Lewis, R.A., Johnson, A.R., "A Scanning Laser Rangefinder for a Robotic Vehicle," 5th International Joint Conference on Artificial Intelligence, pp. 762-768, 1977.

Massa, "E-201B & E-220B Ultrasonic Ranging Module Subsystems Product Selection Guide," Product Literature 891201-10M, Massa Products Corporation, Hingham, MA, undated.

Moravec, H. P., Elfes, A., "High Resolution Maps from Wide Angle Sonar," IEEE International Conference on Robotics and Automation, St. Louis, MO, pp. 116-121, March, 1985.

NASA, "Fast, Accurate Rangefinder, *NASA Tech Brief*, NPO-13460, Winter, 1977.

National, "LM1812 Ultrasonic Transceiver," *Special Purpose Linear Devices Databook*, National Semiconductor Corp., Santa Clara, CA, Section 5, pp. 103-110, 1989.

National, "Electrostatic Transducers Provide Wide Range Ultrasonic Measurement," *Linear Applications Handbook*, National Semiconductor Corp., Santa Clara, CA, pp. 1172-1173, 1991.

Olson, R.A., Gustavson, R.L., Wangler, R.J., McConnell, R.E., "Active Infrared Overhead Vehicle Sensor," IEEE Transactions on Vehicular Technology, Vol. 43, No. 1, pp. 79-85, February, 1994.

Pin, F. G., Watanabe, Y., "Using Fuzzy Behaviors for the Outdoor Navigation of a Car with Low-Resolution Sensors," IEEE International Conference on Robotics and Automation, Atlanta, GA, pp. 548-553, 1993.

Pletta, J.B., Amai, W.A., Klarer, P., Frank, D., Carlson, J., Byrne, R., "The Remote Security Station (RSS) Final Report," Sandia Report SAND92-1947 for DOE under Contract DE-AC04-76DP00789, Sandia National Laboratories, Albuquerque, NM, October, 1992.

Polaroid, "Polaroid Ultrasonic Ranging System User's Manual," Publication No. P1834B, Polaroid Corporation, Cambridge, MA, December, 1981.

Polaroid, "Technical Specifications for Polaroid Electrostatic Transducer," 7000-Series Product Specification ITP-64, Polaroid Corporation, Cambridge, MA, June, 1987.

Polaroid, "6500-Series Sonar Ranging Module," Product Specifications PID 615077, Polaroid Corporation, Cambridge, MA, 11 October, 1990.

Polaroid, "Polaroid Ultrasonic Ranging Developer's Kit," Publication No. PXW6431 6/93, Polaroid Corporation, Cambridge, MA, June, 1993.

RIEGL, "Laser Distance, Level, and Speed Sensor LD90-3," Product Data Sheet 3/94, RIEGL Laser Measurement Systems, RIEGL USA, Orlando, FL, March, 1994.

SEO, "LRF-X Laser Rangefinder Series,", Product Literature, Schwartz Electro-Optics, Inc., Orlando, FL, October, 1991a.

SEO, "Scanning Laser Rangefinder,", Product Literature, Schwartz Electro-Optics, Inc., Orlando, FL, October, 1991b.

SEO, "Helicopter Optical Proximity Sensor System,", Product Literature, Schwartz Electro-Optics, Inc., Orlando, FL, October, 1991c.

SEO, Process Report for US Army Contract DAAJ02-91-C-0026, Schwartz Electro-Optics, Inc., Orlando, FL, December, 1991d.

Siuru, B., "The Smart Vehicles Are Here," *Popular Electronics*, Vol. 11, No. 1, pp. 41-45, January, 1994.

Vuylsteke, P., Price, C.B., Oosterlinck, A., "Image Sensors for Real-Time 3D Acquisition, Part 1," ," in *Traditional and Non-Traditional Robotic Sensors*, T.C. Henderson, ed., NATO ASI Series, Vol. F63, Springer-Verlag, pp. 187-210, 1990.

6

Phase-Shift Measurement and Frequency Modulation

6.1 Phase-Shift Measurement

The *phase-shift measurement* (or *phase-detection*) ranging technique involves continuous-wave (CW) transmission as opposed to the short-duration pulsed outputs used in the time-of-flight systems discussed in Chapter 5. (The transmission of short pulses may also be used if synchronized to a continuous-wave reference against which the phase of the returning signal is measured.) One advantage of continuous-wave systems over pulsed methods is the ability to measure the direction and velocity of a moving target, in addition to its range. In 1842, an Austrian by the name of Johann Doppler published a paper describing what has since become known as the Doppler effect: the frequency of an energy wave reflected from an object in motion is a function of the relative velocity between the object and the observer. This subject will be discussed in more detail in Chapter 8.

In practice, a beam of amplitude-modulated laser, RF, or acoustical energy is directed towards the target as illustrated in Figure 6-1. A small portion of this wave (potentially up to six orders of magnitude less in amplitude) is reflected by the object surface back to the detector (Chen, et al., 1993). Improved measurement accuracy and increased range can be achieved when cooperative targets are attached to the objects of interest to increase the power density of the reflected signal. The returned energy is compared to a simultaneously generated reference that has been split off from the original signal, and the relative phase shift between the two is measured to ascertain the round-trip distance the wave has traveled. As with time-of-flight rangefinders, the paths of the source and the reflected beam are essentially coaxial, preventing the *missing parts* problem.

For high-frequency RF- or laser-based systems, detection is usually preceded by heterodyning the reference and received signals with an intermediate frequency (the relative phase shift is preserved) to allow the phase detector to operate at a more convenient lower frequency (Vuylsteke, 1990). The phase shift expressed as a function of distance to the reflecting target surface is (Woodbury, et al., 1993):

$$\phi = \frac{4\pi d}{\lambda}$$

where:

ϕ = phase shift
d = distance to target
λ = modulation wavelength.

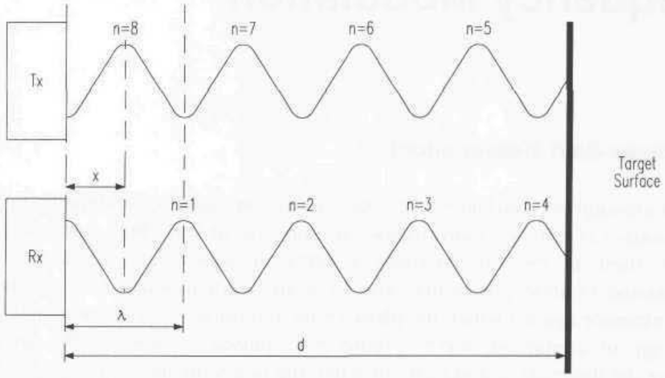

Figure 6-1. Relationship between outgoing and reflected waveforms, where x is the distance corresponding to the differential phase ϕ (adapted from Woodbury, et al., 1993).

The desired distance to target d as a function of the measured phase shift ϕ is therefore given by:

$$d = \frac{\phi\,\lambda}{4\pi} = \frac{\phi\, c}{4\pi f}$$

where:

c = speed of light.
f = modulation frequency.

The phase shift between outgoing and reflected sine waves can be measured by multiplying the two signals together in an electronic mixer, then averaging the product over many modulation cycles (Woodbury, et al., 1993). This integrating process can be relatively time consuming, making it difficult to achieve extremely rapid update rates. The result can be expressed mathematically as follows (Woodbury, et al., 1993):

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \sin\left(\frac{2\pi c}{\lambda} t + \frac{4\pi d}{\lambda}\right) \sin\left(\frac{2\pi c}{\lambda}\right) dt$$

which reduces to:

$$A \cos\left(\frac{4\pi d}{\lambda}\right)$$

where:

t = time
T = averaging interval
A = amplitude factor from gain of integrating amplifier.

From the earlier expression for ϕ, it can be seen that the quantity actually measured is in fact the *cosine* of the phase shift and not the phase shift itself (Woodbury, et al., 1993). This situation introduces a so-called *ambiguity interval* for scenarios where the round-trip distance exceeds the modulation wavelength λ (i.e., the phase measurement becomes ambiguous once ϕ exceeds 360 degrees). Conrad and Sampson (1990) define this ambiguity interval as the maximum range that allows the phase difference to go through one complete cycle of 360 degrees:

$$R_a = \frac{c}{2f}$$

where:

R_a = ambiguity range interval.

Referring again to Figure 6-1, it can be seen that the total round-trip distance $2d$ is equal to some integer number of wavelengths $n\lambda$ plus the fractional wavelength distance x associated with the phase shift. Since the cosine relationship is not single-valued for all of ϕ, there will be more than one distance d corresponding to any given phase-shift measurement (Woodbury, et al., 1993):

$$\cos\phi = \cos\left(\frac{4\pi d}{\lambda}\right) = \cos\left(\frac{2\pi(x + n\lambda)}{\lambda}\right)$$

where:

$d = (x + n\lambda)/2$ = true distance to target
x = distance corresponding to differential phase ϕ
n = number of complete modulation cycles.

Careful re-examination of Figure 6-1, in fact, shows that the cosine function is not single-valued even within a solitary wavelength interval of 360 degrees.

Accordingly, if only the cosine of the phase angle is measured, the ambiguity interval must be further reduced to half the modulation wavelength, or 180 degrees (Scott, 1990). In addition, the slope of the curve is such that the *rate of change* of the non-linear cosine function is not constant over the range of $0 \leq \phi \leq 180$ degrees, and is in fact zero at either extreme. The achievable accuracy of the phase-shift measurement technique thus varies as a function of target distance, from best-case performance for a phase angle of 90 degrees to worst case at 0 and 180 degrees. For this reason, the useable measurement range is typically even further limited to 90 percent of the 180-degree ambiguity interval (Chen, et al., 1993).

A common solution to this problem involves taking a second measurement of the same scene but with a 90-degree phase shift introduced into the reference waveform, the net effect being the sine of the phase angle is then measured instead of the cosine. This additional information (i.e., both sine and cosine measurements) can be used to expand the phase angle ambiguity interval to the full 360 degree limit previously discussed (Scott, 1990). Furthermore, an overall improvement in system accuracy is achieved, as for every region where the cosine measurement is insensitive (i.e., zero slope), the complementary sine measurement will be at peak sensitivity (Woodbury, et al., 1993).

Nevertheless, the unavoidable potential for erroneous information as a result of the ambiguity interval is a detracting factor in the case of phase-detection schemes. Some applications simply avoid such problems by arranging the optical path in such a fashion as to ensure the maximum possible range is always less than the ambiguity interval (Figure 6-2). Alternatively, successive measurements of the same target using two different modulation frequencies can be performed, resulting in two equations with two unknowns, allowing both x and n (in the previous equation) to be uniquely determined. Kerr (1988) describes such an implementation using modulation frequencies of 6 and 32 MHz.

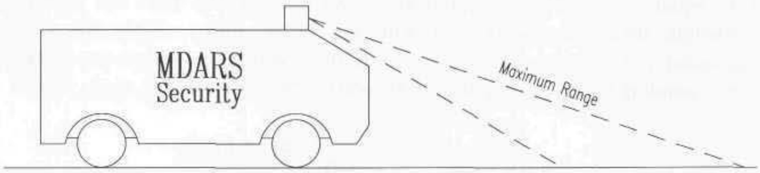

Figure 6-2. By limiting the maximum distance measured to be less than the range *ambiguity interval* R_a, erroneous distance measurements can be avoided.

For square-wave modulation at the relatively low frequencies typical of ultrasonic systems (20-200 KHz), the phase difference between incoming and outgoing waveforms can be measured with the simple linear circuit shown in Figure 6-3 (Figueroa & Barbieri, 1991a). The output of the *exclusive-or* gate goes high whenever its inputs are at opposite logic levels, generating a voltage across capacitor C_1 that is proportional to the phase shift. For example, when the two signals are in phase (i.e., $\phi = 0$), the gate output stays low and V is zero; maximum output voltage occurs when ϕ reaches 180 degrees. While easy to implement, this simplistic approach is limited to very low frequencies and may require frequent calibration to compensate for drifts and offsets due to component aging or changes in ambient conditions (Figueroa & Lamancusa, 1992).

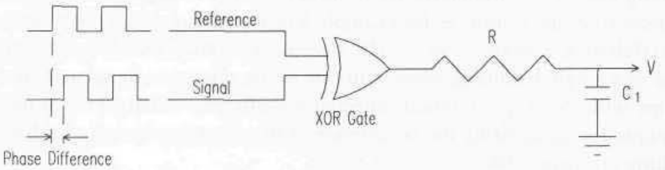

Figure 6-3. At low frequencies typical of ultrasonic systems, a simple phase-detection circuit based on an *exclusive-or* gate will generate an analog output voltage proportional to the phase difference seen by the inputs (adapted from Figueroa & Barbieri, 1991a).

Figueroa and Barbieri (1991a; 1991b) report an interesting method for extending the ambiguity interval in ultrasonic phase-detection systems through frequency division of the received and reference signals. Since the span of meaningful comparison is limited (best case) to one wavelength, λ, it stands to reason that decreasing the frequency of the phase detector inputs by some common factor will increase λ by a similar amount. The concept is illustrated in Figure 6-4 below. Due to the very short wavelength of ultrasonic energy (i.e., about 0.25 inches for the Polaroid system at 49.1 KHz), the total effective range is still only 4 inches after dividing the detector inputs by a factor of 16. Due to this inherent range limitation, ultrasonic phase-detection ranging systems are not extensively applied in mobile robotic applications, although Figueroa and Lamancusa (1992) describe a hybrid approach used to improve the accuracy of TOF ranging for three-dimensional position location (see Chapter 15).

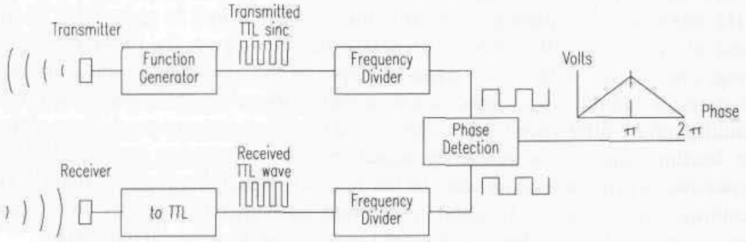

Figure 6-4. Dividing the input frequencies to the phase comparator by some common integer value will extend the ambiguity interval by the same factor, at the expense of resolution (adapted from Figueroa & Barbieri, 1991a).

Laser-based continuous-wave ranging originated out of work performed at the Stanford Research Institute in the 1970s (Nitzan, et al., 1977). Range accuracies approach those achievable by pulsed laser TOF methods. Only a slight advantage is gained over pulsed TOF rangefinding, however, since the time-measurement problem is replaced by the need for fairly sophisticated phase-measurement electronics (Depkovich & Wolfe, 1984). In addition, problems with the phase-shift measurement approach are routinely encountered in situations where the outgoing energy is simultaneously reflected from two target surfaces at different distances from the sensor, as for example when scanning past a prominent vertical edge (Hebert & Krotkov, 1991). The system electronics are set up to compare the phase of a single incoming wave with that of the reference signal and are not able to cope with two superimposed reflected waveforms. Adams (1993) describes a technique for recognizing the occurrence of this situation in order to discount the resulting erroneous data.

6.1.1 ERIM 3-D Vision Systems

The *Adaptive Suspension Vehicle (ASV)* developed at Ohio State University (Patterson, et al., 1984) and the *Autonomous Land Vehicle (ALV)* developed by Martin Marietta Denver Aerospace were the premier mobile robot projects sponsored by the Defense Advanced Research Projects Agency (DARPA) in the 1980s under the *Strategic Computing Program.* In support of these efforts, the Environmental Research Institute of Michigan (ERIM) was tasked to develop an advanced three-dimensional vision system to meet the close-in navigation and collision avoidance needs of a mobile platform. The initial design, known as the *Adaptive Suspension Vehicle Sensor* (Figure 6-5), operates on the principle of optical radar and determines range to a point through phase-shift measurement using a CW laser source (Beyer, et al., 1987).

The ranging sequence begins with the transmission of an amplitude-modulated laser beam that illuminates an object and is partially reflected back to the detector, generating a representative signal that is amplified and filtered to extract the 16-MHz modulation frequency. The amplitude of the signal is picked off at this point to produce a reflectance video image for viewing or for two-dimensional image processing. A reference signal is output by the modulation oscillator; both the detector and reference signals are then sent to the comparator electronics. The resulting phase difference is determined by a time-measurement technique, where the leading edge of the reference signal initiates a counting sequence that is terminated when the leading edge of the returned signal enters the counter. The resulting count value is a function of the phase difference between the two signals and is converted to an 8-bit digital word representing the range to the scene.

Three-dimensional images are produced by the *ASV* sensor through the use of scanning optics. The mechanism consists of a nodding mirror and a rotating polygonal mirror with four reflective surfaces as shown in Figure 6-6. The polygonal mirror pans the transmitted laser beam in azimuth across the ground, creating a scan line at a set distance in the front of the vehicle. The scan line is deflected by the objects and surfaces in the observed region and forms a contour of the scene across the sensor's horizontal field of view. The third dimension is added by the nodding mirror which tilts the beam in discrete elevation increments. A complete image is created by scanning the laser in a left-to-right and bottom-to-top raster pattern.

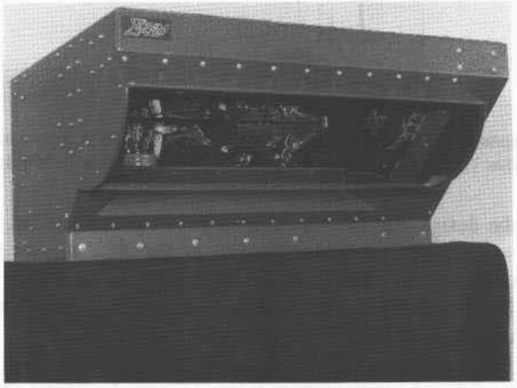

Figure 6-5. The *Adaptive Suspension Vehicle Sensor* (courtesy Environmental Research Institute of Michigan).

The returning signals share the same path through the nodding mirror and rotating polygon (actually slightly offset) but are split through a separate glass optical chain to the detector. The scan rate of 180 lines per second is a function of the field of view and desired frame rate, determined by the vehicle's maximum forward velocity (10 feet/second in this case). The size, weight, and required velocities of the mirrors precluded the use of galvanometers in the system design; the rotating and nodding mirrors therefore are servo driven.

An 820-nanometer gallium arsenide (GaAs) laser diode with collimating and expansion optics is used to produce a 6-inch diameter laser footprint at 30 feet. The detector is a silicon avalanche photodiode, optically filtered to match the laser wavelength. The laser source, detector, scanning optics, and drive motors are housed in a single enclosure situated at a height of 8 feet, looking down upon the field of view. The scanning laser beam strikes the ground between 2 and 30 feet in front of the vehicle, with a 22-foot wide horizontal scan line at the maximum distance of 30 feet. (The major factor limiting the useful range of the system is the measurement ambiguity that occurs when the phase difference between the reference and returned energy exceeds 360 degrees.) The 2-Hz system update rate creates a new image of the scene for every 5 feet of forward motion at the vehicle's maximum speed of 10 feet/second.

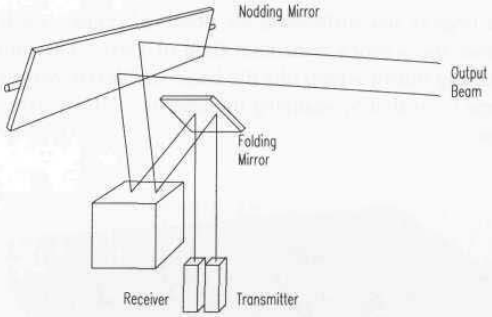

Figure 6-6. Scanning and nodding mirror arrangement in the ERIM laser rangefinder for the *Adaptive Suspension Vehicle* (courtesy Environmental Research Institute of Michigan).

Following the design and fabrication of the *ASV sensor*, ERIM undertook the task of developing a similar device known as the *ALV sensor* for DARPA's autonomous land vehicle. The two instruments were essentially the same in configuration and function but with modified performance specifications to meet the needs of the individual mobile platforms (Table 6-1).

Table 6-1. Selected specifications for the *Adaptive Suspension Vehicle* and *Autonomous Land Vehicle* scanning laser rangefinders.

Parameter	ASV	ALV	Units
Horizontal FOV	80	80	degrees
Vertical FOV	60	30	degrees
Beamwidth	1	0.5	degrees
Frame rate	2	2	Hz
Scan lines per frame	128	64	
Pixels per scan line	128	256	
Maximum range	32	64	feet
Vertical scan	10	20	degrees
Wavelength	820	820	nanometers
Power	24	24	volts
	450	450	watts
Size	14 by 26 by 22	14 by 29 by 22	inches
Weight	85	85	pounds

An advanced ranging device known as the *Multispectral ALV Sensor* was later developed for exterior applications addressing rugged cross-country traversal as opposed to the relatively uniform road surfaces seen in the initial tests of the autonomous land vehicle concept. The variations in terrain, surface cover, and vegetation encountered in off-road scenarios require an effective means to distinguish between earth, rocks, grass, trees, water, and other natural features.

The scanner mechanism for the multispectral sensor was essentially identical to the scanners developed for the earlier *ASV* and *ALV* sensors, the only significant difference being the substitution of a hexagonal rotating mirror instead of a square mirror for panning the beam in azimuth. This configuration caused the transmitted and returned signals to impinge on separate mirrored surfaces, resulting in reduced crosstalk and simplified sensor alignment (Figure 6-7). The nodding mirror for tilting the beam in elevation remained largely unchanged.

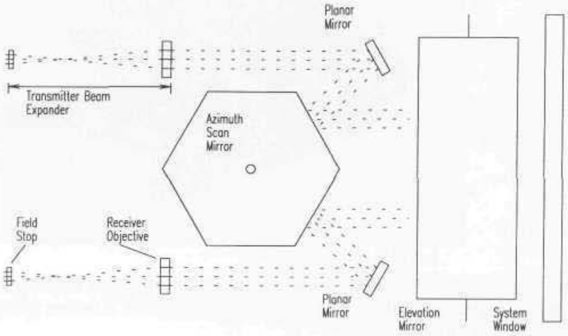

Figure 6-7. Hexagonal rotating mirror used in the multispectral scanner reduces crosstalk and simplifies mirror alignment (courtesy Environmental Research Institute of Michigan).

The mass of the scanning mechanism plus the plurality of lasers, optics, and detectors made the multispectral sensor large (12 by 3 by 2 feet) and heavy (600 pounds), increasing the complexity of the control and analysis required to produce results. The multiple frequency sources, corresponding detectors, detector cooling system, and scanner resulted in significant power consumption: 15 kilowatts!

6.1.2 Perceptron *LASAR*

Perceptron Corporation, Farmington Hills, MI, has developed and is currently in production of *LASAR*, the AM-modulated 3-D laser scanner shown in Figure 6-8. Intended for industrial machine vision applications, versions of this device have already been used in navigational guidance, bin-picking, hazardous inspection, and mining scenarios. The sensor employs a nodding mirror in conjunction with a rotating-polygon assembly to achieve a 45-degree symmetrical field of view. At full-frame (1024 x 1024) resolution, a single update takes 6.4 seconds, with increased frame rates possible at lower resolutions. The maximum operating range of the *LASAR* system is around 40 meters, with an advertised single-frame range accuracy of ±2 millimeters at a distance of 2 meters. Frame rates up to 10 Hz and operating ranges in excess of 100 meters have been demonstrated in specially configured versions of the device.

Figure 6-8. The *LASAR* 3-D scanner achieves a range-measurement accuracy of 2 millimeters over a 45- by 45-degree field of view at a stand-off distance of 2 meters (courtesy Perceptron Corp.).

6.1.3 Odetics Scanning Laser Imaging System

Odetics, Inc., Anaheim, CA, developed an adaptive and versatile scanning laser rangefinder in the early 1980s for use on *ODEX 1*, the six-legged walking robot shown in Figure 6-9 (Binger & Harris, 1987; Byrd & DeVries, 1990). The system determines distance by phase-shift measurement, constructing three-dimensional range pictures by panning and tilting the sensor across the field of view. The phase-shift measurement technique was selected over acoustic-ranging, stereo-vision, and structured-light alternatives because of the inherent accuracy and fast update rate.

The imaging system is broken down into the two major subelements depicted in Figure 6-10: the *scan unit* and the *electronics unit*. The *scan unit* houses the laser source, the photodetector, and the scanning mechanism. The laser source is a GaAlAs laser diode emitting at a wavelength of 820 nanometers, with power output adjustable under software control between 1 to 50 milliwatts. Detection of the returned energy is achieved through use of an avalanche photodiode whose output is routed to the phase-measuring electronics.

The second subelement, the *electronics unit*, contains the range calculating and video processor as well as a programmable frame buffer interface. The range and video processor is responsible for controlling the laser transmission, activation of the scanning mechanism, detection of the returning energy, and determination of range values. Distance is calculated through a proprietary phase-detection scheme, reported to be fast, fully digital, and self-calibrating with a high signal-to-noise ratio. The minimum observable range is 1.5 feet, while the maximum range without ambiguity due to phase shifts greater than 360 degrees is 30.74 feet.

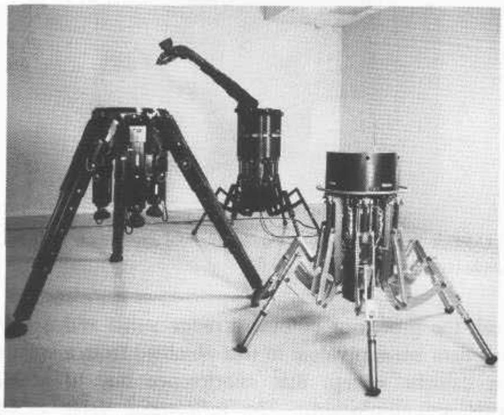

Figure 6-9 The *Scanning Laser Imaging System* was initially developed for use on the *Odex Series* of six-legged walking robots (courtesy Odetics, Inc.).

The scanning hardware consists of a rotating polygonal mirror that pans the laser beam across the scene and a planar mirror whose back-and-forth nodding motion tilts the beam for a realizable field of view of 60 degrees in azimuth and 60 degrees in elevation. The scanning sequence follows a raster-scan pattern and can illuminate and detect an array of 128 by 128 pixels at a frame rate of 1.2 Hz (Boltinghouse, et al., 1990).

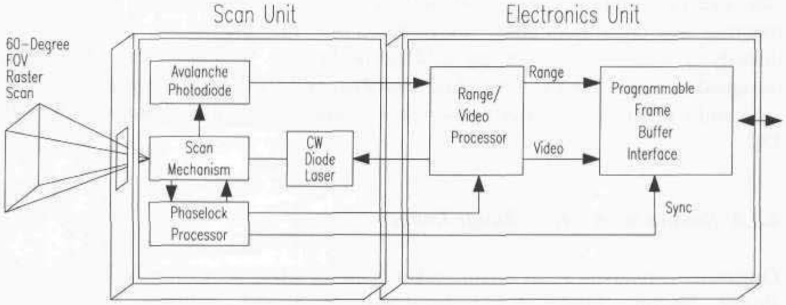

Figure 6-10. Block diagram of the Odetics scanning laser rangefinder (courtesy Odetics, Inc.).

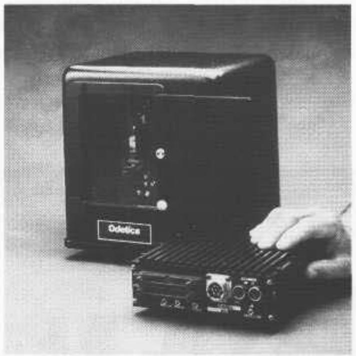

Figure 6-11. The Odetics *Scanning Laser Imaging System* captures a 128- by 128-pixel image every 835 milliseconds (courtesy Odetics, Inc.).

