D. Manolakis, et. al.. "Position, Location, Altitude Measurement."
Copyright 2000 CRC Press LLC. <http://www.engnetbase.com>.

Position, Location, Altitude Measurement

Dimitris E. Manolakis
Technological Education Institute

Mark Stedham
University of Alabama in Huntsville

Partha P. Banerjee
University of Alabama in Huntsville

Seiji Nishifuji
Yamaguchi "University

Shogo Tanaka
Yamaguchi University

Halit Eren
Curtin University of Technology

C.C. Fung
Curtin University of Technology

Jacob Fraden
Advanced Monitors Corporation

10.1 Altitude Measurement

Dimitris E. Manolakis

Accurate monitoring of aircraft cruising height is required in order to reduce vertical separation to a minimum standard. Interest here focuses on the measurement of the distance between aircraft level and the sea surface level. This distance can be estimated onboard via barometric altimeters or it can be measured — either onboard or in ground stations — via electronic radio wave systems. The indication of the first equipment is referred to as pressure altitude, or simply altitude, whereas that of the second category is referred to as geometric height or simply height.

The altitude information at air traffic control (ATC) centers is based on pressure altitude measurement that the aircraft transponder system sends after it receives an appropriate interrogation — known as mode C interrogation — transmitted by a secondary surveillance radar. Actually, the altitude information is an atmospheric pressure measurement transformed to altitude indication through a formula expressing the pressure/altitude relationship. When a flight level is cleared for an aircraft, it actually means that the pilot must keep flying on an isobaric surface. However, the altimetry system may present systematic errors (biases) that are different for each airplane, and that significantly affect safety. Thus, the altimetry

© 1999 by CRC Press LLC

system performance as well as the aircraft height keeping performance must be monitored by an independent radar or satellite system.

Radar or satellite systems determine the position of an object through algorithms that are fed with range, or range difference, or range sum, or range and bearing measurements, and they estimate the object position vector employing appropriate techniques such as triangulation or trilateration. The primary radar measurements are contaminated by two kinds of errors: random and systematic errors. The effect of random errors can be reduced by the use of appropriate noise rejection filters, such as a Kalman filter. The second kind of error is usually removed by calibrating the instrument if this is possible; otherwise, suitable algorithms must be invented to anticipate it.

The estimation may be derived in ground stations or onboard the aircraft according to where the data acquisition and processing is performed. In the latter case, the vertical position estimation has to be downlinked to the appropriate ATC center. Also, the estimation may be performed off-line or on-line. Ground-based methods or systems are the Navigation Accuracy Measurement System (NAMS), the height estimation method with a single air traffic control radar, the method with one Secondary Surveillance Radar (SSR) and one omnidirectional radar, the Dual synchronized Autonomous Height Monitoring System (DAMS) and methods that use multiple SSRs and estimate the height by quadrilateration, or by the use of pseudorange measurements, or by the use of range difference measurements. On-board height measurement methods derive their estimates by trilateration using the Distance Measuring Equipment (DME) or the Global Positioning System (GPS) signals.

Ground-Based Height Estimation

The radars used to derive the original measurements are either primary or secondary surveillance radars. A primary radar sends a signal and scans for the arrival of its reflection. The range to the object reflecting the signal is derived from the time elapsed between transmission and reception. With secondary radar, the radar sends an interrogation to aircraft — to all aircraft or to a selected one — and the appropriate aircraft sends a reply via its transponder. The range to the aircraft is computed from the time elapsed between the signal transmission and the signal arrival, taking into account the nominal delay time of the transponder. Most of the methods estimating the aircraft height make use of the SSR equipment because it is cost effective, the transponder reply signal is stronger than that reflected to a primary radar, and the system can operate more reliably in dense traffic areas.

Any systematic errors in the primary radar and in the ground equipment of the SSR can be corrected by calibration. However, the problem encountered with SSR is that it involves the transponder delay time in the range measurement process. Thus, any systematic error in the transponder delay time causes range bias errors that are different for each aircraft and thus suitable methods must be used to anticipate for it in the subsequent measurement data processing.

Navigation Accuracy Measurement System

Nagaoka has proposed an off-line height estimation system [1–3]. It is composed of a primary marine radar located under an airlane and measures range R and depression angle β. Figure 10.1 shows the geometry of the system. The antenna rotates about a vertical axis and scans the area above it with rate equal to 1 rotation per 3 s. The principle exploited to derive the height estimate is that the range varies as the aircraft passes through the data acquisition area. The rate of change of range is mainly a function of the flight height z and secondarily of depression angle. It is easily derived from Figure 10.1, that the relation between the above quantities and the position x along the x axis at time t is:

$$R(t) = \sqrt{x^2(t) + R'(t)^2} = \sqrt{x^2(t) + \frac{z^2}{\cos^2 \beta(t)}} \tag{10.1}$$

Let x_0 denote the position of the aircraft at time t_0. Assuming the aircraft flies in straight and level flight, this means that the velocity V_x, the depression angle, and the height h remain constant during the data

© 1999 by CRC Press LLC

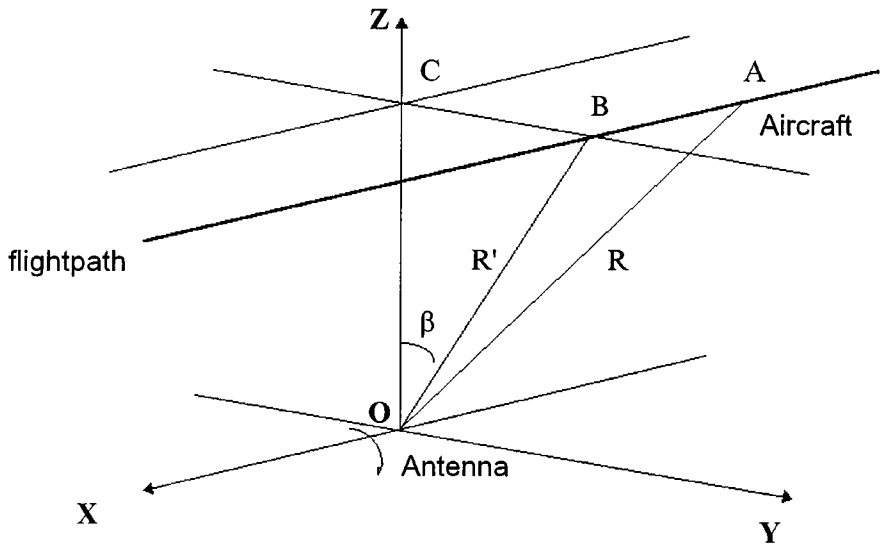

FIGURE 10.1 Geometry of the Navigation Accuracy Measurement System (NAMS). BA = *x*, CB = *y*, OC = *z*, R′ = *z*/cos(β).

collection period. The above quantities and the range measurement at time t_i are related through Equation 10.2.

$$R_i = \sqrt{\left[x_0 + V_x\left(t_i - t_0\right)\right]^2 + \frac{z^2}{\cos^2\beta}} = f\left(t_i, \beta, \mathbf{q}\right) = f_i\left(\mathbf{q}\right) \tag{10.2}$$

where $\mathbf{q}$ is the unknown quantities vector, $\mathbf{q} = [x_0, V_x, z]^T$.

Measurements of R_i and β are collected at times $t_0, t_1, \ldots, t_n$, and their set is briefly expressed as a vector function of the unknown quantities with the following matrix equation:

$$\mathbf{R} = \mathbf{f}\left(\mathbf{q}\right) \tag{10.3}$$

where

$$\mathbf{R} = \begin{bmatrix} R_0 \\ R_1 \\ \cdot \\ R_n \end{bmatrix} \qquad \mathbf{f} = \begin{bmatrix} f_0 \\ f_1 \\ \cdot \\ f_n \end{bmatrix}$$

Equation 10.3 is nonlinear. Thus, a nonlinear least square method, such as the Gauss-Newton iterative method, must be used to estimate the unknown vector. Let $\mathbf{q}_k$ be the estimate at the *k*th iteration. Then, the next estimate is:

$$\hat{\mathbf{q}}_{k+1} = \hat{\mathbf{q}}_k + \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\left(\mathbf{R} - \mathbf{f}\left(\hat{\mathbf{q}}_k\right)\right) \tag{10.4}$$

where **F** is the partial derivatives (Jacobian) matrix; that is:

© 1999 by CRC Press LLC

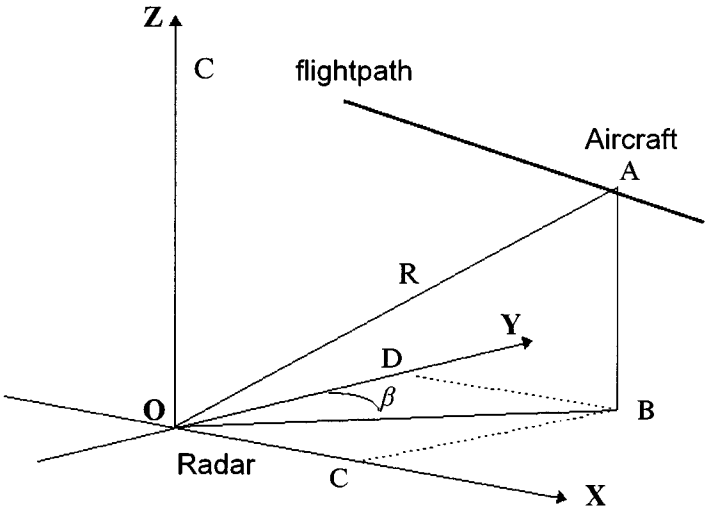

FIGURE 10.2 Geometric configuration of the height measurement system with a single air-traffic control radar. OC = x, OD = y, AB = z, $\beta = \angle$ BOD.

$$\mathbf{F} = \frac{\vartheta \mathbf{f}}{\vartheta \mathbf{q}} = \begin{bmatrix} \frac{\vartheta f_0}{\vartheta x_0} & \frac{\vartheta f_0}{\vartheta V_x} & \frac{\vartheta f_0}{\vartheta z} \\ . & . & . \\ . & . & . \\ \frac{\vartheta f_n}{\vartheta x_0} & \frac{\vartheta f_n}{\vartheta V_x} & \frac{\vartheta f_n}{\vartheta z} \end{bmatrix} \tag{10.5}$$

Assuming typical aircraft velocities and flight levels, and taking into account the antenna fan beam and rotation rate, it is determined that the observation time is approximately 1 min and the number of measurements is approximately 20. The standard deviation (SD) of the height estimation errors is σ_z = 10 m when the SD of the measurement errors of range and depression angle are σ_R = 10 m and σ_B = 0.08°, respectively.

The system has been developed and operates at the Electronic Navigation Institute of Tokyo, Japan. To estimate and anticipate for systematic errors in the depression angle and range measurements, it has been proposed to perform 40 experimental flights above the radar range. The basic assumption of the NAMS system (i.e., that the aircraft flies at constant height) cannot be validated by the NAMS itself. Thus, an SSR is required to confirm that the flight is indeed performed at the same level via the mode C interrogations [3].

Height Estimation with a Single Air-Traffic Control Radar

Nagaoka at the Electronic Navigation Institute of Tokyo investigated the possibility to derive a height estimation with a very simple and inexpensive method [4]. The idea is to use a single radar normally used for air traffic control. The radar antenna rotates around a horizontal axis once every 4 s and measures range R and bearing β, as it is shown in Figure 10.2.

Let (x,y,z) be the 3-D position of the aircraft. The bearing angle is defined as:

$$\beta = \cos^{-1}\left(\frac{y}{\sqrt{x^2 + y^2}}\right) \tag{10.6}$$

© 1999 by CRC Press LLC

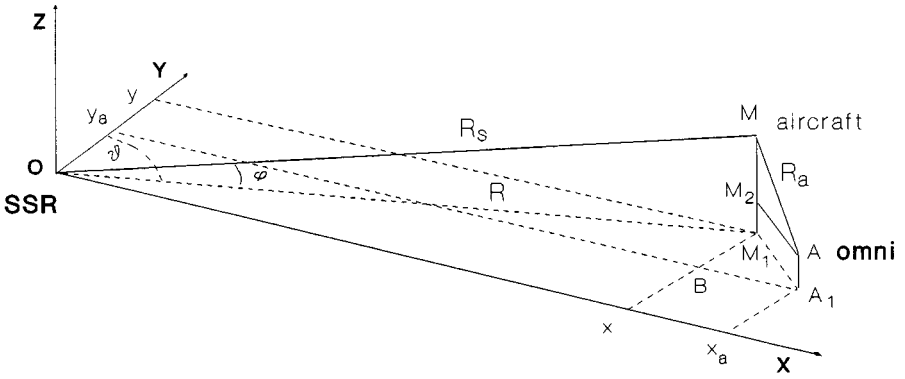

FIGURE 10.3 Geometry of the height estimation system with one SSR and one omnidirectional radar (SSROR). The distance x_a of the radars typically ranges from 80 km up to 160 km. $M_1M = z$.

Thus, the relation between the measurements and the unknown quantities is:

$$R = \left(z^2 + x^2 + y^2\right)^{1/2} = \left(z^2 + \frac{y^2}{\cos^2 \beta}\right)^{1/2} \tag{10.7}$$

The measurement principle is the same as that of the NAMS system, i.e., that during the time the aircraft passes through the data acquisition area, the range varies as a function of both the horizontal position and the height. When straight and level flight is assumed, the unknown quantities are the height z, the y-axis initial position y_0, and the velocity V_y. The expression relating the measurements R_i, and β_i and the unknown quantities vector $\mathbf{q} = [y_0, V_y, z]^T$ at time t_i is:

$$R_i = \left(z^2 + \frac{\left(y_o + V_y\left(t_i - t_o\right)\right)^2}{\cos^2 \beta_i} \right)^{1/2} = f\left(t_i, \beta_i, \mathbf{q}\right) = f_i\left(\mathbf{q}\right) \tag{10.8}$$

The measurements are collected as the aircraft flies close to the radar at a distance of 30 km. Approximately, the data acquisition area is 15 km, the observation time is 1 min and the total number of scans is 15. The total set of measurements is expressed as a nonlinear matrix function in the form of Equation 10.3; that is, $\mathbf{R} = \mathbf{f}(\mathbf{q})$ where $\mathbf{R}$ denotes the measurement vector. The unknown vector is estimated using a nonlinear least squares method like that of the NAMS system expressed by Equation 10.4.

The system performance is evaluated with Monte Carlo simulations. The best accuracy achieved is $\sigma_z = 100$ m under the assumption that the SD errors of range and bearing measurements are $\sigma_R = 100$ m and $\sigma_B = 0.06°$. The method can use either primary or secondary radar. However, in the latter case, the method has to include systematic delays by the transponder.

Height Estimation with one SSR and One Omni Radar (SSROR)

Manolakis et al. investigated the possibility to estimate the aircraft flight height with a system that consists of one standard SSR plus a secondary radar equipped with an omnidirectional antenna located far away from the SSR; hence, this system could be referred to as SSROR [5]. Figure 10.3 shows a typical geometric configuration of the system. The SSR measures bearing θ and slant range R_s, which is transformed to horizontal range R with correcting look-up tables. The omni radar measures range R_a. Let (x,y,z), (x_a, y_a, z_a) and $(0,0,0)$ be the coordinates of aircraft, omni radar, and SSR, respectively. Then, from Figure 10.3 it follows that the theoretical height function is:

© 1999 by CRC Press LLC

$$z = f\left(R_a, x, y\right) = z_a + \left[R_a^2 - \left(x - x_a\right)^2 - \left(y - y_a\right)^2\right]^{1/2} \tag{10.9}$$

$$x = R\sin\theta \tag{10.10}$$

$$y = R\cos\theta \tag{10.11}$$

However, the height measurement obtained through the above formula will be biased because of the transponder delay systematic deviation from its nominal value that affects the range measurements. Since this bias cannot be removed by calibration, a suitable algorithm has to be applied to anticipate for it. The approach proposed is to augment the unknown vector by incorporating the bias term. Let b denote the bias in range measurements. The biased measurements R_b and R_{ab} of R and R_a, respectively, are expressed as:

$$R_b = R + b\cos\varphi \tag{10.12}$$

$$R_{ab} = R_a + b \tag{10.13}$$

where φ is the elevation angle. Let s denote the squared height:

$$s = \left(z - z_a\right)^2 \tag{10.14}$$

and s_b denote the corresponding quantity derived from (biased) range measurements. Then, after some manipulations of the above equations, the following relation is obtained:

$$s_b = s + b\,a \tag{10.15}$$

$$a = 2\left[R_a - \cos\varphi\left(R - x_a\sin\theta - y_a\cos\theta\right)\right] \tag{10.16}$$

The term a is the bias multiplying factor determined by the relative geometry of the system. Equation 10.15 is a linear relation between measurement s_b and the unknown quantities s and b. Figure 10.4 shows that the effect of the bias varies as the aircraft passes through the surveillance area. Its form is mainly determined by the flight height. Consequently, it is possible to estimate both the height and the bias by collecting data during the period the aircraft remains in the surveillance area. The measurement equation at time t_i is:

$$s_{bi} = s_i + b\,a_i + e_i \tag{10.17}$$

where e_i represents the effect of the random measurement errors; hence, it could be referred to as the equation error. Assuming level flight, $s_i = s$, and the set of collected data is expressed with the following linear matrix equation:

$$\mathbf{s_b} = \mathbf{A}\,\mathbf{q} + \mathbf{e} \tag{10.18}$$

$$\mathbf{s_b} = \begin{bmatrix} s_{b0} \\ s_{b1} \\ \cdot \\ s_{bn} \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 1 & a_0 \\ 1 & a_1 \\ \cdot & \cdot \\ 1 & a_n \end{bmatrix} \qquad \mathbf{e} = \begin{bmatrix} e_0 \\ e_1 \\ \cdot \\ e_n \end{bmatrix} \qquad \mathbf{q} = \begin{bmatrix} s \\ b \end{bmatrix}$$

© 1999 by CRC Press LLC

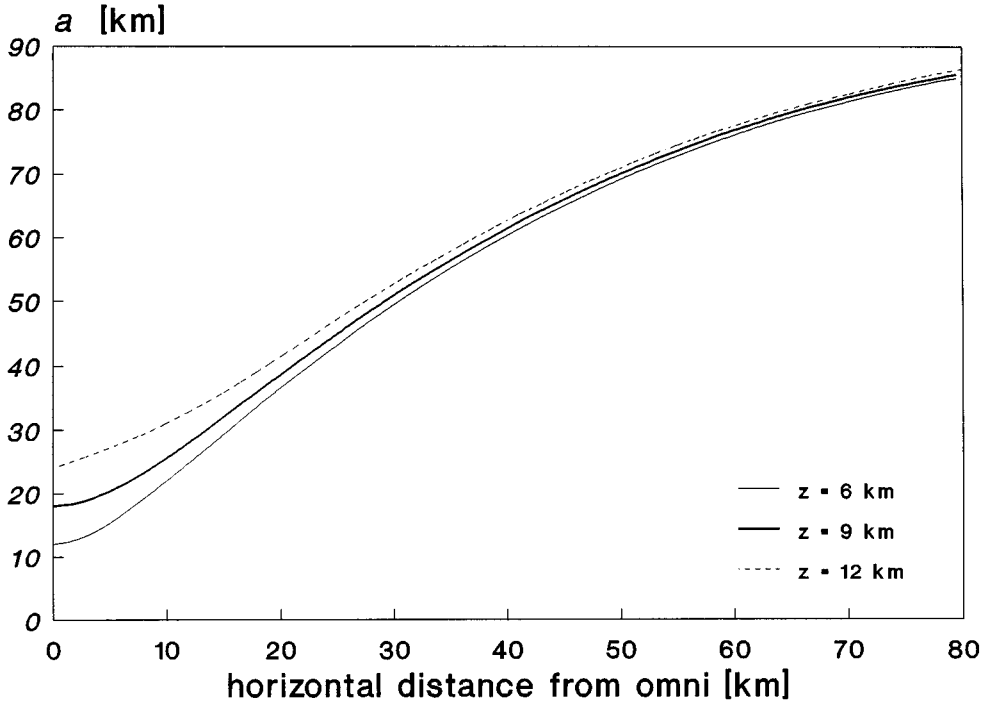

FIGURE 10.4 Range bias multiplier in squared height measurements as a function of the aircraft distance from the omni radar. Three different flight heights are examined. The omni radar is at position (150 km, 0, 0).

The best estimate $\hat{\mathbf{q}}$ of $\mathbf{q}$, minimizing the weighted sum of the squared errors $\mathbf{e}^T\,\mathbf{W}\,\mathbf{e}$, is:

$$\hat{\mathbf{q}} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{s}_b \tag{10.19}$$

where $\mathbf{W}$ is the weighting matrix defined as the inverse of the equation error covariance matrix:

$$\mathbf{W}^{-1} = \mathrm{E}\left\{\mathbf{e}\,\mathbf{e}^T\right\} \tag{10.20}$$

where the notation E{ } stands for the expected value operation. Thus, the estimate of the geometric height is:

$$\hat{z} = \sqrt{\hat{q}_1} = \sqrt{\hat{s}} \tag{10.21}$$

In the case of nonlevel flights, an augmented unknown vector is used that comprises bias b, initial height z_0, and vertical velocity V_z [6]. In this case, however, the measurement equation is nonlinear in terms of the unknown quantities; hence, a nonlinear least squares iterative algorithm must be employed.

The system performance is a function of: (1) the range and bearing measurement standard deviation errors, σ_R and σ_θ, respectively; (2) the aircraft velocity, which affects the number of scans; (3) the omni radar scan rate; and (4) geometric terms such as the flight level and the distance between the radars. By assuming that σ_R = 70 m, σ_θ = 0.08°, the SD of the height estimation error will be between 50 m and 100 m when the values of the other factors lie in reasonable ranges.

© 1999 by CRC Press LLC

Dual Synchronized Autonomous Monitoring System (DAMS)

The research has been conducted in the National Aerospace Laboratory NLR, the Netherlands [7]. The system is composed of two primary marine radars. Each radar has a rotating antenna. The antennas scan, with different directions, the same volume — that is an area 3 km × 10 km above the location site. The primary surveillance area is 3 km × 3 km. The two antennas (A and B) are separated by 2.5 m, are mechanically synchronized, and scan once every 2.5 s. Their data extractors produce measurements of slant ranges and elevation angles, (R_A, φ_A) and (R_B, φ_B), respectively, for each radar. The tracking software derives estimation of aircraft position and trajectory in off-line mode. At each scan, a combination of four measurements $(R_A, \varphi_A, R_B, \varphi_B)$ is available for use in the estimation of the unknown vector $\mathbf{q} = [x_0, y_0, z_0, V_x, V_y, V_z]^T$. A local Cartesian frame is used to perform the calculations. Finally, 16 equations are available to be solved for the 6 unknowns with a weighted least squares method. The weight of each measurement is the measured amplitude of the radar pulse. The maths of the tracker are not presented in [7]. However, one approach could be the following. If a coordinate system is defined such that radar A is at the origin (0, 0, 0) and radar B is at $(x_B, 0, 0)$, then the measurements obtained at scan time t_i can be expressed as:

$$R_{Ai} = \sqrt{x_i^2 + y_i^2 + z_i^2} = \sqrt{\left(x_0 + V_x\left(t_i - t_0\right)\right)^2 + \left(y_0 + V_y\left(t_i - t_0\right)\right)^2 + \left(z_0 + V_z\left(t_i - t_0\right)\right)^2} \quad (10.22a)$$

$$\varphi_{Ai} = \tan^{-1}\left(\frac{z_i}{\sqrt{x_i^2 + y_i^2}}\right) = \tan^{-1}\left(\frac{z_0 + V_z\left(t - t_i\right)}{\sqrt{\left(x_0 + V_x\left(t - t_i\right)\right)^2 + \left(y_0 + V_y\left(t - t_i\right)\right)^2}}\right) \quad (10.22b)$$

$$\begin{aligned} R_{B_i} &= \sqrt{\left(x_i - x_B\right)^2 + y_i^2 + z_i^2} \\ &= \sqrt{\left(x_0 + V_x\left(t_i - t_0\right) - x_B\right)^2 + \left(y_0 + V_y\left(t_i - t_0\right)\right)^2 + \left(z_0 + V_z\left(t_i - t_0\right)\right)^2} \end{aligned} \quad (10.22c)$$

$$\varphi_{Bi} = \tan^{-1}\left(\frac{z_i}{\sqrt{\left(x_i - x_B\right)^2 + y_i^2}}\right) = \tan^{-1}\left(\frac{z_0 + V_z\left(t - t_i\right)}{\sqrt{\left(x_0 + V_x\left(t - t_i\right) - x_B\right)^2 + \left(y_0 + V_y\left(t - t_i\right)\right)^2}}\right) \quad (10.22d)$$

Notice that x_B has a small value (2.5 m) compared to the magnitude of the other quantities; consequently, it can be neglected. The above set of equations at four different times yields 16 equations that can be solved with a nonlinear least squares method such as the Gauss–Newton iterative method presented in Equation 10.4.

The SD of the height estimation error will be less than 15 m in the primary surveillance area, and 30 m at the edges of the area when the SD of range and elevation angle measurements are smaller than 10 m and 0.1°, respectively.

Height Measurement by Quadrilateration

Rice proposed a system consisting of four synchronized receiving SSR stations S_i, i = 0, 1, 2, 3 that use SSR transmissions from the aircraft transponder and estimate the height by quadrilateration [8]. Figure 10.5 shows a typical configuration of systems composed of N SSRs. One of them is an active station, which means that this station has both an interrogator and a receiver. Let (x, y, z) and (x_i, y_i, z_i) denote the Cartesian coordinates of aircraft and station S_i, respectively. Also let R_i denote the range from station S_i to aircraft, c denote the velocity of the light, and T_s denote the time of signal transmission from aircraft transponder. The stations measure the time of arrival (TOA) T_i, i = 0, 1, 2, 3, of the aircraft transponder signal at each site. The following relations hold:

© 1999 by CRC Press LLC

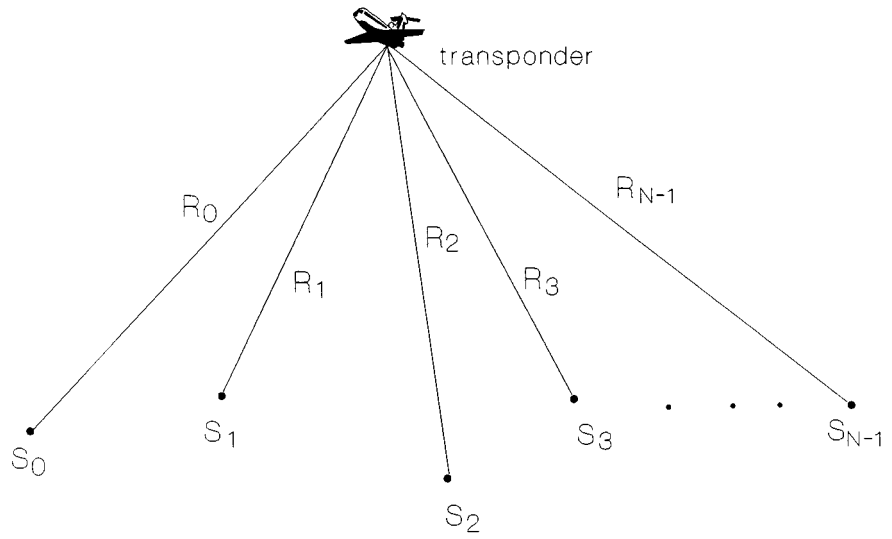

FIGURE 10.5 Typical configuration of the height estimation systems that are based on *N* SSR stations. One of the stations is active, i.e., it both transmits the interrogations and receives the replies, whereas the other stations are receivers only.

$$T_i = T_s + \frac{R_i}{c} \quad i = 0, 1, 2, 3 \tag{10.23}$$

$$R_i^2 = \left(x - x_i\right)^2 + \left(y - y_i\right)^2 + \left(z - z_i\right)^2, \quad i = 0, 1, 2, 3 \tag{10.24}$$

The above system of eight equations can be solved for the unknown quantities. The unknown quantities used by Rice are $(R_0, R_1, R_2, R_3, x, y, z, T_s)^T$. However, an equivalent approach is to substitute for R_i in Equation 10.23, which becomes:

$$T_i = T_s + \frac{1}{c}\sqrt{\left(x - x_i\right)^2 + \left(y - y_i\right)^2 + \left(z - z_i\right)^2} = f_i\left(\mathbf{q}\right) \quad i = 0, 1, 2, 3 \tag{10.25}$$

where $\mathbf{q} = [x, y, z, T_s]^T$ is the unknown vector. Thus, there are four nonlinear equations to be solved for **q**. One suitable method, for example, is the Newton–Raphson method, which iteratively approximates the solution via the following formula:

$$\mathbf{q}_{k+1} = \mathbf{q}_k + \mathbf{F}\left(\mathbf{q}_k\right)^{-1}\left(\mathbf{T} - \mathbf{f}\left(\mathbf{q}_k\right)\right) \tag{10.26}$$

where $\mathbf{T} = [T_0, T_1, T_2, T_3]^T$ is the measurement vector and **F** is the Jacobian matrix:

$$\mathbf{F} = \frac{\partial \mathbf{f}}{\partial \mathbf{q}} = \begin{bmatrix} \frac{\partial f_0}{\partial x} & \frac{\partial f_0}{\partial y} & \frac{\partial f_0}{\partial z} & \frac{\partial f_0}{\partial T_s} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} & \frac{\partial f_3}{\partial T_s} \end{bmatrix} \tag{10.27}$$

Notice that the time of interrogation transmission, as well as the transponder nominal delay time, are not involved in the measurements. The measured quantities are only the TOAs at the station sites. Thus, the height estimate is not affected by any transponder bias.

© 1999 by CRC Press LLC

The theoretical and experimental research has been conducted at the GEC Marconi Research Center, U.K. The optimum station arrangement is to locate the three of them equispaced on a circle and the fourth in the middle. The typical circle radius is 35 km. The same magnitude holds for the measurement range. The method may be implemented in on-line or off-line mode. In the first case, there must be transmitters at the stations to transmit their measurements of TOA to the height monitoring center. The Vertical Dilution Of Precision (VDOP) is a performance index defined by the ratio:

$$VDOP = \frac{\sigma_z}{c\sigma_{rte}} \tag{10.28}$$

where σ_{rte} is the SD of the relative timing errors. The VDOP expresses the effect of the relative geometry to system performance. The VDOP of this system achieves a typical value of 3.

