

INTELLIGENT



TRANSPORTATION



SYSTEMS



# Vehicle Location and Navigation Systems

Yilin Zhao

# **VEHICLE LOCATION AND NAVIGATION SYSTEMS**

**Yilin Zhao**

**Artech House, Inc.  
Boston • London**

Library of Congress Cataloging-in-Publication Data

Zhao, Yilin.

Vehicle location and navigation systems /Yilin Zhao.

p. cm.

Includes bibliographical references and index.

ISBN 0-89006-861-5 (alk. paper)

1. Intelligent Vehicle Highway Systems. 2. Motor vehicles—Automatic location systems. 3. Electronics in navigation.

I. Title.

TE228.3Z45 1997

629.2'7—dc21

97-4200

CIP

British Library Cataloguing in Publication Data

Zhao, Yilin

Vehicle location and navigation systems

1. Intelligent Vehicle Highway Systems 2. Motor vehicles—Automatic location systems

I. Title

625.7'94

ISBN 0-89006-861-5

Cover design by Jennifer Makower

© 1997 ARTECH HOUSE, INC.

685 Canton Street

Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

International Standard Book Number: 0-89006-861-5

Library of Congress Catalog Card Number: 97-4200

10 9 8 7 6 5 4 3

To everyone who

information, the attributes, as well as road access. For an on-line map, the addresses are fully addressed. But addresses can be entered into the map does not use database concepts to find a known address?

precise and Completely  
39, No. 2, May 1986,

Bristol, PA: Taylor &

, 1987.

, 1995.

issues: CEN, Oct. 1995.

at Status," *Proc. IEEE*,

July 1995, pp. 368-373.

ards, Version 2.0, Jan.

p Database in Japan,"

VNIS '93), Oct. 1993,

merican Cartographer,

n Map Databases, SAE

Journal, Vol. 1, No. 2,

or Automotive Naviga-

tance (VNIS '89), Oct.

Springer-Verlag, 1988.

MA: Addison-Wesley,

Topological Structures

nt No. 1277043, Nov.

ilips Technical Review,

o. 860125, Society of

D (Called CD Informa-

## CHAPTER 3

▼▼▼

# POSITIONING MODULE

### 3.1 INTRODUCTION

The positioning module is a vital component of any vehicle location and navigation system. To either help users obtain vehicle location or provide users with proper maneuver instructions, the vehicle location must be determined precisely. Therefore, accurate and reliable vehicle positioning is an essential prerequisite for any good vehicle location and navigation system.

Positioning involves determination of the coordinates of a vehicle on the surface of the Earth. Location involves the placement of the vehicle relative to landmarks or other terrain features such as roads. These tasks are accomplished by the positioning module in conjunction with other modules. In this chapter, we introduce the basic positioning technologies and sensors and then study sensor fusion technologies for integrating the sensor inputs used to estimate position and location.

Three positioning technologies are most commonly used: stand-alone, satellite based, and terrestrial radio based. Dead reckoning is a typical stand-alone technology. A common satellite-based technology involves equipping a vehicle with a global positioning system (GPS) receiver. Dead reckoning and GPS technologies have been used widely in vehicles. Therefore, we will concentrate on these two commonly used and easily accessed technologies in this chapter and leave the terrestrial radio technology to a later section (Section 9.2) following our discussion of radio

communications in Chapter 8. Although some of these radio technologies, like GPS, do not require users to construct central host facilities or complicated communications systems and infrastructures, this arrangement will quickly direct readers' attentions to the more popular technologies first. Readers may refer to Table 10.1 for a performance comparison of the main positioning technologies discussed in the book.

As we will learn shortly, no single sensor is adequate to provide position and location information to the accuracy often required by a location and navigation system. The common solution (frequently the only way of obtaining the required levels of reliability and accuracy) is to fuse information from a number of different sensors, each with different capabilities and independent failure modes. Therefore, a positioning module typically integrates multiple sensors, which compensate for one another to meet overall system requirements. This in turn requires that we study a variety of sensors (Figure 3.1), fusion methods, and algorithms. Because a GPS receiver functions like an absolute sensor, for convenience of study we discuss it together with the compass. Bear in mind that this is a satellite-based technology and is based on an operating principle quite different from that of other sensors (as we will learn shortly).

The positioning module is the most fundamental of the various modules needed for vehicle location and navigation. As seen from Figure 3.1, it is based on a variety of different positioning sensors. Because of the key roles played by these sensors, we will discuss commonly used low-cost sensors in depth. Readers who are not interested in the details can skip the in-depth discussions but still should be able to grasp the basic concepts in this chapter. We do not pretend to cover all possible sensors in this chapter. With the rapid advances of technology, it is difficult to cover even a particular type of sensor completely in a limited space since a variety of implementation methods can be used to manufacture a sensor based on the same theoretical principle. A similar difficulty also occurs with sensor integration and fusion technologies. However, the emphasis placed on sensors in current use and

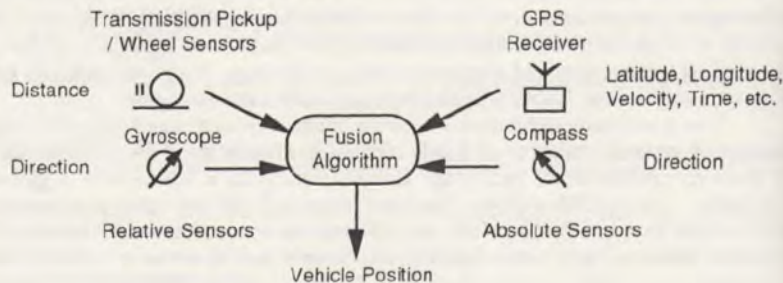


Figure 3.1 A generic positioning module.

the fundamental fusion technology used in vehicle location and navigation. For available linear and two tables summarized in Appendix development and automotive sensor discussion of sensor technologies, be found in [5,6]. In-depth coverage found in [7].

### 3.2 DEAD RECKONING

A very primitive positioning technique is dead reckoning, which is based on a two-dimensional planar space, in any instance provided that starting position is known. Dead reckoning increment of travel relative to a known location to determine vehicle location (or coordinates) requires the direction of the vehicle  $\theta$  and the distance traveled.

The sensors discussed in the previous section have been sampled and integrated and the resulting position  $(x_n, y_n)$  and orientation  $(\theta_n)$  are given by

$$x_n = x_0 + \Delta x_n$$

$$y_n = y_0 + \Delta y_n$$

$$\theta_n = \theta_0 + \Delta \theta_n$$

where  $(x_0, y_0)$  is the initial vehicle position,  $\Delta x_n$  and  $\Delta y_n$  are the magnitude of the displacement vector in the  $x$  and  $y$  directions, and  $\Delta \theta_n$  is the heading of the displacement vector.

When the sampling period is  $\Delta t$ , the changes in vehicle velocity, the above equations can be written as

$$x_n = x_0 + \Delta x_n$$

$$y_n = y_0 + \Delta y_n$$

the fundamental fusion technologies available should be sufficient for anyone interested in vehicle location and navigation to have a basic understanding of this particular area. For available linear and angular automotive sensor technologies, see the two tables summarized in Appendix C. Reviews of the latest automotive sensor development and automotive sensor technologies can be found in [1–4]. A detailed discussion of sensor technologies, sensor principles, and sensor interface circuits can be found in [5,6]. In-depth coverage of sensor integration and fusion can be found in [7].

### 3.2 DEAD RECKONING

A very primitive positioning technique is dead reckoning. For vehicles traveling in a two-dimensional planar space, it is possible to calculate the vehicle position at any instance provided that starting location and all previous displacements are known. Dead reckoning incrementally integrates the distance traveled and direction of travel relative to a known location. In short, it is a technique that determines the vehicle location (or coordinates) relative to a reference point.

The sensors discussed in the following sections can be used to measure the direction of the vehicle  $\theta$  and the distance traveled  $d$ . After the sensor data have been sampled and integrated and sensor fusion has been performed, the vehicle position  $(x_n, y_n)$  and orientation  $(\theta_n)$  at time  $t_n$  can be calculated from the equation

$$\begin{aligned}x_n &= x_0 + \sum_{i=0}^{n-1} d_i \cos \theta_i \\y_n &= y_0 + \sum_{i=0}^{n-1} d_i \sin \theta_i \\ \theta_n &= \sum_{i=0}^{n-1} \omega_i\end{aligned}$$

where  $(x_0, y_0)$  is the initial vehicle position at time  $t_0$ ,  $d_i$  is the distance traveled or the magnitude of the displacement between time  $t_{n-1}$  and time  $t_n$ ,  $\theta_i$  is the direction (heading) of the displacement vector, and  $\omega_i$  is the angular velocity for the same time period. A drawing outlining this method is shown in Figure 3.2.

When the sampling period is constant (and short relative to the time scale for changes in vehicle velocity), the above equation can be written as

$$\begin{aligned}x_n &= x_0 + \sum_{i=0}^{n-1} v_i T \cos(\theta_i + \omega_i T) \\y_n &= y_0 + \sum_{i=0}^{n-1} v_i T \sin(\theta_i + \omega_i T)\end{aligned}$$



### 3.3.1 Transmission Pickups

Transmission pickup sensors are used to measure the angular position of the transmission shaft. Different technologies, such as variable reluctance, the Hall effect, magneto-resistance, and optically based technologies, are used to convert the mechanical motion into electronic signals. Although some transmission pickup sensors may have analog output, it is not difficult to design an interface to convert the signal to pulses indicating fractional revolution. Knowing the number of pulse counts per revolution and a proper conversion scale factor, one can convert the output of the sensor into distance traveled. To avoid mechanical wear and corresponding changes in accuracy of the measurement, it is desirable to measure the angular position of the shaft with a noncontact sensor. Because there is always an air gap between the sensor and the component being monitored, a noncontact sensor is not subject to friction wear. Magnetic fields and optical methods are the two most common methods used for noncontact coupling to a rotating shaft. In this section, we discuss only the magnetic field method.

There are two common places in the vehicle where the signal from the transmission pickup sensor can be obtained. One is from a connector behind the speedometer. The other is from a connector on the transmission housing. In general, a two-wire connector indicates a variable-reluctance position sensor and a three-wire connection indicates a Hall-effect position sensor.

These variable-reluctance and Hall-effect sensors are not limited to transmission pickups. They can be used to detect the position and speed of rotating toothed or notched wheels in crankshaft-, camshaft-, and wheel-monitoring applications. Some of these applications are discussed in the next section.

#### Variable-Reluctance Position Sensor

The principle of the variable-reluctance sensor is that the total electromagnetic force (emf) induced in a closed circuit is equal to the time rate of decrease in the total magnetic flux linking the circuit. The total magnetic flux (number of lines of magnetic force) is inversely proportional to total reluctance. Any change in the reluctance of the magnetic circuit will cause a change in the magnetic flux. If the changing flux lines pass through a coil, the change in flux will induce an emf on the coil. The output emf is measured as a voltage proportional to the rate of change of the reluctance for the path as shown in the following equation:

$$e = - \frac{d\Phi}{dt}$$

where  $d\Phi/dt$  is the rate of change in the magnetic flux in webers per second, and  $e$  is the emf in volts. As a result, the output voltage goes to zero as the rate of change

approaches zero. Therefore, variable-reluctance sensors cannot be used at speeds near zero. Additional detail on magnetic field measurements and magnetic flux is provided in Section 3.4.1.

The variable-reluctance position sensor consists of a permanent magnet with a coil of wire wound around it (Figure 3.3). A ferrous wheel or iron/steel disk mounted on the rotating shaft has teeth or tabs that pass between the pole pieces of this magnet. The voltage output from the sensor is typically a sine wave. When a tooth on the ferrous wheel begins to pass between the magnet pole pieces, the coil voltage starts increasing from zero, reaches a maximum, and falls to zero again when the tooth is exactly between the pole pieces. At this point the rate of change in the magnetic flux is zero (even though the magnetic flux is at its maximum value). Therefore, the induced voltage across the sensing coil goes to zero. As the tooth continues its movement, the voltage output from the coil once again increases from zero, reaches a maximum, and falls to zero again as the tooth passes out of the gap between the pole pieces, and thus finishes a complete sine cycle.

Despite their simple, rugged construction and low cost, variable-reluctance sensors do have drawbacks [2,8]. First, it is difficult to output a constant voltage envelope. As the speed of the ferrous wheel changes, the voltage level and frequency change as well. Second, the signal-to-noise ratio can be seriously degraded by vibrations or resonance. Third, the output of the sensor is sensitive to the gap between the sensor and the disk. In other words, the output is inversely proportional to the gap size. Proper alignment and a rigid sensor mounting are required. Fourth, there is no sensor output when the disk rotation speed is below a certain threshold. This means that no output is available at very low speeds. Fifth, the range of this sensor tends to flatten at high speeds. The extent of this effect varies based on the sensor design. Finally, electromagnetic interference or radio-frequency interference can

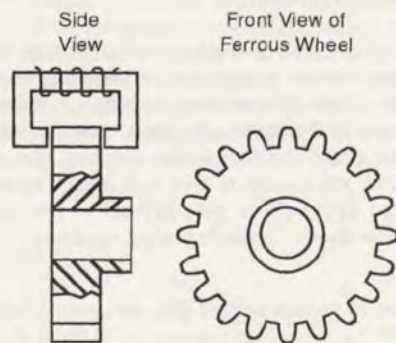


Figure 3.3 Variable-reluctance position sensor.

introduce false signals if the sensor is not properly aligned. The coil in the reluctance sensor has a rather high inductance, which can be added to the sensor package to obtain a more stable output and reduce the interference.

