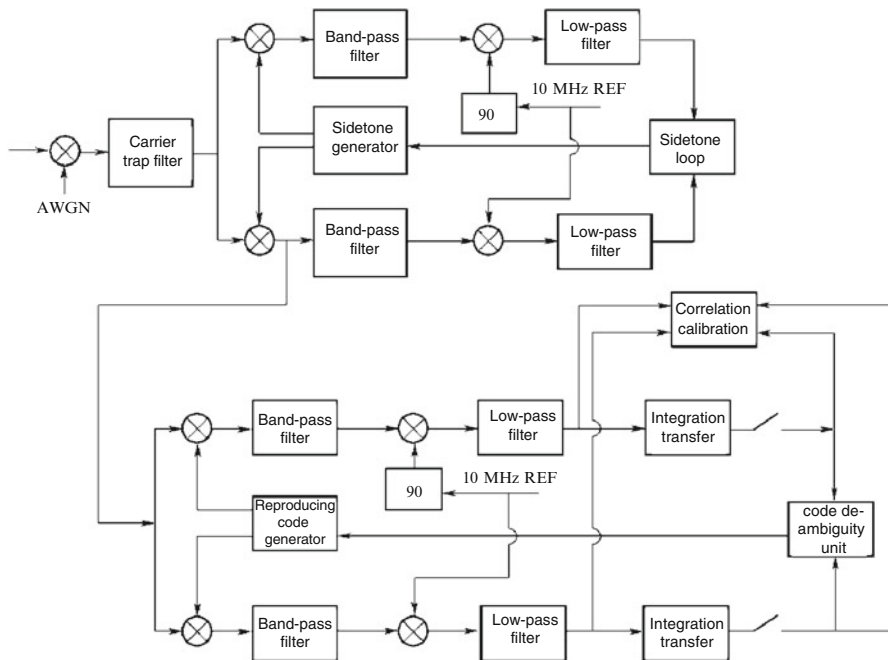


Fig. 2.42 Terminal block diagram

where Q_i can be the output of 1:2 frequency dividing trigger link. Symbol is used to represent sum of “Module 2.”

The first trigger of frequency dividing link is driven by the ranging tone. Code signal $r_n(t)$ can also be represented by a recursion expression:

$$r_n(t) = \begin{cases} r_{n-1}(t), & 0 \leq t \leq T_n/2 \\ -r_{n-1}(t - T_n/2), & T_n/2 < t \leq T_n \end{cases} \quad (2.177)$$

2) Working principle

Figure 2.42 shows a block diagram of IF correlation process. The upper branch is a PLL, while the lower branch is a ranging signal reproducer.

10 MHz IF signal from coherent receiver is sent to a narrowband trap filter for reducing the unwanted carrier signal in input end. Following the trap filter is a phase quadrature tone demodulator, which multiplies the input modulated IF carrier with in-phase and quadrature phase local tone signal. It is then multiplied by in-phase carrier component recovered by carrier loop and thus is converted to baseband.

The digital tone phase-locking loop filters filter out phase noise, the filtered signal is sent to control tone frequency. A frequency compensation generated by measured carrier Doppler is also sent to tone generator and makes such phase-locking loop work in narrow loop bandwidth. Send signal after quadrature tone component correlator to code correlator. In code correlator, signal is firstly multiplied with local I and Q reproducing codes and then converted to baseband after multiplied with recovered carrier.

3) Capture process

The capture process is similar to above “signal reproducing method” and starts from transmitting and receiving tone only. The phase-locked tone is received in a preset time interval. It is used to reproduce sequence codes and the first code is used to modulate the uplink tone. As the first code is synchronous with tone and its period is 4 times of tone period, correlator can find out 4 phase ambiguities of the 4 frequency divider and solve the ambiguities in-phase or out-off-phase with tone, which are used to adjust the state of frequency divider to make reproducing code in the same phase as the input code. Then the second code is used to replace the first code, which can make alternate judgment on ambiguity again. In order to reduce the necessary correlation times of range de-ambiguity, 4 frequency division step is employed for code cycle. Based on correlation feature of 2 quadrature codes, we can use 2 correlators to calculate the 4 phase ambiguities of the 4-frequency divider. At the end of capture, the longest reproducing code is the same as longest received code, which determines the maximum non-ambiguity range. Sequence code is also transmitted to detect whether there is ranging error such as mismatch error. As it is required only for testing, the needed energy is small. Modulation degree of sequence code on tone can be decreased.

4) Ranging

Uplink code generator produces “start pulse” and reduplication code generator produces “end pulse.” The time interval between start pulse and end pulse is measured by a high resolution timer/counter. Ranging accuracy is determined by high tone while non-ambiguity range is determined by longest code.

2.4.2 Vector Analysis Method of Ranging Error

Ranging accuracy is determined by sinusoidal wave signal of the higher frequency (e.g., high tone, code clock, subcarrier signal), so the accuracy of sinusoidal wave signal is the basis of analyzing CW ranging accuracy and it will get analyzed and calculated in this section. For other ranging signal derived from sinusoidal wave signal, such as hybrid signal of code and tone, its ranging spectrum is broadened to a frequency band, whose amplitude-/phase-frequency characteristics will introduce additional ranging system errors. In this case, the total system error can be treated as ranging error of a single sinusoidal wave attached with a loss factor. In case of pure PN code or spread-spectrum signal, the system error resulted from zero change of code loop time discriminator shall be considered.

Based on above reason, this section will focus on analyzing the ranging error of sinusoidal wave ranging signal. Ranging error includes system error and random error.

System error of equipment includes error caused by R/T channel delay change, range calibration residual error, dynamic error, multipath effect error, and transmission line delay error, among which the error resulted from delay is the fundamental component, especially for the transponder working in space, which may have a considerable delay change during its several years of life cycle. It is the main factor that restricts the increase of equipment accuracy.

As group delay is derived from the differential of phase-frequency characteristic, phase-frequency characteristic is of first importance in network characteristics. So this book will first analyze the effect of frequency characteristics on ranging accuracy.

(1) Frequency characteristics analysis methods of ranging error

Nowadays with the development of digital technology, the low-pass filter has been digitalized and its impact has become ignorable. So the following will only analyze the impact of band-pass frequency response on ranging error.

Tone delay resulted from channel phase-frequency characteristic, amplitude-frequency characteristic, and non-orthogonal product demodulation [11].

Channel phase-frequency characteristic and amplitude-frequency characteristic mentioned in this section refer to the frequency characteristics of any part in the channel (e.g., filter, modulator, AGC attenuation network), or frequency response introduced by any abnormal condition (e.g., effect of bad matching), or the total frequency characteristic, or asymmetry of signal spectrum line or phase shift caused by other factors.

Let the sinusoidal tone signal input to filter is:

$$B_i(t) = B_{im} \sin \Omega t \quad (2.178)$$

Carrier for tone phase modulation:

$$A_{ip}(t) = \sin[(\omega_0 t) + m \sin(\Omega t)] \quad (2.179)$$

where ω_0 is carrier frequency, Ω is tone frequency, m is phase modulation index.

For convenience sake, suppose the initial phases of carrier and tone both are 0, carrier amplitude $A_m = 1$, then Bessel expansion of Expression (2.179) is:

$$\begin{aligned} A_{ip}(t) = & J_0(m) \sin \omega_0 t + J_1(m) \sin(\omega_0 + \Omega)t - J_1(m) \sin(\omega_0 - \Omega)t \\ & + J_2(m) \sin(\omega_0 + 2\Omega)t + J_2(m) \sin(\omega_0 - 2\Omega)t + \cdots \end{aligned} \quad (2.180)$$

There are $(\omega_0 \pm n\Omega)$ side frequencies but we only take the first order side frequency for discuss. The reason is that in actual scheme, narrowband filter with center frequency at Ω is used after product demodulation, so only Ω component is extracted and other 2Ω , 3Ω , ..., $n\Omega$ components are all filtered. In Expression (2.180), only first order sideband component is the useful component. In case of small modulation coefficient, components above first order are rather small and their combined components produced after passing through non-linear system is even smaller, so taking first order component only into consideration is allowed by engineering design practice. When only first order component is taken, mathematic processing is simple and Expression (2.180) becomes

$$A_{ip}(t) = J_0(m) \sin \omega_0 t + J_1(m) \sin(\omega_0 + \Omega)t - J_1(m) \sin(\omega_0 - \Omega)t \quad (2.181)$$

After passing through a four-terminal network with amplitude-frequency characteristic $H(\omega)$ and phase-frequency characteristic $\Phi(\omega)$, such phase-modulated wave is to be product demodulated (i.e., sent to a multiplier after digitized). Phase demodulation model is as shown in Fig. 2.43. Corresponding $H(\omega)$ and $\varphi(\omega)$ characteristic curves are as shown in Fig. 2.44.

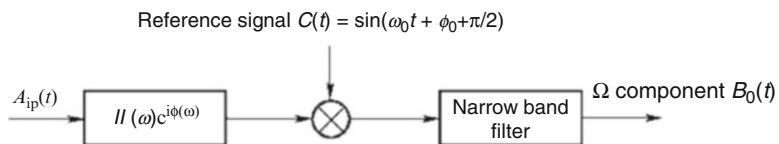


Fig. 2.43 Phase demodulation model

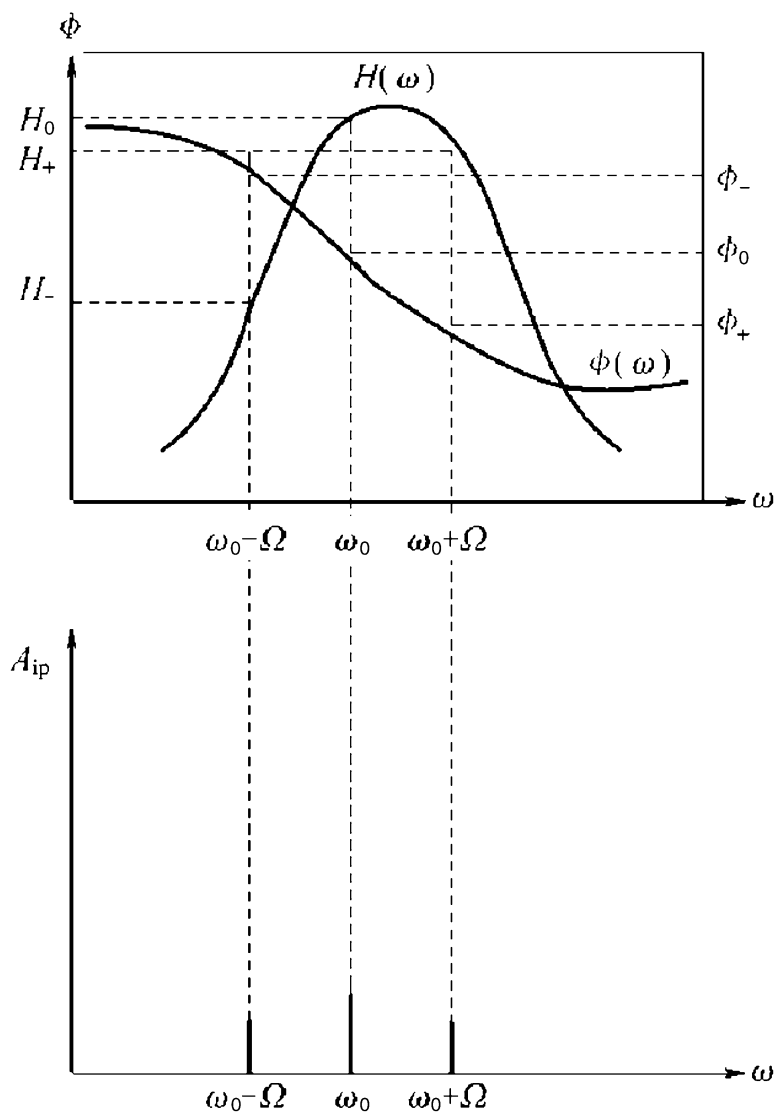


Fig. 2.44 Curves of amplitude-frequency characteristic and phase-frequency characteristic

Obviously, spectrum lines will not be symmetrical after being transmitted through the four-terminal network with amplitude-frequency characteristic and phase-frequency characteristic as shown in Fig. 2.44. Its output signal will be:

$$A_{\text{op}}(t) = H_0 J_0(m) \sin(\omega_0 t + \phi_0) + H_+ J_1(m) \sin[(\omega_0 + \Omega)t + \phi_+] H_- J_1(m) \sin[(\omega_0 - \Omega)t + \phi_-] \quad (2.182)$$

Product demodulation is to realize

$$A_{\text{op}}(t) \times C(t)$$

where $C(t)$ is the reference signal that is correlated to and orthogonal with residual carrier $H_0 J_0(m) \sin(\omega_0 t + \phi_0)$.

In case of no strict orthogonal intersection, there is non-orthogonal phase difference $\Delta\Phi$

$$C(t) = \sin\left[(\omega_0 t + \phi_0) + \frac{\pi}{2} + \Delta\phi\right] \quad (2.183)$$

where $\pi/2$ is orthogonal requirement, $\Delta\Phi$ is non-orthogonal phase difference, which is caused by static phase error, dynamic phase error, phase jitter, circuit phase shift, etc., of residual carrier phase-locked loop.

$$\begin{aligned} A_{\text{op}}(t) \times C(t) &= \frac{1}{2} H_0 J_0(m) \cos\left(-\frac{\pi}{2} - \Delta\phi\right) - \frac{1}{2} H_0 J_0(m) \cdot \\ &\cos\left[2\omega_0 t + 2\phi_0 + \frac{\pi}{2} + \Delta\phi\right] + \frac{1}{2} H_+ J_1(m) \cdot \\ &\cos\left[\Omega t + (\phi_+ - \phi_0) - \Delta\phi - \frac{\pi}{2}\right] - \frac{1}{2} H_+ J_1(m) \cdot \\ &\cos\left[(2\omega_0 + \Omega)t + \phi_+ - \phi_0 + \frac{\pi}{2} + \Delta\phi\right] - \frac{1}{2} H_- J_1(m) \cdot \\ &\cos\left[\Omega t + (\phi_0 - \phi_-) + \Delta\phi + \frac{\pi}{2}\right] + \frac{1}{2} H_- J_1(m) \cdot \\ &\cos\left[(2\omega_0 + \Omega)t + \phi_0 + \phi_- + \Delta\phi + \frac{\pi}{2}\right] \end{aligned} \quad (2.184)$$

Take and arrange Ω component after narrowband filtering:

$$B_0(t) = \frac{1}{2}H_+J_1(m) \sin [\Omega t + (\phi_+ - \phi_0) - \Delta\phi] + \frac{1}{2}H_-J_1(m) \sin [\Omega t + (\phi_0 - \phi_-) + \Delta\phi] \quad (2.185)$$

Expression (2.185) represents: two components are created for Ω frequency after demodulation. One component creates $[(\phi_+ - \phi_0) - \Delta\phi]$ phase shift for input tone signal $B_i(t) = B_m \sin \Omega t$, the other component creates $[(\phi_0 - \phi_-) + \Delta\phi]$ phase shift. Output tone $B_0(t)$ is the sum of such two components.

For Fig. 2.44, its vector diagram is as shown in Fig. 2.45.

As shown in Fig. 2.45, Φ is the phase shift of $B_0(t)$ for $B_i(t)$, i.e., tone phase shift. In such vector relationship, we can clearly see physical significance of Φ and reason of Φ change.

According to Fig. 2.45 and by use of simple triangle relation, we can obtain:

$$\phi = \arctan \frac{H_- \sin [(\phi_0 - \phi_-) + \Delta\phi] + H_+ \sin [(\phi_+ - \phi_0) - \Delta\phi]}{H_- \cos [(\phi_0 - \phi_-) + \Delta\phi] + H_+ \cos [(\phi_+ - \phi_0) - \Delta\phi]} \quad (2.186)$$

$$B_{om} = \frac{J_1(m)}{2} \{H_+^2 + H_-^2 + 2H_+H_- \cos [(\phi_0 - \phi_-) - (\phi_+ - \phi_0) + 2\Delta\phi]\}^{\frac{1}{2}} \quad (2.187)$$

Output tone:

$$B_0(t) = B_{om} \sin (\Omega t + \phi)$$

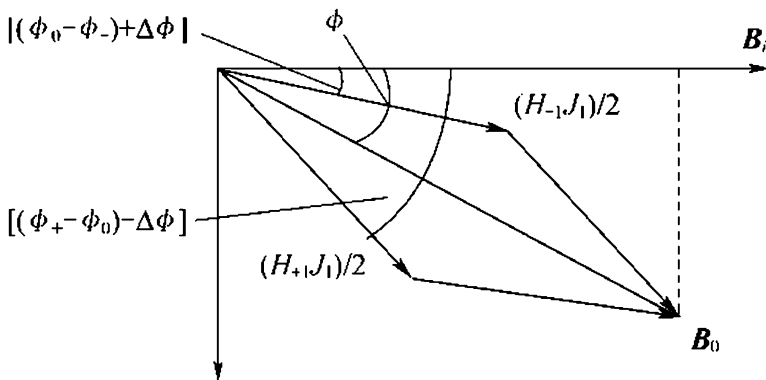


Fig. 2.45 Vector relationship of tone components output from product demodulator

It can be converted into tone delay τ from phase shift:

$$\begin{aligned}\tau &= \frac{\phi}{\Omega} \\ &= \frac{1}{\Omega} \arctan \frac{H_- \sin [(\phi_0 - \phi_-) + \Delta\phi] + H_+ \sin [(\phi_+ - \phi_0) - \Delta\phi]}{H_- \cos [(\phi_0 - \phi_-) + \Delta\phi] + H_+ \cos [(\phi_+ - \phi_0) - \Delta\phi]} \quad (2.188)\end{aligned}$$

where ϕ unit is rad, Ω unit is rad/s, τ unit is s.

Expression (2.188) is the complete expression of tone delay, which is the basis of tone ranging analysis. On the basis of derivation process of Expression (2.188), phase difference of two side frequency phase shift to carrier phase shift and asymmetry of two side frequency amplitude may impact tone delay, which are impacted by amplitude-/phase-frequency characteristics of filter, as well as spectrum line asymmetry from nonlinear, parasitic amplitude modulation, etc.

Expression (2.188) represents that in addition to phase-frequency characteristics, there are following factors that may impact tone delay.

1) Symmetry (H_+/H_-) of two-side frequency, in case of even symmetry (constant value of amplitude-frequency characteristics is a special case)

Substitute $H_+ = H_-$, $H_+/H_- = 1$ into Expression (2.186) and obtain:

$$\phi = \arctan \frac{\phi_+ - \phi_-}{2} = \frac{\phi_+ - \phi_-}{2}$$

According to the above expression, ϕ is not related to $\Delta\phi$. When amplitude-frequency characteristics are symmetrical, non-orthogonal product demodulation does not introduce tone delay. But according to Expression (2.187), output tone amplitude will be reduced, SNR is reduced and random error is increased. If there is amplitude-phase conversion in ranging unit, it may cause ranging error.

2) Odd symmetry ($\phi_+ - \phi_0 = \phi_0 - \phi_-$) of phase-frequency characteristics and quadrature demodulation ($\Delta\phi = 0$), substitute into Expression (2.186) to obtain

$$\phi = \arctan \frac{\sin (\phi_+ - \phi_0)}{\cos (\phi_+ - \phi_0)} = \frac{\phi_+ - \phi_-}{2} \quad (2.189)$$

In case of quadrature demodulation, amplitude-frequency characteristics does not impact tone delay.

Such (1) and (2) are for most actual cases, under most cases, $\phi = (\phi_+ - \phi_-)/2$, $\tau = (\phi_+ - \phi_-)/2\Omega$

3) In the above cases, if phase-frequency characteristic has an odd symmetry and linearity, refer to Fig. 2.46 for linear phase-frequency characteristics one can get

$$\tau = \frac{\phi_+ - \phi_-}{2\Omega} = \tan \theta = \frac{d\phi}{d\omega}$$

i.e., tone delay is equal to group delay, which is a special case of Expression (2.186).

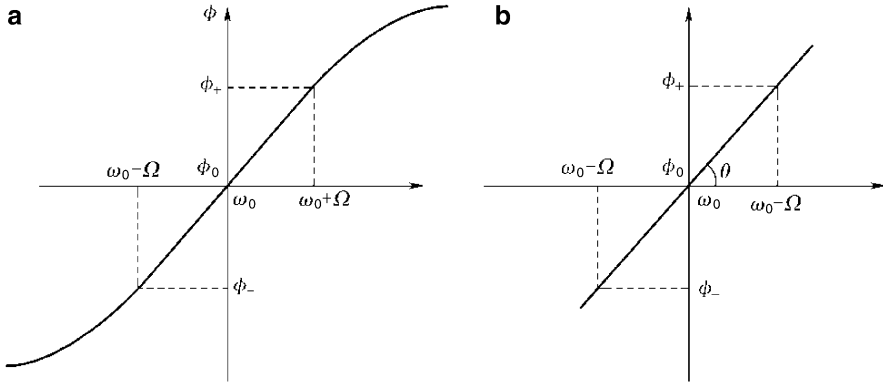


Fig. 2.46 Odd symmetry of phase-frequency characteristics

4) If $H_+ \neq H_-$, according to Expression (2.186), nonorthogonal demodulation still introduces delay, i.e., odd symmetry of phase-frequency characteristics; in case of non-even symmetry of amplitude-frequency characteristics, nonorthogonal demodulation still introduces additional delay. According to Expression (2.187), amplitude is decreased.

In conclusion, such factors can be summarized: ① if phase-frequency characteristics is not constant, it will introduce tone delay; ② if phase-frequency characteristics is not odd symmetry, amplitude-frequency characteristics will impact tone delay; ③ if amplitude-frequency characteristics is not even symmetry, nonorthogonal demodulation will impact tone delay.

To clarify some concepts, it is necessary to make clear following points.

In one four-terminal network, tone delay $\neq$ phase delay $\neq$ group delay (carrier delay is also called phase delay and equal to $[\phi(\omega)/\omega]$, but there are two special cases:

If phase-frequency characteristics are linear, we have tone delay = group delay = constant. If phase-frequency characteristics is linear and through the origin, we have tone delay = group delay = phase delay = $\tan\theta$ = constant. This is the case of non-dispersive medium.

When amplitude-frequency characteristics is given, phase delay, group delay, and tone delay are solely determined by phase-frequency characteristics, phase-frequency characteristics is of prior importance.

In the non-dispersive medium (phase delay does not vary with frequency), group velocity of wave is equal to phase velocity. For one tone-modulated carrier signal, its modulated tone velocity is equal to carrier velocity. In the non-dispersive medium (similar to the free space of non-dispersive medium), carrier delay and tone delay are the same, with velocity equal to velocity of light. Tone delay can be used for ranging. In the dispersive medium, group velocity is not equal to phase velocity, and group delay is not equal to phase delay. Group velocity is also not tone velocity and calibration or remedy is needed.

(2) Ranging error introduced by phase-frequency characteristics, amplitude-frequency characteristics, and product demodulation orthogonality.

Inherent delay of equipment on tone can be reduced by range calibration, but delay change may cause ranging error. Such delay error is a major component of ranging accuracy.

External factors causing delay change: temperature change, Doppler frequency shift, signal level change (AGC impact), power supply voltage change and aging, etc. Such factors can change tone delay through phase-frequency characteristics, amplitude-frequency characteristics and nonorthogonal product demodulation of channel. The quantitative relation of such impact can be obtained through total differentiation on Expression (2.186), i.e., total error expression:

$$\begin{aligned}
 \Delta\Phi = d\phi &= \frac{\partial\phi}{\partial(\Delta\phi)}d(\Delta\phi) + \frac{\partial\phi}{\partial\left(\frac{H_+}{H_-}\right)}d\left(\frac{H_+}{H_-}\right) + \frac{\partial\phi}{\partial(\phi_+ - \phi_0)}d(\phi_+ - \phi_0) + \frac{\partial\phi}{\partial(\phi_0 - \phi_-)}d(\phi_0 - \phi_-) \\
 &= M \left\langle \left\{ 1 - \left(\frac{H_+}{H_-}\right)^2 \right\} d(\Delta\phi) + \left\{ \cos[(\phi_0 - \phi_-) + \Delta\phi] \sin[(\phi_+ - \phi_0) - \Delta\phi] \right. \right. \\
 &\quad \left. \left. - \sin[(\phi_0 - \phi_-) + \Delta\phi] \cos[(\phi_+ - \phi_0) - \Delta\phi] \right\} \times d\left(\frac{H_+}{H_-}\right) \right. \\
 &\quad \left. + \left\{ \frac{H_+}{H_-} \cos[(\phi_0 - \phi_-) + \Delta\phi] \sin[(\phi_+ - \phi_0) - \Delta\phi] + \left(\frac{H_+}{H_-}\right)^2 \right\} d(\phi_+ - \phi_0) \right. \\
 &\quad \left. + \left\{ \frac{H_+}{H_-} \sin[(\phi_0 - \phi_-) + \Delta\phi] \sin[(\phi_+ - \phi_0) - \Delta\phi] + \left(\frac{H_+}{H_-}\right)^2 \right\} d(\phi_+ - \phi_0) \right. \\
 &\quad \left. + \left\{ 1 + \frac{H_+}{H_-}, \cos[(\phi_+ - \phi_-) - \Delta\phi], \cos[(\phi_0 - \phi_-) + \Delta\phi] \right. \right. \\
 &\quad \left. \left. + \left(\frac{H_+}{H_-}\right), \sin[(\phi_+ - \phi_0) - \Delta\phi], \sin[(\phi_0 - \phi_-) + \Delta\phi], \right\} d(\phi_0 - \phi_-) \right\rangle \\
 M &= \left\langle \left\{ \cos[(\phi_0 - \phi_-) + \Delta\phi] + \frac{H_+}{H_-} \cos[(\phi_+ - \phi_0) - \Delta\phi] \right\}^2 \right. \\
 &\quad \left. + \left\{ \sin[(\phi_0 - \phi_-) + \Delta\phi] + \frac{H_+}{H_-} \sin[(\phi_+ - \phi_0) - \Delta\phi] \right\}^2 \right\rangle^{-1}
 \end{aligned} \tag{2.190}$$

where $\Delta\Phi$ is tone phase shift increment; $\Delta\phi$ is nonorthogonal phase difference.

Thus, reducing M will decrease delay error. Minimum value of M shall be sought out in design.

From discussion on Expression (2.186) we can conclude:

1) Impact of phase-frequency characteristics. The third item in Expression (2.186) is the tone delay change when relative phase difference $(\phi_+ - \phi_0)$ of $(\omega_0 + \Omega)$ side frequency to carrier changes in case of no change for $\Delta\phi$, (H_+/H_-) and $(\phi_0 - \phi_-)$.

For special case: $(H_+/H_-) = 1$, $\phi_0 - \phi_- = \phi_+ - \phi_0$, substitute into Expression (2.186) to obtain $M = 1/4$, $d\phi(\phi_0 - \phi_-) = \frac{1}{2}d(\phi_0 - \phi_-) = \frac{1}{2}\Delta(\phi_0 - \phi_-)$, which is equal to 1/2 of phase difference change of side frequency to carrier.

Similarly

$$d\phi(\phi_+ - \phi_0) = \frac{1}{2}d(\phi_+ - \phi_0) = \frac{1}{2}\Delta(\phi_+ - \phi_0)$$

Tone delay variation is:

$$\begin{aligned} \Delta\phi_1 &= d\phi(\phi_0 - \phi_-) + d\phi(\phi_+ - \phi_0) = \Delta(\phi_+ - \phi_-) \\ \Delta R_1 &= \frac{c}{720f_r} \bullet \Delta(\phi_+ - \phi_-) \end{aligned} \quad (2.191)$$

2) Impact of amplitude-frequency characteristics asymmetry. The second item in Expression (1.186) is the delay change $d\phi(H_+/H_-)$ when amplitude-frequency characteristics non-even symmetry (H_+/H_-) changes in case of no change for $\Delta\phi$, $(\phi_+ - \phi_0)$ and $(\phi_0 - \phi_-)$.

$$\begin{aligned} \Delta\phi_2 &= d\phi\left(\frac{H_+}{H_-}\right) \\ &= M \left\{ \frac{\cos[(\phi_0 - \phi_-) + \Delta\phi] \sin[(\phi_+ - \phi_-) - \Delta\phi]}{\sin[(\phi_0 - \phi_-) + \Delta\phi] \cos[(\phi_+ - \phi_-) - \Delta\phi]} - \right\} \Delta\left(\frac{H_+}{H_-}\right) \end{aligned} \quad (2.192)$$

$\Delta\phi_2 = 0$ can be obtained by making $d_\phi(H_+/H_-) = 0$

By resolve Expression (2.190) we can get:

$$\begin{aligned} \frac{\sin[(\phi_+ - \phi_0) - \Delta\phi]}{\cos[(\phi_+ - \phi_0) - \Delta\phi]} &= \frac{\sin[(\phi_0 - \phi_-) + \Delta\phi]}{\cos[(\phi_0 - \phi_-) + \Delta\phi]} \\ \begin{cases} (\phi_+ - \phi_0) - \Delta\phi = (\phi_0 - \phi_-) + \Delta\phi \\ \Delta\phi = \frac{1}{2}[(\phi_+ - \phi_0) - (\phi_0 - \phi_-)] \end{cases} \end{aligned} \quad (2.193)$$

Above analysis shows: ① when adjusting product demodulation non-orthogonal phase difference $\Delta\phi$ to equal to Expression (2.193), change of amplitude-frequency characteristics even symmetry does not cause tone delay change; ② if product demodulator is orthogonal, i.e., $\Delta\phi = 0$, substitute into Expression (2.193)

$$\phi_+ - \phi_0 = \phi_0 - \phi_-$$

It is odd symmetry of phase-frequency characteristics.

When product demodulation is orthogonal, if phase-frequency characteristics are odd symmetry, change of amplitude-frequency characteristics asymmetry does not introduce delay change, which matches above discussed results.

3) Error caused by quadrature demodulation. The first item of Expression (2.189) is the delay error when non-orthogonal phase difference $\Delta\phi$ changes in case of no change for (H_+/H_-) , $(\phi_+ - \phi_0)$ and $(\phi_0 - \phi_-)$,

$$\Delta\phi_3 = d\phi(\Delta\phi) = M \left[1 - \left(\frac{H_+}{H_-} \right)^2 \right] \Delta(\Delta\phi) \quad (2.194)$$

When $H_+/H_- = 1$, $d\phi(\Delta\phi) = 0$, i.e., in case of amplitude-frequency characteristics even symmetry, nonorthogonal product demodulation will not introduce tone delay, which is the same as above conclusion.

On the basis of above analysis, phase-frequency characteristics is the most serious impacting factor on tone delay and other factors' impact can be properly controlled for decrease.

2.4.3 Group Delay Characteristics Analysis Method of Ranging Error [12]

According to above analysis, phase-frequency characteristics is the most direct and accurate method for ranging accuracy analysis. But group delay index or curve is often given in engineering, it is necessary to analyze ranging error as per group delay characteristics.

2.4.3.1 Group Delay Analysis Method of Tone Ranging Error [12]

(1) Several definition methods of channel delay characteristics

1) Group delay τ_g : The delay of channel on wave group signal is called group delay. According to Fourier analysis, one complex signal is composed of many sinusoidal wave components. Use such signal to modulate (including PM, FM, AM, BPSK, QPSK, etc.) carrier to produce one modulated wave, group delay is the delay of its modulated signal (integrated signal of wave group), its mathematic expression:

$$\tau_g = \frac{d\Phi(\omega)}{d\omega}$$

2) Ranging tone delay τ : Ranging tone is one single-frequency sine wave and the delay of channel on such single-frequency sine wave is called ranging tone delay. Reference [1] provides the relationship of ranging tone delay and channel transmission characteristics

$$\tau = \frac{\Delta\Phi}{\omega_m} = \frac{1}{\omega_m} \arctan \frac{H_{-1} \sin(\Phi_0 - \Phi_{-1}) + H_{+1} \sin(\Phi_{+1} - \Phi_0)}{H_{-1} \cos(\Phi_0 - \Phi_{-1}) + H_{+1} \cos(\Phi_{+1} - \Phi_0)} \quad (2.195)$$

where ω_m is ranging tone frequency; Φ_0 is phase shift produced by carrier ω_0 in phase-frequency characteristics; Φ_{+1} and Φ_{-1} are the phase shift corresponding to $\omega_0 + \omega_m$ and $\omega_0 - \omega_m$ in phase-frequency characteristics; H_{+1} and H_{-1} are the amplitude value corresponding to $\omega_0 + \omega_m$ and $\omega_0 - \omega_m$ in amplitude-frequency characteristics.

3) Test delay τ_e : The delay tested by test apparatus on channel is called test delay or envelope delay. In ordinary test apparatus, use one single-frequency sine wave to conduct amplitude modulation on carrier, measure the delay of such amplitude-modulation envelope after passing channel. Its mathematic expression is:

$$\tau_e = \frac{\Phi_{+1} - \Phi_{-1}}{2\Omega} \quad (2.196)$$

Such digital expression is similar to the ranging tone expression. Only when Ω is equal to ranging tone and modulation mode is phase modulation (or frequency modulation), it can represent ranging tone delay.

In Expression (2.196), when $\Omega \rightarrow 0$, then

$$\tau_e = \frac{\Delta\Phi}{\Delta\Omega} = d\Phi(\omega)/d\omega = \tau_g \quad (2.197)$$

But in actual test, $\Delta\Omega \rightarrow 0$ is impossible. Hence, test delay τ_e is similar to group delay.

In addition, “phase delay,” “phase truncation delay,” etc., are used to describe delay characteristics. New test apparatus can directly measure $\phi(\omega)$; microprocessor in test apparatus directly calculate differentiation $d\phi(\omega)/d\omega$ to obtain $\tau(\omega)$. Such various delay characteristics are hidden in and solely determined by frequency characteristics $H(j\omega)$.

(2) Expression calculation method of tone ranging error

Expression of group delay $\tau(\omega)$

$$\tau(\omega) = \frac{d\Phi}{d\omega} \quad (2.198)$$

Use Expression (2.198) to obtain the phase difference between phase Φ_{+1} of frequency point $(\omega_0 + \Omega_1)$ and phase Φ_{-1} of frequency point $(\omega_0 - \Omega_1)$

$$(\Phi_{+1} - \Phi_{-1}) = \int_{\omega_0 - \Omega_1}^{\omega_0 + \Omega_1} \frac{\tau(\omega)}{d\omega} \quad (2.199)$$

where $(\omega_0 \pm \Omega_1)$ unit is rad/s; $\tau(\omega)$ unit is s; $(\Phi_{+1} - \Phi_{-1})$ unit is rad.

1) Ranging error from linearity distortion item and parabolic distortion item. Nonlinear item of group delay characteristics can be approximated with Taylor series (above third order can be omitted), group delay fluctuation item can be approximated with Fourier series (above second order can be omitted), total group delay characteristics can be represented:

$$\tau(\omega_\Delta) = a_0 + a_1\omega_\Delta + a_2\omega_\Delta^2 + \tau_m \sin \left[\frac{2\pi\omega_\Delta}{\Omega_T} + \theta_0 \right] \quad (2.200)$$

where ω_Δ is the frequency difference of offset carrier center frequency ω_0 , $\omega_\Delta = \omega - \omega_0$, $\tau(\omega_\Delta)$ is the group delay characteristics function with ω_Δ as variable.

Expression (2.200) includes two parts; one part changes with frequency, i.e., second, third and fourth items, in which $a_1\omega_\Delta$ and $a_2\omega_\Delta^2$ are respectively the first order item (linearity distortion item) and second order item (parabolic distortion item) of group delay characteristics provided in engineering design; the fourth item is group delay fluctuation item with fluctuation amplitude of τ_m ; the above three items are not respectively provided and called in-band group delay change. The other part is a_0 , a_1 and a_2 , which is group delay change varying with time change, i.e., group delay stability provided in index (group delay change in how many hours, factors causing change includes temperature, span, power supply, input level, etc.).

Substitute Expression (2.199) into Expression (2.200) to obtain:

$$(\Phi_{+1} - \Phi_{-1}) = \int_{-\Omega}^{+\Omega} \frac{\tau(\omega_\Delta)}{d\omega_\Delta} = 2a_0\Omega + \frac{2a_2}{3}\Omega^3 \quad (2.201)$$

where Ω is tone frequency.

In most actual cases, ranging tone delay is determined by Expression (2.189).

$$\tau = \frac{(\Phi_{+1} - \Phi_{-1})}{2\Omega} = a_0 + \frac{1}{3}a_2\Omega^2 \quad (2.202)$$

By above expression, the change of τ caused by changes of a_0 and a_2 can be calculated.

Ranging error from Doppler frequency shift ω_d can be obtained from change $(\Delta\phi_{+d} - \Delta\phi_{-d})$ of $(\phi_{+1} - \phi_{-1})$ (general ω_d can be extended as other frequency

change factors, including signal center frequency drift, filter center frequency drift, etc.). In case of $(\omega_0 + \Omega)$, $\Delta\Phi_{+d}$ from change is:

$$\begin{aligned}\Delta\Phi_{+d} &= \int_{+\Omega}^{\Omega+\omega_d} \tau(\omega_\Delta) d\omega_\Delta = \int_{+\Omega}^{\Omega+\omega_d} (a_0 + a_1\omega_\Delta + a_2\omega_\Delta^2) d\omega_\Delta \\ &= a_0\omega_d + \frac{a_1}{2}[(\Omega + \omega_d)^2 - \Omega^2] + \frac{a_2}{3}[(\Omega + \omega_d)^3 - \Omega^3]\end{aligned}\quad (2.203)$$

Similarly, at $(\omega_0 - \Omega)$, $\Delta\Phi_{-d}$ from change ω_d is represented as follows:

$$\begin{aligned}\Delta\Phi_{-d} &= \int_{-\Omega}^{-\Omega+\omega_d} \tau(\omega_\Delta) d\omega_\Delta \\ &= a_0\omega_d + \frac{a_1}{2}[(-\Omega + \omega_d)^2 - \Omega^2] + \frac{a_2}{3}[(-\Omega + \omega_d)^3 + \Omega^3]\end{aligned}\quad (2.204)$$

Ranging error from Doppler frequency shift is:

$$\Delta\tau_d = \frac{\Delta\Phi_{+d} - \Delta\Phi_{-d}}{2\Omega} = a_1\omega_d + a_2\omega_d^2 \quad (2.205)$$

where a_1 and a_2 are usually provided in specification. If the specification provides group delay characteristics curve, a_1 and a_2 can be calculated from such curve, as shown in Fig. 2.47.

In Fig. 2.47, dotted line is the linearity distortion curve including a_1 , solid line is the total curve including a_1 and a_2 .

The method for calculating a_1 , a_2 with above curves is as follows:

① Linear component a_1 (first order item). Obtain from Fig. 2.47:

$$a_1 = \frac{\tau_A - \tau_B}{B_s} \quad (2.206)$$

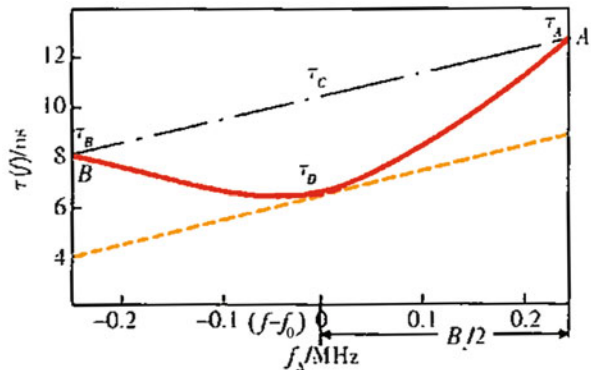


Fig. 2.47 Group delay characteristics curve including a_1 and a_2 distortion

It represents with the time delay of unit band with unit of ns/MHz; B_s is the bandwidth occupied by ranging signal spectrum.

② Parabolic curve component a_2 (second order item): In Fig. 2.47, τ_c is the group delay value of straight line connecting point A and B at center frequency f_s .

According to Fig. 2.47,

$$\tau_D = a_0 \text{ (corresponding to } \omega_\Delta = 0)$$

$$\tau_A = a_0 + a_1 \left(\frac{B_s}{2} \right) + a_2 \left(\frac{B_s}{2} \right)^2 \text{ (corresponding to } \omega_\Delta = \frac{B_s}{2})$$

then

$$\begin{aligned} \tau_c &= \tau_A - a_1 \left(\frac{B_s}{2} \right) \\ (\tau_c - \tau_D) &= a_2 \left(\frac{B_s}{2} \right)^2 \\ a_2 &= \frac{(\tau_c - \tau_D)}{(B_s/2)^2} \end{aligned} \quad (2.207)$$

2) Ranging error from group delay fluctuation. Group delay fluctuation index is provided in engineering design and its fluctuation characteristics curve can be various. We can use Fourier series to decompose it into the sum of many sinusoidal fluctuation harmonics. When first order item is taken, it is represented:

$$\tau(\omega_\Delta) = \tau_m \sin \left[2\pi \frac{\omega_\Delta}{\Omega_T} + \theta_0 \right] \quad (2.208)$$

where Ω_T is the frequency interval of one cycle sinusoidal fluctuation of group delay with unit of rad/s (or Hz); ω_Δ is independent variable with the same unit as Ω_T ; τ_m is the maximum group delay peak value with unit of s; θ_0 is the initial phase with unit of rad.