Height Estimation with SSRs and Pseudoranges

This approach has been investigated by Nagaoka at Electronic Navigation Institute of Tokyo [9]. The system consists of N SSR receiving stations; see Figure 10.5. One of them, say S_0, must be active to transmit interrogations to aircraft. The time of interrogation transmission, T_t, and the times of signal arrival at the receiving stations T_i, i = 0, 1, …, N – 1, are measured. Thus, N pseudorange measurements r_i are obtained where $r_i = c\,(T_i - T_t)$. Let T_D denote the transponder delay and D denote the distance corresponding to this delay, $D = c\,T_D$. Then, for each pseudorange measurement r_i, the following relation holds:

$$\begin{aligned} r_i &= D + R_i + R_0 = D + \sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} + \sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2} \\ &= f_i(\mathbf{q}) \qquad i = 0, 1, \ldots, N-1 \end{aligned} \tag{10.29}$$

where $\mathbf{q} = [x, y, z, D]^T$ is the unknown vector. The set of N measurements $\rho = [\rho_0, \rho_1, \ldots \rho_{N-1}]^T$, $N \geq 4$, and the unknown vector are related through Equation 10.30.

$$\mathbf{r} = \mathbf{f}(\mathbf{q}) \tag{10.30}$$

The unknown vector $\mathbf{q}$ can be obtained from the solution of Equation 10.30 with a nonlinear weighted least squares method. Thus, the best estimate of $\mathbf{q}$ is iteratively calculated as:

$$\hat{\mathbf{q}}_{k+1} = \hat{\mathbf{q}}_k + \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\left(\mathbf{r} - \mathbf{f}\left(\hat{\mathbf{q}}_k\right)\right) \tag{10.31}$$

where $\mathbf{F}$ is the Jacobian matrix

$$\mathbf{F} = \frac{\partial \mathbf{f}}{\partial \mathbf{q}} = \begin{bmatrix} \frac{\partial f_0}{\partial x} & \frac{\partial f_0}{\partial y} & \frac{\partial f_0}{\partial z} & \frac{\partial f_0}{\partial D} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_{N-1}}{\partial x} & \frac{\partial f_{N-1}}{\partial y} & \frac{\partial f_{N-1}}{\partial z} & \frac{\partial f_{N-1}}{\partial D} \end{bmatrix} \tag{10.32}$$

The estimate of $\mathbf{q}$ is free of the transponder delay systematic error because the estimation is based not on the nominal delay, but on the actual delay time, which is one of the parameters to be estimated, whereas the rest of the parameters are the aircraft 3-D position coordinates x, y, z.

© 1999 by CRC Press LLC

The station arrangement proposed by Nagaoka, when there are four stations, is an equilateral triangle formed by the three stations, whereas the fourth station is located in the center. The VDOP, defined as the ratio σ_z/σ_R (where σ_R is the observation error) has a typical value of 4 when the aircraft is above the center at a height equal to the baseline radius. The VDOP increases as the aircraft flies higher and longer and as the baseline radius becomes smaller.

Height Measurement with SSRs and Range Differences

This approach has been proposed by Manolakis and Lefas [10, 11]. The system consists of $N-1$ receiving SSR stations S_i, i = 1, $N-1$, and one station, say S_0, which is both receiver and interrogator, see Figure 10.5. The stations receive the reply and the time difference of arrival (TDOA) between a reference station, say S_0, and station S_i is measured. A set of $N-1$ TDOA or equivalently range difference (RD) measurements is collected at each time the transponder sends a reply signal. The height estimation derived from this set of measurements is not affected by any transponder delay systematic error since this error is inherently subtracted from the measurements used. This system could be referred to as RD height monitoring unit (RDHMU). The systems that derive the position fix based on this kind of measurement are known as TDOA or RD or hyperbolic systems.

Let τ_i denote the TDOA between stations S_i and S_0, and d_i denote the corresponding RD measurement, $d_i = c\,\tau_i$. The following relation holds:

$$\begin{aligned} d_i &= R_i - R_0 = \sqrt{\left(x-x_i\right)^2+\left(y-y_i\right)^2+\left(z-z_i\right)^2} - \sqrt{\left(x-x_0\right)^2+\left(y-y_0\right)^2+\left(z-z_0\right)^2} \\ &= f_i\left(\mathbf{q}\right) \qquad i = 1, 2, \ldots, N-1 \end{aligned} \tag{10.33}$$

where $\mathbf{q} = [x, y, z]^T$ is the unknown aircraft position vector. The vector of RD measurements $\mathbf{d} = [d_1, d_2, \ldots, d_{N-1}]$ is expressed as:

$$\mathbf{d} = \mathbf{f}\left(\mathbf{q}\right) \tag{10.34}$$

A commonly employed method to solve for $\mathbf{q}$ in this nonlinear equation is the Taylor series method or equivalently the Gauss-Newton iterative method. The best estimate of $\mathbf{q}$ is iteratively approximated as:

$$\hat{\mathbf{q}}_{k+1} = \hat{\mathbf{q}}_k + \left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\left(\mathbf{d} - \mathbf{f}\left(\hat{\mathbf{q}}_k\right)\right) \tag{10.35}$$

where $\mathbf{F}$ is the Jacobian matrix:

$$\mathbf{F} = \frac{\vartheta \mathbf{f}}{\vartheta \mathbf{q}} = \begin{bmatrix} \dfrac{\vartheta f_1}{\vartheta x} & \dfrac{\vartheta f_1}{\vartheta y} & \dfrac{\vartheta f_1}{\vartheta z} \\ \dfrac{\vartheta f_{N-1}}{\vartheta x} & \dfrac{\vartheta f_{N-1}}{\vartheta y} & \dfrac{\vartheta f_{N-1}}{\vartheta z} \end{bmatrix} \tag{10.36}$$

In the case of four stations the best arrangement is an equilateral triangle with the fourth station in the center. The SD of height estimation error σ_z will be 15 m when the baseline radius is 6 km, the flying height is 9 km, and σ_{TDOA} is 10 ns.

Work on proof of principles and system development of a HMU based on the concept of TDOA measurement of SSR signals has been conducted by Roke Manor Research Ltd., U.K. [12].

© 1999 by CRC Press LLC

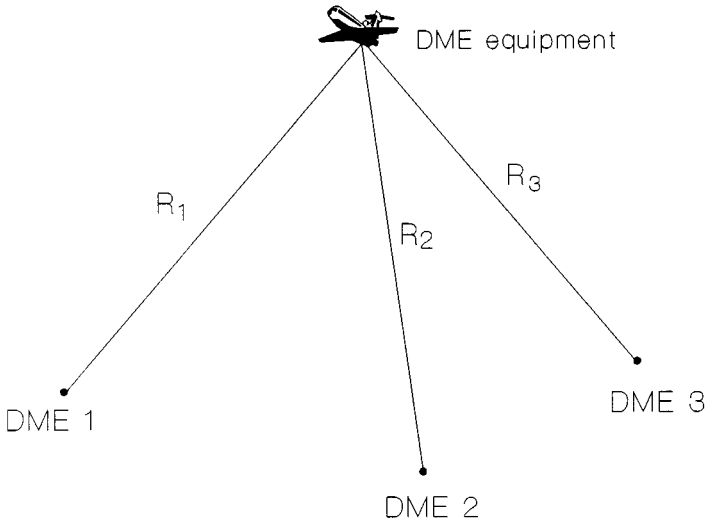

FIGURE 10.6 Configuration of the on-board height estimation system that utilizes the distance measurements derived from the DME equipment.

Onboard Derived Height Estimation

Height Measurement with Distance Measuring Equipment (DME)

This approach for deriving the geometric height onboard the aircraft using DME equipment was first reported by Rekkas et al. [13], whereas more efficient and general techniques have been proposed by Manolakis [14, 15]. Using the DME interrogation equipment, the distance from three DME ground stations is measured onboard (see Figure 10.6). The three stations are located under an airway. The height is then computed from the range measurement vector $\mathbf{R} = [R_1, R_2, R_3]^T$ by trilateration. An exact and efficient solution of the nonlinear measurement equation was derived in [15]. Specifically, the height is computed from the closed form:

$$z = g(\mathbf{R}) = \frac{-b(\mathbf{R}) + \sqrt{d(\mathbf{R})}}{2a} \tag{10.37}$$

where $b(\mathbf{R})$ and $d(\mathbf{R})$ are the following simple polynomial-type functions

$$b(\mathbf{R}) = b_0 + b_1 R_1^2 + b_2 R_2^2 + b_3 R_3^2 \tag{10.38}$$

$$d(\mathbf{R}) = d_{00} + d_{01} R_1^2 + d_{02} R_2^2 + d_{03} R_3^2 + d_{11} R_1^4 + d_{22} R_2^4 + d_{33} R_3^4 + d_{12} R_1^2 R_2^2 + d_{13} R_1^2 R_3^2 + d_{23} R_2^2 R_3^2 \tag{10.39}$$

The coefficients a, b_i, d_{ij} are analytically defined in the Appendix of [15]. An important aspect of these coefficients is that they are completely defined by the ground stations' coordinates (x_i, y_i, z_i), which are fixed. Thus, the coefficients are calculated only once at the moment the aircraft enters the data acquisition area. Then, every time a new set of range measurements is available, the height is computed from the above equations using the range measurements and the stored coefficients. Define the ratio σ_z/σ_R as the VDOP of this technique, where σ_R is the SD of the ranging error. The VDOP is 1 in the case where the stations form an equilateral triangle inscribed in a circle with 10 km radius and the aircraft is above the triangle center at a height of 8 km.

© 1999 by CRC Press LLC

Estimation of Vertical Position with the Global Positioning System (GPS)

The research and development of the GPS has been coordinated by the U.S. Department of Defense. Another similar system is the Global Navigation Satellite System (GLONASS) developed by the former Soviet Union. The GPS is a satellite system providing users with accurate timing and ranging information. The system is available with reduced accuracy to civilian users. Many companies, mainly from the U.S., produce GPS receivers. Let (x, y, z) and (x_i, y_i, z_i) be the coordinates of the user and satellite s_i. The GPS receiver of the user derives the pseudorange measurement D_i, and the corresponding measurement equation is:

$$D_i = R_i + cT_b = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} + b = f_i(\mathbf{q}) \tag{10.40}$$

where T_b is the user clock bias, and $\mathbf{q} = [x, y, z, b]^T$ is the unknown vector that incorporates the bias term b. Thus, in order to estimate the 3-D position of the aircraft, four pseudorange measurements are required at least; consequently, four satellites must be visible from the receiver. The set of N pseudorange measurements $\mathbf{D} = [D_1, D_2, \ldots, D_N]^T$ defines the following matrix measurement equation:

$$\mathbf{D} = \mathbf{f}(\mathbf{q}) \tag{10.41}$$

which is solved for $\mathbf{q}$ with the Gauss–Newton least squares iterative method, that is:

$$\hat{\mathbf{q}}_{k+1} = \hat{\mathbf{q}}_k + \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\left(\mathbf{D} - \mathbf{f}(\hat{\mathbf{q}}_k)\right) \tag{10.42}$$

where $\mathbf{A}$ is the partial derivatives matrix:

$$\mathbf{A} = \frac{\partial \mathbf{f}}{\partial \mathbf{q}} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} & \frac{\partial f_1}{\partial b} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_N}{\partial x} & \frac{\partial f_N}{\partial y} & \frac{\partial f_N}{\partial z} & \frac{\partial f_N}{\partial b} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{x1} & \mathbf{a}_{y1} & \mathbf{a}_{z1} & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \mathbf{a}_{xN} & \mathbf{a}_{yN} & \mathbf{a}_{zN} & 1 \end{bmatrix} \tag{10.43}$$

The elements a_{xi}, a_{yi}, a_{zi}, of the partial derivatives matrix $\mathbf{A}$ are the direction cosines from the receiver to the satellite s_i. The weighting matrix is the inverse of the covariance matrix of the pseudorange measurement errors, $\mathbf{W}^{-1} = \mathbf{E}(\delta\mathbf{D}\,\delta\mathbf{D}^T)$. The weighting is generally used to take into account the possible different performances of each satellite, although usually the same performance is assumed for all satellites; that is, $\mathbf{W} = \mathbf{I}$. The VDOP, defined as σ_z/σ_D, depends on the geometry which varies continuously, even in the case of a fixed receiver, because the satellites are not geostationary but move in such orbits as to complete a rotation in 12 h. The world mean value of VDOP is about 2 [16]. Typical VDOP values range from 1.5 to 7, depending on the area of the receiver and on the time of day. The ranging error for the precision positioning service (available only to U.S. military users) has been specified to be less than 6 m (SD). For the standard positioning service, normally available to civilian users, the specified ranging error is double (12 m, SD), whereas it will be about 40 m when selective availability is activated by the Department of Defense. The corresponding measured ranging errors found to be smaller than the specified ones. Namely, for the three operating conditions mentioned, the corresponding values for ranging errors were found to be 2.3 m, 6 m, and 20 m, respectively [17]. The multiplication of the ranging SD error by the VDOP yields the standard deviation of the height estimation error.

© 1999 by CRC Press LLC

To anticipate for the error intentionally induced by the Department of Defense, the Differential GPS (DGPS) method has been developed. A station located at a precisely known position receives the satellite signals, computes its own position on the basis of pseudorange measurements, compares this position with the known position in order to estimate the included error in satellite signals, and finally transmits the appropriate corrections to the receivers in its neighborhood. The achieved accuracy is in the order of a few meters.

Special Topics

The performance analysis of the various parameter estimation systems is usually restricted to the variance analysis, and the estimation error is usually assumed to have zero mean value. However, it is proven that in all of the above systems, the estimation error does not have zero mean value due to the nonlinearity of the measurement equation. Another important aspect is that there are cases where it is not possible to obtain a solution due to the relative geometry of aircraft vs. stations that leads to large errors. In these cases, the successive iterations applied to solve the system of nonlinear equations may not converge. Also, even after convergence, the solutions need not necessarily be "the best" or the "correct ones."

Inherent Bias

The nonlinearity of the systems, joined with the measurement random errors, causes inherently biased estimations although the measured quantities are unbiased. For example, take the case of height estimation with DME measurements. The function $g(\mathbf{R})$ in Equation 10.37, which determines that the height z is nonlinear. In addition, the range measurements will be contaminated by additive zero mean value random errors. Let $\mathbf{R_m}$ denote the noisy measurement vector and z_m denote the height measurement derived from $g(\mathbf{R_m})$, i.e., from the measurement function when it is fed with noisy measurements. Extending $g(\mathbf{R_m})$ in a Taylor series around the actual values of ranges R_1, R_2, R_3 up to second-order terms and taking the expected values, it is derived that the expected value of z_m will not be the actual value z, but it will differ by an amount b_z, which is called inherent bias. Specifically, for the DME case, the inherent bias is evaluated as:

$$b_z = E\left[z_m\right] - z = \frac{1}{2}\left(\frac{\partial^2 g}{\partial R_1^2}\sigma_{R_1}^2 + \frac{\partial^2 g}{\partial R_2^2}\sigma_{R_2}^2 + \frac{\partial^2 g}{\partial R_3^2}\sigma_{R_3}^2\right) \tag{10.44}$$

Figure 10.7 shows the inherent bias generated in the DME system. The inherent bias becomes larger as the magnitude of the measurement errors and the system nonlinearity becomes stronger. This bias error is inherently generated in all position estimation algorithms and must be taken into account when precise position estimation is required. Biased height estimates have also been reported in [18] for the SSROR system, in [10] for the RDHMU, and in [3] for the NAMS system.

Existence and Uniqueness of Position Fix

In some systems, there are singular cases for which it is not possible to achieve a position fix. This fact has been reported by Abel and Chafee for the GPS system in [19], where it is shown that for some satellites/aircraft relative geometries, it is not possible to solve the relevant equation or there is more than one solution. Also, for the RDHMU, it has been shown by Manolakis and Lefas that there are some station arrangements for which it is not possible to derive height estimation when the aircraft is at specific areas [11]. For example, in the case of four stations, when the quadrilateral defined by the stations is inscribed in a circle, it is not possible to estimate the height when the aircraft is above the center. Also, when the quadrilateral is a rectangle or symmetric trapezoid, it is not possible to derive a position fix when the aircraft is above the line that passes from the middle of the parallel sides. From a mathematical point of view, this singularity is expressed by the singularity of the Jacobian matrix; consequently, this matrix cannot be inverted as is required by the relevant position estimation algorithm. The algorithm in this case diverges from the actual height and finally collapses. Notice that height estimation is achieved everywhere except at this singular point. However, when the aircraft is close to the singular region,

© 1999 by CRC Press LLC

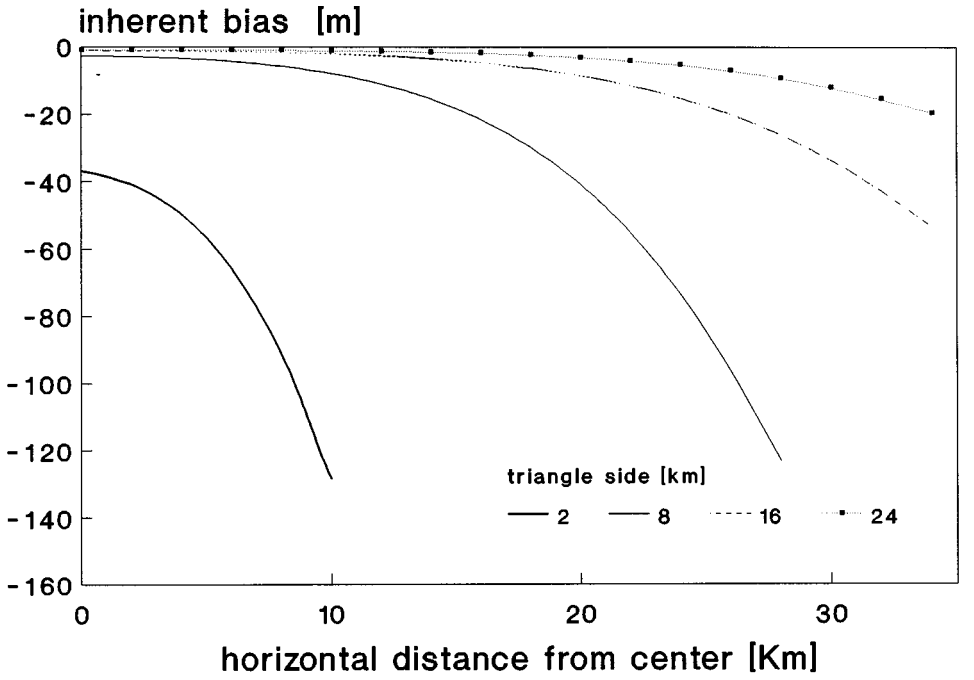

FIGURE 10.7 Bias inherently generated by the height estimation algorithm of the system based on DME measurements. The stations' sites form an equilateral triangle. The inherent bias is shown as function of the horizontal distance from the center for various magnitudes of the triangle side. The flight height is $z = 9$ km, and the SD of the distance measurement error is $\sigma_R = 90$ m.

although a position fix is achieved, it is not actually reliable since it is affected by large errors; for example, the VDOP could be larger than 600 in regions close to the singular region.

Table 10.1 presents the institutes and companies that either investigate and develop prototype height monitoring units or provide relevant systems in the market.

TABLE 10.1 Height Monitoring Systems and Companies/Institutes that Develop and Provide Them

System	Company/Institute
Navigation Accuracy Measurement System (NAMS)	Electronic Navigation Research Institute Ministry of Transport 6-38-1 Shinkawa, Mitaka Tokyo, 181, Japan Tel: +81 422 413171 Fax: 81-422-413176
DAMS height monitoring unit	National Aerospace Laboratory NLR Anthony Fokkerweg 2 1059 CM Amsterdam The Netherlands Tel: +31 (0)20 511 3113 Fax: +31 (0)20 511 3210
SSR and quadrilateration technique	GEC-Marconi Electronics Ltd. Marconi Research Laboratories West Hanningfield Road Great Baddow, Chelmsford Essex, England

© 1999 by CRC Press LLC

TABLE 10.1 (continued) Height Monitoring Systems and Companies/Institutes that Develop and Provide Them

System	Company/Institute	
SSR and TDOA technique (RDHMU)	Roke Manor Research Ltd. Roke Manor, Romsey Hampshire SO51 0ZN U.K. Tel: +44(0)794 833000 Fax: +44(0)794 833433	
GPS providers	Trimble Navigation 585 North Mary Avenue Sunnyvale, CA 94086 Tel: (408) 730-2900	Trimble Navigation Europe Ltd. 79-81 High Str. West Malling Kent ME19 6NA U.K. Tel: +44(0)732 849242 Fax: +44(0)732 847437
	Rockwell International Digital Communication Division 4311 Jamboree Road Newport Beach, CA 92660-3095 Tel: (714) 221-4600 Fax: (714) 221-6375	Rockwell Semiconductor Systems Berkshire Court, Western Road Bracknell, Berkshire RG12 1RE England Tel: +44(0)1344 48644 Fax: +44(0)1344 48655

References

1. S. Nagaoka, E. Yoshioka, and P. T. Muto, Radar estimation of the height of a cruising aircraft, *J. Navigation*, 32(3), 352-356, 1979.
2. S. Nagaoka, E. Yoshioka, and P. T. Muto, A simple radar for navigation accuracy measurements, *J. Navigation*, 34(3), 462-469, 1981.
3. S. Nagaoka, Possibility of detecting a non-level-flight aircraft by the navigation accuracy measurement system (NAMS), *ICAO, Review of the General Concept of Separation Panel, 7th Meeting,* Montreal, RGCSP-WP/180, 30/10/90.
4. S. Nagaoka, Height estimation of a cruising aircraft via a radar for air traffic control, *Electronics and Communications in Japan, Part 1 (Communications), USA*, 71(11), 95-105, 1988.
5. D. E. Manolakis, C. C. Lefas, G. S. Stavrakakis, and C. M. Rekkas, Computation of aircraft geometric height under radar surveillance, *IEEE Trans. Aerosp. & Electr. Systems*, AES-28(1), 241-248, 1992.
6. D. E. Manolakis, Computation of aircraft geometric height under radar surveillance for non level flights, *Int. J. Systems Sci.*, 25(4), 619-627, 1994.
7. J. Brugman, J. Verpoorte, and A. J. L. Willekens, DAMS Height Monitoring Unit-Phase One, Report CR 92328 C, NLR, The Netherlands, 1992.
8. D. E. Rice, Height measurement by quadrilateration, *The Marconi Review*, XLVI, (228), 1-17, 1983.
9. S. Nagaoka, Possibility of geometric height measurement by using secondary surveillance radars, *ICAO, Review of the General Concept of Separation Panel, 7th Meeting,* Montreal, RGCSP-WP/181, 30/10/90.
10. D. E. Manolakis and C. C. Lefas, Aircraft geometric height computation using secondary surveillance radar range differences, *IEEE Proc.-Radar, Sonar and Navigation*, 141(2), 119-124, 1994.
11. D. E. Manolakis and C. C. Lefas, Station arrangement effects on ground referenced height computation by using time differences, *Navigation, J. Inst. Navigation*, 42(2), 409-420, 1995.
12. L. G. Hopkins, D. Sherry, and D. C. Rickard, Geometric Height Monitor Unit (HMU) Programme — Final Report on Phase 1, Proof of Principles, Roke Manor Research Ltd., Report No. 72/91/R161|U, Roke Manor, U.K.,1991.

© 1999 by CRC Press LLC

13. C. M. Rekkas, C. C. Lefas, and N. J. Krikelis, Improving the accuracy of aircraft absolute aircraft altitude estimation using DME measurements, *Int. J. Systems Sci.*, 21(7), 1381-1392, 1990.
14. D. E. Manolakis, Efficient solution and performance analysis of 3-D position estimation by trilateration, *IEEE Trans. Aerosp. & Electr. Systems*, AES-32(4), 1239-1248, 1996.
15. D. E. Manolakis and A. I. Dounis, Advances in aircraft height computation using distance measuring equipment, *IEEE Proc.-Radar, Sonar and Navigation*, 143(1), 47-52, 1996.
16. J. L. Leva, Relationship between navigation vertical error, VDOP, and pseudorange error in GPS, *IEEE Trans. Aerosp. & Electr. Systems*, AES-30(4), 1138-1142, 1994.
17. B. W. Parkinson, History and operation of NAVSTAR, the Global Positioning System, *IEEE Trans. Aerosp. & Electr. Systems*, AES-30(4), 1145-1161, 1994.
18. D. E. Manolakis, C. C. Lefas, and A. I. Dounis, Inherent bias in height computation employing mixed type radar data, *IEEE Trans. Aerosp. & Electr. Systems*, AES-30(4), 1045-1049, 1994.
19. J. S. Abel and J. W. Chaffee, Existence and uniqueness of GPS solutions, *IEEE Trans. Aerosp. & Electr. Systems*, AES-27(6), 952-956, 1991.

10.2 Attitude Measurement

Mark A. Stedham, Partha P. Banerjee, Seiji Nishfuji, and Shogo Tanaka

In many practical situations, it is important to determine and measure the attitude of a particular vehicle, such as a ship, an airplane, a piece of mechanical equipment such as a crane lifter, or a spacecraft. For this reason, many attitude sensors have been developed with advanced computer and semiconductor technologies. This section first introduces the various attitude sensors with an explanation of their operating principles and then presents several methodologies for attitude measurement and determination, including ships and crane lifters, aircraft, and spacecraft applications.

Attitude Sensors for Ships, Aircraft, and Crane Lifters

There are many types of gyroscopes that, corresponding to the physical measurement mechanisms used, may be classified as two-axes *freedom gyro* and single-axis freedom gyro using precession, *vibratory gyro* using Coriolis' force, and *optic gyro* using Sagnac's effect. Among them, the two-axes freedom gyro has the longest history. It consists of a high-speed rotating rotor around a spin axis supported by two orthogonal axes. This type of gyro is generally classified as either a *free gyro*, a *vertical gyro* (VG), or a *directional gyro* (DG).

The single-axis freedom gyro has only one output axis in addition to the spin axis. Depending on the specifications (in which) the gyro is designed, there are two types of gyros, the *rate gyro and* the *rate integrating gyro*. Related to these rotating-type gyros is another type of gyro known as the *electrostatic gyro*, which makes use of a high-speed rotating sphere in a vacuum cavity. Because of its resistance-free property, the electrostatic gyro has the highest accuracy among existing gyros. There are also rotorless gyros. The first one is a vibratory gyro that uses Coriolis' force as the measurement principle. The second type is an optical one. Among optical gyros, there are two types: the *ring laser gyro* and the *fiber optic gyro*. Both rely on the Sagnac effect in their measurement mechanisms. The performance of gyros is evaluated by their drift rates, and the performance of various gyros is shown in Table 10.2, for reference, with their primary usages.

Recently, with the development of computer technology, many types of three-axes gyros have been developed that can measure not only the tilt angles but also the angular velocities and the accelerations along the three axes by combining several gyros and accelerometers. Accelerometers are often coupled with gyros to provide flight and ship navigation systems as well as attitude sensors for dynamic objects such as crane lifters. Examples include the *attitude and heading reference system* (AHRS), *inertial navigation system* (INS), *inertial measurement unit* (IMU), and *gyro compass* (GC), as well as the VGs and DGs discussed above [1].

© 1999 by CRC Press LLC

TABLE 10.2 Performance of Different Types of Gyros

Type of the gyro	Degrees of freedom	Quantities to be detected	Accuracy (° h^{-1})
Free gyro	2	Angle	1
Vertical gyro	2	Declination from horizontal plane	1
Directional gyro	2	Shift from reference direction	1
Rate gyro	1	Angular velocity	10
Rate integrating gyro	1	Angle	0.001–1
Ring laser gyro	1	Angular velocity	0.003
Fiber optic gyro	1	Angular velocity	0.01
Electrostatic gyro	2	Angle	0.00001–0.01

The principle of a servo-type accelerometer is explained below (see Figure 10.8). As soon as the shift of the beam caused by the acceleration α is detected by the deflection pickup, the current i is generated by the servo-amplifier, which produces a torque to keep the beam at the principle axis of the sensor. Since the torque and the current that generates the torque are proportional to α, the acceleration can be measured using the current. The measurement process forms a closed-loop system, so that the sensor is not only robust to disturbances, but also achieves a high measurement accuracy (see Table 10.3).

Similarly, an inclinometer is another inertial sensor that measures tilt angle to provide attitude information (see Figure 10.9). The principle of servo-type inclinometers is the same as that of the servo-type accelerometer, except that the beam in the accelerometer is replaced by a pendulum suspended from the supporting point in the sensor. When the sensor is placed on the inclined static surface of tilt angle β, the pendulum takes the angle β against the principle axis of the sensor, assuming the sensor has no force other than gravity acting on it. The sensor can, however, generate a torque $T_c = mgl\sin\beta \cong mgl\beta$ to keep the pendulum at the principle axis, then the tilt angle β can be accurately measured using the torque (and consequently the current producing the torque), where m and l are the pendulum mass and length of the pendulum to its mass center, respectively. One must note, however, that such a sensor is essentially designed to measure the tilt angles of static inclined surfaces. Thus, when applied to dynamic inclined surfaces, the accelerations will affect the torque, making the sensor unreliable. An intelligent attitude

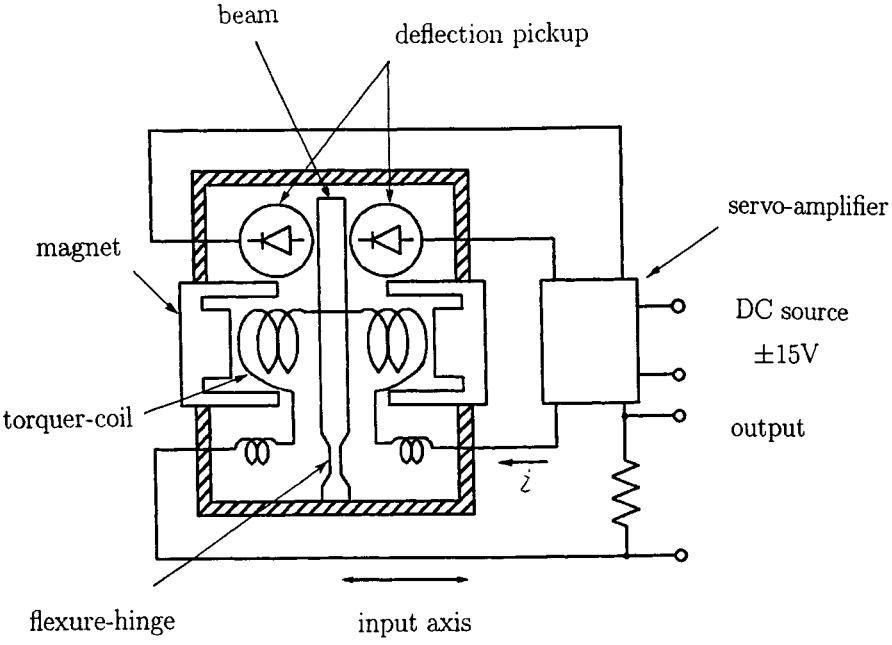

FIGURE 10.8 Servo-type accelerometer.

© 1999 by CRC Press LLC

TABLE 10.3 Specification of a Servo-Type Accelerometer

Measurement range	±5 g
Resolution	Less than 5 μ*g* (dc)
Sensitivity	2 V g^{-1}
Output resistance	560 Ω
Torquer current	3.5 mA g^{-1}
Case alignment	Less than ±1°
Frequency response	450 Hz (±3 dB)
Temperature range	–25 to +70°C
Power source	±15 V (dc)
Consumption current	Less than 15 mA
Size	28.4 mm × 24.5 mm
Mass	46 *g* (including the cable 10 *g*)

Note: g: gravitational acceleration (according to the type TA-25D-05 by TOKIMEC).

sensing system that overcomes such difficulty will be introduced later. Although application is limited to static inclined surfaces with minute tilt angles, a dielectric-type inclinometer employing electrodes and a bubble kept in an electrolyte can achieve high accuracies on the order of 10^{-4}°.