#### Hall-Effect Position Sensor

The Hall effect is the effect by which a voltage is induced in a conductor in a magnetic field. It is based on the interaction between the magnetic field and the current in a conductor. When a current-carrying conductor is placed in a magnetic field, the current (which consists of moving electrons) will experience a force in a direction perpendicular to both the direction of motion and the magnetic field. A voltage can be detected. Hall-effect devices can be made from various materials.

A Hall-effect position sensor with a ferrous wheel is shown in Figure 3.4. The Hall-effect device is assembled on a circuit board. The Hall-effect device actually lies at the end of a probe that it always lies between the magnet and the ferrous wheel. As the ferrous wheel passes through this device, a voltage develops. The direction of current flow and to the direction of rotation of the wheel rotates, the reluctance of the magnetic field is changed.

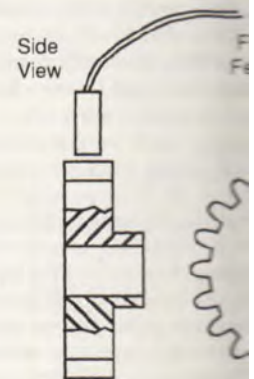


Figure 3.4 Hall-effect position sensor.

used at speeds  
magnetic flux is

at magnet with  
iron/steel disk  
the pole pieces  
the wave. When  
pieces, the coil  
to zero again  
rate of change  
maximum value).  
As the tooth  
increases from  
out of the gap

variable-reluctance  
constant voltage  
and frequency  
graded by vibra-  
the gap between  
portional to the  
d. Fourth, there  
threshold. This  
ge of this sensor  
ed on the sensor  
interference can

introduce false signals if the sensor is not properly shielded or packaged, since the coil in the reluctance sensor has a rather high impedance. Signal processing circuitry can be added to the sensor package to obtain a high signal level and to minimize the interference.

### *Hall-Effect Position Sensor*

The Hall effect is the effect by which a voltage is generated in a conductor moving in a magnetic field. It is based on the interaction between moving electric carriers in a conductor and an external magnetic field. As a conductor is moved through a magnetic field, the current (which consists of moving electric charges) generated in the conductor will experience a force in a direction perpendicular to both the direction of motion and the magnetic field. A voltage due to the movement of these charges can be detected. Hall-effect devices can be manufactured in silicon or semiconductor materials.

A Hall-effect position sensor with a ferrous wheel or iron/steel disks is shown in Figure 3.4. The Hall-effect device is assembled into a probe with a biasing magnet and a circuit board. Figure 3.4 does not show the configuration inside the probe. The Hall-effect device actually lies at the end of the probe, close to the wheel, so that it always lies between the magnet and the wheel. When an external current is passed through this device, a voltage develops across the device perpendicular to the direction of current flow and to the direction of magnetic flux. As the ferrous wheel rotates, the reluctance of the magnetic field changes as the teeth or tabs pass the probe.

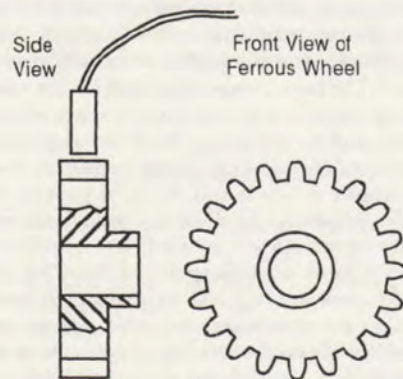


Figure 3.4 Hall-effect position sensor.

Unlike the variable-reluctance sensor, the Hall-effect sensor is a zero-speed sensor. No matter what the velocity, once the Hall-effect device lines up with the strongest magnetic field, there will be a maximum waveform output. Therefore, the voltage output (square wave) from this sensor is independent of the speed of the rotating shaft.

Sensors that utilize Hall-effect technology also have some limitations. The response degrades as the upper and lower limits of operating temperature are approached. They cannot be used in severe temperatures, lower than  $-40^{\circ}\text{C}$  ( $-40^{\circ}\text{F}$ ) or higher than  $65.6^{\circ}\text{C}$  ( $150^{\circ}\text{F}$ ). The presence of other close, strong magnetic fields can cause errors. As in the case of the reluctance sensor, Hall-effect sensors require careful alignment and mounting because of sensitivity to the gap size between sensor and disk. Unlike the passive reluctance sensor, the Hall-effect sensor is an active sensor which requires an external electrical input. In addition, unlike the variable reluctance sensor, the output voltage of the Hall-effect sensor remains constant as the disk rotation frequency is varied. This sensor can easily be combined with a signal processing circuit to output a high-level voltage. In general, it is more desirable to use a Hall-effect sensor than a variable-reluctance sensor for transmission pickup sensors or other position sensors used in vehicle location and navigation systems. Further discussion of these and other automotive sensors can be found in [2,4].

### 3.3.2 Wheel Sensors

#### *Wheel Sensors for Anti-Lock Braking Systems*

A typical anti-lock braking system (ABS) consists of wheel speed sensors, an electronic control unit, a brake pressure modulator, wiring, relays, hydraulic tubing, and connectors to link the whole system together. Most ABS systems currently in use offer four-wheel control. The basic components used for the wheel speed measuring systems include a sensing element, a ferrous wheel (exciter wheel, target wheel, tone wheel, or iron/steel disk), and the mechanical hardware required to mount the above elements with the appropriate spacings to ensure correct operation.

The variable-reluctance sensor is still the most popular for ABSs because of its low cost and relative reliability. As discussed earlier, this sensor is not reliable at low speeds, especially below 0.45 to 1.34 m/s (1 to 3 miles/hr); this makes accurate vehicle positioning very difficult when using these sensors for input to a positioning module. Hall-effect position sensors can also be used as ABS speed sensors. We have learned that both sensors are electromagnetic pulse pickups using toothed wheels (exciter rings) mounted directly on the rotating components of the drive train, shaft, or wheel hubs. They provide a digital signal whose frequency is proportional to the rotational velocity of the wheel. Furthermore, any technology with a sensing element that converts mechanical motion into an electrical signal can also be used as a wheel

sensor. Table 3.1 compares several different sensors. (The items in this table are from [9]). From the above, it is clear that the reluctance sensor is fundamentally inaccurate. The only exception is that a Hall-effect sensor can be used in a feedback circuit that has a capacitor).

#### *Differential Odometer*

Differential odometry is a technique to determine the change in position by integrating the change in distance traveled with respect to an initial position. The number of rotation counts (pulses) generated by the left and right wheel rotation counts and the distance traveled by the vehicle to the next position can be used to obtain the distance traveled by the axle length can be used to obtain the vehicle position. Both the distance traveled and the heading can be used to obtain the vehicle position.

One example of a differential odometer is a Hall-effect sensor (toothed gear) is attached to a ferrous wheel (because acceleration and deceleration on nondriven wheels). A Hall-effect sensor outputs a pulse for each tooth that passes the sensor. One for the left wheel and one for the right wheel. The right tooth count can be translated to distance by multiplying the count by the scale factor (distance per count).

Table 3.1  
Wheel Speed Sensors

| Sensor              | Frequency Range |
|---------------------|-----------------|
| Eddy current        | 0-500 kHz       |
| Hall effect         | 0-1 MHz         |
| Magnetic transistor | 0-500 kHz       |
| Magnetoresistive    | 0-5 MHz         |
| Optical             | 0-10 MHz        |
| Reed switch         | 0-600 Hz        |
| Variable reluctance | 1-100 kHz       |
| Wiegand effect      | 0-20 kHz        |

zero-speed  
up with the  
erefore, the  
peed of the

ations. The  
erature are  
0°C (-40°F)  
agnetic fields  
sors require  
ween sensor  
is an active  
the variable  
constant as  
ined with a  
ore desirable  
ssion pickup  
ion systems.  
d in [2,4].

an electronic  
tubing, and  
rently in use  
ed measuring  
t wheel, tone  
unt the above

ss because of  
s not reliable  
akes accurate  
a positioning  
sors. We have  
othed wheels  
re train, shaft,  
ortional to the  
nsing element  
sed as a wheel

sensor. Table 3.1 compares several different wheel sensor technologies (most of the items in this table are from [9]). From this table we notice that only the variable-reluctance sensor is fundamentally incapable of measuring at zero speed (another exception is that a Hall-effect sensor cannot function as a zero-speed sensor if its feedback circuit has a capacitor).

### Differential Odometer

Differential odometry is a technique to provide both distance traveled and heading change information by integrating the outputs from two odometers, one each for a pair of front or rear wheels. An odometer is a relative sensor that measures distance traveled with respect to an initial position. A wheel odometer typically measures the number of rotation counts (pulses) generated by a rotating wheel. Averaging the left and right wheel rotation counts and multiplying by a proper scale factor enables the distance traveled by the vehicle to be determined. The difference between the counts for the left and right wheels multiplied by the same scale factor and divided by the axle length can be used to obtain the change in the heading of the vehicle. Both the distance traveled and the heading change can be used in calculating the vehicle position.

One example of a differential odometer is shown in Figure 3.5. A ferrous wheel or iron/steel disk (toothed gear) is attached to each of the nondriven wheels of the vehicle (because acceleration and deceleration have less effect on the output of sensors on nondriven wheels). A Hall-effect sensor is located close to the toothed gear. This sensor outputs a pulse for each tooth that passes across its sensing tip. Two counters, one for the left wheel and one for the right wheel, count these pulses. This left or right tooth count can be translated to the distance each wheel has traveled by multiplying the count by the scale factor (which is a predetermined value for the distance per count).

Table 3.1  
Wheel Speed Sensor Comparison

| Sensor              | Frequency Range | Magnet | Ferrous Wheel | Signal/Noise  |
|---------------------|-----------------|--------|---------------|---------------|
| Eddy current        | 0-500 kHz       | No     | No            | High          |
| Hall effect         | 0-1 MHz         | Yes    | Yes           | Moderate      |
| Magnetic transistor | 0-500 kHz       | Yes    | Yes           | Moderate/high |
| Magnetoresistive    | 0-5 MHz         | Yes    | Yes           | Moderate      |
| Optical             | 0-10 MHz        | No     | No            | Very high     |
| Reed switch         | 0-600 Hz        | Yes    | Yes           | High          |
| Variable reluctance | 1-100 kHz       | Yes    | Yes           | Low           |
| Wiegand effect      | 0-20 kHz        | Yes    | Yes           | High          |

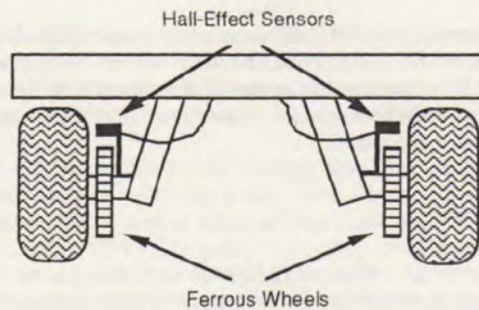


Figure 3.5 Differential odometer.

After resetting the counters and driving the vehicle along a known straight distance  $D$  while recording the counter numbers, the left and right scale factors can be calculated from

$$K_L = \frac{D}{C_L}$$

$$K_R = \frac{D}{C_R}$$

where  $C_L$  and  $C_R$  are the number of counts for the left wheel and right wheel, respectively. Knowing the scale factor for each wheel and the previous measurement at  $t-1$ , we can easily derive the distance traveled up to time  $t$  as follows:

$$d(t) = d(t-1) + \frac{K_L C_L + K_R C_R}{2}$$

When the vehicle heading is changed, the outer wheel travels farther than the inner wheel by  $|K_L C_L - K_R C_R|$ . Knowing the distance  $L$  between the left and right wheels (axle length), the vehicle heading can be obtained from this equation:

$$\theta(t) = \theta(t-1) + \frac{|K_L C_L - K_R C_R|}{L}$$

Note that as long as the ratio  $K_L/K_R \approx 1$ , we know that the vehicle is traveling on a straight-line segment road (provided  $K_L$  and  $K_R$  are accurate). On the other hand, if we know that the vehicle is traveling on a straight-line segment based on a given digital map, we can use this information to calibrate these two scale factors

dynamically. Moreover, the outputs (Section 3.4.2) can be used to calibrate the travels above a certain speed and the track fix is good, the velocity and heading info very good. The travel distance extracted calibrate the scale factor. Now we can use sensors in a location and navigation system.

Depending on the technology used, it has an operational error similar to that of effect sensors in general. In addition, the dynamic and nonsystematic errors such as unequal and the finite sampling rate and count resolution modes involving one or more wheels, etc. A vehicle that decelerates too rapidly or travels on a surface with incorrect calculated distance and heading. The contact point between each wheel and the road surface between the left and right wheels will be used in the preceding equation to derive the heading change over time, as may the wheel diameter change in temperature. Therefore, the scale factors must be recalibrated after initial calibration. These factors contribute to the error of the differential odometer, which lead to accumulative error.