For each pixel, the processor outputs a range value and a video reflectance value. The video data are equivalent to that obtained from a standard black-and-white television camera, except that interference due to ambient light and shadowing effects are eliminated. The reflectance value is compared to a prespecified threshold to eliminate pixels with insufficient return intensity to be properly processed, thereby eliminating potentially invalid range data; range values are set to maximum for all such pixels (Boltinghouse & Larsen, 1989). A three-by-three *neighborhood median filter* is used to further filter out noise from data qualification, specular reflection, and impulse response (Larson & Boltinghouse, 1988).

The output format is a 16-bit data word consisting of the range value in either 8 or 9 bits, and the video information in either 8 or 7 bits, respectively. The resulting range resolution for the system is 1.44 inches for the 8-bit format, and 0.72 inch with 9 bits. A buffer interface provides interim storage of the data and can execute single-word or whole-block direct-memory-access transfers to external host controllers under program control. Information can also be routed directly to a host without being held in the buffer. Currently, the interface is designed to support *VAX, VME-Bus, Multibus*, and IBM-*PC/AT* equipment. The scan and electronics unit together weigh 31 pounds and require 2 amps at 28 volts DC.

6.1.4 Sandia *Scannerless Range Imager*

Originally conceived as an active seeker head for smart weapons, the *Scannerless Range Imager* (Figure 6-12) developed at Sandia National Laboratories, Albuquerque, NM, computes three-dimensional range information without need for mechanical or solid-state scanning. A laser diode or LED array is used to illuminate an entire scene in similar fashion to Robotic Vision System's pulsed TOF *Long Optical Range and Detection System* described at the end of the previous chapter. The Sandia approach, however, employs an amplitude-modulated continuous-wave source in conjunction with a single CCD camera, and determines ranges to all pixel elements in essentially simultaneous fashion based on the perceived round-trip phase shift (Frazier, 1994). Phase-shift measurement is rather elegantly accomplished by converting the phase difference to a more easily quantified intensity representation through use of a *microchannel-plate image intensifier* as shown in the block diagram of Figure 6-13.

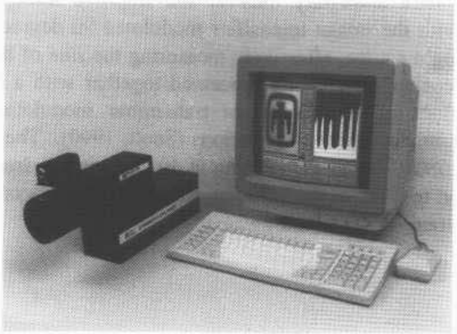

Figure 6-12. The Sandia *Scannerless Range Imager* employs an amplitude-modulated CW laser source in conjunction with a single CCD camera (courtesy Sandia National Laboratories).

Reflected energy from the illuminated scene is focused by the receiver optics upon a photocathode element that creates a stream of electrons modulated in accordance with the amplitude variations of the incoming light. The sinusoidal laser-modulation signal f_m is coupled to a thin conductive sheet (i.e., analogous to the *grid* of a vacuum tube), as shown in the above figure, to control the flow of electrons from the photocathode into the microchannel plate (Scott, 1990). The electron stream is amplified through secondary emissions as it passes through the microchannel plate, and converted back to optical energy upon striking the phosphor screen as illustrated. Since the gain of the image intensifier stage is in this fashion modulated at the same frequency as the outgoing optical energy, the magnitude of phosphor radiance is thus a function of the *cosine* of the range-dependent phase angle (i.e., due to constructive and destructive interference). A 210-frames/second Dalsa CCD camera is coupled to the phosphor screen by way of a coherent fiber-optic bundle to serve as an integrating 256-by-256 detector array (Weiss, 1994).

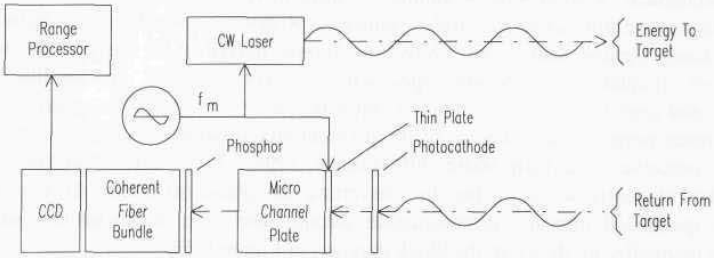

Figure 6-13. Range values are computed for all pixels in the CCD detector array based on the observed phase shift (adapted from Scott, 1990).

To expand the phase ambiguity interval and improve resolution, a second image is obtained with the image intensifier modulated 90 degrees out of phase with respect to the light source, effectively measuring the *sine* of the phase angle. These "sine" and "cosine" images are processed together with a baseline image taken under conditions of no receiver or transmitter modulation in order to eliminate non-range-related intensity variations (Scott, 1990). The current system update rate using a 68040-based PC running at 40 MHz is one frame per second, but will be expanded to 8 Hz in the very near future through incorporation of *TI-C40* digital signal processor (DSP) hardware.

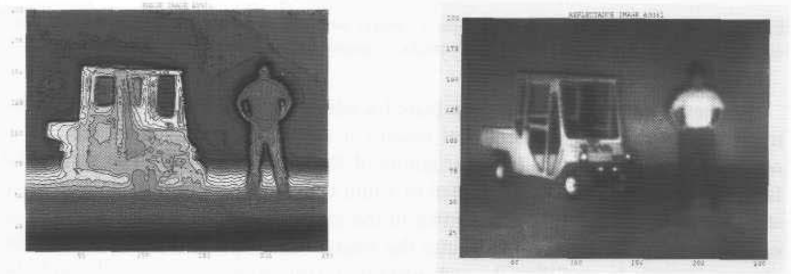

Figure 6-14. Resulting range image (left) and reflectance image (right) for a typical outdoor scene using an array of 660-nanometer (red) LEDs (courtesy Sandia National Laboratories).

Due to its structural simplicity, relatively low cost, and demonstrated potential for high-bandwidth, medium-resolution range data, the Sandia *Scannerless Range Imager* is being investigated for use on a number of robotic platforms, including the MDARS-Exterior system. One existing prototype of the sensor employs a 20-watt laser diode modulated at 5 MHz, resulting in a 90-foot ambiguity interval with a range resolution of 1 foot and a maximum range of 2,000 feet at night

(Weiss, 1994). Nighttime operation using eye-safe LED emitters has also been demonstrated out to 200 feet; representative range and reflectance images at a distance of approximately 60 feet are presented in Figure 6-14. Potential problems still being investigated include the significant power and cooling requirements for the laser source, and attainment of sufficient signal-to-noise ratio for reliable daytime operation.

6.1.5 ESP *Optical Ranging System*

The *Optical Ranging System (ORS-1)* is a low-cost near-infrared rangefinder (Figure 6-15) developed in 1989 by ESP Technologies, Inc., Lawrenceville, NJ, for use in autonomous robotic cart navigation in factories and similar environments. A 2-milliwatt 820-nanometer LED source, 100-percent modulated at 5 MHz, forms a collimated 1-inch diameter transmit beam that is unconditionally eye-safe. Reflected radiation is focused by a 4-inch diameter coaxial Fresnel lens onto a photodetector; the measured phase shift is proportional to the round-trip distance to the illuminated object. An adaptation of an earlier prototype developed by AT&T (Miller & Wagner, 1987), the *ORS-1* provides three outputs: range and angle of the target and an automatic gain control (AGC) signal. Range resolution at 20 feet is approximately 2.5 inches, while angular resolution is about 1 inch at a distance of 5 feet.

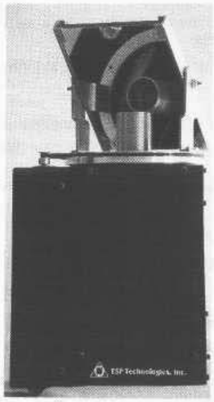

Figure 6-15. The *ORS-1 Optical Ranging System* determines range through phase-shift measurement using an eye-safe near-infrared LED source (courtesy ESP Technologies, Inc.).

The AGC output signal is inversely proportional to the received signal strength and provides information about a target's near-infrared reflectivity, warning against insufficient or excessive signal return (ESP, 1992). Useable range results are produced only when the corresponding gain signal is within a predetermined operating range. A rotating mirror mounted at 45 degrees to the optical axis provides 360-degree polar-coordinate coverage (Figure 6-16). It is driven at one to two revolutions per second by a motor fitted with an integral incremental encoder and an optical indexing sensor that signals the completion of each revolution. The system is capable of simultaneous operation as a wideband optical communication receiver (Miller & Wagner, 1987).

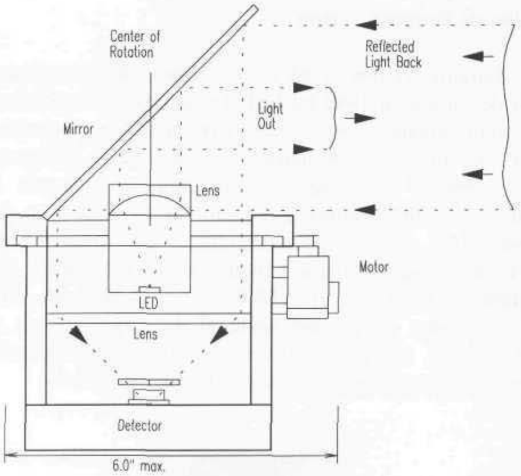

Figure 6-16. Schematic drawing of the *ORS-1* ranging system (courtesy ESP Technologies, Inc.).

Table 6-2. Selected specifications for the LED-based near-infrared *Optical Ranging System*.

Parameter	Value	Units
Maximum range	20	feet
Minimum range	2	feet
Accuracy	< 6	inches
AGC output	1-5	volts
Output power	2	milliwatts
Beamwidth	1	inch
Dimensions	6 x 6 x 12	inches
Weight		pounds
Power	12	volts DC
	2	amps

A representative ranger scan taken in a laboratory environment is shown in Figure 6-17. The ranger is mounted on the robot cart and located at the (0,0) position marked by a cross in the center of the plot. The data collection for this plot corresponds to a single mirror rotation taking approximately one second. Note the absence of any data points in regions (i.e., segment AB in the lower right corner), where the return signal is outside the AGC window; and therefore no data are accumulated. All of the objects in the room are found to correspond accurately to their positions as indicated by the range measurements (Miller & Wagner, 1987).

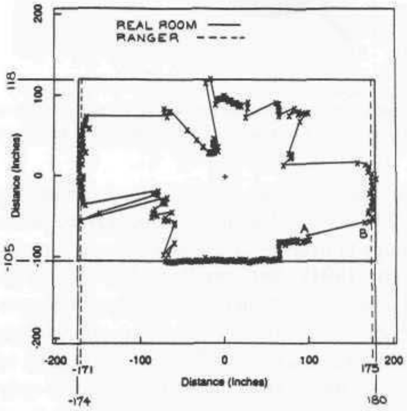

Figure 6-17. This actual scan of a laboratory environment resulted from a single mirror rotation taking approximately one second (courtesy ESP Technologies, Inc.).

6.1.6 Acuity Research *AccuRange 3000*

Acuity Research, Inc., Menlo Park, CA, has recently introduced an interesting product capable of acquiring unambiguous range data from 0 to 20 meters using a proprietary technique wherein the optical beam path is part of an oscillatory feedback loop. The *AccuRange 3000* (Figure 6-18) projects a collimated beam of near-infrared or visible laser light, amplitude modulated with a non-sinusoidal waveform at a 50-percent duty cycle (Clark, 1994). A 2.5-inch collection aperture surrounding the laser diode emitter on the front face of the cylindrical housing gathers any reflected energy returning from the target and subsequently disables the laser source. Once the sensed energy disappears as a consequence of this action, the laser is re-energized, whereupon the cycle repeats. The net effect of this innovative approach, which requires minimal circuitry, is a square-wave output with a period of oscillation proportional to the measured range.

Figure 6-18. The *AccuRange 3000* distance measuring sensor provides a square-wave output that varies inversely in frequency as a function of range (courtesy Acuity Research, Inc.).

The frequency of the output signal varies from approximately 50 MHz at zero range to 4 MHz at 20 meters. Distance to target can be determined through use of a frequency-to-voltage converter, or by measuring the period with a hardware or software timer (Clark, 1994). Separate 0- to 10-volt analog outputs are provided for returned signal amplitude, ambient light, and temperature to facilitate dynamic calibration for optimal accuracy in demanding applications. The range output changes within 250 nanoseconds to reflect any change in target distance, and all outputs are updated within a worst-case time frame of only three microseconds. This rapid response rate (up to 312.5 KHz for all outputs with the optional SCSI interface) allows the beam to be manipulated at a 1000- to 2000-Hz rate with the mechanical-scanner option shown in Figure 6-19 below. A 45-degree balanced-mirror arrangement is rotated under servo control to deflect the coaxial outgoing and incoming beams for full 360-degree planar coverage.

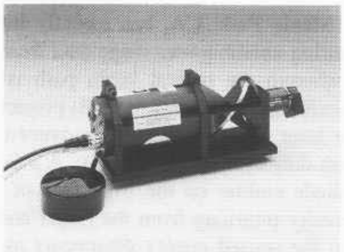

Figure 6-19. A 360-degree beam-deflection capability is provided by an optional single axis rotating scanner (courtesy Acuity Research, Inc.).

Table 6-3. Selected specifications for the Acuity *AccuRange 3000* distance measurement sensor.

Parameter	Value	Units
Laser output	5	milliwatts
Beam divergence	0.5	milliradians
Wavelength	780/670	nanometers
Maximum range	20	meters
Minimum range	0	meters
Accuracy	2	millimeters
Sample rate	up to 312.5	KHz
Diameter	3	inches
Length	5.5	inches
Weight	18	ounces
Power	5 and 12	volts DC
	250 and 50	milliamps

6.1.7 TRC *Light Direction and Ranging System*

Transitions Research Corporation (TRC), Danbury, CT, offers a low-cost *lidar* system for detecting obstacles in the vicinity of a robot and/or estimating position from local landmarks or retroreflective targets (see Chapter 15), based on the previously discussed Acuity Research *AccuRange 3000* unit. TRC adds a two-degree-of-freedom scanning mechanism employing a gold front-surfaced mirror specially mounted on a vertical pan axis that rotates between 200 and 900 rpm (Figure 6-20). The tilt axis of the scanner is mechanically synchronized to nod one complete cycle (down 45 degrees and back to horizontal) per 10 horizontal scans, effectively creating a protective spiral of detection coverage around the robot (TRC, 1994). The tilt axis can be mechanically disabled if so desired for 360-degree azimuthal scanning at a fixed elevation angle.

Table 6-4. Selected specifications for the TRC *Light Direction and Ranging System.*

Parameter	Value	Units
Maximum range	12	meters
Modulation frequency	2	MHz
Accuracy (range)	25	millimeters
Resolution (range)	5	millimeters
(azimuth)	0.18	degrees
Sample rate	25	KHz
Size (scanner)	13 x 13 x 35	centimeters
(electronics)	30 x 26 x 5	
Weight	2	kilograms

A *68HC11* microprocessor automatically compensates for variations in ambient lighting and sensor temperature, and reports range, bearing, and elevation data via an Ethernet or RS-232 interface. Power requirements are 500 milliamps at 12 volts DC and 100 milliamps at 5 volts DC. Typical operating parameters are listed in Table 6-4.

Figure 6-20. The TRC *Light Direction and Ranging System* incorporates a two-axis scanner to provide full-volume coverage sweeping 360 degrees in azimuth and 45 degrees in elevation (courtesy Transitions Research Corp.).

6.2 Frequency Modulation

A closely related alternative to the amplitude-modulated phase-shift-measurement ranging scheme is frequency-modulated (FM) radar. This technique involves transmission of a continuous electromagnetic wave modulated by a periodic triangular signal that adjusts the carrier frequency above and below the mean frequency f_0 as shown in Figure 6-21. The transmitter emits a signal that varies in frequency as a linear function of time:

$$f(t) = f_o + at$$

where:

a = some constant
t = elapsed time.

This signal is reflected from a target and arrives at the receiver at time $t + T$.

$$T = \frac{2d}{c}$$

where:

T = round-trip propagation time
d = distance to target
c = speed of light.

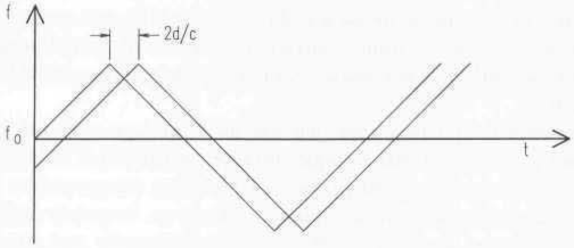

Figure 6-21. The received frequency curve is shifted along the time axis relative to the reference frequency.

The received signal is compared with a reference signal taken directly from the transmitter. The received frequency curve will be displaced along the time axis relative to the reference frequency curve by an amount equal to the time required for wave propagation to the target and back. (There might also be a vertical displacement of the received waveform along the frequency axis, due to the Doppler effect.) These two frequencies when combined in the mixer produce a beat frequency F_b:

$$F_b = f(t) - f(T+t) = aT$$

where:

a = constant.

This beat frequency is measured and used to calculate the distance to the object:

$$d = \frac{F_b c}{4 F_r F_d}$$

where:

d = range to target
c = speed of light
F_b = beat frequency
F_r = repetition (modulation) frequency
F_d = total FM frequency deviation.

Distance measurement is therefore directly proportional to the difference or beat frequency, and is as accurate as the linearity of the frequency variation over the counting interval.

Advances in wavelength control of laser diodes now permit this radar ranging technique to be used with lasers. The frequency or wavelength of a laser diode can be shifted by varying its temperature. Consider an example where the wavelength of an 850-nanometer laser diode is shifted by 0.05 nanometers in four seconds: the corresponding frequency shift is 5.17 MHz per nanosecond. This laser beam, when reflected from a surface 1 meter away, would produce a beat frequency of 34.5 MHz. The linearity of the frequency shift controls the accuracy of the system.

The frequency-modulation approach has an advantage over the phase-shift-measurement technique in that a single distance measurement is not ambiguous. (Recall phase-shift systems must perform two or more measurements at different modulation frequencies to be unambiguous.) However, frequency modulation has several disadvantages associated with the required linearity and repeatability of the frequency ramp, as well as the coherence of the laser beam in optical systems. As a consequence, most commercially available FMCW ranging systems are radar based, while laser devices tend to favor TOF and phase-detection methods.

6.2.1 VRSS Automotive Collision Avoidance Radar

One of the first practical automotive collision avoidance radar systems was developed by Vehicle Radar Safety Systems (VRSS) of Mt. Clemens, MI. This modified Doppler radar unit is intended to alert drivers to potentially dangerous situations. A grill-mounted miniaturized microwave horn antenna sends out a narrow-beam signal that detects only those objects directly in the path of the vehicle, ignoring targets (such as road signs and parked cars) on either side. When the radar signal is reflected from a slower-moving or stationary target, it is detected by the antenna and passed to an under-the-hood electronic signal processor (VRSS, 1983).

The signal processor continuously computes the host vehicle speed and acceleration, distance to the target, relative velocity, and target acceleration. If these parameters collectively require the driver to take any corrective or precautionary action, a warning buzzer and signal light are activated on a special dashboard monitor. An *alert* signal lights up when an object or slower-moving vehicle is detected in the path of the host vehicle. If the target range continues to decrease, and the system determines that a collision is possible, a *warning* light and buzzer signal the driver to respond accordingly. If range continues to decrease with no reduction in relative velocity, then a *danger* light illuminates indicating the need for immediate action.

A filter in the signal processor provides for an optimum operating range for the system, based on the relative velocity between the vehicle and the perceived object. The response window corresponds to a calculated difference in speed of between 0.1 and 30 miles per hour (VRSS, 1983). If the speed differential exceeds 30 miles per hour, the filter circuit delays signals to the dashboard monitor. This helps to eliminate false signals and signals that might otherwise be caused by approaching vehicles when passing another vehicle on a two-lane highway.

The VRSS collision warning system has been tested in over a million miles of driving conditions in fog, rain, snow, and ice with good results. The present model was perfected in 1983 after 36 years of research, and approved by the FCC in 1985. Although aimed initially at the bus and trucking industries, the low-cost unit offers convincing proof that small, low-power radar systems offer a practical alternative to ultrasonic rangefinders for the collision avoidance needs of a mobile robot operating in outdoor scenarios.

Table 6-5. Selected specifications for the VRSS automotive radar.

Parameter	Value	Units
Effective range	1-300	feet
Accuracy	1.5	percent
Update rate	200	Hz
Operating frequency	24	GHz
RF power	10	milliwatts
Beamwidth (horizontal)	6	degrees
(vertical)	6	degrees
Size (antenna)	3 x 4	inches
(electronics unit)	4 x 5 x 2	inches
Weight (total)	4	pounds
Power	12	volts DC
	12	watts

6.2.2 VORAD Vehicle Detection and Driver Alert System

VORAD (Vehicle Onboard Radar) Safety Systems, Inc., San Diego, CA, has also developed a commercial millimeter-wave FMCW Doppler radar system designed for use on a motor vehicle (VORAD, undated). The *Vehicle Collision Warning System* employs a 5- by 5-inch antenna/transmitter/receiver package mounted on the front grill of a vehicle to monitor speed and distance to other traffic or obstacles on the road (Figure 6-22). The flat etched-array antenna radiates approximately 0.5 milliwatts of power at 24.725 GHz directly down the roadway in a narrow directional beam. A GUNN diode is used for the transmitter, while the receiver employs a balanced-mixer detector (Woll, 1993).

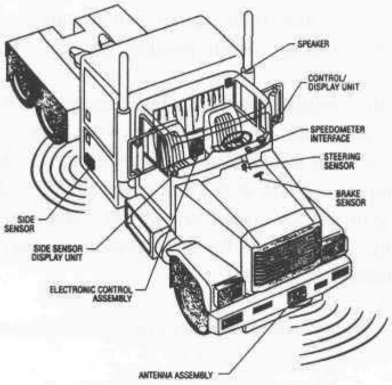

Figure 6-22. The forward-looking antenna/transmitter/receiver module is mounted on the front of the vehicle at a height between 50 and 125 centimeters, while an optional side antenna can be installed as shown for blind-spot protection (courtesy VORAD Safety Systems, Inc.).

The *Electronics Control Assembly* located in the passenger compartment or cab (see again Figure 6-22) can individually distinguish up to 20 moving or stationary objects (Siuru, 1994) out to a maximum range of 350 feet; the closest three targets within a prespecified warning distance are tracked at a 30-Hz rate. A Motorola *DSP 56001* and an Intel *87C196* microprocessor calculate range and range-rate information from the RF data and analyze the results in conjunction with vehicle-velocity, braking, and steering-angle information. The *Control Display Unit* alerts the operator if warranted of potentially hazardous driving situations with a series of caution lights and audible beeps.

As an optional feature, the *Vehicle Collision Warning System* offers blind-spot detection along the right-hand side of the vehicle out to 15 feet. The *Side Sensor* transmitter employs a *dielectric resonant oscillator* operating in pulsed Doppler mode at 10.525 GHz, using a flat etched-array antenna with a beamwidth of about 70 degrees (Woll, 1993). The system microprocessor in the *Electronics Control Assembly* analyzes the signal strength and frequency components from the *Side Sensor* subsystem in conjunction with vehicle speed and steering inputs, and activates audible and visual LED alerts if a dangerous condition is thought to exist.

A standard recording feature stores 20 minutes of the most recent historical data on a removable EEPROM memory card for post-accident reconstruction, including steering, braking, and idle time. VORAD Safety Systems also offers complete trip reporting for commercial vehicle operators, to include engine rpm, idle time, number and severity of warnings, over-rpm, over-speed conditions, etc.

Greyhound Bus Lines recently completed installation of the VORAD radar on all of its 2,400 buses (Bulkeley, 1993), and subsequently reported a 25-year low accident record (Greyhound, 1994). The entire system weighs just 6.75 pounds and operates from 12 or 24 volts DC with a nominal power consumption of 20 watts. An RS-232 digital output is available. Selected specifications are listed in Table 6-6 below.

Table 6-6. Selected specifications for the *Collision Warning System*.

Parameter	Value	Units
Effective range	1-350	feet
Accuracy	3	percent
Update rate	30	Hz
Host platform speed	0.5-120	miles per hour
Closing rate	0.25-100	miles per hour
Operating frequency	24.725	GHz
RF power	0.5	milliwatts
Beam width (horizontal)	4	degrees
(vertical)	5	degrees
Size (antenna)	6 x 8 x 1.5	inches
(electronics unit)	8 x 6 x 5	inches
Weight (total)	6.75	pounds
Power	12-24	volts DC
	20	watts
Mean-time-between-failure	17,000	hours

Along with their new joint-venture partner, Eaton Corporation, VORAD is now shipping "ruggedized" commercial systems to the heavy truck industry, with an extended temperature range of -40 to +85°C, 100 volts/meter EMI susceptibility, and full compliance to environmental standard SAE J1455 (VORAD, 1994).

6.2.3 Safety First Systems *Vehicular Obstacle Detection and Warning System*

Safety First Systems, Ltd., Plainview, NY, and General Microwave, Amityville, NY, have teamed to develop and market a 10.525-GHz microwave unit for use as an automotive blind-spot alert for drivers when backing up or changing lanes (Siuru, 1994). The narrow-band (100-KHz) modified-FMCW technique uses patent-pending phase discrimination augmentation for a 20-fold increase in achievable resolution. For example, a conventional FMCW system operating at 10.525 GHz with a 50-MHz bandwidth is limited to a best-case range resolution of approximately 10 feet, while the improved approach can resolve distance to within 0.6 feet out to 40 feet (SFS, undated). Even greater accuracy and maximum ranges (i.e., 160 feet) are possible with additional signal processing.

A prototype of the system delivered to Chrysler Corporation uses conformal bistatic microstrip antennae mounted on the rear side panels and rear bumper of a minivan and can detect both stationary and moving objects within the coverage patterns shown in Figure 6-23. Coarse range information about reflecting targets is represented in four discrete range bins with individual TTL output lines: 0 to 6 feet, 6 to 11 feet, 11 to 20 feet, and greater than 20 feet. Average radiated power is about 50 microwatts with a three-percent duty cycle, effectively eliminating adjacent-system interference. The system requires 1.5 amps from a single 9- to 18-volt DC supply.

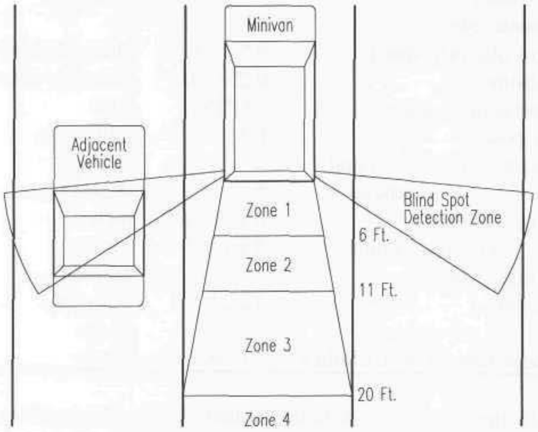

Figure 6-23. The *Vehicular Obstacle Detection and Warning System* employs a modified FMCW ranging technique for blind-spot detection when backing up or changing lanes (courtesy Safety First Systems, Ltd.).

6.2.4 Millitech Millimeter Wave Radar

Millitech Corporation, Deerfield, MA, has designed a number of millimeter-wave FMCW systems aimed at satisfying the short-distance non-contact ranging needs for robotic collision avoidance. These sensors operate at wavelengths of 3.2 millimeters (94 GHz) to 8.6 millimeters (35 GHz), and are superior to infrared devices under all weather conditions, since performance is not significantly degraded by environmental conditions such as fog, rain, dust, and blowing sand (see Chapter 9). Figure 6-24 shows a scanned imaging and data acquisition system in which four vertically stacked beams are mechanically scanned in azimuth to produce a 256-pixel frame of range data at a 5-Hz rate. Each individual pixel contains 512 range bins spaced 0.5 meters apart.