Attitude Sensors for Spacecraft Applications

Attitude measurement for spacecraft usually requires two or more sensors for detecting the reference sources needed to satisfy attitude requirements. The choice of which sensors to employ is primarily influenced by the direction the spacecraft is usually pointing as well as the accuracy requirements for attitude determination [2]. Table 10.4 summarizes some performance parameters for these sensors as well as typical manufacturers.

Inertial measurement units generally consist of gyroscopes coupled with accelerometers, which together measure both rotational and translational motion. These IMUs may be either gimbal mounted (movement about a gimbal point, independent of the spacecraft) or a strapdown system (rigidly mounted to the spacecraft body), where expansive software is used to convert sensor outputs into reference frame measurements. IMUs tend to suffer gyro drift and other bias errors and, when used for spacecraft attitude measurements, are often used with one or more of the sensors discussed below.

Sun sensors detect the visible light from the sun, measuring the angle between the sun's radiation and the detector's photocell. The sun is a commonly chosen attitude reference source since it is by far the

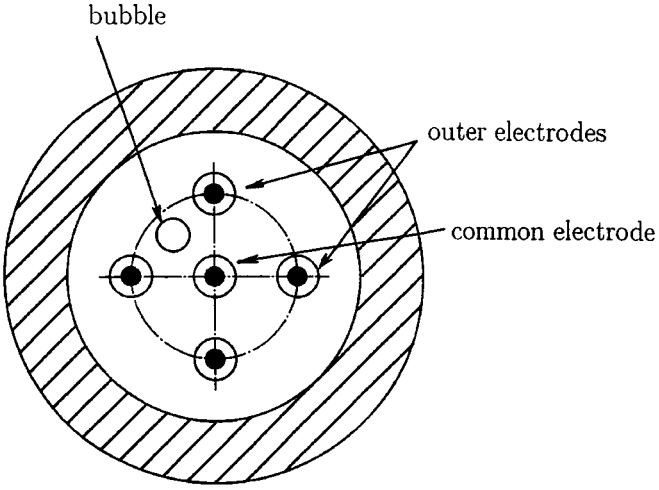

FIGURE 10.9 Dielectric-type inclinometer (front view).

© 1999 by CRC Press LLC

TABLE 10.4 Spacecraft Attitude Determination Sensors

Sensor	Accuracy	Mass (kg)	Typical vendors
IMU	1 to 5×10^{-6} *g*	3 to 25	Northrop Grumman, Bendix, Kearfott,Honeywell, Hamilton, Standard, Litton, Teledyne
Sun Sensor	10^{-2} to 3°	0.5 to 2	Adcole, TRW, Ball Aerospace
Horizon Sensor	10^{-1} to 1°	2 to 5	Barnes, Ithaco, Lockheed Martin, Lockheed Barnes
Star Sensor	10^{-3} to 10^{-2}°	3 to 7	Ball Aerospace, Bendix, Honeywell, Hughes
Magnetometer	0.5 to 3°	~1	Schonstedt, Develco

Adapted from Larsen, W. J. and Wertz, J. R., Eds., *Space Mission Analysis and Design,* Torrance, CA: Microcosm Inc. and Dordrecht, The Netherlands: Kluwer Academic Publishers, 1992, p.360.

visually brightest object in the sky, having a total radiation per unit area of 1353 W m^{-2} at Earth distances [3]. Also, it is generally accepted as a valid point source for most attitude applications, having an angular radius of 0.25° at Earth distances. Increased measurement accuracy can be obtained by determining its centroid. Even though sun sensors are quite accurate (0.01° to 3.0°), they do require clear fields of view, and sometimes suffer periods of eclipse from both the Earth and the moon [4]. Also, sensitive equipment (such as imaging devices) must be protected from the powerful radiation of direct sunlight. When the sun is available, the angle between it and the sensor's primary axis is referred to as the *sun angle.*

For spacecraft in near-Earth orbits, the Earth is the second brightest object in the sky and covers as much as 40% of the sky. Earth *horizon sensors* detect the interface between the Earth's edge (or limb) and the space background. Horizon sensors can detect either of the Earth's visible limb (albedo sensor), infrared limb, or air glow. The infrared limb is the edge between the warm Earth and the cold space background. The air glow is a region of the atmosphere around the Earth that is visible to the spacecraft when it is on the night side of the Earth. Accuracies for horizon sensors are in the 0.1° to 1.0° range. Increased accuracy requires Earth oblate spheroid modeling [4]. Some problems associated with albedo detection include the distortion effects of the Earth's atmosphere, falsely identifying the day/night terminator crossing as the true Earth limb, and the considerable variability of the Earth's albedo in the visible spectrum (varies from land, sea, ice).

Most sensors used to detect the Earth's horizon are scanning sensors with narrow fields of view that measure the time between horizon crossings. In general, two horizon crossings occur per sensor scan period: one crossing when the sensor scans from the space background onto the Earth, followed by a second crossing when the sensor scans from the Earth back to space. The combination of horizon crossing times, scan rate, and spacecraft altitude allows for the computation of the Earth's apparent *angular radius.* The apparent angular radius will be smaller than the real (or physical) angular radius if the spacecraft is tilted away from the Earth nadir vector. The nadir vector is defined as the vector connecting the center of the spacecraft to the center of the Earth. To see this effect, one needs to compute the Earth's physical radius ρ, which for a given spacecraft altitude h (in kilometers), is given by $\rho = \sin^{-1}[(6371)/(6371 + h)]$.

If the spacecraft horizon sensor is pointing exactly *nadir*, then the apparent angular radius as measured by the sensor will agree with the physical radius given by the above relation for ρ. However, if the horizon sensor is pointed away from nadir, the horizon crossing times will be smaller than when pointing exactly nadir. This results in an apparent angular radius that is smaller than the physical radius by an amount proportional to the angle between the sensor axis and the nadir vector. This angle is referred to as the *nadir angle.*

Star sensors are used when extreme accuracy requirements are necessary. This high degree of sensor accuracy (0.003° to 0.01°) can be attributed mainly to the point source nature and precise fixed location of stars in space. Star sensors may be categorized as either star trackers or star mappers. A star tracker utilizes a wide field of view in order to search for a given star of specific brightness. A star mapper is similar to a tracker, except that it scans over many stars, recording their relative positions and angular separations. By comparing the recorded data with that from a *star catalog* (database), exact spacecraft

© 1999 by CRC Press LLC

orientation can be obtained. The angle between the star line-of-sight and the sensor's primary axis is referred to as the *star angle*.

The accuracy of star sensors is obtained with higher costs, however. Star sensors are generally heavier and consume more power than other types of attitude sensors. In addition, star sensors are quite sensitive to stray light sources such as sunlight reflected from the spacecraft or the Earth and sunlight scattered from dust particles and jet exhausts [4]. Most rely on optical shielding to reduce the effects of stray light.

Magnetic sensors (called *magnetometers*) measure both the magnitude and direction of the Earth's magnetic field. The difference in orientation between the measured field and the true field translates into attitude determination. Magnetometer accuracies (0.5° to 3.0°) are usually less than the other sensor types because of the uncertainty in the Earth's true field, which tends to change or shift over time. In addition, the Earth's magnetic field decreases with increasing altitude, and magnetometers are generally limited to altitudes of about 6000 km. For this reason, magnetometers are often used with one of the other sensor types already discussed for improved measurement accuracy [2].

Automatic On-Line Attitude Measurement for Ships and Crane Lifters

For on-line attitude measurement for ships and crane lifters, the first thing that comes to mind is to use gyros. However, because they often suffer from drifts, accurate attitude measurements might not be achieved using the gyros. Accordingly, one uses attitude on-line measurement systems that do not utilize gyros but servo-type accelerometers and inclinometers. The philosophy of the measurement systems introduced here is to make the best use of the system dynamics of the object and the sensors and to apply Kalman filters or adaptive filters to achieve high measurement accuracy.

Attitude Measurement for Ships

On-line accurate measurement of a ship's attitude is extremely important in exact search of the seabed patterns with sonars [5, 6]. It is also required by high-performance ships like hovercrafts from the viewpoint of suppressing swings by the waves. The measurement of a ship's attitude can usually be reduced to that of the heaving, rolling, and pitching of the ship. For such a measurement, a heave sensor has been used, whose output is given by double integration of the output of an accelerometer vertically directed with a gyroscope. However, since the initial values of heaving displacement and its velocity are unknown, the output will contain a bias that increases with time, and the accuracy of the sensor deteriorates considerably. From this viewpoint, one introduces a strapdown-type on-line measurement system that adequately processes the outputs of the two servo-type inclinometers and one accelerometer mounted on the ship [7].

Location of Sensors and Outputs

The two servo-type inclinometers and one servo-type accelerometer are located on the deck (at the point A) of vertical distance L from O, the intersection of rolling and pitching axes (see Figure 10.10).The two inclinometers are set in such a way that the rolling and the pitching angles are measured respectively. The accelerometer is set upward to the deck to obtain the information on the heaving. Because inclinometers were originally developed for the measurement of the tilt angles of static inclined surfaces, the rigid pendulum inside the sensor is considerably affected by the ship's acceleration other than the gravitational one. Applying Lagrange's equations of motion [8, 9] to rigid pendulums and calculating the torques to keep their deflections from the principal axes almost zero yields the sensor outputs [7]:

$$z_1(t) = \theta(t) - \frac{L}{g}\ddot{\theta}(t) + v_1(t) \tag{10.45}$$

$$z_2(t) = p(t) - \frac{L}{g}\ddot{p}(t) + v_2(t) \tag{10.46}$$

© 1999 by CRC Press LLC

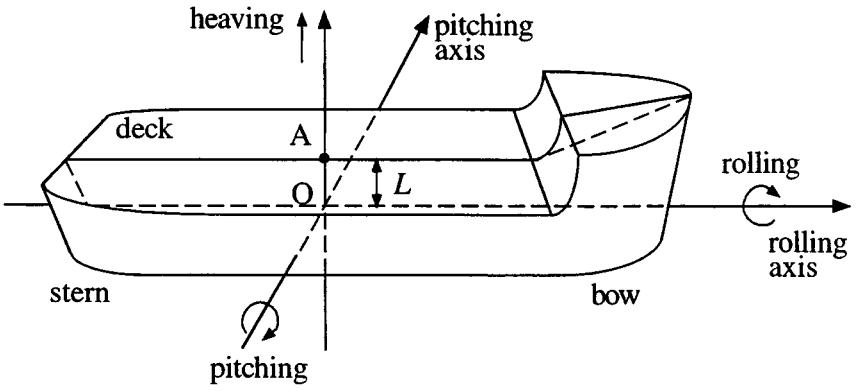

FIGURE 10.10 Location of sensors.

where $z_1(t)$, $z_2(t)$, $\theta(t)$, $p(t)$, and g denote, respectively, the outputs of the two inclinometers, the rolling and the pitching angles, and the gravitational acceleration ($v_1(t)$ and $v_2(t)$: noises of the outputs, including the approximation errors in deriving the outputs).

On the other hand, the accelerometer output is expressed as:

$$z_3(t) = \left(g + \alpha(t)\right)\cos\theta(t)\cos p(t) + v_3(t) \tag{10.47}$$

where $\alpha(t)$ and $v_3(t)$ represent, respectively, the heaving acceleration and the accelerometer noise.

Dynamics of Attitude Signals

It is well known that each of the heaving, rolling, and pitching in inshore seas has two dominant waves in a short interval. That is, a sinusoidal wave of long periodic length (in the range of 6 s to 10 s) and a sinusoidal wave of short periodic length (in the range of 2 s to 3 s) [10–12]. Thus, one model each of the signals in a short interval by a composite wave of the two dominant sinusoidal waves. For the heaving (in a short interval), the displacement is modeled by:

$$x(t) = a_1 \sin(\omega_1 t + \varphi_1) + a_2 \sin(\omega_2 t + \varphi_2) \tag{10.48}$$

with the parameters $\{a_i\}$, $\{\varphi_i\}$, and $\{\omega_i\}$ unknown. From the 4th-order differential equation satisfied by the $x(t)$, we obtain the linear dynamic equation [7]:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t), \qquad A \equiv \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\omega_1^2\omega_2^2 & 0 & -(\omega_1^2 + \omega_2^2) & 0 \end{bmatrix} \tag{10.49}$$

where $\mathbf{x}(t) \equiv (x_1, x_2, x_3, x_4)^{\mathrm{T}}$ ($x_n = d^{n-1}x/dt^{n-1}$ ($n = 1, \ldots, 4$)). On the other hand, the rolling and pitching angles can be modeled by:

$$x(t) = a_1 \sin(\omega_1 t + \varphi_1) + a_2 \sin(\omega_2 t + \varphi_2) + b \tag{10.50}$$

because there are usually some biases associated with them. From the 5th-order differential equation which Equation 10.50 satisfies, we get the similar state variable representation of the model as

© 1999 by CRC Press LLC

Equation 10.49. In practice, the heaving, rolling, and pitching signals have many nondominant sinusoidal waves in addition to the dominant ones. Therefore, Equation 10.49 is modified by introducing a white Gaussian noise $w(t)$ with zero mean and adequate variance σ^2 as follows:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \Gamma w(t) \tag{10.51}$$

where $\Gamma = (0,1,0,0)^T$ for the heaving and $\Gamma = (0,1,0,0,0)^T$ for the rolling and pitching. The higher the order of the models, the better the measurement accuracy will be. If we consider the on-line measurement of the signals, Equation 10.51 will be sufficient.

On-Line Attitude Measurement

The observation Equations 10.45 and 10.46 are expressed using their own state vector $\mathbf{x}(t)$. The observation equations in a discretized form are:

$$y_k = H\mathbf{x}_k + v_k \tag{10.52}$$

where $H = [1,0,-L/g,0,0]$ and y_k, $\mathbf{x}_k$, and v_k, respectively, denote $y(t)$, $\mathbf{x}(t)$, and $v(t)$ of the corresponding signals at the k-th sampling instant [7, 9]. The discretized form of the dynamic Equation 10.51 is:

$$\mathbf{x}_{k+1} = F\mathbf{x}_k + \mathbf{w}_k \tag{10.53}$$

where

$$F \equiv \Phi(t)\Big|_{t=\Delta T}, \quad \Phi(t) \equiv L^{-1}\left\{(sI - A)^{-1}\right\}. \tag{10.54}$$

Here, L^{-1} and ΔT, respectively, denote the inverse Laplace transformation and the sampling period. The discretized transition noise $\mathbf{w}_k$ becomes a white Gaussian noise with zero mean and covariance:

$$W = \sigma^2 \int_0^{\Delta T} \Phi(\Delta T - \tau)\Gamma\Gamma^T\Phi^T(\Delta T - \tau)\,d\tau \tag{10.55}$$

The measurement of the rolling and pitching can thus be reduced to the state estimation of the linear discrete dynamic systems (Equations 10.52 and 10.53), if the angular frequencies ω_1 and ω_2 are given and v_k is assumed to have a white Gaussian property. The state estimation is achieved by a Kalman filter [7, 13]. However, difficulties in implementing the filter are that the exact values of the two angular frequencies are a priori unknown and also time variant. To overcome the difficulty, adequate candidates $\{(\omega_1^i, \omega_2^i);\ 1 \le i \le M\}$ for the parameters $\{\omega_1, \omega_2\}$ are set and a bank of Kalman filters is used. Then, the final estimate is obtained as the conditional expectation of the state estimate as follows:

$$\hat{\mathbf{x}}^0_{k/k} \equiv \sum_{i=1}^{M} p_k^i \hat{\mathbf{x}}^i_{k/k} \tag{10.56}$$

where $\hat{\mathbf{x}}^i_{k/k}$ represents the state estimate $\hat{\mathbf{x}}_{k/k}$ for the i-th candidate $\Omega_i = (\omega_1^i, \omega_2^i)$, and p_k^i denotes the conditional posteriori probability of the i-th candidate calculated based on the Bayesian theorem:

$$p_k^i = \frac{p(y_k / \Omega_i, Y^{k-1})\, p_{k-1}^i}{\sum_{j=1}^{M} p_{k-1}^j\, p(y_k / \Omega_j, Y^{k-1})} \tag{10.57}$$

© 1999 by CRC Press LLC

Here, $p(y_k/\Omega_i, Y^{k-1})$ represents the conditional Gaussian probability density function of y_k under Ω_i and $Y^{k-1} \equiv \{y_j; j \le k-1\}$, whose mean and variance are calculated recursively [7].

The proposed measurement system can adaptively and automatically select the most appropriate candidate versus time. It thus enables an accurate on-line measurement of the rolling and pitching whose dominant angular frequencies vary with time. The first, second, and third components of the final estimate $\hat{\mathbf{x}}^{o}_{k/k}$ represent, respectively, the estimates of the displacement, velocity, and acceleration. The proposed system thus has an advantage in that it can measure not only the displacements, but also the velocities and the accelerations of the three signals. In order to improve the measurement accuracy of the rolling and pitching, one should place the inclinometers near the intersection O of the rolling and the pitching axes.

Finally, the dynamics of the heaving is given by Equation 10.51 similar to that of the rolling and pitching. Substituting the estimates $\hat{\theta}(t)$ and $\hat{p}(t)$ obtained above into Equation 10.47 and subtracting the effect of the gravitational acceleration, one can derive a linear observation equation for $\alpha(t)$:

$$\begin{aligned} y_k &= \left[z_3(t) - g\cos\hat{\theta}(t)\cos\hat{p}(t) \right]\Big|_{t=k\Delta T} \\ &= H_k \mathbf{x}_k + v_k \end{aligned} \tag{10.58}$$

where $H_k = [0,0,\cos\hat{\theta}(t)\cos\hat{p}(t),0]\big|_{t=k\Delta T}$

$$v_k = v_3(k\Delta T)$$

Thus, the on-line measurement of the heaving is also realized by executing the same procedure as described before. The location of the rolling and pitching axes were assumed to be known; however, even when they are unknown, the attitude measurement system described above is effective, if we introduce the candidates on the location of the axes adding to the angular frequencies.

Attitude Measurement for Crane Lifters

Dynamics of Attitude Signals

An illustrative diagram of a crane lifter system is shown in Figure 10.11. One of the easiest ways to measure the attitude of the lifter is to set up a high-resolution camera on the bottom of the trolley and to track a mark on the top of the lifter. However, it increases the cost and also the difficulty in maintenance. Furthermore, sometimes the scheme does not work because of shadows and light reflection. As previously mentioned, for gyros not offering sufficiently accurate measurement, a high-sensitivity servo-type accelerometer is used to extract the attitude signals. When setting up the sensor on the lifter, however, there is a secondary swing signal adding to the primary one, due to the free suspension of the lifter and the structure of the lifter. Despite its small amplitude, the secondary one has a higher frequency and for this reason has a large magnitude on the sensor output. The important signal for practical applications, such as the attitude control of the lifter, is the primary one, which has a larger amplitude with a lower angular frequency of $\omega = \sqrt{g/\ell}$ (g: the gravitational acceleration; ℓ: the wire length from the primary supporting point to the center of gravity of the pulley). If we try to attenuate the secondary swing signal by passing the output through a low-pass filter, the phase lag is also introduced into the primary swing signal and the signal can no longer be used for the accurate attitude control of the lifter.

For the above reasons, we introduce an autonomous measurement system that measures both the primary and the secondary swings by modeling the lifter system with a double pendulum and applying a Kalman filter to it [14]. The dynamics of the trolley-lifter system is derived using Lagrange's equations of motion [8, 9].

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{x}}\right) - \frac{\partial T}{\partial x} + \frac{\partial V}{\partial x} = u - z\dot{x} \tag{10.59}$$

© 1999 by CRC Press LLC

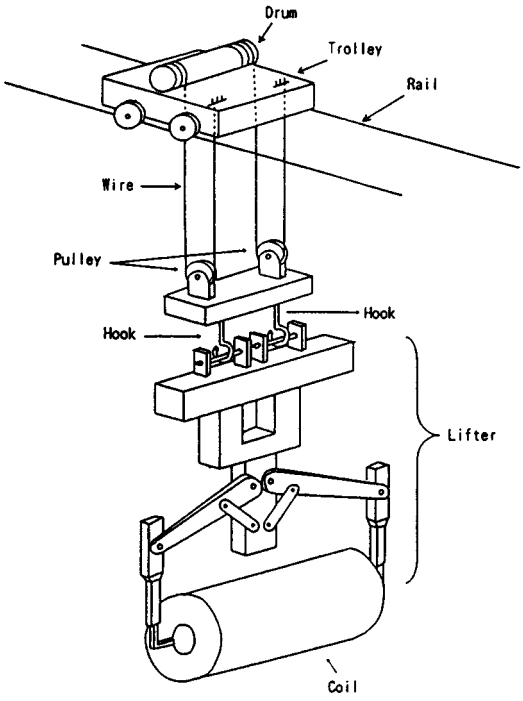

FIGURE 10.11 A crane lifter system.

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{\theta}_i}\right) - \frac{\partial T}{\partial \theta_i} + \frac{\partial V}{\partial \theta_i} = 0 \quad (i = 1, 2) \tag{10.60}$$

where T and V represent, respectively, the kinetic and the potential energies of the trolley-lifter system, and θ_1, θ_2 denote, respectively, the angles that the primary and the secondary pendulums take against the vertical line. The other variables x, u, and a represent, respectively, the location of the trolley, the driving force, and the coefficient of friction between the trolley and the rail. Considering that θ_i, $\dot{\theta}_i$ ($1 \le i \le 2$) are small, the dynamic equation of the trolley-lifter system can be expressed as [14]:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{b}u(t) \tag{10.61}$$

where $\mathbf{x}(t)$ is the state vector $\mathbf{x}(t) = x, \dot{x}, \theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2)^T$. Taking into account the approximation errors in deriving Equation 10.61, air resistance, friction in the wires, and microscopic swings at the other connection points, it is reasonable to introduce white Gaussian noises $w(t)$ ($1 \le i \le 3$) with zero mean and appropriate variances to the dynamic Equation 10.61 as in Equation 10.49 as follows [14]:

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{b}u(t) + \Gamma\mathbf{w}(t) \tag{10.62}$$

where

$$\Gamma = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T, \qquad \omega(\mathbf{t}) = \left[\omega_1(t),\ \omega_2(t),\ \omega_3(t)\right]^T \tag{10.62a}$$

© 1999 by CRC Press LLC

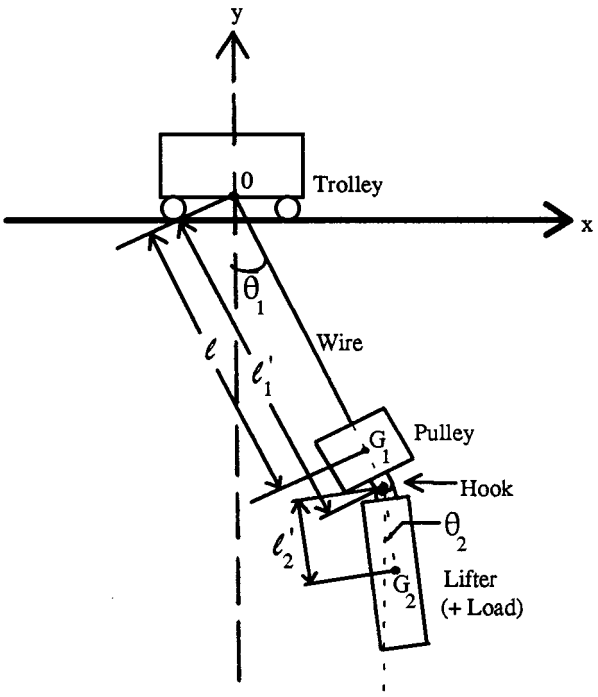

FIGURE 10.12 Dynamics of a trolley lifter system.

Sensor Outputs and On-Line Attitude Measurement

When a servo-type accelerometer is set up on the lifter (in the direction of the swing) at the place of the distance ℓ_2' from the secondary supporting point, the output of the sensor becomes [14]:

$$y(t) \cong -\ddot{x} - \ell_1'\ddot{\theta}_1 - \ell_2'\ddot{\theta}_2 - g\theta_2 \tag{10.63}$$

where ℓ_1' is the distance between the primary and the secondary supporting points (see Figure 10.12). Substitution of Equation 10.61 into Equation 10.63 yields an output expressed in terms of the state vector $\mathbf{x}(t)$, as in Equation 10.52. Using a rotary-encoder to measure the location and the velocity of the trolley, and then combining these three sensor outputs with the dynamic Equation 10.62 and applying a Kalman filter enables the state vector to be estimated on-line. Using this approach, both angular displacement and velocity of the deflections θ_1, θ_2 of the two pendulums can be measured exactly.

Aircraft Attitude Determination

The determination of aircraft attitude requires the measurement of angles about three independent body axes. These angles are the roll, pitch, and yaw angles. There are two primary means employed today for measuring these angles; the first method uses VGs to measure the roll and pitch angles, and a DG to measure the yaw angle. The second method, more commonly used today, employs an IMU for full three-axis attitude determination coupled with a baro-altimeter to correct for vertical drift errors in the IMU. Both methods are described below.

Vertical and Directional Gyro Analysis

A VG is a two degree-of-freedom gyro with its spin axis mounted nominally vertical. It employs two specific force sensors mounted nominally horizontal on the inner gimbal. The two angles measured by the VG — roll and pitch — require nearly identical analyses [1]. Consider the situation shown in

© 1999 by CRC Press LLC

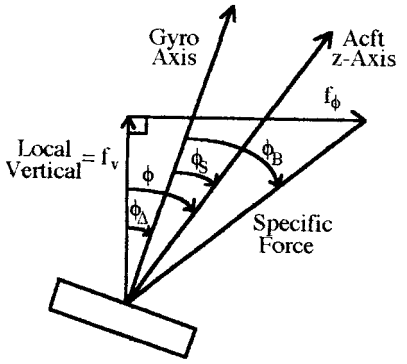

FIGURE 10.13 Vertical gyro analysis.

Figure 10.13, depicting an aircraft with a roll angle of ϕ with respect to the local vertical. The sensed roll angle ϕ_s is given by the difference in the actual roll angle and the gyro roll drift error ϕ_Δ:

$$\phi_s = \phi - \phi_\Delta \tag{10.64}$$

In order to compensate for this drift error, gyros employ a specific force sensor such as an electrolytic bubble device, which senses drift error. This correction device senses the angular difference between the specific force vector **f** acting on the aircraft roll axis and the gyro axis, as shown in Figure 10.13. Thus,

$$\phi_B = \tan^{-1}\left[\left(f_\phi / f_v\right) - \phi_\Delta\right] \cong \left(f_\phi / f_v\right) - \phi_\Delta \tag{10.65}$$

where f_ϕ is the side horizontal component of **f** and f_v = force of gravity is the vertical component. A similar analysis for the pitch angle θ yields:

$$\theta_s = \theta - \theta_\Delta \tag{10.66}$$

$$\theta_B = \tan^{-1}\left[f_\theta / f_v - \theta_\Delta\right] \cong f_\theta / f_v - \theta_\Delta \tag{10.67}$$

where f_θ is the back horizontal component of **f**. Next, define the gyro angular momentum vector by:

$$\mathbf{H}_{VG} = \left[J_x \dot{\phi}_\Delta, J_y \dot{\theta}_\Delta, -h\right] \tag{10.68}$$

where J_x and J_y are the sensor moments of inertia and h is the gyro spin angular momentum. In addition, define the inner gimbal axes angular velocity vector as:

$$\omega_{VG} = \left[\dot{\phi}_\Delta, \dot{\theta}_\Delta, 0\right] \tag{10.69}$$

Finally, define the gimbal torque vector by:

$$\mathbf{Q}_{VG} = \left[Q_{cx} + Q_{dx}, Q_{cy} + Q_{dy}, 0\right] \tag{10.70}$$

where

$$Q_{cx} = \textit{gimbal roll control torque} = -k_c \theta_B \tag{10.71a}$$

© 1999 by CRC Press LLC

$$Q_{cy} = \textit{gimbal pitch control torque} = k_c \phi_B \tag{10.71b}$$

$$Q_{dx} = \textit{gimbal roll disturbance torque} = -k_d\left(\dot{\phi}_\Delta - \dot{\phi}\right) + \textit{random torques} \tag{10.71c}$$

$$Q_{dy} = \textit{gimbal pitch disturbance torque} = -k_d\left(\dot{\theta}_\Delta - \dot{\theta}\right) + \textit{random torques} \tag{10.71d}$$

and the k_c and k_d are constant scaling factors related to each torque component.

Using the vectors defined in Equations 10.68 through 10.70, the gyro equations of motion are given by:

$$\frac{\partial}{\partial t}\left(\mathbf{H}_{VG}\right) + \left(\boldsymbol{\omega}_{VG} \times \mathbf{H}_{VG}\right) = \mathbf{Q}_{VG} \tag{10.72}$$

Taking the Laplace transform of the expansion of Equation 10.72, with the assumption that $J_x \cong J_y = J$, yields the following gyro equations of motion in the Laplace domain:

$$\begin{bmatrix} J_x s^2 + k_d s & -\left(hs + k_c\right) \\ hs + k_c & J_y s^2 + k_d s \end{bmatrix} \begin{bmatrix} \phi_\Delta(s) \\ \theta_\Delta(s) \end{bmatrix} \cong \begin{bmatrix} -k_c \theta_B + k_d s\phi(s) + \textit{random torques} \\ \left(k_c / g\right) f_\phi(s) + k_d s\theta(s) + \textit{random torques} \end{bmatrix} \tag{10.73}$$

For normal gyro operation, $J_x \cong J_y \cong 0$ and $k_d/h \ll 1$; so these factors may be ignored in Equation 10.73. Thus, solving for the desired roll and pitch angles under these assumptions gives [1]:

$$\phi_s = \begin{cases} \phi & \omega \gg k_c / h \\ \phi - f_\phi / g & \omega \ll k_c / h \end{cases} \tag{10.74}$$

$$\theta_s = \begin{cases} \theta & \omega \gg k_c / h \\ \theta - f_\theta / g & \omega \ll k_c / h \end{cases} \tag{10.75}$$

A DG is a two degree-of-freedom gyro with its spin axis mounted nominally horizontal and pointing in the direction of magnetic north. It employs a single specific force sensor mounted on the inner gimbal [1]. The DG measures the third required aircraft angle, yaw, generally denoted by ψ. The sensed yaw angle ψ_s is given by the difference in the actual yaw angle ψ (angle between the aircraft z-axis and true north) and the gyro heading angle drift error ψ_Δ (angle between the gyro axis and true north):

$$\psi_s = \psi - \psi_\Delta \tag{10.76}$$

Define the gyro angular momentum vector by:

$$\mathbf{H}_{DG} = \left[J_y \dot{\theta}_\Delta,\ J_z \dot{\psi}_\Delta,\ -h\right] \tag{10.77}$$

and the inner gimbal axes angular velocity vector as:

$$\boldsymbol{\omega}_{DG} = \left[\dot{\theta}_\Delta,\ \dot{\psi}_\Delta,\ 0\right] \tag{10.78}$$

© 1999 by CRC Press LLC

and the gimbal torque vector as

$$Q_{DG} = \left[Q_{cy} + Q_{dy}, Q_{cz} + Q_{dz}, 0 \right] \tag{10.79}$$

Here, the torque vector components are given by:

$$Q_{cy} = k_c \left(M_\Delta - \psi_\Delta \right) \tag{10.80a}$$

$$Q_{cz} = -k_c \theta_B \tag{10.80b}$$

$$Q_{dy} = -k_d \left(\dot{\theta}_\Delta - \dot{\theta} \right) + \textit{random torques} \tag{10.80c}$$

$$Q_{dz} = -k_d \left(\dot{\psi}_\Delta - \dot{\psi} \right) + \textit{random torques} \tag{10.80d}$$

where M_Δ = magnetic compass heading error (from true north). Therefore, the DG equations of motion are given in Laplace domain as:

$$\begin{bmatrix} J_y s^2 + k_d s & -\left(hs + k_c \right) \\ hs + k_c & J_z s^2 + k_d s \end{bmatrix} \begin{bmatrix} \theta_\Delta(s) \\ \psi_\Delta(s) \end{bmatrix} \cong \begin{bmatrix} k_c M_\Delta(s) + k_d s\theta(s) + \textit{random torques} \\ -k_c \theta_B + k_d s\psi(s) + \textit{random torques} \end{bmatrix} \tag{10.81}$$

The desired yaw angle measurement for the DG is thus given as [1]:

$$\psi_s = \begin{cases} \psi & \omega \gg k_c / h \\ \psi - M_\Delta & \omega \ll k_c / h \end{cases} \tag{10.82}$$

As indicated in Table 10.2, the accuracies of both VGs and DGs are approximately 1°. An improvement of over 2 orders of magnitude can be obtained through the use of inertial measurement units, which are described next.