### 3.3.3 Gyroscopes

Rate-sensing gyroscopes measure angular velocity and determine attitude. At the present time, most gyroscopes are used to measure the angular rate.

Gyroscopes include mechanical, optical, and pneumatic types. A comparison of these gyroscope types is presented in Table 3.3.1. A gyroscope is classified as a pneumatic gyroscope if it uses the explosive development of design and manufacturing.

Table 3.3.1  
Gyroscope Performance

| Type       | Performance |
|------------|-------------|
| Mechanical | Very good   |
| Optical    | Very good   |
| Pneumatic  | Good        |
| Vibration  | Good        |

dynamically. Moreover, the outputs from other sensors (such as GPS, see Section 3.4.2) can be used to calibrate the scale factor. For instance, when the vehicle travels above a certain speed and the tracked satellite geometry used for the position fix is good, the velocity and heading information received from the GPS sensor is very good. The travel distance extracted from the velocity can then be used to calibrate the scale factor. Now we can see the benefit of using multiple positioning sensors in a location and navigation system.

Depending on the technology used, an odometer or differential odometer often has an operational error similar to that experienced by variable-reluctance and Hall-effect sensors in general. In addition, the differential odometer is subject to systematic and nonsystematic errors such as unequal wheel diameter, misalignment of wheels, the finite sampling rate and count resolution, running over objects on road, slips or skids involving one or more wheels, etc. A slip or skid occurs if the vehicle accelerates or decelerates too rapidly or travels on a snowy/icy/wet road. These errors may lead to incorrect calculated distance and heading data. Furthermore, in sharp turns, the contact point between each wheel and the road can change, so that the actual distance  $L$  between the left and right wheels will be different from the one used in the preceding equation to derive the heading. The tire pressure in each wheel may change over time, as may the wheel diameter due to the tire conditions or change in temperature. Therefore, the scale factors used to calculate position may change after initial calibration. These factors contribute to the error sources of the differential odometer, which lead to accumulative errors discussed in Section 3.2.

### 3.3.3 Gyroscopes

Rate-sensing gyroscopes measure angular rate, and rate-integrating gyroscopes measure attitude. At the present time, most location and navigation systems use gyroscopes to measure the angular rate.

Gyroscopes include mechanical, optical, pneumatic, and vibration devices. A comparison of these gyroscope types is presented in Table 3.2. (The gas-rate gyroscope is classified as a pneumatic gyroscope for lack of a better word.) As the explosive development of design and manufacturing technologies continues, the

Table 3.2  
Gyroscope Comparison

| Type       | Performance | Cost      | Size     |
|------------|-------------|-----------|----------|
| Mechanical | Very good   | Expensive | Large    |
| Optical    | Very good   | Moderate  | Moderate |
| Pneumatic  | Good        | Moderate  | Moderate |
| Vibration  | Good        | Cheap     | Small    |

characteristics of each type of gyroscope in Table 3.2 will improve over the years, and other new technologies may mature to the point where they can be used to further expand this table.

The vibration gyroscope is currently the device of choice for many vehicle location and navigation systems. Discussions of other techniques can be found in [5,10,11]. There are four different types of vibration gyroscope based on the bar (or beam) shape: tuning fork, rectangle, triangle, and cylinder as shown in Figure 3.6. A small electronic circuit is commonly used to control the vibration, convert measurements to an angular velocity output, and provide an interface to the user.

A vibration gyroscope measures the Coriolis acceleration—a phenomenon discovered by G. G. de Coriolis. An object moving perpendicular to a rotating frame will cause an observer on the rotating frame to see an apparent acceleration of the object. The Coriolis acceleration is generated by the angular rotation of a vibrating bar or fork. A change in the direction around one axis of a driving element (transducer) induces a vibration in a detection element (transducer) on another axis. Vibrating piezoelectric ceramics are often used as transducers in a vibration gyroscope and are attached to the bars shown in Figure 3.6.

In general, the apparent Coriolis acceleration  $a$  is given by

$$a = 2\omega v \sin \phi$$

where  $\omega$  is the angular velocity of the rotating axes relative to a fixed set of axes,  $v$  is the velocity of the bar, and  $\phi$  is the angle between the vectors  $\omega$  and  $v$ . The

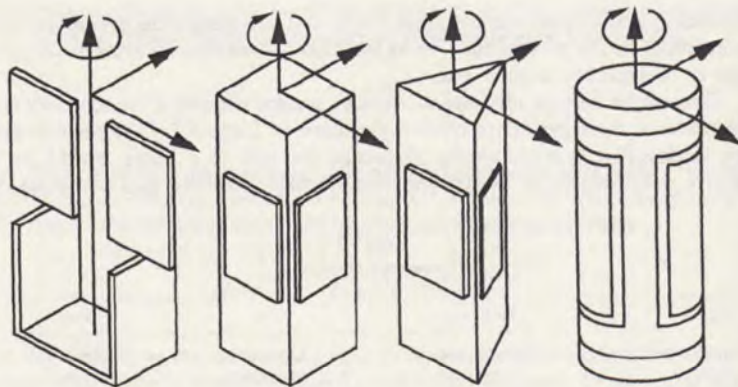


Figure 3.6 Vibration gyroscopes.

acceleration is at right angles to both  $\omega$  and  $v$ . If  $\sin \phi = 1$ , we derive the Coriolis force  $F =$

$$F =$$

Notice that the Coriolis force is proportional to the angular velocity. The Coriolis force is detected and converted to a voltage by a simple circuit. Since the mass of the bar is constant, the Coriolis force is proportional to the angular velocity  $\omega$  may be the output of the gyroscope. This angular velocity is used to produce the change in the vehicle direction.

$$\theta =$$

Gyroscope performance is measured by its accuracy (bias and offset) are major parameters used. The scale factor describes the capability of the gyroscope to convert angular rate into an angular rate. This is done by the gyroscope into an angular rate of small variations in the scale factor. These variations are the gyroscope. Gyroscope drift characterizes all rate measurements to the nominal value on the gyroscope output, so any actual value will cause the angular errors to accumulate. This is a considerable problem between different sensor models, with adjustable parameters  $C_1, C_2, C_3$  for bar vibration gyroscope [12] ( $t$  is the time).

$$\epsilon(t) = C_1(1$$

Recommended methods for reducing the drift or other factors include using an analog-to-digital converter, measuring the offset voltage at zero angular rate, and a sample-and-hold circuit [13,14].

In addition to the methods mentioned, dynamic models to dynamically calibrate a gyroscope. The scale factor involves using other sensors to measure the scale factor. Another approach involves using a known turn maneuver occurs based on the output of calibrating the drift involves calibrating

acceleration is at right angles to both  $\omega$  and  $v$ . Assuming the bar has mass  $m$  and  $\sin \phi = 1$ , we derive the Coriolis force as

$$F = 2m\omega v$$

Notice that the Coriolis force is proportional to  $\omega$  in the preceding equation. This Coriolis force is detected and converted to a voltage by the detector element and a simple circuit. Since the mass of the bar and the vibrational velocity  $v$  of the bar excited by the driving element (and the associated oscillator circuit) are both known quantities, the angular velocity  $\omega$  may easily be derived and provided to the user as the output of the gyroscope. This angular velocity must be sampled and integrated to produce the change in the vehicle direction  $\theta$ :

$$\theta = \int \omega dt$$

Gyroscope performance is measured by a variety of factors. Scale factor and drift (bias and offset) are major parameters used to characterize gyroscope accuracy. The scale factor describes the capability of the gyroscope to accurately sense angular velocity at different angular rates. This value is used to convert the voltage output from the gyroscope into an angular rate or direction. System imperfections can cause small variations in the scale factor. These variations appear as a factor at the output of the gyroscope. Gyroscope drift characterizes the ability of the gyroscope to reference all rate measurements to the nominal zero point. It appears as an additive term on the gyroscope output, so any actual drift away from a given or predetermined value will cause the angular errors to accumulate. The gyroscope drift model varies considerably between different sensor designs. For instance, the following error model, with adjustable parameters  $C_1$ ,  $C_2$ , and  $T$ , has been suggested for a triangle-bar vibration gyroscope [12] ( $t$  is the time since initialization):

$$\epsilon(t) = C_1(1 - e^{-t/T}) + C_2$$

Recommended methods for reducing the gyroscope errors caused by temperature or other factors include using an analog-to-digital (A/D) converter and filters and measuring the offset voltage at zero angular velocity, just before the output of the sample-and-hold circuit [13,14].

In addition to the methods mentioned, users can also build compensatory models to dynamically calibrate a gyroscope. For instance, one method of calibrating the scale factor involves using other sensors to sense the angular rate and modify the scale factor. Another approach involves recalibrating the scale factor whenever a known turn maneuver occurs based on the digital map or other sensors. One method of calibrating the drift involves calibrating the predetermined or manufacturer-given

value whenever the vehicle is stationary. Another method is to correct the value whenever the vehicle is not turning, as determined by other sensors or the map-matching algorithm (Chapter 4).

### 3.4 ABSOLUTE SENSORS

Absolute heading and position sensors are very important in solving location and navigation problems. A relative sensor alone cannot provide an absolute direction or position with respect to a reference coordinate system. In contrast, an absolute sensor can provide information on the position of the vehicle with respect to the Earth. The most commonly used technologies for providing absolute position information are the magnetic compass and GPS.

#### 3.4.1 Magnetic Compasses

A magnetic compass measures the Earth's magnetic field. When used in a positioning system, a compass measures the orientation of an object (such as a vehicle) to which the compass is attached. The orientation is measured with respect to magnetic north. Magnetic field sensors are also known as magnetometers. The intensity of a magnetic field can be measured by the magnetic flux density  $B$ . The units of measurement are the tesla (T), the gauss (G), and the gamma ( $\gamma$ ), where 1 tesla =  $10^4$  gauss =  $10^9$  gamma.

The Earth's magnetic field has an average strength of 0.5 G. It can be represented by a dipole that fluctuates in both time and space. The best-fit dipole location is approximately 440 km off center and is inclined approximately 11 degrees to the Earth's rotational axis [5]. The difference between true north and magnetic north is known as *declination*. This difference varies with both time and geographical location. To correct the difference, declination tables are provided on maps or charts for any given locale. For a vehicle location and navigation system, this does not cause a problem as long as all directions used in the system are referred to the same (magnetic or geographic) north.

Because of portability, high vibration durability, and quick response, the electronic compasses have many advantages over conventional compasses. Among the electronic compasses, the fluxgate compass is the most popular one. The sensitivity range of this sensor is from  $10^{-6}$  G to 100 G and the maximum frequency is about 10 kHz [15]. The term *fluxgate* is obtained from the gating action imposed by an ac-driven excitation coil that induces a time-varying *permeability* in the sensor core. Permeability is the property of a magnetizable substance that determines the degree to which it modifies the magnetic flux of the magnetic field in the region occupied by the magnetizable substance. This property in the magnetic field is analogous to conductivity in an electrical circuit. The absolute permeability  $\mu$  of a given material

is a measure of how well it serves as a path for magnetic flux  $H$  applied to a high-permeability material.

where the magnetic flux density  $B$  is an absolute measure of the magnetic field in a circuit. A magnetization curve can be defined as the permeability  $\mu$  is the slope (Figure 3.7).

High-permeability materials have a tendency to reach magnetic saturation. Hysteresis refers to the fact that the magnetic flux lags behind the change in the magnetizing force, the current value of  $H$ , and the history of  $dH/dt$ . This causes the curve to follow a different path when  $H$  is increased or decreased to form a hysteresis loop (Figure 3.7). The increase in magnetic flux becomes important when a material such as a core is forced into saturation. The flux lines of the external field of this saturated core. The fluxgate compass is used to directly measure the strength of a magnetic field.

The permeable core is driven into saturation by a drive coil (drive winding) on the same core. When the core is in saturation, the magnetic flux lines from the drive coil are also drawn into and out of the core.

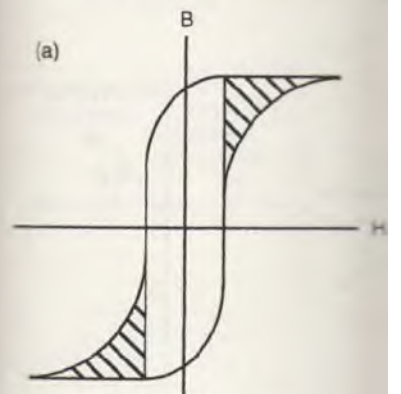


Figure 3.7 Magnetization (hysteresis) curves. (a) typical hysteresis loop. (b) out of saturation.

is a measure of how well it serves as a path for magnetic flux. A magnetic field force  $H$  applied to a high-permeability material induces a magnetic flux density

$$B = \mu H$$

where the magnetic flux density  $B$  is analogous to the current density in an electrical circuit. A magnetization curve can be plotted from the preceding equation, where the permeability  $\mu$  is the slope (Figure 3.7).

High-permeability materials have two interesting properties: hysteresis and saturation. Hysteresis refers to the fact that the magnetic flux through the material lags behind the change in the magnetic field. Due to hysteresis,  $\mu$  depends both on the current value of  $H$ , and the history of previous values for  $H$ , as well as the sign of  $dH/dt$ . This causes the curve to follow a different path as  $B$  increases and decreases to form a hysteresis loop (Figure 3.7). Materials saturate (in other words, further increases in magnetic flux become impossible) at high magnetic flux densities. If a material such as a core is forced into saturation by some additional magnetizing force, the flux lines of the external field will be relatively unaffected by the presence of this saturated core. The fluxgate compass uses the phenomenon of core saturation to directly measure the strength of a surrounding static magnetic field.