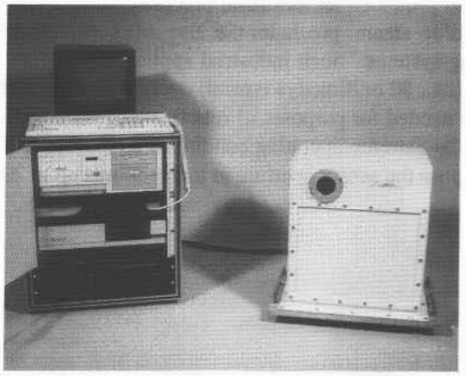

Figure 6-24. Four vertically stacked millimeter-wave beams are mechanically scanned in azimuth to produce a 4- by 64-pixel image over a 12- by 64-degree field of view (courtesy Millitech Corp.).

An innovative feature of Millitech's design is the use of closed-loop control of the oscillator to generate the basic transmitter waveform, yielding stable low-cost performance that will not degrade over time. Operation in the millimeter-wave region of the RF spectrum allows higher-resolution performance in smaller-package configurations than can be achieved with lower-frequency microwave counterparts. A two-degree field of view, for example, requires an aperture size of only 110 millimeters at 94 GHz, as compared to 300 millimeters at 35 GHz. Representative range data collected by the scanning sensor shown above is depicted in Figure 6-25 for two sets of target objects.

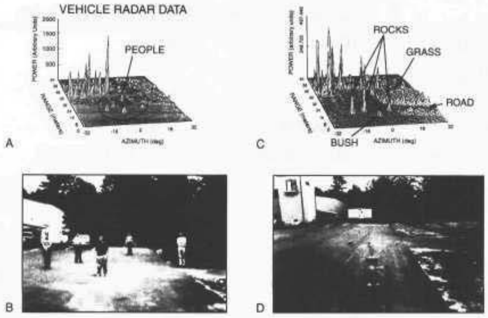

Figure 6-25. Range data acquired by the 256-pixel scanned sensor is shown for (A & B) human targets, and, (C & D) inanimate objects such as rocks and grass (courtesy Millitech Corp.).

Fixed-orientation single-beam versions of Millitech's FMCW radar (Figure 6-26) have also been produced for industrial process-control scenarios where severe dust, smoke and/or steam preclude the use of conventional laser-based or acoustical ranging systems. Such industrial applications generally require higher range resolution (i.e., 50 millimeters typical) over shorter operating distances (i.e., ≤30 meters) than needed for purposes of robotic collision avoidance (see Table 6-7). In addition, the extreme operating conditions associated with some industrial processes can expose the sensor aperture to temperatures up to 200°C.

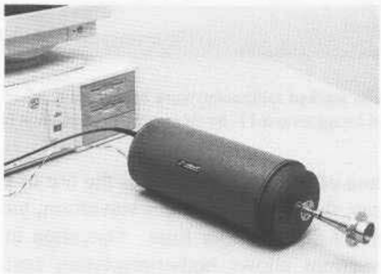

Figure 6-26. Typical industrial millimeter-wave FMCW range sensors are often exposed to ambient temperatures in excess of 200°C (courtesy Millitech Corp.).

Table 6-7. Selected specifications for Millitech millimeter-wave radar prototypes.

Parameter	256-Pixel Scanned Sensor	Fixed-Beam Industrial Sensor	Units
Maximum range:	100	30	meters
Minimum range	0.5	0.2	meters
Output power	10	5-10	milliwatts
Field of View	12 by 64	1 to 2	degrees
Radar cross-section	-40 (minimum)	-30 (minimum)	dBsm
Resolution: range	50	5	centimeters
azimuth	1	1-2	degrees
elevation	3	1-2	degrees
Center frequency	77	94	GHz
Sweep bandwidth	300	400	MHz
Frame rate	5	--	Hz
Data output	Digital	External A/D	
Power	24	±18 to ±28	volts DC
	3	0.5	amps

6.3 References

Adams, M.D., "Amplitude Modulated Optical Range Data Analysis in Mobile Robotics," IEEE International Conference on Robotics and Automation, Atlanta, GA, pp. 8-13, 1993.

Beyer, J., Jacobus, C., Pont, F., "Autonomous Vehicle Guidance Using Laser Range Imagery," SPIE Vol. 852, Mobile Robots II, p. 34-43, Nov., 1987.

Binger, N, Harris, S.J., "Applications of Laser Radar Technology," *Sensors*, pp. 42-44, April, 1987.

Boltinghouse, S., Larsen, T., "Navigation of Mobile Robotic Systems Employing a 3D Laser Imaging Radar," ANS Third Topical Meeting on Robotics and Remote Systems, Section 2-5, pp. 1-7, Charleston, SC, March, 1989.

Boltinghouse, S., Burke, J., Ho, D., "Implementation of a 3D Laser Imager Based Robot Navigation System with Location Identification," SPIE Vol. 1388, Mobile Robots V, Boston, MA, pp. 14-29, November, 1990.

Bulkeley, D., "The Quest for Collision-Free Travel," *Design News*, Oct. 4, 1993.

Byrd, J.S., DeVries, K.R., "A Six-Legged Telerobot for Nuclear Applications Development, *International Journal of Robotics Research*, Vol. 9, pp. 43-52, April, 1990.

Chen, Y.D., Ni, J., Wu, S.M., "Dynamic Calibration and Compensation of a 3D Lasar Radar Scanning System," IEEE International Conference on Robotics and Automation, Atlanta, GA, Vol. 3, pp. 652-664, May, 1993.

Clark, R.R., "A Laser Distance Measurement Sensor for Industry and Robotics," *Sensors*, pp. 43-50, June, 1994.

Conrad, D.J., Sampson, R.E., "3D Range Imaging Sensors," in *Traditional and Non-Traditional Robotic Sensors*, T.C. Henderson, ed., NATO ASI Series, Vol. F63, Springer-Verlag, pp. 35-47, 1990.

Depkovich, T., Wolfe, W., "Definition of Requirements and Components for a Robotic Locating System," Final Report No. MCR-83-669, Martin Marietta Aerospace, Denver, CO, February, 1984.

ESP, "ORS-1 Optical Ranging System," Product Literature, ESP Technologies, Inc., Lawrenceville, NJ 08648, 23 March, 1992.

Figueroa, F., Barbieri, E., "Increased Measurement Range Via Frequency Division in Ultrasonic Phase Detection Methods," *Acustica*, Vol. 73, pp. 47-49, 1991a.

Figueroa, J.F., Barbieri, E., "An Ultrasonic Ranging System for Structural Vibration Measurements," *IEEE Transactions on Instrumentation and Measurement*, Vol. 40, No. 4, pp. 764-769, August, 1991b.

Figueroa, J.F., Lamancusa, J.S., "A Method for Accurate Detection of Time of Arrival: Analysis and Design of an Ultrasonic Ranging System," *Journal of the Acoustical Society of America*, Vol. 91, No. 1, pp. 486-494, January, 1992.

Frazier, K., "Innovative Range Imager Sees How Targets Measure Up," *Sandia LabNews*, Vol. 46, No. 19, September 16, 1994.

Greyhound, "Accident Rate Keeps Falling," *Greyhound Connections*, Vol. 4, No. 2, March/April, 1994.

Hebert, M., Krotkov, E., "3-D Measurements from Imaging Laser Radars: How Good Are They?" International Conference on Intelligent Robots and Systems, pp. 359-364, 1991.

Kerr, J.R., "Real Time Imaging Rangefinder for Autonomous Land Vehicles," SPIE Vol. 1007, Mobile Robots III, , pp. 349-356, November, 1988.

Larson, T.R., Boltinghouse, S., "Robotic Navigation Within Complex Structures," SPIE Vol. 1007, Mobile Robots III, Cambridge, MA, pp. 339-348, November, 1988.

Miller, G.L., Wagner, E.R., "An Optical Rangefinder for Autonomous Robot Cart Navigation," Proceedings of the Advances in Intelligent Robotic Systems: SPIE Mobile Robots II, 1987.

Nitzan, D. et al. "The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis," Proceedings of IEEE, vol. 65, no. 2, pp. 206-220, February, 1977.

Patterson, M.R., Reidy, J.J., Rudolph, R.C., "Guidance and Actuation Systems for an Adaptive-Suspension Vehicle," Final Technical Report, Battelle Columbus Division, OH, AD #A139111, 20 March, 1994.

Scott, M.W., "Range Imaging Laser Radar," US Patent 4,935,616, June 19, 1990.

SFS, "Vehicular Obstacle Detection and Warning System," Product Literature, Safety First Systems, Ltd., Plainview, NY, undated.

Siuru, B., "The Smart Vehicles Are Here," *Popular Electronics*, Vol. 11, No. 1, pp. 41-45, January, 1994.

TRC, "TRC Light Ranger," Product Literature 940823, Transitions Research Corp., Danbury, CN, August, 1994.

VORAD, "The VORAD Vehicle Detection and Driver Alert System," Product Literature, VORAD Safety Systems, Inc., San Diego, CA, undated.

VORAD, "EATON VORAD Collision Warning System," Product Literature VOSL-0100, VORAD Safety Systems, Inc., San Diego, CA, 1994.

VRSS, "Rashid Radar Safety Brake," Product Literature, Vehicle Radar Safety Systems, Inc., Mount Clemens, MI, 1983.

Vuylsteke, P., Price, C.B., Oosterlinck, A., "Image Sensors for Real-Time 3D Acquisition, Part 1," ," in *Traditional and Non-Traditional Robotic Sensors*, T.C. Henderson, ed., NATO ASI Series, Vol. F63, Springer-Verlag, pp. 187-210, 1990.

Weiss, S.A., Lasar-Radar Imaging Without Scanners," *Photonics Spectra*, pp. 28-29, April, 1994.

Woll, J.D., "A Review of the Eaton VORAD Vehicle Collision Warning System," Reprinted from International Truck and Bus Meeting and Exposition, Detroit, MI, SAE Technical Paper Series 933063, ISSN 0148-7191, pp. 1-4, November, 1993.

Woodbury, N., Brubacher, M., Woodbury, J.R., "Noninvasive Tank Gauging with Frequency-Modulated Laser Ranging," *Sensors*, pp. 27-31, Sep., 1993.

7

Other Ranging Techniques

This chapter discusses three final non-contact ranging techniques of potential interest: 1) *interferometry*, 2) *range from focus*, and 3) *return signal intensity*.

7.1 Interferometry

One of the most accurate and precise distance ranging methods known, *interferometry* has existed for many years in laboratory scenarios that afforded the necessary controlled or otherwise structured environment (Brown, 1985). Under such non-turbulent atmospheric conditions, laser interferometers can achieve fractional wavelength accuracy. Recent developments in optical technologies are making possible applications of this technique outside of laboratory scenarios.

The concept is based on the resulting interference patterns that occur when two energy waves caused to travel different paths are compared. If the length of one of the paths is changed, the two beams will interact in such a way that clearly visible constructive and destructive interference fringes are produced. (Fringes are patterns or disturbances in the combined waveform that alternate between maximum and minimum intensity.) Figure 7-1 shows a typical system consisting of a laser emitter, a series of beam splitters and directional mirrors, and a fringe counter. The output of a single coherent light source is split into a reference beam and an output beam. The reference beam is immediately directed into the fringe counter for future recombination with the reflected beam.

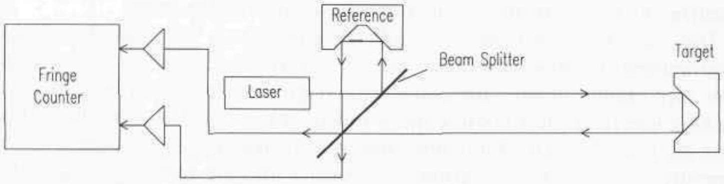

Figure 7-1. Block diagram of typical interferometer ranging system.

The second beam exits the instrument and travels through the air to a retroreflector located on the object of interest. (Retroreflectors must be employed to provide a reliable return signal for the interferometer.) The reflected energy is optically recombined with the reference beam in the fringe counter. Movement of the target through a distance equal to half the source wavelength results in the detection of one fringe (Beesley, 1971). By counting the number of fringes passing the detector it is possible to calculate with extreme accuracy the distance the retroreflector (and thus the object) has traveled along the line of the source beam.

Interferometers do not measure absolute range, but the relative distance an object has moved from its previous location; therefore, the distance from the sensor to the target is not directly known. By initializing the retroreflector to a specified reference point, however, it becomes possible to determine absolute distance to an object. All subsequent measurements will be relative to this reference point, provided the beam is never broken and the target momentarily lost.

In conventional laser interferometry, target displacement of 1 centimeter can result in the movement of approximately 10 million fringes past a detector capable of measuring changes on the order of one tenth of a fringe (Beesley, 1971). Potential accuracies over a distance of 10 meters can approach one part in 1,000,000, provided similar accuracy is available for the wavelength of the energy source. The maximum distance that can be measured by such instruments is therefore dependent on the coherent qualities of the source used. In theory, distances of hundreds of kilometers can be measured; however, this goal cannot be practically achieved using current technology (Beesley, 1971).

While extremely precise, limiting factors of interferometry include the relatively high cost, the need for a continuous line of sight between the source and retroreflector, and the limitation to relative only (as opposed to absolute) distance measurement. Air turbulence effectively reduces the practical range of such systems to around 10 meters (Beesley, 1971). The turbulence causes sufficiently large variations in the path lengths of the light beams so that no spatial coherence exists between the interfering beams; therefore, there are no fringes produced. Temperature changes and microphonic disturbances can cause fluctuations in components of the light source delivery system that alter the wavelength and intensity of the output (Beesley, 1971). The laser output must therefore be well stabilized to realize the full potential of interferometric measuring.

The use of interferometers in robotic applications was initially limited to measurement of single-axis linear motion. Recent developments have expanded their applicability to three-dimensional six-degree-of-freedom systems, known as tracking interferometers because the returning beam is also used by the system to track the lateral motion of retroreflective mirrors mounted on the object. Systems currently in existence are capable of precision tracking of robotic manipulators performing non-rectilinear motions in six degrees of freedom (Everett, 1985; Brown, 1985; Lau, et al., 1985).

7.1.1 CLS Coordinate Measuring System

Chesapeake Laser Systems began development in 1983 of the *CMS-1000*, a laser-based tracking interferometer system that can measure the location of a moving object to better than 10 microns over a volume of 3 by 3 by 3 meters (Brown, 1985; Cleveland, 1986; Brown, et al., 1987). The system employs a servo-controlled beam-steering mechanism to track a randomly moving retroreflective target with a 50-Hz update rate. After a brief calibration routine, three tracking interferometers are used to continuously measure the distance to a number of retroreflectors as shown in Figure 7-2 to calculate the X-Y-Z coordinates of the robotic end effector through trilateration (Everett, 1988).

Figure 7-2. The *CMS-1000* uses multiple HeNe laser beams to track retroreflective targets attached to the moving end effector of the ***Intelligent Robotic Inspection System*** built for the Navy by MTS Systems Corp. (courtesy Applied Research Lab, Penn State University).

The trilateration solution (i.e., superposition of three range arcs) is inherently more accurate than conventional triangulation using angle measurement alone, in which case position errors appear as:

$$\text{error} = r\,d\theta$$

where:

r = radial distance from tracker to retroreflector
$d\theta$ = angular error.

The position error for trilateration, on the other hand, shows up as:

$$\text{error} = r(1 - \cos d\theta) = r\,d\theta^2$$

which is orders of magnitude smaller than $r\,d\theta$.

An improved system, the *CMS-2000*, combines laser interferometry with servo-controlled trackers to measure movement with submicron resolution at ranges up to 35 feet. The *CMS-2000* was initially designed for use by the US Air Force as part of the Strategic Defense Initiative to track (with a 100-Hz update rate) a retroreflector mounted on a hovering rocket. The data obtained by the *CMS-2000* is then used to check the vehicle's onboard control systems (CLS, 1991). This application was unsuccessful due to the excessive heat and dust present in the hanger during launch.

7.2 Range From Focus

The driving thrust behind most of the early work in *range from focus* has been the automatic-focus interest within the commercial video and still camera industry (Marnheim, 1980; Goldberg, 1982; Ctein, 1982; Stauffer & Wilwerding, 1982; Wolpert, 1987). Some more recent efforts have concentrated on the development of sensor subsystems intended specifically for robotic applications (Krotkov & Martin, 1986; Farsaie, et al., 1987). The principle of operation is based on the *Gauss thin lens law* (Conrad & Sampson, 1990):

$$\frac{1}{r}=\frac{1}{f}-\frac{1}{s}$$

where:

r = distance from lens to object viewed
s = distance from lens to image plane
f = focal length of the lens.

Rearranging in terms of the desired distance to object r yields:

$$r=\frac{s\,f}{s-f}.$$

Krotkov and Martin (1986) summarize the *range-from-focus* technique as follows:

- The system is initially calibrated with a point source to establish the relationship between focus motor shaft position and lens focal length f.
- The sharpness of focus is maximized by varying f until the detector plane coincides with the image plane.
- The resulting value of f is read.

- The measured value of f and the constant s are substituted into the thin lens equation, which is then solved for r.

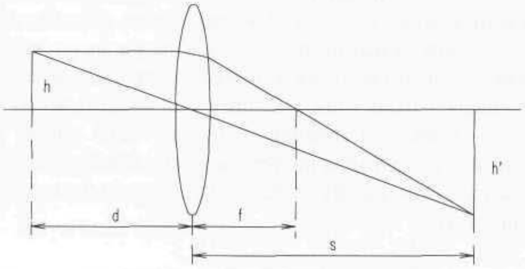

Figure 7-3. The object and image positions (h) and (h') are shown with respect to the focal plane for a thin lens (adapted from Brain, 1979).

The fundamental technical challenge in the above procedure is how to best determine the point of optimal focus; the variety of techniques reported can generally be divided into two broad classes: 1) dedicated in-focal-plane *image analysis* hardware and 2) external signal analysis of conventional composite video output.

A very common approach in the latter case involves examining the high-frequency spatial content of the video signal. The *optical transfer function (OTF)* of a lens describes how each spatial frequency component in a viewed object is attenuated by the lens as it forms an image. Severe aberration (i.e., departure of the wavefront from its ideal spherical form) due to defocusing will significantly reduce the high-frequency portion of the OTF (Krotkov & Martin, 1986). The obvious visual effect is a blurring of the resulting image. More importantly, the loss of high-frequency components in the video signal can be electronically detected to form the basis of a *focus criterion function*.

Deriving distance information from camera focus has an advantage over stereoscopic ranging in the sense that there is no need to solve the computationally intensive and sometimes error-prone correspondence problem. The principle disadvantage is the technique applies only to one observed point in the scene at any given time. Other limitations arise from measurement errors associated with the quantification of both f and s, spatial quantitization of the detector, the performance of the focus criterion function, the validity of the thin lens model describing a compound lens, and the nonlinearity of the lens equation itself (Krotkov & Martin, 1986).

7.2.1 Honeywell Autofocus Systems

The first practical autofocus system for lens-shutter cameras was developed by the Honeywell Visitronic Group in 1976. The system represents a variation of the stereoscopic ranging technique, nicely optimized for low-cost implementation through the development of a special purpose integrated circuit (IC) for autocorrelation. Two five-element photosensitive arrays were located at each end of the IC, measuring about 0.1 by 0.25 inches in size (Stauffer & Wilwerding, 1982). Figure 7-4 shows a pair of mirrors reflecting the incoming light from two viewing windows at either end of the camera housing onto these arrays. One of these images remained fixed while the other was scanned across its respective array through the mechanical rotation of the associated mirror. The angular orientation of the moving mirror at the precise instant that the IC indicated the two images were matched was directly related to the range to the subject and used to position the camera lens.

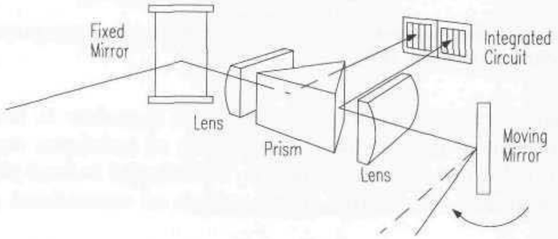

Figure 7-4. The original Visitronic *Autofocus System* employed a pair of five-element detector arrays to establish correlation (courtesy Honeywell Corp.).

The photocurrents from corresponding elements in each array were passed through a string of diodes on the IC and thus converted to voltages proportional to the log of the current. The resulting pair of voltages was then fed to a differential amplifier, which produced a difference signal proportional to the ratio of the two light intensities as seen by the respective detectors (Stauffer & Wilwerding, 1982). For four of the five-element array pairs, the absolute values of these difference signals were summed and the result subtracted from a reference voltage to yield the correlation signal. The better the scene match, the lower the differential signal for each array pair, and the higher the correlation signal.

The peak value of the correlation signal corresponded to the best scene match. An operational amplifier on the IC performed a continuous comparison between the correlation output and the previous highest value stored in a capacitor. The output from this comparator was high as long as the correlation signal was lower than the previous peak value. The last low-to-high transition represents the mirror angle corresponding to the highest peak. A potentiometer on the moving mirror produced a voltage that varied as a linear function of mirror position. The output of this potentiometer was sampled and stored when the IC indicated the peak correlation signal was present. A similar potentiometer coupled to the camera lens positioning mechanism was used to stop the lens travel when its output matched the stored voltage signifying mirror position at best focus.

The Honeywell *Through-the-Camera-Lens (TCL)* autofocus system, a second-generation refinement of the *Visitronic System*, compared the signatures of light passing through two different sectors of the camera lens as opposed to two separate viewing windows. Instead of five, there were 24 pairs of detectors arranged in an array about 5 millimeters long. Two complete arrays were provided to accommodate camera lenses with different aperture sizes (Stauffer & Wilwerding, 1982). Light from any given point in the field of view of a camera passes through all sectors of the camera lens, and subsequently arrives at the image plane from many different angles. If the lens is in focus, these components all converge again to a single point in the image plane. If the lens is not in focus, these components are displaced from one another, and the image becomes fuzzy.

Similarly, light from every point in the scene of interest passes through each sector of the lens. Thus, each sector of the lens will contribute a recognizable signature of light to the image plane, in keeping with the image viewed. (Early pinhole cameras made use of this principle; essentially there was only one sector, and so there was only one image, which was always in focus.) Practically speaking, these signatures are identical, and if the lens is in focus, they will be superimposed. As the lens moves out of focus, the signatures will be displaced laterally, and the image blurs. The Honeywell *TCL* system detects this displacement for two specific sectors (A and B) located at opposite sides of the lens as shown in Figure 7-5.

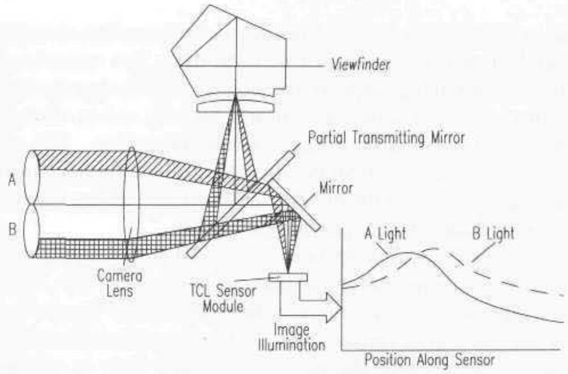

Figure 7-5. Light from two separate sectors of the same lens is compared to determine the position of best focus in the TCL Autofocus System (courtesy Honeywell Corp.).

Light from these two sectors falls upon a series of 24 microlenses mounted on the surface of the integrated circuit in the camera image plane. An array of sensors is positioned within the IC at a specified distance behind the image plane in such a fashion that light incident upon the row of microlenses and their associated image sampling apertures will diverge again to isolate the respective components arriving from each of the two lens sectors (Figure 7-6). Within each aperture image in the detector plane are two detectors, one for each of the two sectors (A and B). Output of all 24 of the A-detectors is used to construct the A-signature; the 24 B-detectors are read to form the B-signature (Stauffer & Wilwerding, 1982).

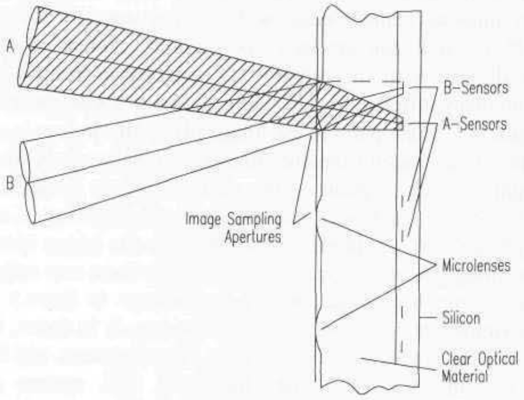

Figure 7-6. A row of microlenses focuses light on the pairs of detectors, forming two separate signatures for comparison (courtesy Honeywell Corp.).

The signatures of light passing through the two camera lens sectors can then be compared and analyzed. The distance between these lens sectors is the base of triangulation for determining range to the subject. Which signature appears to be leading the other and to what degree indicates how far and in what direction the lens must be moved to bring the images into superposition. The output of the CCD detector array is fed to a CMOS integrated circuit which contains the CCD clock circuitry and an A/D converter that digitizes the analog output for further processing by a dedicated-logic algorithm processor.

The Honeywell *TCL* circuitry operates on a 5-volt power supply, and the sensor and companion ICs together draw less than 60 milliwatts. The *TCL* system can sense if the image is in focus to where the plane of the image is within 0.05 millimeter of the position of correct focus, and could potentially provide a passive sensing capability for robotic applications provided there is adequate ambient illumination and scene contrast. The detector pairs in the *TCL* system can discriminate light differences of one part in 100, whereas the human eye is limited to one part in 10 (Stauffer & Wilwerding, 1982).

7.2.2 Associates and Ferren Swept-Focus Ranging

The swept-focus ranging technique uses a conventional video camera with a single lens of very shallow depth of field to produce an image in which only a narrow interval of object space is in focus at any given time. By means of a computer-controlled servo drive (Figure 7-7), the lens can be positioned with great accuracy over a series of discrete stops to view different range "slices." (Some systems operate with a fixed-location lens, and vary the position of the detector element to achieve the same effect.) The distance between the lens and the image plane at the detector is related to the range at which the camera is focused in accordance with the thin lens equation. Thus, if the lens is mechanically positioned to bring the desired object into focus, then the range to that object can be derived from the position of the lens.

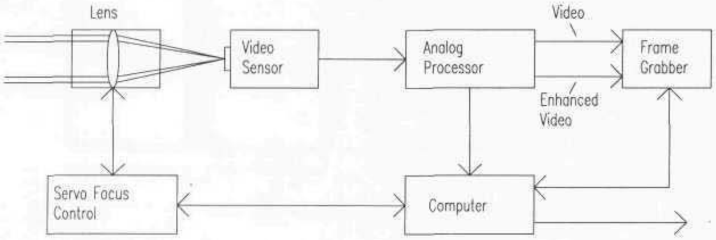

Figure 7-7. Block diagram of a typical swept-focus three-dimensional vision system (courtesy Associates and Ferren, Inc.).

An analog signal processor filters the video signal from the camera to obtain only the high-frequency portion representing information that changes rapidly across the scene, such as in-focus edges or textured material (Figure 7-8). The out-of-focus portions of an image do not contribute to the high-frequency information. This filtered signal is integrated during each video field time.

To perform ranging, the lens is successively positioned at a multitude of discrete precalculated positions, reading and storing the integrated high-frequency data as it becomes available at each position before moving to the next. At the end of this process, the resultant profile of high-frequency response with range is processed to reduce noise effects and then analyzed to determine the locations of all significant peaks. Each peak in high-frequency response represents the best-focus location of a target. The distance to each target can be found simply by reading from a look-up table the object range corresponding to the lens position where the peak occurred.

The swept-focus vision system developed by Associates and Ferren, Wainscott, NY, (Figure 7-9) was specifically intended to address the collision avoidance needs of a mobile robotic platform (Farsaie, et al., 1987; Ferren, 1986). The design therefore employed special optical preprocessing techniques to minimize the onboard computational requirements for image understanding. The system consists of a swept-focus sensor mounted on a robotic vehicle, in communication with a remote host computer and frame grabber. To determine the range to objects in the sensor's field of view, the lens is swept through hundreds of discrete focal positions, remaining at each position for $^1/_{60}{}^{\text{th}}$ of a second, or one video field time. During this time, the analog signal processor integrates the high-frequency response in that field. This summation is a measure of the amount of edge information in the associated range slice and representative of the relative degree of focus.