Inertial Measurement Units (IMUs)

Inertial measurement units consist of gyroscopes and accelerometers that together provide full three-axis attitude measurements. Most are mounted on stable gimbaled platforms that remain locally horizontal via torquing devices. An IMU aboard an aircraft cannot measure exactly the local vertical due to the fact that the specific force acting on the aircraft has a horizontal component due to vehicle motion. In addition, since the vehicle is moving with respect to the inertial reference frame, the Earth's magnetic pole cannot be determined precisely [1].

These problems (errors) are minimized by aligning the IMU to be exactly horizontal and north pointing while the aircraft is stationary. Once platform motion begins, the IMU may be constantly realigned by sensing changes in the direction of vertical and north, and then applying appropriate torques to the platform to keep it properly aligned. This realignment is accomplished by integrating the two orthogonal accelerometer outputs to determine the components of horizontal velocity. This data, combined with the Earth's rotation rate, yields the desired rates of change in local vertical and true north at the vehicle's current latitude and longitude. Performing a second integration of the sensor outputs yields an estimate of relative position.

© 1999 by CRC Press LLC

Analysis in Bryson et al. [1], has shown that the pitch angle (variation in platform horizontal position) is given by the IMU sensor output as:

$$\theta(t) = \frac{-\varepsilon}{\omega_s} \sin(\omega_s t) - \frac{b}{g} \tag{10.83}$$

where ε = gyro drift rate error, b = specific force sensor error, and $\omega_s \equiv$ Schuler frequency = $\sqrt{g/R}$, [g = force of gravity, R = Earth's radius]. Thus, the platform root-mean-square pitch angle becomes:

$$\theta_{rms} = \left[\frac{1}{2} \left(\frac{\varepsilon}{\omega_s} \right)^2 + \left(\frac{b}{g} \right)^2 \right]^{\frac{1}{2}} \tag{10.84}$$

Using typical values for ε ($\cong 0.015° \text{ h}^{-1}$), ω_s ($\cong 0.71° \text{ h}^{-1}$), and b ($\cong 0.01$) yields an rms pitch angle error of $\theta_{rms} = 0.01°$. Thus, it is apparent that under normal operating conditions the IMU provides a two orders-of-magnitude improvement in sensor accuracy when compared to the VG and DG.

Spacecraft Attitude Determination

Most spacecraft attitude determination techniques rely upon finding the orientation of a single axis in space (e.g., the spacecraft z-axis) plus the spacecraft rotation about this axis. This provides a full three-axis attitude solution. In order to achieve this, reference sources that are external to the spacecraft must be used. Specifically, full three-axis spacecraft attitude determination requires at least two external vector measurements. Commonly used reference sources for these external vector measurements include the sun, Earth, moon, stars, planets, and the Earth's magnetic field. In addition, IMUs are also used to provide the necessary attitude measurements.

Attitude Determination Methodology

The first step in attitude determination is to determine the angles between the spacecraft's primary axis and the two (or more) attitude reference sources. For example, suppose a particular spacecraft is using the sun and the Earth for attitude reference. The two angles in this case are referred to as the sun angle β_S and the nadir angle Γ_N. Since the orientation of even a single spacecraft axis is unknown at this point, these angles establish two *cones* along which the attitude vector **A** must lie. Since the attitude vector must lie on both cones, it must lie along the intersection between the two cones [4] (See Figure 10.14). The two vectors, notably $\mathbf{A}_1$ and $\mathbf{A}_2$, resulting from the intersection of these two cones may be determined by the following method derived by Grubin [15]. Let **S** represent the sun vector, **E** the spacecraft nadir vector, and **A** the desired attitude vector, each defined in Cartesian space as follows:

$$\mathbf{S} = (S_x, S_y, S_z) \tag{10.85}$$

$$\mathbf{E} = (E_x, E_y, E_z) \tag{10.86}$$

$$\mathbf{A} = (A_x, A_y, A_z) \tag{10.87}$$

Let the vectors **S**, **E**, and **N** define a set of base unit vectors with:

$$\mathbf{N} = \frac{\mathbf{S} \times \mathbf{E}}{|\mathbf{S} \times \mathbf{E}|} = (N_x, N_y, N_z) \tag{10.88}$$

© 1999 by CRC Press LLC

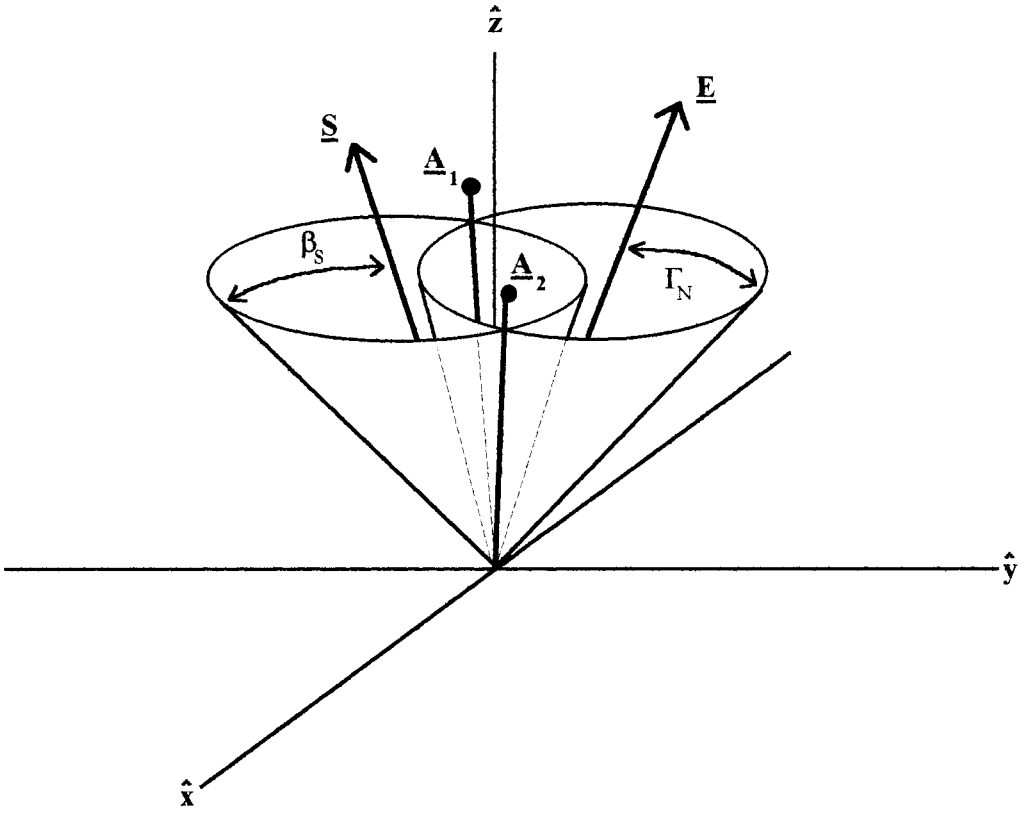

FIGURE 10.14 Relationship between reference vectors and single-axis attitude cones.

If we introduce a proper set of scaling factors as follows:

$$I_x = \frac{\left[\cos\beta_S - (\mathbf{S}\cdot\mathbf{E})\cos\Gamma_N\right]}{1-(\mathbf{S}\cdot\mathbf{E})^2} \tag{10.89a}$$

$$I_y = \frac{\left[\cos\Gamma_N - (\mathbf{S}\cdot\mathbf{E})\cos\beta_S\right]}{1-(\mathbf{S}\cdot\mathbf{E})^2} \tag{10.89b}$$

$$I_z = \sqrt{1 - I_x\cos\beta_S - I_y\cos\Gamma_N} \tag{10.89c}$$

then the two possible attitude vectors $\mathbf{A}_1$ and $\mathbf{A}_2$ are found to be:

$$\mathbf{A}_{1,2} = \left[\left(I_x S_x + I_y E_y \pm I_z N_x\right), \left(I_x S_y + I_y E_y \pm I_z N_y\right), \left(I_x S_z + I_y E_z \pm I_z N_z\right)\right] \tag{10.90}$$

In Equations 10.88 through 10.90, $\mathbf{S}\times\mathbf{E}$ represents the Cartesian vector product, and $\mathbf{S}\cdot\mathbf{E}$ represents the Cartesian scalar product. The radicand in Equation 10.89c may be negative, thus producing imaginary

© 1999 by CRC Press LLC

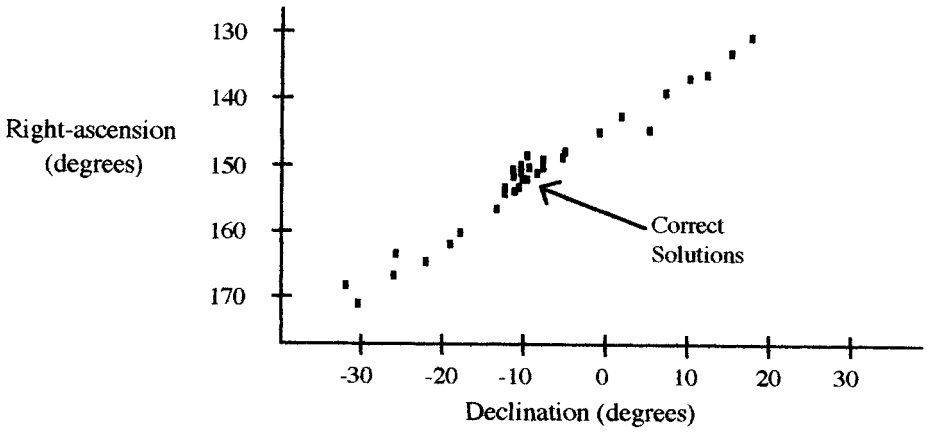

FIGURE 10.15 Method of trace averaging for resolving ambiguous attitude solutions.

values for I_z whenever the two cones do not intersect. Such occurrences are usually attributed to sensor error or random noise fluctuations. In this case, one can add a predetermined sensor bias to both sensors in order to "broaden" the cone angles, thus forcing the cones to intersect.

It should be noted that for most applications involving spacecraft attitude determination, the principle coordinate system used is the *celestial sphere* coordinate system. This coordinate system has the *z*-axis aligned with the Earth's polar axis, and the *x*-axis aligned with the intersection of the Earth's equatorial plane and the Earth's orbital plane around the sun (i.e., aligned with the *vernal equinox*). In this coordinate system, all vectors are considered unit vectors and the two principle measurements describing a vector's position are the *right-ascension* and *declination angles*, denoted Ω and Δ, respectively. Thus, the sun vector **S** and the Earth nadir vector **E** used in Equations 10.85 and 10.86 will, in general, be given as right-ascension and declination angles that can be converted to Cartesian coordinates via the following set of transformations:

$$x = \cos(\Omega)\cos(\Delta); \quad y = \sin(\Omega)\cos(\Delta); \quad z = \sin(\Delta) \tag{10.91a}$$

$$\Omega = \tan^{-1}(y/x); \quad \Delta = \sin^{-1}(z) \tag{10.91b}$$

The final step in measuring three-axis attitude is to determine which attitude solution is correct, $\mathbf{A}_1$ or $\mathbf{A}_2$, and then measure the rotation about this axis. The two ambiguous attitude solutions may be resolved by comparison with a priori attitude information, if available, or through the use of *trace averaging* [4]. Trace averaging is a method of plotting each attitude solution on a right-ascension versus declination plot and choosing the area of greatest concentration as the correct solution, as demonstrated in Figure 10.15. Since the attitude is assumed to change more slowly than the attitude sensor's sample rate, over short time intervals the data for the correct solution usually form a "cluster" near the correct attitude; the data for the incorrect solution are usually much more scattered.

Once the correct attitude vector has been obtained, the orientation of the remaining two orthogonal axes may be found by measuring the rotation, or phase angle, of the spacecraft about the preferred axis. Any sensor measurement that provides this phase angle may be used. An example of this technique is provided by the panoramic annular lens attitude determination system (PALADS), described in the next section. This imaging system uses a unique "three-dimensional" lens that provides simultaneous detection of two (or more) reference sources [16]. This information, combined with the orientation of the single axis, uniquely determines three-axis attitude.

© 1999 by CRC Press LLC

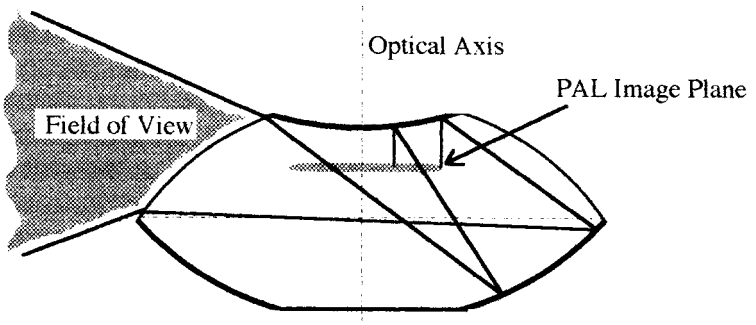

FIGURE 10.16 Panoramic annular lens ray diagram.

The three angles derived above, which are commonly referred to as *Euler angles*, define the orientation of the three spacecraft axes with respect to the chosen reference frame. A more formal treatment of the attitude solution usually requires specifying the components of the 3×3 attitude matrix **A**. Each component of the attitude matrix defines the angular relationship between a given spacecraft axis and a reference frame axis. Various methods exist for computing the attitude matrix **A** (see [4]); the preferred method depends on the particular application at hand.

PALADS

The primary component of PALADS is the panoramic annular lens (PAL), a single-element lens made from a high index of refraction glass with certain portions of the lens coated with a mirrored surface. Hence, it relies on both refraction and reflection in forming an image (Figure 10.16). The lens is unique in that it images a three-dimensional object field onto a two-dimensional image plane, whereas a "normal" lens is capable of only imaging two-dimensional object space onto an image plane. The combination of high index of refraction glass and mirrored surfaces provides the PAL with a field of view extending from approximately 65° to 110° from the optical axis. This 45° field of view covers the entire 360° surrounding the optical axis [17]. Any ray originating from outside the 45° field of view will not form a part of the image. The PAL may be attached to any high-quality imaging system using an appropriate *transfer lens*. As currently configured, the PALADS imaging system utilizes a Sony XC-73 charged-couple device (CCD), a black and white video camera coupled to the PAL via a Nikon *f*/1.4 transfer lens.

The hemispherical view provided by PALADS allows for single-sensor detection of multiple attitude reference sources, such as the Earth and the sun or moon. The position of each reference source in the image plane translates into a unique azimuth elevation angle between the PAL's optical axis and the reference source. Since the PAL has a 360° field of view surrounding the optical axis, it may detect several reference sources simultaneously. The data points associated with each source are extracted from the image plane using digital image processing techniques. Thus, it is easy to see how a single image from PALADS (containing two or more reference sources) provides the necessary angle data to determine three-axis spacecraft attitude.

References

1. A. E. Bryson, *Control of Spacecraft and Aircraft*, Chapter 10, Princeton, NJ: Princeton University Press, 1994.
2. W. J. Larson and J. R. Wertz (eds.), *Space Mission Analysis and Design*, Chapter 11, Torrance, CA: Microcosm Inc. and Dordrecht, The Netherlands: Kluwer Academic Publishers, 1992.
3. NASA Technical Memorandum NASA TM X-64757, *Terrestrial Environment (Climatic) Criteria Guidelines for Use in Aerospace Vehicle Development (1973 Revision)*, Marshall Space Flight Center, AL, 1973.

© 1999 by CRC Press LLC

4. J. R. Wertz (ed.), *Spacecraft Attitude Determination and Control*, Chapters 11 and 12, The Netherlands: Reidel Publishing Company, 1980.
5. R. D. Angelari, A deterministic and random error model for a multibeam hydrographic sonar system, *Proc. OCEANS'78. The Ocean Challenge*, 1978, 48-53.
6. C. de Moustier, T. Hylas, and J. C. Phillips, Modifications and improvements to the Sea Beam system on board R/V Thomas Washington, *Proc. OCEANS'88 — A Partnership of Marine Interests*, 1988, 372-378.
7. S. Tanaka and S. Nishifuji, Automatic on-line measurement of ship's attitude by use of a servo-type accelerometer and inclinometers, *IEEE Trans. Instrum. Meas.*, 45, 209-217, 1996.
8. D. G. Shultz and J. L. Melsa, *State Functions and Linear Control Systems*, New York: McGraw-Hill, 1967.
9. Y. Takahashi, M. J. Rabins, and D. M. Auslander, *Control and Dynamic Systems*, Reading, MA: Addison-Wesley, 1971.
10. S. Tanaka and S. Nishifuji, On-line sensing system of dynamic ship's attitude by use of servo-type accelerometers, *IEEE J. Oceanic Eng.*, 20, 339-346, 1995.
11. S. Tanaka, On automatic attitude measurement system for ships using servo-type accelerometers (in Japanese), *Trans. SICE*, 27, 861-869, 1991.
12. D. E. Cartwright and M. S. Longuet-Higgins, The statistical distribution of the maxima of a random function, *Proc. Roy. Soc. London*, Ser. A, 237, 212-232, 1956.
13. R. E. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME, J. Basic Eng.*, 82, 35-45, 1960.
14. S. Tanaka, S. Kouno, and H. Hayashi, Automatic measurement and control of attitude for crane lifters (in Japanese), *Trans. SICE*, 32(1), 97-105, 1996.
15. C. Grubin, Simple algorithm for intersecting two conical surfaces, *J. Spacecraft Rockets*, 14(4), 251-252, 1977.
16. M. A. Stedham and P. P. Banerjee, The panoramic annular lens attitude determination system, *SPIE Proceedings, Space Guidance, Control, and Tracking II,* Orlando, FL, 17-18 April, 1995.
17. J. A. Gilbert, D. R. Matthys, and P. Greguss, Optical measurements through panoramic imaging systems, *Proc. Int. Conf. Hologram Interferometry and Speckle Metrology*, Baltimore, MD, November 4-7, 1990.

10.3 Inertial Navigation

Halit Eren and C. C. Fung

The Principles

The original meaning of the word *navigation* is "ship driving." In ancient times when sailing boats were used, navigation was a process of steering the ship in accordance with some means of directional information, and adjusting the sails to control the speed of the boat. The objective was to bring the vessel from location A to location B safely. At present, navigation is a combination of science and technology. No longer is the term limited to the control of a ship on the sea surface; it is applied to land, air, sea surface, underwater, and space.

The concept of inertial-navigator mechanization was first suggested by Schuler in Germany in 1923. His suggested navigation system was based on an Earth-radius pendulum. However, the first inertial guidance system based on acceleration was suggested by Boykow in 1938. The German A-4 rocket, toward the end of World War II, used an inertial guidance system based on flight-instrument type gyroscopes for attitude control and stabilization. In this system, body-mounted gyro-pendulum-integrating accelerometers were used to determine the velocity along the trajectory. The first fully operational inertial auto-navigator system in the U.S. was the XN-1 developed in 1950 to guide C-47 rocket. Presently, inertial navigation systems are well developed theoretically and technologically. They find diverse applications,

© 1999 by CRC Press LLC

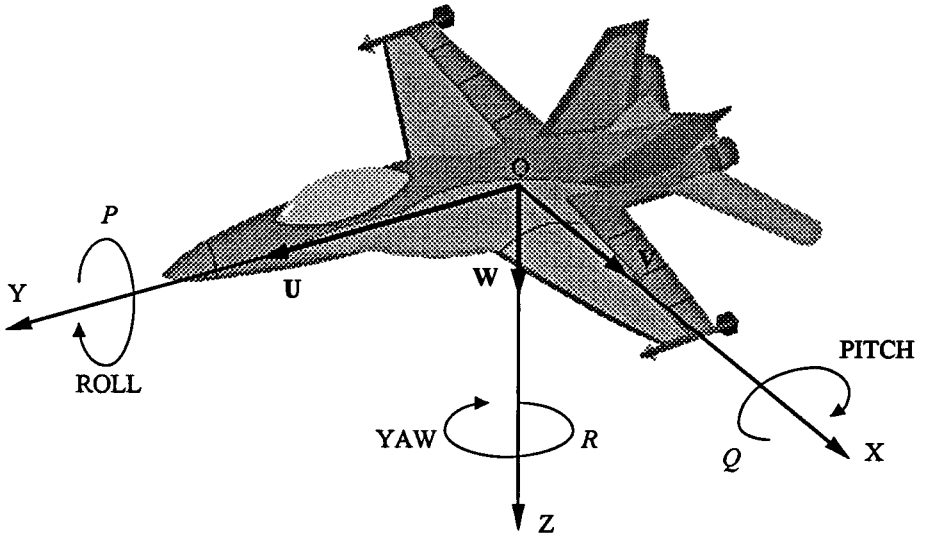

FIGURE 10.17 In inertial navigation, the movement of a vehicle, rocket, ship, aircraft, robot, etc. with respect to a reference axis is monitored. On the Earth's surface, the conventional reference is the Earth's fixed axes — North, East, and Down. A vehicle such as an aircraft or a marine vessel will have its own local axes, known as roll, pitch, and yaw.

allowing the choice of appropriate navigation devices, depending on cost, accuracy, human interface, global coverage, time delay, autonomy, etc.

Inertial navigation is a technique using a self-contained system to measure a vehicle's movement and determine how far it has moved from its starting point. Acceleration is a vector quantity involving magnitude and direction. A single accelerometer measures magnitude but not direction. Typically, it measures the component of accleration along a predetermined line or direction. The direction information is usually supplied by gyroscopes that provide a reference frame for the accelerometers. Unlike other positional methods that rely on external references, an *inertial navigation system* (INS) is compact and self-contained, as it is not required to communicate to any other stations or other references. This property enables the craft to navigate in an unknown territory.

Inertial navigation can be described as a process of directing the movement of a vehicle, rocket, ship, aircraft, robot, etc., from one point to another with respect to a reference axis. The vehicle's current position can be determined from "dead reckoning" with respect to a known initial starting reference position. On the Earth's surface, the conventional reference will be North, East, and Down. This is referred to as the *Earth's fixed axes*. A vehicle such as an aircraft or a marine vessel will have its own *local axes*: roll, pitch, and yaw, as shown in Figure 10.17.

The inertial sensors of the INS can be mounted in such a way that they stay leveled and pointing in a fixed direction. This system relies on a set of gimbals and sensors attached on three axes to monitor the angles at all times. This type of INS is based on a *navigational platform*. A sketch of a three-axis platform is shown in Figure 10.18. Another type of INS is the *strapdown system* that eliminates the use of gimbals. In this case, the gyros and accelerometers are mounted to the structure of the vehicle. The measurements received are made in reference to the local axes of roll, pitch, and yaw. The gyros measure the movement of angles in the three axes in a short time interval (e.g., 100 samples per second). The computer then uses this information to resolve the accelerometer outputs into the navigation axes. A schematic block diagram of the strapdown system is shown in Figure 10.19.

The controlling action is based on the sensing components of acceleration of the vehicle in known spatial directions, by instruments which mechanize Newtonian laws of motion. The first and second integration of the sensed acceleration determine velocity and position, respectively. A typical INS includes

© 1999 by CRC Press LLC

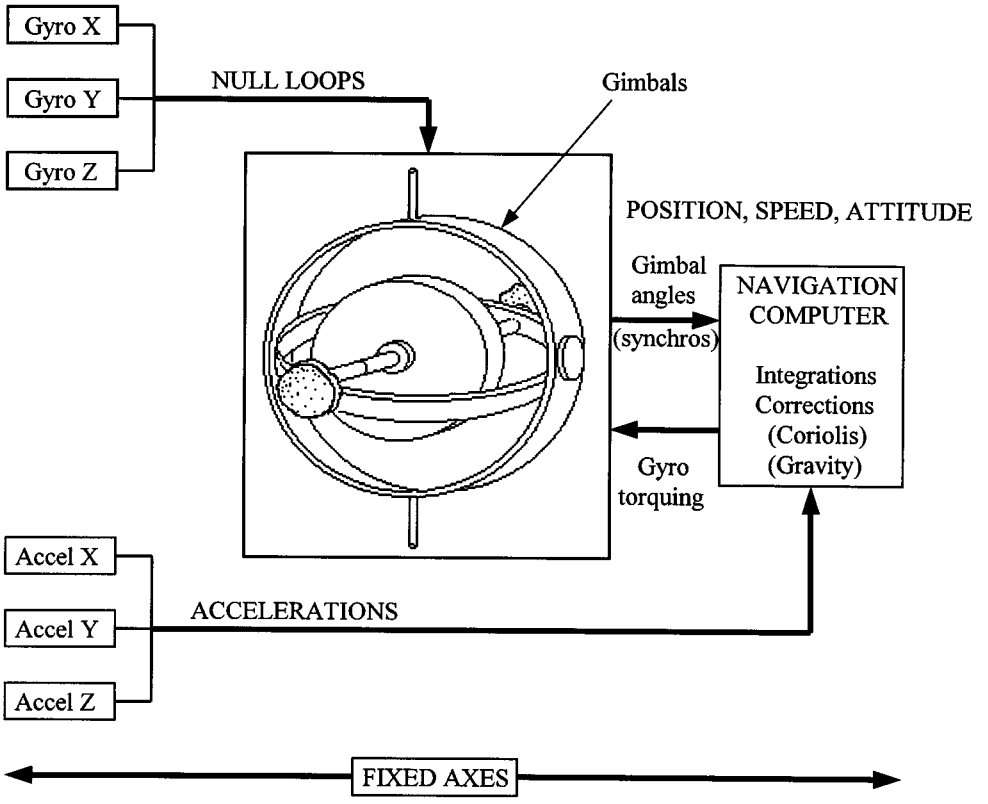

FIGURE 10.18 Some Inertial Navigation Systems, INS, are based on a navigational platform. The inertial sensors are mounted in such a way they can stay leveled at all times, pointing in a fixed direction. This system uses a set of gimbals and sensors attached on three axis in the *x*, *y*, and *z* directions to monitor the angles and accelerations constantly. The navigation computer makes corrections for coriolis, gravity, and other effects.

a set of gyros, a set of accelerometers, and appropriate signal processing units. Although the principle of the systems may be simple, the fabrication of a practical system demands a sophisticated technological base. The system accuracy is independent of altitude, terrain, and other physical variables, but is limited almost purely by the accuracy of its own components. Traditional INSs mainly relied on mechanical gyros and accelerometers, but today there are many different types available, such as optical gyroscopes, piezoelectric vibrating gyroscopes, active and passive resonating gyroscopes, etc. Also, micromachined gyroscopes and accelerometers are making an important impact on modern inertia navigation systems. A brief description and operational principles of gyroscopes and accelerometers suitable for inertial navigation are given below.

Major advances in INS over the years include the development of the *electrostatic gyro* (ESG) and the laser gyro. In ESG, the rotor spins at a speed above 200×10^3 rpm in a near-vacuum environment. The rotor is suspended by an electrostatic field; thus, it is free from bearing friction and other random torques due to mechanical supports. Hence, its operation results in a superior performance compared to others, closely resembling the peformance of a theoretical gyro. Although no system can claim to reach perfection, an ESG requires less frequent updates as compared to other mechanical gyros.

Gyroscopes

There are two broad categories: (1) mechanical gyroscopes and (2) optical gyroscopes. Within both of these categories, there are many different types available. Only the few basic types will be described to

© 1999 by CRC Press LLC

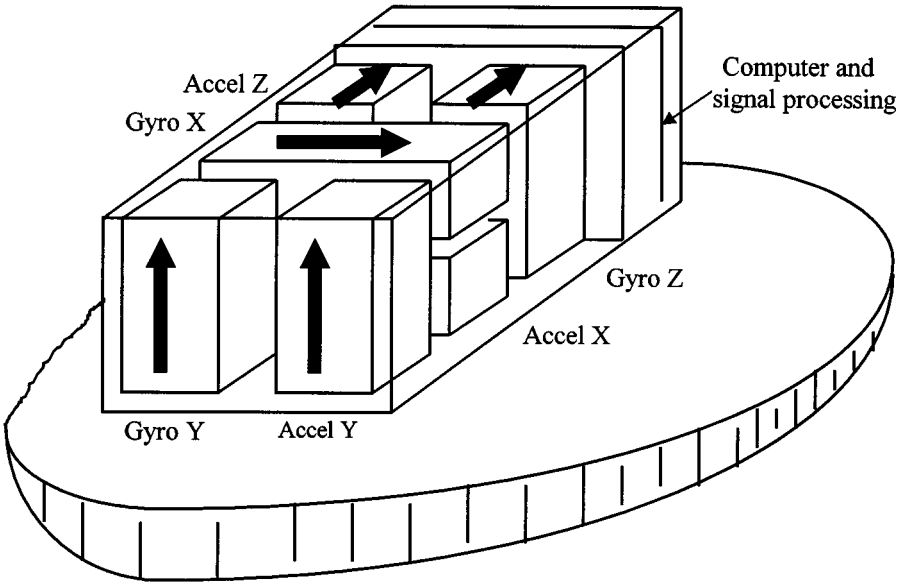

FIGURE 10.19 The use of a strapdown system eliminates the need for gimbals. The gyros and accelerometers are mounted rigidly on the structure of the vehicle, and the measurements are referenced to the local axes of roll, pitch, and yaw. The gyros measure the movement of angles in the three axes in short time intervals to be processed by the computer. This information is used, together with the accelerometer outputs, for predicting navigation axes.

illustrate the operating principles; detailed information may be found in the references listed at the end of this chapter.