The permeable core is driven in and out of saturation by a signal applied to a drive coil (drive winding) on the same core. As the core is driven into and out of saturation, the magnetic flux lines from the external magnetic field being measured are also drawn into and out of the core. These varying flux lines induce positive

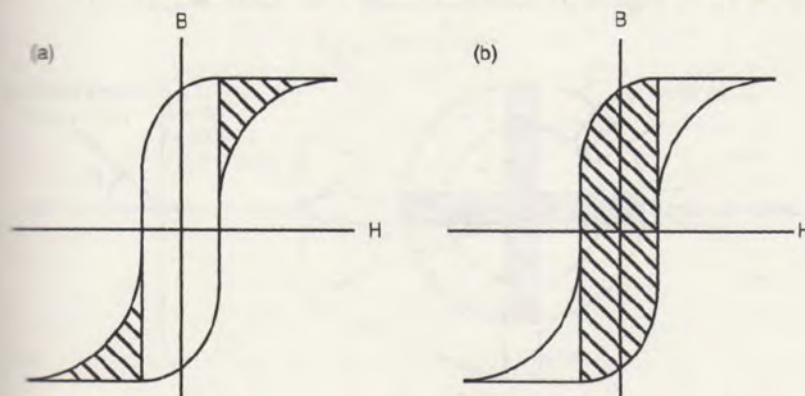


Figure 3.7 Magnetization (hysteresis) curves for high-permeability materials: (a) in saturation and (b) out of saturation.

and negative electrical current spikes in a sense coil (sense winding). Whenever the flux lines are forced out of the core, positive spikes are generated, and vice versa. The amplitude of these spikes is directly proportional to the intensity of the flux vector. The flux vector, whose direction is determined by the polarity of the pulse, runs parallel to the axis of the sense coil. The greater the difference between the saturated and unsaturated states, the more sensitive the compass will be.

Two sensing coils at right angles to each other have been developed to measure all 360 degrees of the Earth's magnetic field. These two orthogonal pickup coils can be configured in a symmetrical fashion around a common core. A toroidal drive winding is used to excite the permeable core into and out of saturation. More complete flux linkage throughout the core can be achieved using a ring geometry, which in turn requires a smaller drive excitation power and reduces interference in the sense coil(s). For these reasons, as well as ease of manufacture, the many low-cost fluxgate compasses use toroidal ring-core configuration. A sketch of such a compass is shown in Figure 3.8.

The integrated dc output voltages  $V_x$  and  $V_y$  from the orthogonal coils vary as sine and cosine functions of  $\theta$ . The instantaneous value of  $\theta$  can be easily derived by performing two successive A/D conversions on these voltages, and taking the arctangent of the quotient:

$$\theta = \arctan \frac{V_y}{V_x}$$

where  $\theta$  is the angle of the sensor unit relative to the Earth's magnetic field.

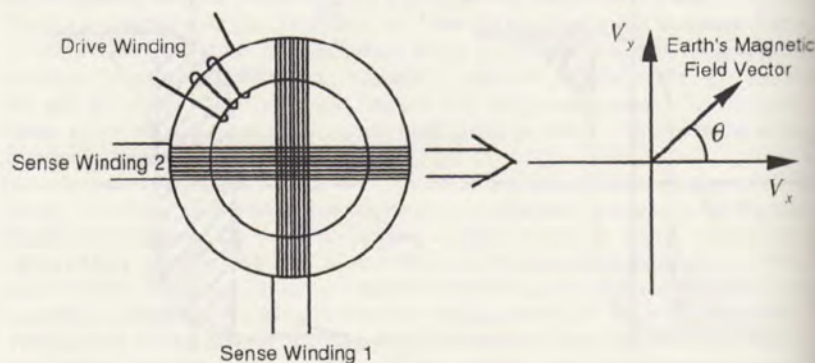


Figure 3.8 Fluxgate compass.

When the compass is installed in a vehicle relative to the magnetic field, the vehicle's own magnetic field and other materials generate their own magnetic field. To measure the vector sum of all magnetic fields, if we ignore the effects of any magnet, the equation just given can be used. Therefore, the equation just given can be used to measure the Earth's magnetic field by subtracting the magnet's field from the measurement.

Determining the magnetic field vector as the vehicle moves and changes its orientation relative to the Earth's magnetic field vector (the vehicle's own magnetic field vector moves with the vehicle's orientation, the compass will remain constant) and a changing Earth's magnetic field (the Earth's magnetic field completes a 360 degree closed loop, the coordinates of the center of the circle will be determined).

Several methods are available for determining the Earth's magnetic field vector. One is to drive the vehicle along a closed loop and measure the minimum values measured by the x and y components generated from an actual trial. The two minimum values should be calculated.

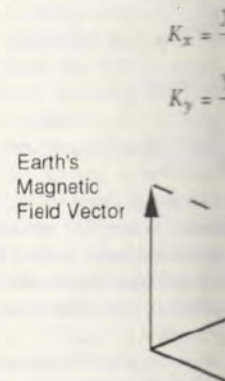


Figure 3.9 Measurement of vector sum.

When the compass is installed in a vehicle it measures the orientation of the vehicle relative to the magnetic field. Any vehicle partially made of iron-based materials generates its own magnetic field, and compasses cannot distinguish between the Earth's magnetic field and other magnetic fields present. In short, compasses measure the vector sum of all magnetic fields present (Figure 3.9 shows the situation if we ignore the effects of any magnetic fields other than the two larger vectors.) Therefore, the equation just given cannot be directly used to derive the direction of the vehicle. The Earth's magnetic field vector may be extracted from the compass measurement by subtracting the magnetic field vector of the vehicle from the compass measurement.

Determining the magnetic field vector of the vehicle is simple. As the vehicle moves and changes its orientation relative to the Earth's magnetic field vector, the vehicle's own magnetic field vector moves with the vehicle. That means that regardless of vehicle orientation, the compass will "see" the sum of the vehicle vector (which remains constant) and a changing Earth vector. If the vehicle is moved through a complete 360 degrees closed loop, the compass output will trace a circle. The coordinates of the center of the circle will describe the vehicle magnetic field vector.

Several methods are available for finding the center of the circle. The simplest one is to drive the vehicle along a closed-loop path and record the maximum and minimum values measured by the x and y coils. Figure 3.10 shows the values generated from an actual trial. The two compass coils might not have identical gain, so the individual gains should be calculated using the following equations:

$$K_x = \frac{X_{\max} - X_{\min}}{2}$$

$$K_y = \frac{Y_{\max} - Y_{\min}}{2}$$

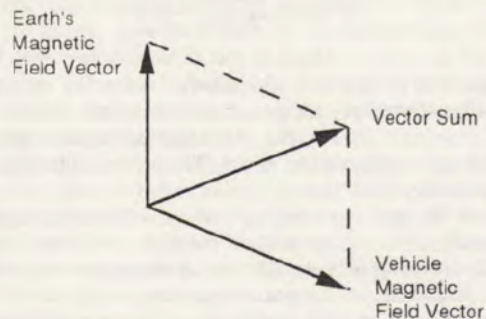


Figure 3.9 Measurement of vector sum.

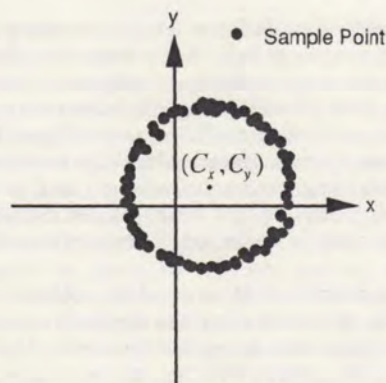


Figure 3.10 Constructing a magnetization circle to determine the magnetic field vector of the vehicle.

The coordinates of the center of the circle are midway between the maximum and minimum points and can be derived from the following equations:

$$C_x = K_x + X_{\min}$$

$$C_y = K_y + Y_{\min}$$

The direction of the vehicle may be computed using:

$$\theta = \arctan \frac{K_y(V_y - C_y)}{K_x(V_x - C_x)}$$

This method is sensitive to magnetic anomalies or noise. For example, if a truck is positioned along the calibration path such that the magnetic field of the truck affects the maximum or minimum values, the calculated coordinates determined for the center of the circle will contain these errors. Therefore, calibration should be performed in a magnetically quiet area.

In general, a fluxgate compass is very sensitive to magnetic anomalies. Figure 3.11 shows direction change data collected in a very noisy magnetic environment for a vehicle traveling in a straight line. For comparison, we also show the direction change data collected simultaneously using a differential odometer. The driver has tried to follow a straight line. However, it is very difficult to control the steering wheel to force the wheels to follow a perfectly straight line. This fact is reflected in the data collected by the odometer. For the compass data, these direction

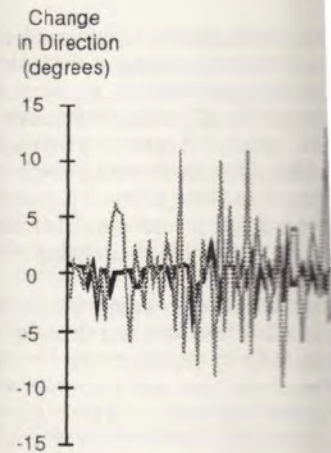


Figure 3.11 Effect of noisy magnetic environment on direction change data.

errors result in positioning errors. This is crucial for any location and navigation system.

Fluxgate compass measurements are affected by short-term magnetic anomalies and long-term anomalies (nearby power lines, big trucks, steel structures, tunnels), reinforced concrete buildings, and long-term magnetic anomalies. Long-term anomalies can be detected during calibration, electrical/magnetic noise can be filtered out.

Filters such as the low-pass filter can be used to remove short-term magnetic anomalies. Section 3.5 and other valid methods can be used to remove long-term magnetic anomalies. These methods involve comparing the measured vector magnitude against the physical open field. When the external field is strong, the measured vector magnitude falls far from the curve defined by the circle. When the external field is weak, the measured vector magnitude falls close to the magnetization circle. The magnitude of errors will be eliminated. Relative to the physical open field (anomalies). When a digital map is available, the direction change data and segment direction are not likely to be affected by the abnormal compass change relative to the abnormal compass change.

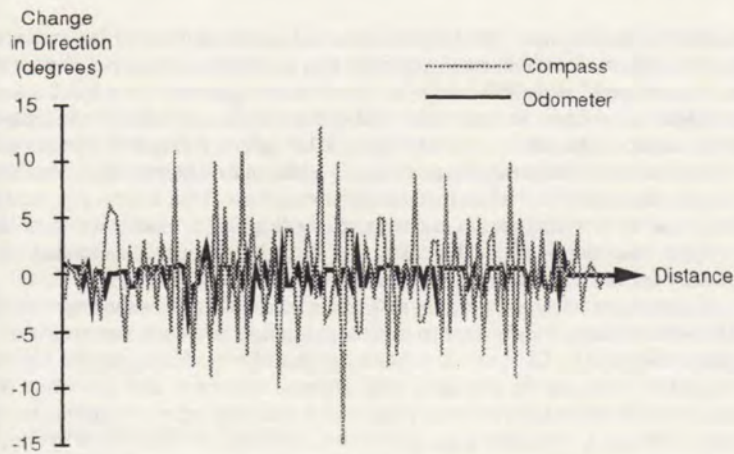


Figure 3.11 Effect of noisy magnetic environment on direction measurements.

errors result in positioning errors. Therefore, minimization of measurement errors is crucial for any location and navigation system using a fluxgate compass.

Fluxgate compass measurements generally contain two types of errors: short-term magnetic anomalies and long-term magnetic anomalies. Deviations caused by nearby power lines, big trucks, steel structures (such as freeway underpasses and tunnels), reinforced concrete buildings, and bridges are typical examples of short-term magnetic anomalies. Long-term magnetic anomalies result from inaccuracies in calibration, electrical/magnetic noise, or magnetization of the vehicle body.

Filters such as the low-pass filter or complementary filter discussed in Section 3.5 and other valid methods can be used to detect and eliminate short-term magnetic anomalies. These methods include checking the heading change reported by the compass against the physical operating envelope of the vehicle and comparing the measured vector magnitude against the expected range of the Earth's magnetic field. When the external field is strong, it produces a measurement vector that falls far from the curve defined by the circle (Figure 3.10). If we use only samples that fall close to the magnetization circle curve to determine the heading, this source of errors will be eliminated. Relative sensors can also be used to detect blunders (anomalies). When a digital map is available, road segment direction is also a very good reference for detection and correction of compass blunders. Relative sensor data and segment direction are not likely to change dramatically. These blunders are not difficult to detect using if-then rules with proper direction change thresholds relative to the abnormal compass changes associated with a blunder [16].

Errors in calibration are deterministic, and can be eliminated by introducing correction offsets. It is important to realize that calibration errors can change over time. For example, when the vehicle is run through a car wash, the automatic car wash brushes can cause the magnetic field of the vehicle to change for few days so that the center of the circle is shifted (Figure 3.12), which will in turn severely affect the performance of the positioning module. In addition to large moving or stationary steel objects, electronic devices in the vehicle may also cause errors. For instance, when a rear window defroster is in operation, the large current generated may affect a compass mounted nearby. Therefore, care must be taken when deciding where the compass should be installed in the vehicle.