Group delay fluctuation characteristics curve is as shown in Fig. 2.48.

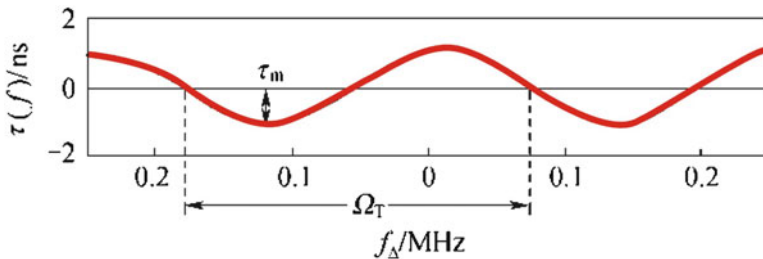


Fig. 2.48 Group delay fluctuation characteristics

From Fig. 2.48 we can obtain

$$\theta_0 = 180^\circ - \left[\frac{(\Omega_1 - \Omega_2)}{\Omega_T} \right] \times 360^\circ$$

Expression of phase difference $\Delta\Phi_{+d}$ and $\Delta\Phi_{-d}$ from Doppler frequency shift ω_d is as follows:

$$\begin{aligned} \Delta\Phi_{-d} = & \int_{-\Omega}^{-\Omega+\omega_d} \tau_{\max} \sin \left[2\pi \frac{\omega_d}{\Omega_T} + \theta_0 \right] d\omega_d = -\tau_{\max} \frac{\Omega_T}{2\pi} \left[\cos \frac{2\pi}{\Omega_T} (\Omega - \omega_d) - \cos \frac{2\pi}{\Omega_T} \Omega \right] \cos \theta_0 \\ & + \tau_{\max} \frac{\Omega_T}{2\pi} \left[\sin \frac{2\pi}{\Omega_T} (\omega_d - \Omega) + \sin \frac{2\pi}{\Omega_T} \Omega \right] \sin \theta_0 \end{aligned}$$

Be same as that,

$$\Delta\Phi_{+d} = \int_{+\Omega}^{+\Omega+\omega_d} \tau_{\max} \sin \left[2\pi \frac{\omega_d}{\Omega_T} + \theta_0 \right] d\omega_d$$

Ranging error from Doppler frequency shift ω_d is:

$$\begin{aligned} \Delta\tau_d &= \frac{\Delta\Phi_{+d} - \Delta\Phi_{-d}}{2\Omega} \\ &= \tau_{\max} \frac{\Omega_T}{2\pi\Omega} \sin \frac{2\pi\Omega}{\Omega_T} \left[\sin \frac{2\pi\omega_d}{\Omega_T} \cos \theta_0 + \left(\cos \frac{2\pi\omega_d}{\Omega_T} - 1 \right) \sin \theta_0 \right] \quad (2.209) \end{aligned}$$

Ranging error can be calculated according to Expression (2.209). It can be seen that the ranging error resulted from group delay fluctuation characteristics is related to fluctuation amplitude $\tau_{\max}$ and initial phase θ_0 of fluctuation. Different ω_d/Ω_T and Ω/Ω_T may cause different ranging errors. When $\Omega \gg \Omega_T$, $\Delta\tau_d \rightarrow 0$ and Ω_T gradually decreases to be less than $\Omega/2$, it evolves to be near linearity distortion or square distortion. When $\Omega \ll \Omega_T$ and $\theta_0 = 0^\circ$, $\Delta\tau_d = \tau_{\max}$ is transformed into Δa_0 , it does not reach the maximum when ω_d is maximum, but reaches the maximum in case of some ω_d . Under some conditions, it may not introduce ranging error (e.g., when $\theta_0 = 0^\circ$ and $2\pi\omega_d/\Omega_T = n\pi$, its physical significance is that plus-minus half-cycle integration of sine wave counteracts each other). On the basis of such complexity, group delay index of engineering design does not include such index as Ω_T and θ_0 . Use group delay characteristics curve and following graphic(al) method for resolution.

3) Ranging error from group delay stability. Definition of group delay stability: assuming with the change of time, environment conditions, AGC circuit, etc., a_0 has a change of Δa_0 and a_2 has a change of Δa_2 , they will cause a ranging error of $\Delta\tau_0$ which can be calculated from Expression (2.202), i.e.,

$$\Delta\tau_0 = \Delta a_0 + \frac{\Delta a_2}{3} \Omega^2 \quad (2.210)$$

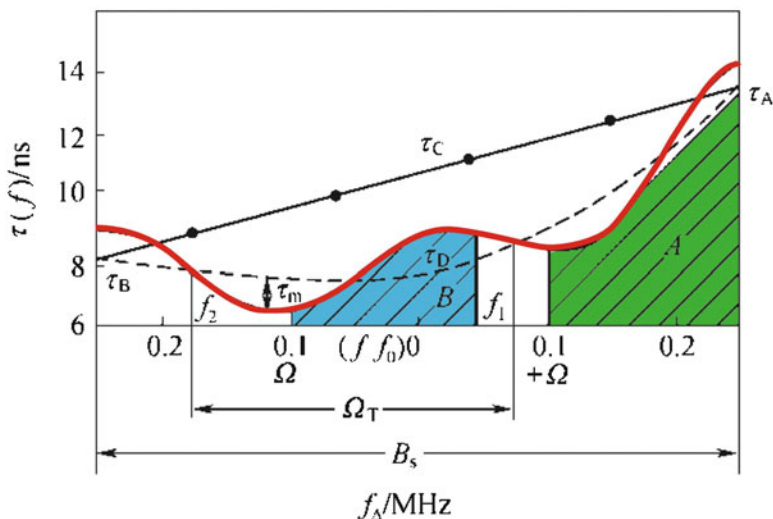


Fig. 2.49 Graphic method of ranging error

$\Delta\tau_0$ is the major component in ranging error and decreased by “range calibration,” but change between the two “zero calibration” may cause ranging error.

(3) Graphic calculation method of tone ranging error

If actual $\tau(\omega)$ curve is complex and Taylor series and Fourier series cannot be used for representation, according to the measured curve diagram $\tau(\omega)$, use graphic area solution method for integration of Expressions (2.203) and (2.204), which can also calculate the ranging error. Figure 2.49 is a $\tau(f)$ curve, take it as an example for graphic calculation of ranging error when $\Omega = 100$ kHz, $f_d = 150$ kHz.

In Fig. 2.49, the dotted line is the peak-peak average value curve of the fluctuation.

According to Expressions (2.203) and (2.204), integration of $\Delta\Phi_{+d}$ is the area A under $\tau(f)$ envelope in an interval of f from $+0.1$ MHz to $(+0.1 + 0.15)$ MHz = 0.25 MHz. Similarly, $\Delta\Phi_{-d}$ is the area B under $\tau(\omega)$ envelope in an interval of f_d from -0.1 MHz to $(-0.1 + 0.15)$ MHz = 0.05 MHz. They are represented by the shaded area in Fig. 2.49, from which one can obtain:

$$\Delta\Phi_{+d} = \text{area A} = 4.4 \text{ rad}$$

$$\Delta\Phi_{-d} = \text{area B} = 1.95 \text{ rad}$$

From Expression (2.211) we can also obtain:

$$\Delta\tau_d = \frac{\Delta\Phi_{+d} - \Delta\Phi_{-d}}{2\Omega} = \frac{(4.4 - 1.95) \times 10^{-3} \text{ rad}}{2 \times 0.1 \times 10^6 \times 2\pi \text{ rad/s}} = 1.95 \text{ ns}$$

Below, we will calculate them again by formula method and make a comparison of the two calculation results.

1) Ranging error $\Delta\tau'_d$ from first order item and second order item can be obtained from Fig. 2.49:

$$B_s = 0.5 \text{ MHz},$$

$$\tau_A - \tau_B = 7.4 \text{ ns}, \quad \tau_C - \tau_D = 3 \text{ ns}$$

From Expression (2.207), we get

$$a_1 = 5.2 \text{ ns}/0.5 \text{ MHz} = 10.4 \text{ ns/MHz}$$

From Expression (2.208), we get

$$a_2 = \frac{3 \text{ ns}}{(0.25 \text{ MHz})^2} = 48 \text{ ns/MHz}^2$$

From Expression (2.206), we get

$$\Delta\tau'_d = a_1\omega_d + a_2\omega_d^2$$

Substitute into a_1 , a_2 and ω_d , etc., the calculation result is:

$$\Delta\tau'_d = 2.64 \text{ ns}$$

2) Ranging error $\Delta\tau''_d$ from group delay fluctuation. From Fig. 2.49, we get:

$$\Omega_T = (f_1 - f_2) = 250 \text{ kHz}, \quad \tau_m = 1.2 \text{ ns},$$

$$\theta_0 = 180^\circ - [(f_1 - f_0)/\Omega_T] \times 360^\circ = 72^\circ$$

$\Delta\tau''_d$ can be obtained by substituting these values into Expression (2.211) and results are as listed in Table 2.9.

The total ranging error $\Delta\tau_d = \Delta\tau'_d + \Delta\tau''_d = 2.1 \text{ ns}$ when $\omega_d = 0.15 \text{ MHz}$ is close to 1.95 ns which is obtained by graphic method. Moreover, we can conclude that ranging error from group delay fluctuation does not increase when ω_d increases.

Table 2.9 Calculation results of ranging error

$\omega_d(\text{MHz})$	0	0.03	0.06	0.09	0.12	0.13	0.14	0.15
$\Delta\tau''_d(\text{ns})$	0	-0.015	-0.12	-0.375	-0.525	-0.547	-0.55	-0.54

2.4.3.2 Group Delay Characteristics Analysis Method of PN Code Ranging Error and Spread Spectrum Ranging Error [13]

(1) Principle of generation

PN code ranging and spread spectrum ranging is usually implemented by delay locking loop, and the null point of error discriminator in the loop is symbol measurement point of the PN code ranging delay. The change of such null point on the time axis reflects that of the ranging delay. Its physical mechanism is: the output voltage of the time error discriminator in the loop is zero in case of stable tracking of PN code. If influenced by other factors, the original time discrimination characteristics become a new time discrimination characteristics, and its output voltage changes by ΔE accordingly. Such ΔE will make the frequency and phase of the local PN code clock change and, in the meantime, make phase (i.e., delay) of PN code change simultaneously. After it changes by the delay $\Delta\delta_0$, the output voltage becomes zero on the new time discrimination characteristics; at this time, the delay locking loop will become stable again. $\Delta\delta_0$ is the ranging error arising from change of time discrimination characteristics, and is corresponding to zero point of output voltage on new time discrimination characteristics. Therefore, if the value of δ'_0 corresponding to the output zero point of new (changed) time discrimination characteristics is solved and then minus by δ_0 of the original time discriminator, the ranging delay error $\Delta\delta_0 = \delta'_0 - \delta_0$ can be obtained, and then solve the range error through delay error.

Change of channel group delay will result in change in time discrimination characteristics. The ranging error arising out of change in group delay characteristics can be solved according to the foregoing mechanism. However, it is very difficult to get its mathematical expression via analytical method, so simulation method is recommended. The section below will set out its impact upon ranging error through simulation.

(2) Simulation block diagram

The simulation block diagram is as shown in Fig. 2.50:

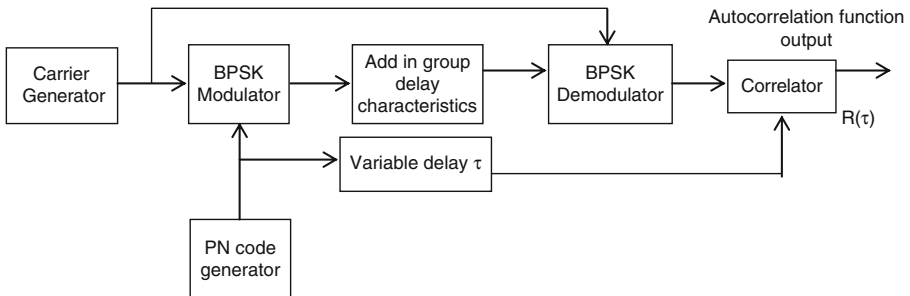


Fig. 2.50 Simulation block diagram of PN code ranging error

In Fig. 2.50 “add in group delay characteristics” include first degree, quadratic term, sinusoidal variation term, and constant-value term of group delay characteristics set out in Sect. 2.4.3.1. It simulates the group delay characteristics in the following expression:

$$\tau(\omega_{\Delta}) = a_0 + a_1\omega_{\Delta} + a_2\omega_{\Delta}^2 + \tau_m \sin \left[\frac{2\pi\omega_{\Delta}}{\Omega T} + \theta_0 \right]$$

where ω_{Δ} is the frequency difference deviating from the central frequency ω_0 .

Group delay characteristics can be simulated by full-pass filter to solve the auto-correlated function $R(t)$ while t is changing, and then calculate the time discrimination characteristics $D(t)$ of advanced code and lagging code at an interval of a code element width T_c : $R(\tau) - R(\tau + T_c) = D(\tau)$.

Then, further solve the value of τ'_0 corresponding to $D(\tau) = 0$ in case of different group delay characteristics and the value of τ_0 in absence of group delay characteristics (namely, the group delay is zero). Calculate $\delta_0 = (\tau'_0 - \tau_0)$ to obtain the ranging signal delay δ_0 (i.e., the delay in absence of group delay) arising out from different group delay characteristics; the change $\Delta\delta_0$ caused by change in such δ_0 along with the group delay characteristics is the ranging error introduced by the group delay characteristics.

(3) Simulation results [13]

When the rate of PN code is 3 and 10 Mbps, the simulation result is as shown in Fig. 2.51.

In Fig. 2.51, the subfigure (a) is PN code ranging delay caused by sinusoidal group delay variation, B_I means the bandwidth of filter (unit: radians/second), usually as $B_I = 1/T_c$. The simulation result reveals that the ranging delay $\delta_0 = \bar{\tau}_{B_I}$, namely,

$$\delta_0 = \bar{\tau}_{B_I} = \frac{1}{B_I} \int_{-B_I/2}^{+B_I/2} \tau_m \sin \left[\frac{2\pi\omega_{\Delta}}{\Omega T} + \theta_0 \right] d\omega_{\Delta} \quad (2.211)$$

Where, $\bar{\tau}_{B_I}$ is the ratio of average value of group delay to the value of B_I within B_I . According to the expression above, it is shown that δ_0 is relevant to amplitude of the group delay variation, bandwidth of B_I , starting phase of variation within the bandwidth (similarly, it is also relevant to the ending phase). But there is a special circumstance, under which, the positive semicircle number of the variation is equal to the negative semicircle number when both starting phase and ending phase is zero; therefore, in case of $\bar{\tau}_{B_I} = 0$, whatever is amplitude and times of the variation, no PN code delay is caused.

Any group delay variation curve can be decomposed to be the sum of n times of harmonic curve; respectively solve δ_{0n} and then solve δ_{Σ} .

In Fig. 2.51, the subfigure (b) is PN code ranging delay caused by the linear group delay, a_1 is the slope of group delay characteristics (unit: ns/MHz), and t_B means the value of group delay at the passband edge (namely, at $\omega_0 \pm B_I/2$, $B_I = 1/T_c$)

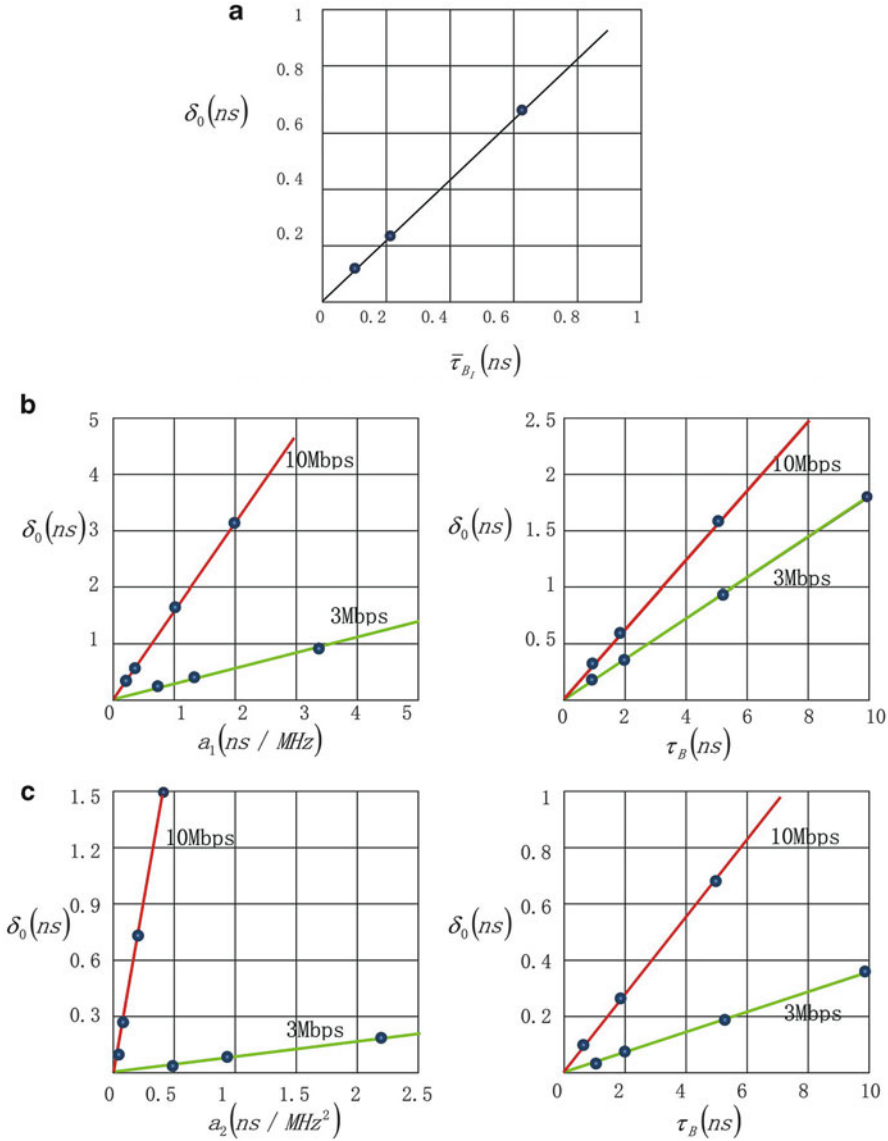


Fig. 2.51 Simulation results when PN code delay is caused by group delay. (a) Relationship between PN code delay and sinusoidal variation group delay characteristics. (b) Relationship between PN code delay and linear group delay characteristics. (c) Relationship between PN code delay and parabolic group delay characteristics

(called as “boundary delay”). The simulation result shows that δ_0 is in linear relationship with a_1 which can be fit to a mathematical expression:

$$\begin{aligned} \delta_0 &= 0.25a_1, \delta_0 = 0.17\tau_B \quad (\text{in case of 3Mbps code rate}) \\ \delta_0 &= 1.5a_1, \delta_0 = 0.3\tau_B \quad (\text{in case of 10Mbps code rate}) \end{aligned} \quad (2.212)$$

In Fig. 2.51, the subfigure (c) is PN code ranging delay caused by the parabolic group delay, a_2 is the coefficient of the parabola (unit: ns/MHz²), τ_B is the value of group delay at the passband edge (namely, at $\omega_0 \pm B_I/2$, $B_I = 1/T_c$). The simulation result shows that the greater a_2 is, the greater τ_B becomes; as a result, the greater δ_0 so caused. δ_0 is approximately in linear relationship with a_2 , which can be fit into a mathematical expression:

$$\begin{aligned} \delta_0 &= 0.085a_2, \quad \delta_0 = 0.038\tau_B \quad (\text{in case of 3Mbps code rate}) \\ \delta_0 &= 3.3a_2, \quad \delta_0 = 0.15\tau_B \quad (\text{in case of 10Mbps code rate}) \end{aligned} \quad (2.213)$$

The PN code ranging signal delay introduced by constant-value group delay is equal to the constant-value coefficient of group delay characteristics, namely, $\delta_0 = a_0$.

Expressions (2.211), (2.212), and (2.213) reveal that a_1 , a_2 , a_0 , τ_B , and $\bar{\tau}_{B_I}$ will introduce PN code ranging signal delay. It introduces range zero value which can be deducted by range calibration. After being calibrated, the change in a_1 , a_2 , a_0 , τ_B , and $\bar{\tau}_{B_I}$ will be introduced into the ranging error.

2.4.4 Random Ranging Error

Thermal noise and short-term stability of ranging signal will cause the random ranging error. More details are discussed below.

2.4.4.1 Ranging Error Introduced by Thermal Noise

According to CW ranging principle $R = \frac{c\Phi}{4\pi f_R}$, we can obtain that

$$\sigma_{R_1} = \frac{C}{4\pi} \frac{\delta\Phi}{f_R} \quad (2.214)$$

where $\delta\Phi$ is the phase measurement error of the equipment, f_R is the frequency of high tone, and σ_{R_1} is the required ranging accuracy.

The phase measurement error of the equipment $\delta\Phi$ is limited by the SNR of the tone signal to be modulated. For the purpose of improving the phase measurement accuracy for the high tone, the phase-locked loop (PLL) has to be used. Because the closed-loop narrowband filtering of the phase-locked loop is able to improve the

SNR of tone, under the function of AWGN $n(t)$, when the SNR of phase-locked loop $(S/N)_L$ is greater than the threshold value (≥ 7 dB), the relation between the phase jitter output by the main tone loop and the loop SNR may be expressed that

$$\delta_\Phi = \sqrt{\delta_\Phi^2} = \sqrt{\delta_\Phi^2(t)} = \frac{1}{\sqrt{2(S/N)_L}} = \frac{1}{\sqrt{\frac{S}{\Phi B_L}}} \quad (2.215)$$

If $(S/N)_L < 7$ dB, the losing lock probability of the phase-locked loop increases, and the pulse noise will be found in the phase noise δ_Φ , consequently making the result greater.

Substitute the δ_Φ in Expression (2.215) into Expression (2.214), then we obtain

$$\sigma_{R_1} = \frac{c}{4\pi f_R \sqrt{2}} \frac{1}{\sqrt{(S/N)_L}} \approx \frac{c}{18f_R \sqrt{(S/N)_L}} = \frac{c}{18f_R \sqrt{(\frac{S}{\Phi}) \frac{1}{2B_L}}} \quad (2.216)$$

where $(S/N)_L < 7$ dB in general, $N = 2\Phi B_L$, Φ is noise power spectral density, B_L is the equivalent noise bandwidth for a single side of loop, $B_L = \int_0^\infty |H(j2\pi f)|^2 df$, which is related to the type and exponent number of the phase-locked loop. See Table 2.7 for calculation.

2.4.4.2 Ranging Error Caused by Short-Term Stability of Ranging Signal

The phase noise of the ranging signal (shown as short-term stability in time domain) may introduce additional ranging error. The reason is that the phase noise has frequency modulation flicker noise and frequency drift noise, and then these noises may be radiated as time extended. Therefore, for the deep-space two-way ranging with long time delay, if the short-term stability of a signal is not required strictly, the non-ignorable ranging error will appear. However, due to a short R/T time delay in CC&T for the orbit satellites of the Earth, such error may be neglected.

The ranging signal is expressed that

$$u_R(t) = A \cos [\omega_s t + \varphi(t)]$$

where ω_s is the nominal frequency, and $\varphi(t)$ is the phase noise.

In a basic model, the ranging distance is the phase of R/T CW ranging signal, which has a space time delay T . Assuming that the additional phase noises may not be introduced by the uplink/downlink channel for ranging and the transponder, that is, the frequency fluctuation (short-term stability) should only be caused by the main vibration source, then the two-way ranging model is simplified as the model shown in Fig. 2.52.

Fig. 2.52 Ranging model

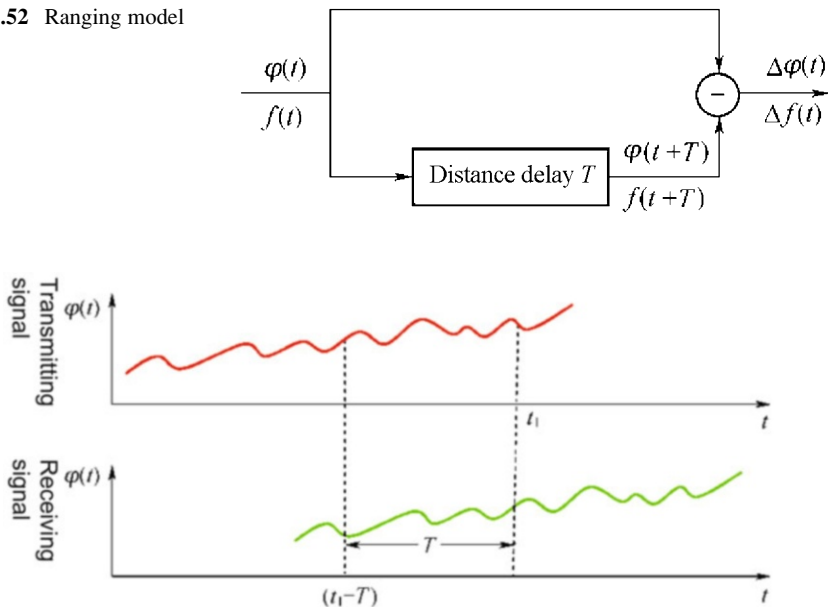


Fig. 2.53 Phase and frequency relationship of R/T signals

In Fig. 2.52, the phase difference $\Delta\varphi(t)$ relative to ranging distance is equal to the difference between the phase $\varphi(t)$ and the phase $\varphi(t+T)$ of transmitter, which have T time difference then

$$\Delta\varphi(t) = \varphi(t+T) - \varphi(t) \quad (2.217)$$

Correspondingly, the difference of frequency fluctuation is expressed that

$$\Delta f(t) = f(t+T) - f(t) \quad (2.218)$$

In Expression (2.217), the first term and the second term are the two measured values of two times of transmitting signals respectively at T time interval; the corresponding time relationship is as shown in Fig. 2.53.

In Fig. 2.53, the measurement starts at t_1 , and the time delay of measured R/T signal is T . The measurement shown in Fig. 2.53 is equivalent to the measurement for the transmitting signal shown in Fig. 2.54.

In Fig. 2.54, for the R/T signal with delay time T , the measurements of phase difference $\Delta\varphi$ and frequency difference Δf are equivalent to those of phase difference and frequency difference between $(t_1 - T)$ and t_1 .

Distance ranging is implemented by measurement of phase difference. Therefore, the relationship of the short-term stability with the phase fluctuation difference $\Delta\varphi$ shall be derived to obtain the result. The relationship between the phase and the frequency is expressed that

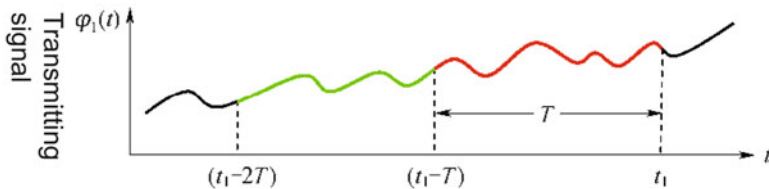


Fig. 2.54 Equivalent measurement for transmitting signal

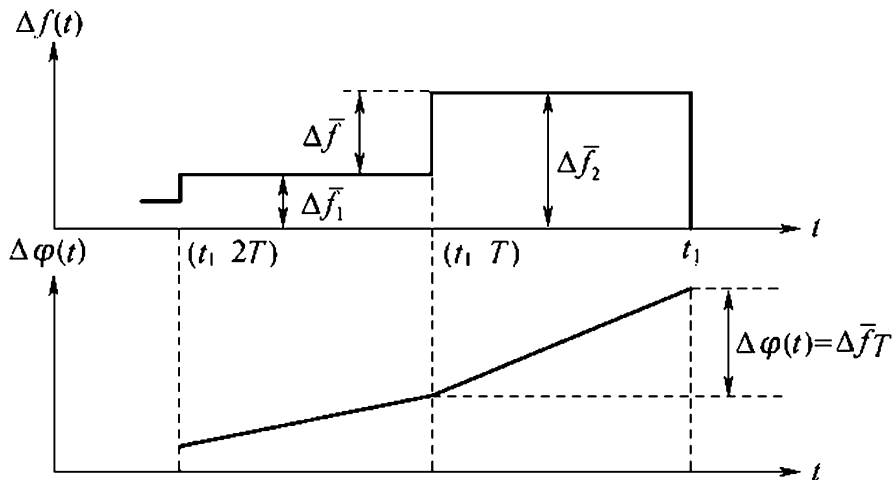


Fig. 2.55 Relationship of phase and frequency fluctuation

$$\Delta\varphi(T) = 2\pi\langle\Delta\bar{f}\rangle T \quad (2.219)$$

where $\Delta\varphi(T)$ is the phase difference of R/T signal caused by frequency fluctuation at delay time T , $\Delta\bar{f}$ is the change of frequency fluctuation averages within T time against the previous time, and $\langle \dots \rangle$ refers to average value of multiple times for improving measurement accuracy.

It is assumed that the previous time is T (so it can be related to Allan variance in this way), as shown in Fig. 2.55.

In Fig. 2.55, $\Delta\bar{f}_2$ is the average value of the difference $\Delta f_2(t) = f(t) - f_s$ of the frequency fluctuation $f(t)$ and the frequency nominal value f_s within $(t_1 - T) - t_1$, $\Delta\bar{f}_1$ is the average value of the difference $\Delta f_1(t)$ of the frequency fluctuation $f(t)$ and f_s within $(t_1 - 2T)$, and the $\Delta\bar{f}_2 - \Delta\bar{f}_1 = \Delta\bar{f}$ expresses the change of average values for adjacent two intervals that occurred at $(t_1 - T)$ and t_1 , so that the phase change $\Delta\varphi(T)$ shall evolve as $\Delta\varphi(T) = \Delta\bar{\omega}T = 2\pi\langle\Delta\bar{f}\rangle T$, which is the ranging error relative to the frequency fluctuation (short-term stability). Following two physical concepts shall be noted in this analytical method.

(1) T is not the integration time of the circuit but the delay between T/R signals which exists objectively in a physical relationship when the phase difference is derived from the frequency fluctuation (Expression (2.219)).

(2) And take the previous time segment also as T , for the purpose of measurement with Allan variance, then the relationship between the ranging error introduced by short-term stability and the Allan variance may be obtained (or using other methods to gain that, however, the results obtained with different methods shall be the same).

According to Expression (2.55), we can obtain that:

$$\begin{aligned}\Delta\bar{f} &= \Delta\bar{f}_2 - \Delta\bar{f}_1 \\ \langle (\Delta\bar{f})^2 \rangle &= \langle (\Delta\bar{f}_2 - \Delta\bar{f}_1)^2 \rangle \\ \frac{\langle (\Delta\bar{f})^2 \rangle}{f_s^2} &= \frac{\langle (\Delta\bar{f}_2 - \Delta\bar{f}_1)^2 \rangle}{f_s^2}\end{aligned}$$

Substitute the relative frequency instability $y = \Delta f/f_s$ into the expression above; then

$$\frac{\langle (\Delta\bar{f})^2 \rangle}{f_s^2} = \langle (\bar{y}_2 - \bar{y}_1)^2 \rangle$$

Because the Allan variance for zero clearance is

$$\sigma_y^2(T) = \frac{1}{2} \langle (\bar{y}_2 - \bar{y}_1)^2 \rangle$$

Therefore,

$$\langle (\Delta\bar{f})^2 \rangle = 2f_s^2 \sigma_y^2(T) \quad (2.220)$$

Substitute Expression (2.220) into Expression (2.219) to obtain:

$$\Delta\varphi(T) = 2\pi\sqrt{2}f_s\sigma_y(T)T$$

The one-way ranging error caused by short-term stability shall be:

$$\sigma_{R_2} = \frac{c \cdot \Delta\varphi}{2 \cdot 2\pi f_s} = \frac{CT}{\sqrt{2}} \sigma_y(T) \quad (2.221)$$

The two-way ranging error caused by short-term stability is $2\sigma_{R_2}$.

Normally $\sigma_y(T)$ does not decrease linearly with increasing of T when T value is large. Therefore, the ranging error caused by it will be larger because of longer ranging distance. The result calculated from the expression above shows that such error is negligible [2], if the distance is not so far (e.g., to the Mars), but for extreme deep-space exploration, in case of Pluto along the solar system or entering into the Galaxy outside solar system, if the short-term stability of a signal is good enough, the non-negligible ranging error shall be introduced. Let's take an example of Pluto exploration. The Pluto is the farthest planet away from the Earth among the nine planets in solar system, and the time delay of electric wave between the Pluto and the Earth is about 27 h, that is,

$$T \approx 27 \times 3,600 = 9.72 \times 10^4(s)$$

If hydrogen clock is used for measurement, by substituting $\sigma_y(9.72 \times 10^4s) \approx 8 \times 10^{-15}$ into Expression (2.221), we can obtain $\Delta R = 0.16$ m, which is not the major component in the total error.

If cesium clock is used for measurement, by substituting $\sigma_y(9.72 \times 10^4s) \approx 8 \times 10^{-14}$ into Expression (2.221), we can obtain $\Delta R = 1.6$ m, which is the major component in the total error. Therefore, in extreme deep-space ranging, the effect of short-term stability of a signal on ranging accuracy shall be considered if the distance is extremely far, and a high-quality atomic clock is preferred.

The analysis above is an objective process in physics without effects of additional circuit, and integration filter is usually used in the practical circuit. The frequency domain method is recommended to analyze the effect of that filter. The transmission function diagram is as shown in Fig. 2.56.

As shown in Fig. 2.56, $\phi_i(\omega)$ is the power spectrum of input phase noise, and $\phi_0(\omega)$ is the power spectrum of output phase noise. $|H_1(\omega)|^2$ is the power transmission function of delay canceling.

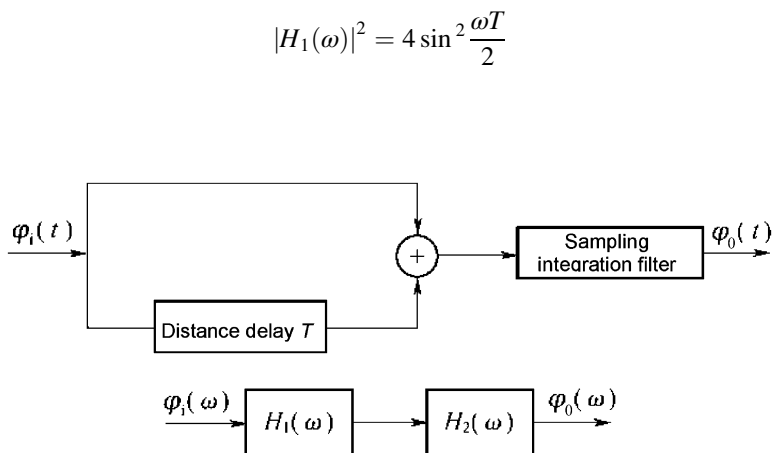


Fig. 2.56 Ranging transmission function block diagram

$|H_2(\omega)|^2$ is the power transmission function of integration filter when the sampling smoothing time is τ .

$$|H_2(\omega)|^2 = \frac{[\sin^2(\omega\tau/2)]}{(\omega\tau/2)^2}$$

The variance of the output phase noise is:

$$\sigma_\varphi^2 = \frac{1}{2\pi} \int_0^\infty \left\{ \varphi_i(\omega) \cdot 4 \sin^2\left(\frac{\omega T}{2}\right) \cdot \left[\sin^2\left(\frac{\omega\tau}{2}\right) \right] / \left(\frac{\omega\tau}{2}\right)^2 \right\} d\omega \quad (2.222)$$

The frequency characteristics of $\varphi_i(f)$, $|H_1(f)|^2$ and $|H_2(f)|^2$ are as shown in Fig. 2.57.

As shown in Fig. 2.57, in case of good short-term stability (i.e., the $\varphi_i(f)$ in the figure is narrow), for deep-space TT&C with long time T , the integration filter works only when τ is close to T or τ is greater than T . If the sampling integration time τ is short, so the integration filter may not takes a well effect. In a practical circuit, the integration time is usually less than the delay time T of deep-space TT&C.

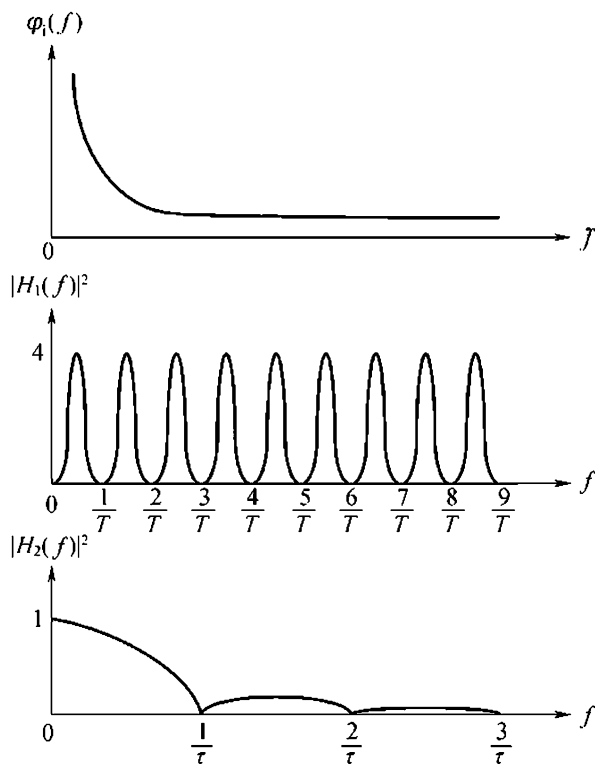


Fig. 2.57 Frequency domain analysis of ranging error

2.4.5 Theoretical Calculation of Ranging Accuracy

The ranging accuracy is mainly restricted by the random and system measurement errors which appear as random variable or system variable in measuring round-trip time of the ranging signal. Error calculation expressions are given below.

(1) Random error

The one-way random error that affects the equipment ranging accuracy has following error sources.

1) The error caused by thermal noise (see Sect. 2.4.5).

$$\sigma_{R_1} = \frac{c}{18f_r} \left[\frac{S}{2\varphi B_n} \right]^{-\frac{1}{2}} \quad (2.223)$$

where c is the velocity of light, f_r is the frequency of ranging principal tone, S/φ is the ratio of the power spectral density of principal tone signal and the noise spectral density, and B_n is the equivalent noise bandwidth of single-side principal tone loop.

2) One way error caused by instability of short-term frequency of the ranging principal tone (see Sect. 2.4.5).

$$\sigma_{R_2} = \frac{cT}{\sqrt{2}} \sigma_y(T) \quad (2.224)$$

where $\sigma_y(T)$ is the Allan variance of the instability of principal tone short-term frequency and T is the time delay in transmitting/receiving signal.

3) Error caused by multiple communication paths (see Sect. 5.4.4).

$$\sigma_{R_3} = \frac{\left(\frac{U_{m_2}}{2\pi U_{m_1}} \right)}{2f_R} \quad (2.225)$$

where U_{m_2} is the amplitude of multipath reflection signal, and U_{m_1} is the amplitude of direct signal.

4) Error caused by the phase noise of principal tone ring VCO.

$$\sigma_{R_4} = \frac{\sigma_\varphi c}{4\pi f_R} \quad (2.226)$$

where σ_φ is the phase noise (rad) of principal tone ring VCO.

5) Error caused by pulse jitter in case of range counter ON/OFF.

$$\sigma_{R_5} = \frac{\sqrt{2}}{2} c \sigma_T \quad (2.227)$$

where σ_T is the RMS value of pulse time jitter in case of counter ON/OFF.

6) Quantization error. The maximum error of the pulse measured time between range counter ON and OFF is ± 1 count clock, then

$$\sigma_{R_6} = \frac{c}{2} \sigma_T = \frac{c}{2} \frac{1}{\sqrt{6} f_R} \quad (2.228)$$

where f_R is the frequency of count clock.

7) Total random error.

$$\sigma_R = \sqrt{\sigma_{R_1}^2 + \sigma_{R_2}^2 + \sigma_{R_3}^2 + \sigma_{R_4}^2 + \cdots} \quad (2.229)$$

σ_R is determined by σ_{R_1} . If f_r and S/φ are given, the bandwidth of principal tone loop B_n may be reduced to satisfy the requirement of system random error.

(2) System error

The one-way system error that affects the equipment ranging accuracy has following error sources.

1) Ranging error introduced by variation of system frequency characteristics.