Mechanical gyroscopes: The first mechanical gyroscope was built by Foucault in 1852, as a gimbaled wheel that stayed fixed in space due to angular momentum while the platform rotated around it. They operate on the basis of *conservation of angular momentum* by sensing the change in direction of an angular momentum. There are many different types, which are:

1. *Single degree of freedom gyroscopes*: include the rate, rate integrating, spinning rotor flywheel, electron, and particle gyros.
2. *Two degree of freedom gyroscopes*: incorporate the external gimbal types, two-axis floated, spherical free-rotor, electrically suspended, gas-bearing free-rotor gyros.
3. *Vibrating gyroscopes*: include the tuning fork, vibrating string, vibrating shell, hemispherical resonating, and vibrating cylinder gyros.
4. *Continuous linear momentum gyroscopes*: incorporate a steady stream of fluid, plasma, or electrons, which tends to maintain its established velocity vector as the platform turns. One typical example is based on a differential pair of hot-wire anemometers to detect the apparent lateral displacement of the flowing air column.

The operating principle of all mechanical gyroscopes is based on the conservation of angular momentum, as shown in Figure 10.20. The angular momentum is important since it provides an axis of reference. From Newton's second law, the angular momentum of a body will remain unchanged unless it is acted upon by a torque. The rate of change of angular momentum is equal to the magnitude of the torque, in vectorial form as:

$$T = \mathrm{d}H/\mathrm{d}t \tag{10.92}$$

where H = angular momentum (= inertia × angular velocity, $I\omega$).

© 1999 by CRC Press LLC

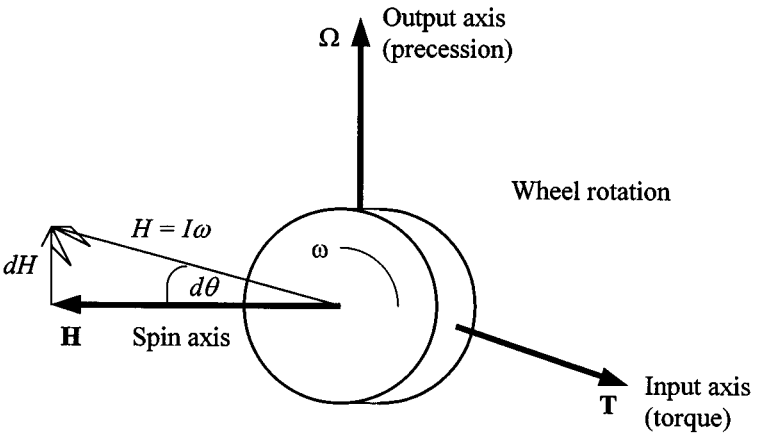

FIGURE 10.20 The operation principle of gyroscopes is based on the angular momentum of a carefully constructed rotating body. The angular momentum stabilizes the system. The angular momentum of a body will remain unchanged unless it is acted upon by a torque. If the torque is orthogonal to the spin axis, it cannot change the velocity, but it can change the direction in the same direction as the torque. The spin axis always tries to align with the external torque.

If a torque acts about the axis of rotation, it changes the angular velocity by:

$$T = I\,\mathrm{d}\omega/\mathrm{d}t = I\alpha \tag{10.93}$$

where I = inertia about the spin axis
α = angular acceleration

If the torque is orthogonal to the spinning axis, it cannot change the magnitude of the angular velocity vector, but it can change direction in the same direction as torque T; then:

$$\mathrm{d}H = H\,\mathrm{d}\theta \tag{10.94}$$

where θ = angle of rotation.

Therefore, from Equations 10.94 and 10.92:

$$T = \mathrm{d}H/\mathrm{d}t = H\,\mathrm{d}\theta/\mathrm{d}t = H\,\Omega \tag{10.95}$$

where Ω is the precession rate or the angular velocity of the spinning wheel about the axis normal to the plane of the spin and the input torque. Generally, the spin axis tries to align with the external input torque.

These equations can be elaborated to describe the operating principles of mechanical gyros by taking into account the angular momentum in *x*, *y*, and *z* directions, nutation, coriolis accelerations, directions of other influencing torques and linear forces, etc. Here, the operation of the well-known *flywheel gyroscope* will be described as the basis for further discussions on inertial navigation systems.

An example of a double-axis flywheel gyro is shown in Figure 10.21. In this type of gyroscope, an electrically driven rotor is suspended in a pair of precision low-friction bearings at both ends of the rotor axle. The rotor bearings are supported by a circular ring known as an *inner gimbal ring*, which in turn pivots on a second set of bearings that is attached to the *outer gimbal ring*. The pivoting action of the inner gimbal defines the horizontal axis of the gyro, which is perpendicular to the spin axis of the rotor. The outer gimbal ring is attached to the instrument frame by a third set of bearings that defines the vertical axis of the gyro that is perpendicular to both the horizontal axis and the spin axis. This type of

© 1999 by CRC Press LLC

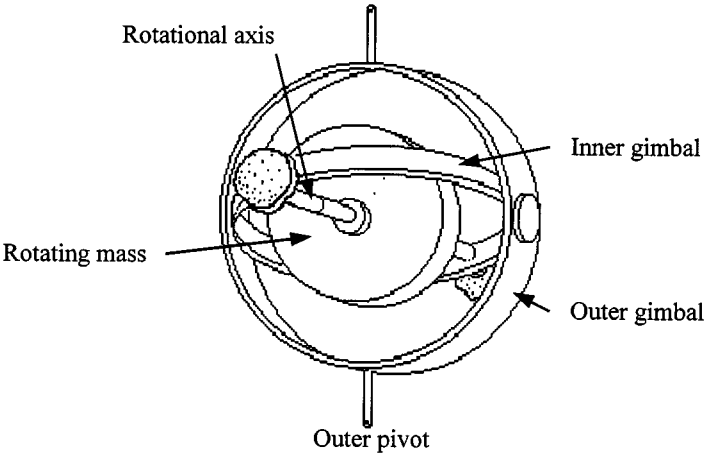

FIGURE 10.21 In a double-axis flywheel gyro, an electrically driven rotor is suspended by a pair of precision low-friction bearings at the rotor axle. The rotor bearings are supported by a circular inner gimbal ring. The inner gimbal ring in turn pivots on a second set of bearings attached to an outer gimbal ring. The pivoting action of the inner gimbal defines the horizontal axis of the gyro, which is perpendicular to the spin axis of the rotor. The outer gimbal ring is attached to the instrument frame by a third set of bearings. This arrangement always preserves the predetermined spin-axis direction in inertial space.

suspension has the property of always preserving the predetermined spin-axis direction in inertial space. Equations governing the two degrees of freedom gyroscope can be written using Equations 10.92 to 10.95. The torque with respect to an inertial reference frame can be expressed as:

$$T = \dot{H}_{\mathrm{I}} \tag{10.96}$$

If the Earth is taken as a moving reference frame, then:

$$\dot{H}_{\mathrm{I}} = \dot{H}_{\mathrm{E}} + \omega_{\mathrm{IE}}\, H \tag{10.97}$$

If the gyroscope itself is mounted on a vehicle (e.g., aircraft) that is moving with respect to the Earth, then:

$$\dot{H}_{\mathrm{E}} = \dot{H}_{\mathrm{B}} + \omega_{\mathrm{EB}}\, H \tag{10.98}$$

The case of the gyroscope can be mounted on a platform so that it can rotate relative to the platform; then:

$$\dot{H}_{\mathrm{B}} = \dot{H}_{\mathrm{C}} + \omega_{\mathrm{BC}}\, H \tag{10.99}$$

Finally, the inner gimbal can rotate relative to the case, hence:

$$\dot{H}_{\mathrm{C}} = \dot{H}_{\mathrm{G}} + \omega_{\mathrm{GC}}\, H \tag{10.100}$$

Substituting Equations 10.97 to 10.100 into Equation 10.96 yields:

$$T = \dot{H}_{\mathrm{G}} \left(\omega_{\mathrm{GC}} + \omega_{\mathrm{BC}} + \omega_{\mathrm{EB}} + \omega_{\mathrm{IE}} \right) H \tag{10.101}$$

© 1999 by CRC Press LLC

But:

$$\left(\omega_{GC} + \omega_{BC} + \omega_{EB} + \omega_{IE}\right) = \omega_{IG} \tag{10.102}$$

Therefore,

$$T = \dot{H}_G + \omega_{IG}\, H \tag{10.103}$$

By carefully constructing the gyroscope and maintaining the spin velocity of the rotor constant, H_G can be made to be zero. Thus, the law of gyroscopes can be written as:

$$T = \omega_{IG}\, H \tag{10.104}$$

This means that if an external torque T is applied to the gyroscope, the inner gimbal will precess with respect to the inertial frame with a velocity ω such that Equation 10.104 is satisfied.

In most designs (e.g., rate gyros), the gimbal is hermetically sealed in a liquid and liquid is floated in the case, to unload the gimbal bearings and to provide viscous damping. A pick-off senses gimbal deflection by means of position transducers and it controls a servo system, with a servomotor driving the case to maintain pick-off null.

Optical gyroscopes are based on the inertial properties of light instead of Newton's law of motion. They operate on the Sagnac effect, which produces interferometer fringe shift against the rotation rate. In this case, two light waves circulate in opposite directions around a path of radius R, beginning at source S. A typical arrangement for the illustration of operation principles is shown in Figure 10.22. When the gyro is stationary, the two beams arrive at the detector at the same time and no phase difference will be recorded. Assume that the source is rotating with a velocity ω so that light traveling in the opposite direction to rotation returns to the source sooner than that traveling in the same direction. Thus, any rotation of the system about the spin axis causes the distance covered by the beam traveling in the direction of rotation to lengthen, and the distance traveled by the beam in the opposite direction to shorten. The two beams interfere to form a fringe pattern and the fringe position may be recorded, or the phase differences of the two beams may be sensed. This phase difference is directional and proportional to the angular velocity. Usually, photodetectors are used to measure the phase shift.

Two different types of optical gyros can be categorized: either passive or active, and resonant or nonresonant. In passive gyrosensors, the Sagnac phase is measured by some external means; whereas in active gyros, the Sagnac phase causes a frequency change internal to the gyro that is directly proportional to the rotation rate.

The Sagnac interferometer is the basis of the *interferometric fiber-optic gyro* (IFOG). A typical fiber-optic gyroscope is shown in Figure 10.22. However, the most widely used gyro is the active resonant *ring laser gyro* (RLG), which is applied extensively in aircraft navigation. Two different types of resonant passive gyros, the *resonant fiber-optic gyro* (RFOG) and the *micro-optic gyro* (MOG), are lower cost devices commonly used and comparable to RLGs.

Accelerometers

In inertial navigation, the absolute acceleration is measured in terms of three mutually perpendicular components of the total acceleration vector. Integrating these acceleration signals twice gives the displacement from an initial known starting location. Details of the acceleration and accelerometers are given elsewhere in this book (see Acceleration, Vibration, and Shock). Accelerometers are made from three basic elements: proof mass, suspension mechanism, and pick-off mechanism. Some accelerometers require electric or magnetic force generators and appropriate servo loops. Accelerometers measure not only real vehicular acceleration, but also respond to gravitational reaction forces. Acceleration due to gravity is a function of position — in particular, latitude and altitude — and is compensated by computers.

© 1999 by CRC Press LLC

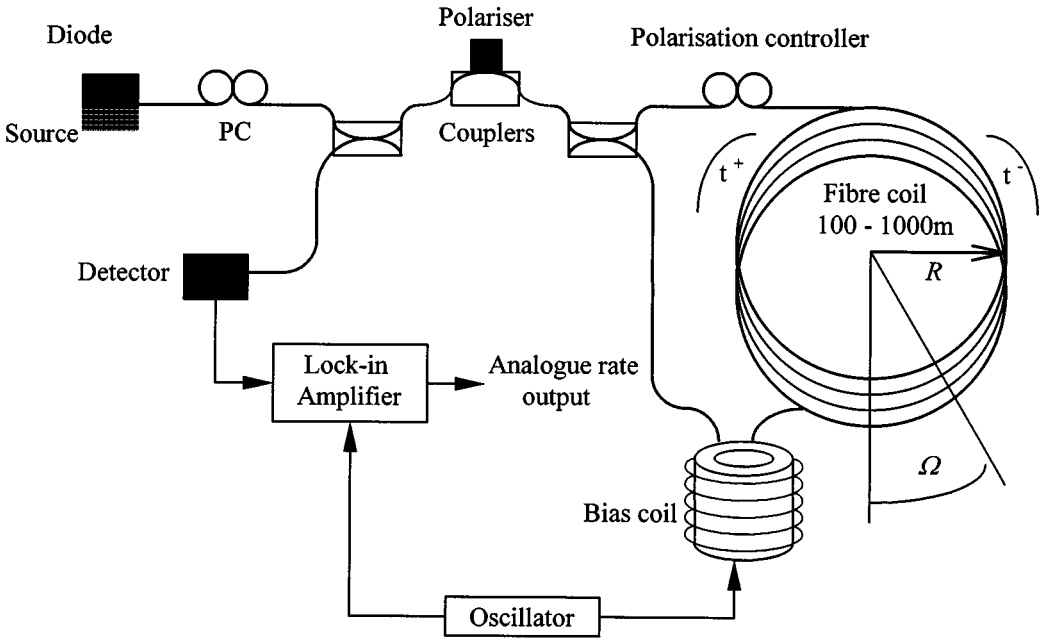

FIGURE 10.22 A typical fiber-optic gyroscope. This gyroscope is based on the inertial properties of light, making use of the Sagnac effect. The Sagnac effect describes interferometer fringe shift against rotation rate. Two light waves circulate in opposite directions around a path of radius *R*, beginning at source *S*. When the gyro is stationary, the two beams arrive at the detector at the same time and no phase difference is recorded. If the optical path is rotating with a velocity, the light traveling in the opposite direction to rotation returns to the source sooner than that traveling in the same direction. The two beams interfere to form a fringe pattern and the fringe position may be recorded, or the phase differences of the two beams may be sensed by photodetectors.

The most commonly used accelerometer in navigation systems is based on pendulous types. These accelerometers can be classified as:

1. Generic pendulous accelerometer
2. Q-flex type accelerometers
3. Micromachined accelerometers (A typical example of a modern micromachined accelerometer is given in Figure 10.23.)

Accelerations in the three axes are measured by suitably positioned accelerometers. Since accelerometers contain errors, the readings must be compensated by removing fixed biases or by applying scaling factors. The errors may be functions of operating temperature, vibration, or shock. Measurement of time must be precise as it is squared within the integration process for position determination. The Earth's rotation must also be considered and gravitational effects must be compensated appropriately.

Errors and Stabilization

Errors

In general, inertial navigation is an initial value process in which the location of the navigating object is determined by adding distances moved in known directions. Any errors in the system cause misrepresentation of the desired location by being off-target. The major disadvantage of an inertial guidance system is that its errors tend to grow with time. These errors in the deduced location are due to a number of reasons, including: imperfect knowledge of the starting conditions, errors in computation, and mainly errors generated by gyros and accelerometers.

© 1999 by CRC Press LLC

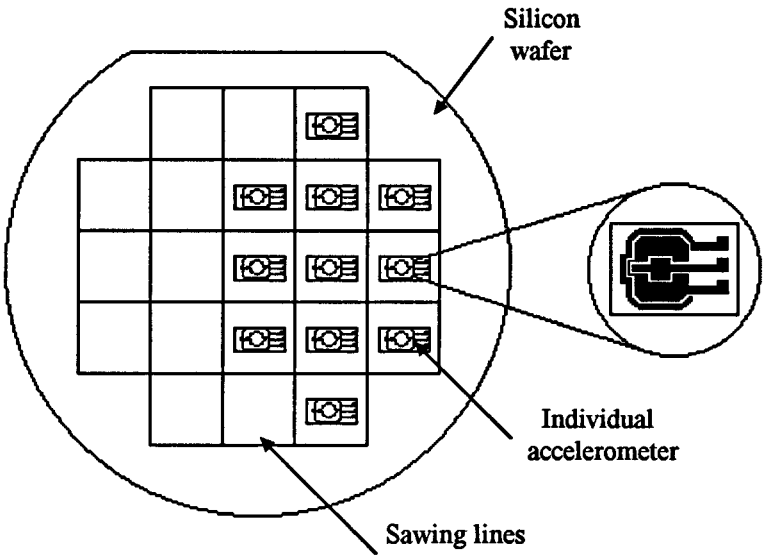

FIGURE 10.23 A typical example of a modern micromachined accelerometer. Multiple accelerometers can be mounted on a single chip, sensing accelerations in the *x*, *y*, and *z* directions. The primary signal conditioning is also provided in the same chip. The output from the chip is usually read in digital form.

If the error build-up with time becomes too large, external aids (e.g., LORAN, OMEGA) may be used to reset or update the system. Optimal use of the data from external aids must account for the geometry of the update and also for the accuracy of the update relative to the accuracy of the inertial system. The Kalman filter, for example, is one of the computational procedures frequently applied for optimally combining data from different sources.

Errors can broadly be classified as:

1. *System heading error:* A misalignment angle in the heading of an object traveling with a velocity can cause serious errors. For example, a vehicle traveling with velocity of 500 km h^{-1} in the same direction with 0.1° initial heading error will be off the target by approximately 873 m at the end of 1 h travel.
2. *Scale error:* Error in scaling can accumulate. In order to minimize scale errors, a scale factor is used. The scale factor is the ratio between changes in the input and output signals. It simply translates the gyro output (counts per second in the case of RLG) into a corresponding angle rotation. Some instruments may have different scale factors for positive and negative inputs, known as *scale factor asymmetry.* (Scale factors are measured in $^\circ\,h^{-1}\,mA^{-1}$, $^\circ\,h^{-1}\,Hz^{-1}$, or $g\,Hz^{-1}$.)
3. *Nonlinearity and composite errors:* In most cases, scale factors are not constant, but they can have second- or higher-order terms relating the output signals to the input. Statistical techniques can be employed to minimize these errors.
4. *Bias errors:* Zero offset or bias error is due to existence of some level of output signal for a zero input. Bias errors exist in accelerometers, gyros, tilt misalignments, etc.
5. *Random drift and random walk errors:* In some cases, the outputs of the devices can change due to disturbances inside the sensors, such as ball bearing noise in mechanical gyros. These disturbances may be related to temperature changes, aging, etc. White noise in optical gyros can cause a long-term accumulation in angle error known as the *random walk.*
6. *Dead band, threshold, resolution, and hysteresis errors:* These errors can be related to inherent operation of accelerometers and gyros. They can be due to stiction, minimum input required for an output, minimum measurable outputs, and nonrepeatability of variations in the output versus variations in the input.

© 1999 by CRC Press LLC

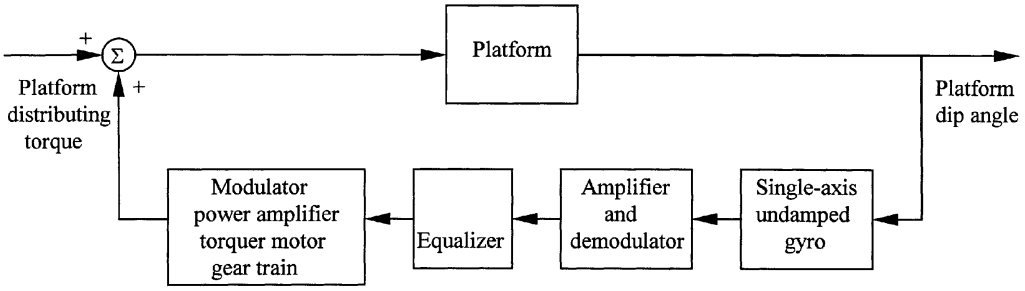

FIGURE 10.24 Stabilization is obtained using platforms designed to accurately maintain accelerometers and gyros leveled and oriented in the azimuth direction. In some cases, the platform is driven around its axis by servo amplifiers and electric motors. Sensitive pick-offs on the gyroscopes fed error signals are used to maintain a desired stability of the platform in the presence of disturbing torques.

It should be pointed out that this list is by no means exhaustive. Detailed error analysis can be found in the references cited.

Stabilization

The inertial navigation sensors must maintain angles within specified limits in spite of the disturbances imposed by the moving object. Accuracy requirements demand that the system must provide reliable and stable information in spite vibrations and other disturbing factors. One way of achieving stabilization is by using a stabilized platform. These platforms are designed to maintain accelerometers and gyros accurately leveled and oriented in the azimuth direction. In some cases, the platform is driven around its axis by servo amplifiers and electric motors. Usually, outputs of doubly integrating accelerometers are used directly to control the level-axis gyroscope precession rates. Sensitive pick-offs on the gyroscopes fed error signals are used to maintain a desired stable platform in the face of disturbing torques. The operation of a typical system, in block diagram form, is shown in Figure 10.24.

Unlike platform models, in a strapped-down system, gyroscopes and accelerometers are rigidly mounted to the vehicle structure so that they move with the vehicle. The accelerometers and gyroscopes are manufactured to measure accelerations and angles up to the maximum expected values. As the vehicle travels, the measured values are frequently transmitted to a computer. The computer uses these values to resolve the readings into the navigation axis sets and make deductions on the body axis sets.

Vehicular Inertial Navigation

In modern vehicular navigation, computerized maps and mobile communication equipment are integrated together with inertial and/or other electronic navigation systems. In recent years, in the wake of low-cost GPS systems, the vehicular navigation system has attracted much attention due to its large potential markets for consumer as well as business vehicles.

Automobile navigation systems are based on dead-reckoning, map matching, satellite positioning, and other navigational technologies. Map intelligent systems achieve high relative accuracy by matching dead-reckoned paths with road geometry encoded in a computerized map. This is also used to perform other functions such as vehicle routing and geocoding. Satellite-based navigation systems achieve high absolute accuracy with the support of dead-reckoning augmentation.

The capabilities and functions of automobile navigation systems depend on:

Choosing the necessary technology
Integrating the overall system
Resolving driver interface
Providing map data basis
Coordinating mobile communications

© 1999 by CRC Press LLC

Digital maps and mobile data communications combine together for full usefulness and effectiveness. The navigation systems are greatly enhanced in conjunction with stored digital maps combined with effective communications.

The usefulness of a navigation system is related to the accuracy in position determination. There are a number of methods available with varying accuracy; these include the following:

Dead-reckoning

Dead-reckoning is the process of determining vehicle location relative to an initial position by integrating measured increments and directions of travel. The devices include the odometer, the differential odometer, and a magnetic compass. Gyros and inertial systems prove to have limited applications in harsh automotive environments. Although, dead-reckoning systems suffer from error accumulation, they are widely used inertial navigation systems, particularly in robotics and vehicular applications. Even the most precise navigation system requires periodic reinitialization and continuous calibrations by computers.

Radiolocation

In radiolocation, the global positioning system (GPS) is used extensively. Nevertheless, LORAN is gaining popularity as means of tracking land vehicle location from a central location. But its modest accuracy limits its global application in automotive navigation.

Map Matching

Artificial intelligence concepts are applied to match dead-reckoned vehicle paths, which are stored in computers. In map matching, sensed mathematical features of the vehicle paths are continuously associated with those encoded in a map database. Thus, a vehicle's dead-reckoned location can be initialized automatically at every turn to prevent accumulation of dead-reckoning errors.

The first application of map matching technology was in the Automatic Route Control System (ARCS), which used a differential odometer for dead-reckoning. In another system, the Etak map matching system, a solid-state flux gate compass is used as well as a differential odometer to dead-reckon paths for matching with digitized maps and aerial photographs. Further details on these technologies can be found in the references given at the end of this chapter.

In a map matching system, as long as the streets and road connectivities are accurately defined, the process identifies position relative to the road network as visually perceived by the vehicle driver.

Most of the dead-reckoning equipment commercially available is sufficiently robust to support map matching when operating in a defined road network. However, a good dead-reckoning accuracy is required to achieve reinitialization through map matching upon returning to the road network after off-road operations.

Proximity Beacon

This approach uses strategically located short-range transmitters, and reception of their location coded signal infers the receiving vehicle's instantaneous location. There are several variations of the proximity approach; some versions involve two-way communications with the equipped vehicle. Typically, the driver enters the destination code on the vehicle panel, for automatic transmission to the roadside unit, as the vehicle approaches an instrumented intersection. The roadside unit, which can be networked with a traffic management system, analyzes the destination code and transmits route instructions to the display on the vehicle panel. Proximity beacon systems are being tested in Germany and Japan. One of the most popular system is the ALI-SCOUT (see references) proximity beacon system, which uses dead-reckoning and map matching techniques between beacons to download updated map and traffic data in Berlin.

The approach to the interface between an on-board navigation system and a vehicle operator must take into account ergonomics and safety considerations as well as functional requirements. As a result of intensive research, especially in the aerospace industry, display of information for the operator is a well-developed area. In a well-known European system, Philips' CARIN, a color CRT map display is used to show vehicle location relative to the surroundings. Many other systems use short visual messages, symbolic graphics, and voice.

© 1999 by CRC Press LLC

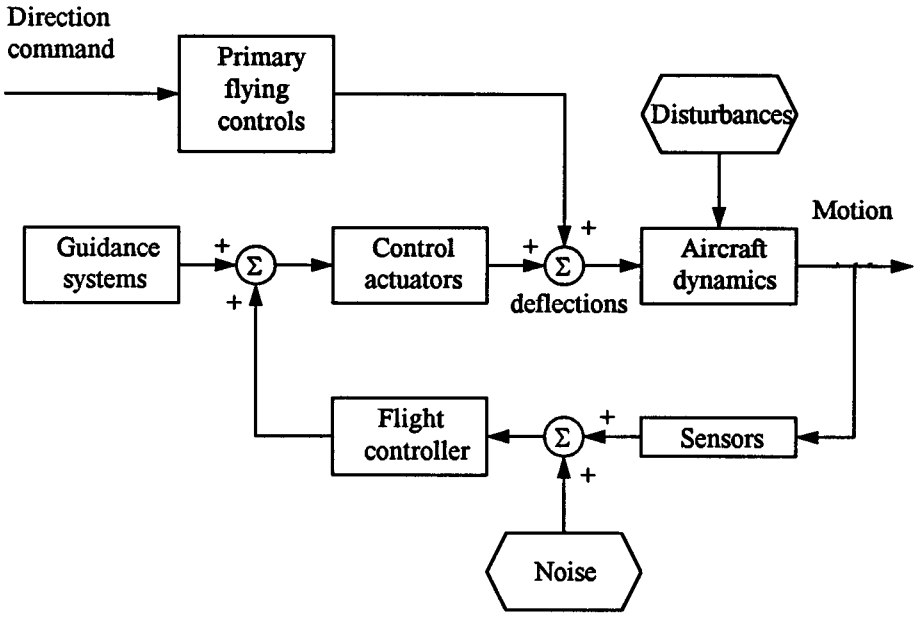

FIGURE 10.25 In many commercial aircraft, suitably located gyroscopes and accelerometers give signals for controlling the stability of the aircraft. Due to various instrumental errors and in-flight random disturbances such as gyro drift, scale factor errors, and accelerometer bias errors, the errors in the desired orientation increase with time. These accumulated errors need to be compensated periodically by external information such as Omega navigation systems.

Major potential roles for data communications in future automobile navigation need to provide current updates (road additions, closures, detours, etc.) for on-board map databases, and also provide real-time information on traffic conditions for on-board route generation.

Aircraft

The primary navigation aid for civil aircraft flying in the airspace of most of the developed countries is the VOR/DME system. The VOR (Very high frequency Omni Range) and the DME (Distance Measuring Equipment) enable on-board determination of an aircraft's bearing relative to North at the fixed ground station and slant range from the station, respectively. Further details can be found in references given at the end of this chapter.

Many commercial aircraft are equipped with precision Inertial Navigation Systems (INS) not only for navigation purposes, but also for stabilization of the aircraft at all times. Suitably located gyroscopes and accelerometers give signals to control the stability of the aircraft, as shown in Figure 10.25. Many aircraft INS utilize a gyro-stabilized platform on which the accelerometers are mounted. The platform is aligned before take-off to the desired orientation. Due to alignment errors and in-flight random disturbances such as gyro drift, scale factor errors, and accelerometer bias errors, the errors in the desired orientation increase with time. The errors need to be compensated periodically by external information such as Omega navigation systems.

Omega is a 10 kHz to 14 kHz navigation system that was primarily intended for updating submarine inertial navigators because of its very low frequencies penetrating beneath the ocean surface. It is also used by many general-aviation and military aircraft because of its worldwide coverage. Some airlines (e.g., American Airlines) have equipped their entire fleet with Omega receivers. The U.S. Coast Guard maintains two Omega stations and other countries maintain six more. Each station transmits eight consecutive sinusoidal tones with a 10 s repetition rate. Four are navigation tones common to all stations;

© 1999 by CRC Press LLC

the other four tones uniquely identify the station. Each station has a cesium clock that is calibrated within 1 μs by exchanging portable atomic clocks with other Omega stations and with the U.S. Naval Observatory.

GPS systems give all vehicles on or near the Earth unprecedented navigation accuracy. A number of international airlines are equipped with confined GPS-Glonoss receivers. Many experiments are now in progress to balance the cost versus accuracy of various combinations of inertial, Omega, Loran, GPS, and Transit equipment. Airborne receivers are designed that combine navaids operating in a common radio band (e.g., GPS, DME, JTID, and IFF).

INMARSAT's "Marec" communication satellites serve ship traffic but are configured to serve air traffic by directly communicating with aircraft for approximately $10 per call.

Underwater

The Ship's Inertial Navigational System (SINS) was originally developed for precision position-finding required by ballistic missile submarines in the late 1950s and early 1960s. The first deployment was on-board U.S. George Washington in 1960, and SINS are used today in submarines, aircraft carriers, and other surface warships. As the cost and size are continually decreasing, the system is also deployed in naval as well as merchant vessels. Another development of INS for underwater application is in the area of the autonomous underwater vehicle (AUV). In this section, a number of such products are described.

AUVs are used extensively for military and civilian purposes. Application examples are mapping, surveillance, ocean exploration, survey, and mining, all of which require precise position determination. The desired features of such systems are: low power, high accuracy, small volume, light weight, and low cost. Two typical examples are the LN family produced by Litton Guidance and Control Systems, and the system developed by the Harbor Branch Oceanographic Institution Inc. (HBOI). The specifications of some of these systems are briefly described below to give examples of underwater INS.

The Litton LN-100 System

Litton's LN-100 is an example of the strapdown INS. The LN-100 system consists of three Litton Zero-Lock Gyros (ZLG), a Litton A4 accelerometer triad, power supply, supporting electronics, and a JIWAG standard 80960 computer. The single-board computer performs all the control, navigation, and interface functions.

The HBOI System

The HOBI system was developed with Kearfott Guidance and Navigation (KGN) and utilizes a Monolithic Ring Laser Gyroscope (MRLG), motion sensors, GPS input, and control unit. The inertial measurement unit is based on the Kearfott's T16-B three-axis ring laser gyro and three accelerometers.