Errors due to certain long-term magnetic anomalies, such as the magnetization of the vehicle body, are difficult to detect. A heading reference that does not use magnetic effects (e.g., GPS) can be used to detect and remove such errors. Electrical or magnetic noise can be averaged over distance and time, and the white noise variance can be estimated *a priori* or estimated in real time when an external heading change reference is available (e.g., differential odometer). As discussed before, the actual magnetic field direction measured by a compass is determined by the combined magnetic field from the Earth and the vehicle itself. If the vehicle is not magnetized, the normalized coil voltages form a circle with the center at the origin. If it is magnetized, the center of the magnetization circle will drop away from the origin (Figure 3.10). A shift curve to 10% of the circle radius will cause a maximum heading error of 6 degrees. The heading error due to the shift in the center of the circle is a function of the actual center and the true heading.

Two methods can be used to correct for the error due to a shift in the center of the magnetization circle. One method is to continuously search for new maximum

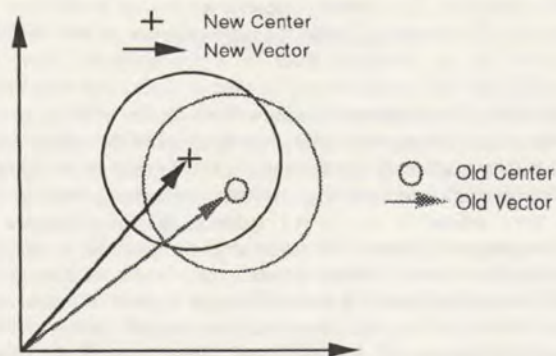


Figure 3.12 Shift of the vehicle magnetic field center.

and minimum points  $V_x$  and  $V_y$  to re method is to introduce a correction to information for any heading. For in of -4 degrees at 100 degrees, thus t the cell for a 100-degree heading. Ad will eliminate the error. Independent digital map) can be used for determ These heading sources are often more on a continuous basis.

When the vehicle is traveling o parallel to the plane of the Earth's sur of the vector components. This typ check on the magnitude of the mag

### 3.4.2 Global Positioning System

The GPS is a satellite-based radio r affordable means of determining pos was designed and paid for by the U.S. is guaranteed through an agreement Transportation (DOT).

GPS consists of three main parts (receivers), and the control segment

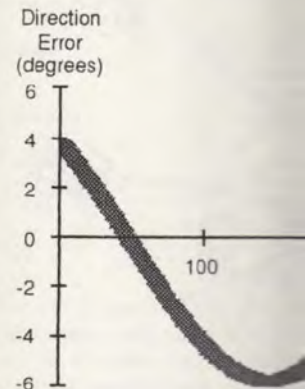


Figure 3.13 Errors due to the shift of the vehicle magnetic field center.

and minimum points  $V_x$  and  $V_y$  to reconstruct the circle center dynamically. Another method is to introduce a correction table. The correction table contains the correction information for any heading. For instance, in Figure 3.13 the compass has an error of  $-4$  degrees at  $100$  degrees, thus the table will contain a  $+4$ -degree correction in the cell for a  $100$ -degree heading. Adding the correction to the compass measurement will eliminate the error. Independent sources of heading information (i.e., GPS or a digital map) can be used for determining the correction and generating the table. These heading sources are often more accurate than the compass but are not available on a continuous basis.

When the vehicle is traveling over a hilly road, the compass plane will not be parallel to the plane of the Earth's surface. The compass measures only the projection of the vector components. This type of error can be detected and removed via a check on the magnitude of the magnetic field.

### 3.4.2 Global Positioning System

The GPS is a satellite-based radio navigation system. It provides a practical and affordable means of determining position, velocity, and time around the globe. GPS was designed and paid for by the U.S. Department of Defense (DOD). Civilian access is guaranteed through an agreement between the DOD and the Department of Transportation (DOT).

GPS consists of three main parts: the space segment (satellites), the user segment (receivers), and the control segment (management and control). From the location

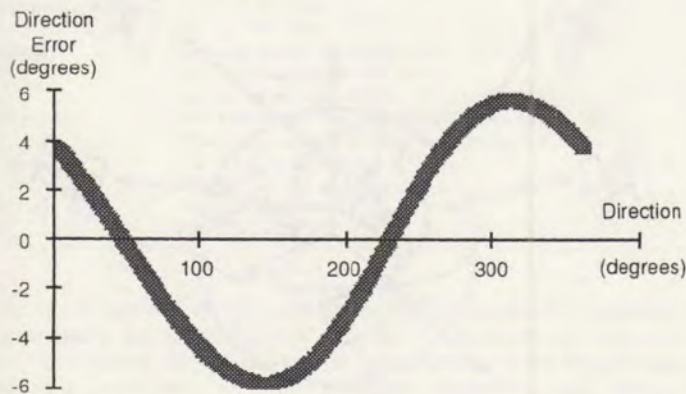


Figure 3.13 Errors due to the shift of the vehicle magnetic field center.

and navigation point of view, only the first two segments are of interest. For more detailed descriptions of all the segments, as well as various theoretical and practical aspects of GPS, readers may consult [17–19].

The observation of at least four satellites simultaneously will permit determination of the 3D coordinates of the receiver (Figure 3.14) and the time offset between the receiver and GPS time. The 3D coordinates can be computed in an ECEF Cartesian coordinate frame or in a geodetic coordinate system such as WGS 84. Using transformation equations, positions can be computed in other geodetic frames or mapping planes (see Section 2.3).

The GPS constellation consists of 24 satellites arranged in six orbital planes with 4 satellites per orbital plane as shown in Figure 3.15. It is designed to provide worldwide coverage 24 hours per day. The characteristics of GPS are summarized in Table 3.3 (some items in Table 3.3 are from [20]). The design life span of each GPS satellite is 7.5 years. The on-board time and frequency standards are the most likely functions to fail, so each satellite carries four atomic clocks on board.

In Table 3.3, SPS stands for Standard Positioning Service, PPS stands for Precise Positioning Service, and 2dRMS stands for two times the root mean square distance error (twice the standard deviation). The term 100m 2dRMS means that 95% of the time the horizontal positions will be within a circle of a 100m radius. Both the C/A code and P code are pseudorandom noise (PRN) ranging codes. PRN is a sequence of digital 1's and 0's that appear to be randomly distributed like noise, but which can be exactly reproduced. The navigation message broadcast includes satellite orbital position data and satellite health status.



Figure 3.14 ECEF Cartesian coordinate system for GPS.



Figure 3.15 Satellite constellation (space segment).

Table 3.3  
GPS Characteristics

| Item                |                                                                                           |
|---------------------|-------------------------------------------------------------------------------------------|
| Satellites          | 24 satellites broadcast signals                                                           |
| Orbits              | Six planes, at 55-degree inclination, with 4 satellites at 20,231-km altitude             |
| Carrier frequencies | L1: 1575.42 MHz;<br>L2: 1227.60 MHz                                                       |
| Digital signals     | C/A code (coarse acquisition)<br>P code (precise code): 10.23 MHz                         |
| Navigation accuracy | Navigation message: 50 bps<br>SPS: 100m horizontal (2dRMS)<br>PPS: 21m horizontal (2dRMS) |
| Velocity accuracy   | SPS: 0.5–2 m/s observed<br>PPS: 0.2 m/s                                                   |
| Time accuracy       | SPS: 340 ns (95%)<br>PPS: 200 ns (95%)                                                    |

The SPS is based on the C/A code and is subject to SA. SA is an accuracy degradation scheme to reduce accuracy to a level within the national security requirements achieved by a combination of orbital parameter dithering. This results in an intentional error in the location (ephemeris) of the satellite.

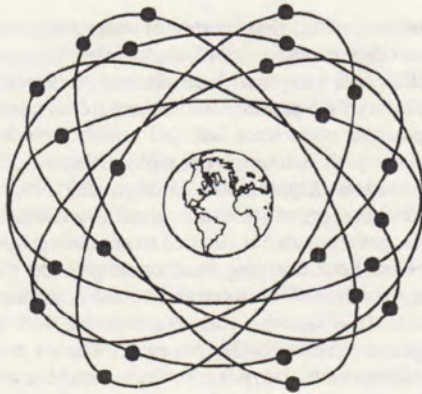


Figure 3.15 Satellite constellation (space segment).

Table 3.3  
GPS Characteristics

| Item                | Characteristic                                                                                                               |
|---------------------|------------------------------------------------------------------------------------------------------------------------------|
| Satellites          | 24 satellites broadcast signals autonomously                                                                                 |
| Orbits              | Six planes, at 55-degree inclination, each orbital plane includes four satellites at 20,231-km altitude, with a 12-hr period |
| Carrier frequencies | L1: 1575.42 MHz;<br>L2: 1227.60 MHz                                                                                          |
| Digital signals     | C/A code (coarse acquisition code): 1.023 MHz<br>P code (precise code): 10.23 MHz<br>Navigation message: 50 bps              |
| Position accuracy   | SPS: 100m horizontal (2dRMS) and 140m vertical (95%)<br>PPS: 21m horizontal (2dRMS) and 29m vertical (95%)                   |
| Velocity accuracy   | SPS: 0.5–2 m/s observed<br>PPS: 0.2 m/s                                                                                      |
| Time accuracy       | SPS: 340 ns (95%)<br>PPS: 200 ns (95%)                                                                                       |

The SPS is based on the C/A code and is subject to selective availability (SA). SA is an accuracy degradation scheme to reduce the accuracy available to civilian users to a level within the national security requirements of the United States. It is achieved by a combination of orbital parameter degradation and satellite clock parameter dithering. This results in an intentional error in the navigation message regarding the location (ephemeris) of the satellites and the intentional error in the

timing source (carrier frequency) that creates an uncertainty in the Doppler velocity measurements. For civilian users, only SPS is available, therefore, SA degrades the performance of civilian GPS receivers. A simple way to detect whether SA is on or off is to record stationary GPS position and velocity data over a period of time. If SA is off, the 2D position error trace will fall within a circle with a 25m radius (versus 100 m), and the peak velocity error will be about 0.25 m/s (versus 1 m/s). The satellites also broadcast a parameter known as user range accuracy (URA) as part of the navigation message, which can be used as an indication of the status of SA. Recently, the U.S. government has decided to conditionally phase out SA placed on the civilian users (without charging fees) starting in the year 2000, subject to yearly review by the president. This is certainly good news for GPS civilian users.

Differential GPS (DGPS) techniques can improve GPS performance significantly. The technique involves two GPS receivers. One is a master receiver at a reference station with known (surveyed) coordinates and the other is a receiver at a remote location whose coordinates are to be determined (Figure 3.16). The calculated position solution for the master receiver can be compared to the known coordinates to generate a differential correction for each satellite. In one such commonly used technique, the correction information can then be transmitted to the remote receiver for calculating the corrected local position. This effectively reduces the position error

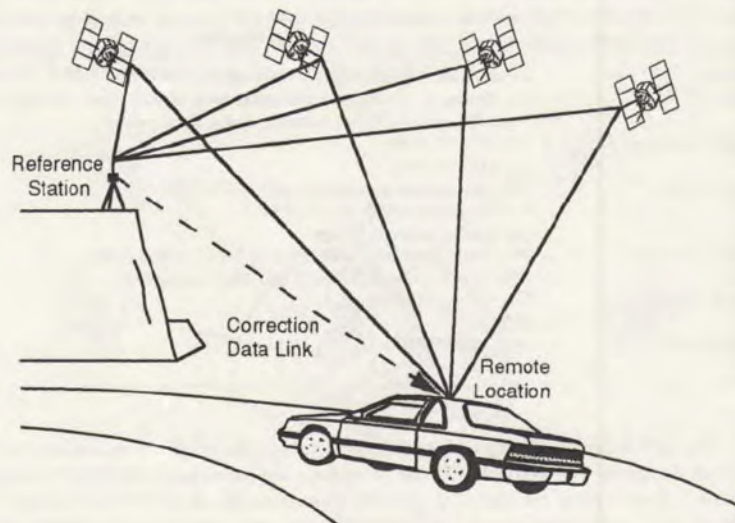


Figure 3.16 Differential GPS.

for commercial systems to under 15m. This assumes that the range errors for the remote receiver. It is easy to see that this assumption is only valid when the two receivers are far apart. Accuracy is achieved for receiver separations of up to 1000 km.

Typical DGPS characteristics are shown in Table 3.3. Table 3.3 indicates that DGPS is much better than PPS. A more detailed description of the basic positioning concepts.

We now briefly discuss the measurement of position. As shown in Figure 3.15, a constellation of satellites transmits signals to radio receivers at or near the surface of the earth. These receivers then calculate their position from the signals. This is a one-way broadcast system; the receivers do not send any signals back to the satellites. Since there is an unlimited number of users, much like a radio broadcast, the system is not limited by the number of users.