① In case of calculating with amplitude-/phase-frequency characteristic: In engineering design, if the variations of phase-frequency characteristic and amplitude-frequency characteristic are given, the expression below may be used to calculate:

The ranging error introduced by the variation of phase-frequency characteristic (see Sect. 2.4.2), without considering effect of amplitude-frequency characteristic:

$$\Delta R_1 = \frac{c}{720 f_R} \Delta(\Phi_+ - \Phi_-) \quad (2.230)$$

where Φ_+ is the phase shift of the phase-frequency characteristic corresponding to $(f_0 + f_R)$, Φ_- is phase shift of the phase-frequency characteristic corresponding to $(f_0 - f_R)$, and $\Delta(\Phi_+ - \Phi_-)$ is the variation of $(\Phi_+ - \Phi_-)$.

The ranging error caused by the change of amplitude-/phase-frequency characteristics is:

$$\tau = \frac{\phi}{\Omega} = \frac{1}{\Omega} \text{tg}^{-1} \frac{H_- \sin [(\phi_0 - \phi_-) + \Delta\phi] + H_+ \sin [(\phi_0 - \phi_-) - \Delta\phi]}{H_- \cos [(\phi_0 - \phi_-) + \Delta\phi] + H_+ \cos [(\phi_0 - \phi_-) - \Delta\phi]} \quad (2.231)$$

where τ is the time delay of ranging tone, Ω is the frequency of ranging tone, $\Delta\phi$ is the non-orthogonal phase difference by orthogonal demodulation, ϕ_0 is the phase-frequency characteristic value relative to the carrier frequency ω_0 , ϕ_+ is the phase-frequency characteristic value relative to the upper sideband of $(\omega_0 + \Omega)$, ϕ_- is the phase-frequency characteristic value relative to the lower sideband of $(\omega_0 - \Omega)$, H_+ is the amplitude-frequency characteristic value relative to the upper sideband of $(\omega_0 + \Omega)$, and H_- is the amplitude-frequency characteristic value relative to the lower sideband of $(\omega_0 - \Omega)$.

Use Expression (2.231) to calculate τ_0 and τ_1 respectively before/after change of (H_+ , H_- , ϕ_+ , ϕ_- , ϕ_0 , $\Delta\phi$); then the variation of one-way distance caused by change of time delay $\Delta\tau = (\tau_1 - \tau_0)$ is that

$$\Delta R_1 = \frac{c(\tau_1 - \tau_0)}{2} \quad (2.232)$$

② In case of calculating with group time-delay characteristic (see Sect. 2.4.3): In engineering design, if the variation of group time-delay characteristic is given, then the following expressions are used to calculate:

The ranging error caused by stability of group time delay is:

$$\Delta R_1 = \frac{C}{2} \left(\Delta a_0 + \frac{\Delta a_2}{3} \Omega^2 \right) \quad (2.233)$$

where a_0 is the zero-order term of group time-delay characteristic; a_1 and a_2 are the first-order term (linear distortion term) and the second-order term (parabola distortion term), respectively; Δa_0 and Δa_2 are the variations of a_0 and a_2 ; and Ω is the frequency of fine tone.

The ranging error caused by the change of Doppler frequency ω_d is:

$$\Delta R_1 = \frac{(a_1 \omega_d + a_2 \omega_d^2) c}{2} \quad (2.234)$$

where ω_d is the Doppler frequency shift (or the change of signal frequency and filter center frequency caused by other factors).

2) Dynamic error ΔR_2 . The dynamic error is the error that affected by the target velocity, acceleration, etc., which is mainly caused by the phase difference $\Delta\theta_e$ of the principal tone loop, and the expression is shown below:

$$\Delta R_2 = \frac{c}{4\pi f_R} \Delta\theta_e \quad (2.235)$$

For each type and exponent of phase-locked loop, $\Delta\theta_e$ is calculated with the expressions in Table 2.10 [14].

As shown in Table 2.10, B_L is the noise equivalent loop bandwidth of the (single-side) principal tone ring, α is the Doppler rate (Hz/S), β is the Doppler acceleration (Hz/S²), t is the time since Doppler acceleration starts (if the Doppler acceleration starts before the principal tone is locked, then t is the time since the loop is acquired and locked). If the Doppler acceleration appears continuously, then the periodic cycle skipping will be found in the Type II loop.

For the Type II loop shown in the table, the phase transmission function is:

$$H(s) = \frac{sK_1 + K_2}{s^2 + sK_1 + K_2} \quad (2.236)$$

Table 2.10 Phase error of principal tone ring $\Delta\theta_e$ (rad)

Loop	Constant range rate	Derivative of constant range rate	Second derivative of constant range rate
	Constant Doppler frequency shift	Constant Doppler rate	Constant Doppler acceleration
Type II loop with standard underdamping	0	$\frac{9\pi}{16B_L^2} \cdot a$	$\left\{ \frac{9\pi\beta}{16B_L^2} \right\} t - \frac{27\pi\beta}{64B_L^3}$
Type II loop with supercritical damping	0	$\frac{25\pi}{32B_L^2} \cdot a$	$\left\{ \frac{25\pi\beta}{32B_L^2} \right\} t - \frac{125\pi\beta}{128B_L^3}$
Type III loop with standard underdamping	0	0	$\frac{12,167\pi}{8,000B_L^3} \cdot \beta$
Type III loop with supercritical damping	0	0	$\frac{35,937\pi}{16,384B_L^3} \cdot \beta$

where $K_1 = (8/3)B_L$, $K_2 = (1/2)K_1^2$ (standard underdamping); $K_1 = (16/5)B_L$, $K_2 = (1/4)K_1^2$ (supercritical damping).

For the Type III loop shown in the table, the phase transmission function is:

$$H(s) = \frac{s^2 K_1 + s K_2 + K_3}{s^3 + s^2 K_1 + s K_2 + K_3} \quad (2.237)$$

where $K_1 = (60/23)B_L$, $K_2 = (4/9)K_1^2$, $K_3 = (2/27)K_1^3$ (standard underdamping); $K_1 = (32/11)B_L$, $K_2 = (1/3)K_1^2$, $K_3 = (1/27)K_1^3$ (supercritical damping).

In the Type II loop, because $\sigma_{R_1} \propto \sqrt{B_n}$ and $\Delta_{R_1} \propto (1/B_n^2)$, the requirements of the random error and the system error cannot be satisfied simultaneously under the limit conditions of threshold level, maximum speed $\dot{R}_{\max}$, and maximum acceleration $\dot{R}_{\max}$. The method below is recommended.

① The variable bandwidth method is applied to the principal tone loop. The flight vehicle has a large acceleration in low Earth orbit, but the s/φ value is much greater than the threshold value, so the bandwidth of the principal tone loop is widened to satisfy the requirements of σ_{R_1} and Δ_{R_1} . When the flight vehicle is operated in high Earth orbit, the s/φ is low, and the acceleration is small, so the bandwidth of principal tone loop is narrowed to satisfy the requirements of σ_{R_1} and Δ_{R_1} .

② If the Type III loop is taken as the principal tone loop, then the error caused by the Doppler rate of change becomes zero.

③ The group time delay of the principal input filter varies with the change of temperature and time, and then drift error Δ_{R_5} is introduced.

Table 2.11 lists the ranging errors caused by different phase changes of the principal tone.

Table 2.11 Effects on ranging errors by the input filter of principal tone loop

Principal tone (kHz)	Phase drift (°)	Ranging error (m)
100	1	4.17
27	1	15

Following methods may be used to minimize the error.

- A thermostat is added on the principal tone input filter.
- When the bandwidth of input filter if the SNR of principal tone input is satisfied, so as to even the phase characteristic of the filter.
- A digital filter is used. The error is fundamentally eliminated after using the digital ranging terminal.
- Zero residual Δ_{R_4} . Because the pattern of each factor that affects the system error is hard to obtained, as a result, the zero value of the equipment is hard to adjust, which consequently leads to ranging error.
- Ranging error Δ_{R_5} introduced by up-link. The phase jitter and drift caused by the phase modulator may change the phase of the baseline signal, and the amplitude-/phase-frequency characteristics (or group time delay) of the filter next to the modulator may cause the phase drift in case of being affected by temperature, so that the ranging error may be introduced. Its calculation has been described above.

The power amplifier is a wideband component that hardly affects the ranging error, however, the change of time delay shall be minimized in design.

- Ranging error Δ_{R_6} introduced by down-link. The ranging error may be caused by following reasons: phase jitter of carrier loop caused by thermal noise and combination intervene, change of receiver input level, the change of amplitude-/phase-frequency characteristics (or group time delay) of the down-link channel caused by changes of frequency and temperature and aging. Its calculation has been described above.
- Ranging error Δ_{R_8} caused by the change of antenna phase center. Radio ranging uses the center of antenna phase as reference; if the center is changed, it will cause ranging error.
- The total system error is calculated by square and extraction of a root, that is,

$$\Delta R_{\Sigma} = \sqrt{\Delta R_1^2 + \Delta R_2^2 + \Delta R_3^2 + \Delta R_4^2 + \Delta R_5^2 + \Delta R_6^2 + \Delta R_7^2 + \Delta R_8^2} \quad (2.238)$$

2.4.6 One-Way Ranging Technique

For the two-way ranging technique introduced above, the flight vehicle requires a transponder to cooperate with the ground station and to form an uplink/downlink loop to measure the delay difference between uplink and downlink signals, thus obtaining the range. In that method, the transponder increases the load on flight

vehicle, and the range error brought about by it is a main factor limiting the improvement of range accuracy in the current ranging system. The one-way ranging technique, however, needs no transponder, and so there's no such shortcoming to overcome.

2.4.6.1 One-Way Ranging by Interferometer

One-way ranging by interferometer refers to ranging by utilizing downlink signals. In this method, the flight vehicle transmits a random waveform signal, and the range then can be worked out by using the two range differences measured by three signal-receiving-only ground stations located in a straight line. This method is also called “Ranging by Interferometer” [15]. “One-way ranging” can be applied provided that the target is not too far from ground stations, and the radio waves which are allowed to be received by ground stations are either plane waves not spheric waves. That is to say, from the perspective of geometric optics, the three rays received by the three ground stations are not parallel to each other but form a triangle $\triangle A_1SA_2$, as shown in Fig. 2.58. Angle measuring by interferometer, however, is usually for targets at a far distance and the three rays received are assumed to be parallel to each other. This is exactly the main distinction between “ranging by interferometer” and “angle measuring by interferometer.”

In the left of Fig. 2.58, the space flight vehicle S is within the 3D coordinates of $OXYZ$ and the three ground stations A_0 , A_1 , and A_2 are on OY axis. The length of baseline between stations is B . A_0 is the master station and A_1 and A_2 are slave stations. The distances from S to A_0 , A_1 , and A_2 are R_0 , R_1 , and R_2 , respectively. Parameter to be measured is the distance R_0 between master station A_0 and target S . Take out the plane formed by S and OY axis from the 3D coordinates in Fig. 2.58 and a 2D coordinate system A_1SY is got, as shown in the right figure above. L is the vertical distance between S and OY axis. From the geometric relationship in the figure above, we can get:

$$R_0^2 = (B + b)^2 + L^2 = (B + b)^2 + (R_2^0 - b^2) = B^2 + 2Bb + R_2^2 \quad (2.239)$$

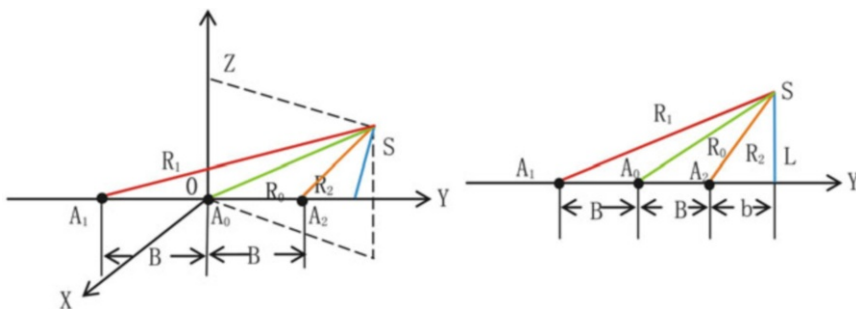


Fig. 2.58 Geometric relationship of rays for ranging by interferometer

Since

$$L^2 = R_1^2 - (2B + b)^2 = R_2^2 - b^2$$

Therefore,

$$2Bb = \frac{R_1^2 - R_2^2}{2} - 2B^2 \quad (2.240)$$

Substitute Expression (2.240) into Expression (2.239) to obtain:

$$R_0^2 = \frac{R_1^2 + R_2^2}{2} - B^2 \quad (2.241)$$

Expression (2.241) shows the relationship between distances R_0 and $(R_1 + R_2)$. However, distance $(R_1 + R_2)$ is hard to measure in engineering, so it is necessary to transform Expression (2.241) to a range difference equation, making it easy for measurement in engineering. From Expression (2.241), we get:

$$\begin{aligned} 2R_0^2 &= R_1^2 + R_2^2 - 2B^2 = [(R_1^2 - 2R_0R_1 + R_0^2) - (-2R_0R_1 + R_0^2)] \\ &\quad + [(R_2^2 - 2R_0R_2 + R_0^2) - (-2R_0R_2 + R_0^2)] - 2B^2 \\ 2R_0^2 &= (R_1 - R_0)^2 + (R_2 - R_0)^2 + 2R_0R_1 + 2R_0R_2 - 2R_0^2 - 2B^2 \\ 2R_0(2R_0 - R_1 - R_2) &= (R_1 - R_0)^2 + (R_2 - R_0)^2 - 2B^2 \\ R_0 &= \frac{(R_1 - R_0)^2 + (R_2 - R_0)^2 - 2B^2}{2[(R_0 - R_1) + (R_0 - R_2)]} = \frac{2B^2 - [(R_1 - R_0)^2 + (R_2 - R_0)^2]}{2[(R_1 - R_0) + (R_2 - R_0)]} \end{aligned} \quad (2.242)$$

Expression (2.242) is the one for solving range R_0 by using range difference. There are two methods for measuring range difference:

(1) Delay difference measurement method

Convert range difference into delay difference:

$$(R_1 - R_0) = (\tau_1 - \tau_0)C = \Delta\tau_{10}C, (R_2 - R_0) = (\tau_2 - \tau_0)C = \Delta\tau_{20}C$$

Substitute the equation above into Expression (2.242) to obtain:

$$R_0 = \frac{2B^2 - [\Delta\tau_{10}^2 + \Delta\tau_{20}^2]C^2}{2[\Delta\tau_{10} + \Delta\tau_{20}]C} \quad (2.243)$$

The delay difference can be measured by using the marking signals of downlink signals, such as the frame synchronization signal of telemetry signal, “all 1” signal of PN code signal, etc.

(2) Phase difference measurement method

To have a higher measurement accuracy, the delay difference can be measured by utilizing the phase difference of carriers. The relationship between delay difference and phase difference is as below:

$$\Delta\tau_{10} = \frac{\Delta\varphi_{10}}{\omega_0} \quad \Delta\tau_{20} = \frac{\Delta\varphi_{20}}{\omega_0}$$

Substitute the equation above into Expression (2.243) to obtain:

$$R_0 = \frac{2B^2 - [(\Delta\varphi_{10}^2 + \Delta\varphi_{20}^2)/\omega_0^2]C^2}{2[(\Delta\varphi_{10} + \Delta\varphi_{20})/\omega_0]C} \quad (2.244)$$

Although carrier phase difference measurement method can get a higher measurement accuracy, it can easily cause phase ambiguity. In that case, the value measured by “delay difference” can be used for resolving such ambiguity. So it is suggested that “delay difference” be used for coarse ranging and “carrier phase difference” for fine ranging.

The measurement accuracy of one-way ranging method relates to the length and accuracy of baseline, delay and phase-shift stability of measuring instruments, wave length and signal form of measuring signal, as well as propagation characteristics of radio waves. Delay, phase calibration, and wired baseline transmission (connected end interferometry) are usually the methods for improving ranging accuracy. With the same delay difference measuring accuracy, the ranging accuracy decreases with the increase of range and deviation angle of target.

2.4.6.2 One-Way Ranging Technique by the Time Synchronization Between Satellite and Earth

The two-way ranging technique aforesaid is to generate the ranging signal to be transmitted based on the clock of ground station, and to measure the two-way delay of the received/transmitted signals based on the same clock to obtain higher ranging accuracy. If the satellite and Earth have high-accuracy synchronization in time, the ranging signal to be transmitted then can be generated by the satellite and the ground station can utilize the highly synchronized clock signal to measure the one-way delay, thus obtaining the one-way range. However, error in the time synchronization between satellite and ground station may directly lead to range error, so its engineering application will significantly influence the development of satellite-Earth high-accuracy synchronization technology. This method is really a very promising ranging technique.

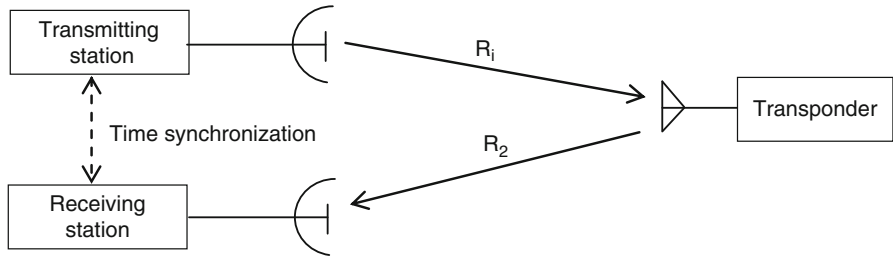


Fig. 2.59 Schematic diagram of a three-way ranging

2.4.7 Three-Way Ranging Technique

For the two-way ranging technique described above, ranging signals are transmitted and received by the same ground station. The technique allowing ranging signals to be transmitted by one ground station and received by another is called three-way ranging. Its working principle block diagram is as shown in Fig. 2.59.

As shown in Fig. 2.59, high-accuracy inter-station synchronization technology is utilized to synchronize the time between receiving and transmitting stations. Assuming that the synchronization error is $\Delta\tau$ and if the time synchronized in the receiving station is used as the reference time for measuring the delay of downlink signal, the delay τ measured by receiving station will be as below:

$$\tau = \frac{R_1}{C} + \frac{R_2}{C} + \Delta\tau \quad (2.245)$$

$$(R_1 + R_2) = \tau C - (\Delta\tau)C = C(\tau - \Delta\tau)$$

As shown in Expression (2.245), the value measured is the sum of ranges.

The accuracy of three-way ranging is lower than that of two-way ranging for the following reasons:

(1) The time synchronization error $\Delta\tau$ between stations: usually solved by utilizing high-accuracy inter-station synchronization. The main solutions include time synchronization by common-view GPS satellite and time synchronization by two-way transmission comparison.

(2) The ranging error caused by the uplink delay of transmitting station equipment and the downlink delay of the receiving station equipment: for two-way ranging, the uplink and downlink equipment delays are generated in the same ground station and can be deducted through range calibration; while for three-way ranging, however, the equipment delays exist in the one-way delay of two stations respectively, which makes three-way range calibration impossible. One solution is that the two stations calibrate their own two-way equipment delay $\Delta\tau_1$

and $\Delta\tau_2$ respectively and then work out the average uplink delay and downlink delay of the two station equipment: $(\frac{\Delta\tau_1 + \Delta\tau_2}{2})$, that is,

$$\begin{aligned}\Delta\tau_1 &= (\Delta\tau_{\text{uplink}})_1 + (\Delta\tau_{\text{downlink}})_1 \\ \Delta\tau_2 &= (\Delta\tau_{\text{uplink}})_2 + (\Delta\tau_{\text{downlink}})_2 \\ \frac{\Delta\tau_1 + \Delta\tau_2}{2} &= \frac{(\Delta\tau_{\text{uplink}})_1 + (\Delta\tau_{\text{uplink}})_2}{2} + \frac{(\Delta\tau_{\text{downlink}})_1 + (\Delta\tau_{\text{downlink}})_2}{2}\end{aligned}\quad (2.246)$$

Obviously, this can only reach the approximate range calibration. If the consistency between the two station equipment is good, the calibration error will be small accordingly.

Three-way ranging technique is usually used in deep-space TT&C. Under that circumstance, the receiving/transmitting signal delay is very long, and due to the Earth's rotation, the antenna beam of a ground station cannot receive its return signal after it is transmitted, and when the signal returns to the Earth's surface, it is needed to have another beam to receive it, thus realizing three-way ranging.

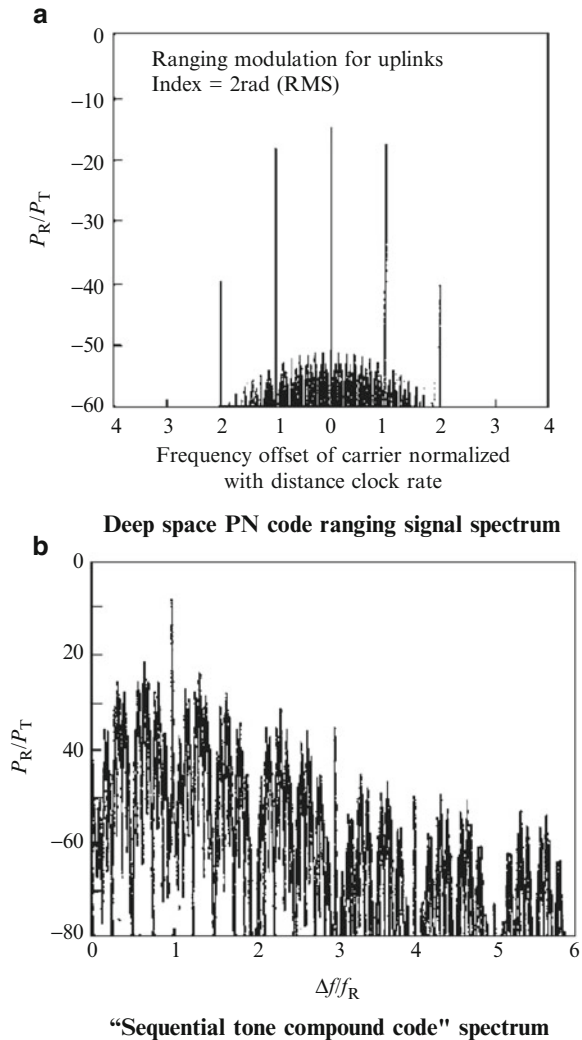
2.4.8 Deep-Space Ranging System

Currently, three main systems are used for deep-space ranging, that is, sequential tone ranging, PN code ranging, and tone and code mixed ranging. The analytical results for each system are discussed below.

(1) Sequential tone ranging [12, 16, 17]. The signal used for sequential tone ranging is a correlated tone sequence with frequency ratio of 2. It has one ranging clock and multiple ambiguity-resolving tones, of which the previous one decides the accuracy, and the latter determines the distance without ambiguity. When a tone sequence is transmitted, every tone may last for a while so that the correlated integration is conducted to improve the SNR. Under operating condition, a ranging data can be obtained by transmitting a tone sequence. Therefore, n ranging data will be obtained by transmitting periodically for n times. Strictly speaking, sequential tone ranging is discrete ranging. Due to that the ranging data is determined by the code clock; thus, the method aforesaid may be used to calculate the ranging error by substituting the code clock for the tone.

(2) PN code ranging [12, 16]. The signal used for deep-space PN code ranging is a compound code ranging signal logically combining a ranging clock and several PN codes ("and" and "or" logic), wherein the ranging clock decides the accuracy, and the compound code resolves the distance ambiguity. Because such signal is a continuous signal, so that PN code ranging is continuous ranging. For such PN code signal only used for ranging, the compound code ranging signal is designed to obtain higher accuracy and longer distance ambiguity, that is, the major part of power of a signal concentrates on the ranging clock for improving the ranging

Fig. 2.60 Deep-space ranging signal spectrum.
 (a) Deep-space PN code ranging signal spectrum.
 (b) “Sequential tone compound code” spectrum



accuracy and acquiring signal easily, therefore, such signal basically is a periodic sequence with a period of 2 chips. In case of resolving ambiguity, other PN codes make the ranging clock to reverse in a half-period occasionally, but such reversion is not happened frequently. One specific example for such compound PN code is a regenerative compound code for ranging used in the Deep-Space Network of the United States. The frequency spectrum of the code is as shown in Fig. 2.60a [12].

As shown in Fig. 2.60a, P_T is the total power of signal, P_R is the power of ranging signal, f_R is the frequency of ranging clock, and the Δf is the frequency offset relative to the center frequency of carrier.

From Fig. 2.60a, the ranging clock has a great first-order sideband component, and the spectrum line between the carrier frequency and the second-order component is small, the reason is that the component for resolving ambiguity is produced by the semi-cycle conversion occurred occasionally with the compound PN code.

The main spectrum component of ranging signal still is the ranging clock component, and other components rarely affect the ranging clock. Therefore, the method aforesaid can be remained for estimating the ranging error.

The important thing is that the method aforesaid is not applicable to spread spectrum TT&C system. Because the spread spectrum PN code is not only used for ranging but also used for anti-jamming, anti-interception, code division multiple access (CDMA), etc., so that the spectrum of the ranging signal is similar with that of “white noise.”

(3) Tone and code mixed ranging. Here is a typical example presented in *ESA Standards* (ESA PSS-04-104). For tone and code mixed ranging, the ranging signal is such signal that obtained by 45° phase modulation with one Sequential tone code against the ranging clock (i.e., high tone), wherein the ranging clock determines the ranging accuracy, and the Sequential tone code is used for resolving the distance ambiguity, it can be expressed as

$$C_n = Q_1 \oplus Q_2 \oplus Q_3 \oplus \cdots \oplus Q_n$$

where C_n is the n th code, $\oplus$ represents for XOR logic, and Q_i stands for the two-divided frequency sequence square wave generated by ranging clock.

From this expression, the code is compounded by the Sequential tones, not the PN codes. The difference between such code and the Sequential tone is that, since the ranging clock always exists, the code is not acquired by searching using relevant functions but acquired in sequence by modulation of different tones; therefore, this kind of code acquisition is more simple and reliable, which is the reason why such a code is better than PN compound code. The spectrum of such ranging signal after distance acquisition is shown in Fig. 2.60a.

Observed from Fig. 2.60b, because the ranging clock only undertakes 45° phase modulation, therefore, residual clock component is great, so that other components hardly affect the ranging clock. Consequently, the method aforesaid can be used for estimating the ranging error.

The analytical results obtained above may be used to calculate the ranging error caused by group delay. If a_1 , a_2 , τ_m , Ω_T , and θ_0 are given, the ranging error may be calculated with equation; if the group delay characteristic curve is given, the ranging error may be calculated in graphic way. The latter is more accurate. For PN code ranging signal, the method aforesaid also can be used to calculate the ranging error only when the component of code clock is the main component of ranging signal spectrum.

2.5 Angle Measurement Theory and Technology

The angle measurement using antenna-tracking and the interferometer angle measurement are usually adopted for TT&C system. For angle measurement using antenna-tracking, the antenna-feed-servo subsystem is used to automatically track the target, so that the electronic boresight of antenna aims to the target correctly, then the displayed value of the antenna pedestal position, after compensation, is the angle measured value. It is a common method used for TT&C system. However, this method has limited measurement accuracy due to several factors including the processing accuracy of antenna-feed-servo system, antenna rotating and tracking receiver. In this case, the interferer system shall be adopted for further improving the accuracy of the height and angle measurement. The reason is that the aperture of a single antenna is limited, but the interferer may improve the accuracy of the height and angle measurement by extending the distance (or referred to as “baseline”) between the two measurement units (or referred to as “terminal station”). Moreover, with the development of electrical measuring technology, the measuring element of the phase interferer is the phase or time delay of a signal, consequently, the angle measurement in high accuracy and long distance recently adopts the Very Long Baseline Interferometry (VIBI) technology.

Angle tracking is a technology combining auto control and wireless radio, in which a lot of edge technical problems exist. We will discuss the specific problems below.

2.5.1 Angle Measurement Using Antenna Tracking – Three-Channel, Dual-Channel, and Single-Channel Monopulse Technologies

Since the conical scanning radar was invented in 1942, its reliability and accuracy always were the major issues for further improvement. The rotating joint and the conical scanning motor of the radar are the main weak links for its reliability. In the 1950s, a three-channel monopulse system is presented. It significantly improved the accuracy and reliability, making the precision tracking radar entered into a new era. This system is still in use today but complicated. After that, a dual-channel monopulse (DCM) system which was easier than the three-channel monopulse system was presented and often been used in recent TT&C system. However, it has some problems in cross coupling and angle error caused by phase shift inconsistency of the sum and difference channels, and still complicated. Then a single-channel monopulse (SCM) system was presented in 1960s, which simplified the system and was more reliable than conical scanning but had a poorer accuracy than the two systems aforesaid. Therefore, most medium-accuracy systems (e.g., telemetry receiver, guidance system, etc.) adopted this system from the 1960s to 1970s, but it also had some issues such as conversion loss caused by the single-channel

converter, lower tracking accuracy and higher sidelobe. Then, the radiator scanning radar system (RADSCAN), a new conical scanning system, was appeared in 1982. It uses one waveguide radiator (or bevel waveguide), which is deviated from the rotating axis, to move around the rotating axis, so as to produce a conical scanning beam. It has not the defects that the single-channel converter in SCM has. Due to the transmitter weighs light (or even can be made of carbon fiber), so the rotating motor only requires a small driving power. The brushless DC motor is recommended with some advantages like high reliability, high G/T value, lower sidelobe and light weight. Since then, several angle tracking methods similar with such “sequence lobe” system had been presented, for example, electronic scanning (ESCAN) antenna, etc.

The amplitude comparison and the phase comparison are included in simultaneous lobe angle measurement system. For amplitude comparison, it has three-channel monopulse, dual-channel monopulse, single-channel monopulse, conical scanning, and amplitude extreme tracking. For phase comparison, it includes phase monopulse and interferer system.

Some angle tracking systems that are commonly used in TT&C system are discussed below.

2.5.1.1 Three-Channel Monopulse (TCM) System

The working principles of amplitude comparison and equisignal method are normally used for this system, as shown in Fig. 2.61.

For measuring the angle of the two planes in azimuth and elevation, the four antennas (or feed horn) shown in Fig. 2.61a, b may be used to measure it with amplitude measurement. Theoretically, the sum signal E_{Σ} produced by the comparator, the error signal E_A of azimuth angle and the error signal E_E of the elevation angle form the amplitude monopulse diagram and the sum-difference directional diagram as shown in Fig. 2.61d, e respectively. Because there are three signals, E_{Σ} , E_A , and E_E , shall be processed, it is so called TCM system. In Fig. 2.61c, beam 1 and beam 2 are same but overlapped in part, and the line from the intersecting point of the two beams to the origin O is the equisignal line (or referred to as electronic boresight). When the target places on the OA line, the two beams receive the signals with same amplitude and phase, it gets zero by subtracting, which is the origin point of angle tracking. The equisignal method has a higher accuracy of angle measurement, but the antenna feeds are more complicated. Two beams may be alternatively produced by sequential rotating of one lobe (sequential lobing), or simultaneously produced by two lobes (simultaneously lobing).

The block diagram of tracking servo system is as shown in Fig. 2.61f. The angle error signal obtained using the difference directional pattern is converted into the DC error signal through processing and modulation in receiver, and its voltage amplitude is in direct proportional to the azimuth (or elevation angle) that the antenna deviates from the target. Then such signal is transmitted to the servo system for amplification and correction, driving the execution mechanism consisting of

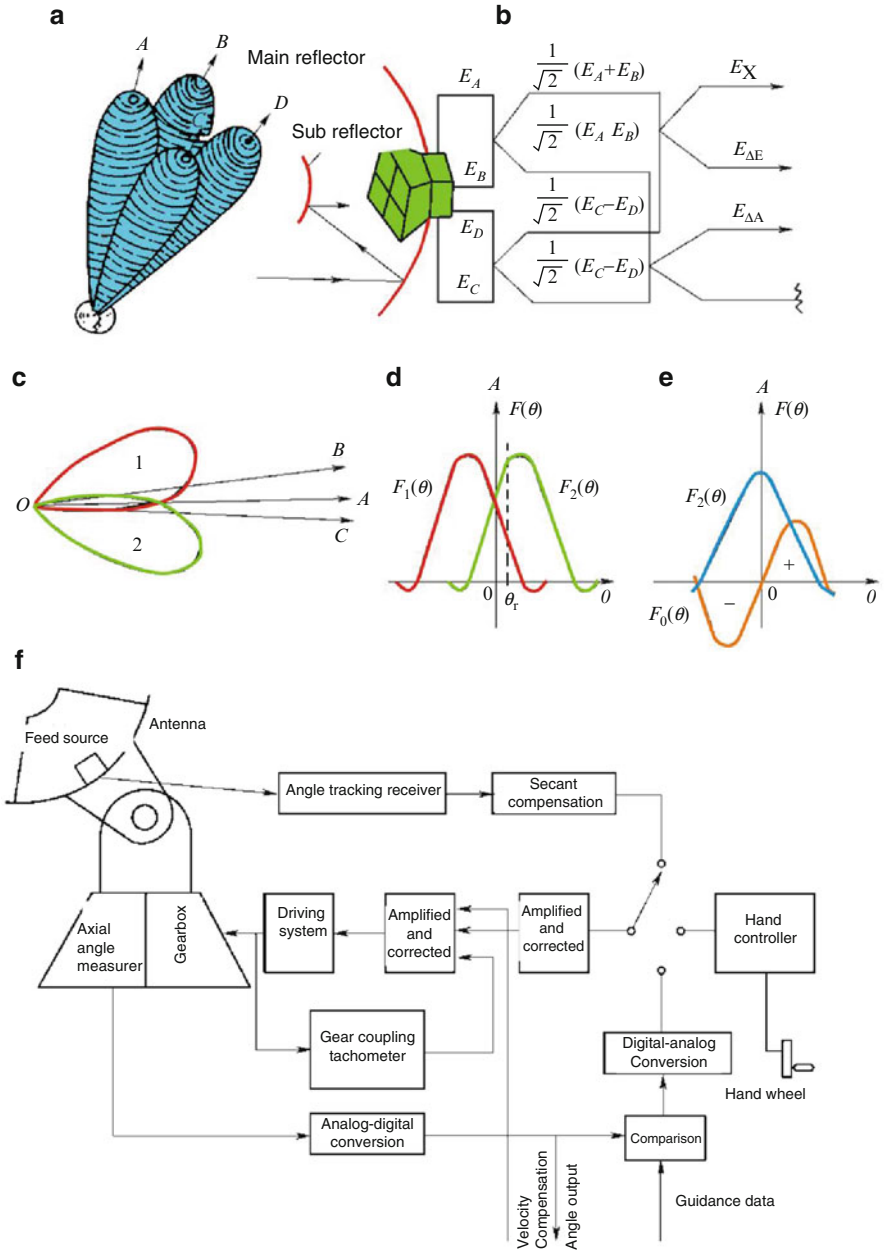


Fig. 2.61 Three-channel monopulse system concept diagram

motor and reduction gearbox to make the antenna to move along the direction of reducing sighting error, so that the electronic boresight of antenna points the target. For angle tracking, the error that the target deviates from the electronic boresight is referred to as tracking error. The intersection point of the antenna azimuth and

elevation rotating shafts is considered as the reference point, then the rotating position of the antenna azimuth shaft and the elevation shaft reflects the angle of target. Therefore, for angle measurement, the shaft angle measurer installed on the rotating shaft of antenna is used to measure the target angle in the coordinate system, and the measuring error is referred to as angle measurement error.

2.5.1.2 Dual-Channel Monopulse System (DCM)

For simplifying system, the azimuth error signal and the elevation error signal in TCM are usually combined as one signal with the orthogonal phase multiplexing method, which becomes the dual-channel monopulse system. As shown in Fig. 2.62, $E_{\Delta E}$ and $E_{\Delta A}$ are combined to E_{Δ} , which can be realized by multiple horn feeds and 90° phase shifter, or by multimode feed with circular waveguide TE_{11} and TE_{21} .

As shown in Fig. 2.62, K_{Σ} and Φ_{Σ} , respectively, are the gain and phase shift of the sum channel, and K_{Δ} and Φ_{Δ} are the gain and phase shift of the sum channel, which are normalized to the amplitude with AGC in the sum channel and the orthogonal phase are divided by two coherent multipliers with phase difference of 90° to modulate the azimuth error signal and the elevation error signal. The reference signal of the multiplier is coherent to the sum-channel signal with a phase-locked loop. If the amplitude of the sum-channel signal is set to 1:

$$E_{\Delta} = \mu\theta_A \cos \omega_0 t + \mu\theta_E \cos \left(\omega_0 t + \frac{\pi}{2} \right) = \mu\theta_A \cos (\omega_0 t - \varphi) \quad (2.247)$$

$$E_{\Sigma} = \cos \omega_0 t$$

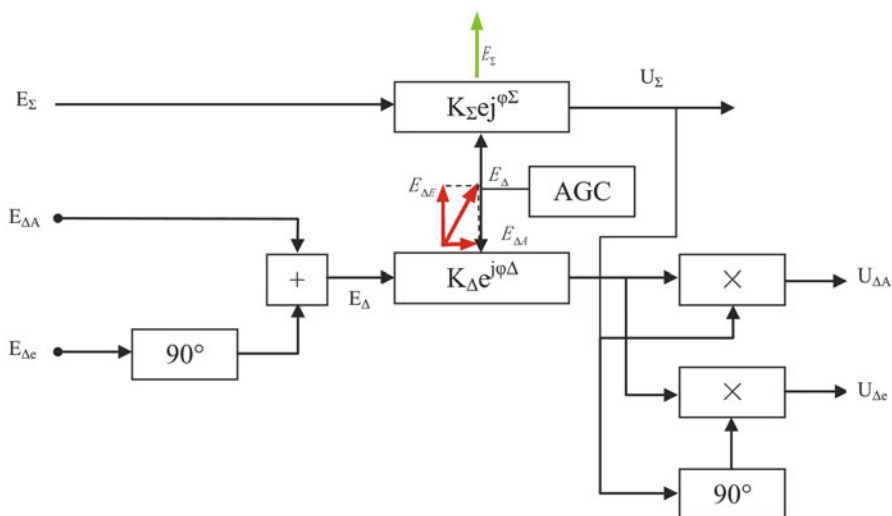


Fig. 2.62 Schematic diagram of dual-channel monopulse

where θ_A and θ_E are the azimuth deviation angle and the elevation deviation angle, respectively, and μ is the slope of relative angle error of the antenna.

$$\theta = \sqrt{\theta_A^2 + \theta_E^2}, \quad \Phi = \arctan \frac{\theta_E}{\theta_A}$$

where Φ is the phase of combined carrier and θ is the included angle between the target and the electronic boresight of antenna.

After magnification and AGC normalization through respective channel,

$$U_\Delta = \frac{K_\Delta}{K_\Sigma} \mu \theta_A \cos \omega_0 t + \frac{K_\Delta}{K_\Sigma} \mu \theta_E \cos \left(\omega_0 t + \frac{\pi}{2} \right) \quad (2.248)$$

$$U_\Sigma = \cos(\omega_0 t + \Delta\varphi)$$

where $\Delta\varphi = \varphi_\Sigma - \varphi_\Delta$ is the phase inconsistency of the sum channel and the difference channel when the difference channel is taken as a reference channel. The reference signal is coherent with the sum-channel signal, i.e., $\cos(\omega_0 t + \Delta\varphi + \frac{\pi}{2})$, and when it is multiplied with the difference-channel signal U_Δ by LPF, we can obtain that

$$U_{\Delta A} = \frac{K_\Delta}{K_\Sigma} \mu \theta_A \cos \Delta\varphi + \frac{K_\Delta}{K_\Sigma} \mu \theta_E \sin \Delta\varphi \quad (2.249)$$

According to the expression above, the first term is the effective azimuth error output signal, and there is a second term if $\Delta\varphi$ exists, which is related to the elevation deviation angle, making the azimuth error signal has the elevation deviation signal; consequently, a cross coupling will be formed for elevation against azimuth. Therefore, when the antenna is tracking a target, the spiral tracking will be produced with slow velocity of convergence and poor dynamic tracking performance. After 90° phase shift of reference signal, $\cos(\omega_0 t + \Delta\varphi + \frac{\pi}{2})$ is multiplied with the difference channel signal U_Δ through LPF, then we can obtain that

$$U_{\Delta E} = \frac{K_\Delta}{K_\Sigma} \mu \theta_E \cos \Delta\varphi + (K_\Delta/K_\Sigma) \mu \theta_A \sin \Delta\varphi \quad (2.250)$$

According to the expression above, the first term is the effective elevation error output signal, and when the phase shift of the sum channel and the difference channel are different, the cross coupling signal for azimuth against elevation will be produced. The cross coupling for azimuth against elevation K_{cE} is defined as: when $\theta_E = 0$ but θ_A exists, then the ratio of the cross-coupling interfering signal output from the elevation channel and the output signal of the azimuth channel is obtained from Expressions (2.249) and (2.250):

$$K_{cE} = \tan \Delta\varphi \quad (2.251)$$

And the cross coupling for elevation against is also be solved in a same way with the identical expressions.