Robotics

Closely related to the autonomous underwater vehicles, autonomous mobile robots also use INS extensively as a self-contained, independent navigation system. Typical applications are mining, unknown terrain exploration, and off-line path planning. There are many commercially available inertial navigation systems suitable for cost-effective utilization in the navigation of robots. Some of these are: gyrocompasses, rate gyros, gyrochip, piezoelectric vibrating gyros, ring laser gyros, interferometric, and other types of fiber-optic gyros. Three popular systems will be explained here.

The Honeywell Modular Azimuth Position System (MAPS)

Honeywell's H-726 Modular Azimuth Position System (MAPS) is a typical example of an inertial navigation system for land-based vehicles. It consists of a Dynamic Reference Unit (DRU) that provides processed information from the inertial sensor assembly, a Control Display Unit (CDU) that is used for human-machine interface, and a Vehicle Motion Sensor (VMS) that monitors the vehicle's directional and distance information. The inertial sensor assembly comprises three Honeywell GG1342 ring-laser

© 1999 by CRC Press LLC

gyros and three Sundstrand QA2000 accelerometers mounted to measure movements in three local axes. The inertial processor translates the information to the navigation processor that resolves the vehicle movement information from the VMS. The CDU provides mode selection, data display, waypoint information, and general operator interface.

Hitachi Fiber-Optic Gyroscopes

The Hitachi products are significant as they are relatively inexpensive and were designed for automotive applications. The open-loop Interferometric Fiber-Optic Gyros (IFOG) HOFG-4FT received the "Most Technologically Significant New Products of the Year" award in 1993, and is now installed in one of the production models from Toyota. The subsequent models of IFOG are HOFG-X, HOFG-1, and HGA-D. The HOFG-1 has been employed extensively in industrial mobile robots. The output of the system can be in serial form in RS-232 standard or as analog signals. Specifications of HOFG-1 include a range of $\pm 60^\circ\ s^{-1}$, an update rate of 15 ms, and linearity of $\pm 0.5\%$. The power requirement is 10 to 16 V dc and 0.5 A.

References

K. R. Britting, *Inertial Navigation Systems Analysis,* New York: Wiley-Interscience, 1971.

M. Kayton, *Navigation-Land, Sea, Air and Space,* New York: IEEE Press, 1990.

A. Lawrance, *Modern Inertial Technology — Navigation, Guidance, and Control,* New York: Springer-Verlag, 1993.

Appendix — List of Manufacturers/Suppliers

Aerodata Co.
5550 Merric Rd.
Suit 205
Massapequa, NY
Tel: (616) 798-1873

AGV Products, Inc.
9307-E Monroe Rd.
Charlotte, NC 28270
Tel: (704) 825-1110
Fax: (704) 825-1111

American GNC Corporation
9131 Mason Avenue
Chatsworth, CA 91311
Tel: (818) 407-0092
Fax: (818) 407-0093
e-mail: agnc@kincyb.com

Astronautics Co.
P.O. Box 523
518 W Cherry St.
Milwaukee, WI

Cybermotion, Inc.
115 Sheraton Drive
Salem, VA 24153
Tel: (703) 562-7626
Fax: (703) 562-7632

First State Map Co.
12 Marry Ella Dr.
Wilmington, CT 19805
Tel: (800) 327-7992

Hitachi Cable America, Inc.
50 Main Street,
White Plains, NY 10606-1920
Tel: (914) 993-0990
Fax: (914) 993-0997

Honeywell, Inc.
Technology Center
3660 Technology Drive
Minnepolis, MN 55418
Tel: (612) 951-7715
Fax: (612) 951-7438

Ketema Inc.
790-T Greenfield Dr.
P.O. Box 666
El Cajon, CA

NASA Goddard Space Flight Center
Robotics Branch, Code 714.1
Greenbelt, MD 20771
Tel: (301) 286-4031
Fax: (301) 286-1613

Naval Command Control Center
RDT&E Division 5303
San Diego, CA 92152-7383
Tel: (619) 553-3672
Fax: (619) 553-6188

Navigation Science Co.
31127 Via Clinas
Suite 807
Westlake Village, CA 91326
Tel: (818) 991-9794
Fax: (818) 991-9896

Navigations Technologies Corp.
740 Arques Avenue
Sunnyvale, CA 94086
Tel: (408) 737-3200
Fax: (408) 737-3280

NSC
P. O. Box 4453
Thousand Oaks, CA 91359
Tel: (818) 991-9794
Fax: (818) 991-9896

Romarc Co.
512 Scott Rd.
Plumsteadville, PA 18949
Tel: (800) 755-2572

© 1999 by CRC Press LLC

Schwartz Electro-Optics, Inc.
3404 N. Orange Blossom Trail
Orlando, FL 32804
Tel: (407) 298-1802
Fax: (407) 297-1794

Siemens Co.
1301 Avenue of the Americas
New York, NY 10019

Southern Avionics Co.
5000-T Belmont
Beaumont, TX 77707
Tel: (800) 280-0322
Fax: (409) 842-2987

Sperry Marine Inc.
1070 T Seminole Tr.
Charlottesville, VA 22901
Tel: (804) 974-2000
Fax: (804) 974-2259

Systron Donner Inertial Division
BEI Electronics
2700 Systron Drive
Concord, CA 94518-1399
Tel: (510) 682-6161
Fax: (510) 671-6590

Trackor Inc.
6500-T Trackor Lane
Austin, TX

Warren-Knight Inst. Co.
2045 Bennet Dr.
Philadelphia, PA 19116
Tel: (215) 484-9300

10.4 Satellite Navigation and Radiolocation

Halit Eren and C. C. Fung

Modern electronic navigation systems can be classified by range, scope, error, and cost. The range classifications are short, medium, and long ranges, within which exact limits are rather indefinite. The scope classifications can be either self-contained or externally supported, and active (transmitting) or passive (not transmitting) mode of operation.

Short-range systems include radiobeacons, radar, and Decca. Medium-range systems include Decca and certain types of extended-range radars. The long-range systems include Loran-C, Consol, and Omega. All these systems depend on active radio frequency (RF) transmissions, and all are externally supported with respect to the object being navigated, with the exception of the radar. In addition to these, there is another category of systems which are called *advanced navigation systems*; the transit satellite navigation systems, Glonass, and the Global Positioning Systems (GPS) are typical examples.

Utilization of electromagnetic radio waves is common to all navigation systems discussed here. Understanding of their behavior in the Earth's atmosphere is very important in the design, construction, and use of all kinds of navigation equipment — from satellites to simple hand-held receivers.

When an FM radio wave is generated within the Earth's atmosphere, the wave travels outward. The waves may be absorbed or reflected from surfaces of materials they encounter. The absorption and scattering of electromagnetic waves take place for many reasons, one of which is caused by excitation of electrons within the molecules in the propagation media. The behavior of an electromagnetic wave is dependent on its frequency and corresponding wavelength. Figure 10.26 shows the frequency spectrum of electromagnetic waves. They are classified as *audible waves* at the lower end of the spectrum, *radio waves* from 5 kHz to 300 GHz, and *visible light* and various other types of rays at the upper end of the spectrum.

For practical purposes, the radio wave spectrum is broken into eight bands of frequencies; these are: *very low frequency* (VFL) less than 30 kHz, *low frequency* (LF) 30 kHz to 300 kHz, *medium frequency* (MF) 300 kHz to 3 MHz, *high frequency* (HF) 3 MHz to 30 MHz, *very high frequency* (VHF) 30 MHz to 300 MHz, *ultra high frequency* (UHF) 300 MHz to 3 GHz, *super high frequency* (SHF) 3 GHz to 30 GHz, and *extremely high frequency* (EHF) 30 GHz to 300 GHz.

For easy identification, the frequencies above 1 GHz are further broken down by letter designators, as: L-band (1–2 GHz), S-band (2–4 GHz), C-band (4–8 GHz), X-band (8–12.5 GHz), and K-band (12.5–40 GHz). Since severe absorption of radar waves occurs near the resonant frequency of water vapor at 22.2 GHz, the K-band is subdivided into lower K-band (12.5–18 GHz) and upper K-band (26.5–40 GHz). Most navigation radars operate in the X- and S-bands, and many weapons fire control radars operate in the K-band range.

© 1999 by CRC Press LLC

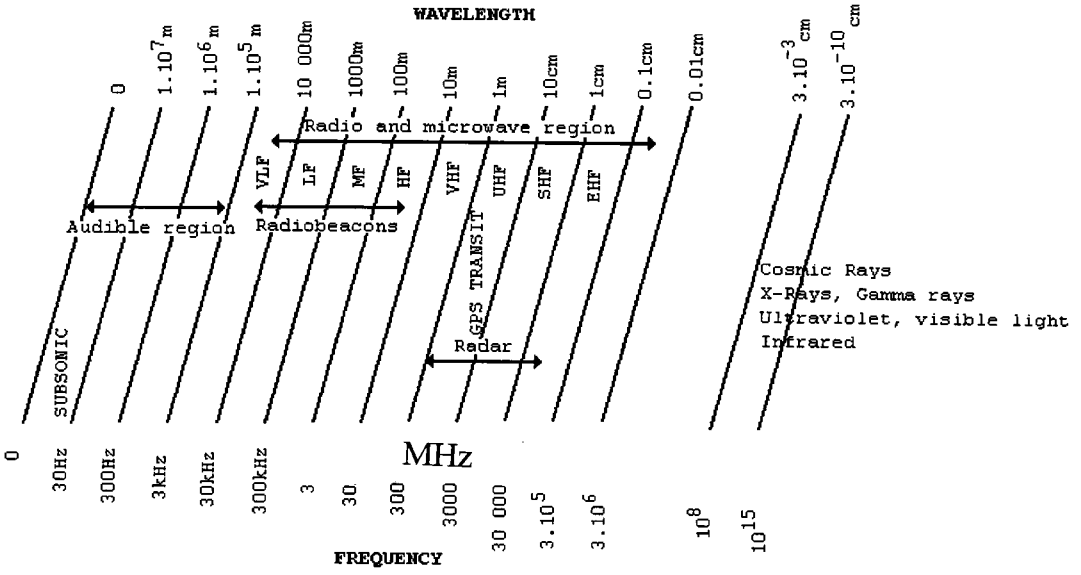

FIGURE 10.26 Electromagnetic wave frequency spectrum. Audible range can be heard if converted to sound waves. Radiobeacons operate in the VLF, LF, and MF ranges. Omega operating at VLF covers the entire world with only eight transmission stations. GPS, Transit, and Glonass use UHF frequencies. Wavelengths less than 10 cm are not suitable for satellite systems, but they are used in radars.

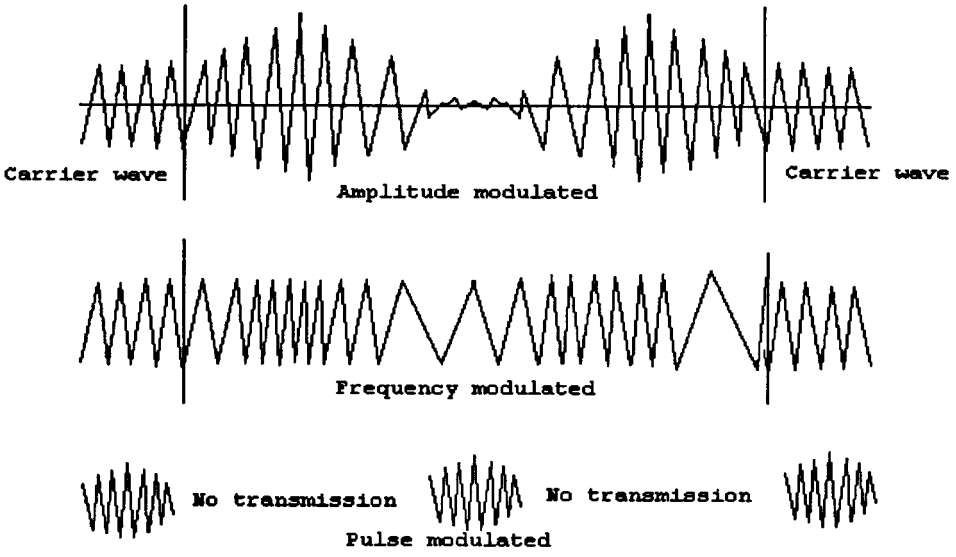

FIGURE 10.27 Amplitude, frequency, and pulse modulation of RF carrier waves. Amplitude modulation is suitable for broadcasting radio stations. Frequency modulation is used in commercial radio broadcasts. The pulse modulation is used in satellite systems, radars, and long-range navigation aids.

The radio waves are transmitted as continuous or modulated waves. A carrier wave (CW) is *modulated* to convey information in three basic forms: amplitude, frequency, and pulse modulation, as shown in Figure 10.27. The amplitude modulation (AM) modifies the amplitude of the carrier wave with a modulating signal. In frequency modulation (FM), the frequency of the carrier wave is altered in accordance with the frequency of the modulating wave. FM is used in commercial radio broadcasts and the sound

© 1999 by CRC Press LLC

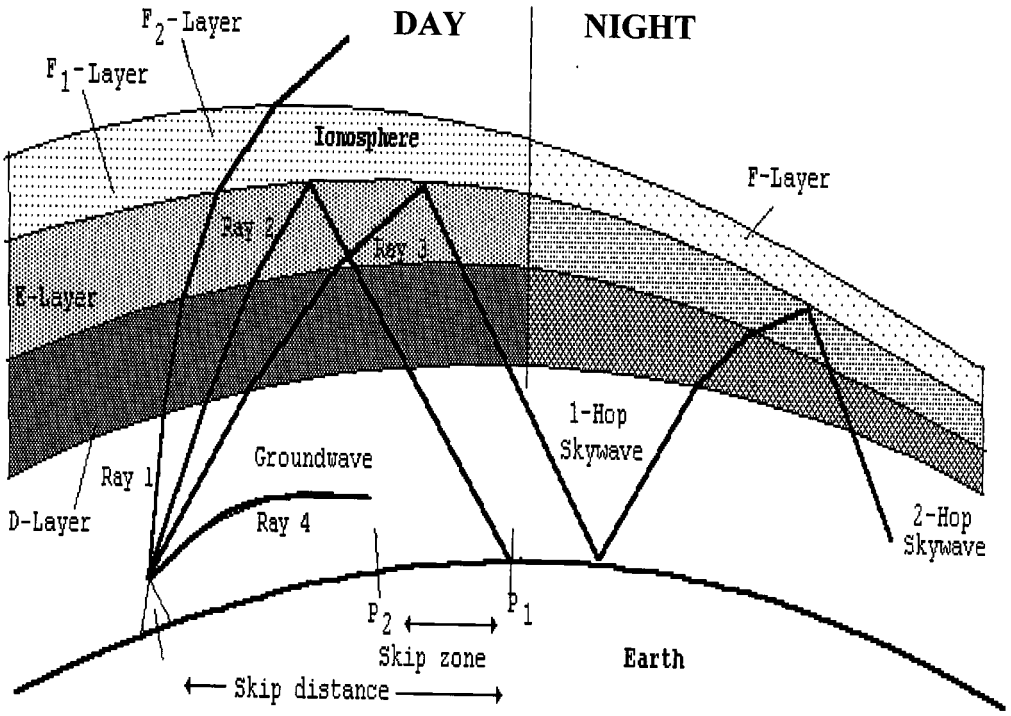

FIGURE 10.28 The four layers of the ionosphere and its effect on radio propagation. The four layers are produced by the ionization of molecules of particles in the atmosphere by ultraviolet rays of sun. The effect of ionosphere on the radio waves is shown by reflections, also termed hops. The frequency of the electromagnetic wave is important in its behavior through ionosphere.

portion of the television broadcast. Pulse modulation is different from AM and FM in that there is usually no impressed modulation wave employed. In this modulation, the continuous wave is broken up into very short bursts or "pulses," separated by periods of silence during which no wave is transmitted. This is used in satellite navigation systems, surface search radars, and long-range radio navigation aids such as Loran.

When an electromagnetic wave encounters an obstruction, *diffraction* takes place marked by a weak reception of the signal within the "shadow" zone. Two waves acting on the same point will also result in *interference*. The degree of interference depends on the phase and frequency relationship. For example, two waves of the same frequency with a 180° phase difference will result in a null at that point. Also, under certain conditions, a portion of the electromagnetic energy in radio waves may reflect back toward the Earth's surface to form the *ionosphere*. The ionosphere is a layer of charged particles located about 90 to 400 km high from Earth's surface; such reflected waves are called *sky waves*.

In the study of radio wave propagation, there are four *ionosphere layers* of importance, as shown in Figure 10.28. The D-layer is located about 60 km to 90 km and is formed during daylight. The E-layer is about 110 km. It persists through the night with decreased intensity. The F_1-layer is between 175 km and 200 km; it occurs only during daylight. The F_2-layer is between 250 km and 400 km; its strength is greatest in the day but it combines with the F_1-layer later to form a weak F-layer after dark. The layers in the ionosphere are variable, with the pattern seeming to have diurnal, seasonal, and sun spot periods. The layers may be highly conductive or may entirely hinder transmissions, depending on the frequency of the wave, its angle of incidence, height, and intensity on various layers at the time of transmission. In general, frequencies in the MF and HF bands are most suitable for ionosphere reflections during both day and night.

© 1999 by CRC Press LLC

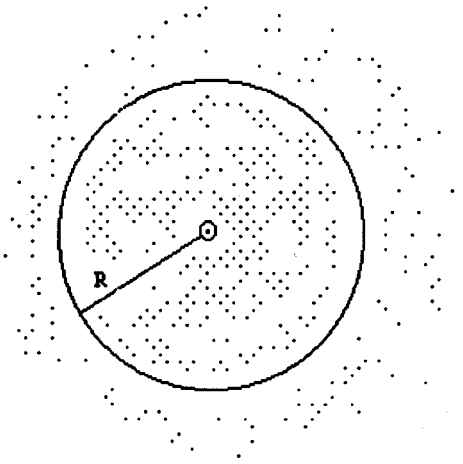

FIGURE 10.29 The rms radius circle that encompasses 68% of all measured positions. The variations in the measurements are due to a number of factors, including: ionosphere conditions, precise location of satellites, and inefficiencies in electronic circuits. (2 rms encompasses 95% of all indicated positions.)

Because of the higher resistance of the Earth's crust as compared to the atmosphere, the lower portions of radio waves parallel to the Earth's surface are slowed down, causing the waves to bend toward Earth. A wave of this type is termed a *ground wave.* The ground waves exist because they use the Earth's surface as a conductor. They occur at low frequency since LF causes more bending in conformity to Earth's shape. The ultimate range of such ground waves depends on the absorption effects. Sometimes, in the lower atmosphere, *surface ducting* occurs by multiple hopping, thus extending the range of a ground wave well beyond its normal limits. It is associated with higher radio and radar frequencies. This phenomenon is common in tropical latitudes. Behavior patterns of waves transmitted at various angles are illustrated in Figure 10.28.

Accuracy of Electronic Fix

There are a number of random effects that influence the accuracy of an electronic position determination; atmospheric disturbances along the transmission path, errors in transmitters and receivers, clocks, inaccuracy in electronic circuitry, gyro errors, etc. As a result, a series of positions determined at a given time and location usually results in a cluster of points near the true position. There are two measures commonly used to describe the accuracy: the first is the *circular error probable* (CEP) — a circle drawn on the true position whose circumference encompasses 50% of all indicated positions, and the second technique, more common, is the *root mean square* (rms), where:

$$\text{rms} = \sqrt{\sum_{n=1}^{N} \left(E_n\right)^2 \Big/ N} \tag{10.105}$$

where E = the distance between actual and predicted positions
N = the number of predicted positions

A circle, shown in Figure 10.29, with one rms value is expected to contain 68% of all the indicated positions. Another circle of radius equal to 2 rms should contain 95% of all the indicated positions, for isotropic scattering, or errors.

© 1999 by CRC Press LLC

In electronic navigation systems, three types of accuracy are important: (1) *predictable* or *absolute accuracy* — the accuracy of a position with respect to the geographic coordinates of the Earth; (2) *repeatable accuracy* — the accuracy with which the user can return to a position whose coordinates have been determined at a previous time with the same navigation system; and (3) *relative accuracy* — the accuracy with which a user can measure position relative to that of another user of the same system at the same time.

Radionavigation Systems

In the 1930s, improved radio direction-finding techniques and better equipment led to the establishment of systems called *radiobeacons*. These systems consisted of small radio transmitters located in fixed places to provide radio bearings that could be used in all weather conditions. Position findings by these beacons became known as *radionavigation*. Continued refinements in the state-of-the-art electronics technology and a better understanding of electromagnetic wave propagation led to the subsequent development of radar and longer-range radionavigation systems.

Essentially, radiobeacons are single transmitters, transmitting a continuous wave at low power, usually modulated by audible Morse code characters for identification. The transmitted signal is received by an on-board receiver incorporating a radio direction finder (RDF) to be processed further.

Short- and Medium-Range Radiolocation Systems

Most short- to medium-range navigation systems are designed to provide either a bearing to a fixed transmitter site, as in the case of radiobeacons, or a range and bearing from the transmitter to a natural or manufactured navigation aid, as in the case of radar.

Long-Range Hyperbolic Navigation Systems

Long-range electronic navigation systems are based on *hyperbolic systems*. In these systems, the lines of positions yield in segments of hyperbolas rather than as radial lines. The line connecting the master and secondary stations transmitting the same signal simultaneously is called the *baseline*. Hyperbola lines represent the locus of all points of the arrival of master and secondary pulses at specified time differences. This is illustrated in Figure 10.30. Any change in the position of the receiver near the baseline corresponds to a relatively large change in time difference of reception of pulses.

In practice, the secondary stations transmit pulses at fixed intervals, called *coding delays*, after having received a pulse from the master station. Currently, hyperbolic systems employ atomic time standards to regulate both master and secondary station transmissions to increase the system accuracy by eliminating random time errors due to atmospheric conditions.

In some systems, as an alternative to short pulses, continuous waves are propagated with the leading edges of each cycle representing a given time and distance interval. The corresponding hyperbolas then represent loci of constant phase differences, called *isophase lines*. The space between the hyperbolas are referred to as *lanes*. The position of the receiver within a lane is determined by the phase difference between the master and secondary signals. A disadvantage of this system is that it is not possible to distinguish one lane from another by the use of phase comparison alone. Hence, the lanes must be either counted as they are traversed from some fixed position, or they must be identified by some other means. For good accuracy, the user's distance from the stations can seldom exceed about six times the length of the baseline.

Radiobeacons

Most radiobeacon signals are limited to less than 320 km (200 miles), with a majority not receivable beyond about 32 km (20 miles). Often, radiobeacons located in a given area are grouped on a common frequency, such that each transmitter transmits only during a segment of a time-sharing plan. Radio bearings to the site of transmitter are determined by the use of a radio receiver equipped with a Radio Direction Finder (RDF). There are moderately priced, manually operated RDF receivers, and several more expensive fully automated models are also available. As a general rule, RDF bearings are normally

© 1999 by CRC Press LLC

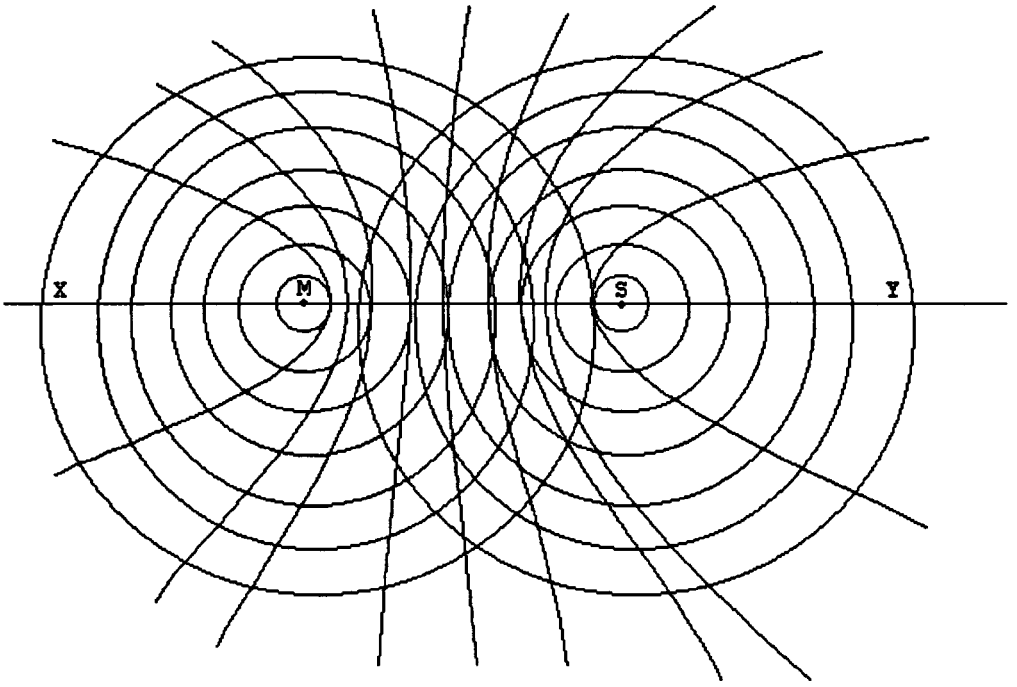

FIGURE 10.30 Hyperbolic patterns of two interacting radio waves propagated in opposite directions. These lines represent the locus of all points of a specified time difference between master and secondary pulses. The straight line MS is called the baseline. The maximum distance of the target object should not exceed 6 times the length of the baseline.

considered accurate only to within ±2°; for example, under 200 km (120 miles) to the transmitter in favorable conditions and ±5° to 10° when conditions are unfavorable.

Information on the locations, ranges, and using procedures of radio beacons are given in a number of publications such as DMAHTC Publication No. 117 *The Radio Navigation Aids.* Correct radiobeacon bearings are plotted on Mercator charts for position estimation. Because of possible inaccuracies, radiobeacons are not used universally. Navigators such as small boats and merchant ships not equipped with other systems use radiobeacons.

Loran-C

Loran was developed in the 1940s to be one of the first systems implementing a long-range hyperbolic system for both ships and aircraft. The system used master and slave stations transmitting sequential radio waves in the upper MF band with frequencies between 1850 kHz and 1950 kHz. Loran-A featured ground wave coverage out to between 700 km and 1250 km from the baseline by day, and up to 2200 km by night. It was the most widely used electronic navigation system until 1970. Later, a system employing synchronized pulses for both time-difference and phase comparison measurements was developed, known as Loran-C. Loran-C was configured to operate in a *chain* form consisting of more than one slave station usually located in triangles.

All stations in the system transmit a signal on a common carrier frequency in mid-LF band of 100 kHz ± 10 kHz. The ground wave range is about 1900 km. One-hop sky waves have a range of about 3600 km, and two-hop signals were noted to have been received about 6250 km from the ground station. One-hop sky waves are produced both by day and by night, while two-hop sky waves are formed only at night. Present Loran-C chains have baseline distances between 1500 km and 2300 km. The accuracy of the system varies from about ±200 m rms near the baseline to ±600 m rms near the extreme ranges

© 1999 by CRC Press LLC

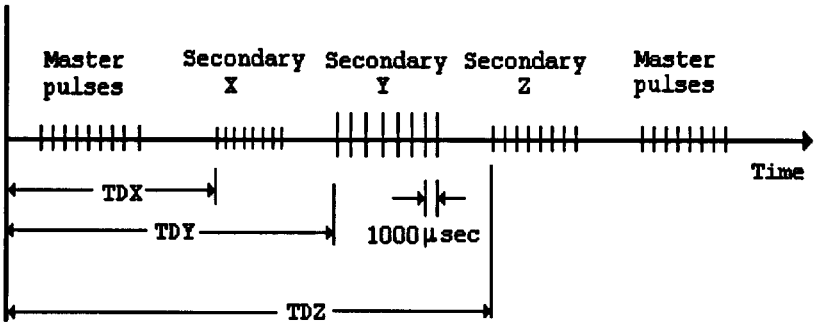

FIGURE 10.31 Loran-C pulse sequence. The nine pulses from the master station are separated by 1 ms intervals, except the ninth one, which has a 2 ms interval. The secondary stations X, Y, and Z transmit pulses some time after they receive the information from the master station. The receivers picking up these pulses provide direct latitude and longitude information, usually by a digital readout.

of the system. Differential techniques employed in recent years have increased the accuracy. The low frequency and high-power transmitters allow ground waves to penetrate the surface layer of the sea water, enabling submerged submarines to receive them.

To lessen the large power requirements, *multipulsed* signals are used. The multipulsed signals of each master station and its associated secondary stations are transmitted in a predetermined sequence, as shown in Figure 10.31. Eight of the nine pulses are separated by 1 ms intervals and the ninth one by a 2-ms interval. Integrated master and secondary pulses are compared at a sampling point at exactly 30 μs from their leading edges.

Loran-C receivers are fully automatic, suitable to be employed for marine, land vehicle, and aircraft applications. Most receivers provide direct lat/long digital readout, precise to a tenth of a minute of arc. They also provide auxiliary features such as destination selection, course and speed overground, etc. Once initialized, they automatically select the best chain and the most suitable master/secondary pulses.

There are 13 Loran-C chains worldwide. Each chain uses a different basic pulse repetition rate (PRR) and pulse repetition interval (PRI). The specific PRI used in a given chain is referred to as group repetition interval (GRI), often called *rate*. This system has enjoyed great expansion since 1970s, attracting many users. It has found applications in ships, aircraft, as well as land vehicles. Land vehicles are equipped with automatic vehicle location systems (AVLS). In one application in the U.S., the AVLS system is integrated with the city's emergency telephone system, a computer-aided dispatch system to control the position of emergency vehicles such as ambulances and fire trucks.

Nevertheless, the progress made in satellite-based global positioning systems (GPS) calls for termination of Loran-C in the very near future. However, complete system shut-down may not occur immediately, due to the vast number of current users of the system.

Decca

Decca is similar to Loran-C in that each chain is composed of one master and three slave stations arranged in *star* pattern, at an angle of about 120° between the baselines. It uses unmodulated continuous waves rather than the pulsed waves of Loran. The characteristic hyperbolic grid pattern is formed by phase comparisons of master and slave signals. All stations transmit in the LF-band between 70 kHz and 130 kHz. The nominal range is about 400 km both by day and by night. The system is extremely accurate within the operating range. The signals transmitted by each of the stations are all harmonics of a single fundamental frequency.