Position measurement is based on the time interval taken for a signal to travel from the satellite to the receiver. The time interval taken for a signal to travel from a known location to reach a receiver is used to obtain the emitter-to-receiver range (distance). By measuring the range from multiple emitters at known locations, the position of the receiver can be determined. However, it is impossible to measure the actual range because of the clock offset between satellite and receiver. The clock offset is the constant difference in time between the satellite and receiver clock. To determine the position, the receiver must know the pseudoranges to the satellites. Pseudoranges are obtained by multiplying the time by the speed of light. The signal processing requires the time offset required to match the received signal to a replica of the code. This is known as "code correlation." The number of independent observations needed to be solved.

Table 3.3  
DGPS Characteristics

| Item                  |
|-----------------------|
| Receiver separation   |
| Data link update rate |
| Position accuracy     |
| Velocity accuracy     |
| Time accuracy         |

for commercial systems to under 15m even when SA is in effect. DGPS technology assumes that the range errors for the remote receiver are the same as for the master receiver. It is easy to see that this assumption may not be true all the time, especially when the two receivers are far apart. An accuracy level of 5m can commonly be achieved for receiver separations of up to 50 km when low-noise receivers are used.

Typical DGPS characteristics are listed in Table 3.4. Comparison with Table 3.3 indicates that DGPS is much better than SPS and even (surprisingly) better than PPS. A more detailed description of DGPS techniques follows the presentation of the basic positioning concepts.

We now briefly discuss the measurement principles used by the GPS receiver. As shown in Figure 3.15, a constellation of satellites is maintained in Earth orbit. Radio receivers at or near the surface of the planet are used to decode the transmitted satellite signals. These receivers then calculate position, velocity, and time from the signals. This is a one-way broadcast system, so receivers do not need to transmit any signals back to the satellites. Since it is a passive system, GPS can support an unlimited number of users, much like a television broadcast system.

Position measurement is based on the principle of time of arrival (TOA) ranging. The time interval taken for a signal transmitted from an emitter (in our case satellite) at a known location to reach a receiver is multiplied by the speed of the signal to obtain the emitter-to-receiver range (distance). Multiple signals received by a receiver from multiple emitters at known locations are used to determine its location. Because of clock offset between satellite and receiver, propagation delays, and other errors, it is impossible to measure the actual range, so a pseudorange is measured. The clock offset is the constant difference in the time reading between the satellite clock and receiver clock. To determine the position of the receiver, the receiver needs to know the pseudoranges to the satellites being tracked and the positions of those satellites. Pseudoranges are obtained by multiplying the apparent signal propagation time by the speed of light. The signal propagation time is determined by measuring the time offset required to match the received satellite code to an internally generated replica of the code. This is known as "correlation." As in all surveying situations, the number of independent observations required depends on how many unknowns need to be solved.

Table 3.4  
DGPS Characteristics

| <i>Item</i>           | <i>Real-Time Characteristic</i>    |
|-----------------------|------------------------------------|
| Receiver separation   | <50 km                             |
| Data link update rate | 5–10 sec                           |
| Position accuracy     | 15m (2–5m for low-noise receivers) |
| Velocity accuracy     | 0.1 m/s                            |
| Time accuracy         | 100 ns                             |

In the 2D case, at least two satellites are required to derive the receiver position. If the positions of the satellites are known and the pseudorange from the receiver to each satellite can be measured, the receiver position is constrained to be one of two points of intersection between circles having radii equal to the pseudoranges (Figure 3.17).

The unknown receiver position  $(x, y)$  can be calculated using the following equations. Note that these equations use a 2D Cartesian coordinate system for the reference coordinate frame.

$$p_1 = \sqrt{(x - x_1)^2 + (y - y_1)^2}$$

$$p_2 = \sqrt{(x - x_2)^2 + (y - y_2)^2}$$

where  $(x_1, y_1)$  and  $(x_2, y_2)$  are the known coordinates of the satellites, and  $p_1$  and  $p_2$  are the measured pseudoranges. After solving these two simultaneous equations, a rough initial estimate (or a third satellite) can be used to determine which of the two solutions should be used.

For the 3D case, three satellites are required (Figure 3.18). Now instead of circles, the pseudoranges yield spheres of position. Clearly, a unique receiver position  $(x, y, z)$  of the receiver can be determined by solving the following equations:

$$p_1 = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2}$$

$$p_2 = \sqrt{(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2}$$

$$p_3 = \sqrt{(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2}$$

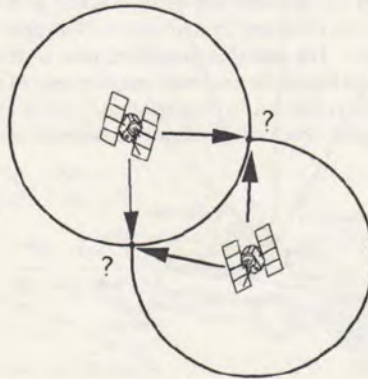


Figure 3.17 Two-pseudorange intersection.

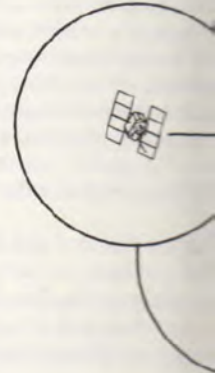


Figure 3.18 Three-pseudorange intersection.

where  $(x_1, y_1, z_1)$ ,  $(x_2, y_2, z_2)$  and  $(x_3, y_3, z_3)$  are the known coordinates of the satellites, and  $p_1$ ,  $p_2$  and  $p_3$  are the measured pseudoranges.

The receiver clock used to measure time must be synchronized to GPS time. Instead, the clock of the receiver must be determined. This parameter can be determined. By design, all of the satellite clocks are synchronized to GPS time. If the receiver clock could be precisely synchronized, the measurement would be simple. However, receivers do not use atomic clocks, so instead, inexpensive quartz clocks introduce a time offset (clock bias) between the receiver and the satellites. This offset must be considered in the computation of the receiver position, and it is the same for all satellites. The receiver position and clock offset can be derived by solving the following equations:

$$p_1 = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2}$$

$$p_2 = \sqrt{(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2}$$

$$p_3 = \sqrt{(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2}$$

$$p_4 = \sqrt{(x - x_4)^2 + (y - y_4)^2 + (z - z_4)^2}$$

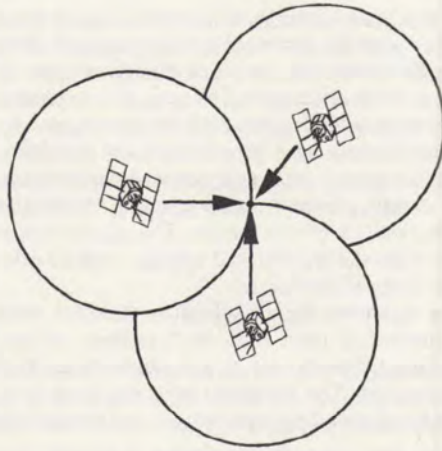


Figure 3.18 Three-pseudorange intersection.

where  $(x_1, y_1, z_1)$ ,  $(x_2, y_2, z_2)$  and  $(x_3, y_3, z_3)$  are the known positions of the satellites, and  $p_1$ ,  $p_2$  and  $p_3$  are the measured pseudoranges.

The receiver clock used to measure the signal propagation times is not synchronized to GPS time. Instead, the clock offset between receiver time and GPS time must be determined. This parameter can be calculated by adding a fourth satellite. By design, all of the satellite clocks are synchronized using very precise atomic clocks. If the receiver clock could be precisely synchronized with the satellite clocks, time measurement would be simple. However, it is economically impractical for receivers to use atomic clocks, so instead, inexpensive crystal oscillators are used. These clocks introduce a time offset (clock bias) between the receiver and the GPS clocks so we must consider their effects in the computation. The receiver clock bias is the time offset of the receiver, and it is the same for each satellite. Thus both the receiver position and clock offset can be derived from the following equations:

$$p_1 = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2} + c(dt - dT_1)$$

$$p_2 = \sqrt{(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2} + c(dt - dT_2)$$

$$p_3 = \sqrt{(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2} + c(dt - dT_3)$$

$$p_4 = \sqrt{(x - x_4)^2 + (y - y_4)^2 + (z - z_4)^2} + c(dt - dT_4)$$

where  $(x_1, y_1, z_1)$ ,  $(x_2, y_2, z_2)$ ,  $(x_3, y_3, z_3)$ , and  $(x_4, y_4, z_4)$  are the known satellite positions,  $p_1, p_2, p_3$ , and  $p_4$  are measured pseudoranges,  $c$  is the speed of light,  $dT_1, dT_2, dT_3, dT_4$  are the known satellite clock bias terms from GPS time, and  $dt$  is the unknown receiver clock offset from GPS time. The satellite clock bias terms are calculated by the receiver from the broadcast navigation message. In the preceding equations, several error terms have been left out for simplicity. For instance, the range errors due to ionospheric delay and tropospheric delay can both be estimated using atmospheric models. However, receiver noise, multipath propagation error, satellite orbit errors, and SA effects remain. The square-root term represents the geometric range between the receiver and satellite, and all other terms contribute to the measurement being a pseudorange.

The preceding equations can be solved by standard techniques such as the Newton-Raphson method. If more than four satellites are used for a solution, a least squares algorithm (Appendix D) or a multiple-state Kalman filter algorithm (Appendix E) can be used. The algorithm used and how it is used have a large impact on the accuracy of position determination and the smoothness of the receiver trajectory.

Let us use four satellites as an example to explain briefly how to solve these equations and to derive measurements of the satellite geometry with respect to the receiver. After linearization, the above equations can be written in a vector-matrix form as

$$Y = AX$$

where  $Y$  is the known vector of measurements and  $X$  is the unknown vector. Linear algebra can be used to solve for the unknown vector  $X$ . A by-product of the computation is the cofactor matrix

$$P = (A^T A)^{-1}$$

where  $T$  stands for the matrix transpose operation and  $-1$  stands for matrix inversion. For a 3D GPS solution, the cofactor matrix is a  $4 \times 4$  matrix where three components are contributed by the receiver position  $(x, y, z)$  and one component by the receiver clock. The matrix can be written as

$$P = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} & \sigma_{xt} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} & \sigma_{yt} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 & \sigma_{zt} \\ \sigma_{tx} & \sigma_{ty} & \sigma_{tz} & \sigma_t^2 \end{bmatrix}$$

To use a GPS receiver's output properly, users must understand various potential GPS errors such as the uncertainty in satellite positions and the effects of

propagation conditions. In addition, the respect to the satellites being tracked measurement errors. Various dilution of precision (DOP) terms characterize the geometric contribution to the total error. The DOP is based on the square root of the estimated position along the three coordinates shown below. We usually consider the GDOP (position DOP) is less than 3 or PDOP (position

$$\text{GDOP} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_t^2}$$

$$\text{PDOP} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$$

Velocity measurements are based on the Doppler shift in the frequency of each satellite signal. The velocity of the receiver and the satellite is very high relative to a stationary observer on the Earth's surface. The velocity solution can be derived from the velocity solution (pseudorange rate) with respect to time.

To have a general idea of the complexity of a GPS receiver, Figure 1 depicts a block diagram of a Motorola 68000-based application specific integrated circuit (ASIC) for application specific integrated circuit (ASIC) with an erasable programmable ROM. See [21] for more details on the design and some practical receiver examples.

Readers who are interested in evaluating GPS receiver performance in specific applications should consult [22]. In many applications, GPS receiver performance in urban areas is particularly important. Overhead of satellite signals, and high-rise buildings block satellite signals. Therefore, a receiver with multipath and parallel channel tracking (versus multiplexing) is more than one without such features (multiplexing) for all of the available channels among the multiplexed signals. Under these circumstances, see [24] for more details on receiver performance. Receiver performance is receiver autonomous mode detection, isolation, and removal of multipath. It indicates that an eight-channel low-cost receiver is needed and performance needed to track satellite signals.

As mentioned earlier in this section, several errors are associated with the

propagation conditions. In addition, the geometrical position of the receiver with respect to the satellites being tracked plays a considerable role in GPS receiver measurement errors. Various dilution of precision (DOP) measures are used to characterize the geometric contribution to the receiver error. The calculation of these DOPs is based on the square root of the sum of the variances of the errors in the estimated position along the three coordinate axes, and of the estimated time as shown below. We usually consider the GPS geometry good if the GDOP (geometrical DOP) is less than 3 or PDOP (position DOP) is less than 12.

$$\text{GDOP} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2 + \sigma_t^2}$$

$$\text{PDOP} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$$

Velocity measurements are based on the Doppler frequency shift principle. The Doppler shift in the frequency of each satellite is a direct measure of the relative velocity of the receiver and the satellite along the line between them. Each satellite has a very high velocity relative to a stationary receiver, both because of the orbital motion of the satellite and because of the motion of the receiver associated with the Earth's rotation. The velocity solution can be calculated by differentiating the four-dimensional navigation solution (pseudorange equations) introduced earlier with respect to time.

To have a general idea of the composition of a GPS receiver, Figure 3.19 depicts a block diagram of a Motorola Oncore GPS receiver, where ASIC stands for application specific integrated circuit and EEPROM stands for electronically erasable programmable ROM. See [21] for more information about the receiver design and some practical receiver examples.