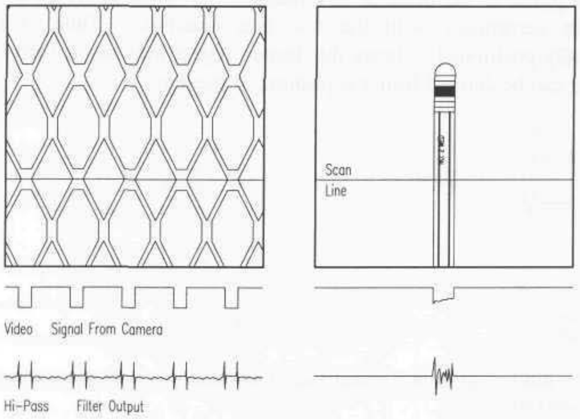

Figure 7-8. Video and high-pass filter output when viewing a piece of expanded metal and a pencil (courtesy Associates and Ferren, Inc.).

Good accuracy (about 1 inch) and repeatability are obtained with a 600-position scan over a 25-foot range interval, which takes approximately 12 seconds (50 millimeter/Fl.0 lens). Accuracy and resolution vary with range and are greatest at closer range, using the current exponential scan profile. Ranging accuracy and the ability to separate targets closely spaced in range are bounded by the physical constraints of the lens. The greater the desired accuracy and resolution, the shorter the required depth of field, which can be achieved by using a lens of longer focal length or larger aperture as illustrated by the following equation (Krotkov & Martin, 1986):

$$DOF = \frac{2\,a\,f\,c\,d\,(d-f)}{a^2 f^2 - c^2 (d-f)^2}$$

where:

DOF = depth of field
a = aperture diameter
f = focal length
c = smallest dimension of detector photoreceptor
d = distance to object.

The tradeoffs involved are reduced field of view and increased size and weight. In practice, the two lenses found to be most useful are a 50 millimeter/F1.0 and a 105 millimeter/F1.8, both good quality photographic lenses. The longer lens offers better ranging accuracy and resolution but has a narrower field of view than the shorter lens.

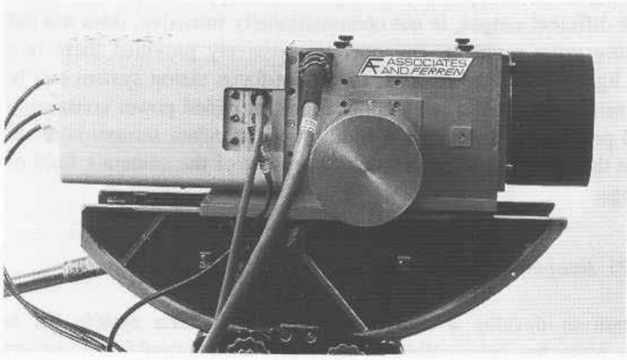

Figure 7-9. Swept-focus camera system developed for passive three-dimensional vision applications (courtesy Associates and Ferren, Inc.).

The swept-focus vision system described has been used as the primary sensor for a mobile robot with good success (Ferren, 1986). The main factor limiting the speed of this technique is the standard 60-Hz video field rate. The system supplies accurate range data and can generate a floor-plan map of its environment that is used in map-based path planning. For such an imaging task, a quick full-range scan could be executed to find the gross location of a target. The lens could then be scanned through the identified range space at smaller increments, saving the entire video field at each position in a large bit-mapped model. A three-dimensional representation of the edge-enhanced object could thus be generated and stored in memory.

During subsequent motion of the robotic vehicle, the onboard video camera can be used as a visual proximity detector by positioning the lens at a fixed focus and monitoring the change in high-frequency content of the scene as the robot travels. A significant rise in this high-frequency information is indicative of a target coming into focus at the range that the lens is imaging. When this condition arises, the robot pauses until it can determine whether or not a collision is imminent. In this application, the 50-millimeter lens has been most useful. The accuracy of the l05-millimeter lens is superior, but its 110-degree field of view is too restrictive.

The use of optical preprocessing in the swept-focus sensor gives it some advantages over other sensing techniques. There is no *missing parts* problem since there is only a single lens, and periodic mechanical alignment is not necessary. The preprocessing action of the short depth-of-field lens also allows for ranging that is not computationally intensive. The system operates passively under normal ambient lighting conditions, responding well to all target objects except those which present a flat field, such as painted walls with no visible texture or markings.

Swept focus has acceptable accuracy for most applications, will locate multiple targets at different ranges, is not computationally intensive, does not suffer from the *missing parts* problem, and operates passively provided there is sufficient ambient light. For these reasons, the swept-focus vision system can be a good primary sensor for mobile robot applications, provided power consumption is not a critical problem. However, the addition of redundant sensors is recommended to ensure the detection of objects which are out of the camera's field of view at close range.

7.2.3 JPL *Range-from-Focus* System

A program to develop a semi-autonomous navigation system for use on a planetary rover has been underway at NASA's Jet Propulsion Laboratory (JPL) since late 1988. The initial testbed vehicle used a passive vision-based navigation technique that required a great deal of computation. To reduce the computational overhead, researchers at JPL are working an alternate approach in the form of a *range-from-focus* optical system (Figure 7-10). The goal is to minimize the necessary computation so that navigation of the rover can be practically performed on board rather than remotely from earth (Wilcox, 1990).

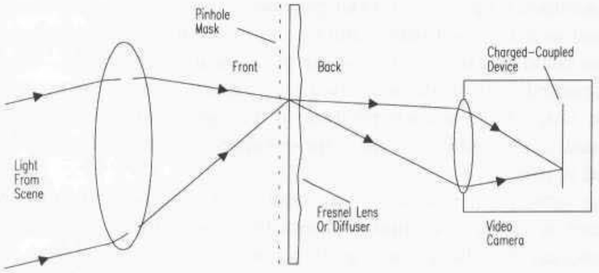

Figure 7-10. Diagram of the JPL *Range-From-Focus* system (courtesy Jet Propulsion Laboratory).

This focus-based ranging system uses a large-aperture short-focal-length lens with a pinhole mask at the prime focus. This mask is transparent only in an array of pinholes at or near the diffraction-limiting spot size of the lens. A diffuser or

Fresnel lens is placed behind the mask to direct the light coming through the holes back toward a CCD camera. The camera is focused on the mask such that there is a one-to-one correspondence between each pixel and pinhole. When successive frames from the CCD camera are differenced and the magnitude of that difference averaged, the only significant signal remaining will be in those parts of the image where the terrain is in focus.

The initial prototype will use a 75-millimeter, F1.9 lens. By way of example, if the lens is focused at a range of 10 meters, the corresponding focal distance is 75.567 millimeters; at 10.5 meters it is 75.540 millimeter. The difference in focal distance is 27 microns, which leads to a 14-micron circle of confusion for a point source at 10.5 meters. The pinhole array is focused for 10 meters, and the diffraction limiting spot is approximately 2 microns. Even a highly textured surface at 10.5 meters will not produce strong difference values between successive frames (assuming the image moves less than 14 microns across the array), whereas an object at 10 meters will produce a 100-percent contrast change with only 2 microns of image motion.

Two or three different range planes could be mixed on different video scan lines in the same sensor. To accomplish this, the pinhole array could be corrugated so that alternate scan lines represent different range distances. A practical implementation would be made from layers of photographic film, with stripes of clear film alternating with the pinhole arrays. For robotic collision avoidance purposes, it is generally not required to have a range map as dense as a standard video image (approximately 500 by 500); several pixels can be averaged horizontally, assuming they will be at approximately the same range. This approach yields two 250- by 250-pixel range maps at two different ranges from the single sensor.

Another variation would incline the image pinhole array to match the flat-earth ground plane. Using the corrugated approach previously described, one could mix the resulting images to produce a single video image depicting elevation deviations. This method would allow obstacle detection from brightness changes in the video image alone; no postprocessing would be required.

7.3 Return Signal Intensity

Ranging techniques involving ***return signal intensity*** determine the distance to an object based on the amplitude of energy (usually light) reflected from the object's surface. The inverse square law for emitted energy states that as the distance from a point source increases, the intensity of the source diminishes as a function of the square of the distance. If Lambertian surfaces are assumed (see Chapters 8 and 9), then this principle results in a computationally simple algorithm for range calculation. Numerous prototypes have been developed using this technique, but few have found their way to commercial products. Ctein (1982), however, reports that Kodak incorporated an active near-infrared rangefinder in the *Kodamatic*

980L instant camera that measures the brightness of the reflected flash to estimate distance for focusing purposes.

7.3.1 Programmable Near-Infrared Proximity Sensor

A custom-designed near-infrared proximity sensor was developed for use on ROBART I (Everett, 1982) to gather high-resolution geometric information in support of navigational routines as will be discussed in Chapter 10. The primary purpose of this head-mounted sensor was to provide precise angular location of prominent vertical edges such as door openings. An improved programmable version was incorporated on ROBART II (Figure 7-11) to complement range information provided by a Polaroid sonar sensor (Flynn, 1985). Adams, et al. (1990) report additional follow-up work of a similar nature at the Oxford University.

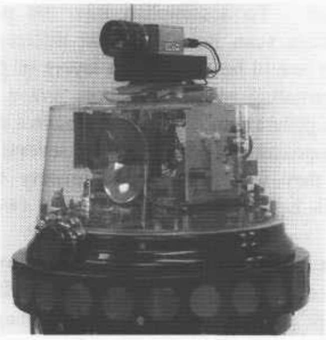

Figure 7-11. The near-infrared proximity sensor developed for ROBART II used a programmable array of four high-powered LED emitters with a variable threshold receiver. The PIN-photodiode detector is situated at the focal point of the parabolic reflector.

An astable multivibrator produces a square wave of short-duration pulses, driving high-power XC-880-A gallium-aluminum arsenide LEDs that emit energy in the near-infrared spectrum. The system uses an array of adjacent LEDs for increased range and sensitivity, with reflected energy focused on the lens of a TIL413 photodiode by a parabolic reflector. The output of this photodiode is passed through an L/C differentiator network, amplified, and fed to four separate follow-on threshold detector stages (Figure 7-12). The receiver sensitivity is broken into four discrete levels by these individually adjustable threshold comparators. A strong return will cause all four channels to go low, whereas a weak return will cause only the most sensitive channel to indicate detection.

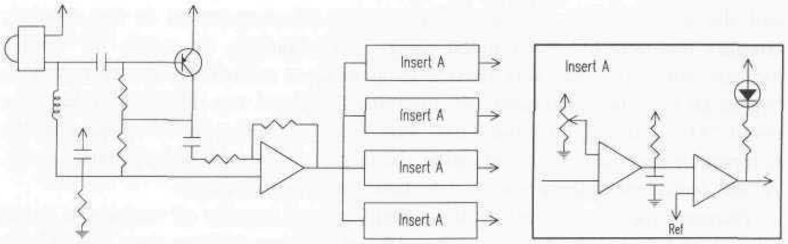

Figure 7-12. Schematic diagram of the receiver portion of the programmable near-infrared proximity sensor.

Effective range is controlled by firing combinations of LEDs, thereby emitting regulated amounts of energy (i.e., the more LEDs illuminating the scene, the farther the detection range). The number of LEDs in the array that are enabled at any given time is specified by a microprocessor, providing programmable control over the amount of transmitted energy, which in turn fixes the maximum range of the sensor (Everett & Flynn, 1986). The total number of active emitters can be any value between one and four (Figure 7-13). The robot "feels around" out to a distance of 5 or 6 feet, and notes any detected obstruction. If no reflected energy is sensed, an additional LED is activated to extend the range of the sensor a few more feet, and the area is probed again. This process is repeated as the head pans to map the entire region in terms of range discontinuities as a function of relative bearing.

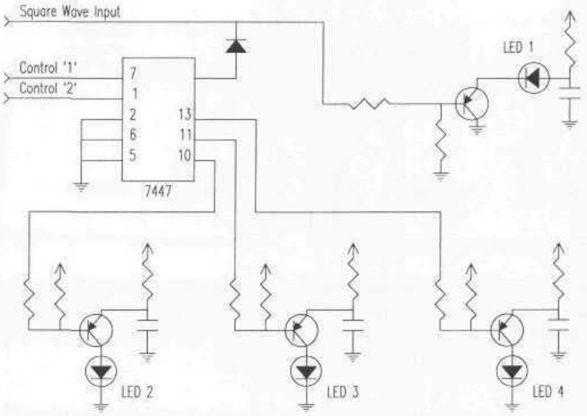

Figure 7-13. Schematic diagram of the transmitter portion of the programmable proximity detector.

The data protocol employed for communicating the information is in the form of a single byte in which the upper nibble represents the number of LEDs fired, and the lower nibble represents the number of comparators in the receiver circuitry that detected the returned energy. For example, the result "14" would indicate only one LED was needed to generate a sufficiently strong return to trigger all four threshold detectors, implying the target was relatively close. The result "41" would signify four LEDs were required, with only the most sensitive comparator responding, an indication the target was considerable distance away. A "40" would mean there was no detectable target within range.

During experimental testing the system proved capable of seeing out to an average of 6 feet with one LED active, 10 feet with two LEDs active, 13 feet with three, and a maximum average range of 15 feet attainable with all four (Everett & Flynn, 1986). Figure 7-14A shows a sonar plot of a small room generated by 256 range readings taken by the head-mounted Polaroid sensor shown in Figure 7-11, with the robot situated as shown. Note the broadened representation of the right side of the upper doorway due to the 30-degree effective beamwidth of the Polaroid sonar. Figure 7-14B shows the same data overlaid with crosses to mark the state transitions for the near-infrared proximity sensor. Examination of the recorded data shows the biggest reflectance discontinuities for the two crosses closest to the edges of door opening.

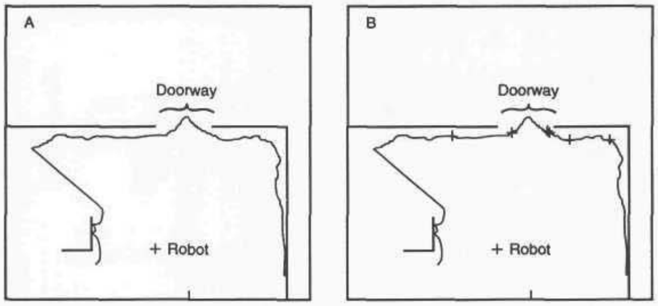

Figure 7-14. A total of 256 sonar range measurements, taken by ROBART II while stationary at the point marked by the cross, are overlaid on actual room measurements in (A). State transitions in the near-infrared sensor data marked with crosses in (B) clearly identify the open doorway boundaries at the top of the figure (courtesy MIT AI Lab).

Unfortunately, not all objects in the real world are ideally Lambertian in nature, and the varying reflectivities of typical surfaces preclude simple measurement of return signal strength from being a reliable indicator of distance under most conditions.

7.3.2 Australian National University Rangefinder

A monocular ranging technique developed at the Australian National University attempted to resolve the reflectivity problem by simultaneously measuring the return signal intensity of a pair of light sources (Jarvis, 1984). The two sources were arranged with a two-dimensional camera detector along a common optical axis that was focused on the target surface as illustrated in Figure 7-15. The displacement between the sources resulted in differing magnitudes of returned energy, each related to distance by the inverse square law. For identical collocated sources, the intensity of the return signal as sensed by the receiver should ideally be the same. However, in this configuration one emitter was closer to the scene than the other, resulting in a difference in the return signal intensity produced by the two sources. This measurable difference was exploited to yield absolute range values, and the effects of surface reflectivity (which similarly attenuates returned energy for both sources) subsequently cancel out.

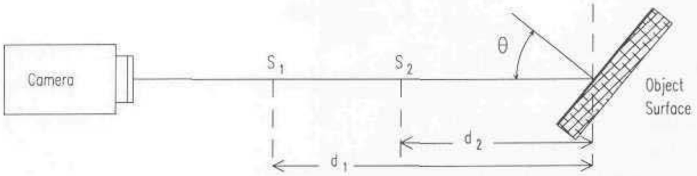

Figure 7-15. By using a pair of emitters at S_1 and S_2 with a common detector, two equations are generated with two unknowns to eliminate the influence of surface reflectivity (adapted from Jarvis, 1984).

The experimental system developed for evaluation of this technique used 35-millimeter slide projectors as the point sources. Sensitivity improved as the distance between projectors was increased. Color images of the scene (256 successive frames) were captured for three separate lighting conditions: 1) background lighting only, 2) background lighting with the far source energized, and 3) background lighting with the near source energized (Jarvis, 1984). The background illumination components were subtracted pixel by pixel from each of the actively illuminated scenes to isolate the effects of the point-source illuminators. Image data from the red, blue, and green channels were then separately processed on a VAX 11/780 to compute the range estimates for each pixel in a 128 by 128 spatial array. Of the approximately 60 seconds required for one complete ranging operation, 75 percent involved image acquisition (Jarvis, 1984). The prototype was capable of measuring range over uniform textured or colored surfaces, but encountered difficulty when observing multicolor non-planar targets.

7.3.3 MIT Near-Infrared Ranging System

A one-dimensional implementation of this ranging technique was developed by Connell at the Massachusetts Institute of Technology (MIT) Artificial Intelligence Lab. The MIT system used a pair of identical point-source LEDs positioned a known distance apart, with their incident light focused on the target surface. The emitters were individually fired in a sequential manner, with the reflected energy in each case detected by a phototransistor and digitized with an analog-to-digital converter.

By the inverse square law, the observed intensity is inversely proportional to the square of the round trip distance traveled. Furthermore, the difference in the resulting intensities caused by the offset in the distance between the LED emitters can be used to solve for the range value:

$$r = \frac{d}{\sqrt{\frac{B_1}{B_2}} - 1}$$

where:

r = the range to the target
d = the distance between emitters
B_1 = intensity of return for LED 1
B_2 = intensity of return for LED 2.

The basic assumptions made in the design are that all surfaces are Lambertian in nature and that the observed objects are wider than the field of view of the LEDs. Ambient light interference is reduced by blinking the LEDs and synchronizing the detector to look for this *on-and-off* sequence of energy returning from the observed scene.

7.3.4 Honeywell Displaced-Sensor Ranging Unit

Honeywell Visitronics has developed a prototype return-signal-intensity ranging system using a single near-infrared LED source and two displaced silicon detectors. A momentary pulse of near-infrared radiation is projected onto the target surface, while the reflected flux is simultaneously detected with two sensors that are displaced along the measurement axis (Figure 7-16). The signal from each sensor may be represented by the following:

$$S_1 \propto \frac{FR}{D^2} \quad \text{and} \quad S_2 \propto \frac{FR}{(D+d)^2}$$

where:

S_1 and S_2 are the detected signals
F = projected spot flux
R = surface reflectivity
D = distance to target
d = displacement seen by S_2.

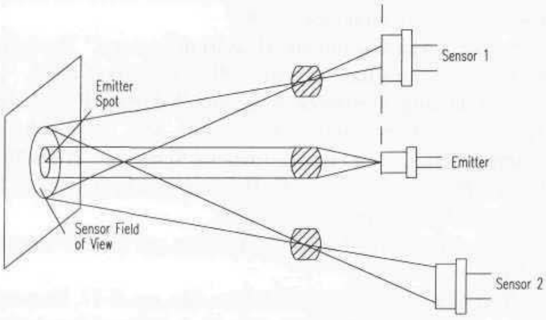

Figure 7-16. Two displaced sensors are used with a common LED source in the Honeywell return-signal-intensity prototype (courtesy Honeywell Corp.).

The detected signal intensities thus provide a means to determine range independent of the surface reflectivity. The use of twin displaced detectors as opposed to displaced emitters eliminates the need to alternately fire the LEDs and offers the advantage of matched stable response and excellent linearity. (LED emitters are temperature sensitive and their performance changes with age, thus making it difficult to maintain identical output.)

The Honeywell prototype provides a pulse-repetition frequency that is proportional to range with a 0- to 5-volt DC linear output. The system can sense objects with a surface reflectivity of 10 to 90 percent out to a distance of 5 meters, with a resolution of 6 millimeters at a distance of 1 meter. System response is less than 5 milliseconds. The prototype is packaged in an enclosure 51 by 51 by 150 millimeters with a weight of 0.65 kilograms.

7.4 References

Adams, M.D., Hu, H., Probert, P.J., "Towards a Real-Time Architecture for Obstacle Avoidance and Path Planning in Mobile Robots," IEEE International Conference on Robotics and Automation, Cincinnati, OH, pp. 584-589, May, 1990.

Beesley, M.J., *Lasers and Their Applications*, pp. 137-147, Taylor and Francis LTD, London, 1971.

Brain, A.E., "Lenses for Industrial Automation, Part One: A Brief Review of Basic Optics," SRI Technical Note 201, SRI International, Menlo Park, CA, November, 1979.

Brown, L.B., "A Random-Path Laser Interferometer System," Proceedings of the Laser Institute of America's International Congress of Applications of Lasers and Electro-Optics, San Francisco, CA, 11-14 November, 1985.

Brown, L.B., Merry, J.B., Wells, D.N., "Tracking Laser Interferometer," U.S. Patent No. 4,790,651, 30 September, 1987.

CLS, "The CMS-2000 Laser Coordinate Measuring System," Product Literature, Chesapeake Laser Systems Inc., Lanham, MD, December, 1991.

Cleveland, B.A. "An Intelligent Robotic Inspection System (IRIS)," Technical Report, MTS Systems Corporation, Minneapolis, MN, 31 March, 1986.

Conrad, D.J., Sampson, R.E., "3D Range Imaging Sensors," in *Traditional and Non-Traditional Robotic Sensors*, T.C. Henderson, ed., NATO ASI Series, Vol. F63, Springer-Verlag, pp. 35-47, 1990.

Ctein, "Autofocus Looks Sharp," *High Technology*, pp. 53-56, November-December, 1982.

Everett, H.R., "Robotics in the Navy," *Robotics Age*, pp. 6-11, November, 1985.

Everett, H.R., Flynn, A.M., "A Programmable Near-Infrared Proximity Detector for Mobile Robot Navigation", Proceedings SPIE Mobile Robots I, Cambridge, MA, pp. 221-230, October, 1986.

Everett, H.R., "Survey of Collision Avoidance and Ranging Sensors for Mobile Robots," NOSC Technical Report 1194, Naval Ocean Systems Center, San Diego, CA, March, 1988.

Farsaie, A., McKnight, T.R., Ferren, B., Harrison, C.F., "Intelligent Controllers for an Autonomous System," IEEE International Symposium on Intelligent Control, 1987.

Ferren, "3-D Computer Vision System for Robots," Final Report for Phase II, NSWC Contract No. N60921-85-D-0064, Associates and Ferren, Wainscott, New York, June, 1986.

Flynn, A.M., "Redundant Sensors for Mobile Robot Navigation," Technical Report 859, MIT Artificial Intelligence Laboratory, Cambridge, MA, October, 1985.

Goldberg, N., "Inside Autofocus: How the Majic Works," Popular Photography, pp. 77-83, February, 1982.

Jarvis, R.A., "Range from Brightness for Robotic Vision," Proceedings of 4th International Conference on Robot Vision and Sensory Controls, London, U.K., pp. 165-172, 9-11 October, 1984.

Krotkov, E., Martin, J., "Range from Focus," IEEE International Conference on Robotics and Automation, San Francisco, CA, pp.1093-1098, April, 1986.

Lau, K., et al., "Robot End Point Sensing Using Laser Tracking System," Proceedings of the NBS Sponsored Navy NAV/CIM Robot Standards Workshop, Detroit, MI, pp. 104-111, June, 1985.

Marnheim, L.A., "Autofocus: What's it all about?" *Modern Photography*, pp. 102-178, June, 1980.

Stauffer, N., Wilwerding, D., "Electronic Focus for Cameras," *Scientific Honeyweller*, Vol. 3, No. 1, March, 1982.

Wilcox, B.H., "Vision-based Planetary Rover Navigation," SPIE International Conference on Visual Communications and Image Processing, Lausanne, Switzerland, October, 1990.

Wolpert, H.D., "Autoranging/Autofocus: A Survey of Systems," *Photonics Spectra*, pp. 165-168, June, 1987.

8

Acoustical Energy

All sensors, whether active or passive, perform their function by detecting (and in most cases quantifying) the change in some specific property (or properties) of energy. Active sensors emit energy that travels away from the sensor and interacts with the object of interest, after which part of the energy is returned to the sensor. For passive sensors, the source of the monitored energy is the object itself and/or the surrounding environment. In the case of acoustical systems, it must be recognized that the *medium of propagation* can sometimes have significant influence, and such effects must be taken into account.

Sound is a vibratory mechanical perturbation that travels through an elastic medium as a longitudinal wave. For gases and liquids the velocity of wave propagation is given by (Pallas-Areny & Webster, 1992):

$$s = \sqrt{\frac{K_m}{\rho}}$$

where:

s = speed of propagation
K_m = bulk modulus of elasticity
ρ = density of medium.

Since the introduction of sonar in 1918, acoustic waves have been successfully used to determine the position, velocity, and orientation of underwater objects in both commercial and military applications (Ulrich, 1983). It therefore seems only logical we should be able to take advantage of this well-developed sonar technology for deployment on mobile robotic vehicles. This seemingly natural carry-over from underwater scenarios has been somewhat lacking, however, for a number of reasons. The speed of sound in air (assume sea level and 25°C) is 1138 feet/second, while under the same conditions in sea-water sound travels 5,034 feet/second (Bolz & Tuve, 1979). The wavelength of acoustical energy is directly proportional to the speed of propagation as shown below:

$$\lambda = \frac{s}{f}$$

where:

λ = wavelength
s = speed of sound
f = operating frequency.

This relationship means the wavelength for an underwater sonar operating at 200 KHz would be approximately 0.30 inches, while that associated with operation in air at the same frequency is in contrast only 0.07 inches. As we shall see later, the shorter the wavelength, the higher the achievable resolution. So in theory, better resolution should be obtainable with sonar in air than that associated with operation in water. In practice, however, the performance of sonar operating in air seems poor indeed in comparison to the success of underwater implementations, for several reasons.

For starters, water (being basically incompressible) is a much better conductive medium (for sound) than air. In fact, sound waves originating from sources thousands of miles away are routinely detected in oceanography and military applications. One such example involves monitoring global warming (as manifested in long-term variations in average sea-water temperature) by measuring the associated change in the speed of wave propagation over a transoceanic path. Secondly, the mismatch in acoustical impedance between the transducer and the conducting medium is much larger for air than water, resulting in reduced coupling efficiency. The high acoustic impedance of water allows for conversion efficiencies from 50 to 80 percent, depending on the desired bandwidth (Bartram, et al., 1989). In addition, underwater systems are generally looking for fairly large discrete targets in relatively non-interfering surroundings, with the added benefit of intensely powerful pulse emissions. And finally, one should keep in mind that untold millions of defense dollars have been invested over many decades in the research and development of sophisticated underwater systems that individually cost millions of dollars to procure, operate, and maintain. In contrast, most robotic designers begin to balk when the price of any sensor subsystem begins to exceed a few thousand dollars.

The range of frequencies generally associated with human hearing runs from about 20 Hz to somewhere around 20 KHz. Although sonar systems have been developed that operate (in air) within this audible range, ultrasonic frequencies (typically between 20 KHz and 200 KHz) are by far the most widely applied. It is interesting to note, however, ultrasonic frequencies as high as 600 MHz can be produced using piezoelectric quartz crystals, with an associated wavelength in air of 500 nanometers (Halliday & Resnick, 1974). (This wavelength is comparable to electromagnetic propagation in the visible light region of the energy spectrum.) Certain piezoelectric films can be made to vibrate in the gigahertz range (Campbell, 1986).