Even though the dual-channel monopulse system has the cross coupling, but it simplifies the receiver, with the simple sum-difference signals produced by multimode feeds, so this system is widely applied in TT&C system.

If the sum phase and the difference phase have a good consistency, then the accuracy of angle measurement in DCM is similar with that in TCM, where the tracking error caused by thermal noise is:

$$\sigma_1 = \frac{\theta_{0.5}}{K_m \sqrt{\frac{S}{N} \frac{B_{1F}}{B_n}}} = \frac{1}{\mu \sqrt{\frac{S}{N} \frac{B_w}{B_n}}} \quad (2.252)$$

where $\theta_{0.5}$ is the beam width of half power point, S/N is the SNR of receiver (bandwidth is B_{1F}), B_{1F} represents the equivalent noise bandwidth of medium-frequency amplifier of the receiver, B_n is the equivalent noise bandwidth of the servo, K_m represents the normalized difference slope (generally ranging between 1.2 and 1.7), $K_m = K(\theta_{0.5}/\sqrt{G_\Sigma})$, $K = (\partial F(\theta)/\partial \theta)|_{\theta=0^\circ}$ is the absolute slope of difference beam voltage gain of the antenna (when $\theta = 0^\circ$), which is determined by the antenna aperture, wherein, $F(\theta)$ is the difference voltage pattern, G_Σ is the gain of sum beam, $\sqrt{G_\Sigma}$ is the sum beam voltage gain, so that K_m is defined as: the slope of the difference beam voltage gain is normalized by the sum beam voltage gain, with measurement unit of 3 dB beam width. (Note: Another normalization difference slope $\mu = \left[\frac{\partial F(\theta)}{\partial \theta} / \sqrt{G_\Sigma} \right] = K_m / \theta_{0.5}$ is often used in antenna. Do not confuse the former one with the latter).

$K = K_m \sqrt{G_\Sigma} / \theta_{0.5}$ is obtained from K_m expression, then the voltage gain of the difference beam deviated from the electronic boresight of $\Delta\theta$ is $\sqrt{G_\Sigma} = \Delta\theta K$, and the voltage gain ratio of the sum beam and the difference beam is $\sqrt{G_\Delta} / \sqrt{G_\Sigma} = K_m (\Delta\theta / \theta_{0.5})$, and then the voltage of the receiver's sum/difference input signal may be simulated base on these expressions.

2.5.1.3 Single-Channel Monopulse System (SCM)

For further simplifying the system, the dual-channel or the three-channel described above are combined as single-channel monopulse respectively. The combination of the azimuth error signal and the elevation error signal is based on time division multiplexing, while the combination of the angle error signal and the sum-channel signal is normally implemented by conducting $(0/\pi)$ phase modulation and then adding sum-channel signal to generate the amplitude modulation signal.

More and more equipment provided with single-channel monopulse tracking system are used in the system that has low accuracy requirement for tracking system, such as business TT&C station, RSGS, satellite reconnaissance signal receiving station, and telemetering ground station.

(1) Dual-channel monopulse combines as single-channel monopulse: On the basis of dual-channel monopulse shown in Fig. 2.62, the difference signal is processed by quadri-phase modulation with square wave modulation signal and then combines

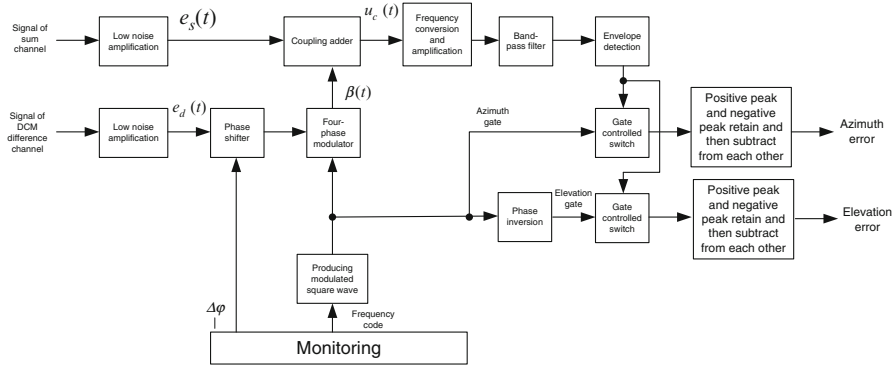


Fig. 2.63 Schematic block diagram of combination of DCM into SCM

with sum signal, thereby becoming one channel output signal, that is, a combined signal. The error voltage for such signal may be obtained only through envelope detection [18]. Since the channels are combined, the phase and gain after combination of sum channel and difference channel will not be inconsistent for this system. All equipment except LNA are allowed to be installed at the generator room, so as to improve the reliability, usability, and maintainability of the equipment; meanwhile, the cost is greatly decreased with reduction of equipment numbers.

Refer to Fig. 2.63 for the theory of combining dual-channel monopulse into single-channel monopulse. For ensuring the G/T value of the system, the combination of the sum signal and difference signal shall be implemented after LNA.

Suppose the signal from the antenna feed is single-frequency signal, after the feed output and the signal instantaneous value is amplified for K_Σ through LNA, then we can obtain that

$$e_s = K_\Sigma A m \cos \omega t$$

The difference-channel signal after combination of the two phase-quadrature channels is amplified for K_Δ through LNA, and then we can obtain that:

$$e_d = (\mu\theta) K_\Delta A m \cos(\omega t - \Phi)$$

$$\varphi = \arctan \frac{A}{E}$$

After four-phase modulation, the difference signal is:

$$e_d = (\mu\theta) K_\Delta A_m \cos(\omega t - \Phi + \beta(t) + \Delta\varphi) \quad (2.253)$$

where E is the elevation deviation angle, A is the azimuth deviation angle, $\Delta\varphi$ is the relative phase difference of the sum signal and the difference signal, and $\beta(t)$ is the time-varying function of the phase modulation angle, which is controlled by the square wave modulation signal $\alpha(t)$ to produce the change of phase $\beta(t)$; it is expressed as below in four-phase modulation:

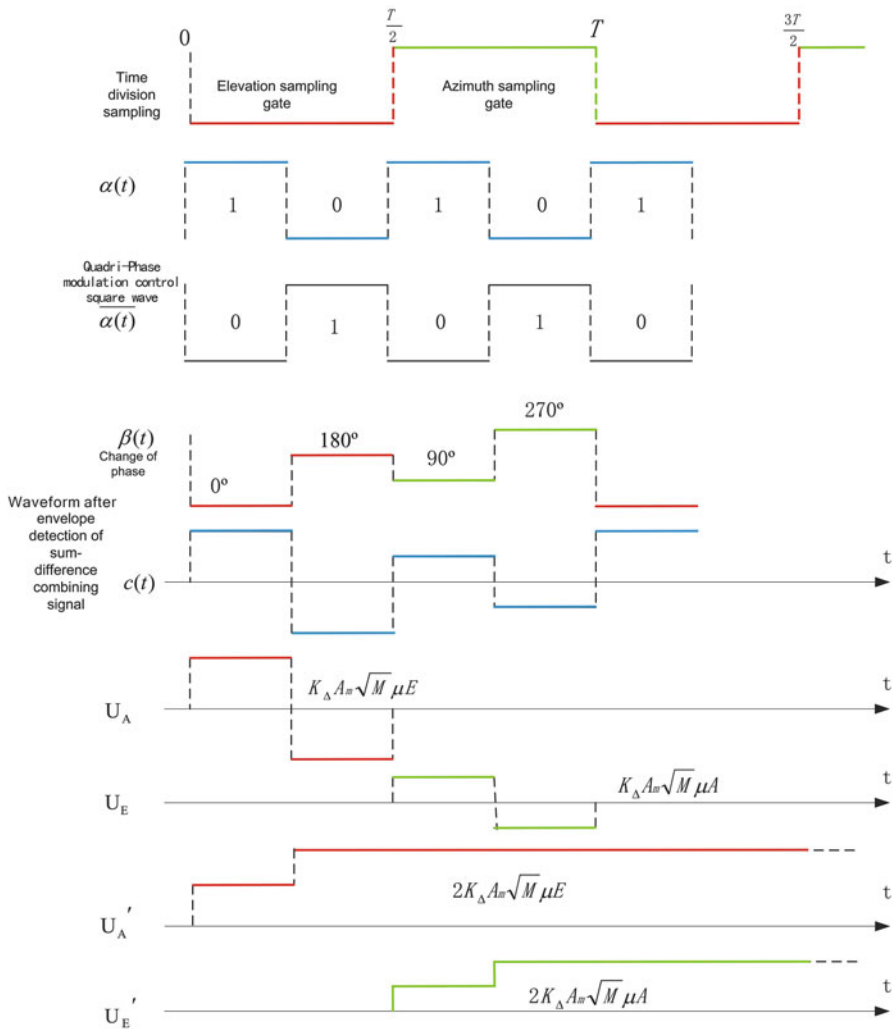


Fig. 2.64 Synchronous demodulation waveform of single-channel monopulse angle error signal

$$\beta(t) = \begin{cases} 0 & t = 0 - \frac{T}{4} \\ \pi & t = \frac{T}{4} - \frac{T}{2} \\ \frac{\pi}{2} & t = \frac{T}{2} - \frac{3T}{4} \\ \frac{3\pi}{2} & t = \frac{3T}{4} - T \end{cases}$$

The change of phase is as shown in Fig. 2.64.

After modulation, the difference signal is combined with the sum signal by the directional coupler with power coupling coefficient of M , then the combined signal is expressed as:

$$\begin{aligned} u_c &= K_\Sigma A_m \cos \omega t + K_\Delta A_m \sqrt{M}(\mu\theta) \cos [\omega t - \Phi + \beta(t) + \Delta\varphi] \\ &= c(t) \cos [\omega t + \psi(t)] \end{aligned} \quad (2.254)$$

The voltage of angle error that is in relation to the amplitude $C(t)$ of $u_c(t)$ is extracted by envelope detector. Using the formula $a \cos \omega t + b \cos(\omega t + \alpha) = a[1 + (b^2/a^2) - (2b/a)\cos(180^\circ - \alpha)]\cos[\omega t + \psi(t)]$ to expand Expression (2.254), and as θ is a small value under tracking situation, we can obtain the following expression through derivation:

$$C(t) = K_\Delta A_m \sqrt{M}(\mu\theta) \cos [\beta(t) - \Phi + \Delta\varphi] \quad (2.255)$$

Within one period of four-phase modulation, according to different value of $\beta(t)$, then we can obtain that:

$$C(t) = \begin{cases} K_\Delta A_m \sqrt{M} \mu \theta \cos [\Phi - \Delta\varphi] & t = 0 - \frac{T}{4} \\ -K_\Delta A_m \sqrt{M} \mu \theta \cos [\Phi - \Delta\varphi] & t = \frac{T}{4} - \frac{T}{2} \\ K_\Delta A_m \sqrt{M} \mu \theta \sin [\Phi - \Delta\varphi] & t = \frac{T}{2} - \frac{3T}{4} \\ -K_\Delta A_m \sqrt{M} \mu \theta \cos [\Phi - \Delta\varphi] & t = \frac{3T}{4} - T \end{cases} \quad (2.256)$$

In equipment calibration, the relative phase difference before combination of the sum signal and the difference signal shall be adjusted to close to 0, i.e., $\Delta\varphi \approx 0$. As $E = \theta \cos \Phi$, $A = \theta \sin \Phi$, then Expression (2.251) becomes:

$$\begin{aligned} t = 0 - \frac{T}{4} \quad U_{E1} &= K_\Delta A_m \sqrt{M} \mu E \\ t = \frac{T}{4} - \frac{T}{2} \quad U_{E2} &= -K_\Delta A_m \sqrt{M} \mu E \\ t = \frac{T}{2} - \frac{3T}{4} \quad U_{A1} &= K_\Delta A_m \sqrt{M} \mu A \\ t = \frac{3T}{4} - T \quad U_{A2} &= -K_\Delta A_m \sqrt{M} \mu A \end{aligned} \quad (2.257)$$

As shown in Expression (2.252), the elevation error signal E is transmitted by segmenting within $(0 - T/2)$; similarly, the azimuth error signal A is transmitted by

segmenting within $(T/2 - T)$, so as to implement time division multiplexing of dual-channel. Meanwhile, the servo system for angle tracking is becoming a discrete self-regulation system, which approximates a continuous self-regulation system only if $1/T$ is greater than (5–10) times of self-regulation loop bandwidth and the peak-holding circuit is used to retain the error voltage within $T/2$ interval. As shown in Fig. 2.63, the “gate controlled switch” is used for time division demultiplexing, and the “peak-holding circuit” is used to convert the discrete signal into continuous signal, and subtract the positive peak voltage from the negative peak voltage, so as to double the error voltage. After processing, the output of elevation error angle is:

$$U_E = 2K_{\Delta}A_m\sqrt{M}\mu E \quad (2.258)$$

Similarly, the output of azimuth error signal is:

$$U_A = 2K_{\Delta}A_m\sqrt{M}\mu A \quad (2.259)$$

If there has a time-delay error Δt between the modulating square wave and the demodulated envelop's square wave U_A , the video cross coupling will be produced. Consequently, Δt must be aligned to 0 at factory acceptance test. Because the value of Δt changes very little, there is no need to calibrate again on site.

The other kind of cross coupling is “RF cross coupling,” which is caused by phase shift inconsistency $\Delta\varphi$ of DCM sum/difference RF signals before combination in SCM, and the result is shown as in Expression (2.251).

Meanwhile, $\Delta\varphi$ causes the loss of error signal, i.e., $L_{\Delta} = \cos \Delta\varphi$; as a result, the SNR of error signal is decreased, thus making the error of angle tracking increase.

For mitigating the effect on angle tracking of these two factors, the RF phase calibration shall be performed, to make $\Delta\varphi = 0$. If the index can be controlled to enable that $\Delta\varphi \leq \pm 14^\circ$ is true under all conditions, then $K_{\Delta} = \tan 14^\circ = 0.25$, $L_{\Delta} = \cos 14^\circ = 0.9703$, so that RF calibration shall not be performed in case of lower tracking accuracy. As shown in the result, the cross coupling is the main affecting factor.

For both DCM and SCM, the phase inconsistency between sum channel and difference channel will cause cross coupling and static tracking error. The reason is: when the phase difference between the sum and difference signals is $\Delta\varphi = 90^\circ$, there will be no DC error signal output represented by first term of Expression (2.249) even when there is input of IF signal in the difference channel, which will bring in a static tracking error corresponding to the IF input in difference channel; when the phase difference between the sum and difference signals is $\Delta\varphi \neq 0^\circ$, there is a DC error signal output represented by first term of Expression (2.249), which will cause a cross coupling. For DCM, its error detector is a phase discriminator with discrimination characteristic of $\cos(\Delta\varphi)$. If the phase difference between the sum and difference signals is $\Delta\varphi = 90^\circ$, the output is 0; if $\Delta\varphi \neq 0^\circ$, the output is not equal to 0. For SCM, the error signal detector is the envelope detector for amplitude

modulation. If the phase difference between the sum and difference signals is $\Delta\varphi = 90^\circ$, through $0/\pi$ modulation and addition, the difference signal will not form amplitude modulation to the sum signal, so the output of the envelope detector is 0; if $\Delta\varphi \neq 0$, through $0/\pi$ modulation and addition, the difference signal will form amplitude modulation to the sum signal, so the output of the envelope detector is not equal to 0. Since SCM and DCM have the same principles of cross coupling and static tracking error caused by phase inconsistency of the sum channel and difference channel (refer to Sect. 2.5.5.1 for principle of static tracking error, and refer to Sect. 2.5.1.2 for principle of cross coupling), so the calculation expressions of them are the same (see Expression (2.251)).

Due to the SCM aforesaid is produced from the channel combination for DCM, so the expression of tracking error caused by thermal noise in SCM may be derived from the DCM Expression (2.252), then from Expression (2.252), we obtained that:

$$\delta_1 = \frac{\theta_{0.5}}{K_m \sqrt{\frac{S}{N} \frac{B_{IF}}{B_n}}} = \frac{\sqrt{G_\Sigma}}{K \frac{U_I}{U_n} \sqrt{\frac{B_{IF}}{B_n}}} \quad (2.260)$$

After combined into SCM, the expression above then has following effects:

1) The absolute slope of difference lobe gain is K : it affects the output voltage of the difference channel $U_\Delta = K\theta$ (θ is the angle deviated from the electronic boresight). If other conditions remain the same, the change of U_Δ can be equivalent to that caused by K changed to K' :

Because: $U_\Delta' = K'\theta$, $U_\Delta = K\theta$

then, $K' = K \frac{U_\Delta'}{U_\Delta}$

Factors affecting the signal amplitude of the difference channel are the power coupling coefficient of summit or M ($M < 1$) and the phase inconsistency between sum and difference channels $\Delta\varphi$, that is,

$$U_\Delta' = U_\Delta \sqrt{M} \cos \Delta\varphi$$

Then,

$$K' = K \sqrt{M} \cos \Delta\varphi \quad (2.261)$$

2) Noise after channel combination

The RMS value of the noise voltage of sum and difference channels are both set to U_n . Whereas the noise voltage of difference channel is coupled $\sqrt{M}$ times and then adds the noise of sum channel, so the noise power after addition increases up to $U_n^2 + MU_n^2 = (1 + M)U_n^2$, correspondingly, the noise voltage increases to $\sqrt{1 + M}U_n$.

3) Aliasing noise

The thermal noise of the difference channel will cause the random error of angle tracking. Refer to Fig. 2.65 for noise spectrum of the difference channel in SCM and DCM.

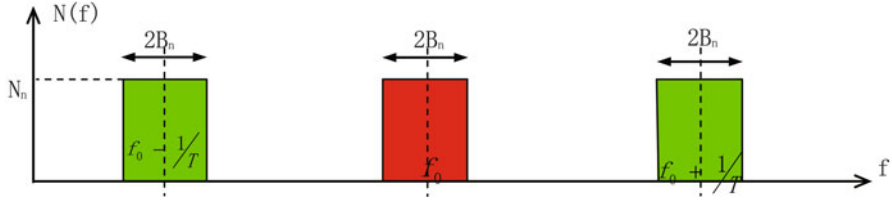


Fig. 2.65 Noise spectrum of difference channel

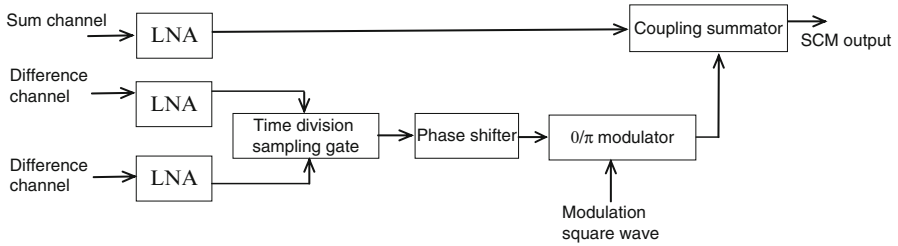


Fig. 2.66 Principle block diagram of combining TCM into SCM

As shown in Fig. 2.65, the spectrum of angle tracking random error in DCM is within $f_0 \pm B_n$ band, and its noise power is $2B_n N_0$, where N_0 is the noise power spectral density. The spectrum of angle tracking random error in SCM is within $2B_n$ band at two symmetrical frequency points $f_0 \pm 1/T$ ($1/T$ is the modulation frequency for forming SCM amplitude modulation), then the two symmetrical noise spectrums are overlapped through envelop detection, and its noise power is $4B_n f_0$ as twice as that of DCM, correspondingly, the random error of angle tracking increases $\sqrt{2}$ times.

Substitute the effects of those three factors into Expression (2.260), then the random error σ_n caused by the thermal noise in SCM is obtained:

$$\begin{aligned}
 \sigma_n &= \frac{\sqrt{G_\Sigma}}{K\sqrt{M}(\cos \Delta\varphi) \frac{U_\Sigma}{\sqrt{1+MU_n}} \sqrt{\frac{B_{IF}}{B_n}}} \\
 &= \frac{\theta_{0.5}}{K_m \sqrt{M}(\cos \Delta\varphi) \sqrt{\frac{1}{(1+M)} \frac{S}{N} \frac{B_{IF}}{B_n}}} \quad (2.262)
 \end{aligned}$$

(2) Three-channel monopulse combines into single-channel monopulse:

Then the antenna outputs three signals of the sum channel, azimuth difference channel, and elevation difference channel. Refer to Fig. 2.66 for one combination method in common use.

As shown in Fig. 2.66, the azimuth difference channel and the elevation difference channel are combined into one channel using TDM method, meanwhile, part of the signals from the two channels are coupled via the same one $0/\pi$ modulator and then adds the signal from the sum channel to become an amplitude-modulated signal, after that, its working principle and the random error analysis are the same as that of combining DCM into SCM. It is easier since that no cross coupling exists, and the phase modulator changes from the four-phase modulation to $0/\pi$ two-phase modulation with period of $T/2$. Its analysis is the same as that of combination of DCM into SCM.

2.5.2 Angle Measurement with Interferometer

The interferometer angle measurement is a method that uses the phase difference or time-delay difference of the signal transmitted from the target received by the two receiving antennas to measure the angle with coherent principles. When it receives single-frequency signal, the interferometer is used to measure the phase, so we called it the phase difference interferometer; when it receives broadband signal (e.g., broadband noise of radio star), the time-delay difference is usually used to measure since the phase measurement is more complicated, and we called it the time-delay-difference interferometer. The target signal can be response signal, beacon signal, or others.

2.5.2.1 Phase Difference Interferometer

Its principle is as shown in Fig. 2.67.

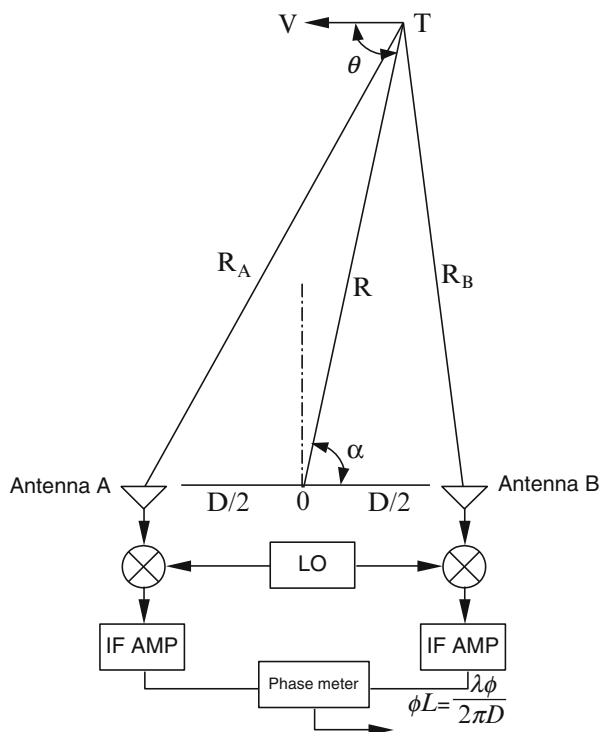
The origin of coordinate is the midpoint of the line between the two antennas (referred to as baseline) O , and α is the angle of target which defines the included angle of the baseline and the target direction. The phase difference of the target signals received from the two antennas is $\Delta\phi$. When the target is far away from the two antennas, the intensities of the signals received from the two antennas are approximately equal and can be considered as a plane wave (i.e., parallel ray); therefore, the phase difference caused by the distance difference between the two antennas is expressed as:

$$\Delta\phi = \left(\frac{2\pi}{\lambda}\right)(R_A - R_B) = \left(\frac{2\pi}{\lambda}\right)D \cos \alpha$$

where D is the distance between the two antennas of receivers, λ is the wavelength of the target signal, and $\cos \alpha$ is the direction cosine.

As long as the phase difference $\Delta\phi$ is measured, then the direction cosine $L(L = \cos \alpha)$ can be solved, and then the angle coordinate of the target α can be

Fig. 2.67 Angle measurement principle of phase difference interferometer



solved. Since increasing the baseline length D is helpful to improve the measurement accuracy of α , it is obvious that the improvement of $\Delta\phi$ measurement accuracy is also useful to improve the angle measurement accuracy.

$\Delta\phi$ measurement method: mix the RF signals received from the two antennas to IF signals, and then, transmit them after amplification to the phase meter to measure $\Delta\phi$. Since the period of phase is 2π , and $\Delta\phi$ has multiple measured values, leading to angle ambiguity (similar to the tone ranging), for which we should take appropriate measures. In addition, in the course of frequency conversion and amplification of the two signals aforesaid, the phase inconsistency of the two channels will introduce the phase error, which also requires appropriate technical means, thus leading to complicated implementation, higher technical specification, and the necessity of requiring the equipment to be stable and fixed, which is inconvenient for mobility. However, the angle measurement accuracy of it is the highest in radio TT&C system.

In the event that the baseline of the interferometer is extreme short, the interferometer becomes phase-comparison monopulse radar. Under this situation, the antenna is fixed on an antenna pedestal and uses the direction cosine as the error signal to track the angle. Four antennas are required for 2D tracking in azimuth and elevation.

Key technologies for interferometer are as follows.

(1) Phase balance of the interferometer phase measurement system. The error of phase measurement shall be minimized in order to provide higher angle measurement accuracy for the interferometer under the entire operation environment. The phase error normally is caused by the phase inconsistency of the interferometer receiving system and system SNR. Following error sources may cause inconsistency of the phases between the channels:

- Polarization error and antenna pattern error
- Mechanical error
- Change of phase center of the antenna
- Phase inconsistency of antenna element and feed waveguide
- Phase inconsistency of each local oscillator and receiving channel
- Phase inconsistency of each IF amplifier and baseband

In order to minimize these errors, the antenna, reference signal, front end, and the local oscillator of balanced mixer shall be consistent as far as possible so as to minimize the phase inconsistency in case of change of environment. The waveguide transmission line shall be as short as possible. The phase matching antenna and balanced mixed shall be used. The IF amplifier with consistent phase shift and the transmission lines are used to minimize the inconsistency of the phases between IF channels. The automatic phase calibration unit is adopted in the system to solve the phase inconsistency caused by change of environment, using frequency division, time division, or code division method.

(2) The length error of the baseline between the end stations shall be measured accurately.

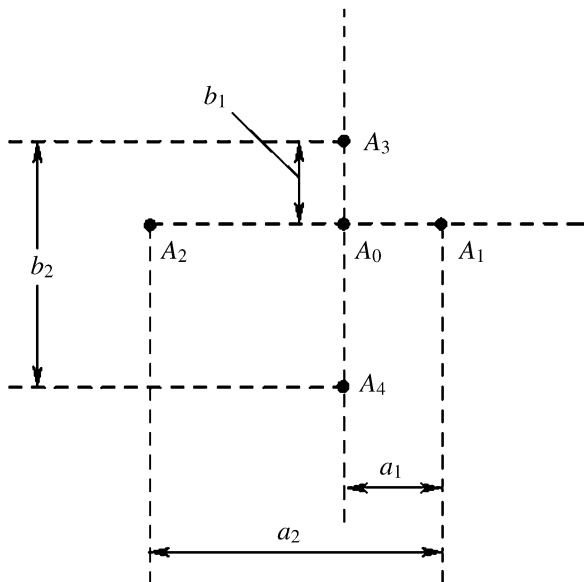
(3) Digitalization: to minimize the drift error caused by phase drift of analog circuit on the phase interferometer.

Refer to Fig. 2.68 for typical 2D deployment of the interferometer antenna [19].

As shown in Fig. 2.68, the design of antenna deployment shows: on elevation plane: $b_1 = 0.577\lambda$, $b_2 = 15b_1$, and antenna A_1, A_2 and A_0 are used for azimuth angle measurement, where antenna A_3, A_4 and A_0 are perpendicular to antenna A_1, A_2 and A_0 . Since there are five antennas in total, with which five channels shall be connected correspondingly. A_1-A_2 is used to accurately measure the azimuth angle, A_1-A_0 is used to solve the azimuth ambiguity, A_3-A_4 is used to accurately measure the elevation angle, and A_3-A_0 is used to solve the elevation ambiguity. In each DF channel as shown in Fig. 2.68, the input signal is converted into a 70 MHz IF signal through LNR and one down-conversion, and then transforms by A/D conversion at 70 MHz to form a digital signal. The digitalized phase measurement is used to improve the measurement accuracy.

Phase calibration technology is adopted to calibrate the long-term drift error so as to improve DF accuracy. The frequency of calibration signal source that is close to the received frequency (frequency division) is generated by the frequency synthesizer which uses high-stability crystal oscillator as frequency reference. The frequency synthesizer also generates Tx/Rx frequency of local oscillator, Tx signal carrier frequency, and sampling frequency of A/D.

Fig. 2.68 Schematic diagram of antenna 2D deployment



The phase error required for angle measurement accuracy may be calculated by the expression below, that is,

$$\begin{aligned}\sigma_\theta &= \frac{1}{2\pi d} \frac{1}{\cos \theta} \sigma_\varphi \\ \sigma_\varphi &= \sigma_\theta \frac{2\pi d \cos \theta}{\lambda}\end{aligned}\tag{2.263}$$

where σ_θ is the RMS index of angular accuracy, σ_φ is the RMS value of inconsistency of the phases between the channels, d is the longest distance of one baseline, that is, $d = 15b_1 = 8.355\lambda$, λ is the working wave length, and θ is the angle in the greatest coverage direction.

1) The measurement error of phase difference required for ambiguity solution is expressed as:

$$\varphi_d < \frac{\pi}{(n_i/n_{i-1}) + 1}\tag{2.264}$$

where n_i/n_{i-1} is the ratio of baselines.

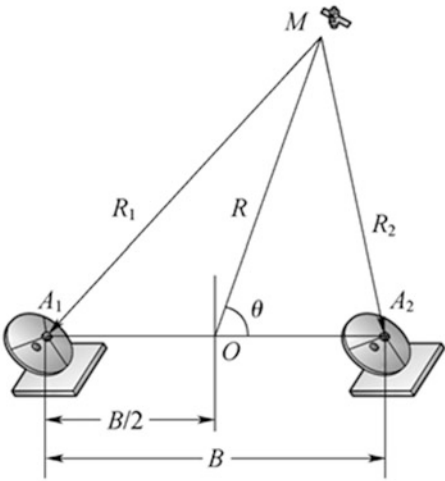
$\varphi_d > \sigma_\varphi$ is required for solving ambiguity.

2) Allocation of system DF errors: Table 2.12 lists all the factors causing DF error and its allocated values.

Table 2.12 System phase errors

Error source	Phase inconsistency error (1σ)/(°)	Error source	Phase inconsistency error (1σ)/(°)
Polarization mode	1	Mixer	2
Center of phase	0.5	Calibration error	1
Mechanical calibration	2	Phase error caused by integrated baseband	3
Waveguide circuit	3	Phase error caused by thermal noise	2
IF amplifier	3	Total error	6.5

Fig. 2.69 Angle measurement principle of time-delay difference interferometer



2.5.2.2 Time-Delay Difference Interferometer

The principle of the time-delay difference interferometer is as shown in Fig. 2.69, in which *M* is the instantaneous position of the target, *A*₁ and *A*₂ are the positions where the two antennas locate in the same coordinate axis, *B* is the distance between the two antennas, the length of baseline, *O* is the center of baseline, *θ* is the included angle from the target direction *OM* and the baseline, *R*₁ and *R*₂ respectively are the distance from the target to each of two antennas, and *R* is the distance from the target to the center of baseline.

According to cosine theorem, there is:

$$R_1^2 = R^2 + \left(\frac{B}{2}\right)^2 + 2R\left(\frac{B}{2}\right)\cos\theta$$
$$R_2^2 = R^2 + \left(\frac{B}{2}\right)^2 - 2R\left(\frac{B}{2}\right)\cos\theta$$

then,

$$R_1^2 - R_2^2 = 4R \left(\frac{B}{2} \right) \cos \theta$$

The direction cosine is:

$$l = (R_1 + R_2)(R_1 - R_2)/(2RB)$$

If $R \gg B$, then $2R = R_1 + R_2$, and

$$l = \cos \theta \approx \frac{(R_1 - R_2)}{R} = \frac{r}{B} \quad (2.265)$$

where r is the range difference.

According to the time-delay difference τ of the same signal received from the two stations (measured with correlation method), we obtain:

$$r = \tau c$$

where c is the light velocity.

$$l = \cos \theta \approx \frac{\tau c}{B} \quad (2.266)$$

By differentiating Expression (2.244), we can obtain that

$$\begin{aligned} \delta l &= (\delta \theta) \sin \theta = \frac{(\delta \tau) c}{B} \\ \delta \theta &= \frac{(\delta \tau) c}{(B \sin \theta)} \end{aligned} \quad (2.267)$$

where δ represents error increment. This expression shows the measurement error caused by time-delay difference interferometer.

2.5.2.3 Deep-Space TT&C Application with Interferometer

Actually, the angle measurement with interferometer is a very old concept. In application to deep-space TT&C, the baseline is required to be extended to improve the measurement accuracy, which forms the Very Long Baseline Interferometry (VLBI). Currently, the baseline has extended to the edges of continents and reached to the ends of the Earth, which is tending to the limits. Although there are points of

view that the baseline shall be extended to the outer space, it hasn't performed yet. Lengthening baseline brings some disadvantages as follows:

- Measurement error of "time difference" caused by different sources between the two terminal stations
- Error caused by different radio wave propagation paths during receiving the signals of the two terminal stations
- Error caused by different equipment used in the two terminal stations
- Errors caused by incorrect baseline measurement
- Loss of some real-time capability due to the post processing of range difference data between the two distanced terminal stations
- Smaller overlapped area covered with one spacecraft by the two terminal stations because of long baseline

Therefore, some technologies like Difference Very Long Baseline Interferometry (Δ VLBI), Same-Beam Interferometry (SBI), and Connected Element Interferometry (CEI) are presented to minimize these errors. Δ VLBI and SBI are designed to put the reference point for measurement into the outer space, and change the measurement of long range to that of relatively short range. If the location of the reference point is known, the position of deep-space probe shall be determined by measuring the relative positions of the deep-space probe and the reference point, with an improved measurement accuracy by mitigating relevant system errors with differential technology. The principle of Δ VLBI is that: Firstly, the two long-distanced DSNs simultaneously observe the deep spacecraft and then generate a time difference by processing of measured data with accuracy of 20 ps, and the time-delay difference at that time is converted to an angle. Then, these two deep-space stations simultaneously take measurements of the same radio star in a known location near the spacecraft, which yields two values, the time difference and angle, from which the equipment errors can be calculated by subtracting the known radio star position, thus improving the accuracy by differentiating this value with the measured data of spacecraft. Spacecraft CEI technology uses the optical fiber for baseline transmission so as to improve measurement accuracy of the time difference between the two end stations and acquire real-time ability. For SBI technology, two antennas in the deep-space station are used to simultaneously measure the two targets within the same beam, thus to obtain the measured data of difference interferometry. The accurate relative positions of the two targets can be obtained with SBI. Moreover, the deep-space interferometer also adopts following methods:

(1) Angle measurement system with Delta differential one-way ranging (Δ DOR)

The time-delay difference measurement may be implemented with one-way range difference; therefore, VLBI will evolve to differential one-way ranging (DOR), and further evolves to Δ DOR plus differential technology.

Δ DOR is used to differentiate the time-delay difference of the target signals received from the two TT&C stations at the ends of baseline with that of the radio star signals received from the calibration receiver, where the position of radio star is as a known reference point (obtained from astronomy). Because all ranging signals are the CW signals, the range-ambiguity-solving method is required (see Sect. 2.4.1

for its basic principle). Thus, the range-ambiguity-solving method and its “big hop counts” algorithm are still the key technologies. For Δ DOR ambiguity-solving method, except for the residual carrier after modulation, the transmit signal of the target shall carry the strong subcarrier (or side tone) and harmonic components of each sub-harmonic, that is, the transmitting spectrum of the target receiver is wide enough, and the ground receiver is able to receive the residual carrier of the transmitting signal and the even harmonic component of the modulated subcarrier of the forwarding signal, thus to solve the range difference with higher harmonic as well as the ambiguity with other harmonics.

(2) CEI angle measurement

The Connected Element Interferometry (CEI) based on the same reference frequency, can implement accurate measurement to the received time-delay signals from two end stations through the fiber communication link between the two tracking-measurement stations distanced from tens to 100 km. With this high precision data the included angle of the baselines between the two stations relative to the radio emission source can be determined. The equivalent measurement accuracy may be implemented on the relatively short baseline without any long-term and continuous arc segment data. The fiber communication link is used to transmit the acquired data to the same correlator for processing in real time.

Its angular accuracy $\delta\theta$ is inversely proportional to the length of baseline B and is proportional to the measurement accuracy $\delta\tau$ of time-delay difference τ . In the event of achieving high accuracy of angle measurement, the longer baseline shall be used (by lengthening baseline B) or the measurement accuracy of time delay for interferometer shall be improved (by reducing $\delta\tau$). In addition, the factor causing a lot of errors is a relevant factor on the short baseline and reduced by cancellation. Since the short baseline prior model has a small error, the high accurate phase delay data can be obtained.

CEI measurement has such advantages as follows:

- 1) Real-time processing
- 2) Coherent processing between the two stations owing to the use of the same clock
- 3) Obviously reduced errors of transmission medium by cancellation of short baseline
- 4) A higher elevation angle and a longer observation time achieved with a shorter baseline

This method has more advantages in deep-space orbit determination, especially for orbit measurement when approaching/falling of the Moon and the planets. Also, this technology has a special advantage in real-time high-accuracy orbit determination of LEO and MEO, implementing measurement with multiple stations jointly without transmission of uplink signals.

The error sources affecting the CEI phase delay measurement mainly include the length of baseline, position of radio source, troposphere, ionosphere, stability of clock, and measurement accuracy. Many error sources may be corrected by simultaneously observing the radio sources (spaceship and radio star) adjunct to the two angle positions and then conducting the differential calculation, for example, the

unknown delay and phase jitter of the clock and local oscillator between the two stations, unknown delay of measuring equipment, equipment phase jitter in the course of measurement, delay caused by antenna deformation, transmission medium delay, and baseline error.

The prior model is used to solve the phase ambiguity in measurement of time delay. For dual mode of S-band and X-band, the prior model is used to solve the integer ambiguity of phase measurements of S-band and then X-band signals. The prior model is developed on the basis of available information, such as station address, source position and transmission medium delay, etc. This model shall be a feature of accuracy for solving phase ambiguity.

The random jitter of troposphere is the main error source affecting the measurement accuracy of CEI phase relay, where the statics error is inversely proportional to the elevation angle while observing the spaceship.

(3) Angle measurement with same beam interferometer

In case two spaceships' angle location, are very approximate they can be in the same beam of the ground antenna, and two antennas in the two deep-space stations are used to observe these spaceships simultaneously and then generate differential interferometry data, which is called as Same Beam Interferometry (SBI) technology that provides the high accurate measurement for relative angle positions. SBI data supplements the radial direction information from the traditional Doppler data and further improves the orbit determination accuracy. Observing from the Earth, the angle between the two spacecraft is far smaller than that between the spacecraft and the radio star. Since measurement errors are proportional to the angle, hence the interferometry of spacecraft vehicle may be more accurate than the traditional way. When the two spacecraft locate in the same antenna beam, the carrier phase of them can be tracked at the same time, so that the carrier phase is used for further improving the measurement accuracy instead of group time delay. SBI technology is better than the traditional interferometry of spacecraft-radio star on operation. It's not required to compare with the radio star, and the antenna is able to point at the spacecraft without shift, so that real-time phase measurements may be implemented at the same time. The determination of orbit plan combining the same beam interferometry data with dual-way Doppler data or one-way/dual-way Doppler data has many advantages and multiple spacecraft may be tracked with this method.

2.5.3 Theoretical Calculation of Angle Tracking Accuracy [20]

(1) Random error

1) Error caused by thermal noise of the receiver is:

$$\sigma_1 = \frac{\theta_{0.5}}{K_m \sqrt{\frac{S}{N} \frac{B_{IF}}{B_n}}} = \frac{1}{\mu \sqrt{\frac{S}{N} \frac{B_{IF}}{B_n}}} \quad (2.268)$$

where $\theta_{0.5}$ is the beam width of half power point, S/N is signal-to-noise ratio of receiver (bandwidth is B_{IF}), B_{IF} represents IF amplification equivalent noise bandwidth of the receiver, B_n is servo equivalent noise bandwidth, K_m represents normalized difference slope (generally ranging between 1.2 and 1.7), and μ is relative difference slope $\mu = \left[\frac{\partial F(\theta)/\partial \theta}{\sqrt{G_s}} \right] = \frac{K_m}{\theta_{0.5}}$.

2) Multipath reflection. When the elevation is relative low, the target signal passes through the antenna side lobe and enters the receiver after being reflected by ground. The elevation measurement error generated by the jamming servo system is:

$$\sigma_2 = \frac{\rho \theta_{0.5}}{\sqrt{8A_3}} \quad (2.269)$$

where ρ represents reflection coefficient; A_3 is gain ratio of main lobe to side lobe.