There are a wide variety of receivers available for Decca system, including automatic flight logs for aircraft. In general, the systems consist of four separate receivers, each of which can be set to receive one of the four signals transmitted by a given chain. The lane identification is accomplished by a signal transmitted by each master and slave station, transmitted once every 20 s for a duration of 0.6 s. By some

© 1999 by CRC Press LLC

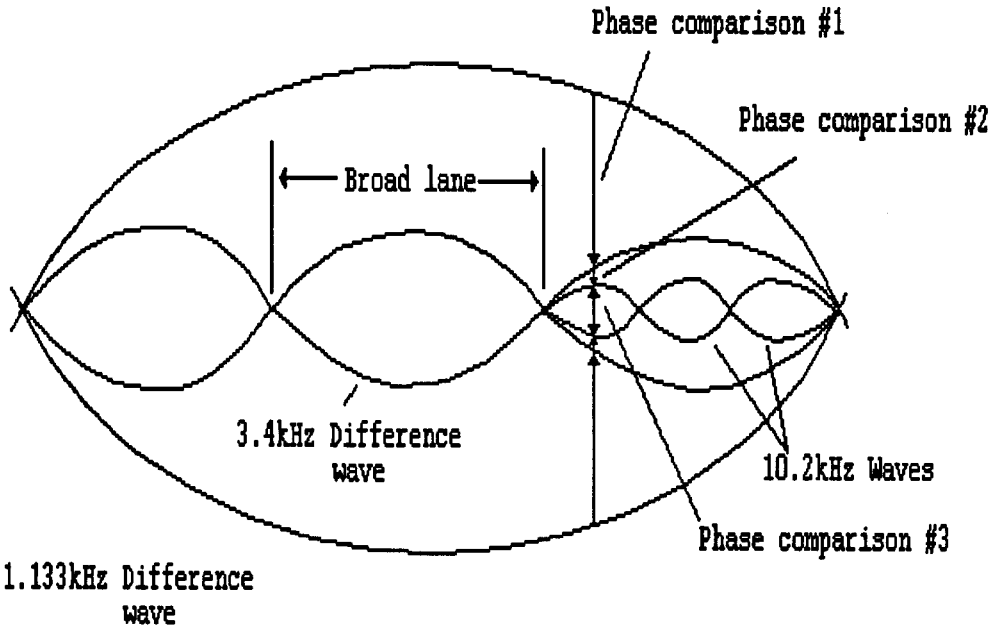

FIGURE 10.32 Three successive phase comparisons for lane resolution in Omega systems. Phase differences are compared in three stages with respect to three different signals transmitted by each station for accurate position finding. One wavelength is 25 km, representing two lanes. Accuracy of the Omega system is limited.

short signal interruptions and transmissions of harmonics simultaneously, zones and lanes are identified in a precise manner.

Consol

Consol is limited to the Eastern and Northern Atlantics. It is a hyperbolic system with extremely short baseline lengths, such that a collapsed hyperbolic pattern is formed. It employs three towers located three wavelengths apart. The operational frequencies are in the MF range between 250 kHz and 370 kHz. The system range is about 1900 km by day and 2400 km by night. The minimum range is about 40 km near the stations. One tower transmits a continuous wave, while other towers transmit waves with 180° phase shift by a *keying cycle*. The signals are modulated by dots and dashes such that receivers determine the position by counting them and printing on Consol grid patterns.

Omega

Omega is a hyperbolic navigation system that covers the entire world with only eight transmission stations located 7500 km to 9500 km apart. It transmits on frequencies in the VLF band from 10 kHz to 14 kHz at a power of 10 kW. The signals of at least three and usually four stations can be received at any position on Earth.

The 10 kHz to 14 kHz frequency band was chosen specifically to take advantage of several favorable propagation characteristics, such as: (1) to use the Earth's surface and ionosphere as a waveguide; (2) to enable submerged submarines to receive the signals; and (3) to form long baselines at 7500 km to 9500 km.

The basic frequency at which all eight stations transmit is 10.2 kHz. Each station transmits four navigation signals as well as a timing signal with atomic frequency standards ensuring that all stations are kept exactly in phase. Two continuous waves are in phase but traveling in opposite directions to produce a series of Omega lanes. Within each lane, a phase difference measurement would progress from 0° to 360° as the receiver moves across, as shown in Figure 10.32. Two Omega lanes complete one cycle, giving a wavelength of 25 km and lane of 12 km expanding as the distance from the baseline increases. Lanes are identified by three other signals transmitted by each station on a multiplexed basis. Omega

© 1999 by CRC Press LLC

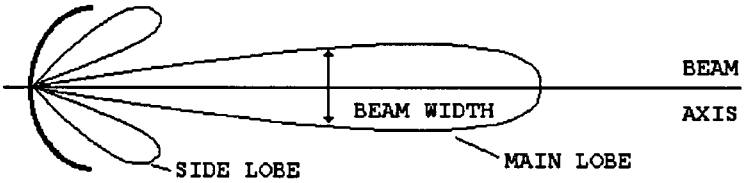

FIGURE 10.33 A surface search radar beam. High-frequency electromagnetic waves are formed by parabolic antenna. The receiving antenna is rotated 360° to scan the entire surrounding area. The location of the target is determined by the reflected back signals and the orientation of the antenna.

fixes have been accurate to within ±1.5 km rms by day and ±3 km rms by night. Differential techniques can greatly reduce this error.

Omega receivers are fully automated sets that provide direct lat/long readout for a cost of around U.S. $1,000 for marine and aviation use. Nevertheless, since the precision of the system is not as good as others, they are mainly used as back-up systems.

Radar

The word is derived from *radio detection* and *ranging*. It works on the basic principle of reflection, which determines the time required for an RF pulse to travel from a reference source to a target and return as an echo. Most surface search and navigation radar high-frequency electromagnetic waves are formed by a parabolic antenna into a beam form, as shown in Figure 10.33. The receiving antenna is rotated to scan the entire surrounding area, and the bearings to the target are determined by the orientation of the antenna at the moment the echo returns. A standard radar is made up of five components: transmitter, modulator, antenna, receiver, and indicator. They operate on pulse modulation.

Radars are extremely important devices for air control applications. Nowadays, airborne beacon radar systems are well developed in traffic alert and collision avoidance systems (TCAS). In this system, each plane constantly emits an interrogation signal, which is received by all nearby aircraft that are equipped appropriately. The signal triggers a transponder in the target aircraft, which then transmits some information concerning 3-D location and identification.

Satellite Relay Systems

The use of satellites is a highly developed technology utilized extensively throughout the world. In the past 2 decades, it has progressed from quasi-experimental in nature to one with routine provisions of new services. They take advantage of the unique characteristics of *geostationary satellite orbits* (GSO). The design of satellite systems is well understood, but the technology is still dynamic. The satellites are useful for long-distance communication services, for services across oceans or difficult terrain, and point-to-multipoint services such as television distribution.

Frequency allocation for satellites is controlled by the International Telecommunication Union (ITU). In the U.S., the Federal Communications Commission (FCC) makes the frequency allocations and assignments for nongovernment satellite usage. The FCC imposes a number of conditions regarding construction and maintenance of in-orbit satellites.

There are many satellite systems operated by different organizations and different countries mainly developed for communications and data transmissions; these include: Iridium of Motorola, Globalstar of Loral Corporation, Intelsat, CS-series of Japan, Turksat of Turkey, Aussat of Australia, Galaxy and Satcom of the U.S., Anik of Canada, TDF of France, etc. Some of the communication satellite systems are suitable for navigation purposes. However, satellite systems specifically designed for navigation are limited in number. The most established and readily accessible by civilian and commercial users are the GPS system of the U.S. and the Glonass of Russia.

The first generation of the satellite system was the *Navy satellite system* (Navsat), which became operational in January 1964, following the successful launch of the first transit satellite into polar orbit.

© 1999 by CRC Press LLC

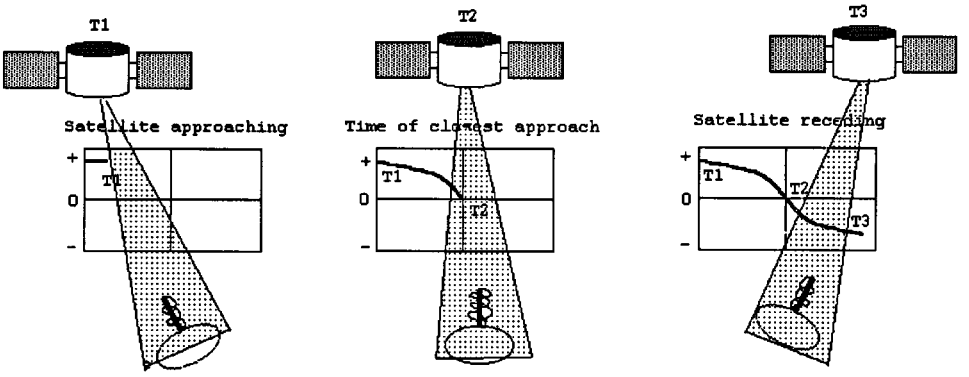

FIGURE 10.34 Transit satellite Doppler curve. As the satellite approaches the receiver, the frequency of the received signal increases due to Doppler shift. At the time of closest approach, the transmitted and received frequencies are the same. The frequencies received from a receding satellite result in lower values. This is also applicable in other position sensing satellites such as GPS, Glonass, Starfix, etc.

The system was declared open for private and commercial use in 1967. Civil designation of the name of the system is *Transit Navigation Satellite System*, or simply *Transit*. Later, this system evolved to become the modern Navsat GPS system, which will be discussed in detail. Most of the operational principles discussed here are inherited by the GPS system.

The Transit system consists of operational satellites, plus several orbiting spares, a network of ground tracking stations, a computing center, an injection station, naval observatory time signals, and receiver-computer combinations. The transit satellites are in circular polar orbits about 1075 km above ground with periods of revolution of about 107 min. Because of the rotation of the Earth beneath the satellites, every position on Earth comes within range of each satellite at least twice a day, at 12 h intervals. As originally intended, if at least five satellites are operational at any given time, the average time between fix opportunities would vary from about 95 min near the equator to about 35 min or less above 70° North and South.

The Transit system is based on the Doppler shift of two frequencies, 150 MHz and 400 MHz, transmitted simultaneously by each satellite moving its orbit at a tangential velocity of about 7.5 km s^{-1}. Two frequencies are used so that the effects of the ionosphere and atmospheric refraction on the incoming satellite transmission can be compensated for by the receivers. Each frequency is modulated by a repeating data signal lasting 2 min, conveying the current satellite time and its orbital parameters and other information. Within the receiver, a single Doppler signal is created by processing the two signals transmitted by the satellite. By plotting the frequency of this signal versus time, a characteristic curve of the type shown in Figure 10.34 is obtained. Since the frequency of the transmitted signal is compressed as the satellite approaches, according to what proportion of the velocity vector is seen by the user receiver, the curve begins at time T_1 at a frequency several cycles higher than the transmitted frequency.

Tracking stations record Doppler observations and memory readout received during each satellite pass to relay them to a computer center. Updated orbital position and time data communications are relayed to an "injection" station from the computer center for transmission to satellite in a burst once each 12 h. Enough data is supplied in this 15 s injection message to last for 16 h of consecutive 2 min broadcasts describing the current orbital positions of the satellite.

The system accuracy depends on the accuracy of the satellite orbit computation, the effect of ionosphere refraction, the precision of the receiver speed, and heading determination. Under optimal conditions, the system is capable of producing fixes with a maximum rms error of about 35 m for the stationary receivers anywhere on Earth. Nevertheless, if a site is occupied for several weeks, an accuracy better than 1 m can be achieved. The time signal transmitted as a "beep" at the end of each 2 min transmission cycle coincides with even minutes of Coordinated Universal Time, which can be used as a chronometer check.

© 1999 by CRC Press LLC

There are other satellite systems, either already in existence or in the planning stages, suitable for navigation. Some of these are: Marec satellites operating at VHF and owned by the intergovernment consortium INMARSAT; privately owned Geostar provides services for oil industry; and many other systems offering transcontinental communication and navigation services as well as position sensing; examples include: SATCOM, ARINC's, Avsat, Starfix, etc.

Transponders

Transponders are transducers that respond to incoming signals by generating appropriate reply messages. Recent developments in technology have made the configuration of transponders possible using elaborate and powerful on-board signal processing. This enhanced the capacity by improving the link budgets, by adjusting the antenna patterns and by making the satellite resources available on a demand basis — called the "switch board in the sky concept."

Increased interest in deep sea exploration has brought acoustic transponders to the forefront as an important navigation tool. They provide three-dimensional position information for subservience vehicles and devices.

Some transponders are based on radar signals that respond to radar illumination. Transponders are programmed to identify friend or foe or, in some cases, simply inform ground stations about the position of aircraft.

Transponders are used for emergency warning. The U.S. and Canadian satellites carry Sarsat transponders, and Russian satellites carry Cospas transponders. They are also used as warning devices in collision avoidance systems in aircraft and land vehicles.

Global Satellite Navigation Systems

The GPS System

The Global Satellite Navigation Systems are second-generation satellites evolved primarily from the Naval Global Positioning System. They provide a continuous three-dimensional position-finding capability (i.e., latitude, longitude, and altitude), in contrast to the periodic two-dimensional information of the Transit system. Twenty-four operational satellites, as shown in Figure 10.35, constitute the system. Each satellite orbit is circular, about 2200 km high, and inclined at angles of 55° with respect to Earth's axis.

The position determination using the GPS system is based on the ability of the receivers to accurately determine the distance to the GPS satellites above the user's horizon at the time of fix. If accurate distances of two such satellites and the heights are known, then the position can be determined. In order to do this, the receiver would need to know the exact time at which the signal was broadcast and the exact time that it was received. If the propagation speeds through the atmosphere are known, the resulting range can be calculated. The measured ranges are called *pseudoranges.* Nowadays, normally, information is received from at least four satellites, leading to accurate calculations of the fix. The time errors plus propagation speed errors result in range errors, common to all GPS receivers. Time is the fourth parameter evaluated by the receiver if at least four satellites can be received at a given time. If a fifth satellite is received, an error matrix can be evaluated additionally.

Each GPS satellite broadcasts simultaneously on two frequencies for the determination and elimination of ionosphere and other atmospheric effects. The Navstar frequencies are at 1575.42 MHz and 1227.6 MHz, designated as L1 and L2 in the L-band of the UHF range. Both signals are modulated by 30 s navigation messages transmitted at 50 bits s^{-1}. The first 18 s of each 30 s frame contain *ephemeris* data for that particular satellite, which defines the position of the satellite as a function of time. The remaining 12 s is the *almanac* data, which define orbits and operational status of all satellites in the system. The GPS receivers store and use the ephemeris data to determine the pseudorange, and the almanac data to help determine the four best satellites to use for positional data at any given time. However, the "best four" philosophy has been overtaken slowly by an all-in-view philosophy.

© 1999 by CRC Press LLC

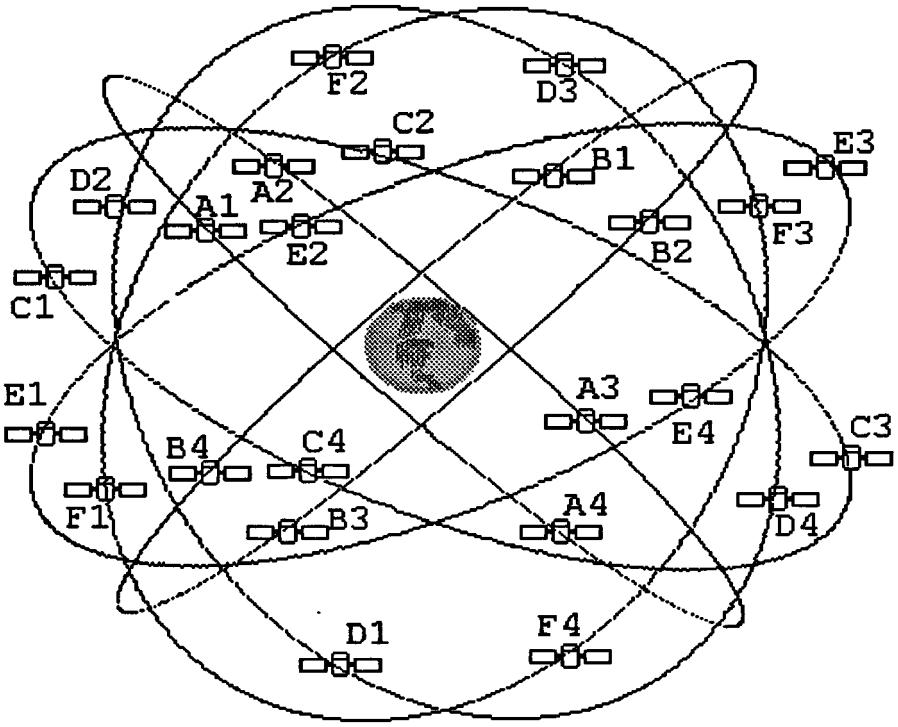

FIGURE 10.35 Operational GPS satellite coverage for navigation. Four satellites orbit in circular form. There are six such orbits inclined at angles of 60° from each other. In this arrangement, any point on Earth can see at least four satellites at any given time. This yields great accuracy in the position determination of the target, even only with C/A codes received.

The L1 and L2 satellite navigation signals are also modulated by two additional binary sequences called *C/A code* for acquisition of coarse navigation and the other *P-code* for precision ranging. The L1 signal is modulated both by the C/A and P-codes, and the L2 only by the P-code. Positional accuracies of about 20 m rms are usual in using C/A codes alone. The P-code, however, is not available for civilian users. The P-code is redesignated to be a Y-code, decipherable only by high-precision receivers having access to encrypted information in the satellite message. Nevertheless, it is fair to comment that civilians have figured out how to benefit the P/Y signals without actually knowing the codes, but at lower SNR. Further, C/A codes are degraded by insertion of random errors such that positional accuracy is limited to 50 m rms for horizontal values and 70 m for vertical values. These errors are intended to be lifted by the year 2006. Civilian users have access to the so-called *Standard Positioning Services* (SPS) accurate to 50 m rms, while U.S. and NATO military users will use *Precise Positioning Service* (PPS).

In enhancing SPS accuracy, differential techniques may be applied, as shown in Figure 10.36, to the encrypted GPS signals. Since the reference receiver is at a known location, it can calculate the correct ranges of pseudoranges at any time. The differences in the measured and calculated pseudoranges give the correction factors. Accuracy less than 1 m can be obtained in the stationary and moving measurements. Recently, differential receivers became commonly available, giving higher accuracy in sub-centimeter ranges. They are getting cheaper day by day and finding applications in many areas such as airplanes, common transport vehicles, cars, geological surveying, orienteering, farming, etc.

There are currently three basic types of GPS receivers designed and built to address various user communities. These are called *slow sequencing, fast sequencing,* and *continuous tracking* receivers. The least complicated and lowest cost receiver for most applications is the slow sequencing type, wherein only one measurement channel is used to receive sequential L1 C/A code from each satellite every 1.2 s,

© 1999 by CRC Press LLC

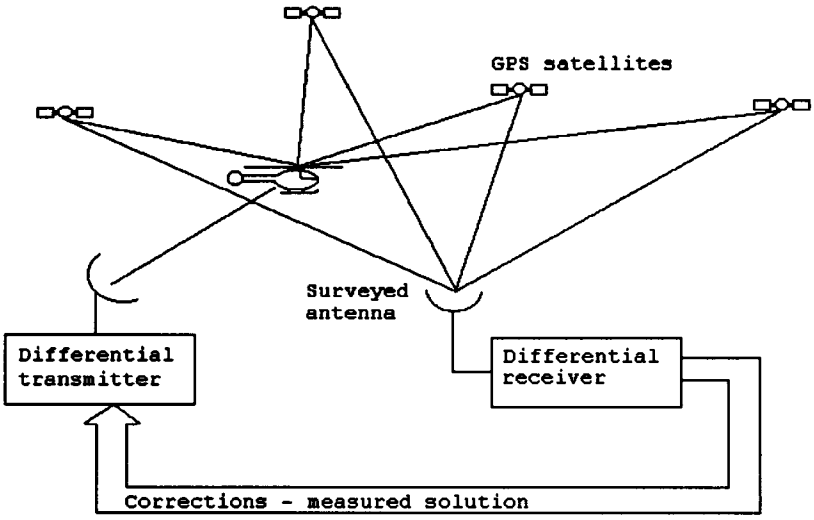

FIGURE 10.36 Differential GPS operation. Satellite and target positions are sensed by ground-fixed receivers or mobile receivers with exact known positions. Errors of the target position due to signals received from the satellites are corrected using the information from the fixed receivers. Using this method, accuracy in submeter range can be obtained.

with occasional interrupts to collect ephemeris and almanac data. Once the data is received, computation is carried out within 5 s, making this system suitable for stationary or near-stationary fixes.

Fast sequencing receivers have two channels: one for making continuous pseudorange measurements, and the other collection of the ephemeris and almanac data. This type is used in medium dynamic applications such as ground vehicles.

Continuous tracking receivers employ multiple channels (at least five) to track, compute, and process the pseudoranges to the various satellites being utilized simultaneously, thus obtaining the highest possible degree of accuracy and making it suitable for high dynamic applications such as aircraft and missiles. The parallel channel receivers are so cost effective nowadays that other types are likely to disappear.

A number of companies produce highly sophisticated GPS receivers. EURONAV GPS receivers operate on two and five channels for military applications. They provide features such as precise time, interfacing with digital flight instruments, RS-422 interface, altimeter input, self initialization, etc.

Software implementation satellite management functions, having different features, are offered by many manufacturers. In the case of DoD NAVSTAR GPS receivers, for example, three functional requirements are implemented: (1) database management of satellite almanac, ephemeris, and deterministic correction data; (2) computation of precise satellite position and velocity for use by navigation software; and (3) using satellite and receiver position data to periodically calculate the constellation of four satellites with optimum geometry for navigation. The DoD receivers are divided for three functions as Satellite Manager (SM), Satellite-Data-Base-Manager (SDBM) SV-Position Velocity Acceleration (SVPVA), and Select-Satellites (SS).

Differential navigation is also applied where one user set is navigating relative to another user set via a data link. In some cases, one user has been at a destination at some prior time and is navigating relative to coordinates measured at that point. The true values of this receiver's navigation fix are compared against the measured values, and the differences become the differential corrections. These corrections are transmitted to area user sets in real-time, or they may be recorded for post-mission use so that position fixes are free of GPS-related biases.

Differential navigation and GPS systems find applications in enroute navigations for commercial and civil aviation, military application, navigation of ships especially in shallow waters, in station keeping of aircraft, seismic geophysical explorations, land surveying, transport vehicles and traffic controls, etc.

© 1999 by CRC Press LLC

The Glonass

There are a number of other satellite navigation systems similar to GPS of the U.S., such as Russian *Glonass*. The Glonass consists of 24 satellites orbiting in circular form 1500 km above the ground. The accuracy of the system is about 10 m rms. Glonass satellites transmit details of their own position and a time reference. The carrier frequencies are in L-band, around 1250 MHz (L2) and 1600 MHz (L1). Only the L1 frequency carries the Civil C/A code. The radio frequency carriers used by Glonass are channelized within bands 1240–1260 MHz and 1597–1617 MHz, the channel spacing being 0.4375 MHz at the lower frequencies and 0.5625 MHz at the higher frequencies. The number of channels is 24. Glonass data message is formatted in frames of 3000 bits, with a duration of 30 s. The ephemeris data are transmitted as a set of position, velocity, and acceleration coordinates in a Cartesian Earth-centered, Earth-fixed (ECEF) coordinate system. The new ephemeris data are available every half hour, valid for the following quarter-hour. The data are sent at a 50 baud rate and superimposed on a pseudorandom noise (PRN) code. The low-precision code has length 511 bits as compared to 1023 bits for Navstar. Glonass accuracy is as good as that for the GPS system. Glonass and GPS have different coordinate frames and different time frames that are being coordinated together.

The Starfix

Another interesting satellite-based system — privately funded, developed, launched, and maintained — is the Starfix positioning system. This system is designed primarily for oil surveying. The system consists of a master site, which generates satellite ephemeris data, and four satellites in geosynchronous orbits. The system said to have a precision of 2.5 m rms.

References

M. Kayton, *Navigation-Land, Sea, Air and Space,* New York: IEEE Press, 1990.

A. F. Inglis, *Electronic Communication Handbook,* New York: McGraw-Hill, 1988.

J. Everet, *VSATs — Very Small Aperture Terminals*, IEEE Telecommunication Series 28, London: Peter Peregrinus Ltd., 1992.

A. Leick, *GPS Satellite Surveying*, New York: John Wiley & Sons, 1990.

B. R. Elbert, *The Satellite Communication Applications Handbook,* Boston: Artech House, 1997.

Appendix — List of Manufacturers/Suppliers

Advanced Videotech Co.
1840 County Line Rd., Dept. G
Huntington Valley, PA 19006
Tel: (800) 221-8930
Fax: (800) 221-8932

AlliedSignal
101 Colombia Rd.
Dept. CAC
Morristown, NJ 07962
Tel: (800) 707-4555
Fax: (602) 496-1001

American GNC Corp.
9131 Mason Avenue
Chatsworth, CA 91311
Tel: (818) 407-0092
Fax: (818) 407-0093

Astroguide
Lasalle, IL 61301
Tel: (815) 224-2700
Fax: (815) 224-2701

Colombia Elect. Int. Inc.
P.O. Box 960-T
Somis, CA 93066
Tel: (805) 386-2312 or
(800) 737-9662
Fax: (805) 386-2314

Comstream
10180 Barnes Canyon Rd.
San Diego, CA 92121
Tel: (800) 959-0811
Fax: (619) 458-9199

GE Co.
3135 Easton Tpke.
Fairfield, CT 06431
Tel: (800) 626-2004
Fax: (518) 869-2828

Orbitron
351-TR-S Peterson St.
Spring Green, WI 53588
Tel: (608) 588-2923
Fax: (608) 588-2257

STI
31069 Genstar Rd.
Hayward, CA 94544-7831
Tel: (800) 991-4947
Fax: (510) 471-9757

© 1999 by CRC Press LLC

10.5 Occupancy Detection

Jacob Fraden

Occupancy sensors detect the presence of people in a monitored area. Motion detectors respond only to moving objects. A distinction between the two is that the occupancy sensors produce signals whenever an object is stationary or not, while the motion detectors are selectively sensitive to moving objects. The applications of these sensors include security, surveillance, energy management (electric lights control), personal safety, friendly home appliances, interactive toys, novelty products, etc. Depending on the applications, the presence of humans may be detected through any means that is associated with some kind of a human body's property or actions [1]. For example, a detector may be sensitive to body weight, heat, sounds, dielectric constant, etc. The following types of detectors are presently used for the occupancy and motion sensing of people:

1. *Air pressure sensors*: detect changes in air pressure resulting from opening doors and windows
2. *Capacitive*: detectors of human body capacitance
3. *Acoustic*: detectors of sound produced by people
4. *Photoelectric*: interruption of light beams by moving objects
5. *Optoelectric*: detection of variations in illumination or optical contrast in the protected area
6. *Pressure mat switches*: pressure-sensitive long strips used on floors beneath the carpets to detect the weight of an intruder
7. *Stress detectors*: strain gages embedded into floor beams, staircases, and other structural components
8. *Switch sensors*: electrical contacts connected to doors and windows
9. *Magnetic switches*: a noncontact version of switch sensors
10. *Vibration detectors*: react to the vibration of walls or other building structures; may also be attached to doors or windows to detect movements
11. *Glass breakage detectors*: sensors reacting to specific vibrations produced by shattered glass
12. *Infrared motion detectors*: devices sensitive to heat waves emanating from warm or cold moving objects
13. *Microwave detectors*: active sensors responsive to microwave electromagnetic signals reflected from objects
14. *Ultrasonic detectors*: similar to microwaves, except that instead of electromagnetic radiation, ultrasonic waves are used
15. *Video motion detectors*: video equipment that compares a stationary image stored in memory with the current image from the protected area
16. *Laser system detectors*: similar to photoelectric detectors, except that they use narrow light beams and combinations of reflectors
17. *Triboelectric detectors*: sensors capable of detecting static electric charges carried by moving objects

One of the major aggravations in detecting occupancy or intrusion is a false positive detection. The term "false positive" means that the system indicates an intrusion when there is none. In some noncritical applications where false positive detections occur once in a while, for example, in a toy or a motion switch controlling electric lights in a room, this may be not a serious problem: the lights will be erroneously turned on for a short time, which will unlikely do any harm. In other systems, especially those used for security purposes, the false positive detections, while generally not as dangerous as false negative ones (missing an intrusion), may become a serious problem. While selecting a sensor for critical applications, consideration should be given to its reliability, selectivity, and noise immunity. It is often good practice to form a multiple sensor arrangement with symmetrical interface circuits; this can dramatically improve the reliability of a system, especially in the presence of external transmitted noise. Another efficient way to reduce erroneous detections is to use sensors operating on different physical principles [2]; for example, combining capacitive and infrared detectors is an efficient combination as they are receptive to different kinds of transmitted noise.

© 1999 by CRC Press LLC

Ultrasonic Sensors

Ultrasonic detectors are based on transmission to the object and receiving reflected acoustic waves. Ultrasonic waves are mechanical — they cover frequency range well beyond the capabilities of human ears, i.e., over 20 kHz. However, these frequencies may be quite perceptive by smaller animals, like dogs, cats, rodents, and insects. Indeed, the ultrasonic detectors are the biological ranging devices for bats and dolphins.

When the waves are incident on an object, part of their energy is reflected. In many practical cases, the ultrasonic energy is reflected in a diffuse manner. That is, regardless of the direction where the energy comes from, it is reflected almost uniformly within a wide solid angle, which may approach 180°. If an object moves, the frequency of the reflected waves will differ from the transmitted waves. This is called the Doppler effect (see below). To generate any mechanical waves, including ultrasonic, the movement of a surface is required. This movement creates compression and expansion of the medium, which can be a gas (air), a liquid, or a solid. The most common type of the excitation device that can generate surface movement in the ultrasonic range is a piezoelectric transducer operating in the so-called *motor* mode [3]. The name implies that the piezoelectric device directly converts electrical energy into mechanical energy.

Microwave Motion Detectors

Microwave detectors offer an attractive alternative to other detectors, when it is required to cover large areas and to operate over an extended temperature range under the influence of strong interferences (e.g., wind, acoustic noise, fog, dust, moisture, etc.). The operating principle of the microwave detector is based on radiation of electromagnetic radio frequency (RF) waves toward a protected area. The most common frequencies are 10.525 GHz (X-band) and 24.125 GHz (K-band). These wavelengths are long enough ($\lambda = 3$ cm at X-band) to pass freely through most contaminants, such as airborne dust, and short enough to be reflected by larger objects.

The microwave part of the detector consists of a Gunn oscillator, an antenna, and a mixer diode. The Gunn oscillator is a diode mounted in a small precision cavity that, on application of power, oscillates at microwave frequencies. The oscillator produces electromagnetic waves, part of which is directed through an iris into a waveguide and focusing antenna that directs the radiation toward the object. Focusing characteristics of the antenna are determined by the application. As a general rule, the narrower the directional diagram of the antenna, the more sensitive it is (the antenna has a higher gain). Another general rule is that a narrow beam antenna is much larger, while a wide-angle antenna can be quite small. A typical radiated power of the transmitter is 10 mW to 20 mW.