Acoustical ranging can be implemented using triangulation, time of flight (Frederiksen & Howard, 1974; Biber, et al., 1980), frequency modulation (Mitome, et al., 1984), phase-shift measurement (Fox, et al., 1983; Figueroa & Barbieri, 1991), or some combination of these techniques (Figueroa & Lamancusa, 1992). Triangulation and time-of-flight methods typically transmit discrete short-duration pulses and are effective for in-close collision avoidance needs (Chapter 10), and at longer distances for navigational referencing (Chapter 15). Frequency-modulation and phase-shift ranging techniques involving the transmission of a continuous sound wave are better suited for short-range situations where a single dominant target is present.

In addition to distance, the radial direction and velocity of a moving object can also be determined with continuous-wave systems by measuring the *Doppler shift* in frequency of the returned energy. Anyone who has ever noticed the change in siren pitch that occurs when an approaching fire truck or ambulance passes by a stationary observation point is familiar with this effect. For such a fixed observer listening to a moving source, the arriving *Doppler frequency* is expressed as (Halliday & Resnick, 1974):

$$f' = f\left(\frac{s}{s \pm v_s}\right)$$

where:

f' = Doppler frequency at observation point
f = frequency of source
s = speed of sound in air
v_s = radial velocity of source.

As the source closes on the observer, the sign of v_s is negative, resulting in a slightly higher apparent frequency. The sign of v_s becomes positive when the source is moving away from the observer, and the Doppler frequency is decreased.

Alternatively, for a moving observer listening to a fixed-location source, the observed frequency is expressed as (Halliday & Resnick, 1974):

$$f' = f\left(\frac{s \pm v_o}{s}\right)$$

where:

v_o = radial velocity of observer.

In this situation, the sign of v_o is negative if the observer is moving away from the source, resulting in a slightly lower frequency. If the observer is closing on the source, the sign of v_o is positive and the frequency is increased. Note the change in frequency for a moving source approaching a stationary observer is different from that of a moving observer approaching a fixed-location source at the same relative velocity.

In the case of a reflected wave, there is a factor of two introduced, since any change x in relative separation affects the round-trip path length by $2x$. Furthermore, in such situations it is generally more convenient to consider the *change* in frequency Δf, known as the *Doppler shift*, as opposed to the *Doppler frequency* f' (Pallas-Areny & Webster, 1992):

$$\Delta f = f_e - f_r = \frac{2 f_e v \cos\theta}{s}$$

where:

Δf = Doppler frequency shift
f_e = emitted frequency
f_r = received frequency
v = velocity of the target object
θ = relative angle between direction of motion and beam axis.

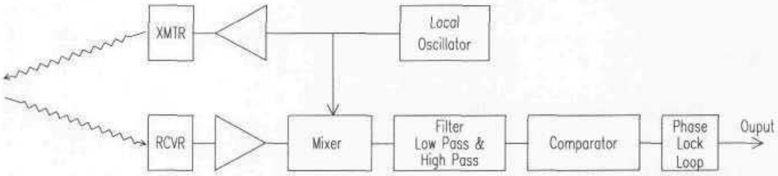

Figure 8-1. Block diagram of an ultrasonic Doppler ground speed sensor used by skiers and joggers (Milner, 1990).

8.1 Applications

Ultrasonic-based measurement systems have found broad appeal throughout the industrial community for a wide variety of purposes such as non-destructive testing (Campbell, 1986), industrial process control (Asher, 1983), stock measurement (Shirley, 1991), liquid level measurement (Shirley, 1989), safety interlocks around dangerous machinery (Irwin & Caughman, 1985), and even intrusion detection in security scenarios (Smurlo & Everett, 1993). In the recreational electronics industry major usage is seen in underwater sonar for depth and fish finding (Frederikson & Howard, 1974), and automatic camera focusing (Biber, et al., 1980). Typical robotic applications include collision avoidance (Everett, 1985), position location (Dunkin, 1985; Figueroa & Mahajan, 1994), and Doppler velocity measurements (Milner, 1990).

8.2 Performance Factors

There are three basic types of ultrasonic transducers: 1) *magnetostrictive*, 2) *piezoelectric*, and 3) *electrostatic*. The first of these categories, *magnetostrictive*, is primarily used in high-power sonar and ultrasonic cleaning applications (Campbell, 1986) and of limited utility from a mobile robotics perspective. *Piezoelectric* and *electrostatic* transducers were treated briefly in Chapter 5, but will be re-examined here from the standpoint of some unique features affecting performance.

Piezoelectric crystals change dimension under the influence of an external electrical potential and will begin to vibrate if the applied potential is made to oscillate at the crystal's resonant frequency. While the force generated can be significant, the displacement of the oscillations is typically very small, and so piezoelectric transducers tend to couple well to solids and liquids but rather poorly to low-density compressible media such as air (Campbell, 1986). Fox, et al. (1983) report using a quarter-wavelength silicon-rubber matching layer on the front face of the transducer in an attempt to achieve better coupling into air at operating frequencies of 1 to 2 MHz. There is also a mechanical inertia associated with the vibrating piezoelectric crystal. As a consequence, such transducers will display some latency (typically several cycles) in reaching full power, and tend to "ring down" longer as well when the excitation voltage is removed.

Electrostatic transducers, on the other hand, generate small forces but have a fairly large displacement amplitude, and therefore couple more efficiently to a compressible medium such as air than do piezoelectric devices. The low-inertia foil membrane allows for quicker turn-on and turn-off in comparison to the slow response of piezoelectrics, facilitating unambiguous short-duration pulses for improved timing accuracy (Campbell, 1986). Since effective operation is not limited to a unique resonance frequency, electrostatic transducers are much more broadband, but with an upper limit of several hundred kilohertz in contrast to megahertz for the piezoelectric variety.

In addition to transducer design considerations, the performance of ultrasonic ranging systems is significantly affected by target characteristics (i.e., absorption, reflectivity, directivity) and environmental phenomena, as will be discussed below.

8.2.1 Atmospheric Attenuation

As an acoustical wave travels away from its source, the signal power decreases according to the inverse square law as illustrated in Figure 8-2, dropping 6 dB as the distance from the source is doubled (Ma & Ma, 1984).

$$I = \frac{I_o}{4\pi R^2}$$

where:

I = intensity (power per unit area) at distance R
I_o = maximum (initial) intensity
R = range.

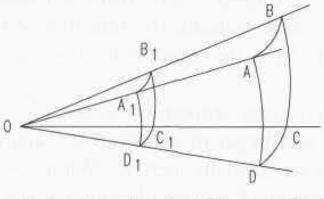

Figure 8-2. Neglecting atmospheric attenuation, the total energy flowing within the cone OABCD is independent of the distance at which it is measured, whereas the intensity per unit area falls off with the square of R (adapted from Feynman, et al., 1963).

There is also an exponential loss associated with molecular absorption of sound energy by the medium itself (Pallas-Areny & Webster, 1992):

$$I = I_o \, e^{-2\alpha R}$$

where:

α = attenuation coefficient for medium.

The value of α varies slightly with the humidity and dust content of the air and is a function of the operating frequency as well (higher frequency transmissions attenuate at a faster rate). The maximum detection range for an ultrasonic sensor is thus dependent on both the emitted power and frequency of operation: the lower the frequency, the longer the range.

The maximum theoretical attenuation for ultrasonic energy (Shirley, 1989) can be approximated by:

$$a_{\max} = \frac{f}{100}$$

where:

a_{max} = maximum attenuation in dB/foot
f = operating frequency in KHz.

For a 20-KHz transmission, a typical absorption factor in air is approximately 0.02 dB/foot, while at 40 KHz losses run between 0.06 and 0.09 dB/foot (Ma & Ma, 1984).

Combining the above *spherical-divergence* and *molecular-absorption* attenuation factors results in the following governing equation for intensity as a function of distance R from the source:

$$I = \frac{I_o \, e^{-2\alpha R}}{4\pi \, R^2} .$$

Note that in this expression, which does not yet take into consideration any interaction with the target surface, intensity falls off with the square of the distance.

8.2.2 Target Reflectivity

The totality of all energy incident upon a target object is either reflected or absorbed, be it acoustical, optical, or RF in nature. The *directivity* of the target surface determines how much of the reflected energy is directed back towards the transducer. Since most objects scatter the signal in an isotropic fashion, the returning echo again dissipates in accordance with the inverse square law (Biber, et al., 1980), introducing an additional $4\pi R^2$ term in the denominator of the previous equation for intensity. In addition, a new factor K_r must be introduced in the numerator to account for the *reflectivity* of the target:

$$I = \frac{K_r \, I_o \, e^{-2\alpha R}}{16\pi^2 \, R^4}$$

where:

K_r = coefficient of reflection.

This *coefficient of reflection* for a planar wave arriving normal to a planar object surface is given by (Pallas-Areny & Webster, 1992):

$$K_r = \frac{I_r}{I_i} = \left(\frac{Z_a - Z_o}{Z_a + Z_o} \right)^2$$

where:

I_r = reflected intensity
I_i = incident intensity
Z_a = acoustic impedance for air
Z_o = acoustic impedance for the target object.

The bigger the impedance mismatch between the two media, the more energy will be reflected back to the source. In industrial applications, this phenomenon allows tank level measurement to be accomplished using an ultrasonic transducer in air looking down on the liquid surface, or alternatively an immersed transducer looking upward at the fluid/air interface.

Most targets are *specular* in nature with respect to the relatively long wavelength (roughly 0.25 inch at 50 KHz) of ultrasonic energy, as opposed to being *diffuse*. In the case of *specular reflection*, the angle of reflection is equal to the angle of incidence, whereas for *diffuse reflection* energy is scattered in various directions by surface irregularities equal to or larger than the wavelength of incident radiation. *Lambertian* surfaces are ideal diffuse reflectors that in theory scatter energy with equal probability in all directions (Jarvis, 1983).

To develop a more or less intuitive appreciation for this relationship to wavelength, it is perhaps helpful to consider the analogy of a pair of rubber balls impacting a hypothetical surface with the sawtooth profile shown below in Figure 8-3. Assume one ball is approximately an inch in diameter, while the other is a much larger basketball. If the sawtooth dimension *d* is in the same neighborhood as the diameter of the smaller ball, then there is a good chance this ball when approaching the surface at some angle of incidence θ will bounce back towards its origin as shown in Figure 8-3A. This is because on the scale of the smaller ball, the surface has a significant normal component. On the other hand, when the basketball impacts this sawtooth surface with the same angle of incidence, the surface irregularities are too small with respect to the ball diameter to be effective. The basketball therefore deflects in a specular manner as shown in Figure 8-3B.

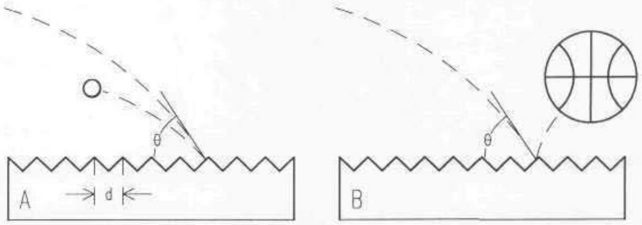

Figure 8-3. A small ball impacting a sawtooth surface as shown will generally bounce back towards its origin, whereas a ball much larger in diameter than the sawtooth dimension *d* will bounce away in specular fashion.

A familiar example of this effect at optical wavelengths can be seen when the beam of an ordinary flashlight is pointed towards a wall mirror at roughly a 45-degree angle. The footprint of illumination on the mirror surface is not visible, because all the light energy is deflected away in a specular fashion. In other words, you can't see the flashlight spot on the mirror itself. Now suppose the flashlight is redirected slightly towards the wall adjacent to the mirror. The spot of light shows up clearly on the wall surface, which is Lambertian in nature with respect to the wavelength of light. The wall is thus a *diffuse reflector* as opposed to a *specular reflector* for optical energy.

When the angle of incidence of a sonar beam decreases below a certain critical value, the reflected energy does not return to strike the transducer. The obvious reason for this effect is the normal component falls off as the angle of incidence becomes more shallow, as illustrated in Figure 8-4. This critical angle is a function of the operating frequency chosen and the topographical characteristics of the target surface.

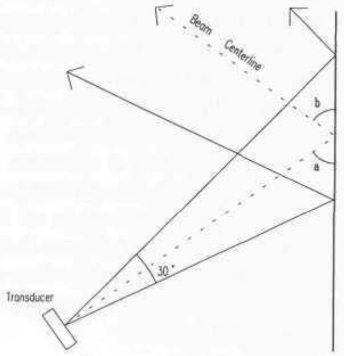

Figure 8-4. As the angle of incidence decreases below a certain critical angle, reflected energy no longer returns to the transducer.

For the Polaroid electrostatic transducers this angle turns out to be approximately 65 degrees (i.e., 25 degrees off normal) for a flat target surface made up of unfinished plywood. Transducer offset from the normal will result in either a false echo as deflected energy returns to the detector over an elongated path, or no echo as the deflected beam dissipates. In Figure 8-4 above, the ranging system would not see the wall and indicate instead maximum range, whereas in Figure 8-5 the range reported would reflect the total round trip through points A, B, and C as opposed to just A and B.

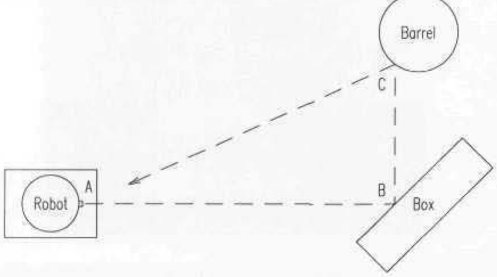

Figure 8-5. Due to specular reflection, the measured range would represent the round trip distance through points A, B, and C as opposed to the actual distance between A and B (adapted from Everett, 1985).

When the first prototype of the MDARS Interior robot was delivered to another government laboratory for formal *Technical Feasibility Testing* in early 1991, the narrow-beam collision avoidance sonar array installed by NCCOSC experienced significant problems in the form of false echo detections. These erroneous sonar readings were quickly seen to correlate with the presence of periodic expansion joints in the concrete floor surface of the test facility. The transducers in the forward-looking array were purposely installed with a 7-degree down angle to increase the probability of detection for low-lying obstructions. This approach had worked very well in our building over months of extended operations, because the smooth floors were very *specular targets* with no significant discontinuities. An overnight field change realigning the sonar beams to a horizontal orientation was required to resolve the problem.

Any significant absorption can result in a reduction of the reflected wave intensity with an adverse impact on system performance. For example, the Polaroid ultrasonic system has an advertised range of 35 feet. In testing the security module on the MDARS Interior robot (Figure 8-6), however, we found it was difficult to pick up an average size person standing upright much beyond a distance of 23 feet. Harder targets of smaller cross-sectional area, on the other hand, could be seen out to the maximum limit.

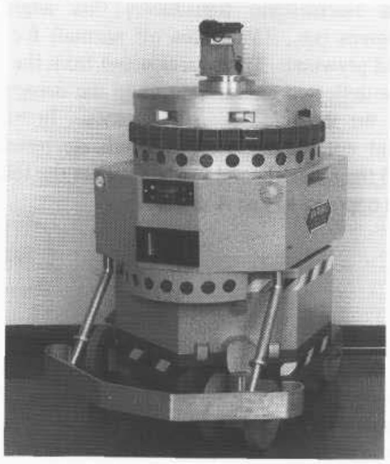

Figure 8-6. The early prototype security sensor suite on the MDARS Interior robot consisted of 24 Polaroid ultrasonic transducers, 48 passive infrared motion detectors, and six microwave motion detectors (courtesy Naval Command Control and Ocean Surveillance Center).

The amount of energy coupled into the target surface (i.e., absorbed) versus that reflected is basically determined by the difference in *acoustic impedance* (*Z*) between the propagation medium (air) and the target object itself. Typical values for *Z* are listed in Table 8-1. Maximum transmission of energy occurs in the case of a fully homogeneous medium where *Z* is uniform throughout. For non-homogeneous situations involving an interface between two dissimilar media, effective coupling falls off (and reflectivity subsequently goes up) as the differential in *Z* increases. The *coefficient of transmission* for a planar wave incident upon a planar target in a direction normal to the target surface is given by (Pallas-Areny & Webster, 1992):

$$K_t = \frac{I_t}{I_i} = \frac{4 Z_a Z_o}{(Z_a + Z_o)^2}$$

where:

K_t = coefficient of transmission (absorption)
I_t = transmitted intensity
I_i = incident intensity
Z_a = acoustic impedance for air
Z_o = acoustic impedance for the target object.

Table 8-1. Typical values of acoustical impedance (*Z*) for various conducting media (adapted with permission from Bolz & Tuve, 1979, © CRC Press, Boca Raton, FL; and Pallas-Areny & Webster, 1992).

Medium	Z	Units
Air	4.3×10^{-4}	million Pascal-seconds/meter
Cork	1.0	million Pascal-seconds/meter
Water	1.5	million Pascal-seconds/meter
Human tissue	1.6	million Pascal-seconds/meter
Rubber	3.0	million Pascal-seconds/meter
Glass	13	million Pascal-seconds/meter
Aluminum	17	million Pascal-seconds/meter
Steel	45	million Pascal-seconds/meter
Gold	62.5	million Pascal-seconds/meter

The original Polaroid ranging module transmitted a 1-millisecond *chirp* consisting of four discrete frequencies: 8 cycles at 60 KHz, 8 cycles at 56 KHz, 16 cycles at 52.5 KHz, and 24 cycles at 49.41 KHz (Biber, et al., 1980). This technique was employed to increase the probability of signal reflection from the target, since certain surface characteristics could theoretically absorb a single-frequency waveform, preventing detection. In actual practice such frequency-dependent effects rarely arose, suggesting this aspect of the absorption problem had been somewhat overestimated. In fact, Polaroid subsequently developed an improved version of the ranging module circuit board, the *SN28827*, that operated at a single frequency of 49.1 KHz.

My daughter Rebecca compiled a significant amount of empirical data in 1993 as part of her high school science fair project entitled *Determining the Accuracy of an Ultrasonic Ranging Sensor*. One of her tests investigated the reflective properties of various target surfaces measuring 16 by 24 inches. The targets were maintained normal to a temperature-compensated Polaroid sensor (a *Digitape* ultrasonic tape measuring unit made by Houseworks) mounted 14 inches above a smooth concrete floor and 35 feet away. Starting at a point beyond the maximum range of detection, the distance between the sensor and target was decreased in 1-foot increments until a valid range reading was obtained. The following table is reproduced here with her permission:

Table 8-2. Maximum detection ranges for standardized 16- by 24-inch cross-sections of various materials.

Surface	Distance	Reading	Units
Plywood	24	24.2	feet
Towel	22	22.3	feet
Underside of rug	16	16.3	feet
Foam	13	13.3	feet
Pillow	9	9.4	feet
Blanket	8	8.4	feet
Top of rug	3	3.6	feet

8.2.3 Air Turbulence

Turbulence due to wind and temperature variations can cause bending or distortion of acoustical energy traveling through air (Shirley, 1989). Wind direction and velocity can have a noticeable push or delay effect on the wave propagation velocity, more relevant in the case of outdoor vehicles. Consideration of wind effect errors must also treat crosswind components in addition to those which travel on a parallel path either with or against the wavefront. Crosswind effects can cause the beam center to be offset from its targeted direction, diminish the intensity of returned echoes, and result in a slightly longer beam path due to deflection.

In general, little effort is made in the case of mobile robotic applications to correct for such errors. This is probably due to the fact that ultrasonic ranging is most widely employed in indoor scenarios where the effects of air turbulence are minimal, unless extreme measurement accuracy is desired. In addition, there is really no practical way to reliably measure the phenomena responsible for the interference, and so compensation is generally limited to averaging over multiple readings. This approach introduces a coordinate transformation requirement in the case of a moving platform, since the slow speed of sound limits effective update rates to roughly 2 Hz (i.e., single transducer, assuming 28 feet maximum range). Faster updates are possible if the system is range-gated to some lesser distance (Gilbreath & Everett, 1988).

8.2.4 Temperature

Recall the earlier expression for wave propagation speed (s) in a gas, as a function of density ρ and bulk modulus of elasticity K_m:

$$s = \sqrt{\frac{K_m}{\rho}} .$$

Since both these parameters change with temperature, the speed of sound in air is also temperature dependent (Pallas-Areny & Webster, 1992), and in fact directly proportional to the square root of temperature in degrees Rankine (Everett, 1985):

$$s = \sqrt{g\,k\,R\,T}$$

where:

s = speed of sound
g = gravitational constant
k = ratio of specific heats
R = gas constant
T = temperature in degrees Rankine (F + 460).

For temperature variations typically encountered in indoor robotic ranging applications, this dependence results in a significant effect even considering the short distances involved. A temperature change over the not unrealistic span of 60° to 90°F can produce a range error as large as 12 inches at a distance of 35 feet. Fortunately, this situation can be remedied through the use of a correction factor based upon the actual ambient temperature, available from an external sensor mounted on the robot. The formula is simply:

$$R_a = R_m \sqrt{\frac{T_a}{T_c}}$$

where:

R_a = actual range
R_m = measured range
T_a = actual temperature in degrees Rankine

T_c = calibration temperature in degrees Rankine.

The possibility does still exist, however, for temperature gradients between the sensor and the target to introduce range errors, since the correction factor is based on the actual temperature in the immediate vicinity of the sensor only. As in the case of air turbulence, there is generally little recourse other than averaging multiple readings. (Some industrial applications provide a temperature-stabilized column of air using a small blower or fan.)

8.2.5 Beam Geometry

Still another factor to consider is the beamwidth of the selected transducer, defined as the angle between the points at which the sound power has been reduced to half (-3 dB) its peak value (Shirley, 1989). This formal definition does not always map directly into any useful parameter in real-world usage, however. What is generally of more concern can be better described as the *effective beamwidth*, or the beam geometry constraints within which objects are reliably detected. (Reliable detection, of course, is also very much dependent on the size and shape of the object.) The width of the beam is determined by the transducer diameter and the operating frequency. The higher the frequency of the emitted energy, the narrower and more directional the beam, and hence the greater the angular resolution. Recall, however, an increase in frequency causes a corresponding increase in signal attenuation in air and decreases the maximum range of the system.

The wavelength of acoustical energy is inversely proportional to frequency as shown below:

$$\lambda = \frac{s}{f}$$

where:

λ = wavelength
s = speed of sound
f = operating frequency.

The *beam-dispersion angle* is directly proportional to this transmission wavelength (Brown, 1985):

$$\theta = 1.22\frac{\lambda}{d}$$

where:

θ = desired dispersion angle
λ = acoustic wavelength
d = transducer diameter.

The above relationship can be intuitively visualized by considering the limiting case where *d* approaches zero. Such a hypothetical device would theoretically function as a point source, emitting energy of equal magnitude in all directions. As *d* is increased, the device can be considered a planar array of point sources clustered together in circular fashion. For this configuration, the emitted energy will be in phase and at maximum intensity only along a surface normal. Destructive interference from adjacent point sources causes the beam intensity to fall off rapidly to either side up to some local minimum value as shown in Figure 8-7. Constructive interference then occurs past this minimum point, resulting in the presence of side lobes.

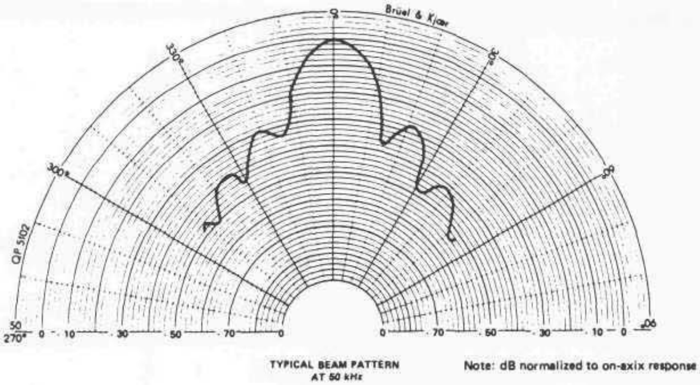

Figure 8-7. Constructive interference results in maximum power in the main lobe along the beam center axis (courtesy Polaroid Corp.).

Shirley (1989) defines the spot diameter that is *insonified* by the ultrasonic beam (i.e., footprint of the incident beam at the target surface) in terms of this beam-dispersion angle θ:

$$D = 2\,R\,\tan\frac{\theta}{2}$$

where:

D = spot diameter
R = target range.

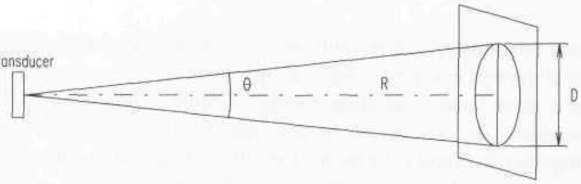

Figure 8-8. The diameter of the insonified footprint at the target surface, assuming normal incidence.

Best results are obtained when the beam centerline is maintained normal to the target surface. As the angle of incidence varies from the perpendicular, note the range actually being measured does not always correspond to that associated with the beam centerline (Figure 8-9). The beam is reflected first from the portion of the target closest to the sensor. For a 30-degree beam-dispersion angle at a distance of 15 feet from a flat target, with an angle of incidence of 70 degrees with respect to normal, the theoretical error could be as much as 10 inches. The actual line of measurement intersects the target surface at point B as opposed to point A.

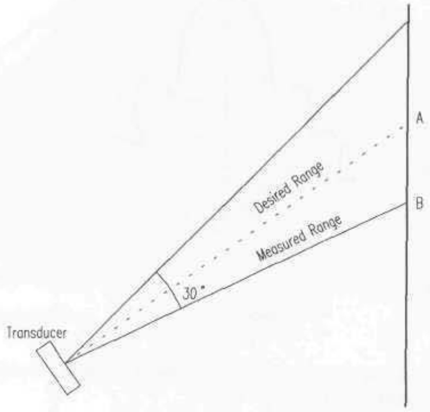

Figure 8-9. Ultrasonic ranging error due to beam divergence results in a shorter range measurement to the target surface at B instead of the desired reading to point A.

Effective beamwidth introduces some uncertainty in the perceived distance to an object from the sensor but an even greater uncertainty in the angular resolution of the object's position. A very narrow target such as a vertical pole would have a relatively large associated region of floor space that would essentially appear to the sensor to be obstructed. Worse yet, a 3-foot doorway may not be discernible at all when only 6 feet away, simply because at that distance the beam is wider than the door opening.

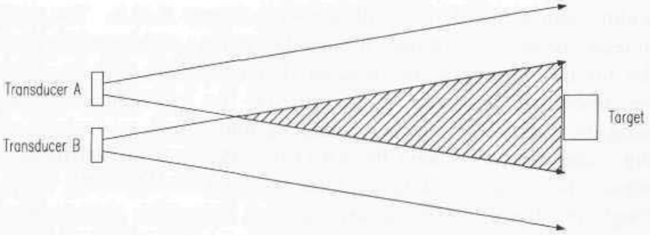

Figure 8-10. Beam-splitting techniques using two or more sensors can improve angular resolution for discrete targets (adapted from Everett, 1985).

Improved angular resolution can sometimes be obtained through beam splitting, a technique that involves the use of two or more transducers with partially overlapping beam patterns. Figure 8-10 shows how for the simplest case of two transducers, twice the angular resolution can be obtained along with a 50-percent increase in coverage area. If the target is detected by both sensors A and B, then it (or at least a portion of it) must lie in the region of overlap shown by the shaded area. If detected by A but not B, then it lies in the region at the top of the figure, and so on. Increasing the number of sensors with overlapping beam patterns decreases the size of the respective regions, and thus increases the angular resolution.

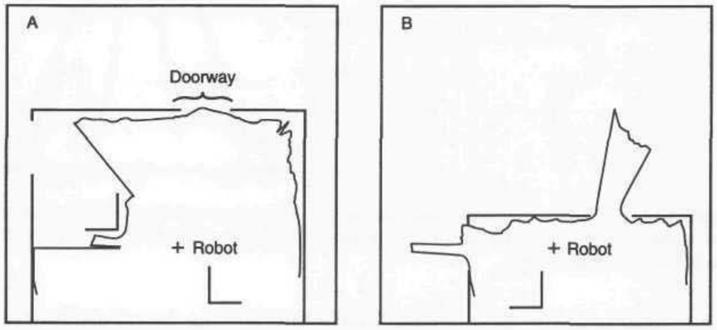

Figure 8-11. With ROBART II situated as shown in (A), the sonar beamwidth is too wide to see the open doorway; relocation of the robot as depicted in (B) allows several readings to penetrate as the head-mounted transducer is scanned left to right (plot courtesy MIT AI Lab).