3) Variation error caused by dynamic delay of the target. The variation error of angle delay caused by the target angular acceleration and angular jerk is the largest when dynamic variation occurs at low orbit and will reduce along with increase in distance. Variation error of the target is:

$$\sigma_3 = \left(\frac{\dot{\omega}}{K_v} + \frac{\ddot{\omega}}{K_d} \right) \frac{1}{B_n} \quad (2.270)$$

where $\dot{\omega}$ represents angular acceleration of the target, $\ddot{\omega}$ is angular jerk of the target, K_v represents speed error constant of the servo system, and K_d is acceleration error constant of the servo system.

4) When azimuth axis is not perpendicular to horizontal plane, the introduced error is:

$$\begin{aligned} \sigma_4 &= \gamma \tan E \sin A \text{ (azimuth angle)} \\ \sigma_4 &= \gamma \cos E \text{ (elevation angle)} \end{aligned} \quad (2.271)$$

where r represents random tilt of the azimuth axis to horizontal plane.

5) When azimuth axis is not perpendicular to elevation, the introduced error is:

$$\sigma_5 = \delta \tan E \text{ (azimuth angle)} \quad (2.272)$$

where δ is random quantity when two axes are not perpendicular.

6) When optical axis is not perpendicular to elevation axis, the introduced error is:

$$\sigma_6 = \Delta \sec E \text{ (azimuth angle)} \quad (2.273)$$

where Δ is random quantity of optical axis deviating from elevation axis.

7) Torque in-equilibrium error generated by gusty wind is:

$$\sigma_7 = 2K_w \bar{v} \left[\int_0^{f_{\max}} \frac{\varphi_w(f)}{K_T^2(f)} df \right]^{1/2} \quad (2.274)$$

where K_w represents wind torque constant, $\bar{v}$ is mean wind speed (m/s), $f_{\max}$ represents the highest frequency of gusty wind, $K_T^2(f)$ is system torque transmission function, and $\varphi_w(f)$ represents gusty wind spectrum.

8) Quantization error of data sensor is

$$\sigma_8 = \frac{g}{\sqrt{12}} \quad (2.275)$$

where σ is the minimum-digit quantization unit.

Additionally, there are angle fluctuation error σ_9 caused by irregular troposphere, angle measurement error σ_{11} introduced by zero drift of phase discriminator of receiver, error σ_{12} introduced by zero drift of servo amplifier, random error σ_{13} introduced by servo noise and blind zone, random error σ_{14} caused by pedestal unlevelness, and angular conversion error σ_{15} .

9) Total random error is

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_{15}^2} \quad (2.276)$$

(2) System error

1) Dynamic delay error is

$$\Delta_1 = \frac{\omega}{K_v} + \frac{\dot{\omega}}{K_d} \quad (2.277)$$

where ω is angular velocity of the target and $\dot{\omega}$ is angular acceleration of the target.

2) Polarization error is

$$\Delta_2 = \frac{\theta_{0.5} \cdot \delta}{2K_m} \sqrt{\frac{3}{8} \left[(1-K)^2 + \frac{4}{3} \cdot K \sin^2 \Delta\varphi \right]} \quad (2.278)$$

where δ is the length difference between major axis and minor axis, which is normalized with respect to the minor axis, K_m is normalized slope of difference lobe, K is degree of inconsistency of axial ratio of two polarizers, and is difference between polar tilt angles of two polarizers, and $\Delta\varphi$ is difference between polar tilt angles of two polarizers.

3) Amplitude unbalance before the comparator is

$$\Delta_3 = \frac{\Delta G/G}{2\mu} \quad (2.279)$$

where $\Delta G/G$ is relative value of antenna gain unbalance.

Inconsistency between the sum and difference gain of tracking receiver will cause open-loop gain change of the servo system, thus causing dynamic error variation (see Expression (2.267)).

4) Phase unbalance before and after the comparator is

$$\Delta_4 = \frac{\tau \cdot \operatorname{tg} \varphi}{2K_m} \theta_{0.5} = \frac{\theta_{0.5} \operatorname{tg} \varphi}{K_m \sqrt{G_n}} \quad (2.280)$$

where τ is phase unbalance before the comparator; φ is phase unbalance after the comparator (inconsistency between sum and difference phases of the tracking receiver); $\sqrt{G_n} = 2/\tau$ is null depth of differential beam, representing the ratio of G_Σ to the minimum gain $G_{\min}$ at differential beam zero, namely, $G_\Sigma/G_{\min}$.

5) Coupling error of sum-difference channel is

$$\Delta_5 = \frac{\theta_{0.5} \cdot \cos \varphi}{K_m \cdot \sqrt{F_1}} \quad (2.281)$$

where φ is signal coupling phase shift angle of sum-difference channel; F_1 is isolation degree of sum-difference channel.

6) Unbalance and zero drift of servo amplifier, namely,

$$\Delta_6 = \frac{\Delta V}{K_\varphi} \quad (2.282)$$

where ΔV is unbalance voltage value of all servo systems converted to the output terminal of the detector; K_φ is directional sensitivity of the antenna to output terminal channel of error discriminator.

7) Due to unbalance in antenna structure, the error between antenna deformation resulting from the torque generated by stable wind and the antenna pointing is

$$\Delta_7 = \frac{\bar{T}_W}{K_T} \quad (2.283)$$

where $\bar{T}_W$ is mean wind torque:

$$\bar{T}_w = K_w \times V^2$$

8) When azimuth axis is not perpendicular to horizontal plane, the error is:

$$\Delta_8 = \gamma' \tan E \sin A (\text{azimuth angle}) \quad (2.284)$$

$$\Delta_8 = \gamma' \cos A (\text{elevation angle}) \quad (2.285)$$

where γ' represents inclination of the azimuth axis to horizontal plane.

9) When azimuth axis is not perpendicular to elevation axis, the error is:

$$\Delta_9 = \delta' \tan E(\text{azimuth angle}) \quad (2.286)$$

where δ' refers to non-perpendicularity between the two axes.

10) When optical axis is not perpendicular to elevation axis, the error is:

$$\Delta_9 = \Delta' \sec E(\text{azimuth angle}) \quad (2.287)$$

where Δ' represents non-perpendicularity between the two axes.

11) Optical-electrical axis non-parallelism error is

$$\Delta_{11} = \psi \sec E(\text{azimuth angle}) \quad (2.288)$$

$$\Delta_{11} = \psi(\text{elevation angle}) \quad (2.289)$$

where ψ is adjustment surplus of optical-electrical axis parallelism.

Total system error is

$$\Delta = \sqrt{\Delta_1^2 + \Delta_2^2 + \cdots + \Delta_{11}^2} \quad (2.290)$$

2.5.4 Theoretical Analysis of Angle Tracking of Wideband Signal – Cross-Correlation Function Method

With exponential increase in high-speed data transfer rate, more and more attention arises to angle tracking of wideband signals. In the past, during angle tracking of FM wideband signal, a small part of signal bandwidth is segmented to perform angle tracking. Namely, the wideband signal becomes narrowband signal. Such method simplifies the proposal but causes loss of signal energy, thereby enlarging angle error and decreasing angle tracking sensitivity. Especially, there are obvious defects if partial bandwidth exists in C/N wideband angle tracking. The better method is to solve cross-correlation function of monopulse sum and difference signal, and demodulate wideband difference signal from low C/N in the case of equal time delay and large integral constant. Additionally, it is required to normalize angle error signal to enable angle servo loop to operate normally. Angle acquisition shall be made before the servo loop is closed, which is usually implemented through detection of sum channel signals (e.g., sum AGC). However, in case of low C/N , the common detection method fails to detect sum channel signal, which involving key technology of cross-correlation method. The updated angle tracking technology is given in the section below.

(1) Amplitude comparison pulse tracking principle of wideband signal

Common angle tracking system is a narrowband system and generally analyzed by single-frequency signal, with a phase discriminator used as the angle error detector. The impact of sum and difference channels being discussed is the

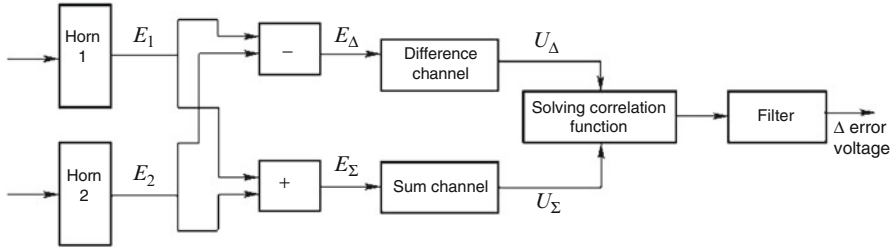


Fig. 2.70 Amplitude comparison monopulse receiver for receiving wideband signal

consistency of phase and gain; for quadrature dual-channel combination only those with 90° phase difference are considered. However, angle tracking system of wideband signal involves tracking to the signal full spectrum, so there are many special problems.

Take typical four-horn amplitude comparison monopulse for instance; refer to Fig. 2.70.

Figure 2.70 shows a pair of horns at elevation direction (another pair of horns are the same in terms of operating principle). Signal E_1 of horn 1 is taken as reference to discuss correlation of all signals in the figure. Since horn 2 and horn 1 receive the same wideband signal, E_1 is correlated with E_2 , and the sum signal E_Σ and difference signal $E_\Delta = E_1 - E_2$ are both cross-correlated with E_1 . When $E_1 > E_2$ (upward deviate corresponding to the target), its cross-correlation value is positive, when $E_1 < E_2$, its cross-correlation value is negative (downward deviate corresponding to the target). Positiveness and negativeness of cross-correlation value can be used to determine direction to which of the target is deviated. The cross-correlation value is determined by the following cross-correlation function expression [21], namely,

$$C_{\Sigma\Delta}(\tau) = U_\Sigma U_\Delta K \cos [(2\pi f_0 + \pi B)\tau] \frac{\sin \pi B \tau}{\pi B \tau} \quad (2.291)$$

Expression (2.291) is corresponding to wideband signals of flat spectrum with reference significance. τ is delay difference between sum channel and difference channel, B is signal bandwidth, U_Σ and U_Δ are amplitude of sum and difference signal, and K is transmission coefficient of correlator. To achieve peak value of $C_{\Sigma\Delta}(\tau)$, an additional time delay $(-\tau)$ shall be added between sum channel and difference channel to achieve $\tau + (-\tau) = 0$, at this time, its peak value is:

$$C_{\Sigma\Delta}(O_g) = U_\Sigma U_\Delta K \quad (2.292)$$

$KE^2(U_\Delta/U_\Sigma)$ (E is sum channel AGC comparison level) can be obtained after the correlation value is normalized. It is angle error signal, which is in direction proportion to the angle deviation but irrelevant to distance. Namely, the angle error slope is normalized, and normalization of its sum and difference signals can

be implemented by conventional AGC. Also, normalization can be achieved through sampling U_{Σ} and U_{Δ} simultaneously, and calculating U_{Δ}/U_{Σ} .

As a result, differences between wideband signal tracking and single-carrier signal (or narrowband signal) tracking are as follows:

- 1) Since there is not residual carrier signal in wideband signal, the conventional carrier phase-locked loop method cannot be used. The angle error extraction method, in which the sum signal and difference signal are in direct cross-correlation, is employed.
- 2) Time delay between sum channel and difference channel is required to be consistent to make difference between time be, thus obtaining relevant peak value.
- 3) Amplitude-/phase-frequency features of sum channel and difference channel are required to be consistent, in order to make sum and difference signals not in distortion or maintain consistent distortion, thereby obtaining large relevant value.

Therefore, time delay of the channel shall be controlled to be in consistence during angle tracking of wideband signals. Additionally, phase shift of each sum and difference LO signal may be inconsistent, and that required to be adjusted. As LO signal is single frequency signal, a phase shifter can be used for adjustment.

According to above, high accuracy of angle measurement may be achieved for wideband signals as long as design method is appropriate. When power spectrum density of signals maintains unchanged, the wider the bandwidth is in the signal spectrum range, the stronger the signal power becomes, and the stronger its cross-correlation value is.

See Fig. 2.71 for block diagram of wideband angle tracking receiver.

In Fig. 2.71, delay adjustment is used to calibrate inconsistency between sum channel and difference channel, phase shifter used to calibrate inconsistency between phase shifts of LOs. Wideband filtering requires consistency between amplitude-/phase-frequency features. Figure 2.71 shows “3-channel” monopulse

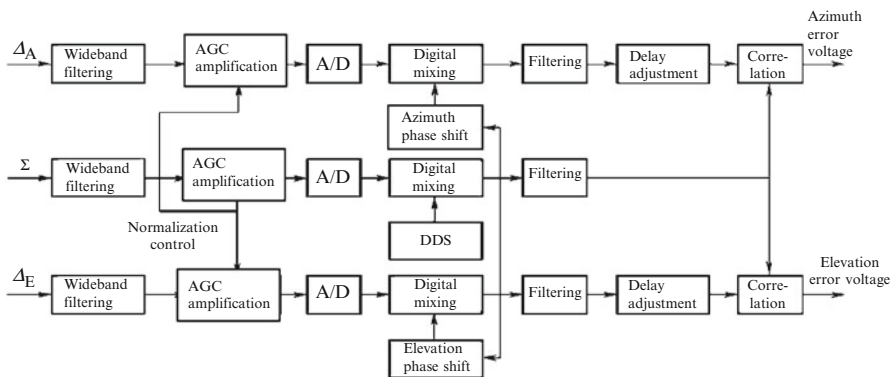


Fig. 2.71 Block diagram of a 3-channel monopulse wideband receiver

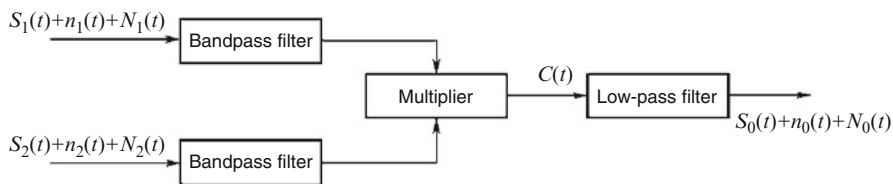


Fig. 2.72 Block diagram of cross-correlation detection

proposal. It is revealed from Expression (2.291) that an additional time delay t_0 must be added between elevation difference signal and azimuth difference signal to implement “dual-channel monopulse” making that the correlation value of error signal from dual-channel to the elevation difference is maximum while that to the azimuth difference is relatively small (vice versa). Application of “single-channel monopulse” proposal will result in more than 6 dB signal-noise ratio loss, which conflicts with low C/N receiving.

(2) Output signal-noise ratio of angle error cross-correlation detection

Block diagram of cross-correlation detection is as shown in Fig. 2.72.

In Fig. 2.72, $S_1(t)$ and $S_2(t)$ are the two-channel signals being correlated (in angle detection, respectively being sum channel and difference channel). $n_1(t)$ and $n_2(t)$ are uncorrelated part of 2-channel noise, and $N_1(t)$ and $N_2(t)$ are correlated part of 2-channel noise. Band-pass filter allows signal spectrum to pass but limit white noise, with bandwidth being B . Low-pass filter filters the relevant output signal, with bandwidth being B_L .

Cross correlator can solve cross-correlation function $R_{xy}(0)$ when $t = 0$, multiplier can multiply 2-channel signal, its output is:

$$\begin{aligned}
 C(t) &= [S_1(t) + n_1(t) + N_1(t)] \cdot [S_2(t) + n_2(t) + N_2(t)] \\
 &= S_1(t)S_2(t) + S_1(t)n_2(t) + S_1(t)N_2(t) + n_1(t)S_2(t) \\
 &\quad + n_1(t)n_2(t) + n_1(t)N_2(t) + N_1(t)S_2(t) + N_1(t)n_2(t) + N_1(t)N_2(t)
 \end{aligned} \tag{2.293}$$

1) In case of low carrier-noise ratio

In case of low carrier-noise ratio, as $N_1(t) \ll n_1(t)$, $S_2(t) \ll n_2(t)$, $N_1(t) \ll n_1(t)$, $N_2(t) \ll n_2(t)$ and the amplitude of $n_1(t)$ and $n_2(t)$ is equal approximately; therefore,

$$C(t) \approx S_1(t)S_2(t) + n_1(t)n_2(t) + N_1(t)N_2(t) \tag{2.294}$$

where $S_1(t)S_2(t)$ is signal part; $n_1(t)n_2(t)$ is non-correlation noise part; $N_1(t)N_2(t)$ is correlated noise part.

① Signal output. Output signal is the cross-correlation function of $S_1(t)$ and $S_2(t)$ in bandwidth B when $= 0$:

$$R_{S_1 S_2}(0) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T S_1(t)S_2(t)dt \tag{2.295}$$

$S_1(t)$ is correlated with $S_2(t)$ but they have different amplitude. Let $\frac{S_1(t)}{S_2(t)} = m$, when $E(n) = 0$, then

$$R_{S_1 S_2}(o) = m \cdot D[S_1^2] = m \cdot \sigma_{S_1}^2 \quad (2.296)$$

where $D[S_1^2] = \sigma_{S_1}^2$ is variance of $S_1(t)$.

② Cross-correlation output of correlated noise $R_{N_1 N_2}(o)$. Since $N_1(t)$ is correlated with $N_2(t)$, according to above, the following expression can be obtained:

$$R_{N_1 N_2}(o) = M \cdot D[N_1^2] = M \cdot \sigma_{N_1}^2 \quad (2.297)$$

where,

$$M = \frac{N_1(t)}{N_2(t)}$$

where $\sigma_{N_1}^2$ is variance of correlated noise $N_1(t)$.

③ Output of non-correlated noise. Typical analytic method of its cross-correlation function $R_{n_1 n_2}(o)$ is that the cross-correlation function, $R_{xy}(t)$ is derived from inverse Fourier transformation of cross-spectral density $R_{xy}(\omega)$ of random function $x(t)$ and $y(t)$. Here, we employ frequency domain method to make analysis, in order to obtain clear understanding on the project development.

Strictly speaking, random signals cannot be decomposed in Fourier method, because random signals are usually not periodic or square integrable. However, Wiener put forward “broad sense harmonic” concept, which points out that: Fourier analysis of random signals can be discussed in limitation significance, namely, take out random signals $x(t)$ within limited time period $(-T \sim T)$, but the time is limited; therefore, there is Fourier transformation in $x(t)$. When $T \rightarrow \infty$, it approaches actual condition. Therefore, Fourier series can approximately express frequency domain of $x(t)$.

To implement digital signal processing, continuous random signals are usually sampled to form discrete random signals which are separated evenly, on account of its periodicity, can be dealt with Fourier series decomposition.

Therefore, $n_1(t)$ and $n_2(t)$ can be approximately expressed as:

$$n_1(t) = \sum_{n=0}^{\infty} \sigma_{1n} \cos(n\Omega t + \varphi_{1n}) \quad (2.298)$$

$$n_2(t) = \sum_{n=0}^{\infty} \sigma_{2n} \cos(n\Omega t + \varphi_{2n}) \quad (2.299)$$

where σ_{1n} and σ_{2n} are root-mean-square value; $\Omega = 2\pi/T_0$.

In Expressions (2.298) and (2.299), all spectral components are uncorrelated with each other and, when $T \rightarrow \infty$, become continuous spectrum.

For flat noise spectrum with $\sigma_{1n} = \text{constant} = \sigma_{10}$ and $\sigma_{2n} = \text{constant} = \sigma_{20}$, then

$$n_1(t) = \sigma_{10} \sum_{n=0}^{\infty} \cos(n\Omega t + \varphi_{1n}) \quad (2.300)$$

$$n_2(t) = \sigma_{20} \sum_{n=0}^{\infty} \cos(n\Omega t + \varphi_{2n}) \quad (2.301)$$

After being bandlimited by the band-pass filter BPF with bandwidth $2\pi B$ (unit is rad/s, unit of B is Hz) and center frequency at ω_0 , they become band-pass noise, namely,

$$\text{BPF}[n_1(t)] = \sigma_{10} \sum_{n=x}^{x+b} \cos(n\Omega t + \varphi_{1n}) \quad (2.302)$$

$$\text{BPF}[n_2(t)] = \sigma_{20} \sum_{n=x}^{x+b} \cos(n\Omega t + \varphi_{2n}) \quad (2.303)$$

where $x\Omega = \omega_0 - (1/2)2\pi B$ is $x = (\omega_0/\Omega) - (\pi B/\Omega)$, $(x+b)\Omega = \omega_0 + (1/2)\pi B$, that is $b = 2\pi B/\Omega$, we can obtain:

$$\begin{aligned} \text{BPF}[n_1(t)n_2(t)] &= \sigma_{10} \cos(n\Omega t + \varphi_{1x}) \cdot \sigma_{20} \sum_{n=x}^{x+b} \cos(n\Omega t + \varphi_{2n}) + \cdots \\ &\quad \sigma_{10} \cos[(x+b)\Omega t + \varphi_{x+b}] \cdot \sigma_{20} \sum_{n=x}^{x+b} \cos(n\Omega t + \varphi_{2n}) \end{aligned} \quad (2.304)$$

There are b terms to be added in Expression (2.304). After the calculation result passes through the low-pass filter LPF with bandwidth being B_L , we can obtain that the spectral component included in $n_0(t)$ is

$$\begin{aligned} n_0(t) &= \text{LPF}\{\text{BPF}[n_1(t)n_2(t)]\} = \sigma_{10}\sigma_{20} \cos(x\Omega t + \varphi_{1x}) \cdot \sum_{n=x}^{x+b} \cos(n\Omega t + \varphi_{2n}) + \cdots \\ &\quad \sigma_{10}\sigma_{20} \cos[(x+b)\Omega t + \varphi_{1(x+b)}] \cdot \sum_{n=(x+b)}^{x+b+L} \cos(n\Omega t + \varphi_{2n}) \end{aligned} \quad (2.305)$$

where $L = 2\pi B_L/\Omega$

There are $\left[\left(\frac{2\pi B}{\Omega} \cdot \frac{2\pi B_L}{\Omega} \right) - \sum_{m=1}^L m = 1 \right]$ terms in expression (2.305). As $B \gg B_L$, it is approximately $\frac{2\pi B}{\Omega} \cdot \frac{2\pi B_L}{\Omega}$ terms. Considering image frequency beat, the terms will be doubled, namely, there are $\sigma_{n_1 n_2} = \sqrt{\frac{8\pi^2 B B_L}{\Omega^2} \cdot \frac{\sigma_{10}^2 \sigma_{20}^2}{4}} = \sqrt{\frac{2\pi^2 B B_L}{\Omega^2}} \sigma_{10} \sigma_{20}$ terms.

According to this, when the bandwidth B of the bandpass filter grows, the beat terms falling in B_L will increase accordingly, that is physical reason why output noise will increase while B increasing. Therefore, the noise increment can be called as “beat loss.”

Amplitude of each beat term is $1/2\sigma_{10}\sigma_{20}$ and is not correlated mutually. Perform operation of mean square and root extraction to $(8\pi^2 BB_L/\Omega^2)$ terms to get output noise root-mean-square value $\sigma_{n_1 n_2}$ of the correlator, namely,

$$\sigma_{n_1 n_2} = \sqrt{\frac{8\pi^2 BB_L}{\Omega^2} \frac{\sigma_{10}^2 \sigma_{20}^2}{4}} = \sqrt{\frac{2\pi^2 BB_L}{\Omega^2}} \sigma_{10} \sigma_{20} \quad (2.306)$$

After $n_1(t)$ and $n_2(t)$ being bandlimited by bandwidth B , the variance $\sigma_{n_1}^2$ and $\sigma_{n_2}^2$ respectively are:

$$\sigma_{n_1}^2 = \frac{2\pi B}{\Omega} \sigma_{10}^2, \quad \sigma_{n_2}^2 = \frac{2\pi B}{\Omega} \sigma_{20}^2$$

therefore,

$$\sigma_{n_1 n_2} = \sigma_{n_1} \sigma_{n_2} \sqrt{\frac{B_L}{2B}} \quad (2.307)$$

④ Signal-noise ratio of output voltage in case of low carrier-noise ratio

(a) Output voltage signal-noise ratio $(S/N)_n$ of non-correlated noise

$$\left(\frac{S}{N}\right)_n = \frac{R_{S_1 S_2}(o)}{\sigma_{n_1 n_2}(o)} = \sqrt{\frac{2B}{B_L} \frac{m \sigma_{s_1}^2}{\sigma_{n_1} \sigma_{n_2}}} = \frac{\sigma_{s_1}}{\sigma_{n_1}} \cdot \frac{\sigma_{s_2}}{\sigma_{n_2}} \sqrt{\frac{2B}{B_L}} = \sqrt{\frac{P_{s_1}}{P_{n_1}}} \sqrt{\frac{P_{s_2}}{P_{n_2}}} \sqrt{\frac{2B}{B_L}} \quad (2.308)$$

(b) Output voltage signal-noise ratio $(S/N)_N$ of correlated noise

$$\left(\frac{S}{N}\right)_N = \frac{R_{S_1 S_2}(o)}{\sigma_{N_1 N_2}(o)} = \frac{\sigma_{s_1}^2}{\sigma_{N_1}} \frac{m}{M} = \frac{\sigma_{s_1}}{\sigma_{N_1}} \cdot \frac{\sigma_{s_2}}{\sigma_{N_2}} = \sqrt{\frac{P_{s_1}}{P_{N_1}}} \sqrt{\frac{P_{s_2}}{P_{N_2}}} \quad (2.309)$$

Quantization loss shall be considered in actual project. It is 0.5 in case of 1 bit quantization, 0.7 in case of 2 bit quantization, and 0.8 in case of the higher quantization.

According to Expression (2.308), in case of white noise, the bigger the bandwidth B of the band-pass filter is, the more the beat component of noise becomes, in which, output noise voltage in B_L increases in directly proportional to $\sqrt{B}$. However, the two signals are correlated and their beat component are added arithmetically within B_L , namely, they increase in direct proportion to B (in strict sense, only uniform signal spectrum is strictly direct proportional to B). Therefore, the output signal-noise ratio is in direct proportion to $\sqrt{B}$ within main energy range of the

signal spectrum. $(B/B_L) = K_B$ is known as bandwidth gain, therefore, a higher K_B can be obtained for wideband signal of B , which shall increase output S/N , and this is the reason why wideband angle tracking system can implement low C/N tracking. Therefore, in case of low C/N , the signal bandwidth shall be used as possible to implement angle tracking. If only a section of signal bandwidth is segmented, the output S/N within B_L will decrease. However, when the bandwidth B increases and exceeds signal spectrum, increment of B will not improve output signal voltage. But, the output noise voltage still increases according to $\sqrt{B}$, at this time, the output signal-noise ratio will decrease. Therefore, there is an optimum bandpass bandwidth B_{onT} to ensure a maximum output S/N . The value of B_{onT} is relevant to the spectral shape of the signal. Namely, there is different B_{onT} for different spectral shape. Such value can be obtained by simulation or experiment.

2) In case of high carrier-noise ratio

At this time, a section of signal bandwidth is segmented to perform angle tracking and thus simplifying the proposal. Within the intercepted narrower bandwidth B , the signal spectrum shall be deemed to be uniform. Since the carrier-noise ratio is extremely high, the Expression (2.293) is simplified as:

$$C(t) = S_1(t)S_2(t) + S_1(t)n_2(t) + S_2(t)n_1(t) + N_1(t)N_2(t) \quad (2.310)$$

where $S_1(t)$ and $S_2(t)$ are correlated, and $N_1(t)$ and $N_2(t)$ are correlated. Their correlated output expression has been solved as shown in expressions (2.296) and (2.297). $S_1(t)$ and $n_2(t)$, $S_2(t)$ and $n_1(t)$ in Expression (2.310) are non-correlated. Because of uniform spectrum, the derivation principle of correlated output is the same as that of non-correlated noise. Its formula is similar to the one output by non-correlated noise expressed by Expression (2.307).

① Correlated output of $S_1(t)$ and $n_2(t)$:

$$\sigma_{s_1} \sigma_{n_2} \sqrt{B_L/2B} \quad (2.311)$$

② Correlated output of $S_2(t)$ and $n_1(t)$:

$$\sigma_{s_2} \sigma_{n_1} \sqrt{\frac{B_L}{2B}} \quad (2.312)$$

③ Total non-correlated noise:

$$\sqrt{\sigma_{s_1}^2 \sigma_{n_2}^2 + \sigma_{s_2}^2 \sigma_{n_1}^2} \sqrt{\frac{B_L}{2B}} \quad (2.313)$$

When tracking is conducted, the sum signal $S_1(t)$ is larger than difference signal $S_2(t)$; the Expression (2.303) is approximately expressed as:

$$\sigma_{s_1} \sigma_{n_2} \sqrt{\frac{B_L}{2B}} \quad (2.314)$$

④ Output voltage signal-noise ratio in case of high carrier-noise ratio:

(a) Output voltage signal-noise ratio of non-correlated noise

$$\left(\frac{S}{N}\right)_n = \frac{\sigma_{S_1} \sigma_{S_2}}{\sigma_{S_1} \sigma_{n_2}} \cdot \sqrt{\frac{2B}{B_L}} = \frac{\sigma_{S_1}}{\sigma_{n_1}} \sqrt{\frac{2B}{B_L}} = \sqrt{\frac{P_{S_2}}{P_{n_2}}} \sqrt{\frac{2B}{B_L}} \quad (2.315)$$

where $\sqrt{P_{S_2}/P_{n_2}}$ is the signal-noise ratio of the segmented bandwidth B .

(b) Output voltage signal-noise ratio of correlated noise

$$\left(\frac{S}{N}\right)_n = \sqrt{\frac{P_{S_1}}{P_{N_1}}} \sqrt{\frac{P_{S_2}}{P_{N_2}}} \quad (2.316)$$

The foresaid noise includes two parts: ① antenna feed noise and LNA noise, ② sky noise (cosmic noise and atmospheric loss noise). The LNA is not shared, so LNA noise is non-correlated. But, sky noise is correlated with some antenna feed noise. In most applications, correlated noise is generated by antenna noise (except for deep-space detection) and determined by correlation of the patterns of sum and difference sidelobes touching to ground. Finally, it will be the minimum value of sum and difference cross-correlated output noise, namely, it determines the threshold value. This correlated noise cannot be deducted by prolonging integral time. However, non-correlated noise can be greatly weakened by prolonging integral time, so the output signal-noise ratio is finally determined by correlated noise.

If correlated noise $\sigma_{N_1}^2$ is $1/L$ of non-correlated noise σ_n^2 , the output noise after long integral time can be obtained through Expression (2.297): $R_{N_1 N_2}(o) = \frac{M}{L} \sigma_{n_1}^2$.

(3) Angle tracking error in case of cross-correlated angle error detection

The precondition of still and balance of tracking antenna is that the servo motor has no drive voltage; in this case, the output of angle error detector shall be 0, namely,

$$E_\Delta - E_n = 0 \quad (2.317)$$

where E_Δ is the signal voltage output by angle detection, E_n is the value of noise voltage output by angle detection; and the negative sign in the expression above indicates the servo system is negative feedback.

According to Expression (2.317), there is:

$$\frac{E_\Delta}{E_n} = 1 \quad (2.318)$$

The meaning of Expression (2.318): if there is noise E_n , the servo system will make the antenna deviate from $\Delta\theta$ to generate error voltage E_Δ , which offsets E_n to make output of angle error detector be 0, so as to enable the antenna achieve a new balance and stable point. When the instantaneous value of the angle detector

$E_{\Delta}/E_n = 1$, its r.m.s value equals to 1, which is determined by Expression (2.308), namely,

$$\frac{S}{N} = \sqrt{\frac{P_{\Sigma}}{P_{n\Sigma}}} \sqrt{\frac{P_{\Delta}}{P_{n\Delta}}} \sqrt{\frac{2B}{B_L}} - \frac{U_{\Sigma}}{U_{n\Sigma}} \cdot \frac{U_{\Delta}}{U_{n\Delta}} \cdot \sqrt{\frac{2B}{B_L}} = 1$$

then,

$$U_{\Delta} = \frac{U_{n\Sigma}}{U_{\Sigma}} \cdot U_{n\Delta} \cdot \sqrt{\frac{B_L}{2B}} \quad (2.319)$$

The angle error $\Delta\theta$ corresponding to U_{Δ} is determined by the following expression in angle measurement analysis (converted to feed outlet), namely,

$$\Delta\theta = \frac{\sqrt{G_{\Delta}}}{K} = \frac{\theta_{0.5}\sqrt{G_{\Delta}}}{K_m\sqrt{G_{\Sigma}}} = \frac{\theta_{0.5}}{1.4} \cdot \frac{U_{\Delta}}{U_{\Sigma}} = 0.7 \times \theta_{0.5} \frac{U_{\Delta}}{U_{\Sigma}}$$

where G_{Σ} is gain of sum beam; G_{Δ} , gain of difference beam; K , absolute difference slope; $K = K_m(\sqrt{G_{\Sigma}}/\theta_{0.5})$, $\theta_{0.5}$ is 3 dB width of sum beam; normalization difference slope $K_m = 1.2\text{--}1.7$, 1.4 is applied.

Substitute it into U_{Δ} in Expression (2.319)

$$\Delta\theta = 0.7\theta_{0.5} \left(\frac{U_{\Delta}}{U_{\Sigma}} \right) \cdot \left(\frac{U_{n\Sigma}}{U_{\Sigma}} \right) \sqrt{\frac{B_L}{2B}} \quad (2.320)$$

When the sum noise temperature is same as the difference noise temperature, i.e., $T_{\Sigma} = T_{\Delta}$, then $U_{n\Sigma} = U_{n\Delta}$ and

$$\Delta\theta = 0.7\theta_{0.5} \left(\frac{U_{n\Sigma}}{U_{\Sigma}} \right)^2 \sqrt{\frac{B_L}{2B}} = 0.7\theta_{0.5} \frac{P_{n\Sigma}}{P_{\Sigma}} \sqrt{\frac{B_L}{2B}} \quad (2.321)$$

(4) Four-channel monopulse proposal

The foregoing angle error signals will not serve as drive signal of angle tracking system until being normalized by sum signal. Additionally, when angle acquisition is conducted, it is required to detect amplitude of sum channel signal to serve as normalized reference signal and angle acquisition identification signal, which is usually extracted by sum channel signal. When the signal-noise ratio of sum channel signal is quite low, it will be a difficult problem. In that case, the advanced modern signal processing technology can be used to extract weak signal against the background of strong noise, such as cyclic spectrum technology and spectral line enhancement technology. In combination with monopulse angle tracking system, four-channel monopulse proposal is applied to extract such signal; namely, on the basis of typical “five-horn” 3-channel monopulse feed, the original four horns are employed to form an additional “reference” signal channel, as shown in Fig. 2.73.

Fig. 2.73 Four-channel monopulse feed

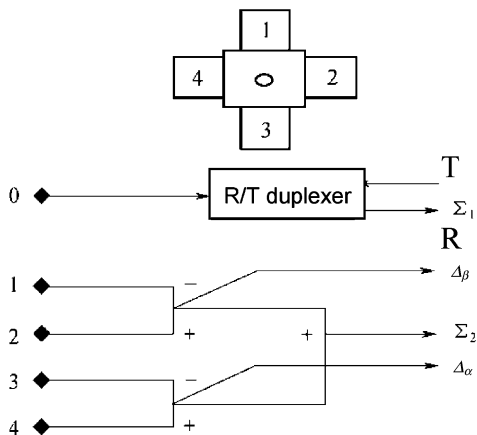


Figure 2.73 shows a five-horn feed, Σ_1 is sum channel, employing Σ_1 , Δ_α , Δ_β to form monopulse; Σ_2 is reference channel, Σ_1 and Σ_2 are used to perform cross-correlation detection of normalized signal and angle acquisition signal; Σ_1 serves as transmitting reference (like safety control signal). Σ_1 outputs P_Σ and P_Δ , $P_{n\Delta}$ outputs P_0 and $P_{n\Sigma 2}$, $P_{n\Sigma 1}$ and $P_{n\Sigma 2}$ are noise of sum channel and reference channel, which contains uncorrelated and correlated parts. Since they are generated by different LNA, horns, and feeders, therefore, the correlated part is relatively low, the SNR solved by cross-correlation are high extremely, and finally limited by relevant part of antenna noise.

When cross-correlation method is applied to extract normalized signals, the output signal-noise ratio expression may replace P_Δ and $P_{n\Delta}$ with P_0 and P_{n0} , we can get

$$\text{SNR} = \sqrt{\frac{P_\Sigma}{P_{n\Sigma}}} \sqrt{\frac{P_0}{P_{n0}}} \sqrt{\frac{2B}{B_L}} \quad (2.322)$$

Additionally, “square method” can be used to recover carrier (such as PSK signal). Square loss may exist when square method is used, however, when P_s/N_0 of the signal is relatively large, the single-carrier C/N_0 recovered by the squarer is still large and can be applied as normalized reference signal after being extracted by the carrier tracking ring, even serve as tracking signal of signal-channel monopulse. Since P_s/N_0 of wideband signal is very large, such method is likely to be used.

Furthermore, the possibility of non-normalization will be discussed in case of low C/N : with decrease in C , the angle error slope will decline and the dynamic performance of angle tracking will be deteriorated. When C is small, the distance is far. Therefore, C changes mildly with change in distance, so it can be coordinated with the servo system to make the servo system in normal operation within the specified variation range of the angle error slope. When C improves to achieve proper normalization, enter normalized operating state.

According to analysis above, for low C/N , angle acquisition and tracking can be implemented through the following proposal:

1) For non-periodic wideband signals: when there is no signal beacon (e.g., there are FM non-periodic signals only), normalization is key point, in this case, “four-channel monopulse” proposal can be employed. The threshold of the minimum receivable signal is limited by the correlated antenna feed noise between sum and difference channels. The improvement approach is to reduce relevance between the sidelobes of the antenna, i.e., the sidelobes of the two antennas. In addition, the sum and difference channel correlated noise cancellation method may be discussed. The other method is not to modulate FM signals at acquisition state, in order to make it be single-frequency carrier signals which can be acquired easily. Such signals will be modulated after being acquired and tracked.

2) For PSK wideband signals: “square method” may be applied to recover “single-frequency carrier,” and then angle acquisition and tracking may be performed for single-frequency carrier. Since there is square loss, therefore, it is required to recover P_{Σ}/N_0 . The acquisition and tracking requirements can be met. In case of the specified carrier-noise ratio, the wider the signal bandwidth is, the higher the corresponding P_{Σ}/N_0 is, so it may be applicable to rapid data transmission signals.

3) For S/Ka-band signals: where there are S-band signals, S-band wide beam may be used to guide Ka-beam narrow beam.

4) For periodic wideband signals: Theoretically speaking, “auto-correlation method” may be used to extract normalized reference signals. “Auto-correlation” required in-phase in carrier, but the frequency at Ka-band is high and there is Doppler frequency shift, so it is difficult to implement such extraction.

5) For single-frequency narrowband signal. Angle error detection of single-frequency signal is actually cross-correlation detection. However, due to tracking and filtering of phase-locked loop, $P_{\Sigma}/N > 1$, therefore, in case of the same P_{Σ}/N_0 , its random error is relative small compared with wideband signal and the acquisition probability is high. Although there is locking time, the technology is matured in terms of USB and deep-space detection, therefore, this method is the best choice in presence of signal beacons. In order to perform angle tracking and improve the system reliability in absence of beacon signal, the system may be provided with two tracking operation modes, namely, “beacon signal” and “wideband signal” modes.

2.5.5 Angle Measurement by Phased Array Tracking and Angle Measurement Accuracy – Space Window Sliding and Projective Plane Methods

2.5.5.1 Angle Measurement Principle by Phased Array Tracking

Angle measurement method used in phased array CW radar, compared with that employed in mechanical scanning radar, has some features due to the fact that the

phased array receiving antenna is of multichannel system and the antenna beam has phased array characteristics.

In terms of principle, the monopulse angle measurement method used in phased array CW radar is the same as monopulse precise tracking radar with mechanical scanning feature. There are three methods: amplitude contrast method, phase contrast method, amplitude and phase contrast method. These methods are used to conduct comparison and inner-angle insertion to the phase or amplitude of receiving signal from different beams, so as to improve angle measurement accuracy. Differently, mechanical tracking monopulse highly-accurate measurement radar commonly uses amplitude sum-difference monopulse system, the phased array CW radar generally employs phase sum-difference monopulse system. Amplitude sum-difference monopulse method requires that distance between two feeds shall meet $d < 2\lambda$, which is difficult to be achieved in constrained feed phased antenna. However, when monopulse phase contrast method to measure angel in phased array, it is easy to divide the entire antenna array into two or four parts, and to employ the entire receiving array to receive beam. In phased array antenna as shown in Fig. 2.74, two sub-antennas consisting of two sub-arrays can conduct phase sum-difference monopulse angle measurement.

In Fig. 2.74, the antenna array is equally divided into two sub arrays, with the distance between the two array phase centers of D , and the electrical boresight direction of θ . There is phase difference φ between the signals received by the two sub-arrays from θ direction, then,

$$\varphi = \frac{2\pi}{\lambda} D \sin \theta \quad (2.323)$$

θ can be obtained from measured φ .