An antenna transmits the frequency f_0, which is defined by the wavelength λ_0 as:

$$f_0 = \frac{c_0}{\lambda_0} \tag{10.106}$$

where c_0 is the speed of light. When the target moves toward or away from the transmitting antenna, the frequency of the reflected radiation will change. Thus, if the target is moving away with velocity v, the reflected frequency will decrease, and it will increase for the approaching targets. This is called the *Doppler effect*, after the Austrian scientist Christian Johann Doppler (1803–1853). While the effect was first discovered for sound, it is applicable to electromagnetic radiation as well. However, in contrast to sound waves that may propagate with velocities dependent on movement of the source of the sound, electromagnetic waves propagate with speed of light, which is an absolute constant. The frequency of reflected electromagnetic waves can be predicted by the theory of relativity as:

$$f_r = f_0 \frac{\sqrt{1-\left(v/c_0\right)^2}}{1+v/c_0} \tag{10.107}$$

© 1999 by CRC Press LLC

For practical purposes, however, the quantity $(v/c_0)^2$ is very small compared with unity; hence, it can be ignored. Then, the equation for the frequency of the reflected waves becomes identical to that for the acoustic waves:

$$f_r = f_0 \frac{1}{1 + v/c_0} \tag{10.108}$$

Due to a Doppler effect, the reflected waves have a different frequency f_r. A mixing diode combines the radiated (reference) and reflected frequencies and, being a nonlinear device, produces a signal that contains multiple harmonics of both frequencies.

The Doppler frequency in the mixer can be found from:

$$\Delta f = f_0 - f_r = f_0 \frac{1}{c_0/v + 1} \tag{10.109}$$

and since $c_0/v \gg 1$, the following holds after substituting Equation 10.106:

$$\Delta f \approx \frac{\nu}{\lambda_0} \tag{10.110}$$

Therefore, the signal frequency at the output of the mixer is linearly proportional to the velocity of a moving target. For example, if a person walks toward the detectors with a velocity of 0.6 m s^{-1}, a Doppler frequency for the X-band detector is $\Delta f = 0.6/0.03 = 20$ Hz.

Equation 10.110 holds true only for movements in the normal direction. When the target moves at angles Θ with respect to the detector, the Doppler frequency is:

$$\Delta f \approx \frac{\nu}{\lambda_0} \cos\Theta \tag{10.111}$$

Micropower Impulse Radar

In 1993, Lawrence Livermore National Laboratory developed a *micropower impulse radar* (MIR), which is a low-cost, noncontact ranging sensor [3]. The operating principle of the MIR is fundamentally the same as a conventional pulse radar system, but with several significant differences. The MIR consists of a noise generator whose output signal triggers a pulse generator. Each pulse has a fixed short duration, while the repetition of these pulses is random, according to triggering by the noise generator. The pulses are spaced randomly with respect to one another in a Gaussian noise-like pattern. It can be said that the pulses have the pulse frequency modulation (PFM) by white noise with maximum index of 20%. In turn, the square-wave pulses cause amplitude modulation (AM) of a radio transmitter. The radio transmitter produces short bursts of high-frequency radio signal that propagate from the transmitting antenna to the surrounding space. The electromagnetic waves reflect from the objects and propagate back to the radar. The same pulse generator that modulates the transmitter, gates (with a predetermined delay) the radio receiver to enable the output of the MIR only during a specific time window. Another reason for gating the receiver is to reduce its power consumption. The reflected pulses are received, demodulated (the square-wave shape is restored from the radio signal), and the time delay with respect to the transmitted pulses is measured. Since the pulses are spaced randomly, practically any number of identical MIR systems can operate in the same space without a frequency division (i.e., they work at the same carrier frequency within the same bandwidth). There is little chance that bursts from the interfering transmitters overlap and, if they do, the interference level is significantly reduced by the averaging circuit.

© 1999 by CRC Press LLC

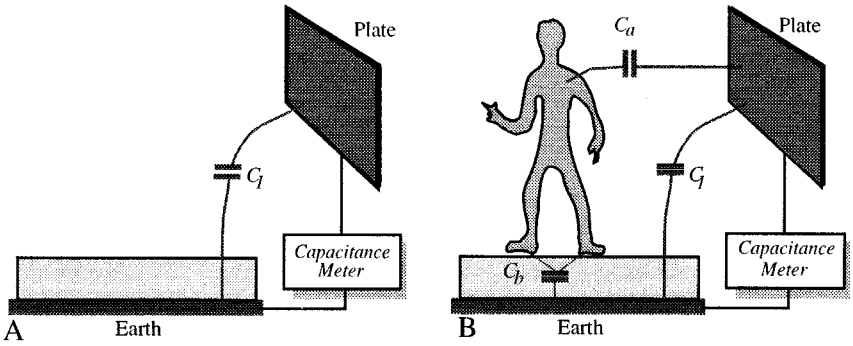

FIGURE 10.37 An intruder brings in additional capacitance to a detection circuit.

Capacitive Occupancy Detectors

Being a conductive medium with a high dielectric constant, a human body develops a coupling capacitance to its surroundings. This capacitance greatly depends on such factors as body size, clothing, materials, type of surrounding objects, weather, etc. However wide the coupling range is, the capacitance can vary from a few picofarads to several nanofarads. When a person moves, the coupling capacitance changes, thus making it possible to discriminate static objects from the moving ones. In effect, all objects form some degree of capacitive coupling with respect to one another. If a human (or for that purpose, anything) moves into the vicinity of the objects whose coupling capacitance with each other has been previously established, a new capacitive value arises between the objects as a result of the presence of an intruding body [3]. Figure 10.37 shows that the capacitance between a test plate and Earth is equal to C_1. When a person moves into the vicinity of the plate, it forms two additional capacitors: one between the plate and its own body C_a, and the other between the body and the Earth, C_b. Then, the resulting capacitance C between the plate and the Earth becomes larger by ΔC.

$$C = C_1 + \Delta C = C_1 + \frac{C_a C_b}{C_a + C_b} \tag{10.112}$$

With the appropriate apparatus, this phenomenon can be used for occupancy detection [3]. What is required is to measure a capacitance between a test plate (the probe) and a reference plate (the Earth).

Figure 10.38 illustrates a circuit diagram for detecting variations in the probe capacitance C_p [4]. That capacitance is charged from a reference voltage source V_{ref} through a gate formed by transistor Q_1 when the output voltage of a control oscillator goes low. When it goes high, transistor Q_1 closes while Q_2 opens. The probe capacitance C_p discharges through a constant-current sink constructed with a transistor Q_3. A capacitor C_1 filters the voltage spikes across the transistor. The average voltage, e_p, represents a value of the capacitor C_p. When an intruder approaches the probe, the latter's capacitance increases, which results in a voltage rise across C_1. The voltage change passes through the capacitor C_2 to the input of a comparator with a fixed threshold V_T. The comparator produces the output signal V_{out} when the input voltage exceeds the threshold value.

When a capacitive occupancy (proximity) sensor is used near or on a metal device, its sensitivity can be severely reduced due to capacitive coupling between the electrode and the device's metallic parts. An effective way to reduce that stray capacitance is to use driven shields [3].

Triboelectric Detectors

Any object can accumulate, on its surface, static electricity. These naturally occurring charges arise from the triboelectric effect; that is, a process of charge separation due to object movements, friction of clothing

© 1999 by CRC Press LLC

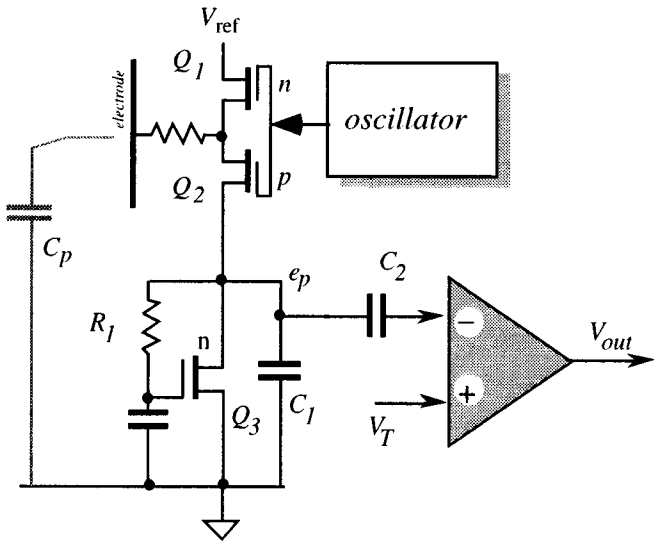

FIGURE 10.38 Circuit diagram for a capacitive intrusion detector.

fibers, air turbulence, atmosphere electricity, etc. Usually, air contains either positive or negative ions that can be attracted to the human body, thus changing its charge. Under idealized static conditions, an object is not charged — its bulk charge is equal to zero. In reality, any object that, at least temporarily, is isolated from the ground can exhibit some degree of its bulk charge imbalance. In other words, it becomes a carrier of electric charges that become a source of electric field. The field, or rather its changes, can be detected by an electronic circuit having a special pick-up electrode at its input [3, 5]. The electrode increases the capacitive coupling of the circuit's input with the environment, very much like in the capacitive detectors described above. The electrode can be fabricated in the form of a conductive surface that is well isolated from the ground.

If a charge carrier (a human or an animal) changes its position — moves away or a new charge carrying an object enters into the vicinity of the electrode — the static electric field is disturbed. The strength of the field depends on the atmospheric conditions and the nature of the objects. For example, a person in dry man-made clothes walking along a carpet carries a million times stronger charge than a wet intruder who has come from the rain.

It should be noted that contrary to a capacitive motion detector, which is an active sensor, a triboelectric detector is passive; that is, it does not generate or transmit any signal. There are several possible sources of interference that can cause spurious detections by the triboelectric detectors. That is, the detector may be subjected to transmitted noise resulting in false positive detection. Among the noise sources are 60 Hz or 50 Hz power line signals, electromagnetic fields generated by radio stations, power electric equipment, lightnings, etc. Most of these interferences generate electric fields that are distributed around the detector quite uniformly and can be compensated for by employing a differential input circuit with a significant common mode rejection ratio.

Optoelectronic Motion Detectors

Optoelectronic motion detectors rely on electromagnetic radiation in the optical range, specifically having wavelengths from 0.4 μm to 20 μm. This covers visible, near- and part of far-infrared spectral ranges. The detectors are primarily used for the indication of movement of people and animals. They operate over distance ranges up to several hundred meters and, depending on the particular need, can have either a narrow or wide field of view.

© 1999 by CRC Press LLC

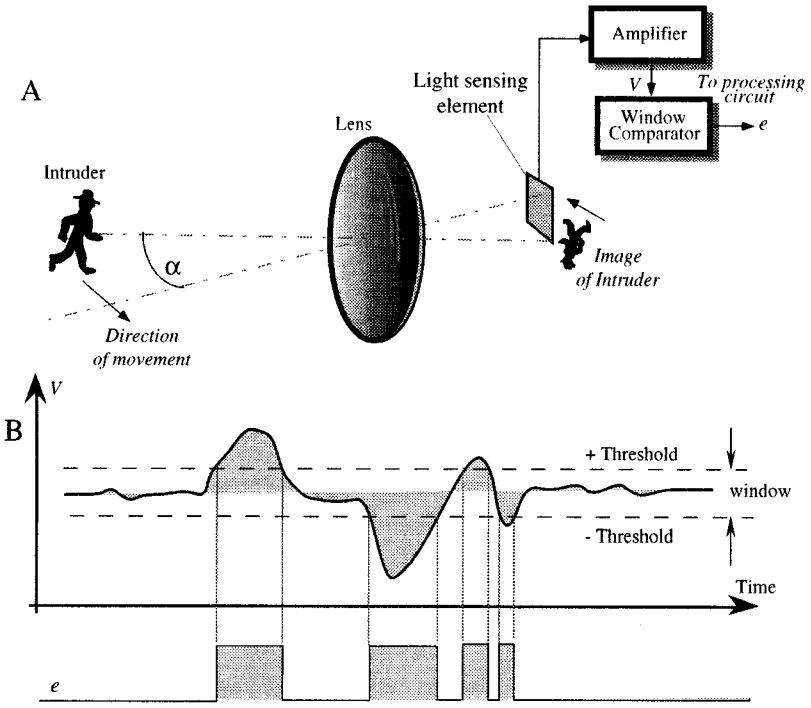

FIGURE 10.39 General arrangement of an optoelectronic motion detector. A lens forms an image of a moving object (intruder). When the image crosses the optical axis of the sensor, it superimposes with the sensitive element (*A*). The element responds with the signal that is amplified and compared with two thresholds in the window comparator (*B*). (From J. Fraden, *Handbook of Modern Sensors,* 2nd ed., Woodburg, NY: AIP Press, 1997. With permission.)

Most of the objects (apart from very hot) radiate electromagnetic waves only in the mid- and far-infrared spectral ranges. Hence, visible and near-infrared light motion detectors must rely on an additional source of light to illuminate the object. The light is reflected by the object's body toward the focusing device for subsequent detection. Such illumination can be sunlight or the invisible infrared light from an additional near-infrared light source (a projector).

The major application areas for the optoelectronic motion detectors are in security systems (to detect intruders), in energy management (to turn lights on and off), and in the so-called "smart" homes where they can control various appliances such as air conditioners, cooling fans, stereo players, etc. They can also be used in robots, toys, and novelty products. The most important advantage of an optoelectronic motion detector is simplicity and low cost.

Sensor Structures

A general structure of an optoelectronic motion detector is shown in Figure 10.39(*A*). Regardless what kind of sensing element is employed, the following components are essential: a focusing device (a lens or a focusing mirror), a light detecting element, and a threshold comparator. An optoelectronic motion detector resembles a photographic camera. Its focusing components create an image of its field of view on a focal plane. While there is no mechanical shutter like in a camera, in place of the film, a light sensitive element is used. The element converts the focused light into an electric signal. A focusing lens creates an image of the surroundings on a focal plane where the light sensitive element is positioned. If the area is unoccupied, the image is static and the output signal from the element is steady stable. When an "intruder" enters the room and keeps moving, his/her image on the focal plane also moves. In a certain moment, the intruder's body is displaced for an angle α and the image overlaps with the element. This

© 1999 by CRC Press LLC

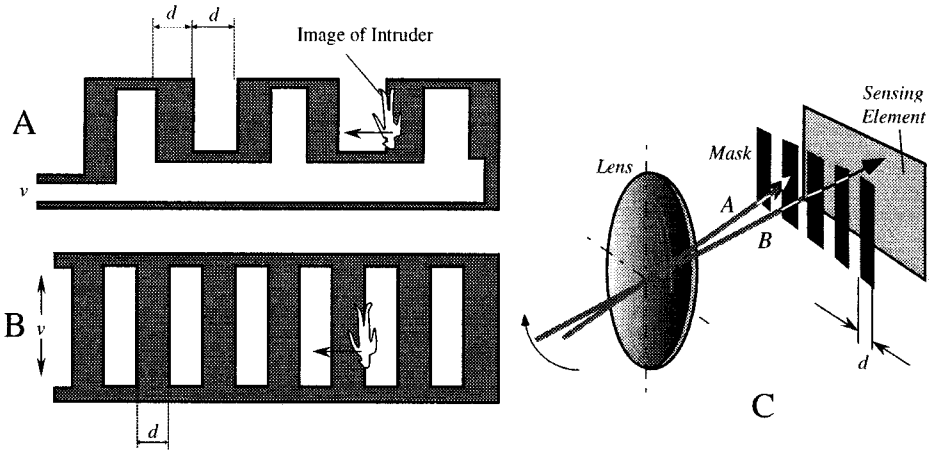

FIGURE 10.40 Complex shapes of light sensing element with series (*A*) and parallel (*B*) connection of segments. (*C*) shows use of a grid mask in front of light sensing element. (From J. Fraden, *Handbook of Modern Sensors,* 2nd ed., Woodburg, NY: AIP Press, 1997. With permission.)

is an important point to understand — the detection is produced only at the moment when the object's image either coincides with the detector's surface or clears it. That is, no overlapping — no detection. Assuming that the intruder's body creates an image with a light flux different from that of the static surroundings, the light-sensitive element responds with deflecting voltage *V.* In other words, to cause detection, a moving image will have a certain degree of optical contrast with its surroundings.

Figure 10.39(*B*) shows that the output signal is compared with two thresholds in the window comparator. The purpose of the comparator is to convert the analog signal *V* into two logic levels: ø, no motion detected and 1, motion is detected.

To increase area of coverage, an array of detectors can be placed in the focal plane of a focusing mirror or lens. Each individual detector covers a narrow field of view, while in combination they protect larger areas. All detectors in the array should either be multiplexed or otherwise interconnected to produce a combined detection signal; that is, they can be made into a complex sensor shape. An alternative solution is the use of a multiple-element focusing system.

Complex Sensor Shape

If the detector's surface area is sufficiently large to cover an entire angle of view, it may be optically broken into smaller elements, thus creating an equivalent of a multiple detector array. To break the surface area into several parts, one can shape the sensing element in an odd pattern, like the interdigitized shape shown if Figure 10.40(*A*) or parallel grid as in Figure 10.40(*B*). Each part of the sensing element acts as a separate light detector.

The parallel or serially connected detectors generate a combined output signal, for example, voltage *v,* when the image of the object moves along the element surface crossing alternatively sensitive and nonsensitive areas. This results in an alternate signal *v* at the detector terminals. Each sensitive and nonsensitive area must be sufficiently large to overlap with most of the object's image. An alternative solution to the complex shape of the sensor is use of the image distortion mask as shown in Figure 10.40(*C*); however, this solution requires a larger sensor surface area.

Facet Focusing Element

A cost-effective way of broadening the field of view while employing a small-area detector is to use multiple focusing devices. A focusing mirror or a lens may be divided into an array of smaller mirrors or lenses called facets, just like in the eye of a fly. Each facet creates its own image resulting in multiple images, as shown in Figure 10.41. When the object moves, the images also move across the element,

© 1999 by CRC Press LLC

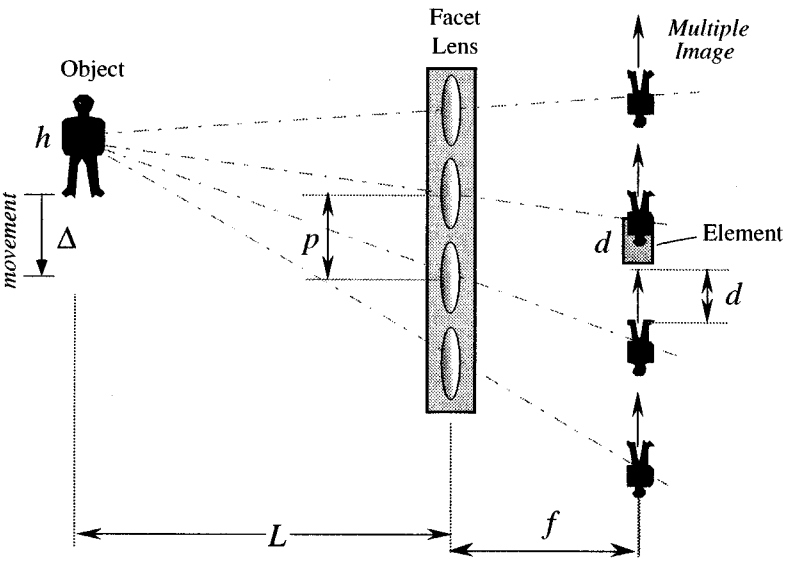

FIGURE 10.41 Facet lens creates multiple images near the sensing element.

resulting in an alternate signal. By combining multiple facets, it is possible to create any desirable detecting pattern in the field of view, in both horizontal and vertical planes. Positioning of the facet lens, focal distances, number, and a pitch of the facets (a distance between the optical axes of two adjacent facets) can by calculated in every case by applying rules of geometrical optics [3]. In the far-infrared spectral range (thermal radiation sensors), the polyethylene facet Fresnel lenses are used almost exclusively, thanks to their low cost and relatively high efficiency.

For the visible portion of the optical spectrum, a simple, very inexpensive, yet efficient motion detector can be developed for nondemanding applications, like light control or interactive toys, using simple photoresistors and pinhole lenses [3, 6, 7].

Far-Infrared Motion Detectors

A motion detector that perceives electromagnetic radiation that is naturally emitted by any object operates in the optical range of thermal radiation, also called far-infrared (FIR). Such detectors are responsive to radiative heat exchange between the sensing element and the moving object. The principle of thermal motion detection is based on the physical theory of emission of electromagnetic radiation from any object whose temperature is above absolute zero (see Chapter 32, Section 6, on *Infrared Thermometers*).

For IR motion detection, it is essential that a surface temperature of an object be different from that of the surrounding objects, so a thermal contrast would exist. All objects emanate thermal radiation from their surfaces and the intensity of that radiation is governed by the Stefan–Boltzmann law. If the object is warmer than the surroundings, its thermal radiation is shifted toward shorter wavelengths and its intensity becomes stronger. Many objects whose movement is to be detected are nonmetals, hence they radiate thermal energy quite uniformly within a hemisphere. Moreover, the dielectric objects generally have a high emissivity. Human skin is one of the best emitters, with emissivity over 90%, while most fabrics also have high emissivities, between 0.74 and 0.95 [3]. Below, two types of far-infrared motion detectors are described. The first utilizes a passive infrared (PIR) sensor, while the second has active far-infrared (AFIR) elements.

PIR Motion Detectors

These detectors became very popular for security and energy management systems. The PIR sensing element must be responsive to far-infrared radiation within a spectral range from 4 μm to 20 μm where most of the thermal power emanated by humans is concentrated. There are three types of sensing elements

© 1999 by CRC Press LLC

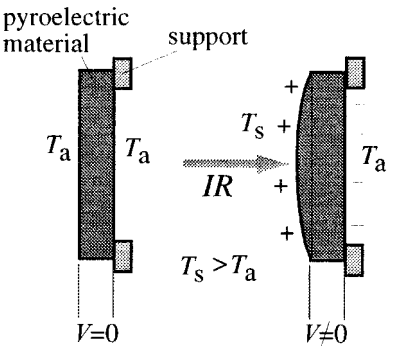

FIGURE 10.42 A simplified model of a pyroelectric effect as a secondary effect of piezoelectricity. Initially, the element has a uniform temperature (*A*); upon exposure to thermal radiation, its front side expands, causing a stress-induced charge (*B*).

that are potentially useful for that detector: thermistors, thermopiles, and pyroelectrics; however, pyroelectric elements are used almost exclusively for the motion detection thanks to their simplicity, low cost, high responsivity, and broad dynamic range. A pyroelectric effect is described in Chapter 32, Section 7 on *Pyroelectric Thermometers*. How this effect can be used in practical sensor design is discussed here.

A pyroelectric material generates an electric charge in response to thermal energy flow through its body. In a simplified way it may be described as a secondary effect of thermal expansion (Figure 10.42). Since all pyroelectrics are also piezoelectrics, the absorbed heat causes the front side of the sensing element to expand. The resulting thermally induced stress leads to the development of a piezoelectric charge on the element electrodes. This charge is manifested as voltage across the electrodes deposited on the opposite sides of the material. Unfortunately, the piezoelectric properties of the element also have a negative effect. If the sensor is subjected to a minute mechanical stress due to any external force, it also generates a charge that in most cases is indistinguishable from that caused by the infrared heat waves. Sources of such mechanical noise are wind, building vibrations, loud sound, etc.

To separate thermally induced charges from the piezoelectrically induced charges, a pyroelectric sensor is usually fabricated in symmetrical form (Figure 10.43(*A*)). Two identical elements are positioned inside the sensor's housing. The elements are connected to the electronic circuit in such a manner as to produce the out-of-phase signals when subjected to the same in-phase inputs. The idea is that interferences produced by, for example, the piezoelectric effect or spurious heat signals are applied to both electrodes simultaneously (in phase) and thus will be canceled at the input of the circuit, while the variable thermal radiation to be detected will be absorbed by only one element at a time, thus avoiding a cancellation.

One way to fabricate a differential sensor is to deposit two pairs of electrodes on both sides of a pyroelectric element. Each pair forms a capacitor that can be charged either by heat or by mechanical stress. The electrodes on the upper side of the sensor are connected together forming one continuous electrode, while the two bottom electrodes are separated, thus creating the opposite-serially connected capacitors. Depending on the side where the electrodes are positioned, the output signal will have either a positive or negative polarity for the thermal influx. In some applications, a more complex pattern of the sensing electrodes is required (for example, to form predetermined detection zones), so that more than one pair of electrodes is needed. In such a case, for better rejection of the in-phase signals (common mode rejection), the sensor should still have an even number of pairs where positions of the pairs alternate for better geometrical symmetry. Sometimes, such an alternating connection is called an interdigitized electrode.

A differential sensing element should be mounted in such a way as to ensure that both parts of the sensor generate the same signal if subjected to the same external factors. At any moment, the optical component must focus a thermal image of an object on the surface of one part of the sensor only, which is occupied by a single pair of electrodes. The element generates a charge only across the electrode pair that is subjected to a heat flux. When the thermal image moves from one electrode to another, the current

© 1999 by CRC Press LLC

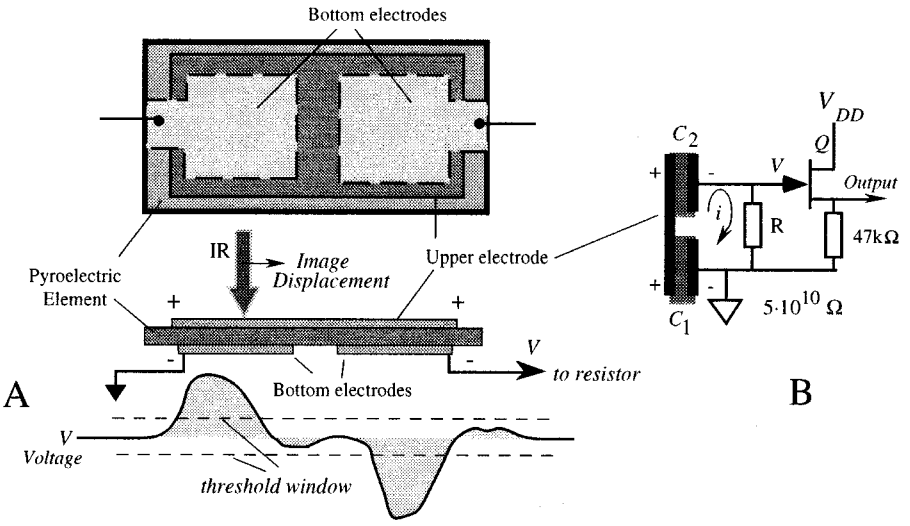

FIGURE 10.43 Dual pyroelectric sensor. (*A*) A sensing element with a front (upper) electrode and two bottom electrodes deposited on a common crystalline substrate. (*B*) A moving thermal image travels from the left part of the sensor to the right, generating an alternating voltage across bias resistor, *R*.

i flowing from the sensing element to the bias resistor *R* (Figure 10.43(*B*)) changes from zero, to positive, then to zero, to negative, and again to zero (Figure 10.43(*A*) lower portion). A JFET transistor *Q* is used as an impedance converter. The resistor *R* value must be very high. For example, a typical alternate current generated by the element in response to a moving person is on the order of 1 pA (10^{-12} A). If a desirable output voltage for a specific distance is $v = 50$ mV, then according to Ohm's law, the resistor value is $R = v/i = 50$ GΩ (5×10^{10} Ω). Such a resistor cannot be directly connected to a regular electronic circuit; hence, transistor *Q* serves as a voltage follower (the gain is close to unity). Its typical output impedance is on the order of several kilohms.

The output current *i* from the PIR sensor can be calculated on the basis of the Stefan–Boltzmann law as [3]:

$$i \approx \frac{2P\sigma a\gamma}{\pi hc} bT_s^3 \frac{\Delta T}{L^2} \tag{10.113}$$

where $\Delta T = (T_b - T_a)$ is the temperature gradient between the object and its surroundings, *P* is the pyroelectric coefficient, σ is the Stefan-Boltzmann constant, *a* is the lens area, γ is the lens transmission coefficient, *h* is the thickness, and *c* is the specific heat of the pyroelectric element, respectively, and *L* is the distance to the object.

There are several conclusions that can be drawn from Equation 10.113. The first part of the equation (the first ratio) characterizes a detector, while the rest relates to an object. The pyroelectric current *i* is directly proportional to the temperature difference (thermal contrast) between the object and its surroundings. It is also proportional to the surface area of the object that faces the detector. A contribution of the ambient temperature T_a is not as strong as it might appear from its third power. The ambient temperature must be entered in kelvin, hence its variations become relatively small with respect to the scale. The thinner the sensing element, the more sensitive the detector. The lens area also directly affects signal magnitude. On the other hand, pyroelectric current does not depend on the sensor's area as long as the lens focuses an entire image on a sensing element.

AFIR Motion Detectors

The AFIR motion detector is a new class of thermal sensors whose operating principle is based on balancing thermal power supplied to the sensing element [8, 9]. Contrary to a passive motion detector

© 1999 by CRC Press LLC

that absorbs thermal radiation from a warmer object, an AFIR motion detector is active; that is, it radiates heat waves *toward* the surroundings. The sensor's surface temperature (T_s) is maintained somewhat above ambient. The element is combined with a focusing system, very much like the PIR detector; however, the function of that system is inverse to that of the passive detectors. A focusing part in the AFIR detector projects a thermal image of the warm sensing element into its surroundings. The AFIR sensors have a significant advantage over the PIR: immunity against many interferences (such as RFI and microphonics).

The output voltage from the AFIR motion detector can be described by the following equation [3]:

$$\Delta V \approx -\frac{R}{V_0}\frac{\sigma a \gamma}{\pi} b T_s^3 \frac{\Delta T}{L^2} \tag{10.114}$$

where R is the resistance of the sensor's heater and V_0 is the heating voltage. The minus sign indicates that for warmer moving objects, the output voltage decreases. There is an obvious similarity between Equations 10.113 and 10.114; however, sensitivity (detection range) of the AFIR sensor can be easily controlled by varying R or V_0. For better sensitivity, the temperature increment above ambient can be maintained on a fairly low level. Practically, the element is heated above ambient by only about 0.2° C.

References

1. S. Blumenkrantz, *Personal and Organizational Security Handbook,* Government Data Publications, Washington, D.C.: 1989.
2. P. Ryser and G. Pfister, Optical fire and security technology: sensor principles and detection intelligence, *Transducers'91. Int. Conf. Solid-State Sensors Actuators,* 1991, 579-583.
3. J. Fraden, *Handbook of Modern Sensors,* 2nd ed., Woodburg, NY: AIP Press, 1997.
4. N. M. Calvin, *Capacitance proximity sensor.* U.S. Patent No. 4,345,167, 1982.
5. J. Fraden, *Apparatus and method for detecting movement of an object,* U.S. Patent No. 5,019,804, 1991.
6. J. Fraden, *Motion discontinuance detection system and method.* U.S. Patent No. 4,450,351, 1984.
7. J. Fraden, *Toy including motion-detecting means for activating same.* U.S. Patent No. 4,479,329, 1984.
8. J. Fraden, *Active infrared motion detector and method for detecting movement.* U.S. Patent No. 4,896,039, 1990.
9. J. Fraden, Active far infrared detectors, in *Temperature. Its Measurement and Control in Science and Industry,* Vol. 6, Woodburg, NY: American Institute of Physics, 1992, Part 2, 831-836.

© 1999 by CRC Press LLC