It should be noted, however, that this increase in angular resolution is limited to the case of a discrete target in relatively uncluttered surroundings, such as a metal pole supporting an overhead load or a lone box in the middle of the floor. No improvement is seen for the case of an opening smaller than an individual beamwidth, such as the doorway illustrated in Figure 8-11A. The entire beam from at least one sensor must pass through the opening without striking either side in order for the opening to be detected (Figure 8-11B), and the only way to improve resolution otherwise is to decrease the individual beamwidths by increasing the operating frequency, changing transducers, or through acoustical focusing. Some designs achieve this effect through use of an attachable horn that concentrates the energy into a tighter, more powerful beam (Shirley, 1989).

Kilough and Hamel (1989) incorporated an innovative arrangement of four Polaroid transducers in a *phased-array* cluster on HERMIES as an alternative method of reducing the effective beamwidth. Figure 8-12A shows how two pairs of transducers connected in parallel are in turn wired in series to preserve the original impedance seen by the driving module. The constructive and destructive interferences resulting from the close proximity of the sensors produced the measured pattern depicted in Figure 8-12B, which agrees with the theoretical predictions for the case of a 38-millimeter center spacing and a 6.6-millimeter wavelength at 50 KHz. Note the nulls in the pattern at 5 degrees either side of centerline, and the associated side lobes. To minimize this effect, the transducers are mounted as closely spaced as possible. Measured beamwidth was reduced from 30 to 18 degrees.

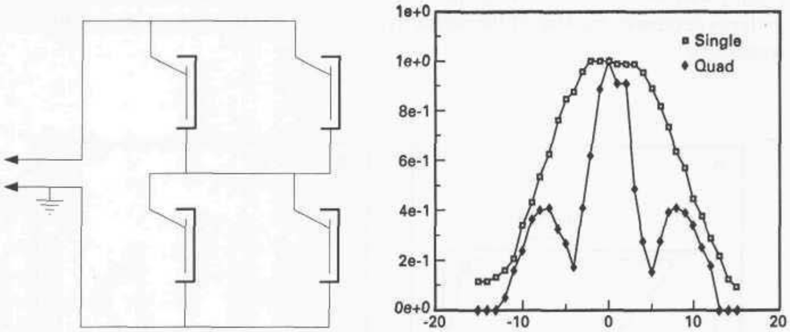

Figure 8-12. Four Polaroid electrostatic transducers wired in parallel-series fashion as shown form a phased-array configuration with the resulting beam pattern shown on the right (adapted from Kilough & Hamel, 1989).

A number of factors must be considered when choosing the optimal beamwidth for a particular application. A narrow beamwidth will not detect unwanted objects to either side, is less susceptible to background noise, and can achieve greater ranges since the energy is more concentrated (Shirley, 1989). One the other hand, for collision avoidance applications it is often desirable to detect any and all objects in front of the robot, and since extremely long ranges are not usually required, a wide-angle transducer may be a more optimal choice (Hammond, 1993). When comparing a single transducer of each type, the use of a wide beamwidth will improve chances of target detection due to the greater likelihood of some portion of the beam encountering a surface normal condition as seen in Figure 8-13. Admittedly this observation is a bit like saying the wider the beam, the more chance of hitting a target. Taken to the extreme, a hypothetical 360-degree field-of-view transducer is clearly of rather limited utility due to the total lack of azimuthal information regarding the target's whereabouts.

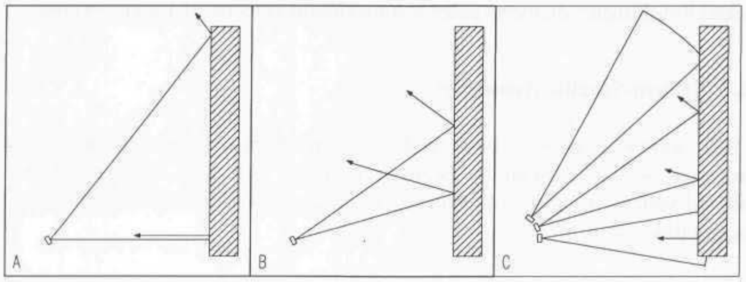

Figure 8-13. A wide-angle transducer (A) has a greater chance of encountering a surface normal condition than a single narrow-beam transducer (B), but at the expense of reduced angular resolution and effective range. A more optimal configuration is presented in (C), at a slight cost to system update rate.

Alternatively, an equivalent surface-normal condition can be realized using a cylindrical array of narrow-beam transducers to achieve the same volumetric coverage as illustrated in Figure 8-13C. This approach offers the added advantage of significantly improved angular resolution but at the expense of a slower overall update rate. The MDARS Interior robot uses a combination of wide-angle piezoelectric sonars operating at a frequency of 75 KHz for timely *obstacle detection* coverage, and a nine-element array of narrow-beam Polaroid electrostatic transducers operating at 49.4 KHz to support intelligent *obstacle avoidance*. Detection of any potential obstructions by either type of sonar causes the platform to slow to a speed commensurate with the narrow-beam update rate, whereupon the high-resolution Polaroid data is used to formulate an appropriate avoidance maneuver.

8.2.6 Noise

Borenstein & Koren (1992) of the University of Michigan Mobile Robotics Lab define three types of noise affecting the performance of ultrasonic sensors:

- *Environmental noise* resulting from the presence of external sources operating in the same space. Typical examples in industrial settings include high-pressure air blasts and harmonics from electrical arc welders.
- *Crosstalk* resulting from the proximity of other sensors in the group, which can be especially troublesome when operating in confined areas.

- *Self noise* generated by the sensor itself.

A noise-rejection measure for each of the components was developed and integrated into a single algorithm (Michigan, 1991), which was in turn combined with a fast sensor-firing algorithm. This software has been implemented and tested on a mobile platform that was able to traverse an obstacle course of densely packed 8-millimeter-diameter poles at a maximum velocity of 1 meter/second.

8.2.7 System-Specific Anomalies

A final source of error to be considered stems from case-specific peculiarities associated with the actual hardware employed. We shall again refer to the Polaroid system, in light of its widespread usage, as an illustrative example in the ensuing discussion.

Pulse Width

The 1-millisecond length of the original four-frequency Polaroid *chirp* was a potential source of range measurement error since sound travels roughly 1100 feet/second at sea level, which equates to about 13 inches/millisecond. The uncertainty and hence error arose from not knowing which of the four frequencies making up the *chirp* actually returned to trigger the receiver, but timing the echo always began at the start of the *chirp* (Everett, 1985). For the initial application of automatic camera focusing, designers were less concerned about absolute accuracy than missing a target altogether due to surface absorption of the acoustical energy. The depth of field of the camera optics would compensate for any small range errors that might be introduced due to this *chirp* ambiguity.

Even with the more recent *SN28827* ranging model operating at a single frequency of 49.1 KHz, the transmission pulse duration is 0.326 milliseconds, giving rise to a maximum theoretical error of approximately 1.7 inches. (This estimate takes into account round-trip distance, and assumes best-case echo detection after just three cycles of returned energy.) The new Polaroid *Ultrasonic Ranging Developer's Kit* allows for programmable pulse duration to alleviate this limitation in demanding applications (Polaroid, 1993).

Threshold Detection

The specific method for detection of the returned pulse can be a significant source of error in any TOF ranging system (Figueroa & Lamancusa, 1992). Kuc and Siegel (1987) point out that the intensity of a typical pulse transmission peaks in the second cycle (Figure 8-14), and so simple thresholding of the received signal can cause late detection of weak echoes. Leonard and Durrant-Whyte (1992) discuss further complications in the specific case of the integrating capacitive threshold detector employed in the Polaroid ranging module. This integrative approach was incorporated by the designers to discriminate against unwanted noise spikes (Biber, et al., 1980). Compared to strong reflections, valid but weak echo returns can take substantially longer to charge up the capacitor to the threshold level required for the comparator to change state (i.e., the *time-walk* problem identified in Chapter 5).

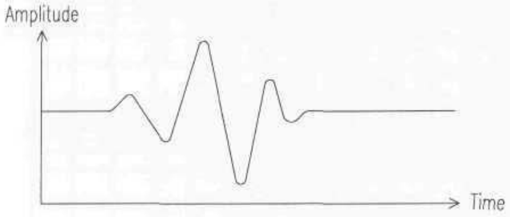

Figure 8-14. A typical pulse waveform for an electrostatic transducer can be approximated by a sinusoid that is modulated by a Gaussian envelope, peaking in intensity during the second cycle (Kuc & Siegel, 1987, © IEEE).

The effect of this charging delay is to make those targets associated with weaker returns appear further away. Ignoring the obvious worst-case scenario of a completely missed echo, maximum theoretical error is bounded by the length of the transmitted burst. The obvious question now becomes, which is more preferable: missing target detection altogether, or being alerted to target presence at the expense of range accuracy? The answer of course depends on the particular priorities of the application addressed. If the ranging sensor is being employed as a presence detector for security purposes (see Chapter 17), precise accuracy is not all that important. On the other hand, if the sensor is being used for navigational referencing (see Chapters 15 and 16), the situation may be somewhat different.

Stepped Gain

Lang, et al. (1989) experimentally confirmed error effects associated with the piecewise 16-step gain ramp employed on the earlier Polaroid *607089* ranging module. In order to precisely counter the effects of signal loss as a function of range to target (i.e., due to atmospheric attenuation and spherical divergence), the actual time-dependent gain compensation would be an exponential function inversely related to the equation presented in Section 8.2.2. A rather coarse piecewise approximation to this ideal gain curve (Figure 8-15) naturally results in a situation where the instantaneous amplifier gain is: 1) correct only for a single point in time over the period represented by a specific step value, 2) excessive prior to this point, and 3) insufficient afterwards. If the gain is too low at the time of reflected pulse train arrival, weak echoes are either missed entirely or delayed in triggering the integrating detector, resulting in an erroneous increase in the perceived range.

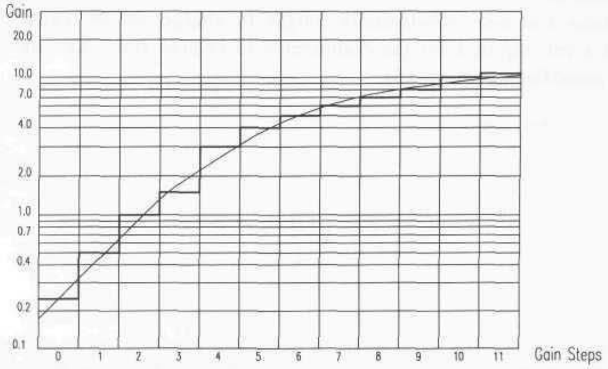

Figure 8-15. The 12-step approximation employed in the new *6500-series* receiver gain ramp results in a situation where the instantaneous gain is either above or below the ideal value for most of the step duration (adapted from Polaroid, 1993). Note the large jump in gain between steps 3 and 4.

8.3 Choosing an Operating Frequency

The operating frequency of an ultrasonic ranging system should be selected only after careful consideration of a number of factors, such as the diameter and type of transducer, anticipated target characteristics, sources of possible interference, and most importantly the nature of the intended task, to include desired angular and range resolution. Resolution is dependent on the bandwidth of the transmitted energy, and greater bandwidth can be achieved at higher frequencies but at the expense of maximum effective range. The minimum ranging distance is also a function of bandwidth, and thus higher frequencies are required in close as the distance between the detector and target decreases. Most man-made background noise sources have energy peaks below 50 KHz (Hammond, 1993), however, and so higher-frequency systems are generally preferred in acoustically noisy environments (Shirley, 1989).

8.4 Sensor Selection Case Study

The Department of Energy's Office of Technology Development has an ongoing environmental restoration effort that, among other things, seeks to develop a robotic inspection capability (Figure 8-16) for stacked 55-gallon drums of hazardous waste stored inside warehouse facilities (Byler, 1993). Aside from the obvious challenges associated with global navigation and collision avoidance, the mobile robot must be able to appropriately position itself next to a stack of barrels before activating the onboard inspection system (Heckendorn, et al., 1993). This *barrel-detection* requirement presents an interesting hypothetical case study for reviewing some of the factors discussed in this chapter that determine the applicability of an ultrasonic ranging sensor for a particular task. Careful consideration of the nature of the target surface in conjunction with the reflective-sensor performance factors (i.e., *cross-sectional area, reflectivity*, and *directivity*) can provide some helpful insights into appropriate sensor selection.

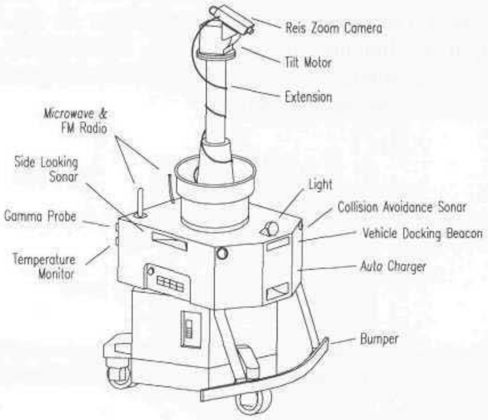

Figure 8-16. The Savannah River Site nuclear surveillance robot is equipped with an extending vertical mast illustrative of the type that could potentially be employed in a stacked barrel inspection task (adapted from Fisher, 1989).

Let's examine first the nature of the object of interest itself, namely the barrel, in terms of its target characteristics. The diameter and height of each barrel is expected to be somewhat standardized, although there may likely be some minor variations encountered. In general, however, the *cross-sectional area* of the barrel will be very large with respect to the sensor due to extremely close proximity to the robot. There is therefore no decided advantage given in this particular application to ultrasonic, RF, or optically based systems from the standpoint of *target cross-sectional area*.

The material from which the barrels are constructed is of particular interest. An all-metal barrel would suggest there might possibly be some advantage to using a low-power microwave-based ranging system to take advantage of the inherent high *reflectivity* with virtual immunity to varying surface conditions. There is no guarantee, however, that plastic barrels will not be used in some cases, perhaps for containment of corrosive agents. Plastic provides poor reflectivity for RF energy, but both plastic and metal surfaces reflect ultrasonic energy well. The color and material condition of the target surface will have little effect on either RF or ultrasonic energy, but could seriously impair performance of an active optical ranging system.

By far the most important target characteristic is the cylindrical geometry, significant in that it represents a common feature for all barrels in the warehouse and directly impacts *directivity*. A cylindrical reflector will scatter energy in all directions (i.e., as is desired in stealth applications) rather than redirect it in a concentrated fashion back towards the sensor (i.e., as in the case of a cooperative retroreflective target). This scattering reduces the strength of the return, but at the same time ensures some energy does get reflected back towards the sensor, provided an imaginary ray passing from the transducer to the cylinder axis intersects the footprint of illumination (Figure 8-17). At extremely close ranges, the resultant loss in signal strength is not a problem, and the effect of a guaranteed surface normal turns out to be highly beneficial.

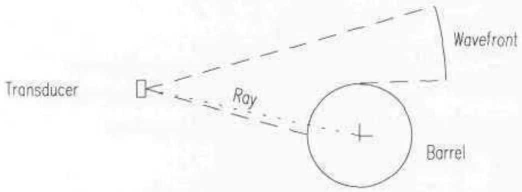

Figure 8-17. A surface normal condition exists as long as the footprint of illumination of the propagating wavefront contains a ray originating at the transducer and passing through the vertical axis of the cylindrical barrel.

The situation illustrated in Figure 8-17 above suggests the fairly wide beamwidth of an ultrasonic ranging system is actually somewhat advantageous when trying to detect a cylindrical target, as it increases the chances of a surface normal condition. In addition, it is illustrative to note that the measured range is always along the path of the ray which passes through the center of the drum, even with the transducer offset as shown. Since the barrel radius is known in advance, this observation allows for distance to be calculated to the barrel centerpoint itself. A number of such measurements made as the robot advances along a known baseline enables the precise location of the barrel to be determined through triangulation.

In summary, an ultrasonic ranging system would be expected to give good results in this particular application for the following reasons:

- Very low cost and easy to interface.
- Barrel surfaces, whether metal or plastic, provide excellent reflectivity regardless of surface color or condition.
- Fairly wide beam increases probability of detection.
- Ranges can easily be derived to barrel centerline.

8.5 References

Asher, R.C., "Ultrasonic Sensors in the Chemical and Process Industries," Journal of Physics E.: Scientific Instruments, Vol. 16, pp. 959-963, 1983.

Bartram, J.F., Ehrlich, S.L., Fredenberg, D.A., Heimann, J.H., Kuzneski, J.A., Skitzki, P., "Underwater Sound Systems," in *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, McGraw Hill, New York, NY, pp. 25.95-25.133, 1989.

Biber, C., Ellin, S., Shenk, E., "The Polaroid Ultrasonic Ranging System," Audio Engineering Society, 67th Convention, New York, NY, October-November, 1980.

Bolz, R.E., Tuve, G.L., *CRC Handbook of Tables for Applied Engineering Science*, CRC Press, Boca Raton, FL, 1979.

Borenstein, J, Koren, Y., "Error Eliminating Rapid Ultrasonic Firing for Mobile Robot Obstacle Avoidance," IEEE International Conference on Robotics and Automation, Nice, France, May, 1992.

Brown, M.K., "Locating Object Surfaces with an Ultrasonic Range Sensor," IEEE Conference on Robotics and Automation, St. Louis, MO, pp.110-115, March, 1985.

Byler, E., "Intelligent Mobile Sensor System for Drum Inspection and Monitoring," Phase I Topical Report, DOE Contract DE-AC21-92MC29112, Martin Marietta Astronautics Group, Littleton, CO, June, 1993.

Campbell, D., "Ultrasonic Noncontact Dimensional Measurement," *Sensors*, pp. 37-43, July, 1986.

Dunkin, W.M., "Ultrasonic Position Reference Systems for an Autonomous Sentry Robot and a Robot Manipulator Arm", Masters Thesis, Naval Postgraduate School, Monterey, CA, March, 1985.

Everett, H.R., "A Multielement Ultrasonic Ranging Array," *Robotics Age*, pp.13-20 July, 1985.

Feynman, R.P., Leighton, R.B., Sands, M., *The Feynman Lectures on Physics*, Vol. 1, Addison-Wesley, Reading, MA, 1963.

Figueroa, F., Barbieri, E., "Increased Measurement Range Via Frequency Division in Ultrasonic Phase Detection Methods," *Acustica*, Vol. 73, pp. 47-49, 1991.

Figueroa, J.F., Lamancusa, J.S., "A Method for Accurate Detection of Time of Arrival: Analysis and Design of an Ultrasonic Ranging System," Journal of the Acoustical Society of America, Vol. 91, No. 1, pp. 486-494, January, 1992.

Figueroa, J.F., Mahajan, A., "A Robust Navigation System for Autonomous Vehicles Using Ultrasonics," *Control Engineering Practice*, Vol. 2, No. 1, pp. 49-59, 1994.

Fisher, J.J., "Application-Based Control of an Autonomous Mobile Robot," American Nuclear Society, Charleston, SC, Sect. 2-6, pp. 1-8, 1989.

Fox, J.D., Khuri-Yakub, B.T., Kino, G.S., "High-Frequency Acoustic Wave Measurements in Air," IEEE Ultrasonics Symposium, pp. 581-584, 1983.

Frederiksen, T.M., Howard, W.M., "A Single-Chip Monolithic Sonar System," *IEEE Journal of Solid State Circuits*, Vol. SC-9, No. 6, December, 1974.

Gilbreath, G.A., Everett, H.R., "Path Planning and Collision Avoidance for an Indoor Security Robot," SPIE Mobile Robots III, Cambridge, MA, pp. 19-27, Novemeber, 1988.

Halliday, D., Resnick, R., *Fundamentals of Physics*, John Wiley, New York, NY, 1974.

Hammond, W., "Smart Collision Avoidance Sonar Surpasses Conventional Systems," *Industrial Vehicle Technology '93: Annual Review of Industrial Vehicle Design and Engineering*, UK and International Press, pp. 64-66, 1993.

Heckendorn, F.M., Ward, C.W., Wagner, D.G., "Remote Radioactive Waste Drum Inspection with an Autonomous Mobile Robot," ANS Fifth Topical Meeting on Robotics and Remote Systems, American Nuclear Society, Knoxville, TN, pp. 487-492, April, 1993.

Irwin, C.T., Caughman, D.O., "Intelligent Robotic Integrated Ultrasonic System," Proceedings, Robots 9, Society of Manufacturing Engineers, Detroit, MI, Sect. 19, pp. 38-47, June, 1985.

Jarvis, R.A., "A Laser Time-of-Flight Range Scanner for Robotic Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 5, pp. 505-512, 1983.

Kilough, S.M., Hamel, W.R., "Sensor Capabilities for the HERMIES Experimental Robot," American Nuclear Society, Third Topical Meeting on Robotics and Remote Systems, Charleston, SC, CONF-890304, Section 4-1, pp. 1-7, March, 1989.

Kuc, R., Siegel, M.W., "Physically Based Simulation Model for Acoustic Sensor Robot Navigation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 6, pp. 766-778, November, 1987.

Lang, S., Korba, L., Wong, A., "Characterizing and Modeling a Sonar Ring," SPIE Mobile Robots IV, Philadelphia, PA, pp. 291-304, 1989.

Leonard, J.J., Durrant-Whyte, H.F., *Directed Sonar Sensing for Mobile Robot Navigation*, Kluwer Academic Publishers, Boston, MA, 1992.

Ma, Y.L., Ma, C., "An Ultrasonic Scanner System Used on an Intelligent Robot," IEEE IECON '84, Tokyo, Japan, pp. 745- 748, October, 1984.

Michigan, "Mobile Robotics Lab," Brochure, University of Michigan Mobile Robotics Lab, Ann Arbor, MI, 1991.

Milner, R., ""Measuring Speed and Distance with Doppler," *Sensors*, pp. 42-44, October, 1990.

Mitome, H., Koda, T., Shibata, S., "Double Doppler Ranging System Using FM Ultrasound," *Ultrasonics*, pp. 199-204, September, 1984.

Pallas-Areny, R., Webster, J.G., "Ultrasonic Based Sensors," *Sensors*, pp. 16-20, June, 1992.

Polaroid, "Polaroid Ultrasonic Ranging Developer's Kit," Publication No. PXW6431 6/93, Polaroid Corporation, Cambridge, MA, June, 1993.

Shirley, P.A., "An Introduction to Ultrasonic Sensing," *Sensors*, pp. 10-17, November, 1989.

Shirley, P.A., "An Ultrasonic Echo-Ranging Sensor for Board Inspection and Selection," *Sensors*, June, 1991.

Smurlo, R.P., Everett, H.R., "Intelligent Sensor Fusion for a Mobile Security Robot," *Sensors*, pp. 18-28, June, 1993.

Ulrich, R., *Principles of Underwater Sound for Engineers*, 1983.

9

Electromagnetic Energy

It is important to realize that *acoustical energy* and *electromagnetic energy* are two fundamentally different phenomena with some very dissimilar properties and characteristics. Perhaps most obvious is the fact that sound travels rather slowly as a longitudinal pressure wave, whereas radio and optical energy propagate as electromagnetic waves at the speed of light. Additionally, sound must be *conducted* through some type of *medium*, whereas none is required for optical or RF energy, as evidenced by their effective transmission through the vacuum of space. In fact, for electromagnetic radiation, the earth's atmosphere is not an enabling mechanism at all but rather an impediment to propagation.

The electromagnetic spectrum is depicted in Figure 9-1. All the waves represented are electromagnetic in nature, travel at the same speed c in free space, and differ only in wavelength and the type of source with which they are generated (Halliday & Resnick, 1974). The *index of refraction n* relates the speed of light in a particular medium to the speed of light in a vacuum as follows:

$$n = \frac{c}{c_m}$$

where:

n = refractive index of medium
c = speed of light in a vacuum
c_m = speed of light in medium.

The value of n for air is 1.0003; light therefore travels slightly slower in air than in the vacuum of space, and slower still in water, for which n is 1.33. Such changes in propagation velocity are responsible for the bending of light rays at an interface between different media in accordance with Snell's law (Figure 9-2):

$$n_1 \sin\theta_i = n_2 \sin\theta_t$$

where:

n_1 = index of refraction in first medium
θ_i = angle of incidence with respect to surface normal
n_2 = index of refraction in second medium
θ_t = angle of refraction with respect to surface normal.

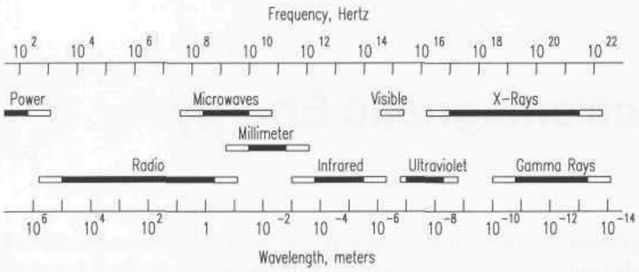

Figure 9-1. The electromagnetic spectrum is divided into several overlapping frequency intervals for which a particular energy source and detection technology exist (adapted with permission from Halliday & Resnick, 1974, © John Wiley and Sons, Inc.).

The index of refraction *n* is a function of the properties of atoms in the medium and the frequency of the electromagnetic radiation (Feynman, et al., 1963):

$$n = 1 + \frac{N q_e^2}{2 \varepsilon_o m (\omega_o^2 - \omega^2)}$$

where:

N = number of charges per unit volume
q_e = charge of an electron
ε_o = permittivity
m = mass of an electron
ω = frequency of the electromagnetic radiation
ω_o = resonant frequency of electron bound in an atom.

In examining the above equation, it can be seen that *n* slowly gets larger as the frequency increases (i.e., as ω approaches ω_o), an effect witnessed by anyone who has ever observed a rainbow in the sky. For example, the index of refraction is higher for blue light than for red. This phenomenon is known as *dispersion*, because the frequency dependence of the index *n* causes a prism to "disperse" the different colors into a distinctive spectrum (Feynman, et al., 1963).

Recall from the earlier discussions on acoustical energy in Chapter 8 that the *Doppler shift* for a moving observer approaching a fixed source was different from that of a moving source closing on a fixed observer at the same rate. In the case of electromagnetic energy, unlike sound, there is no *medium of transmission* relative to which the source and observer are moving (Halliday & Resnick, 1974).

There should therefore be no difference at all in the two cases under consideration: a source moving towards an observer should exhibit the same Doppler shift as an observer moving towards a source with the same relative velocity.

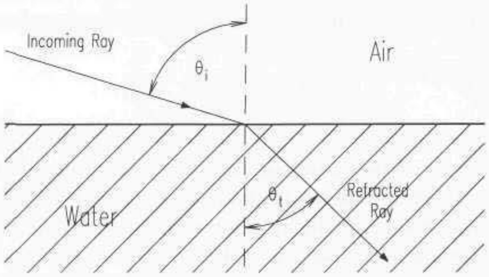

Figure 9-2. Snell's law describes the change in path angle that occurs when light passes a boundary between two media with different values of the refractive index *n*. A ray of light incident upon the surface of water as shown will appear to bend downward.

According to the theory of relativity, the observed frequency is given by (Halliday & Resnick, 1974):

$$f' = f \frac{1 - \frac{v}{c}}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}$$

where:

f' = observed frequency
f = operating frequency of source
v = rate of separation or closure
c = speed of light.

As discussed in Chapter 8, it is usually more convenient to deal with the *Doppler shift frequency*, rather than the observed frequency, which is approximated by (Schultz, 1993):

$$f_d = \pm \frac{2vf \cos\theta}{c}$$

where:

f_d = *Doppler shift* frequency
θ = angular offset from radial path
$v \ll c$.

The sign in the above expression is positive in the case of target closure.