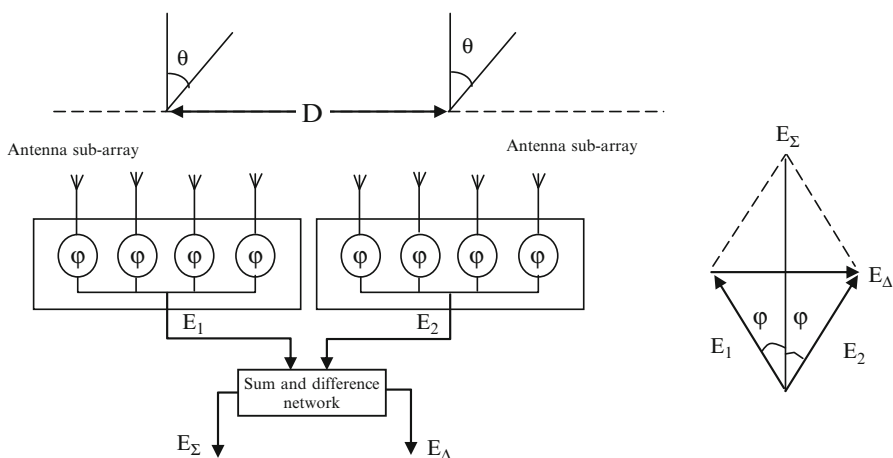


Fig. 2.74 Block diagram of phase sum-difference monopulse

In the phase sum-difference monopulse scheme as shown in Fig. 2.74, phase contrast is conducted between left and right sub-arrays in the sum-difference channel. The output differential signal E_{Δ} is relative to φ and θ . Meanwhile, the output E_{Σ} from the sum-difference channel is relative to θ . After the differential signal is normalized with $E_{\Delta}/E_{\Sigma} = K(\theta)$, the angle-sensitive function of the phase sum-difference monopulse can be obtained:

$$\Delta\theta = \frac{\lambda}{\pi D \cos \theta_0} \text{tg}^{-1} k(\theta) \quad (2.324)$$

where $k(\theta) = \frac{E_{\Delta}}{E_{\Sigma}}$ is known as normalized differential signal.

When phased array antenna performs targets continuous tracking, the targets will be in location nearby θ_0 during most of tracking period; therefore, when $\Delta\theta \ll \Delta\theta_{3\text{dB}}/2$, Expression (2.324) is simplified as:

$$K(\theta) = \frac{E_{\Delta}}{E_{\Sigma}} \approx \frac{\pi}{\lambda} D \cos \theta_0 \times \Delta\theta \quad (2.325)$$

$$\Delta\theta = \frac{\lambda}{\pi D \cos \theta_0} \times K(\theta) \quad (2.326)$$

However, when the phased array antenna conducts discrete tracking to the targets on a time-sharing multi-target basis, especially, when the tracking accuracy stays low and the angle of the target deviating from θ_0 is approximately $\pm \Delta\theta_{3\text{dB}}/2$, Expression (2.324) shall also be used to get deviation value of the target $\Delta\theta$. Accordingly, the target position θ is $\theta_0 \pm \Delta\theta$, where θ_0 is coarse measurement and obtained from phase control code of the phased array. $\Delta\theta$ is fine measurement and obtained from the angle-sensitive function Expression (2.324). The expressions above are the basis of phased array open-loop angle measurement.

In fact, angle measurement is carried out by antenna and receiver. The block diagram of the whole angle measurement system is as shown in Fig. 2.75.

In Fig. 2.75, the angle error output signal of the receiver is $\Delta u = K(\theta) = E_{\Delta}/E_{\Sigma}$, of which the normalization is achieved by sum-channel AGC, and phase change of E_{Δ} is detected by phase discriminator. Phase shift of $\pi/2$ is to make E_{Δ} and E_{Σ} be

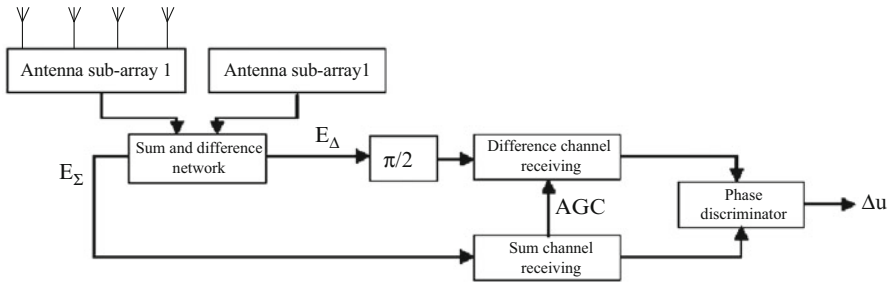


Fig. 2.75 Block diagram of phase sum-difference monopulse angle tracking receiving system

Fig. 2.76 Angle-sensitive feature of tracking receiver's output

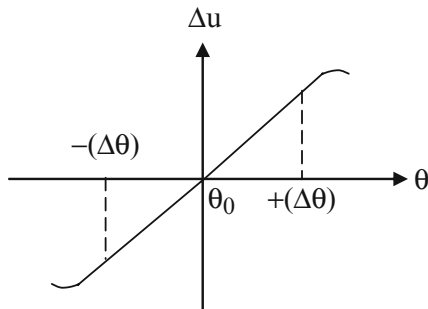
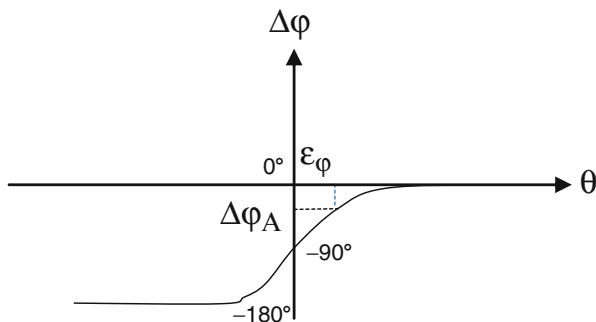


Fig. 2.77 Phase change between sum channel and difference channel



in-phase, so as the phase discriminator can be used to test angle error voltage. When the target is in the direction $\Delta\theta$ deviating from the electrical axis, the differential channel will output E_Δ , and the corresponding sum channel will output E_Σ ; when the target moves from the left side of the electrical axis to right side, phase of E_Δ will change 180° . The angle-sensitive function curve containing receiver is as shown in Fig. 2.76.

At this time, the angle-sensitive function is:

$$\Delta\theta = \frac{\lambda}{\pi D \cos \theta_0} \arctan(\Delta u), \quad (2.327)$$

In this angle measurement method, the phase difference of the signals received by two antennas spatially separated is used to determine deviation of the target from the beam pointing θ , and sum and differential beams are formed to achieve differential signal normalization and to judge the direction of the target deviating from θ , therefore, such method is also known as amplitude-phase monopulse angle measurement method. In such method, after phase-shifting of E_Δ , and when θ is changing from $-(\Delta\theta)$ to $+(\Delta\theta)$, the phase is changing from -180° to 0° , which will be -90° nearby θ_0 . Therefore, even if there is E_Δ , the output from the discriminator is zero. It is concluded that even if the antenna has null depth, output from phase discriminator still be zero to stabilize angle tracking. Phase reversion process is shown in Fig. 2.77, the deeper the null depth is, the smaller the transitions scope from -180° to 0° becomes.

The value of null depth affects position of $\Delta u = 0$ on θ axis, this is because the null depth is generated from inconsistency between sum/difference phase before comparator. The smaller the null depth is, the larger the inconsistency of the phase before comparator becomes. Additionally, the phases of the sum/differential channel of the receiver are also inconsistent. When the sum attains to $\pi/2$, zero point of Δu will exist (namely, angle tracking static stable tracking position). In case the sum-difference phase of the receiver is a give value, $\Delta\varphi_R$, the zero value of Δu will appear when the antenna sum-difference phase is $\Delta\varphi_A = (-\frac{\pi}{2} - \Delta\varphi_R)$. It is revealed from Fig. 2.77 that it is corresponding to ε_φ value, such value is angel tracking static error. $\Delta\varphi_A$ varies from different $\Delta\varphi_R$, ε_φ will change accordingly. The above section illustrates the physical concept of generating angle tracking static error by null depth. In case of amplitude sum-difference monopulse, it is expressed as following:

$$\varepsilon_\varphi = \frac{\theta_{3\text{dB}} \tan(\Delta\varphi_R)}{K_m \sqrt{G_n}} \quad (2.328)$$

where $\theta_{3\text{dB}}$ is 3 dB beam width of the antenna, K_m represents normalized difference slope, G_n is null depth of difference beam.

Based on the foresaid angle error open-loop measurement principle, the principle block diagram of phased array monopulse closed-loop angle tracking can be obtained as shown in Fig. 2.78. In this figure, the negative feedback signals in angle tracking loop is phase control value to each array element output from a beam controller. They control the phase discriminator of each array element to generate phase shift $\phi_1, \phi_2, \dots, \phi_N$, and the phase difference between adjacent array elements is $\Delta\phi_x$. When the target is on the direction of the antenna array electrical boresight, the spatial phase difference between adjacent array elements $\Delta\phi_x = 0$, at this time, the beam controller makes $\Delta\phi_x = 0$, so plus the same phases output from

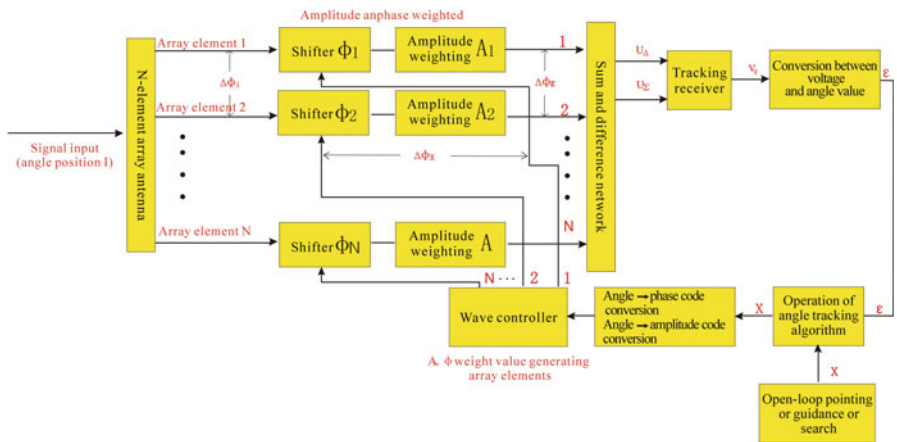


Fig. 2.78 Principle of phased array monopulse tracking

discriminators to synthesize a maximum gain. When the target deviates from the electrical axis, $\Delta\phi_i \neq 0$, the phase difference $\Delta\phi_e$ between the phase shifter output of adjacent array elements is $\Delta\phi_e \neq 0$. Applying the principle of angle error signals generation, the error signal V_Δ will be generated by the sum and difference device, which is amplified, processed and filtered by the tracking receiver to output angle error voltage V_e , which is converted into angle value ε by using angle-sensitive functions, and a new antenna pointing angle X shall be calculated with the angle tracking algorithm. X is converted to the new weight value φ and A of each array element of the phased array by beam controller. Such recursion algorithm shall be performed until $\Delta\phi_i \neq -\Delta\phi_x$, so $\varepsilon=0$, thereby angle tracking implemented. In fact, it is phase-shift tracking, and the phase shift value is corresponding to the angle measurement data.

2.5.5.2 Monopulse Tracking of Planar Phased Array – Projective Plane Method

“Open-loop angle measurement” or “closed-loop angle measurement” can be used for phased array angle measurement. The former, comparatively simpler, is based on angle-sensitive function as shown in Expression (2.327) and implement angle deviation measurement by measuring the output voltage of angle error, which obviously has a relation to accuracy of angle-sensitive characteristics. The latter has no relation to accuracy of angle-sensitive characteristics, because the output of angle error is zero after tracking is stable, and implements angle measurement with the phase value from the phase shifter, which has a high accuracy. For different target angle, the projective plane of planar array on radio wave equiphase surface will roll with the changing of the target angle. It is so called Projective Plane method, wherein gain loss of scan angle exists. The method shall be mainly described in this section. The analysis can be carried out in time domain or frequency domain.

(1) Time domain analysis

It is well known that the time domain analysis adopts difference equation, and if the current output of the equation is related not only to the current and past inputs, but also to the past output, it is the recursive equation. The phased array angle tracking is the case. The phased array angle tracking is usually implemented by a set of recursive equations calculation. The values of recursive functions are calculated based on the previous one or several calculation results, so this is a recursive step tracking method. Different recursive functions can compose different type of systems, such as Type I system (including one integrating element, Type I filter), Type II system (including two integrating elements, Type II filter), Type III system (including three integrating elements, Type III filter) and uses the Kalman filter, etc. There are different dynamic errors according to different types of loops. Such tracking are realized by optimal estimation through following parameters, including: ① error of observation, ② estimated beam position, ③ passing through smooth target location, rate, acceleration, etc. According to design requirements, such tracking algorithm may have different principles, e.g., minimum mean square

error criterion. The following introduces a common tracking algorithm based on recursive function:

1) Recursive tracking equation of Type I system

$$X_{n+1} = X_n + \alpha_0 \varepsilon_n \quad (2.329)$$

where α_0 position smoothing constant ($0 < \alpha_0 < 1$), and may slowly converge if it is too small and may oscillate if it is big; X_n is the angle value calculated from the n th sampling point; ε_n is the angle error measured from the n th sampling point; X_{n+1} is the resulted forecast angle value calculated from the $(n+1)$ -th sampling point.

The above expression is a recursive equation. To perform one recursive calculation within t , $1/t$ is used as sampling clock frequency. Continuous summing up (equal to integral) shall result in an angle error closing to zero, which realizes the angle tracking of no position static state error and is equal to one integration element. The higher the calculation frequency, the better the tracking dynamics is.

2) Recursive tracking equation of Type II system

$$X_{n+1} = X_n + \alpha_0 \varepsilon_n + t \bar{X}_n \quad (2.330)$$

where $\bar{X}_n$ smooth target angular velocity of the n th sampling point; t is sampling interval. The above expression adds velocity estimation component and increases velocity tracking accuracy, equal to 2 integration elements. One calculation method of $\bar{X}_n$:

$$\bar{X}_n = \bar{X}_{n-1} + \frac{\alpha_1}{t} \varepsilon_n \quad (2.331)$$

where $\bar{X}_{n-1}$ smooth target angular velocity of the $(n-1)$ -th sampling point; α_1 velocity smoothing constant, $0 < \alpha_1 < 4$, $0 < \alpha_0 < 1$; $\bar{X}_n$ also can be estimated according to other methods.

3) Recursive tracking equation of Type III system

$$X_{n+1} = X_n + \alpha_0 \varepsilon_n + t \bar{X}_n + \frac{1}{2} t^2 \ddot{X}_n \quad (2.332)$$

where $\ddot{X}_n$ smooth target angular velocity of n th sampling point. The above expression adds acceleration estimation component and increases acceleration tracking accuracy, equal to three integration elements. One calculation method of $\ddot{X}_n$:

$$\ddot{X}_n = \ddot{X}_{n-1} + \frac{\alpha_2}{t^2} \varepsilon_n \quad (2.333)$$

where $\ddot{X}_{n-1}$ smooth target acceleration of $(n-1)$ -th sampling point; α_2 acceleration smoothing constant, $0 < \alpha_2 < 8$, $0 < \alpha_1 < 16/9$; $\ddot{X}_n$ also can be estimated according to other methods.

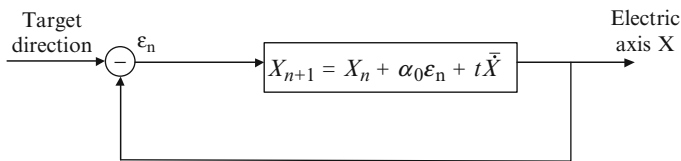


Fig. 2.79 Block diagram of phased array servo system

According to above tracking algorithm, block diagram of phased array angle tracking closed-loop system is as shown in Fig. 2.79.

In Fig. 2.79, $\ominus$ is accomplished by difference beam of phased array antenna. Difference signal, after passing through the filter, shall be processed with the recursive tracking algorithm, which is recursive calculation process (the digital transfer function is available by algorithm as PI, PID, etc.). X_n is the sum-beam electric axis direction θ_n , which is the same as zero direction of difference-beam and shifts according to Expression $\theta_n = \sin^{-1}(\frac{\lambda}{2\pi d}\Delta\phi)$ under the control of phase control code $\Delta\phi$ of phased array. ε_n is the angle difference signal. $\bar{X}_n$ is the smooth target velocity of the n th sampling point. t is sampling interval and α_0 is calculation step size. X_0 initial value of angle at the time of tracking start. When enabling the target to fall into the major beam with guidance, etc., there is X_0 and ε_0 and continuous recursive calculation can be done until tracking is stable.

(2) Frequency domain analysis

As everyone knows, frequency domain analysis of digital loop shall be conducted with Z transform and Z-transfer function. Such frequency domain analysis is more frequently applied because the technical specifications of angle tracking loop are generally reflected in its frequency response. However, time domain analysis is connected with frequency domain analysis by Z transform.

The following is the recursive tracking equation of first-order loop of time domain provided by Expression (2.329):

$$X_{n+1} = X_n + \alpha_0 \varepsilon_n$$

This is a difference equation, and is a digital integrator physically. Under control of digital error signal, digital-control beam is generated. In the equation, X_{n+1} is the $(n+1)$ -th output sampling point of integrator, ε_n is the n th input sampling point, and α_0 is proportion coefficient. In recursive operation, X_{n+1} is stored in integration register. Z transform for integrator with an unit delay is:

$$X(Z) = \frac{\alpha_0 Z^{-1} \varepsilon(Z)}{1 - Z^{-1}}$$

where, $X(Z)$ is the output, $\varepsilon(Z)$ is the input, and transfer function $K_A(Z)$ of corresponding digital integrator is

$$K_A(Z) = \frac{X(Z)}{\varepsilon(Z)} = \frac{\alpha_0 Z^{-1}}{1 - Z^{-1}} \quad (2.334)$$

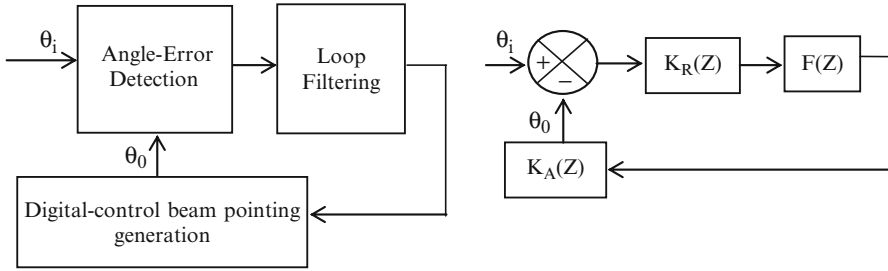


Fig. 2.80 Block diagram of phased array angle tracking loop

High-order loop should be applied to obtain better loop dynamic performance. The following shows a simple block diagram of such loop:

In Fig. 2.80, angle-error detection is conducted by difference beam and tracking receiver; its transform function is represented by $K_R(Z)$. Therefore, when receiver bandwidth is larger than loop bandwidth, it is a proportional element $K_R(z) = k_r = \text{constant}$. Digital-Control Beam Pointing Generation is conducted by the phase shifter that controls such phased array. It is a digital integrator. Its transform function is represented by $K_A(Z)$, while $F(Z)$ is the transform function of loop filter. Closed-loop transfer function $H(Z)$ of angle tracking loop is

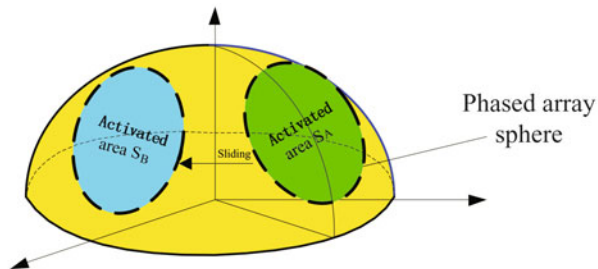
$$H(Z) = \frac{K_R(Z)F(Z)K_A(Z)}{1 + K_R(Z)F(Z)K_A(Z)} = \frac{K_r\alpha_0 Z^{-1}F(Z)}{(1 - Z^{-1}) + K_r\alpha_0 Z^{-1}F(Z)} \quad (2.335)$$

In Eq. (2.235), $K_A(Z)$ is the transform function of digital-control pointing angle generator. Multiple integral elements can be included in $F(Z)$. If only one integral element is included, the loop is called Type II loop. If two integral elements are included in $F(Z)$, it is called Type III loop. If no integral element is included in $F(Z)$, it is called Type I loop. The $F(Z)$ design method is already mature technology in PLL design. The design of phased array angle tracking loop can use many design methods in digital PLL.

2.5.5.3 Angle Tracking of Curved Surface Phased Array – Space Window Sliding Method

There are three defects for planar phased array. Firstly there is limited coverage of scanning angle with theoretical value of $\pm 60^\circ$; secondly, no conformity to carrier, which impacts the aerodynamic characteristics of aircraft and limits increase of phased array area; thirdly, no special beamforming and curved surface phased array is applied. Common curved surface phased arrays include: hemispheric array, paraboloid array, hyperboloid array, cone array, etc. As is well known, since phased array's price is high, cost effectiveness is one of the key factors for options. Cost is

Fig. 2.81 Schematic diagram of empty window sliding of sphere-phased array



mainly determined by array element number, which is related to surface area of antenna array. There are various performance specifications. If we make a comparison with the specifications of continuous azimuth coverage of 360° , elevation of 0° to $+90^\circ$ and the same gain, it shall be found that in the above schemes, when elevation changes from 0° to $+90^\circ$, the performance/price ratio will be: hemispheric array 0.19–0.38, paraboloid 0.19–0.48, hyperboloid array 0.19–0.62, cone 0.15–0.72. For elevation is 0° , the performance/price ratio decreases because of the only use of some area of curved surface, which makes gain decrease. One of the solutions is to expand phased array area downward and keep the effective area unchanged, e.g., on the basis of hemisphere surface, expand one section of cylinder downward so as to keep gain unchanged at low elevation. But for the triangular pyramid composed of three planar arrays, when scanning angle is 0° – 45° – 90° , performance-price ratio 0.15–0.22–0.15 and is below curved surface array. Generally speaking, curved surface array scheme is much better than planar array scheme.

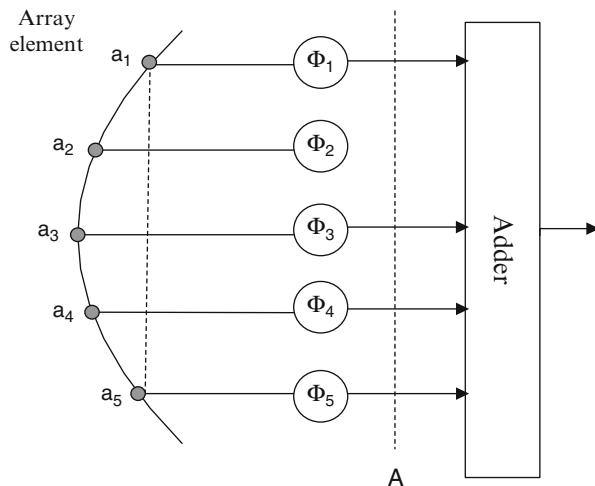
The actual work area (area producing antenna gain) of curved surface array is the area occupied by that array element activated by electromagnetic wave in beam direction, which is called activated area or irradiated area. As shown in Fig. 2.81, when activated surface moves from S_A to S_B (for processing of amplitude/phase signal at rear end of array element, that means a way of window sliding), beam pointing moves from θ_A to θ_B .

The activated area in Fig. 2.81 is equal to one paraboloid area. When it moves from S_A to S_B , it is looked like a parabolic movement, its sum beam, difference beam, and angle error characteristics are sliding correspondingly. So the angle tracking of sphere-phased array is “activated area sliding,” and the activated area is similar to one of the windows of the sphere-phased array. Therefore, it can also be called “space window sliding.” As the value of activated area wouldn’t change when sliding, there won’t be any scanning angle gain loss. So the scan, acquisition, and tracking antenna of the spheric array and parabolic antenna are similar. During spheric array design, we can refer to the parabolic antenna in its scanning, capturing, and tracking.

(1) Monopulse simultaneous lobe method

Different from planar array, in the curved surface array, even if the DOA direction is the same as that of normal line of irradiated side, the signal delay received by array element in irradiated area is different, which causes different output phase value of array elements, which shall be made a phase compensation to ensure in-phase summation of output signal of array elements, as shown in Fig. 2.82.

Fig. 2.82 Phase compensation of spheric array



Spheric array example is as shown in Fig. 2.82, after spatial delay difference of array element a_1, a_2, a_3, a_4, a_5 is compensated by phase shifter $\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_5$, it forms the in-phase signal at cable A and becomes the same as planar array, the subsequent sum and difference beam forms, which is the same as monopulse tracking of planar array in Sect. 2.5.5.2.

For the spheric array, spheric equation is known and according to geometry relationship of bow and string in circle, array element delay difference can be obtained and mathematical function relationship of array element phase shift and beam direction in spheric phased array is achieved.

For better illustrating of physical concept of phased array angle tracking of spheric surface, the following table describes the function relationship explaining the physical process of angle tracking. We use Figs. 2.81 and 2.82 to obtain the relationship of beam electric axis pointing and array element phase shift as shown in the Table 2.13.

The S_A and S_B in Table 2.13 are as shown in Fig. 2.81. When θ_A points, amplitude weighted value of all array elements except S_A is 0 and phase shift value of array elements $a_1 \dots a_5$ in S_A is $\varphi_1 \dots \varphi_5$. When the beam pointing is required to change to θ_B , set the amplitude weighted value of all array elements except S_B to be 0, phase shift value of array elements $b_1 \dots b_5$ in S_B is $\varphi_1 \dots \varphi_5$ (as spheric surface is symmetrical, its phase value is the same as that of S_A). Other pointing can be obtained in a similar way to realize phased beam scanning. For the purpose of monopulse angle tracking, it can form four monopulse subarrays in irradiated area S_A to form sum-difference phase monopulse angle. Then the angle tracking can be implemented with the phased array angle tracking system in Fig. 2.83, i.e., when signal is on the normal line X_{n-1} (on the direction of electric axis) of irradiated area, difference signal is zero; when it deviates from the direction of electric axis, it outputs one difference signal ε_{n-1} , from which we can calculate a new electric axis pointing angle X_n according to expression in Fig. 2.79, which corresponds to θ_n in

Table 2.13 Relationship between spheric beam electric axis pointing and array element phase shift

Electric axis direction	θ_A										θ_B		$\dots$		θ_n		$\dots$	
Illumination area S	S_A										S_B		$\dots$		S_n		$\dots$	
Array element included in illuminated area	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5	$\dots$		n_1		n_2	n_3	n_4	n_5
Compensation phase of array elements to planar array	φ_1	φ_2	φ_3	φ_4	φ_5	φ_1	φ_2	φ_3	φ_4	φ_5	$\dots$		φ_1		φ_2	φ_3	φ_4	φ_5

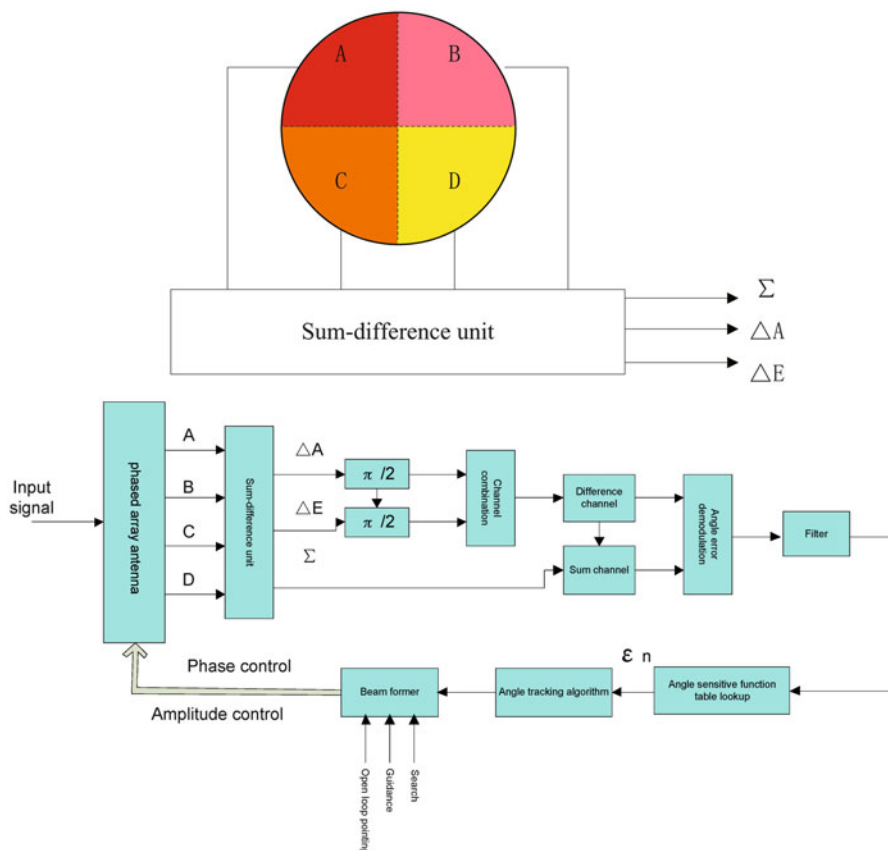


Fig. 2.83 Block diagram of phased array monopulse tracking system

Table 2.13. With the phase shift value and amplitude weighted value of array elements referred to the table, the first recursive tracking calculation can be performed and sum and difference beam can also be formed. Then, the recursive computation continues to make the sum and difference beam slide constantly until the tracking is stable.

In Fig. 2.83, A, B, C, and D are four monopulse subarrays in planar array and composite signal of subarrays form phase sum-difference monopulse, which is different from electromechanical scanning antenna.

The method given above, based on a table to find out the corresponding array element phase shift with the beam pointing θ_n , is clear in physical concept and is feasible, except for a requirement of mass operation. In order to reduce the workload, $\phi_n = f(\theta_n)$ function expression of different curved surfaces can be used. If it is applicable, phase control value ϕ_n can be obtained through real-time calculation of computer, which is the same as planar array. As for the signal amplitude value received by array element, as array element is on the curved

surface, the gain difference of array element antenna pointing makes signal amplitude difference between array elements bigger in comparison with planar array, which can be obtained from pattern. It also needs additional amplitude weighted value to obtain required sidelobe diagram or maximum S/N . Projected area slides with the sliding of electrical boresight of antenna and is realized by movement of amplitude weighted value, which continuously changes in 0–1. Through a method of maximum signal to noise ratio (SNR) synthesis weighting, the best synthesis gain will be available, which can refer to maximum ratio combination in Expression (2.354). When it adopts digital beam forming scheme, such amplitude-phase weight of continuous sliding is easy to be realized, and the phase jerk and amplitude jerk are small, so the velocity measurement, distance measurement, and the angle measurement accuracy are high.

(2) Phase scan sequential lobe method

The advantage of the above monopulse simultaneous lobe method is: for monopulse, using simultaneous lobe method can obtain the real-time, high-precision angle information, so as to obtain a high angle tracking accuracy. Its disadvantage is that its realization is complicated and it has high requirements on the amplitude consistency and phase consistency of four subarrays. The amplitude inconsistency or phase inconsistency before its sum-difference will cause angle tracking error. For an asymmetric curved surface array, such influence is more severe and it requires amplitude and phase compensation which will increase the circuit complexity. In order to simplify the circuit scheme, we can only use the amplitude of the antenna lobe and the sequential lobe scan to obtain the angle error signal if certain sacrifice of thermal noise angle tracking error is allowed. This scheme is similar to the conical scanning scheme of the mechanical scanning antenna. However, it is the electronic phased array scanning that is used in phased array antenna, and in lobe scanning scheme, binary scanning is implemented only in 2D of elevation and azimuth. Its realization is much simpler, and its angle tracking accuracy is higher than that of conical scanning scheme. Its working principle is shown in Fig. 2.84.

What is shown in Fig. 2.84 is the elevation one-dimensional electronic scanning. OA axis in the lobe diagram is the pointing axis, and it is also the measuring axis of angle measurement as well as the tracking axis of angle tracking. When the antenna is not scanning, the electronic axis of the beam will point at A, while in scanning, the beam will deviate upward from the electronic axis to an angle $(+\Delta\theta)$ (indicated by the beam location in full line in the diagram) within the time interval of ΔT_1 , and the beam will deviate downward from the electronic axis to an angle $(-\Delta\theta)$ (indicated by the beam location in dotted line in the diagram) within the time interval of ΔT_2 . Periodic binary $(+\Delta\theta, -\Delta\theta)$ scanning will be conducted according to this time interval. If the target is located at direction A of the pointing axis, it can be seen from the diagram that what is received by the antenna will be continuous wave U_0 within ΔT_1 and ΔT_2 . If the target is located at direction B above the pointing axis, the signal received by the antenna will be U_{1B} within the time interval of ΔT_1 and U_{2B} within ΔT_2 . From the diagram, we can know $U_{1B} > +U_{2B}$, so an amplitude-modulated wave will be formed, with a modulation degree of

$M_B = (U_{1B} - U_{2B}/2U_0)$ which is in direct proportion to the angle of deviation of the target deviating from the pointing axis. Therefore, the error signal can be obtained by detecting the amplitude-modulated wave. Similarly, when the target is located at the direction C below the pointing axis, an amplitude-modulated wave with a modulation degree of $M_C = (U_{1C} - U_{2C}/2U_0)$ will be received by the antenna. The value of M_C is also in proportion to the angle of deviation of the target deviating from the pointing axis, but $U_{1C} < U_{2C}$, so the phase of M_C modulated signal is opposite to that of M_A modulated signal. This means target A and target B are respectively located at the opposite direction of the pointing axis. Thus, the phase indicates the direction of the target deviating from the pointing axis. Therefore, we can obtain the angle error signal with positive and negative polarity by using the sync signal indicated in the diagram as the phase reference to detect the modulated signal synchronously.

The above-mentioned is the elevation one-dimensional situation. When implementing elevation and azimuth 2D angle tracking, the time division method can be used, i.e., implementing electronic scanning at elevation within the time interval of T_E , and taking out the elevation angle error signal from the tracking receiver synchronously, and keeping a sync time interval. While implementing electronic scanning at azimuth within the time interval of T_A , we can use the same method to take out the azimuth angle error signal. Refer to Sect. 2.5.1.3 for the principle of this time division multiplexing.

This scheme is similar to conical scanning scheme. Refer to relevant materials of conical scanning radar and the derivation method of Expression (2.262) for its angle tracking accuracy calculation.

(3) Phase weighting of curved array

In the above list method, it uses angle of spread (θ_n) to check corresponding array element phase shift (ϕ_n), which has a clear physical concept and makes it easy to understand. So next, let's further derive its mathematical expression based on the physical conception.

First, we will derive a mathematical expression of a curved array. Here, we put a curved array in any shape into a coordinate system OYZ that is plane and two-dimensional. In theory, the origin of such coordinate can be chosen by random. However, in consideration of the ease of engineering fulfillment, the choice is better to concentrate in the region which makes the curve seem to be symmetrical. Figure 2.85 shows a sketch diagram of any curved array placed in a coordinate system OYZ.

Then, we will derive, with geometric optics, the phase shift values that should be additionally added to each element in the random curved array. In Fig. 2.85, the equiphase surface (O_2OO_1) is formed by radio wave when propagates freely (i.e., with no sheltering from linear array antenna) in the space. Here, we take such equiphase surface as reference. Since such surface is perpendicular to the direction of radio wave propagation (in coordinate system OYZ, the angle of direction of radio wave is represented by θ), when elements O, O_1 , and O_2 are deployed in such surface, in-phase addition can be conducted for each of them receives the same signal phase information. However, if these elements are moved to positions (a_0 , a_1 ,

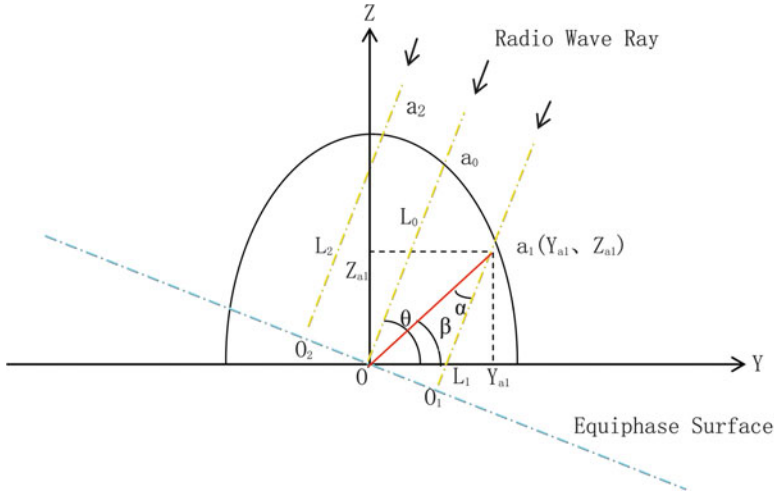


Fig. 2.85 Sketch diagram of geometric position of curved array in a two-dimensional coordinate

and a_2) on the curved array shown in figure above, such in-phase addition cannot be conducted, for there exist signal phase differences in space ($\Delta\phi_0$, $\Delta\phi_1$, $\Delta\phi_2$, corresponding to position a_0 , a_1 , and a_2), because geometric path length $(a_0o) \neq (a_1o_1) \neq (a_2o_2)$. Therefore, in-phase addition for signal at each element can be conducted only when phase shift value for each element is added. Next, let's take $\Delta\phi_1$ and $\Delta\phi_0$ and as examples to derive the phase shift value.

From the geometric relationship in Fig. 2.85, we can get:

$$\begin{aligned} (a_1o_1) &= (a_1o) \cos \alpha = (a_1o) \cos (\theta - \beta) \\ &= (a_1o) (\cos \theta \cos \beta + \sin \theta \sin \beta) = Y_{a1} \cos \theta + Z_{a1} \sin \theta \end{aligned} \quad (2.336)$$

Phase difference in space ($\Delta\phi_{a1}$) from distance is (a_1o) :

$$\Delta\phi_{a1} = (a_1o) \left(\frac{2\pi}{\lambda} \right) = \frac{2\pi}{\lambda} (Y_{a1} \cos \theta + Z_{a1} \sin \theta) \quad (2.337)$$

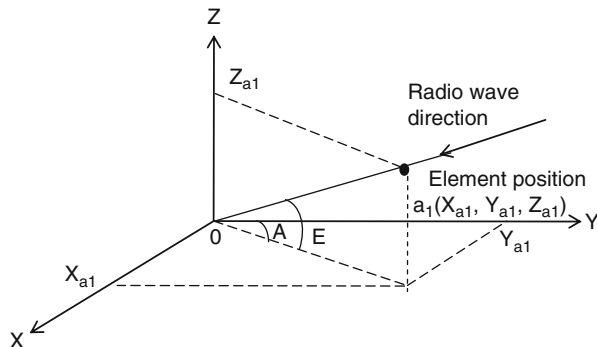
So phase shift value ($-\Delta\phi_{a1}$) must be added in the array to compensate such difference.

Similarly, we can derive:

$$\Delta\phi_{an} = \frac{2\pi}{\lambda} (Y_{an} \cos \theta + Z_{an} \sin \theta) \quad (2.338)$$

where Y_{an} and Z_{an} are the coordinate values of element a_n , and θ is the angle of spread of the beam. Therefore, the phase shift value needed by each element can be calculated with Expression (2.338).

Fig. 2.86 Sketch diagram of curved surface array in a three-dimensional coordinate



For a random curved-surface array, we can establish a three-dimensional coordinate system, and use the same theory to derive [22]:

$$\Delta\phi_{an} = \frac{2\pi}{\lambda} (X_{an} \cos E \cos A + Y_{an} \cos E \sin A + z_{an} \sin E) \quad (2.339)$$

where, $\Delta\phi_{an}$ is the phase shift value for element a_n , X_{an} , Y_{an} , and Z_{an} are the three-dimensional coordinate values of such element. If a random curved surface array contains n elements, there will be n sets of coordinate values in total for these elements. Such values can be determined at design and manufacturing phase, and listed and stored in a memory, from which they can be read during phased array operation and calculated with Eq. (2.339) to get the phase weight value for each element. The three-dimensional relationship of curved surface array is shown in Fig. 2.86.

In Fig. 2.86, (X_{a1}, Y_{a1}, Z_{a1}) are the coordinate values of element, A and E are, respectively, azimuth angle and elevation angle of beam pointing.