Readers who are interested in evaluation and selection of GPS receivers for specific applications should consult [22] and [23]. For location and navigation applications, GPS receiver performance under heavy foliage and in urban canyon areas is particularly important. Overhead foliage may cause significant attenuation of satellite signals, and high-rise buildings in urban canyons may block and reflect satellite signals. Therefore, a receiver with a position filter, fast reacquisition time, and parallel channel tracking (versus multiplexing channel) capabilities is much better than one without such features (multiplexing is the practice of rapidly sequencing any or all of the available channels among the visible satellites). For performance test reports under these circumstances, see [24,25]. Another feature that enhances GPS receiver performance is receiver autonomous integrity monitoring (RAIM). It provides detection, isolation, and removal of faulty satellite signals. Finally, research indicates that an eight-channel low-cost receiver provides an excellent balance of cost and performance needed to track satellite signals reliably.

As mentioned earlier in this section, when determining the receiver position, several errors are associated with the pseudorange measurement. Table 3.5

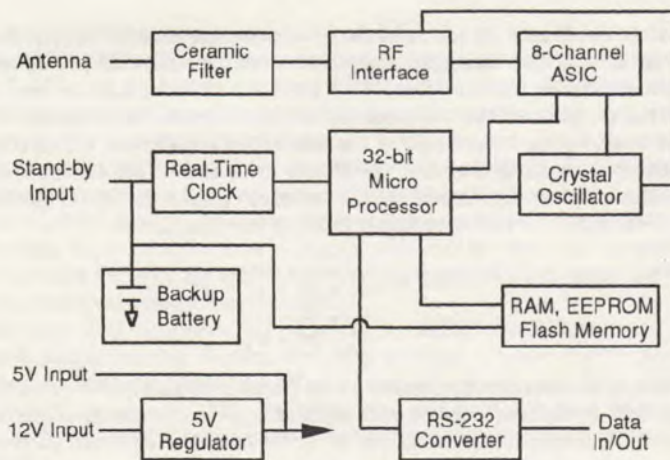


Figure 3.19 Motorola Oncore GPS receiver diagram.

Table 3.5  
GPS Receiver Errors

| Error                | Typical GPS (m) | Worst-Case GPS (m) | Typical DGPS (m) |
|----------------------|-----------------|--------------------|------------------|
| Satellite clock      | 2-3             | 25                 | <0.5             |
| Satellite orbit      | 1-2             | 5                  | <0.5             |
| SA                   | 30-50           | 100                | <0.5             |
| Ephemeris prediction | 3-5             | 15                 | <0.5             |
| Ionospheric delay    | 10-15           | 100                | <0.5             |
| Tropospheric delay   | 3-5             | 30                 | <0.5             |
| Multipath            | 10              | 300                | 10               |
| Receiver noise       | 5               | 15                 | 5                |
| Total effect         | 100 (95%)       | 300 (99.99%)       | 15 (95%)         |

summarizes the typical effects of these errors on the position output by the receiver. From the table, we see that there are other sources of error in addition to SA. The first six errors in Table 3.5 are virtually zero for typical DGPS (receiver separation less than 50 km) because they are highly correlated.

Other techniques can also be used when designing the receiver algorithm to reduce the errors further. For civilian receivers, two methods can be used to calculate receiver position at first lock-on to satellites. One is based on the code phase (1.023 MHz, wavelength  $\approx 293,000$  m). The other is based on the carrier phase

(1575.42 MHz, wavelength  $\approx 0.19$  m) transmitted by the satellite as shown in Figure 3.18. Both methods require the receiver to decide the number of cycles of the code or carrier to correlate with the one in the navigation message. Both methods require the receiver to initially lock on to the signal, which is easily resolved.

Sophisticated algorithms are required to resolve ambiguity. A combination of these two methods (code tracking (carrier-smoothed code)) can reduce the error to 1 to 2m. One study shows that with this data, DGPS accuracy can be made as good as carrier phase differential processing and is easily resolved.

GPS receivers can be used for time synchronization. Recall that the satellites have very expensive clocks (accuracy  $10^{-12}$ ) compared with the receivers. Furthermore, the clock time-keeping is very accurate. Other technologies because all satellite clocks are synchronized to GPS time. The time difference between the satellites is within 60 ns. Therefore, it is very easy to synchronize time to either GPS time or to the satellites. Examples of such applications are to synchronize time among three radio-frequency (Sections 9.2.2 and 10.2). Because GPS transmits a pulse per second or 1 Hz, these pulses can be used for time synchronization. Using a correction circuit, the time can be synchronized very precisely.

GPS receivers have many different applications. A reference station receiver with few mobile receivers can collect raw data in the field, and the receiver can calculate the distances between the receiver locations in the office. A reference station receiver can be used for a static survey. A reference station receiver can be used for a dynamic survey. A reference station receiver can be used for a static survey. A reference station receiver can be used for a dynamic survey. A reference station receiver can be used for a static survey. A reference station receiver can be used for a dynamic survey.

DGPS has different implementations. One is based on the code phase and the receiver positions are known, the receiver can calculate the range by subtracting the known range vector from the

(1575.42 MHz, wavelength  $\approx 0.19$  m). Both the code and the carrier phases are transmitted by the satellite as shown in Table 3.3. The receiver generates its own code or carrier to correlate with the one received from the satellite in order to decode the navigation message. Both methods have associated range ambiguities because the receiver needs to decide the number of cycles transmitted from the satellite in order for the receiver to initially lock on. For the code phase tracking, the ambiguity is easily resolved.

Sophisticated algorithms are required to resolve the carrier phase tracking ambiguity. A combination of these two observations, as is done in carrier-aided code tracking (carrier-smoothed code) produces better results. For instance, typical Motorola Oncore DGPS can reduce both the multipath and receiver noise errors down to 1 to 2m. One study shows that by postprocessing the raw Oncore output data, DGPS accuracy can be made as good as 2 cm [26]. The postprocessing uses carrier phase differential processing and ambiguity resolution techniques.

GPS receivers can be used for time synchronization and frequency stabilization. Recall that the satellites have very expensive on-board atomic clocks with picosecond accuracy ( $10^{-12}$ ) compared with the nanosecond accuracy ( $10^{-9}$ ) of low-cost GPS receivers. Furthermore, the clock time-keeping ability of the satellites is better than other technologies because all satellite clocks are time synchronized by the control segment to GPS time. The time difference between any two GPS receivers is typically within 60 ns. Therefore, it is very economical to use low-cost GPS receivers to synchronize time to either GPS time or universal time coordinated (UTC) time. Examples of such applications are to synchronize time between adjacent cellular telephone cells to smooth the transition of a caller from one cell to another, and to synchronize time among three radio-frequency reference sites to locate a mobile user (Sections 9.2.2 and 10.2). Because GPS receivers can generate steady pulses (one pulse per second or 1 Hz), these pulses can be used for comparison with oscillator frequencies. Using a correction circuit, the frequency of the oscillator can be stabilized very precisely.

GPS receivers have many different applications. In geodetic surveying, a reference station receiver with few mobile receivers at selected surveying locations can collect raw data in the field, and the resulting data can be used to calculate the distances between the receiver locations in the office. This is called a *static postprocessing survey*. A reference station receiver with one mobile receiver can collect data while the mobile receiver is moved to different surveying locations in the field, and the distances are calculated in the office. This is known as a *semi-kinematic* or *rapid static survey*. A reference station receiver or a mobile receiver can broadcast its raw measurements to the other receiver, so that carrier phase differential processing can be done in real time. This method is called a *real-time kinematic (RTK) survey*.

DGPS has different implementations. For instance, if both the satellite positions and the receiver positions are known, the range correction vector can be derived by subtracting the known range vector from the measured range vector. The correction

vector can then be transmitted from the reference station to the mobile receiver. The data transmission rates for the corrections are determined by the rates of variation in the errors. For autonomous vehicle location and navigation, one-way communication of corrections is sufficient. For vehicle tracking, two-way communication may be required if the system needs both the user in the vehicle and a central dispatcher to have corrected positions. VHF (very high frequency) radio, two-way radio, FM subcarrier radio, pager signals, cellular telephone, or satellite-based communications, etc., can be used for this communications link. We discuss RF communication in more detail in Chapter 8.

One cost-effective technique for vehicle tracking is inverse DGPS. In this situation, the vehicle does not have the corrected positions; only the reference station does. This technique requires the vehicle to transmit satellite numbers and ephemeris data for the satellites it is tracking to the reference station along with the vehicle position. The software in the reference station uses these data and other data collected by the reference station using the same satellites to compute a position fix. It then derives the correction vector for that vehicle and applies it. If the system needs to track many vehicles, the reference station needs a very powerful processor or processors because of the computationally intensive nature of the task.

Wide-area DGPS (WADGPS) is another cost-effective method of bringing differential techniques to many users. Recall that DGPS needs to have a receiver separation of less than 50 km to provide very good accuracy. In other words, the main disadvantage of DGPS is the high correlation between the reference-station-to-user distance and the positioning accuracy achieved. To cover a wide area, a substantial increase in the number of reference stations is required. In contrast, WADGPS requires very few reference stations. For example, 10 sparsely located reference stations may be enough to cover the continental United States. Using the GPS receiver data collected by these 10 stations, a national map can be calculated for the entire country. Whenever any local user needs the service, this national map can be used to derive corrections for a virtual reference station in that particular area. Since this virtual reference station is nearly equivalent to a real reference station in the same area, similar position accuracy can be achieved. Research has also been very active in applying similar techniques to determine the exact satellite orbits. If real-time computation can be achieved, this *precise orbits* method should be able to improve the positional accuracy even more. Currently, precise orbits and clocks are available for postmission use only.

In addition to the Global Positioning System developed by the United States, other satellite systems have been developed that could be used for location and navigation, such as GLONASS, which was developed by the former USSR. NAVSAT has been developed by the European Space Agency (ESA), LOCSTAR is being promoted by the French space organization (CNES), and STARFIX has been developed by John E. Chance and Associates of Louisiana for the Gulf of Mexico and continental United States. More information about these systems can be found in [21,27] and

the references cited therein. We discuss in Section 8.3.8; most of the satellite system position determination as a feature. In general, the accuracy of these technologies is worse than that of

### 3.5 SENSOR FUSION

From the preceding discussion, it is clear that accurate vehicle position information. The in order to provide the on-vehicle system with information for its location and navigation, a system has been developed to fuse the complementary information from one representation format. The information from multiple sensors during a single period of time or over a period of time. Integrated multisensor systems provide higher levels of accuracy and fault tolerance.

Multisensor integration and fusion provide many advantages. These may include robust operational performance, extended temporal coverage, an increased level of performance, enhanced spatial resolution, increased dimensionality, full utilization of sensor data. However, we must keep in mind that although sensor fusion, most fusion methods make certain assumptions. If the assumed sensor model does not adequately represent the sensor, a perfect sensor modeling and estimation result.

#### 3.5.1 Simple Filters

Because of error propagation, many different methods are used. One such method is the digital filter. A low-pass filter can be used to eliminate effects of other magnetic fields on the compass. To eliminate errors due to odometer drift, the odometer data can be integrated to obtain the output from the filter. The filter is a deterministic method of linear system analysis and can be written in the form of a transfer function:

$$G(s) = \frac{1}{s\tau + 1}$$

where  $\tau$  is the time constant of the low-pass filter.  $\tau$  could be the product of a resistance and a capacitance. We can derive the transfer function for the

the references cited therein. We discuss other satellite communication technologies in Section 8.3.8; most of the satellite systems in operation or in planning also offer position determination as a feature. In general, the positional accuracy provided by these technologies is worse than that of GPS.

### 3.5 SENSOR FUSION

From the preceding discussion, it is clear that no single sensor can provide completely accurate vehicle position information. Therefore, multisensor integration is required in order to provide the on-vehicle system with complementary, sometimes redundant information for its location and navigation task. Many fusion technologies have been developed to fuse the complementary information from different sources into one representation format. The information to be combined may come from multiple sensors during a single period of time or from a single sensor over an extended period of time. Integrated multisensor systems have the potential to provide high levels of accuracy and fault tolerance.

Multisensor integration and fusion provide a system with additional benefits. These may include robust operational performance, extended spatial coverage, extended temporal coverage, an increased degree of confidence, improved detection performance, enhanced spatial resolution, improved reliability of system operation, increased dimensionality, full utilization of resources, and reduced ambiguity. However, we must keep in mind that although there are many advantages in sensor fusion, most fusion methods make certain assumptions, either explicitly or implicitly. If the assumed sensor model does not adequately describe the data from the real sensor, a perfect sensor modeling and estimation theory may not produce the expected result.