9.1 Optical Energy

The *optical* region of the electromagnetic spectrum is broken up into *ultraviolet*, *visible*, and *infrared* domains as shown in Figure 9-3, where the wavelength is determined by the speed of light (in a vacuum) divided by frequency:

$$\lambda = \frac{c}{f}.$$

The precise limits of the *visible* portion of the spectrum are not well defined because eye sensitivity approaches the horizontal axis in an asymptotic fashion at the upper and lower bounds (Halliday & Resnick, 1974) of approximately 400 and 700 nanometers, respectively (see Figure 9-4 later in this section). Fortunately for us humans, the emission spectrum of the sun conveniently peaks at 555 nanometers, in the very center of the visible range (Herman, et al., 1989).

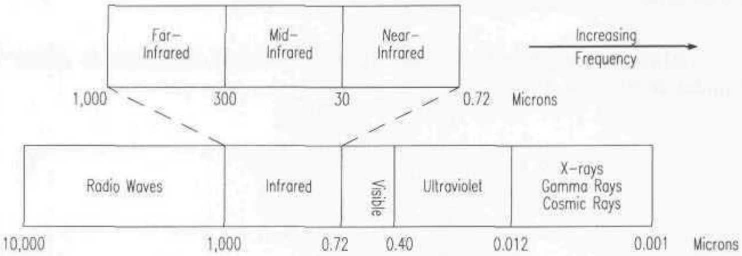

Figure 9-3. The optical portion of the electromagnetic spectrum encompasses wavelengths from 0.012 micron (ultraviolet) up to 1000 microns (infrared); the infrared region can be further subdivided into near-infrared, mid-infrared, and far-infrared (adapted from Banner, 1993, and Buschling, 1994).

The *infrared* portion of the electromagnetic spectrum encompasses wavelengths of 0.72 to 1,000 microns. All objects with an absolute temperature above 0°K emit radiant energy in accordance with the Stephan-Boltzman equation (Buschling, 1994):

$$W = \varepsilon \sigma T^4$$

where:

W = emitted energy
ε = emissivity
σ = Stephan-Boltzman constant (5.67×10^{-12} watts/cm^{2}K^4)
T = absolute temperature of object in degrees Kelvin.

The totality of all energy incident upon an object surface is either absorbed, reflected, or reradiated in accordance with Kirchoff's law. *Emissivity* (ε) is defined as the ratio of radiant energy emitted by a given source to that emitted by a perfect blackbody radiator of the same area at the same temperature under identical conditions (Graf, 1974). *Emissivity* is also a convenient measure of energy absorption. A hypothetical surface with an emissivity value of zero is a perfect reflector, neither absorbing nor emitting radiant energy, whereas in contrast, a theoretical *blackbody* with an ideal emissivity of one would absorb 100 percent of the supplied energy, reflecting none (Buschling, 1994).

9.1.1 Electro-Optical Sources

In 1977 the IEEE redefined a radar as "an electromagnetic means for target location and tracking" (IEEE, 1977). As pointed out by Sundaram (1979), this includes electro-optical devices such as laser radars (lidars) or laser rangefinders in general. Relative to microwave and millimeter-wave systems, electro-optical sensors are characterized by extremely short wavelengths affording much higher resolution, but suffer the greatest attenuation by atmospheric constituents. Active optical sources employed in rangefinding include broadband *incandescent*, broadband *electronic strobes*, narrowband *light emitting diodes*, *super luminescent diodes*, and coherent *lasers*. Each of these devices will be discussed briefly in the following paragraphs.

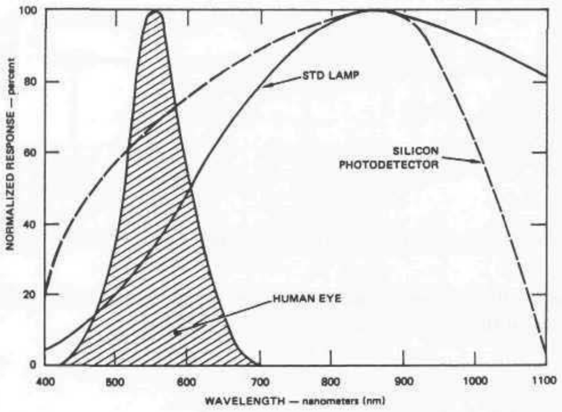

Figure 9-4. Radiation from a tungsten lamp filament is shown in comparison to the spectral sensitivity of silicon (Brain, 1979).

An *incandescent lamp* emits radiation as a consequence of the heating of a conductive filament (typically tungsten) with an electrical current. The higher the operating temperature of the filament, the greater the percentage of radiation falling in the visible portion of the energy spectrum; infrared heat lamps are simply tungsten bulbs operating at low filament temperature (Fox, 1989). The vast majority of energy given off by a conventional tungsten lamp is outside the visible region and in the infrared (Figure 9-4), which explains why fluorescent lighting, with almost no infrared component, is so much more energy efficient. Incandescent lamps transform only about five percent of their filament current into visible light (Mims, 1987). In the case of both infrared and fluorescent sources, the fundamental light emission processes occur in individual atoms, and these atoms do not operate together in a cooperative fashion (Halliday & Resnick, 1974). The resulting light output therefore is said to be *non-coherent*, with random phase variations associated with individual rays.

Electronic strobes consist of a glass or quartz flashtube filled with gas and containing two or more ionizing electrodes connected to a high-voltage DC power supply. Xenon is the preferred fill gas for most applications due to its high white-light spectral peak, but other gases such as argon, neon, krypton, and hydrogen can also be used (Fox, 1989). Figure 9-5 shows a typical trigger circuit in block diagram form. The power supply generally consists of a storage capacitor charged by the rectified output of a step-up transformer driven by an astable multivibrator. A silicon controlled rectifier (SCR) is used as an electronic switch to activate a trigger transformer when a flash is desired, coupling several thousand volts to the trigger electrode of the tube. This trigger pulse ionizes some of the xenon gas, forming a low resistance path that immediately discharges the storage capacitor through the gas-filled tube (Mims, 1987).

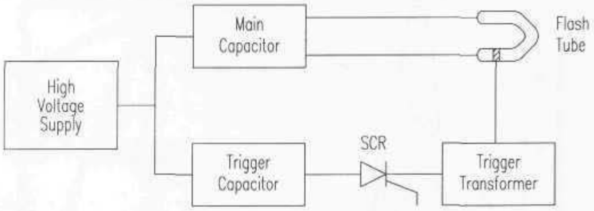

Figure 9-5. Block diagram of a typical xenon flash trigger circuit (adapted from Mims, 1987).

The length of the flash is determined by the resistance of the discharge path and the value of the capacitor as follows (Fox, 1989):

$$T_d = 3RC$$

where:

T_d = duration of flash

R = resistance in ohms
C = capacitance in farads.

Flash duration is typically very short (5 to 200 microseconds), a feature often used to advantage in freezing relative motion in inspection applications (Novini, 1985). The xenon immediately returns to its non-conductive state once the capacitor has discharged, or the flash has been terminated by interrupting the current flow using a solid-state switch (Mims, 1987). The flash can be retriggered as soon as the capacitor has recharged. The average life expectancies of ordinary xenon tubes are in the millions of flashes.

A *light emitting diode* (LED) is a solid-state P-N junction device that emits a small amount of optical energy when biased in the forward direction. LEDs produce *spontaneous emission* with a moderate spectrum (i.e., 40-100 nanometers) about a central wavelength. (As Udd (1991) points out, in sources dominated by *spontaneous emission* there is a low probability of one photon stimulating emission of another photon; such devices have important advantages in fiber-optic sensor applications, including low noise and relative immunity to optical feedback.) LEDs are attractive from the standpoint of durability, shock tolerance, low heat dissipation, small package size, and extremely long life (typically in excess of 100,000 hours).

Due to the spectral response of silicon detectors, near-infrared LEDs are the most efficient, and were the only type offered in photoelectric proximity sensors until around 1975 (Banner, 1993). Green, yellow, red, and blue versions are now readily available (Table 9-1), but near-infrared remains the most popular source in LED-based sensor applications. For example, most optical proximity detectors employ near-infrared LEDs operating between 800 and 900 nanometers. Fast cycle times allow LEDs to be used in a similar fashion to strobed illuminators for freezing motion and other image-synchronization applications, such as frame differencing between successive illuminated and non-illuminated scenes.

Table 9-1. Summary of typical light emitting diode types showing color selectivity (adapted from Fox, 1989).

LED Type	Spectral Peak	Color	Output
Gallium arsenide	540 nm	green	80 ft-lm
	900 nm	IR	10 mw
	980 nm	IR	500 mw
Gallium arsenide phosphide	560 nm	green	300 ft-lm
	610 nm	amber	200 ft-lm
	680 nm	red	450 ft-lm
Gallium aluminum arsenide	800 nm	red	1 mw
	850 nm	IR	5 mw
Silicon carbide	590 nm	yellow	150 ft-lm

Super luminescent diodes (SLDs) are a relatively new development (in the rapidly expanding field of fiber-optic communications and optical-disc technology) that can best be described as midway between the simplistic LED and the more complex coherent laser diode. The construction of all three devices is similar: a forward-biased P-N junction leads to a recombination of excess holes with electrons accompanied by emission of photon energy. While LEDs produce *spontaneous emission* only, laser diodes are physically configured so emissions in the active region oscillate back and forth several times between specially designed front and back facets. A high flux of photons past an excited state results in a high probability the excited state will be "stimulated" to radiate by a passing photon (Udd, 1991). (The principle of *stimulated emission* was first described by Einstein in 1917 and later demonstrated by T.H. Maiman in 1960 (Koper, 1987).) The SLD's characteristic laser "gain" on each forward pass results in a primary wavelength or mode of operation and what is termed a *coherent* output (Dokras, 1987), since each newly created photon has the same phase and wavelength as the stimulating photon (Udd, 1991).

LEDs have no such amplification mechanism; the output intensity simply increases with an increase in current density while the photon flux remains below the threshold required for *stimulated emission* (Udd, 1991). Surface-emitting LEDs have a wide solid-angle output beam with a Lambertian intensity distribution. Edge-emitting LEDs, on the other hand, have a waveguide mechanism built into their structure that results in a narrow Gaussian intensity pattern (Dokras, 1987). A more detailed description of the construction and operation of these two devices is presented by Udd (1991). An SLD is like an edge-emitting LED but with a single-pass gain feature similar to the laser, resulting in a combination of *spontaneous* and *stimulated emission*. The power output is therefore greater than a conventional LED, but as current density is increased, the device is unable to achieve multiple-pass gain as does a laser diode (Dokras, 1987). This limitation is assured by disabling the lasing cavity, typically through: 1) antireflection coating the end facets, 2) using proton bombardment to make one end of the cavity an absorber, or 3) mechanically destroying an end facet (Udd, 1991).

At present, the majority of active optical ranging devices employ laser sources, in equipment based on triangulation, time of flight, phase detection, interferometry, and return signal intensity. Lasers exist in a variety of types: solid-state (Sharp, 1989), liquid (Manzo, 1989), gas (Janney, 1989), and semiconductor (Shurtz, 1989). The more well-known types are gas lasers like helium-neon (HeNe) or the solid-state variety like neodymium: yttrium-aluminum-garnet (Nd:YAG). The recent advent of semiconductor-based laser diodes has had significant impact on the rangefinder instrument community (Depkovich & Wolfe, 1984). Although they typically have reduced power output and poorer spectral quality relative to other lasers, semiconductor devices are compact, rugged, reliable, and efficient, with sufficient quality of performance for most sensing needs. An often used laser of this type is the gallium arsenide (GaAs) laser diode, which emits in the near-infrared region.

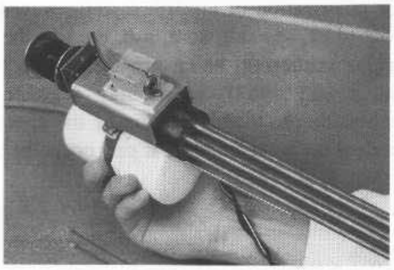

Figure 9-6. A visible-red laser diode is used as a manual sight for the pneumatic dart gun (see Chapter 1) on ROBART III.

This dynamic expansion in usage can be better understood by reviewing some of the inherent qualities of laser light (Depkovich & Wolfe, 1984). Lasers produce an intense well-collimated output, an important property in distinguishing the signal from background illumination, particularly in long-distance applications. The use of energy from the optical portion of the spectrum minimizes the specular reflectance problems encountered with acoustics, with the exception of polished surfaces (Jarvis, 1983). Furthermore, lasers generally transmit spectrally pure light of a single wavelength, void of extraneous signals and noise. This quality can be exploited by placing narrowband optical filters matching the source wavelength in front of the detector component. Filters of this type will reject ambient light, resulting in an improved signal-to-noise ratio for the system.

Along with these advantages there also exist some disadvantages that must be taken into account (Depkovich & Wolfe, 1984). Laser-based systems represent a potential safety problem in that the intense and often invisible beam can be an eye hazard. Gas lasers require high-voltage power supplies that present some danger of electrical shock. Laser sources typically suffer from low overall power efficiency. Lasing materials can be unstable with short lifetimes, resulting in reliability problems. Finally, some laser-based ranging techniques require the use of retroreflective mirrors or prisms at observed points, effectively eliminating selective sensing in unstructured surroundings.

9.1.2 Performance Factors

Atmospheric Attenuation

Attenuation of optical energy due to atmospheric effects can occur in the form of: 1) *scattering* and 2) *absorption*. *Scattering* is a disordered change in the direction of propagation and/or polarization of electromagnetic waves upon encountering molecules and aerosols (dust, sea salt, soot, ash, etc.) in the earth's atmosphere (Figure 9-7). *Molecular scattering*, being proportional to the fourth power of frequency (Feynman, et al., 1963), is strongly wavelength dependent but essentially negligible outside the visible portion of the spectrum; scattering by aerosols is less dependent on wavelength and can occur in both the visible and infrared regions (Herman, et al., 1985).

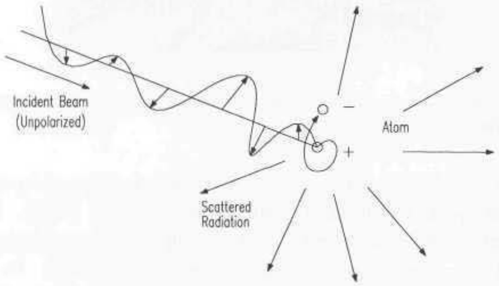

Figure 9-7. In molecular scattering, a beam of incident radiation causes the charges (electrons) in the atom to move; the moving electrons in turn radiate in various directions (Feynman, et al., 1963).

Visibility is a quantitative indicator of atmospheric attenuation. The World Meteorological Organization defines *meteorological optical range* (*MOR*) as the length of a path in the atmosphere required to reduce the luminous flux of a white light of color temperature 2700°K to five percent of its original value (Crosby, 1988). The factor 0.05 (i.e., five percent) is known as the *contrast threshold* ε. The *extinction coefficient* σ is a measure of the reduction of transmitted light due to atmospheric effects, where σ is the summation of the *absorption coefficient* α and the *scattering coefficient* β (Crosby, 1988). The *apparent contrast* C of an object against a uniform horizon sky varies exponentially with the *extinction coefficient* and range, a relationship known as Koschmieder's law:

$$C = e^{-\sigma r}$$

where:

C = apparent contrast

σ = extinction coefficient
r = range to object.

As the *apparent contrast C* approaches the *contrast threshold* ε (0.05), the above equation reduces to (Crosby, 1988):

$$r = \frac{3}{\sigma}$$

where:

r = maximum visibility range (MOR).

Solutions to the actual radiative transfer function describing the effects of atmospheric scattering involve very complex mathematical functions, and so a detailed discussion of atmospheric scattering and absorption is beyond the scope of this chapter. Herman, et al. (1985) provide a comprehensive mathematical treatment of the subject of atmospheric scattering in *The Infrared Handbook* (Wolfe & Zissis, 1985), with appropriate aerosol models and extensive tables of optical constants for various airborne particulates. LaRocca (1985) provides a similar treatment of atmospheric absorption in the same volume.

Surface Reflectivity

As discussed in Chapter 8, the parameter that governs the ratio of reflected versus transmitted sound energy at the interface between two media is the acoustical impedance Z; in the case of optical energy it is the index of refraction n. When light strikes a specular surface as illustrated in Figure 9-8, a portion of the wave is reflected as shown, with the *angle of reflection* θ_r equal to the *angle of incidence* θ_i. The angle θ_t of the *refracted* wave is given by Snell's law.

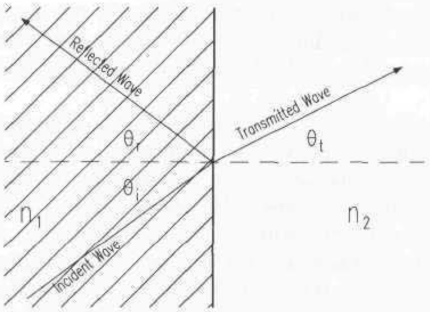

Figure 9-8. For specular surfaces, the angle of reflection is equal to the angle of incidence, while the product $n \sin \theta$ is the same for the incident and refracted beam (Feynman, et al., 1964).

For *specular* surfaces, the intensity of the reflected light depends on the *angle of incidence* and also the *polarization*. For situations where the electric field, ***E***, is perpendicular to the plane of incidence, the *reflection coefficient* is (Feynman, et al., 1964):

$$R_c = \frac{I_r}{I_i} = \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)}.$$

For ***E*** parallel to the plane of incidence, the *reflection coefficient* is (Feynman, et al., 1964):

$$R_c = \frac{I_r}{I_i} = \frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)}.$$

For normal incidence (any polarization):

$$R_c = \frac{I_r}{I_i} = \frac{(n_2 - n_1)}{(n_2 + n_1)}$$

where:

R_c = reflection coefficient
I_r = reflected intensity
I_i = incident intensity.

Due to the shorter wavelengths involved, optical energy is in general less susceptible than microwave or millimeter-wave energy to problems associated with *specular reflection*. As was illustrated in Chapter 8 for the case of a flashlight beam aimed at a mirror, however, specular reflection will occur on highly polished or shiny surfaces. Anyone who has experienced driving at night on a wet road knows firsthand the results in the form of diminished headlight effectiveness. The normally *diffuse* road surface is transformed into a *specular* reflector by the pooling of water in tiny pits and crevices, markedly reducing the surface normal component and consequently the amount of light reflected back to the vehicle.

This last example brings up an important point: the amplitude of a surface reflection is not so much a property of the material, as is the index of refraction, but a "surface property" that depends on the specific topographical characteristics of the target (Feynman, et al., 1964). The above formulas hold only if the change in index is sudden, occurring within a distance of a single wavelength or less. Recall a similar criteria was established in Chapter 8 regarding *specular reflection* for acoustical waveforms. For light, however, the short wavelengths involved dictate that for very smooth surface conditions, the interface transition must occur over a distance corresponding to only a few atoms (Feynman, et al., 1964). For other than highly polished surfaces, the transition happens over a distance of several wavelengths, and the nature of the reflection will be *diffuse* rather than *specular*.

Glass, clear plastic, and other transparent substances with little or no reflectance properties can cause problems. In fact, the unknown reflectivity of observed targets is perhaps the most significant problem in optical range measurement and, coupled with the changing angle of incidence, causes the returned energy to vary significantly in amplitude. As a result, detection capabilities over a wide dynamic range (between 80 to 100 dB) are required, complicating the design of the receiver electronics.

Air Turbulence

Turbulence-induced pressure variations can cause random irregularities in the index of refraction of the atmosphere that will perturb an optical wavefront, causing image motion, image distortion, and added blur (Hufnagel, 1985). As the distorted wavefront continues to propagate, its local irradiance will randomly vary as a consequence of defocusing and spreading effects, a phenomenon known as *scintillation*, an example of which is seen in the twinkling of distant stars. The general consensus is the index of refraction does not vary rapidly in either space or time (Hufnagel, 1985), causing only small-angle refraction and diffraction of the radiation, with no changes in polarization (Saleh, 1967). The effects of turbulence on the index of refraction n are small, however, in comparison to those of temperature, and therefore can usually be ignored (Hufnagel, 1985). This is particularly true in the case of most robotic applications, considering the relatively short distances involved.

Temperature

In addition to wavelength and ambient pressure, the index of refraction is also dependent on temperature and humidity. Of all these parameters, temperature is by far the most significant, influencing the rate of index change as follows (Hufnagel, 1985):

$$\frac{dn}{dT} = \frac{78\,P}{T^2} \times 10^{-6}$$

where:

n = index of refraction
P = local air pressure in millibars
T = absolute temperature in degrees Kelvin.

A condition known as *thermal blooming* occurs when the radiation is strong enough to significantly heat the air along its path, but such effects are generally limited to application of very high-power laser beams (Hufnagel, 1985).

A final consideration should be given to temperature influence on componentry as opposed to the atmosphere itself. The wavelength of a laser diode can be shifted by temperature changes at the P-N junction. This shift in spectral output with temperature is approximately 0.25 nanometers per degree centigrade in the case of gallium-arsenide lasers (Gibson, 1989). A temperature increase will also radically lower the power output of semiconductor emitters, including LEDs and SLDs. Laser diodes can employ tiny closed-loop thermoelectric coolers within the emitter package to compensate for such effects (Dokras, 1987).

Beam Geometry

One of the decided advantages of optical systems relative to acoustical and RF is the ease with which the beam can be focused, using very small and inexpensive lenses. Laser beams in particular are narrow and collimated with little or no divergence, enabling a powerful spatially-selective beam of energy to be concentrated on a small spot at long distances. Typical values for very inexpensive scanning laser systems (see Chapter 6) are in the neighborhood of 0.3 to 0.5 milliradians. Wider beamwidths (i.e., in the neighborhood of 5 degrees) are sometimes employed in proximity sensor (Banner, 1993) and collision avoidance ranging applications (Moser & Everett, 1989) to increase the volumetric coverage and probability of target detection, but at the expense of range.

9.1.3 Choosing an Operating Wavelength

To be used with mobile robotic systems, an optical ranging system must function effectively under normal ambient lighting conditions, which makes the choice of the energy source somewhat critical. Some structured light systems use an incandescent lamp or xenon flash directed through a slit or patterned mask and projected onto the surface. Others use laser beams that are mechanically or electronically scanned at high rates to create the desired illumination. The major criterion for selecting a light source is to be sure that its intensity peaks at a spectral frequency other than that of the ambient energy. The camera (or detector) should be outfitted with a matching narrowband filter to complement the source and improve detection.

For example, ultraviolet light with a wavelength of 0.2 to 0.3 microns is effective outdoors because absorption by atmospheric ozone prevents the transmission of sunlight energy less than 0.3 microns in length. However, an ultraviolet source of the required power density level would be hazardous in indoor environments (not eye safe). Contrast this with infrared energy near 2.8 microns, which is better suited to indoor activities because man-made objects tend to reflect infrared energy well. Infrared loses some of its usefulness outdoors due to the inherent radiation emitted by the natural terrain, roadways, and objects (LeMoigue & Waxman, 1984). There is a strong component of near-infrared energy in sunlight and incandescent light, but ambient noise effects can be reduced by modulating the source over time, then demodulating the received energy at the detector. This technique effectively subtracts the constant illumination of the background.

The design of electro-optical systems must be optimized to extract the necessary information from ambient noise and clutter with a comfortable signal-to-noise ratio. For active systems, the actual source should be chosen according to the following guidelines (Dokras, 1987). It must produce energy:

- With sufficient intensity.
- At the required wavelength (or within an appropriate spectrum).
- With the desired radiation pattern.

9.2 Microwave Radar

The portion of the electromagnetic spectrum considered to be the useful frequency range for practical radar is between 3 and 100 GHz (Miller et al., 1985). Most modern conventional radars use microwave energy and operate in the L, S, C, or X bands (IEEE, 1976). The use of letter designations (see Table 9-2) was adopted for security reasons during World War II and persisted afterwards for sake of convenience. For the most part unaffected by fog, rain, dust, haze, and smoke, radar systems can produce astonishing accuracies in terms of target discrimination and range computation when combined with computerized signal processing (Nowogrodzki, 1983).

Table 9-2. Designated radar frequency bands (IEEE Standard 521-1976) shown in relation to VHF and UHF allocations (Barton, 1989).

Band	Frequency Range	Units
VHF	30-300	MHz
UHF	300-1,000	MHz
P	230-1,000	MHz
L	1,000-2,000	MHz
S	2,000-4000	MHz
C	4,000-8,000	MHz
X	8,000-12,500	MHz
K_u	12.5-18	GHz
K	18-26.5	GHz
K_a	26.5-40	GHz
Millimeter	> 30	GHz

Ranging is accomplished by pulsed TOF methods, CW phase detection, or CW frequency modulation. Pulsed systems can detect targets up to distances on the order of hundreds of miles, relying on the measurement of the round-trip time of a wave propagating at the speed of light. Near-field measurements (less than 100 feet) are more difficult for this type of system because the extremely sharp short-duration signals that must be generated and detected are expensive and complicated to realize, and virtually impossible for distances less than 1 foot (Schultz, 1993). CW systems, on the other hand, are effective at shorter ranges because the phase-detection or frequency-shift measurements are not dependent on the wave velocity and are also well suited for measuring the speed of moving objects by Doppler methods.

9.2.1 Applications

Microwave radars are extensively employed in both military and commercial surveillance, tracking, and navigational applications. Microwaves are ideally suited for long-range sensing because the resolution is generally sufficient, attenuation of the beams in the atmosphere is minimal, and low-mode guiding structures can be constructed. The relatively long wavelengths provide radar systems with an "all weather" capability, overcoming the absorption and scattering effects of air, weather, and other obscurants.

Microwaves are also used for shorter-range sensing needs such as tail-warning radar and ground-control radar for aircraft, typically involving distances in hundreds of feet. Other such uses include tank level indicators (Williams, 1989; Fitch, 1993), traffic control (Schultz, 1993), motion sensors (Hebeisen, 1993), presence detectors (Williams, 1991), and obstacle avoidance radars (Siuru, 1994), operating over ranges from a few feet to a few hundred yards. Microwave systems have been in the developmental stage for quite some time but only came into their prime within the last 20 years or so with the advent of inexpensive, reliable solid-state components as alternatives to the typically fragile, power-consuming thermionic devices (Nowogrodzki, 1983). Equipment for transmitting, receiving, and processing of the waveform is now widely available.

9.2.2 Performance Factors

The basic *radar equation* expresses the relationship between the signal power received at the antenna as a function of antenna size and the emitted power of the system (Ridenour, 1947):

$$S = \left[\frac{PG}{4\pi R^2}\right]\left[\frac{\sigma}{4\pi R^2}\right]\left[\frac{G\lambda^2}{4\pi}\right]$$

where:

S = signal power received
P = transmitted power
G = antenna gain
λ = wavelength
σ = radar cross-section of target
R = range to target.

The quantity in the first set of square brackets represents the power density of the incident wave at the target. For an isotropic antenna distribution radiating energy equally in all directions, the transmitted power P would simply be divided by the area of a sphere ($4\pi R^2$). Since in reality some type of directional antenna is generally employed, the radiated energy is concentrated by the associated antenna gain G.

As was the case for sonar in Chapter 8, we make the assumption here that the incident wave is reflected from the target in an isotropic Lambertian fashion, dispersing yet again in accordance with the inverse square law. This relationship is expressed in the second set of brackets in the radar equation above, where σ (the *radar cross-section*) serves as a combined representation of target *cross-sectional area*, *reflectivity*, and *directivity,* accounting for the fact the reflected energy distribution may not be purely isotropic. (The concept of *radar cross section* will be re-examined in more detail later in the section addressing millimeter-wave radar performance factors.) The first two sets of brackets together therefore give the power density of the returning wave at the receiving antenna, while the last set of brackets represents the cross-sectional area of the receiving antenna.