2.5.5.4 Time-Sharing Multi-target Tracking, Simultaneous Multi-target Tracking, and Adaptive Tracking

Time-sharing multi-target tracking of phased array uses TDMA principle and points antenna beam to different targets at different time slots, each target is distributed with fixed address code and time slots. In such time slot, phased array antenna beam dwells to accomplish angle measurement, distance measurement, remote measurement, remote control and multi-target TT&C. In the period other than beam dwelling target, it uses tracking filtering to predict the track of target in such period as the orbit prediction of next tracking. Beam directly jumps to prediction position for next tracking. When target velocity is V and interval of two beam irradiation is T , for the second tracking starting point, target coordinate will be limited in one spheric space with radius of TV . If such spheric space does not exceed beam width, the second tracking is very simple; if it exceeds, location

correlation and track correlation are needed. It can also use such time slot distribution to realize TWS, i.e., after discovery of first target, it distributes some time slots to track such target and after tracking is stable it distributes some time slots to search the whole tracking airspace. During the search, it uses tracking filter to extrapolate the first target. After searching the second target, it immediately tracks the second target; after tracking is stable, it starts searching on the next target. It does not stop searching and enter full tracking state until searching and tracking of all targets. As phased array TT&C system coordinates the target and target signal is known, orbit is known for orbit aircraft, technical difficulty is smaller. Its main technical problems are tracking filter and orbit prediction. Tracking filter algorithm is key point of multi-target tracking and ordinary contents include: two-point extrapolated filter, Wiener filtering, Kalman filtering, generalized filtering, etc. Simple tracking filter algorithm can be used for low velocity target. When TWS state changes into comprehensive tracking state, work becomes simpler.

Its main technical problems are: ① low sampling rate: low tracking sampling rate is feature of time-shared multi-beam phased array system. It adopts the minimum sampling rate adaptive to tracking filter design and specific target dynamic characteristics to save time. ② “Interpolative-zero value” tracking design: “interpolative-zero value” tracking is the feature of phase scanning radar system. It uses the movement mode of discrete step and low sampling rate to irradiate tracked target. Differential beam zero point hardly points to target on the preset beam position. It needs to insert one corrected increment for zero point to provide accurate angle data. Even if limited interpolation is needed, error slope (by angle error) shall be calibrated, then such corrected increment shall be plus (or minus) discrete beam pointing center. Monopulse is the most accurate technique of beam separation angle measurement and monopulse angle measurement is adopted for interpolation correction. ③ Initial tracking problem: initial tracking after capturing target requires the design of loop filter that makes tracking loop best match. During the initial tracking of target, it moves only with specified coordinate but without target, the specified coordinate is one step input of filter. It requires wider bandwidth and quicker sampling rate than stable work to finish transient process as soon as possible and ensure continuous tracking. ④ Design of loop filter: basis of tracking loop filter: dynamic characteristics of target, accuracy, power, utilization ratio of time (sampling rate required by continuous tracking), and calculation complexity under initial tracking and stable tracking. In various engineering designs, Type II, Type III, and Kalman filter are the commonest design of phased array system.

But for digital transmission of some high speed code (e.g., reconnaissance image), time division method may cause some limitation. Transponder storage and transmission can be used for time sharing transmission, which transmission bit rate is increased by N time (N is the number of multi-targets). If bit rate increase is limited, image compression can be used to decrease bit rate. If it still cannot meet technical requirements, simultaneous multiple beam can be used.

Simultaneous multiple beam tracking: each beam tracks one target and use multiple simultaneous beams for multi-target tracking; there are many tracking

loops and each tracking loop design is the same as single target, which has been introduced in Sects. 5.5.5.2 and 5.5.5.3. Its difference from time-sharing multiple beam tracking: there is continuous sampling for its angle error signal and sampling rate is high (1–100 kHz), time-sharing multiple beam tracking provides time-sharing tracking on multiple targets, but angle error sampling and processing of each target is time-sharing and discrete. Its sampling interval is the time interval on different sampling and sampling rate is low (0.2–10Hz). As time interval is long, “extrapolation,” “interposition,” and other tracking filter of moving targets are required to predict target position. For phased array C&T system, as remote metering, remote control, remote sensing, etc., is of continuous signal, time-sharing multiple beam is not used, but simultaneous multiple beam is used. In simultaneous multiple beam, there may be multiple targets in one beam, CDMA and FDMA can be adopted to differentiate different target signals.

Adaptive angle tracking is mostly used in digital transmission. Minimum mean square, CMA, recursion least square, and other adaptive algorithms are adopted for adaptive angle tracking or spatial filtering [23], according to principle of highest signal-to-noise ratio (or best signal/jamming ratio, etc.).

Adaptive antenna technology includes adaptive multi-beam forming, adaptive beam zero alignment, DOA, spatial spectrum analysis, early automatic alignment and integration of phase, etc. Automatic alignment of phase and adaptive beam forming can automatically accomplish beam forming and target capturing, which is closed loop tracking mode; but phase alignment mode cannot adaptively form spatial filter, which is its disadvantage. Contrarily, adaptive beam forming can adaptively form spatial filter and strengthen anti-jamming capability of equipment, which can obtain the maximum ratio of signal to noise. But calculation is complex, equipment capacity is too big, and a reference signal same as signal to be received is needed; convergence procedure of recursive calculation is too long with possibility of divergence. There are many principles for adaptive beam forming, including highest signal-to-noise ratio principle, maximum signal-to-interference noise ratio principle, LMS, etc. Its angle tracking accuracy is not high and is not used for angle measurement.

Space spectrum means that one time varying signal is composed of incoming wave components (e.g., incoming wave of mainlobe and sidelobe) of different spatial directions. Space spectrum analysis can analyze the components in all spatial directions, which can be used for measurement. The phase of array elements in phased array represents direction and phased array is good for spatial spectrum analysis.

Such above modes can be integrated for application; e.g., DOA or spatial spectrum analysis is used for direction guidance, multi-beam for tracking, LMS for anti-jamming.

2.5.5.5 Angle Measurement Accuracy of Phased Array Tracking

Tracking angle measurement accuracy is an important index in phased array TT&C system. Compared with normal monopulse tracking radar, it has no essential difference, but under the condition of multi-target tracking, it has higher relative

measurement accuracy for the common system errors and time differences are eliminated. As for absolute angle measurement accuracy, phased array decreases the error introduced in electromechanical link, and can quickly and flexibly change the servo loop parameters, so as to improve the dynamic performance. However, it also increases some error factors. In a planar array, the increased angle error factors are as follows.

(1) Scan angle factor

Beam width θ is inversely proportional to the cosine of scan angle Φ of the beam deviating from the array, i.e.,

$$\theta = \frac{\theta_0}{\cos \Phi} \quad (2.340)$$

where θ_0 is the width of the beam at the direction of array normal. When the beam deviates from the direction of array normal, beam width will increase and the measurement error item related to beam width (e.g., angle thermal noise) will directly increase. With the increase of beam width, its corresponding antenna gain will decrease, which will decrease the receiving echo SNR and increase the thermal noise error.

(2) Beam pointing random error

Due to such factors as array deformation, position error of array element, phase shift error of phase shifter, the fixed insertion phase shift error of phase shifter, the amplitude and phase of antenna aperture will have error, which will cause beam deformation and changes of phase center. Such errors are caused by fabrication error, changes of environmental conditions, etc., and will generate beam pointing error.

(3) Quantization error of phase shifter

Digital phase shifter is generally used in a phase shifter unit. It will make the phase distribution of the antenna aperture in a discrete form and generate aperture error which will cause quantization error to the beam pointing.

(4) Changes of phase center

Phase center is the reference point for ranging, angle measurement and velocity measurement, and its changes will induce orbit-measuring error. In a phased array antenna, the phase center will change with the changes of the arrangement of array elements, array deformation, beam direction, array element mutual coupling, fabrication tolerance, etc. Besides reducing the changes of such factors to the maximum extent in design and fabrication to improve the stability of phase center, measurement of phase center is also required. Angle and range calibration can also be applied to eliminate this error.

(5) Consistency and calibration of time delay and amplitude of each channel

When an active phased array scheme is applied, the consistency of time delay and amplitude of the T/R assembly of each array element will have great effect on the synthesis gain, beam pointing, beam shape of the phased array. The stability of time delay can induce ranging errors. Therefore, measures must be taken to reduce their inconsistencies. For a narrowband phased array system, changes of time delay

can be converted into changes of phase, so the amplitude and phase calibration technology is usually applied to ensure the amplitude and phase consistency of each channel. The angle measurement of a phased array angle tracking system is realized by the phase measurement of a phase shifter; therefore, phase calibration is very important in high-precision angle measurement. Generally, phase calibration inside the array is applied for the calibration of phase shift consistency of each array element channel. A better method is phase calibration outside the array. Its calibration includes the array and array element channel, so it is more accurate. Because the amplitude of each array element channel will influence the synthesis result, amplitude calibration is also required, to fully realize vector calibration. In adaptive tracking, S/N is required to be up to its maximum value, while this requirement can be complied with for amplitude and phase calibration can be automatically done under the adaptive function. Thus, phase calibration (or simplified phase calibration) is not required in theory.

For other angle measurement error, formulas related to electromechanical angle tracking system may be used for calculation (Table 2.14).

2.5.6 Independent Guidance, Self-Guidance, and Multi-beam Guidance

The first step of target acquisition by C&T system is to realize the acquisition of flight vehicles by ground station at angles. With the improvement of operating frequency band and the increase of antenna aperture, antenna beam becomes narrower and narrower, which makes the target acquisition by narrow beam a key technical issue. Angle guidance technology is required to be used. Some typical schemes will be introduced in the followings.

2.5.6.1 Independent Guidance

Independent guidance is a scheme which uses special guidance equipment (also called synchronous guidance) or other prior target position data (also called number guidance) to guide the antenna of the main C&T equipment (equipment to be guided) to point at the target spatial coverage, and its block diagram is as shown in Fig. 2.87.

In Fig. 2.87, the wide beam “guiding equipment” will first track the target, and the target position data acquired by it will be sent to the servo system of the “equipment to be guided.” It will be continuously used to guide “the equipment to be guided” to follow. The following duration is called continuous guidance time T . The guidance probability of this scheme is described by continuous guidance

Table 2.14 Angle error sources of phased array tracking

Error source	Error formula and affecting factors
1. Thermal noise	$\sigma_1 = \frac{\theta_{0.5}}{K_m \sqrt{2S/N}}$ where σ_1 is angle thermal noise error (mrad); S/N is SNR; K_m is normalized monopulse difference slope, $K_m = 1.89$ (typical value), $K_m = 1.2-2.0$ (general scope)
2. Multipath reflection	$\sigma_2 = \frac{\theta_{0.5}}{\sqrt{8A_3}}$ where σ_2 is angle multipath error (mrad); ρ is surface reflectivity; A_3 is the ratio of main lobe to side lobe at the direction of reflected ray
3. Dynamic lag	In case of type II tracking loop with critical damping, employ minimum mean square error criterion and exponential weighting: $\alpha_3 = \frac{\ddot{\omega} T_s^2}{a_1} (1 - a_0)$ where σ_3 is angle dynamic lag error (rad); $\ddot{\omega}$ is the maximum angular acceleration (rad/s²); T_s is sampling period (s); α_0 is position smoothing constant; α_1 is velocity smoothing constant
4. Errors caused by changes of amplitude and phase of monopulse feed source and receiver	In case of four-horn monopulse feed source: $\sigma_4 = \left[1 - G_1 \frac{G_2 \cos(a_1 + a_2 + a_3) + \sin(a_1 + a_3)}{G_2 \cos(a_2 + a_3) - \sin a_3} \right]$ where σ_4 is the error (mrad sine) caused by amplitude and phase inconsistency, G_1 is the amplitude inconsistency before comparator (gain ratio), α_1 is the phase inconsistency before comparator (°), G_2 is the amplitude inconsistency after comparator (gain ratio), α_2 is the phase inconsistency after comparator (°), α_3 is the phase inconsistency caused by other factors (°)
5. Error of electrical boresight direction	$\sigma_5 = \frac{\varepsilon \theta_{0.5}}{2\sqrt{N}}$ where σ_5 is pointing error (milliradian), ε is the value of root extraction of the phase error quadratic sum of each antenna unit (rad, which is caused by unit spacing, axis pointing shift, tolerance of phase shifter, cable, and feed), $\theta_{0.5}$ is 3 dB beam width (mrad), and N is the total number of array elements Caused by fabrication tolerance (mechanical and electrical), temperature, wind, sun, attraction, base deviation, and array deviation
6. Quantization of phase shifter	$\sigma_6 = \frac{2.6\theta_0}{N_L 2^P}$ where σ_6 is beam pointing error caused by phase quantization (m rad sine), P is the bit of phase shifter, N_L is the unit number of linear array

(continued)

Table 2.14 (continued)

Error source	Error formula and affecting factors
7. Digital angle quantization	$\sigma_7 = \frac{L_\Delta}{\sqrt{12}}$ where σ_7 is angle quantization error (rad) and L_Δ is the value of minimum effective bit (rad)
8. Scanning factor	Error caused by the changes of monopulse slope leaving the normal with scanning angle and the loss of scanning angle (based on the minimum mean square fitting)

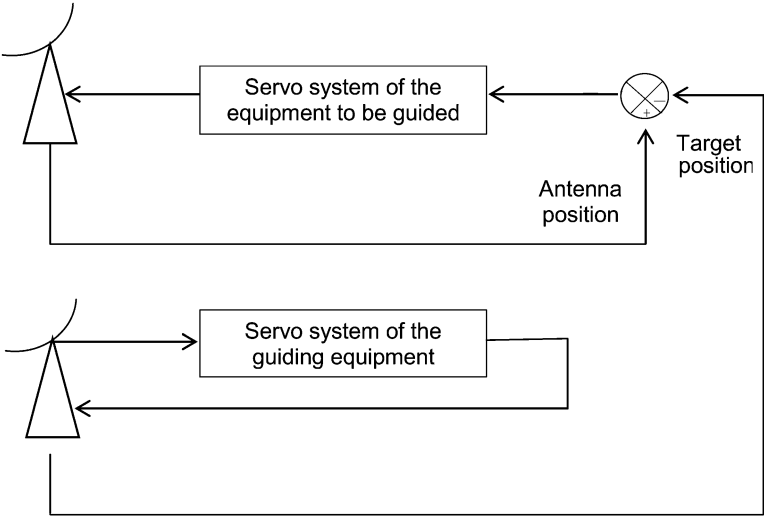


Fig. 2.87 Principle block diagram of independent guidance

probability [16], which can be analyzed according to guiding antenna tracking error along with the following assumptions:

- (1) The error of guidance system is random and stable.
- (2) Guidance probability is close to 1.
- (3) The average number of times that the target exceeds the air space within unit time is small.

When such assumptions are satisfied, the guidance probability that allows exceeding air space K times within guidance time T is approximately subject to Poisson distribution:

$$P_{K,T} = \frac{(\overline{NT})^K}{K!} e^{-\overline{NT}} \tag{2.341}$$

Expression (2.341) indicates: the probability of an event occurring K times in the time interval of $[0, T)$ is $\frac{(\bar{N}T)^K}{K!} e^{-\bar{N}T}$, where $\bar{N}$ is the mathematical expectation of event occurrence times within unit time.

Therefore, the guidance probability when the times that target exceeds the air space is less than j in continuous guidance time T is:

$$P_{j/T} = \sum_{K=0}^j P_{K,T} = \sum_{K=0}^j \frac{(\bar{N}T)^K}{K!} e^{-\bar{N}T} \quad (2.342)$$

It can be concluded from the above expression that in guidance time T , the guidance probability when not even one time that the target exceeds the specified air space is:

$$P_{0/T} = e^{-\bar{N}T}$$

$$\bar{N}_{(R>R_0)} = \frac{1}{\sqrt{2\pi}} \frac{\dot{\sigma}}{\sigma} \frac{R_0}{\sigma} I_0 \left(\frac{R_0 \sqrt{\Delta_A^2 + \Delta_E^2}}{\sigma^2} \right) e^{-\frac{1}{2\sigma^2} (R_0^2 + \Delta_A^2 + \Delta_E^2)} \quad (2.343)$$

where:

R_0 is the radius of specified spatial coverage which actually depends on the antenna beam width of the equipment to be guided;

Δ_A and Δ_E are, respectively, the azimuth systematic error and elevation systematic error of the angle of guiding equipment;

σ is random angle error of the guiding antenna;

$\dot{\sigma}$ is the root-mean-square value $\dot{\sigma} = \frac{2\pi\beta}{\sqrt{3}}\sigma$ of random error time change rate of Δ_A or Δ_E .

β is random error equivalent bandwidth (related to the bandwidth of the servo system of guiding equipment);

I_0 is the modified Bessel function of the first kind.

Independent guidance includes digital guidance, data guidance, program tracking, and so on. The Δ_A , Δ_E , σ , $\dot{\sigma}$ in Expression (2.343) are the errors between the angle guidance data and real angle of the target.

2.5.6.2 Self-Guidance

Self-guidance refers to the use of the wide beam antenna on the same antenna pedestal of the main C&T equipment to complete its narrow beam guidance. When the signal is guided into the narrow beam, the guidance mission is fulfilled. Then the next mission is handed over to the servo tracking system of the narrow beam in

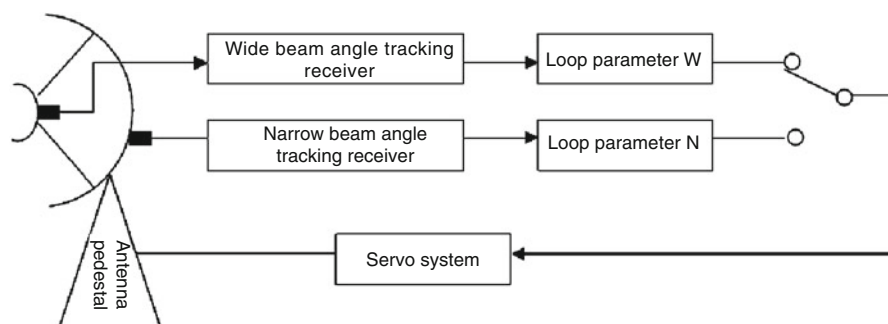


Fig. 2.88 Principle block diagram of self-guidance

“waiting for acquisition” state, by which the narrow beam acquisition and tracking are performed. This “waiting for acquisition” capability is provided by most narrow beam in modern TT&C equipment. The guidance probability of the self-guidance scheme is calculated according to the “hit probability” when the signal hits into the narrow beam. Such wide beam antenna includes: small aperture antenna of the same frequency, low frequency band wide beam of coplanar dual-band antenna, etc. Under ideal conditions, the electrical boresights of wide beam antenna (guiding antenna) and narrow beam antenna (antenna to be guided) are coincident. When wide beam antenna tracks and aims at a target, the angle tracking system will convert into narrow beam from wide beam to realize the self-guidance. However, there are random errors and systematic errors when wide beam antenna tracks a target, and these errors may lead to the result that the signal fails to drop into the narrow beam, which specifies the existence of the probability of guidance success. Take sub-reflector antenna as an example for the analysis of probability of guidance success.

In Fig. 2.88, a small aperture wide beam antenna of the same frequency is installed on the sub-reflector. Good electrical boresight consistency between the wide beam and narrow beam is required in this scheme. When the system starts to work, the target is made to enter the range of wide beam by using waiting for acquisition, scanning for acquisition, digital guidance and other means. The starting position of the switch in Fig. 2.88 is W , so the target will first enter the self-tracking range of wide beam antenna. At the same time, the narrow beam antenna will move along with the movement of the pedestal and stays in “waiting for acquisition” state. When the wide beam antenna tracks and aims at the target, the target is located at the range of narrow beam. The narrow beam tracking receiver will give an indication signal of “signal has been received and normalized” represented by the voltage of the angle error and AGC voltage. According to this signal, the switch in the Figure will be switched to N . The input angle error signal of servo system will be converted to the output of narrow beam tracking receiver from the output of wide beam tracking receiver. At the same time, the servo loop parameter will be changed. In this way, the angle self-tracking of narrow beam antenna is completed.

Fig. 2.89 Diagram of the cross section of the wide and narrow beam when there is no systematic error

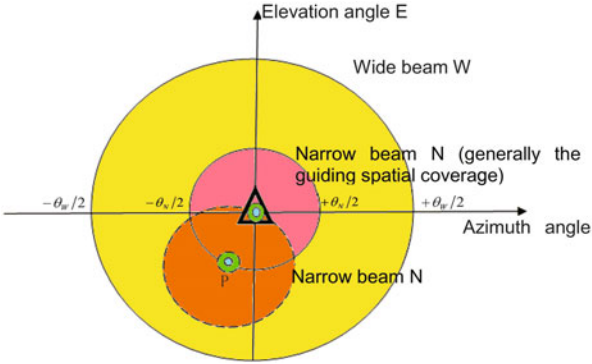


Table 2.15 Self-guidance probability

θ_N/σ	P
0.5	0.1
1	0.2
2	0.8
3	0.88
4	0.95

(1) Self-guidance probability if there is no relative systematic error or only random error existed between electrical boresights of the wide and narrow beam

At this moment, the electrical boresights of the wide and narrow beam is completely coincident, as shown in Fig. 2.89.

In Fig. 2.89, Δ is target position, width of wide beam is θ_w , position of electrical boresight is represented by a big circle “○,” width of narrow beam is θ_N , position of electrical boresight is the center of the narrow beam, represented by a small circle “○.” The required guiding spatial coverage is θ_N . When there is no relative systematic error, the big circle “○” and the small circle “○” are coincident and at the center of the circle. When there is random error of wide beam tracking, the relative position between the target and the electrical boresight will have a relative shaking. If this shaking is still within a circle with a radius of $\theta_N/2$, the lobe width of the narrow beam (marked by the dotted line in the figure) can still cover the target, which means the guidance succeeds, otherwise the guidance fails. The shaking is generally subject to normal distribution, the probability of the error less than $(-\theta_N/2 \sim +\theta_N/2)$ can be calculated based on statistical theory and it is:

$$P = 1 - \exp \left[-\frac{x^2}{2} \right] \tag{2.344}$$

Where P is called “hit probability,” or the probability of self-guidance, $x = \frac{\theta_N}{\sigma}$, σ is root mean square value of the angle tracking random error under the wide beam guidance. Partial calculated results of the above expression are shown in the following Table 2.15:

(2) The guidance probability with systematic error in guidance subsystem

The systematic errors here mainly include:

1) Static relative error θ_1 between the electrical boresight pointing of the narrow beam and the electrical boresight of the wide beam.

2) Sum θ_2 of dynamic lag error and other relevant systematic errors during wide beam tracking. The former is the main part of θ_2 , and the latter includes sum and difference phase inconsistency, amplitude unbalance before comparator, coupling error among channels, zero drift of phase discriminator, dead zone of servo system, unbalance and zero drift of servo amplifier, inconsistency of electrical boresight pointing change of wide and narrow beam caused by gust, etc. However, the conversion error and radio wave propagation error in monopulse angle measurement error will not affect the guidance probability of same-pedestal wide beam, and the systematic error which will maintain the same in a long time can be decreased by calibration.

3) In the conversion period T_s of wide beam guidance to narrow beam tracking, before the conversion of conversion dynamic error θ_3 caused by target movement, the PLL locking and AGC of the main tracking receiver have been completed, but in the period of T_s , the self-tracking loop of the wide beam is off, and the narrow beam loop is not connected yet, so θ_3 will be imported into the movement velocity of the target $\dot{\theta}_0$.

$$\theta_3 = T_s \dot{\theta}_0 \quad (2.345)$$

When there is velocity memory in wide beam angle tracking, and if the memory velocity is $\dot{\theta}_A$, then:

$$\theta_3 = T_s (\dot{\theta}_0 - \dot{\theta}_A) \quad (2.346)$$

For the current angle tracking equipment, the conversion can be quickly completed in computer, so T_s is very small. Besides, there is velocity memory, so θ_3 is a minor component.

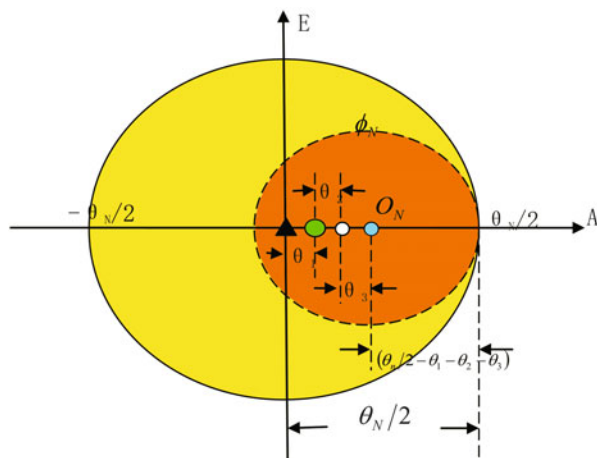
When there are θ_1 , θ_2 , θ_3 , the relationship among all kinds of errors in guiding spatial coverage θ_N is as shown in Fig. 2.90.

It can be seen from Fig. 2.90 that the electrical boresight of narrow beam will be located at O_N point, and the beam will shake around O_N point. In order to ensure 100 % success of guidance, the minimum value for which the shaking cannot exceed O_N is $(\frac{\theta_N}{2} - \theta_1 - \theta_2 - \theta_3)$, and the probability P' for not exceeding it is:

$$P' = 1 - \exp\left(-\frac{x_N^2}{2}\right) \quad (2.347)$$

$$x_N = \left(\frac{\theta_N}{2} - \theta_1 - \theta_2 - \theta_3\right)$$

Fig. 2.90 Diagram of the cross section of the wide and narrow beam when there is systematic error



In the above expression, the physical meaning of p' is the probability that the shaking of O_N point in Fig. 2.90 will not exceed the dotted circle ϕ_N . However, the shaking between the ϕ_N circle and θ_N circle also cannot exceed the guidance range θ_N . Set the probability as P'' , then the total hit probability is:

$$P_s = 1 - (1 - P') (1 - P'') \quad (2.348)$$

Obviously, $P_s > P'$, $P_s < P$ (refer to (2.344) for the expression concerning P). In order to improve P_s , it is required to take measures to decrease θ_1 , θ_2 , θ_3 .

The description above is the qualitative physical concept of P_s decrement. The quantitative computational formula of P_s is: [16].

$$P_s(R \leq R_0) = 1 - \Phi \left(\frac{R_0}{\sigma} \bullet \frac{\sqrt{\Delta_A^2 + \Delta_E^2}}{\sigma} \right) \quad (2.349)$$

where P_s is the hit probability for a single guidance, R is the angle of guiding antenna deviating from the target, R_0 is the guiding spatial coverage required, here it is narrow beam width θ_N , σ is random error of guiding subsystem, and Δ_A and Δ_E are the systematic error of guiding subsystem. $\Phi \left(\frac{R_0}{\sigma} \bullet \frac{\sqrt{\Delta_A^2 + \Delta_E^2}}{\sigma} \right)$ is an integral formula, which can be worked out by the general curve shown in Fig. 2.91.

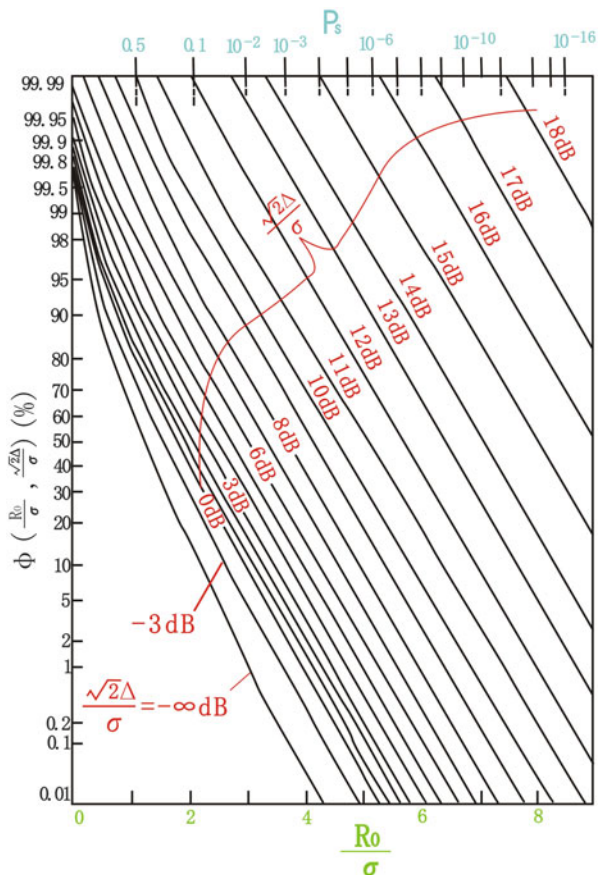
When N independent guidances are repeated, the guidance probability P_N is

$$P_N = 1 - (1 - P)^N \quad (2.350)$$

From the above, N guidances can increase the guidance probability, but it is required to meet the time interval T_G ($T_G = 2/B_L$, and B_L is the bandwidth of servo

Fig. 2.91 $\Phi\left(\frac{R_0}{\sigma} \bullet \frac{\sqrt{2}\Delta}{\sigma}\right)$

Curve (assumed

 $\Delta_A = \Delta_E = \Delta$)

loop) required by N independent guidances. The time spent for N independent guidances is NT_G , which means that the guidance probability is increased at the cost of extending time. But the target cannot fly outside the space domain within NT_G time.

Under certain exceptional circumstances, if the narrow-beam tracking system requires the wide-beam guiding subsystem to provide a relatively long continuous guiding time T_Σ to ensure the narrow-beam tracking shifts to auto-tracking reliably, Expression (2.343) for calculating continuous guidance probability will then be used. The exceptional circumstances include circumstances where the ability of narrow-beam angle tracking system to wait for capture is poor and the system needs to be guided for a period of time T_Σ , or where the delay time of the effective angle error voltage output of the narrow-beam tracking system is relatively long or even longer than the acquisition time of angle servo system.

According to the analysis above, the system error will lower the guidance probability and must, therefore, be reduced as far as possible. The measures for decreasing are: ① in case of long range, using narrow bandwidth for servo to

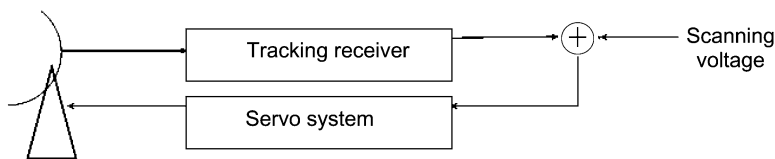


Fig. 2.92 Principle block diagram of “scan while tracking”

decrease random error for angle velocity is small and C/N is low; in case of short range, using wide bandwidth for servo bandwidth to decrease dynamic error for angle velocity is large and C/N is high; ② decreasing T_s and improving the velocity memory performance of the servo system to decrease the dynamic error in conversion period; ③ using fast servo system to decrease dynamic error; ④ improving the processing and installation accuracy of small antenna to ensure their electrical boresight consistency; and ⑤ adopting “calibrate while tracking,” “scan while tracking,” and other schemes to decrease systematic error. The principle block diagram of “scan while tracking” is as shown in Fig. 2.92.

In Fig. 2.92, the servo system of wide beam is in closed-loop tracking target state. It will overlap a small scanning voltage on the error voltage output by the tracking receiver while tracking the target, to make the antenna have a small scanning around the direction of tracking target. When the narrow beam scanning at the same time scans the target, it will convert into narrow beam tracking.

2.5.6.3 Multi-beam Guidance

(1) Overview of multi-beam guidance

Multi-beam guidance refers to using several adjacent narrow beams to realize the wide angle coverage and obtain angle error information, which is represented as the relative angle difference between each beam pointing and the main beam pointing.

The following are schemes for multi-beam guidance:

- 1) Planar phased array multi-beam: use phased array to form multi-beam for angle measurement and then guide into the main beam. It includes feed source phased array and face phased array. Its advantage is that it can realize a wide guiding spatial coverage, while its disadvantages are that the system is complicated and the cost is high.
- 2) Offset-focus array feed source multi-beam: use the principle of feed source offset-focus inducing beam deviation and use the offset-focus arranged array feed source to form multi-beam, i.e., each horn in the array feed source will form a beam. Its pattern is as shown in Fig. 2.93:

The following are options for using offset-focus array feed multi-beam antenna to realize multi-beam guidance:

- ① Open loop angle measurement re-guidance scheme: it uses the angle value of multi-beam open loop measurement based on parameter estimation theory to

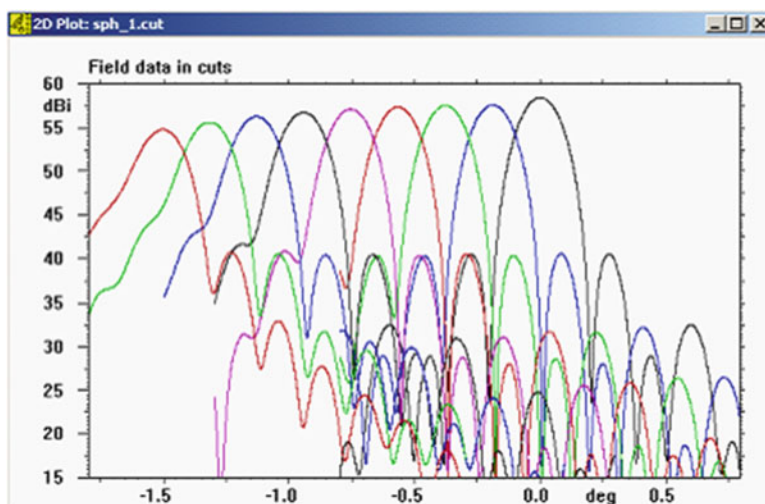


Fig. 2.93 An example of multi-beam pattern

predict the angle value of the target relative to main beam. This value will be regarded as the guidance data and sent to the follow-up servo system which is under guiding state. The system mechanically drives the antenna to make the main beam point at the target.

② Closed loop tracking convergence scheme: use “multi-beam closed loop tracking” to replace “multi-beam open loop measurement,” and the acquisition process of dynamic prediction will be changed into dynamic tracking process. It uses: (1) multi-beam feed source and its antenna; (2) high-speed sampling switch and multichannel Ka-band integrated channel; (3) signal processing unit with wide angle error characteristics, to collectively form a wide angle error characteristics corresponding to multi-beam angle coverage range (equivalent to a wide beam). Its output information is sent to the closed loop of servo system and the wide angle coverage tracking is realized based on discrete automatic regulatory theory. When the tracking converges to the coverage of main beam width, it converts to main beam tracking. Its principle is as shown in Fig. 2.94.

If there is any error of the target not being able to be aimed at when the convergence ends, the above-mentioned “scan while tracking” can be used as a solution.

This method can overcome the disadvantages of open loop measurement which require high accuracy and speed. During tracking process, the requirement on the accuracy of angle error characteristics is not high. As long as there is angle error voltage, the servo system can be driven to converge to the coverage of main beam width. And the sampling rate of signal is determined by the bandwidth of servo system. Due to the narrow servo system bandwidth, the sampling rate is low. Thus, time division multiplexing and equipment simplification can be realized. The scheme is simple and reliable.

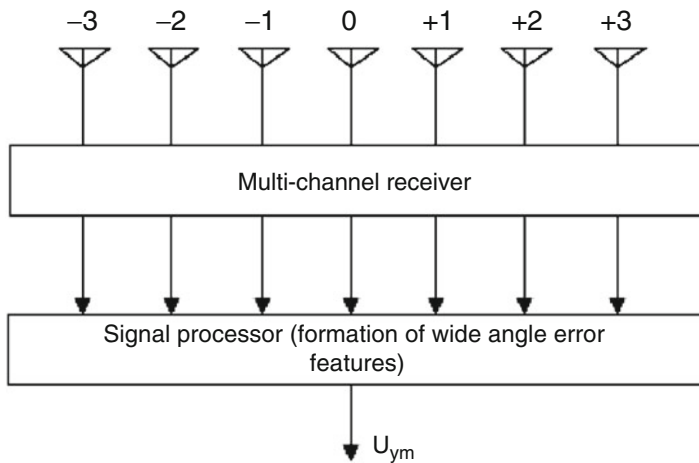


Fig. 2.94 Principle block diagram of multi-beam wide angle tracking

③ Multi-beam solution of phased array feeder: multi-beam can be formed by the parabolic feeder focus-deviated solution described above, and each beam corresponds to one angle information. Besides, angle information that is more precise can be further obtained by comparison of adjacent beams. However, such feeder focus-deviated will cause distortion of phase distribution, which will affect the directivity pattern of parabolic antenna, decrease the gain, widen the lobe and increase the sidelobe. Such negative influence will be decreased as the focus distance (f) gets longer, and relation between the deviation angle of beam ($\Delta\theta$) and focus (f), and relation between distance of horizontal feeder deviation (L) and focus (f) can be determined by the following equation:

$$\Delta\theta = \frac{L/f}{1 + (L/f)^2} \text{rad} \quad (2.351)$$

For energy collection by parabolic focusing, parallel radiations can be focused on the same focus only when their direction parallels with the focus axis; parallel radiations in other directions will be focused on difference focuses and will face defocusing. Therefore, there exists a focus area for acquisition of most energy, and such area can be planar or curved. If a phased array feeder is deployed in such focus area for phase and amplitude weight to match the field distribution of the focus plane, then the distortion of directivity pattern and decrease of gain can be effectively eliminated. Such phased array feeder has the following advantages:

- (a) Having a higher gain and supporting a larger scan angle under requirement of certain scan consumption.
- (b) Supporting continuous beam scan without beam loss at crossover caused by discrete beam in the solution of multi-beam of feeder focus-deviated.

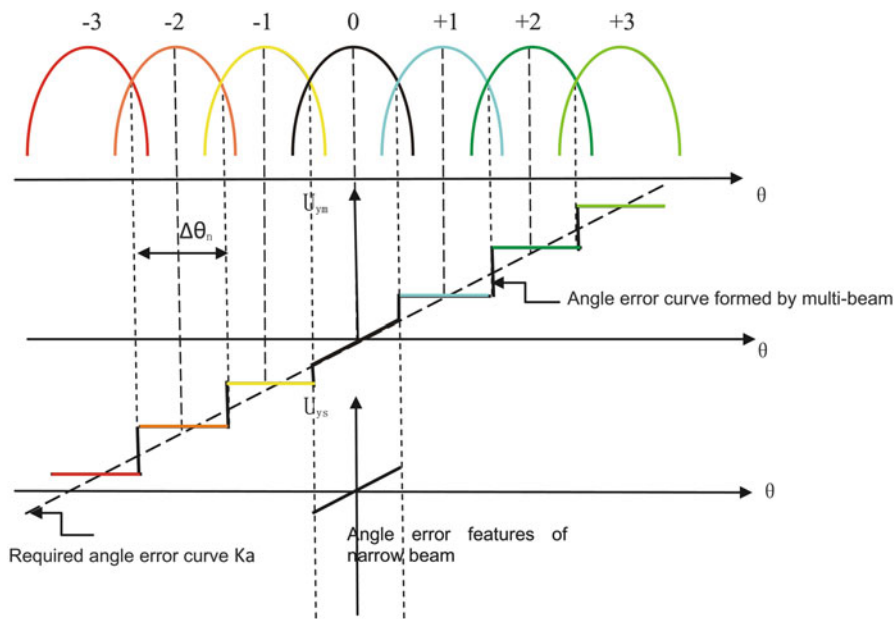


Fig. 2.95 Comparison between the wide angle error characteristics formed by multi-beam and narrow beam error characteristics

- (c) Supporting larger element interval when given a smaller scan angle, thereby decreasing its feeder array element.

However, its disadvantage is a more complex multi-beam forming.

(2) The principle of using multi-beam to generate wide angle error characteristics

Its principle is as shown in Fig. 2.95.

Figure 2.95 is one-dimensional. In the figure, -3 , -2 , -1 , 0 , $+1$, $+2$, and $+3$ are seven multi-beams and 0 is the main beam, which is one element of the multi-beams. The space pointing of each beam includes angle information. The difference between the n th beam pointing θ_n and the electrical boresight of main beam θ_0 is angle error θ . When target signal is detected within the n th beam, the angle error θ of the target relative to the electrical boresight of the antenna will be known. And then, according to requirements of servo system for the slope of angle error characteristics curve, signal processor will give the corresponding angle error voltage U_{ym} . Different beam corresponds to different U_{ym} , so as to form the wide angle error characteristics corresponding to multi-beam. For example, if a signal is received in the $+3$ rd beam (or $+n$ beam), judge whether there is a signal or not in the signal detection units according to amplitude adjustment criterion. If “there is a signal,” it indicates the target is in No. $+3$ beam, and the signal processing circuit will make it output a U_{+3} voltage. When there is a signal in No. $+2$ beam, the signal processing circuit will make it output a $U_{+2} = \frac{2}{3}U_{+3}$ voltage. When there is a signal in No. $+1$ beam, the signal processing circuit will make it output a $U_{+1} = \frac{1}{3}U_{+3}$ voltage. When there is a signal in No. 0 beam, the output is error voltages within the

main beam. Similarly, U_{-3} , U_{-2} , U_{-1} can be obtained. Because the multi-beam does not continuously cover the airspace but be quantized by each beam width, there is quantization error so as to form the angle error characteristics of stepped appearance as shown in Fig. 2.95. It brings two kinds of influences. One is to cause angle quantization random error but it has little effect on the acquisition. “Amplitude comparison angle measurement” may be used as the supplement of decrease of quantization error, i.e., using “beam position number” as large number and the value of amplitude comparison angle measurement as decimal. The other one is nonlinear, which is equivalent to the phenomenon when there is bias between the angle error slope output by angle tracking receiver and the slope required by servo system. When there is such nonlinearity, the tracking system can still converge the tracking, but may affect the dynamic quality.