#### 3.5.1 Simple Filters

Because of error propagation, many different error filtration methods are used. One such method is the digital filter. A low-pass filter can be used to reduce the disturbing effects of other magnetic fields on the compass. A high-pass filter can be used to eliminate errors due to odometer drift. The output of these two filters can then be integrated to obtain the output from the positioning module. From the basic deterministic methods of linear system analysis, a low-pass filter can be written in the form of a transfer function:

$$G(s) = \frac{1}{1 + \tau s}$$

where  $\tau$  is the time constant of the low-pass filter. If a hardware filter is used,  $\tau$  could be the product of a resistance and a capacitance. By substituting  $s$  with  $1/s$ , we can derive the transfer function for the high-pass filter:



ne such implemen-

measurements of the  
without distorting  
( $t$ ) is the measured  
put of the comple-  
sents the dynamic  
very effective when  
ncies, respectively.  
and to obtain an  
rly, from the basic  
ram in Figure 3.20,  
output is

re can select a  $G(s)$

ent model to recur-  
l sense. If a system  
ensor errors can be  
rovide statistically  
ithout storing any  
d to blend location  
Chapter 4) to yield

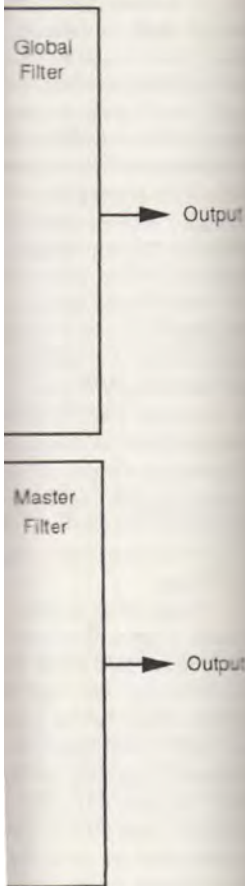
→  $y(t)$

A Kalman filter can be used to fuse measurements from multiple sensors and provide both an estimate of the current state of a system and a prediction of the future state of the system. Two approaches can be taken: a centralized (standard) Kalman filter or a decentralized Kalman filter. A centralized filter takes all the sensor input and outputs a complete solution while a decentralized filter uses local filters for each sensor and a master filter to fuse all the sensors (Figure 3.21). When applied to multisensor systems and/or systems with embedded local filters, research results show that a centralized filter may result in heavy computation loads, poor fault tolerance, and an inability to correctly process prefiltered data in a cascaded filter structure. During the last several years, many decentralized Kalman filters have been introduced in an effort to improve fault tolerance (see [31] and its references).

Many good specialized textbooks have been published that discuss the underlying principles and implementation of Kalman filters [29,32,33]. The recursion equations for discrete Kalman filters are given in Appendix E. As mentioned earlier, the fusion method generally makes certain assumptions. This is certainly the case for a Kalman filter. It assumes that all measurement errors not modeled are uncorrelated and that the variations in these errors are uncorrelated (white noise). Therefore, if these assumptions are not valid, a theoretically perfect filter algorithm might lead to unacceptable results. Since the discrete Kalman filter is recursive, round-off error may lead to problems as the number of steps increases. Sometimes there may be one or more unobservable state variables. All of these situations may cause the system to diverge.

Many other Kalman filters can be used for sensor fusion. In some applications, a linear Kalman filter can be used, provided the nominal trajectory can be linearized in the state space of the system (i.e., the system model is approximately linear) and the trajectory does not depend on the results of the measurement. An extended Kalman filter can be used if the estimated trajectory can be linearized under the assumption that this trajectory can be estimated from the measurements [12]. If the estimated trajectory of the extended Kalman filter is worse than the normal one, this could lead to further error in the trajectory, which causes further errors in the estimates, and so on, leading to the eventual divergence of the filter. Therefore, caution must be used, especially in the situation where the initial uncertainty and measurement errors are both large. An adaptive Kalman filter can be used if all the parameters of the filter are not known initially or if the filter parameters change over time. When numerical stability such as round-off error is of special concern for possible divergence of the filter, a U-D covariance factorization filter can be used [34].  $U$  is an upper triangular matrix (ones along the major diagonal, nontrivial elements in the upper triangular part, and zeros in the lower triangular part) and  $D$  is a diagonal matrix. In practice, many systems have no divergence problems, because they are completely observable and there is adequate process noise feeding into all system states.





them here for interested readers to explore further. Many techniques not mentioned here are often used for pixel processing and image recognition and could be very useful for enhancing drivers' vision or even allowing automatic driving. Some techniques have already been proposed and used in practice.

Fuzzy logic can be used to represent uncertainty in the inference (or fusion) process to make a crisp decision. A distributed blackboard architecture can be used to ease the communication between various sensors in the fusion process. Weighted averaging can be employed to average redundant information. Statistical decision theory can be used to model sensor noise as probability distributions for removal of outliers. Bayesian estimation (Bayesian inference) can be applied to combine sensor information based on the rules of probability theory. Dempster-Shafer evidential reasoning can be used to extend the Bayesian estimation. Production rules can be used to symbolically represent the relation between sensory information and an attribute that can be inferred from the information as part of an expert system, etc. A neural network can be trained for fusing sensors without knowing their characteristics. Neural networks consist of many layered nodes (neurons) that are interconnected. A neural network is a dynamic system and must be trained to rearrange layers and interconnections in order to model the real-world application. One important feature of neural networks is that they simulate a dynamic system without a mathematical model of how the outputs depend on the inputs. Neural networks learn from numerical (and sometimes even linguistic) sample data.

Although various sensor integration and fusion technologies are available, we must keep in mind that any technique used in a vehicle location and navigation system must be efficient enough for real-time derivation of the correct position. Otherwise, the technique is not suitable for actual implementation in an embedded real-time system.

In general, conventional vehicle positioning techniques such as dead reckoning tend to accumulate errors; absolute positioning systems such as GPS are free from error accumulation. However, as discussed, the accuracy and availability of GPS are affected by SA and the vehicle environment. Consequently, the raw vehicle position derived from sensor measurements as described earlier will not be error free. In Chapter 4, we discuss how to use the map-matching module to enhance the vehicle positioning system.

Before turning to the next chapter, can you list the main error sources for each sensor discussed and the compensation methods learned that can eliminate or reduce these errors? Are there any other approaches that can compensate for the errors you listed?

#### References

- [1] K. Jost, "Sensor Technology Review," *Automotive Engineering*, Vol. 103, No. 9, Sep. 1995, pp. 39-49.
- [2] R. Jurgen (Ed.), *Automotive Electronics Handbook*, New York: McGraw-Hill, 1995.

and for sensor fusion [7,35].  
be familiar, we introduce

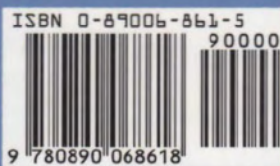
- [3] C. O. Nwagboso (Ed.), *Automotive Sensory Systems*, London/New York: Chapman & Hall, 1993.
- [4] W. B. Ribbens, *Understanding Automotive Electronics*, 4th ed., Carmel, IN: SAMS, 1992.
- [5] J. Fraden, *AIP Handbook of Modern Sensors: Physics, Designs, and Applications*, New York: American Institute of Physics, 1993.
- [6] L. Ristic (Ed.), *Sensor Technology and Devices*, Norwood, MA: Artech House, 1994.
- [7] R. C. Luo and M. G. Kay (Eds.), *Multisensor Integration and Fusion for Intelligent Machines and Systems*, Norwood, NJ: Ablex Publishing Corp., 1995.
- [8] D. A. Foster, "Developments in Wheel Speed Sensing," *ABS Traction Control*, SAE Paper 880325, Society of Automotive Engineers, Aug. 1988, pp. 81-87.
- [9] R. F. Wells, "Non-Contacting Sensors: An Update," *Automotive Engineering*, Vol. 96, No. 11, Nov. 1988.
- [10] S. Oho, H. Kajioaka, and T. Sasayama, "Optical Fiber Gyroscope for Automotive Navigation," *IEEE Trans. Vehicular Technol.*, Aug. 1995, pp. 698-705.
- [11] N. Yumoto, "Onboard Equipment," in *Advanced Technology for Road Transport: IVHS and ATT*, Ian Catling (Ed.), Norwood, MA: Artech House, 1994, pp. 338-342.
- [12] B. Barshan and H. F. Durrant-Whyte, "Inertial Navigation Systems for Mobile Robots," *IEEE Trans. Robotics Automation*, Vol. 11, No. 3, June 1995, pp. 328-342.
- [13] J. Miyazaki, R. Higashimae, T. Kuirhara, and S. Ishino, "Vibratory Gyroscope and Various Applications," *Murata Electronics North America*, 1995.
- [14] T. Oikawa, Y. Aoki, and Y. Suzuki, "Development of Vibrational Rate Sensor and Navigation System," *SAE Paper No. 870215*, Society of Automotive Engineers, 1987.
- [15] L. E. Lenz, "A Review of Magnetic Sensors," *Proc. IEEE*, Vol. 78, No. 6, June 1990, pp. 973-989.
- [16] Y. Zhao, L. G. Seymour, and E. M. Kozikaro, "Absolute Heading Sensor Blunder Detection Using Relative Heading Sensor and Road Segment," *Technical Developments*, Motorola, Vol. 29, Nov. 1996, pp. 104-107.
- [17] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice*, 3rd ed., Berlin: Springer-Verlag, 1994.
- [18] A. Leick, *GPS Satellite Surveying*, New York: John Wiley & Sons, 1995.
- [19] B. W. Parkinson, J. J. Spilker Jr., P. Axelrad, and P. Enge (Eds.), *Global Positioning System: Theory and Applications*, Washington, DC: American Institute of Aeronautics and Astronautics, 1996.
- [20] DoD/DoT, *1994 Federal Radionavigation Plan*, Springfield, VA: National Technical Information Service, Report No. DOT-VNTSC-RSPA-95-1/DOD-4650.5, May 1995.
- [21] B. Forssell, *Radionavigation Systems*, New York: Prentice Hall, 1991.
- [22] J. F. McLellan and J. P. Battie, "Testing and Analysis of OEM GPS Sensor Boards for AVL Applications," *Proc. IEEE Position Location and Navigation Symposium*, Apr. 1994, pp. 512-519.
- [23] Y. Zhao, "Evaluation of GPS Receivers for Automotive Applications," *Proc. Second World Congress on Intelligent Transport Systems*, Nov. 1995, pp. 568-572.
- [24] G. Lachapelle and J. Henriksen, "GPS Under Cover: The Effect of Foliage on Vehicular Navigation," *GPS World*, Mar. 1995, pp. 26-35.
- [25] Motorola, *Navigation Performance Test Report*, Position and Navigation Systems Business, Jan. 1994.
- [26] A. Tabsh, H. Martell, and D. Cosandier, "High Precision Real-Time Positioning Using Low Cost OEM Sensors," *Proc. ION GPS-95*, Sep. 1995, pp. 1433-1441.
- [27] E. Kaplan (Ed.), *Understanding GPS: Principles and Application*, Norwood, MA: Artech House, 1996.
- [28] M. L. G. Thoone, "CARIN, A Car Information and Navigation System," *Philips Tech. Rev.*, Vol. 43, No. 11/12, Dec. 1987, pp. 317-329.
- [29] R. G. Brown and P. Y. C. Huang, *Introduction to Random Signal Analysis and Applied Kalman Filtering*, New York: Wiley, 1992.
- [30] W.-W. Kao, "Integration of GPS and Dead Reckoning for Navigation Systems," *IEEE Trans. Vehicular Technol.*, Vol. 41, No. 3, Fall 1994, pp. 297-322.
- [31] H. A. Carlson, "Federated Kalman Filter," *IEEE Trans. Vehicular Technol.*, Vol. 41, No. 3, Fall 1994, pp. 297-322.
- [32] A. Gelb (Ed.), *Applied Optimal Estimation*, New York: Wiley, 1974.
- [33] P. S. Maybeck, *Stochastic Models, Estimation, and Prediction*, New York: Wiley, 1974.
- [34] E. J. Krakiwsky, C. B. Harris, and R. V. L. "Map Matching, and GPS Positioning," *IEEE Trans. Vehicular Technol.*, Vol. 37, No. 1, Feb. 1988, pp. 39-46, 1988.
- [35] W. Waltz and J. Llinas, *Multisensor Data Fusion*, New York: Wiley, 1990.

- c Hall, 1993.  
1992.  
New York:
- Machines and  
aper 880325.  
96, No. 11.  
Navigation."
- HS and ATT.  
obots," *IEEE*  
rious Applica-  
nd Navigation  
, pp. 973-989.  
etection Using  
rola, Vol. 29,  
n: Theory and
- System: Theory  
utics, 1996.  
al Information
- ards for AVL  
t, pp. 512-519.  
World Congress  
ar Navigation,"  
s Business, Jan.  
Using Low Cost  
Artech House,  
Tech. Rev., Vol.  
Applied Kalman
- [30] W.-W. Kao, "Integration of GPS and Dead-Reckoning Navigation Systems," *Proc. IEEE-IEE Vehicle Navigation & Information Systems Conference (VNIS '91)*, Society of Automotive Engineers, Oct. 1991, pp. 635-643.
- [31] H. A. Carlson, "Federated Kalman Filter Simulation Results," *Navigation: J. Institute of Navigation*, Vol. 41, No. 3, Fall 1994, pp. 297-321.
- [32] A. Gelb (Ed.), *Applied Optimal Estimation*, Cambridge, MA: MIT Press, 1974.
- [33] P. S. Maybeck, *Stochastic Models, Estimation and Control*, New York: Academic Press, 1979.
- [34] E. J. Krakiwsky, C. B. Harris, and R. V. C. Wong, "A Kalman Filter for Integrating Dead Reckoning, Map Matching, and GPS Positioning," *Proc. IEEE Position, Location and Navigation Symposium (PLANS)*, IEEE, pp. 39-46, 1988.
- [35] W. Waltz and J. Llinas, *Multisensor Data Fusion*, Norwood, MA: Artech House, 1990.

I INTELLIGENT

T TRANSPORTATION

S SYSTEMS



**AH** Artech House Publishers BOSTON • LONDON