What are shown in Fig. 2.95 are the angle error characteristics of multi-beam and the angle error characteristics of the main beam. It can be seen that multi-beam largely increases the scope of angle error characteristics so as to increase the angle acquisition scope. When multi-beam antenna angle tracking converges to the main beam, the main beam channel will receive signals. Signal processing unit will judge “there are signals,” and the multi-beam closed loop will be converted into main beam closed loop to realize narrow beam self-tracking. The tracking state of the n th multi-beam and 2D antenna can be obtained in a similar way.

(3) Use sampling discrete servo system to realize wide angle acquisition and tracking—closed loop discrete automatic tracking method

The explanation of the above operating principle is based on that each beam horn can output continuous signals to form a continuous automatic regulatory system. This system requires that each horn shall have 1-channel receiver and its signal detector, which makes the equipment very complicated. A simplified method is to use error signal discrete sampling to form discrete automatic regulatory system, as shown in Fig. 2.96.

In Fig. 2.96, the sampling switch is a fast electronic active switch. Low gain LNA₁ can be used to decrease the effect of insertion loss of sampling switch on G/T value. The HF amplification of reception channel is mainly provided by high gain LNA₂ after the sampling switch. As there's only one LNA₂, the total stages and price of LNAs are largely reduced. Voltage holding time is designed longer than the sampling period T design of each beam. If there is voltage input in the memory time, update the voltage holding value. If there is still no signal at the end of memory time, zero set the holding voltage and make the angle servo system open loop to restart the acquisition.

If signals are detected in a beam when sampling switch is used for the discrete sampling of multi-beam, a pulse signal will be outputted. The width of the pulse signal is the sampling residence time τ at that beam and pulse period is the sampling period T of that beam. In this way, a discrete automatic regulatory system is formed. There is no error signal in the time interval of $(T - \tau)$, and it is required to use a voltage holding circuit to make the continuous output of error signals. If the voltage remains unchanged in $(T - \tau)$, it is called “zero-order hold.” If it presents linear slope, it is called “first-order hold.” The performance of the latter is better than of the former. Discrete automatic regulatory theory points that, if sampling frequency $1/T$

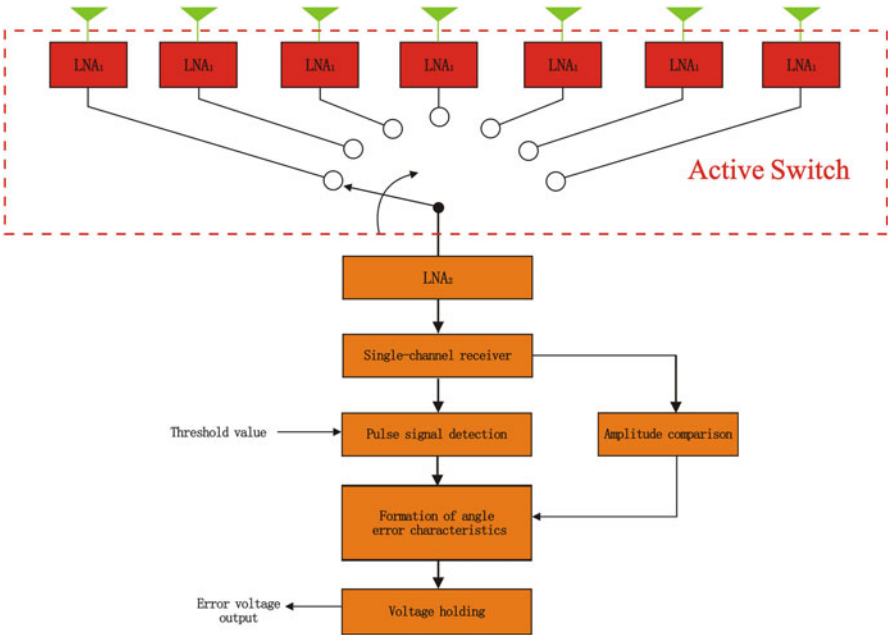


Fig. 2.96 Principle diagram of sampling discrete angle tracking

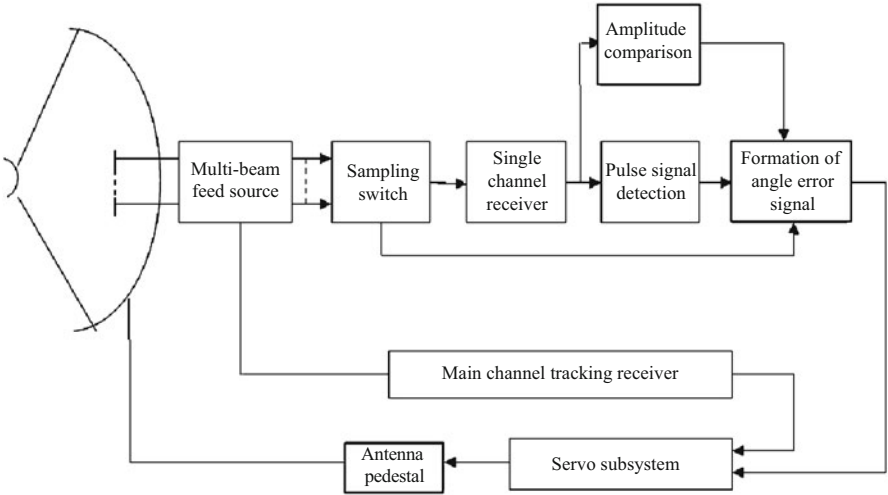


Fig. 2.97 Simplified block diagram of sampling discrete angle tracking system

is too low, large tracking error and poor dynamic quality will be caused. When sampling frequency $1/T$ is (5–10) times larger than servo bandwidth, the phase lag of “zero-order holder” is small and can be omitted. At this moment, the discrete automatic regulatory system can be handled approximately as the continuous automatic regulatory system. Block diagram of this scheme is as shown in Fig. 2.97.

When the signals received are too strong, the adjacent beams have the possibility to receive signals. An amplitude comparison circuit can be used for the amplitude comparison of such signals. The strongest one is determined as within that beam. Meanwhile, the amplitude ratio of adjacent beams can also be used in angle measurement by comparing amplitude, which can be a supplement to acquire more accurate angle error characteristics.

The detection of sampling signal is the detection of a single pulse signal whose width is τ and period is T . Different from radar waveform design, it is required to complete the detection of n multi-beam received signals within T . When switching time of the sampling switch is too short, the residence time of each beam

$$\tau = \frac{T}{n} \quad (2.352)$$

where τ is pulse width. Generally, the signals in τ can be a single carrier or a modulated carrier (e.g., FM, QPSK).

As shown in Fig. 2.97, angle error signal corresponding to a sub-beam pointing is acquired when pulse signal detector detects that sub-beam has an pulse signal output (under the action of sampling by switch, the output is an pulse signal) for the first time. If loop of servo system is closed when that pulse signal output is detected for the first time, the guiding system will then enter into the process of closed-loop tracking convergence. At that moment, under the action of that angle error signal, the loop will start automatic adjustment and tracking will gradually converge to the main beam at the center of multi-beams to finally complete angle guidance. As long as there is an error signal, automatic adjustment is possible. In this scheme, error signal can be acquired and loop can be closed as long as a single pulse signal is detected. So the guidance probability is mainly determined by the detecting probability of single pulse.

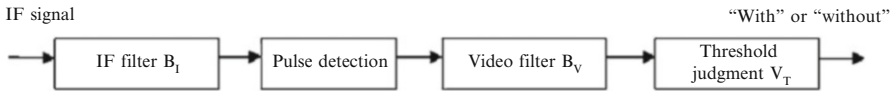
When loop of servo system is closed, a series of pulses (n pulses) will act on the closed-loop system subsequently. $n = T_r f_c$, where T_r is the time for convergence, and f_c is the sampling frequency. The n pulses can generate n angle error voltages and the voltage that is finally applied on the servo system is the mean voltage E of those n error voltages. However, some pulses may be lost due to the impact of single pulse detection probability P_D . The number of lost pulses is $n(1 - P_D)$. The losing of pulses will make E decreased to $\frac{n-n(1-P_D)}{n}E = P_DE$, which will then make the angle error slope K_R decreased to $P_D K_R$. Since open-loop gain of loop is $K_a = K_R K_0$ (K_0 is a known design value), K_a will also be decreased to $P_D K_a$. The decrease of K_a will have the following influences:

1) Servo loop bandwidth B_n becomes narrower: this is due to [19]. For type II servo system, $B_n = \sqrt{K_a/3.5}$.

2) $\sigma\% \leq 10\%$ The time for convergence becomes longer: the fast convergence time of servo system is the rise time t_r (the time for reaching 90 % of the stable value) during its transient process. When overshoot $\sigma\%$ in the transient process of servo system is equal to or less than 10 %, i.e., $\sigma\% \leq 10\%$, $t_r \approx \frac{1}{2B_n} = \sqrt{\frac{3.5}{4K_a}}$; when $\sigma\% \geq 10\%$, $t_r \approx \sqrt{\frac{2.5}{4K_a}}$

Table 2.16 Relationship between detection probability and IF SNR

$(S/N)_I$ (dB) \ P_d	10^{-3}	10^{-4}	10^{-5}	10^{-8}	10^{-10}	10^{-12}
0.8	10.4	11.5	12	13.5	14.5	15.2
0.9	11	12	12.5	14.2	15	15.8
0.95	11.5	12.3	13	14.5	15.5	16.1
0.98	12.2	13	13.5	15.5	16	17
0.999	13.8	14.5	15	16.2	17	17.5

**Fig. 2.98** Pulse signal detection model

3) Dynamic error Δ_a becomes greater: for type II servo system, $\Delta_a = \ddot{\theta}/K_a$, where $\ddot{\theta}$ represents the angular acceleration of target. When Δ_a is less than half of the width of main beam, servo system can shift to main-beam tracking, in which case the guidance is successful; when is less than half of the multi-beam coverage angle, servo system works under the continuous guiding state of tracking convergence process θ_w ; when $\Delta_a > \theta_w/2$, the target is beyond the guiding range, which will lead to guidance failure.

Generally, value of P_D is designed to be a great one and approximate to 1. It has minor impact on convergence and will not cause failure of guidance.

For the same reasons as above, the stepped appearance of angle error as shown in Fig. 2.95 will also make the angle error slope deviate from its optimum value K_a as designed. And the decrease of K_a will have influences similar to those mentioned above. The difference lies in that the stepped error voltage is formed due to the fact that the error characteristics of linear K_a is quantized by sub-beam width $\Delta\theta_n$. For each $\Delta\theta_n$, however, the mean error voltage generated by linear K_a is equal to the stepped error voltage corresponding to that $\Delta\theta_n$. During target tracking and convergence process, the angle of deviation of target is variable, and its error voltage also varies; so, the main component that is finally applied on narrowband servo system is the mean error voltage smoothed by narrowband filter. Therefore, the impact of stepped error on guidance will not cause the failure of guidance.

The acquisition probability of this scheme can be given based on the detection theory of pulse radar signal. The detection theory of pulse radar signal [24] gives the following data of detection probability P_d , false alarm P_f , and IF SNR $(S/N)_I$.

The data given in Table 2.16 is based on the detection model of Fig. 2.98.

In Fig. 2.98, the input IF signal is a single-carrier pulse signal, B_I is the bandwidth of IF filter, B_V is the bandwidth of video filter after detection, and the

optimal bandwidth is $B_I = \frac{1}{\tau}$, $B_V = \frac{0.5}{\tau}$. In radar signal detection analysis, assume $B_V \geq B_I/2$, i.e., mainly decided by IF signal envelope. S/N also corresponds to IF SNR $(S/N)_I$.

When IF signal is a modulated signal, its bandwidth B_I is usually very large, and narrow band filter cannot be added at IF, or signal energy will be lost. The optimal bandwidth of video signal is decided by pulse width τ , $B_V = 1/2\tau$, because $1/2\tau < B_I/2$, the envelope of video signal is mainly decided by $1/2\tau$. Because the narrowing of video bandwidth, the video SNR $(S/N)_V$ is improved. But in the case of low $(S/N)_I$, amplitude detection will cause square loss. The expression of detection output $(S/N)_V$ is:

$$\left(\frac{S}{N}\right)_V = \left(\frac{S}{N}\right)_I^2 \frac{B_I}{2B_V} = \left(\frac{S}{N}\right)_I^2 B_I \tau \quad (2.353)$$

P_d and P_f are determined by video signal envelope. After $(S/N)_V$ is calculated using the above expression, replace $(S/N)_I$ with $(S/N)_V$, then the approximate value of P_d and P_f will also be calculated. In real project design, computer simulation or hardware test are adopted to calculate P_d and P_f .

To sum up, the design method of this scheme is: when servo bandwidth B_n , IF bandwidth B_I , IF SNR $(S/N)_I$ and multi-beam data n are known: ① calculate T by $1/T \geq (5 \sim 10)B_n$; ② calculate τ by $\tau = T/n$; ③ calculate $(S/N)_V$ under the required acquisition probability P_d and false-alarm probability according to Table 2.16; ④ calculate the required $(S/N)_I$ by Expression (2.352).

If the calculated $(S/N)_I$ cannot meet the requirement, increase τ to improve $(S/N)_I$ so as to decrease n . In this way, array feed source will be divided into several sub-arrays for parallel processing and n of each sub-array will be decreased.

(4) Use “sequential sub-lobe” and “double time division” to realize wide angle acquisition

Based on “discrete automatic regulatory theory,” the above scheme uses time division sampling to decrease the number of channels. At the border of acquisition spatial coverage, the sampling interval cannot be longer than the dwell time of the target at narrow beam, or the target will fly out of that beam at the second sampling. There is another thought which uses “sequential lobe method” theory to divide the multi-beam array feed source into several sub-arrays, and each sub-array correspond to one sub-lobe. Conduct time division sequential scan to the sub-lobe to cover a wide acquisition spatial coverage. The time division sampling interval cannot be longer than the dwell time of the target at the sub-beam, so its sampling interval may be long. There are a lot of channels in the sub-beam required, and parallel processing is required.

In system design, comprehensive comparison may be conducted according to the features of the two kinds of “time division multiplexing” schemes, or integrate the advantages of the two by using “double time division multiplexing” method, i.e., using “discrete automatic regulatory theory” for the first time division and the sequential lobe theory for the second time division. One principle block diagram of “double time division multiplexing” scheme is shown in Fig. 2.99.

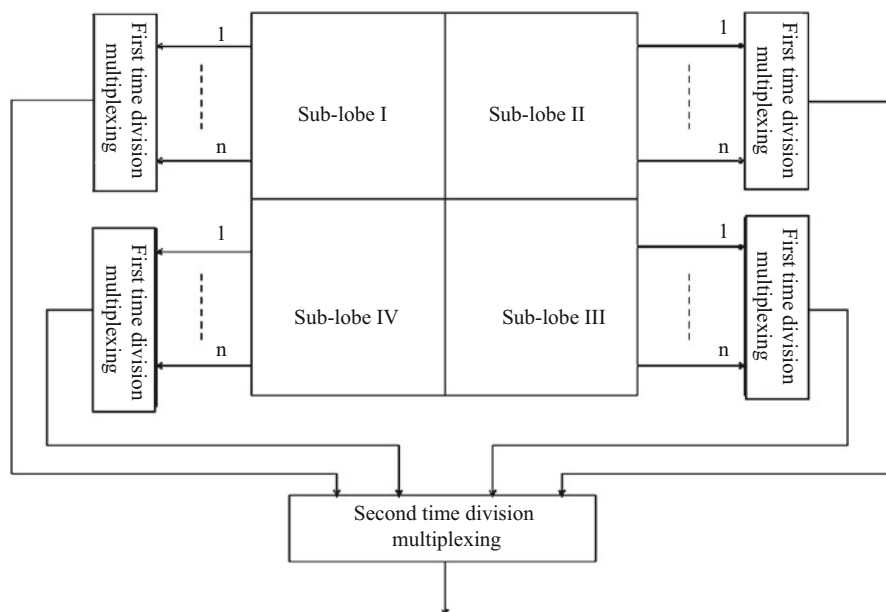


Fig. 2.99 Principle block diagram of “second time division multiplexing”

In Fig. 2.99, divide the $4n$ horn feed source arrays into four sub-arrays, and each sub-array includes n horns. Each sub-array uses “discrete automatic regulatory theory” for the first time division multiplexing to multiplex n -channel into 1-channel output. It corresponds to one sub-lobe. The output of four sub-arrays will go through a 4-channel switch to complete the second time division multiplexing. Then it will multiplex 4-channel into 1-channel output. The 4-channel switch will be switched on, which is equivalent to the sequential scan of four sub-lobes so as to complete wide angle coverage.

2.5.7 Polarization Diversity-Synthesized Technology of Angle Tracking

2.5.7.1 Function of Polarization Diversity-Synthesized Technology

In flight vehicle TT&C and information transmission system, polarization diversity-synthesized technology is generally applied, which functions are as follows:

- (1) Anti-multipath interference, to gain a stable tracking.
- (2) To gain two kinds of orthogonal polarization wave for diversity synthesis to obtain synthesis gain, so as to improve SNR and tracking accuracy.
- (3) The two orthogonal polarization receiving channels back up each other to improve the reliability of the system.

Reasons why there are two kinds of polarization waves in the radio wave received are:

(1) When the spacecraft enters at a low altitude and low elevation, the antenna elevation of the TT&C station is very low, and multipath interference will be caused due to refraction by the ground (or sea surface). Thus, signal fluctuation and fadeout will be caused, and there will be two kinds of orthogonal polarization components.

(2) When the antenna of the spacecraft is polarized by a line, it will be decomposed into two kinds of polarization components, i.e., the left-hand rotation and right-hand rotation, which can be received by the ground station through its dual-polarization antenna for polarization synthesis. 3 dB S/N gain will be obtained in theory.

(3) In order to obtain an omnidirectional antenna free of “dead spot” on a revolving flight vehicle, the space synthesis of the pattern can be changed into polarization synthesis of circuit. Adopt “Four array elements orthogonal polarization synthesis” scheme, i.e., use 4 antennas 90° apart from each other. The two symmetrical antennas 180° apart from each other are co-polarized and the other two antennas are orthogonal polarized. Thus, the polarization synthesis technology can be used to obtain an omnidirectional synthetic beam free of dead spot.

(4) If the thunder, rain, etc., make the atmospheric composition change and form plenty of small refractive bodies and reflective bodies during the process of radio wave propagation, the generated cross polarization (e.g., the influence of rain drops) will suffer irregular random fadeout.

(5) The antenna on the spacecraft and the shell of spacecraft will affect each other. When the antenna is close to the shell, it will change the polarization index (e.g. the axial ratio). In fact, the change of axial ratio is the change of proportion between the left-hand rotation and right-hand rotation circular polarized waves. Figure 2.100 shows the polarization change of the circular polarization satellite antenna.

In ideal conditions, there is only circular polarized wave, while in fact it will change into elliptic polarized wave and even linear polarized wave when the antenna deviates from the direction of the electrical boresight of the antenna. Therefore, polarization attenuation will be caused if a fixed polarization antenna is used for receiving.

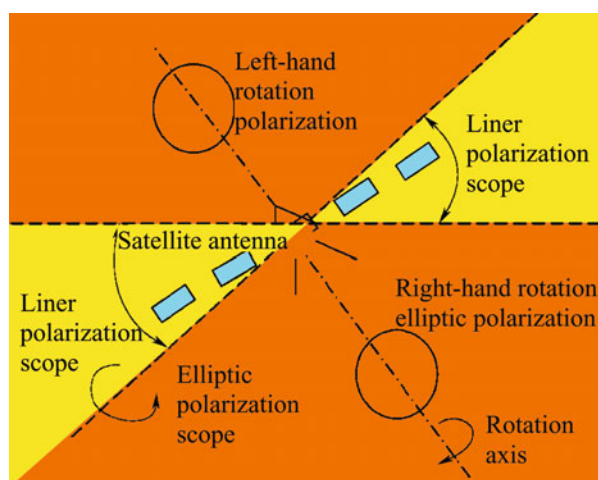


Fig. 2.100 Circular polarization satellite antenna

In addition, during the process of TT&C satellite positioning or when satellite control system fails, rolling phenomenon will occur. If both left-hand and right-hand rotation polarization antennas are installed on the satellite, the signal received by the ground station will be left-hand rotation signal for a moment and then the right-hand rotation signal for another moment.

(6) When the spacecraft is in high-speed movement or rotation, changes of rotating state will cause the rotation of spacecraft antenna polarization direction and zero point of the pattern, which will cause a long-term fadeout of the signal received by the antenna. Such fadeout includes cross polarization fadeout which is different from the fadeout of multipath effect and generally occurs in high SNR conditions. Cross polarization can not only weaken the signal but also have great effect on self-tracking performance.

(7) The flame jetted from the engine of the flight vehicle (during the working of the engine or stage separation) makes the atmosphere ionize to form a plasma area. Therefore, the signal passed through by it will have a rapid random change with respect to its amplitude and polarization.

Since there are two kinds of orthogonal polarized waves in the radio wave from the spacecraft, the ground station shall have the corresponding polarization diversity receiving capability to weaken the effect of signal fading so as to obtain a stable tracking and to improve S/N .

2.5.7.2 Mode and Principle of Signal Synthesis

In diversity synthesis receiving, signals obtained by N different independent signal branches can be used to obtain diversity gain according to different combination schemes. From the point of view of the synthesis position, the synthesis before a detector can be called “synthesis before detection,” i.e., the synthesis at IF and RF and almost the synthesis at IF. The synthesis can also be performed after a detector, called “synthesis after detection,” i.e., the synthesis at video.

(1) Maximum ratio synthesis

N branches will be formed after frequency diversity at the receiving terminal, and the branches will have in-phase addition and be sent to the detector for detection after phase adjustment and multiplied by a proper weighed coefficient. The principle diagram of maximum ratio synthesis is as shown in Fig. 2.101.

It can be proved based on Chebyshev's inequality when the variable gain weighting coefficient is $G_i = \frac{A_i}{\sigma^2}$, the SNR after diversity synthesis can be up to the maximum value, where A_i indicates the signal amplitude of the i th diversity branch and σ^2 is the noise power of that branch, $i = 1, 2, \dots, n$.

Output after synthesis is:

$$A = \sum_{i=1}^N G_i A_i = \sum_{i=1}^N \frac{A_i}{\sigma^2} \times A_i = \frac{1}{\sigma^2} \sum_{i=1}^N A_i^2 \quad (2.354)$$

It can be seen that a branch with a higher SNR has a higher contribution to the signal after synthesis. When the statistics of each branch is independent and all the

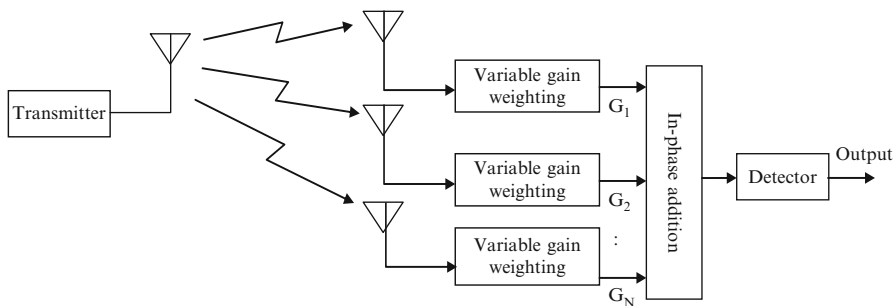


Fig. 2.101 Principle diagram of maximum ratio synthesis receiving of space diversity

branches have a same average SNR $\overline{SNR}$, the average output SNR after maximum ratio synthesis is:

$$\overline{SNR}_M = N \bullet \overline{SNR} \quad (2.355)$$

where $\overline{SNR}_M$ indicates the average output SNR after maximum ratio synthesis, and N indicates the number of diversity branches, i.e. diversity multiplicity.

Synthesis gain is:

$$K_M = \frac{\overline{SNR}_M}{\overline{SNR}} = N \quad (2.356)$$

Therefore, synthesis gain is directly proportional to the number of diversity branches N .

(2) Equal gain synthesis

In the above maximum ratio synthesis, making $A_i = 1$, $i = 1, 2, \dots, n$ will obtain equal gain synthesis, and the average output SNR after equal gain synthesis is:

$$\overline{SNR}_E = \overline{SNR} \left[1 + (N - 1) \frac{\pi}{4} \right] \quad (2.357)$$

Synthesis gain of equal gain synthesis is:

$$K_M = \frac{\overline{SNR}_E}{\overline{SNR}} = 1 + (N - 1) \frac{\pi}{4} \quad (2.358)$$

When N (diversity multiplicity) is large, equal gain synthesis is almost the same as maximum ratio synthesis, only about 1 dB difference. The realization of equal gain synthesis is easy and the required equipment is simple.

(3) Selective synthesis

Principle diagram of selective synthesis is shown in Fig. 2.102.

In Fig. 2.102, the receiving terminals are receivers of N diversity branches $i = 1, 2, \dots, N$. In R_i , use the selection logic to select a channel with the highest SNR $\overline{SNR}_{i=1}$ as the output. The average output SNR of selective synthesis is:

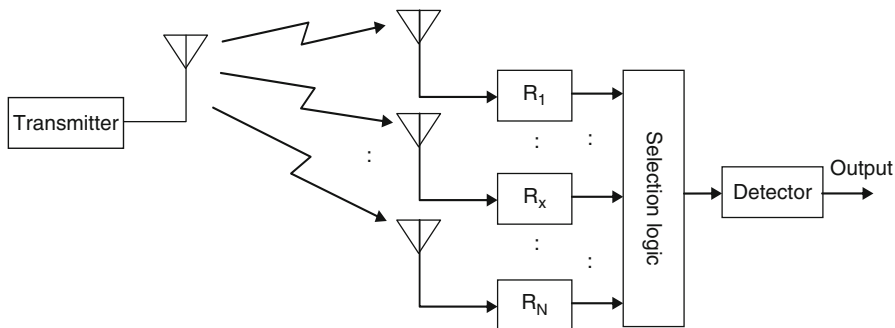


Fig. 2.102 Block diagram of selective synthesis receiver

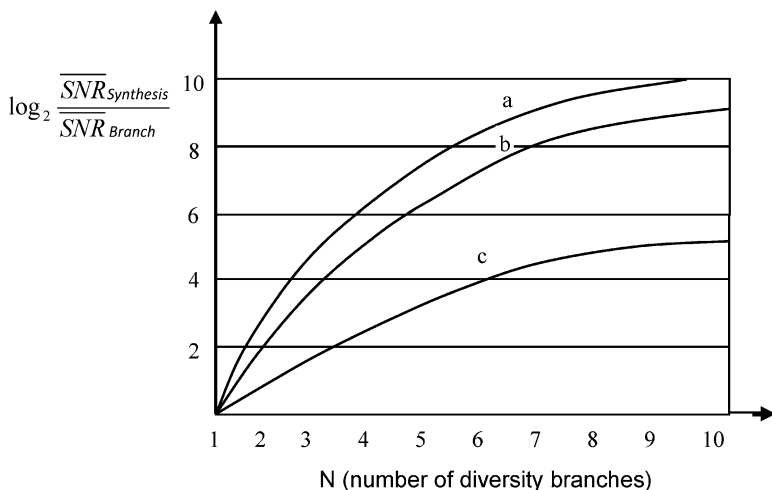


Fig. 2.103 Performance comparisons of three kinds of diversity synthesis. (a) Maximum ratio synthesis, (b) equal gain synthesis, and (c) selective synthesis

$$\overline{\text{SNR}}_S = \overline{\text{SNR}}_{i=1} \sum_{i=1}^N \frac{1}{i} \quad (2.359)$$

It can be seen that the contribution to selective diversity output SNR by adding one diversity branch is only the reciprocal of the total number of diversity branches. Synthesis gain of selective synthesis is:

$$K_s = \frac{\text{SNR}_S}{\text{SNR}_{i=1}} = \sum_{i=1}^N \frac{1}{i} \quad (2.360)$$

(4) Performance comparison of three main synthesis modes

What is shown in Fig. 2.103 is the improvement degree of the average SNR of the three synthesis modes.

When $\overline{\text{SNR}}$ of branches are different, the performance difference among them will be greater.

In addition, selective synthesis scheme is greatly influenced by the correlation of the synthesis signal fading. Only when their fading correlation coefficient is less than 0.3, the influence will be less. The equal gain synthesis scheme is greatly influenced by the fading distribution characteristics.

Therefore, the maximum ratio synthesis scheme is the best and it is frequently adopted in actual engineering.

2.5.7.3 Polarization Synthesis Scheme of Angle Tracking Channel

When there are two kinds of components in the radio wave received, it will cause the signal fluctuation and fading and have great effect on angle tracking performance, which makes the antenna unable to aim at the target and cause jitter. Therefore, polarization diversity receiving is generally adopted for angle measurement channel to improve tracking performance.

(1) IF polarization synthesis

For the monopulse system of three-channel and two-channel, diversity phase lock is unable to be done for there is no signal in difference channel (or S/Φ is very small) during tracking, thus it is difficult to realize it. One IF polarization synthesis scheme is to realize left-hand rotation and right-hand rotation diversity phase lock and maximum ratio synthesis in a sum channel with available signal, and use sum channel VCO as the reference signal of difference channel error detection. The phase of the sum channel and difference channel shall be consistent to ensure the same phase of the left-hand signal and right-hand signal in difference channel. The difference channel shall adopt the same weighted value as that of the sum channel; thus, the difference channel can also realize maximum ratio synthesis. AGC voltage can be used to realize diversity weighting. This scheme uses a diversity synthesis circuit whose sum and difference channels are completely the same to realize the polarization diversity synthesis of difference channel. Since the sum channel can automatically and self-adaptively realize maximum ratio synthesis, and the difference channel is consistent with the sum channel, the polarization diversity synthesis of difference channel can also be realized.

(2) Video polarization synthesis

Monopulse angle error detection is a coherent detection, and it won't cause output SNR loss in theory; therefore, synthesis technology after detection can be used. The angle error signal has a low frequency and is easy for digitization, so this scheme is called video digital polarization synthesis. Its principle diagram is shown in Fig. 2.104. The reference signal of PD in the Figure can be provided by the reference signal of PLL which is included in IF_Σ .

The advantages of this scheme are:

- 1) Do without diversity phase locked loop so as to simplify the circuit.
- 2) Digitalized maximum ratio synthesis, simple scheme, good performance, close to theoretical results.

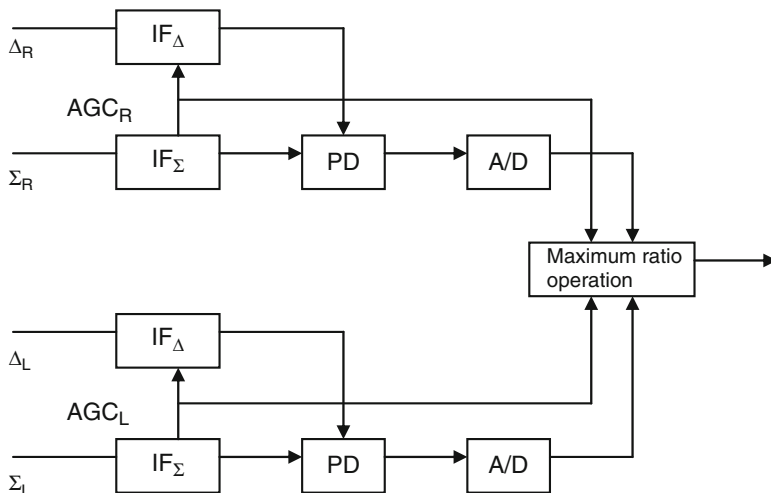


Fig. 2.104 Principle diagram of video digital polarization synthesis scheme of angle tracking channel

- 3) Fully inherit the previous monopulse channel technology and only add the part of maximum ratio synthesis.
- 4) The debugging is simple, stable and reliable.

(3) Polarization combination scheme of circular waveguide TE_{11} , TE_{21} module monopulse—"diversity-combination method"

In C&T system, circular waveguide TE_{11} , TE_{21} mode feed source is generally used to compose a DCM monopulse antenna [25]. It has a high efficiency but also some special problems in polarization diversity synthesis.

The principle of conical multimode corrugated horn is to use the pattern distribution and radiation characteristic of fundamental mode and high order mode in circular waveguide, to form the sum pattern by fundamental mode and the difference pattern by high order mode. When the antenna beam axis aims at a target, the incoming wave signal only excites the fundamental mode (TE_{11} mode) in the conical multimode corrugated horn. When the target deviates from the antenna beam axis, the incoming wave signal can excite the fundamental mode (TE_{11} mode) and the high order mode (TE_{21} mode) in the conical multimode corrugated horn at the same time. Based on the design of smooth-walled circular waveguide mode-selective coupler (tracker), only TE_{11} mode is allowed to pass its direct-through arm which is the sum T/R channel. The side arm can couple two orthogonal TE_{21} modes through eight rectangular waveguide coupling aperture arrays and form the output port for circular polarized different-mode signal through synthesis network. The amplitude of high order TE_{21} mode is directly proportional to the angle of deviation, and forms circular polarized different signal output through synthesis network.

This feed source is different from multi-horn, square horn multimode feed source. The relative phase relationship among its sum/difference signals is shown in Fig. 2.105.

Fig. 2.105 Phase relationship among the sum/difference signals formed by circular waveguide TE11 and TE21 mode feed sources

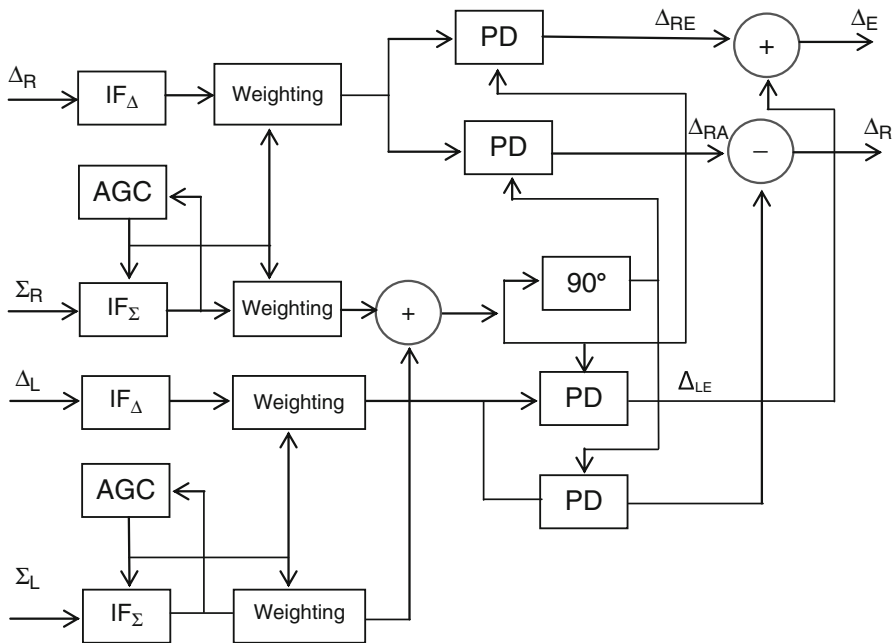
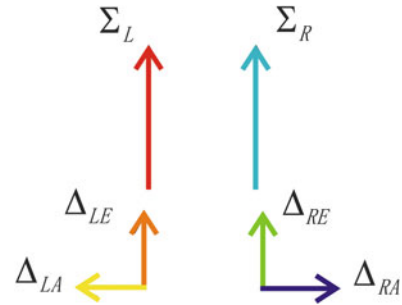


Fig. 2.106 Block diagram of polarization synthesis of circular waveguide TE11 and TE21 mode monopulse

In Fig. 2.105, Σ_L , Δ_{LE} , Δ_{LA} is respectively the sum channel, elevation difference, azimuth difference signal of left-hand rotation signal. The phase difference between Δ_{LE} and Δ_{LA} is 90° , so as to realize DCM channel combination of left-hand polarization in DCM. In a similar way, Δ_{RE} , Δ_{RA} will realize the DCM combination of right-hand rotation polarization channel. It can be seen from that figure that, if the polarization combination is conducted at IF, only Δ_{LE} and Δ_{RE} can have in-phase addition, while Δ_{LA} and Δ_{RA} will have reverse phase subtraction. Therefore, polarization synthesis cannot be realized. In order to solve this problem, it is required to split Δ_{LE} and Δ_{LA} and make the phase of Δ_{LA} component reversed; thus Δ_{LA} and Δ_{RA} can have in-phase addition now. This idea of “synthesis after split” can be realized by video polarization diversity synthesis, as shown in Fig. 2.106.

In Fig. 2.106, difference channel signal is synthesized by video polarization. The subtraction between azimuth right-hand rotation video polarization component and azimuth left-hand rotation video polarization component is used to realize their reverse phase addition. IF polarization synthesis is adopted for sum channel. Because AGC voltage is directly proportional to the voltage of sum channel signal, AGC_R voltage of right-hand sum channel is used for the amplitude weighting of right-hand sum channel signal and its difference channel signal, and the AGC_L voltage of left-hand sum channel is used for the amplitude weighting of left-hand sum channel signal and its difference channel signal, to realize the maximum ratio synthesis.

References

1. Sun Zhongkang et al (1996) Active and passive localization technology with single or multiple base. National Defense Industry Press, Beijing
2. Wu Ping (1996) Optimal station arrangement of typical TT&C system. J Astronaut 17(1):15–23
3. Liu Jiaying (2007) Frequency domain analysis of velocity measure error. Telecommun Eng 2007 Suppl
4. Huang Xuede et al (2000) Missile TT&C system. National Defense Industry Press, Beijing
5. Liu Jiaying (2010) Characterization of short term stability and its transform between $S_1(f)$ [J]. J Spacecr TT&C Technol 6:29–33
6. Bames JA, Chi AR (1971) Characterization of frequency stability. IEEE Trans M-20 No. 1971:10–13
7. Qian Yunmin (1978) Relationship between sampling variance and CW radar velocity measurement accuracy. Telecommun Eng 1978(6)
8. Liu Jiaying (2005) Effect of short term frequency stability of oscillators on the velocity in a TT&C system. Telecommun Eng 2005(1):1–7
9. Gardner FM (2007) Phase-locked loop technology. Post & Telecom Press, Beijing
10. Huang Chengfang (2012) Full coherent frequency folding multistation transponder. Telecommun Eng 2012(3)
11. Liu Jiaying (1980) Tone ranging time delay in linear four terminal network and product demodulator. J CETC 10, 1980.2
12. Liu Jiaying (2009) Group delay characteristics analysis method for range error and its applications in deep space TT&C systems. J Spacecr TT&C Technol 2009.2:22–26
13. Liu Jiaying, Zhou Wenlan (2013) Effect of group delay on distance-measuring error. Telecommun Eng 2013(5)
14. JPL (2007) DSMS telecommunications link design handbook. <http://deepspace.jpl.nasa.gov/dsndocs/810-005/>
15. Cai Tingsong (2013) Short-baseline positioning technology based on phase difference. 10th Institute of China Electronics Technology Group Corporation, Chengdu
16. Chen Fangyun (1992) Satellite TT&C manual. Science Press, Beijing
17. Wang Yuanling (2008) Analysis of deep space ranging signal systems. Telecommun Eng 48(5):1–6
18. Wang Xiaoyan (2005) Study on single channel monopulse angle tracking systems. Telecommun Eng 45(3):117–120
19. Liu Jiaying (2001) A tentative scheme of interferometer MMW rendezvous radar. Telecommun Eng 41(1):13–18

20. Zhao Yefu, Liu Jiaying et al (2001) Radio tracking system. National Defense Industry Press, Beijing
21. Yu Zhijian (2009) Deep space TT&C system. National Defense Industry Press, Beijing
22. Zhang Guangyi (2006) Phased array radar technology. Publishing House of Electronics Industry, Beijing
23. Yi Runtang (1998) Study on the adaptive processing of TDRSS phased array ground multi-beam formation. Telecommun Eng 1998(4)
24. Skolnik M (2012) Introduction to radar systems (trans: Zuo Qunsheng et al.). Publishing House of Electronics Industry, Beijing
25. Ke Shuren (2012) Discussion on polarization diversity receiving of circular waveguide TE₁₁ and TE₂₁ mode true and false monopulse system, C&T, 2012(2)

Spacecraft TT&C and Information Transmission Theory
and Technologies

Liu, J.

2015, XI, 531 p. 265 illus., 53 illus. in color., Hardcover

ISBN: 978-3-662-43864-0