This most basic form of the radar equation assumes a monostatic configuration where the same antenna is used for both transmission and reception. For a more detailed explanation of terms and treatment of bistatic (separate antennae) configurations, see Blake (1990). In addition, any losses due to atmospheric scattering or absorption, which can be quite significant, are not yet taken into account. (This atmospheric attenuation aspect of radar performance will be treated in some detail later in this chapter.) Otherwise, however, the above equation very closely parallels the expression presented in Chapter 8 for the intensity of a reflected acoustical wave, with signal strength falling off (for both radar and sonar) with the fourth power of distance.

A major component consideration in the implementation of radar ranging capability is the configuration of the transmitting and receiving antennae. Conventional long-range monostatic systems typically feature a large *parabolic reflector* with the detector (or feed) positioned at the focal point of the dish (Figure 9-9A). The relationship between transmitted beamwidth and antenna diameter is expressed in the following equation:

$$\theta = 1.22\frac{\lambda}{d}$$

where:

θ = beamwidth
λ = wavelength
d = diameter of the reflector.

Increasing the diameter of the reflector results in improved range capability due to the more powerful (tighter focused) outgoing beam, and the larger antenna surface area with which to intercept the reflected energy (i.e., the gain parameters in the first and third brackets of the radar equation). Disadvantages include the need to manipulate a bulky mechanical system with high inertial loading, with a massive supporting structure required to offset the effects of vibration and wind. These factors have been major impediments to deployment of conventional radar-based ranging systems on board mobile robots.

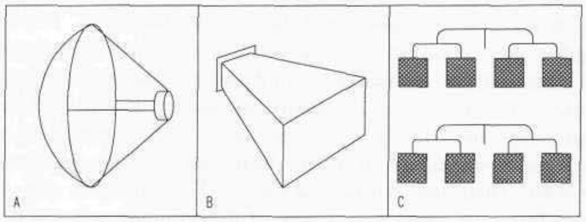

Figure 9-9. Common configurations of microwave antennae include: (A) reflective dish with feed situated at focal point, (B) conventional horn, and (C) two-dimensional microstrip arrays (adapted from Hebeisen, 1993).

To offset some of these drawbacks, many short-range commercial applications use a *horn* antenna configuration. Contrary to popular assumption, the dimensions of the horn opening are inversely proportional to beamwidth. In other words, the longer dimension of the horn results in the more narrow beam dimension, and vice versa. The antenna orientation depicted in Figure 9-9B, for example, would result in a rather broad vertical beam cross-section with a more narrow horizontal profile. Horn size tends to be reasonably small for beamwidths greater than 20 degrees, but fairly unrealistic in applications requiring less than 15 degrees (Schultz, 1993).

Phased-array antenna configurations (Figure 9-9C) present an alternative arrangement in the form of an array of multiple small antennae separated by distances of a few wavelengths. The transmissions from each radiator diverge and overlap with neighboring transmissions in a constructive and destructive fashion based on their phase relationships. By properly adjusting the phases, the antenna array can be tuned to a desired direction and intensity, as well as electronically scanned across the field of view. The small size of the individual transmitter-receivers facilitates low-profile designs and reduces problems due to wind and vibration. The resulting smaller coverage area decreases overall effectiveness, however, while the requirement for electronically variable phase control increases overall system complexity. Flat-plate dipoles and microstrip patch radiators are especially popular in applications involving ranges of 200 feet or less (Schultz, 1993).

Effective detection range is influenced by a number of factors besides antenna design and system performance, one of the more significant being atmospheric attenuation. Under fair weather conditions, clear atmosphere progresses from completely transparent all the way to 100-percent opaque as frequency is increased from 1 to 1000 GHz (Falcone & Abreu, 1979). Rain and snow can cause significant attenuation in signals above 2 GHz, as the physical size of the droplet becomes comparable to the wavelength (Hebeisen, 1989). Other issues for consideration include backscatter from airborne moisture, ground multipath interference, background clutter, reflectivity and directivity of the target surface, and natural cover such as snow or foliage.

Relatively speaking, there are a number of disadvantages associated with microwave radars from a robotics perspective:

- Microwave energy, like ultrasonic, is susceptible to specular reflections at the target surface, requiring receivers and signal processors with wide dynamic ranges.
- Available bandwidth is lower than shorter-wavelength millimeter-wave and electro-optical systems, resulting in reduced resolution.
- Longer wavelengths translate into large and bulky system components.
- Large antenna sizes are required to get narrow directional beams.
- The wide beamwidths result in problems due to side lobes, backscatter, and multipath reflections.
- Less available bandwidth, longer range capability, and wide beamwidth collectively make covert undetected operations difficult, with increased susceptibility to intercept and jamming.

Significant environmental constraints on microwave radar were identified by Barton (1977) and summarized by Johnston (1979). For all the above reasons, conventional radar systems operating in the microwave portion of the energy spectrum have less applicability to the short-range collision-avoidance needs of a mobile robotic platform.

9.3 Millimeter-Wave Radar

Millimeter waves constitute that portion of the electromagnetic spectrum with wavelengths of about 500 micrometers to 1 centimeter (30 to 300 GHz), midway between microwave and electro-optic. Millimeter waves possess several properties which differ substantially from microwave radiation. Relative to microwaves, millimeter-wave systems have significantly less range capability, primarily due to atmospheric attenuation and backscatter, but this generalized observation is given with regard to the more conventional applications of air-search and tracking radars. For the more unorthodox shorter-range needs of a mobile robotic system, this characteristic is not necessarily a limiting disadvantage.

In fact, Strom (1977) points out that for short-range applications, the performance of millimeter wave is actually superior to microwave under poor weather conditions:

> "If we compare a 3-mm radar and a 3-cm radar which are designed for short range usage and are equivalent in all other respects, we note that the equivalence in aperture provides 20 dB more gain at the short wavelength to offset atmospheric attenuation, and the equivalence in signal bandwidth (when considered as a percentage of operating frequency) yields a tenfold improvement in range resolution. Thus, the combined effect of antenna gain and range resolution is to reduce the clutter cell size by 30 dB."

The shorter wavelengths result in a narrow beam, with relatively small-sized antenna apertures for a given bandwidth. While the overall physical size of the system is reduced, the smaller apertures mean less collected energy, which again limits the effective range of the system.

Shorter wavelengths yield more accurate range and Doppler measurements. The ratio of wavelength to target size improves for higher frequencies, enabling better detection of small objects such as poles, wires, trees, and road signs. The high Doppler frequencies (conveniently in the audio range) provide good detection of slowly moving targets (Richard, 1976). The low power requirements of shorter wavelengths are attractive for robotic applications. In addition, shorter wavelengths translate into smaller componentry, which leads to reduced size and weight; the tradeoffs are that atmospheric attenuation increases, and the smaller antenna sizes lead to reduced receiver sensitivities.

Another advantageous feature is the extremely large bandwidth associated with millimeter waves; for the four major bands (35, 94, 140, and 220 GHz), the associated bandwidths are 16, 23, 26, and 70 GHz respectively. This means the entire microwave region could fit into any one of the millimeter bands (Sundaram, 1979). A bandwidth of one percent at 300 GHz is equal to all frequencies below S-band center frequency at 3 GHz (Johnston, 1978). More bandwidth translates into reduced multipath, greater resolution and sensitivity, reduced interference between mutual users of the band, and greater resistance to jamming.

9.3.1 Applications

Millimeter waves have been proposed for numerous applications (Skolnik, 1970) and have been the subject of theoretical studies (and much debate) since the early 1950s (Johnston, 1978). Unfortunately, the technology was not sufficiently developed during the earlier part of this period, and it was not until recent solid-state advances in the past two decades that practical devices could be developed and tested. Likely beneficiaries of this technology include remote environmental sensing, interference-free communications, low-angle tracking radar, high-resolution and imaging radar, spectroscopy (Senitzky & Oliner, 1970), terminal guidance, active fusing, range finding equipment, and automobile braking. Of these, the most common usage today is target tracking and designation in the military. The narrow beamwidth of millimeter-wave transmissions is highly immune to ground reflection problems when following targets at low-elevation angles, making such radars highly effective at low-angle tracking.

In addition to their limited range capabilities, millimeter-wave systems are not used for large-volume search and target acquisition because of their narrow beamwidths. Anyone who has ever tried to find a specific object of interest with a very high-power telescope is familiar with this problem. Some such optical devices have wider-angle "spotter scopes" attached on a parallel axis to assist in first locating the target. Certain military ground-to-air defense systems have adopted the same approach, employing longer-range wide-beam microwave search radars for initial acquisition, and then switching to a millimeter tracking radar for weapons system control.

Short-range, low-power millimeter-wave systems would appear to be well suited to the collision avoidance and navigational ranging needs of an exterior mobile robot. In fact, Johnston (1978) suggested automobile braking as being the largest potential commercial application of millimeter-wave radar. As Weisberg (1977) pointed out, the ability to use smaller antennae is a dominating characteristic influencing selection of millimeter wave over microwave. He cited three platforms most directly affected: satellites, missiles, and mini-RPVs. The envisioned application of millimeter ranging systems on robotic vehicles can be considered an extension of Weisberg's third category.

9.3.2 Performance Factors

Atmospheric Attenuation

All electromagnetic energy is absorbed to some degree by atmospheric gases (water, carbon dioxide, oxygen, and ozone), and from attenuation due to haze, fog, clouds, and rain (Sundaram, 1979). The absorbed energy is converted into heat and then lost to the surrounding atmosphere (Hebeisen, 1989). Frequency selective absorption takes place at the higher frequencies, due to resonances of the atmospheric gases (Van Vleck, 1964), and varies with atmospheric pressure, temperature, and relative humidity (Dyer & Currie, 1978). Two most pronounced effects are due to the magnetic interaction of oxygen and the electric polarity of the water molecule in water vapor (Koester, et al., 1976). These regions of maximum absorption are denoted as *absorption bands*, with the most significant being around 60 GHz (Sundaram, 1979). Similarly, atmospheric absorption by gases is minimal for certain frequencies (Table 9-3), appropriately called *atmospheric windows* (Sundaram, 1979).

Table 9-3. Naturally occurring atmospheric windows (adapted from Sundaram, 1979).

Window	Favored Wavelengths
Visible	0.4 to 0.7 µm
IR	3 to 5 µm, 8 to 12 µm
RF	3 and 10 cm (main microwave bands)
MMW	8.5, 3.2, 2.1, 1.4 mm (35, 94, 140, 220 GHz)

Relative to microwaves, millimeter waves display greater interaction with the environment. This attribute is a bonus in radiometry applications in that sensors can detect small particles and carry on frequency-selective interaction with gases. However, the resulting atmospheric attenuation limits the maximum range and restricts operation of such devices in adverse weather conditions. Millimeter waves interact with atmospheric gases, particulate matter, and suspended moisture droplets (hydrometers) through three primary mechanisms: absorption, scattering, and refraction (Dyer, et al., 1978). Attenuation can occur due to absorption by water vapor and gases, absorption from condensed water droplets in fog and rain, and scattering from water droplets in rain (Richard, 1976). Of these, rain is by far the most significant factor, with its associated attenuation being directly proportional to the size of the water droplets and the rainfall rate (Hebeisen, 1989), and to the 2.5 power of frequency (Barton, 1989). Dry snow, on the other hand (with the exception of very heavy snowfall rates), produces very little attenuation (Dyer & Currie, 1978). Because airborne particulates such as dust, smog, and smoke have dielectric constants much smaller than that of water, their associated attenuation cross section is reduced, resulting in negligible millimeter wave attenuation (Dyer & Currie, 1978).

Richard (1976) characterizes clear weather attenuation as small for 35, 70, and 94 GHz, appreciable at 140 and 240 GHz and prohibitively large at 360 GHz and above. Fair weather ranges are generally restricted to 10 to 20 kilometers (Sundaram, 1979), and typical employment with comparative all-weather performance to 3-cm microwave is more likely limited to between 5 and 10 kilometers (Strom, 1977). Johnston (1979) provides an excellent overview of early work addressing the various factors affecting millimeter-wave propagation, to include attenuation, backscatter, foliage penetration, phase variations, polarization effects, and surface phenomena. Another comprehensive survey of millimeter-wave propagation data treating attenuation, backscatter, and foliage is presented by Dyer and Currie (1978). Clear weather atmospheric as well as calculated rain attenuation curves versus frequency of operation are presented by Richard (1976) and numerous others.

In summary, atmospheric losses in the four millimeter bands, while higher than those for microwave energy, are significantly lower than that associated with optical devices, particularly in the case of rain, fog, smoke, dust, and haze. For the robotic applications envisioned (collision avoidance, navigation, communications), the range limitations impose no significant constraints.

Radar Cross-Section

Radar cross-section, a term used to quantify the ability of a particular target to return energy to the receiving antenna, is a function of three independent parameters (Schultz, 1993):

- *Geometric cross-section* — refers to the amount of incident radiation intercepted by the target's cross-sectional area and is basically influenced by target size and orientation (aspect ratio).
- *Reflectivity* — determines that portion of the incident energy reflected versus absorbed by the target and is primarily influenced by target composition.
- *Directivity* — is a measure of the reflected energy returning to the receiving antenna and is mainly influenced by target geometry.

The *coefficient of reflection* introduced earlier in this chapter in the particular case of optical energy applies to radar reflectivity as well. Recall for normal incidence:

$$R = \frac{I_r}{I_i} = \frac{(n_2 - n_1)}{(n_2 + n_1)}$$

where:

R = reflection coefficient
I_r = reflected intensity
I_i = incident intensity.

The actual value of R depends on target surface conditions, material composition, and the frequency of the incident wave, but in general is greater for lossy materials possessing: 1) higher permittivity (dielectric constant), 2) higher conductivity, and 3) lower permeability (Hebeisen, 1989). In other words, lossless dielectric materials such as plastic and glass reflect little energy, whereas conductive materials such as aluminum or steel make good radar targets. The more magnetic the conductive material, the less its reflectivity, and so one would expect aluminum to reflect more energy than steel. Typical values of permittivity, conductivity, and permeability for several representative materials are listed in Table 9-4.

Table 9-4. Typical values of relative permittivity, conductivity (s/m), and relative permeability for common materials, for low frequencies at room temperature (adapted from Hebeisen, 1989).

Material	Relative Permittivity	Conductivity	Relative Permeability
Air	1.0	-	-
Bakelite	5.0	-	-
Glass	4-10	-	-
Oil	2.3	10^{-11}	-
Dry soil	3-4	10^{-5}	-
Distilled water	80	2×10^{-4}	-
Silver	-	6.17×10^{7}	0.99998
Copper	-	5.80×10^{7}	0.99999
Aluminum	-	3.54×10^{7}	1.000021
Iron	-	10^{7}	4,000

To get an intuitive appreciation for the values listed in Table 9-4 above, try heating a small glass container of baby oil in a microwave oven for a set time interval, and then heat the same amount of water for an identical length of time. The temperature of the oil shows negligible increase, whereas the container of water can easily be brought to a boil. Water, being a better conductor than oil, absorbs more of the RF energy, which is then converted into heat. Non-conductive materials that have a high moisture content, therefore, make relatively good targets (Figure 9-10).

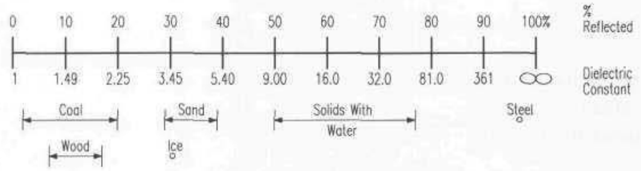

Figure 9-10. The percentage of reflection of microwave energy off various solids, showing the relative reflectivity for high-moisture-content solids in relationship to steel (adapted from Hebeisen, 1993).

All other factors being equal, a general rule of thumb is the radar cross-section of a particular object will increase with size and also with the frequency of incident radiation, up until the point where the object's size is much larger than the operating wavelength, after which little additional improvement is seen (Hebeisen, 1989). This relationship is the reason multipath reflections in RF communications are so much more of a problem at higher frequencies; the radiated energy is more likely to reflect from the various objects in the immediate environment. The longer the wavelength (i.e., the lower the frequency), the easier it is for the energy to penetrate through surrounding structure for non-line-of-sight operation.

Multipath Reflections

In addition to the desired signal return from discrete targets, multipath reflections from intermediate surfaces must also be taken into account in ground-based scenarios. (Multipath reflections are generally not a problem in surface-to-air radar tracking applications.) When a horizontal beam is projected from a mobile robotic platform, there is a good chance the lower part of the beam will illuminate a portion of the ground as well as the target of interest (Figure 9-11). If the ground-reflected signal is not suppressed by the directional characteristics of the receiving antenna, this leads to constructive and destructive interference, causing fades. Ground multipath and clutter, however, do not usually limit the performance of millimeter ground-to-ground systems as seriously as foliage obscuration, background clutter, and terrain masking (Richard, 1976).

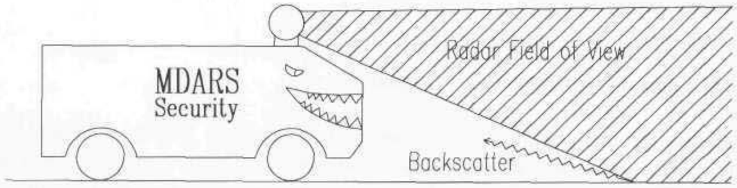

Figure 9-11. Ground backscatter at low elevation angles can be a problem in robotic collision-avoidance radar applications.

When scattered energy is returned in the direction of the receiver (i.e., backscatter), it can seriously contribute to the noise level, sometimes lowering the signal-to-noise ratio to the point target detection is difficult or impossible. Backscatter effects from rain are dependent upon the rainfall rate and the frequency of operation, and appear most significant in the region of 20 to 40 GHz, which includes the 35-GHz K_a band of frequencies (Richard, 1976). Backscatter from fog and clouds is more than two orders of magnitude less than that associated with rain (Lo, et al., 1975). In addition to causing backscatter along the path, an accumulation of snow on the reflecting surface can effectively mask the target altogether. The best way to reduce backscatter is to limit the size of the resolution cell (i.e., the beamwidth). Other techniques include narrow-beam antennae, short pulse durations for TOF systems (Richard, 1976), wide-band frequency modulation (McGee, 1968), frequency optimization, and circular polarization. The use of circular polarization, for example, can reduce rain return by 10 dB or more (Strom, 1977).

Temperature

The speed of light, quite unlike the speed of sound, is not significantly influenced by ambient temperature variations. Temperature differentials and humidity can, however, generate time-varying refractions, producing beam wander that results in short-term fades (Strom, 1977). This topic was previously treated in Section 9.1.2.

Beam Geometry

The shorter wavelength of millimeter-wave radiation results in narrower beamwidths relative to microwave. By way of illustration, a millimeter-wave antenna with a 12-cm-diameter aperture provides a 1.8-degree beamwidth at 94 GHz, as opposed to 18 degrees at 10 GHz (Sundaram, 1979). Narrow beams mean increased range, greater angular resolution, reduced noise and interference, minimized side lobe returns, fewer multipath problems, and less chance of detection and jamming.

9.3.3 Choosing an Operating Frequency

Tradeoffs in the selection of an operating frequency involve antenna size, atmospheric attenuation properties, and available component technology (Koester, et al., 1976). Strom (1977) provides a very comprehensive overview of frequency optimization that takes into account the effects of attenuation, clutter, and system parameters. These factors must be collectively considered from a systems integration point of view when establishing design parameters.

For example, due to narrower beamwidth, side lobes and ground multipath effects are much reduced for 220 GHz. Since operation at that frequency band would also entail the smallest size components, it might seem like the logical choice for most applications. However, Weisberg (1977) points out that on a hot humid day, the required power for a 3-kilometer target acquisition system jumps from a fraction of a watt at 94 GHz to tens of kilowatts if operated at 220 GHz.

9.4 References

Banner, *Handbook of Photoelectric Sensing*, Banner Engineering Corp., Minneapolis, MN, 1993.

Barton, D.K., “Philosophy of Radar Design,” in *Radar Technology*, E. Brookner, ed., ARTECH House Books, Dedham, MA, 1977.

Barton, D.K., "Radar Principles," in *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, New York, McGraw Hill, pp. 25.2-25.53, 1989.

Blake, L., "Prediction of Radar Range," in *The Radar Handbook*, Chapt. 2, Skolnik, M., ed., 2^{nd} edition, McGraw Hill, New York, NY, 1990.

Buschling, R., "Understanding and Applying IR Temperature Sensors," *Sensors*, pp. 32-37, October, 1994.

Crosby, J.D., "Visibility Measurement: An Assessment of Two Techniques," *Sensors*, pp. 32-40, October, 1988.

Depkovich, T., Wolfe, W., "Definition of Requirements and Components for a Robotic Locating System," Final Report No. MCR- 83-669, Martin Marietta Denver Aerospace, Denver, CO, February, 1984.

Dokras, S., "Active Components in Fiber-Optic Sensors," *Sensors*, pp. 20-23, April, 1987.

Dyer, F.B., Currie, N.C., "Environmental effects on Millimeter Radar Performance," AGARD Conference Proceedings, CP 245, *Millimeter and Submillimeter Wave Propagation and Circuits*, pp. 2.1 - 2.9, 1978.

Falcone, V.J., Abreu, L.W., "Atmospheric Attenuation of Millimeter and Submillimeter Waves," IEEE EASCON-79 Conference Record, Vol. 1, pp. 36-41, 1979.

Feynman, R.P., Leighton, R.B., Sands, M., *The Feynman Lectures on Physics*, Vol. 1, Addison-Wesley, Reading, MA, 1963.

Feynman, R.P., Leighton, R.B., Sands, M., *The Feynman Lectures on Physics*, Vol. 2, Addison-Wesley, Reading, MA, 1964.

Fitch, F.M., "Measuring Level with Radar Technology," *Sensors*, pp. 40-41, April, 1993.

Fox, C.S. "Lamps, Luminous Tubes, and Other Noncoherent Electric Radiation Sources," *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, New York, McGraw Hill, pp. 11.4-11.11, 1989.

Gibson, S.B., "Application of Semiconductor Lasers," *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, New York, McGraw Hill, pp. 11.37-11.41, 1989.

Graf, R.F., *Dictionary of Electronics*, Howard W. Sams, Indianapolis, IN, 1974.

Halliday, D., Resnick, R., *Fundamentals of Physics*, John Wiley, New York, NY, 1974.

Hebeisen, S., "Target and Environmental Characteristics Which Affect the Performance of Microwave Sensing Systems," Sensors Expo International, Cleveland, OH, September, 1989.

Hebeisen, S., "Microwave Proximity Sensing," *Sensors*, pp. 22-27, June, 1993.

Herman, B., LaRocca, A.J., Turner, R.E., "Atmospheric Scattering," in *The Infrared Handbook*, Wolfe, W.L., Zissis, G.J., eds., pp. 5.1-5.131, 1985.

Hufnagel, R.E., "Propagation Through Atmospheric Turbulence," in *The Infrared Handbook*, Wolfe, W.L., Zissis, G.J., eds., pp. 6.1-6.56, 1985.

IEEE, "IEEE Standard Letter Designations for Radar Bands," IEEE Standard 521-1976, November 30, 1976.

IEEE, "IEEE Standard Radar Definitions," IEEE Standard 686-1977, November, 1977.

Janny, G.M., "Gas Lasers," *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, New York, McGraw Hill, pp. 11.27-11.31, 1989.

Jarvis, R.A., "A Laser Time-of-Flight Range Scanner for Robotic Vision," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-5, No. 5, pp. 505-512, September, 1983.

Johnston, S.L., "Some Aspects of Millimeter Wave Radar," Proceedings International Conference on Radar, Paris, France, December 4-8, pp. 148-159, 1978.

Johnston, S.L., "A Radar Engineer Looks at Current Millimeter- Submillimeter Atmospheric Propagation Data," IEEE EASCON-79 Conference Record, Vol. 1, pp. 27-35, 1979.

Koester, K.L., Kosowsky, L., and Sparacio, J.F., "Millimeter Wave Propagation," Appendix A to "Millimeter Wave Radar Applications to Weapons Systems," V.L. Richards, pp. 77-105, June, 1976.

Koper, J.G., "A Three-Axis Ring Laser Gyroscope," *Sensors*, pp. 8-21, March, 1987.

LaRocca, A.J., "Atmospheric Absorption," in *The Infrared Handbook*, Wolfe, W.L., Zissis, G.J., eds., pp. 5.1-5.131, 1985.

Le Moigue, J., Waxman, A.M., "Projected Light Grids for Short Range Navigation of Autonomous Robots," Proceedings of 7th IEEE Conference on Pattern Recognition, Montreal, Canada, pp. 203-206, July - August, 1984.

Lo, L.T., Fannin, B.M., Straiton, A.W., "Attenuation of 8.6 and 3.2 mm Radio Waves by Clouds," *IEEE Transactions on Antennae and Propagation*, Vol. AP-23, No. 6, November, 1975.

Manzo, P.R., "Liquid Lasers," *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, New York, McGraw Hill, pp. 11.25-11.27, 1989.

McGee, R., "Multipath Suppression by Swept Frequency Methods," Ballistic Research Laboratories Memorandum Report No 1950, November, 1968.

Miller, D.L. et al., "Advanced Military Robotics," Interim Report No. R84-48603-001, Martin Marietta Denver Aerospace, Denver CO, 26 July, 1985.

Mims, F.M., Forrest Mims' Circuit Scrapbook II, Howard W. Sams, Indianapolis, IN, p. 95, 1987.

Moser, J., Everett, H.R., "Wide-Angle Active Optical Triangulation Ranging System," SPIE Vol. 1195, Mobile Robots IV, Philadelphia, PA, November, 1989.

Novini, A., "Fundamentals of Machine Vision Lighting," Penn Video, December, 1985.

Nowogrodski, M., "Microwave CW Radars in Industrial Applications," *Electro, 1983 Conference Record*, Vol. 8, pp. 1-7, 1983.

Richard, V.W., "Millimeter Wave Radar Applications to Weapons Systems," U.S. Army Ballistic Research Labs Report No. 2631, June, 1976.

Ridenour, L.N., *Radar Systems Engineering*, MIT Radiation Laboratory Series, McGraw Hill, pp. 143-147, 1947.

Saleh, A.A.M., "An Investigation of Laser Wave Depolarization Due to Atmospheric Transmission," *IEEE Journal of Quantum Electronics*, Vol. QE-3, pp. 540-543, November, 1967.

Schultz, W., "Traffic and Vehicular Control Using Microwave Sensors," *Sensors*, pp. 34-49, October, 1993.

Senitzky, B., Oliner, A.A., "Submillimeter Waves - A Transition Region," *Submillimeter Waves*, Microwave Research Institute Symposia Series, Fox, J., ed., Polytechnic Press of the Polytechnic Institute of Brooklyn, NY, 31 March - 2 April, 1970.

Sharp, E.J., "Solid Optically Pumped Lasers," *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, New York, McGraw Hill, pp. 11.18-11.24, 1989.

Shurtz, R.R., "Semiconductor Lasers and LEDs," *Electronic Engineer's Handbook,* D. Christiansen and D. Fink, eds., 3rd edition, New York, McGraw Hill, pp. 11.31-11.36, 1989.

Siuru, B., "The Smart Vehicles Are Here," *Popular Electronics*, Vol. 11, No. 1, pp. 41-45, January, 1994.

Skolnik, M.I., "Millimeter and Submillimeter Wave Applications," *Submillimeter Waves*, Microwave Research Institute Symposia Series, Fox, J., ed., Polytechnic Press of the Polytechnic Institute of Brooklyn, New York, pp. 9-26, March 31 - April 2, 1970.

Strom, L.D., "The Unresolved Problems and Issues," 6th DARPA Tri-Service Millimeter Wave Conference, pp. 10-27, 1977.

Sundaram, G.S., "Millimetre Waves - The Much Awaited Technology Breakthrough?", *International Defense Review*, part of the Jane's Information Group, Vol. 11, No. 2, February, 1979.

Udd, E., "Fiber Optic Sensors Based on the Sagnac Interferometer and Passive Ring Resonator," in *Fiber Optic Sensors: An Introduction for Engineers and Scientists*, E. Udd, ed., John Wiley, New York, pp. 233-269, 1991.

Van Vleck, J.H., "Theory of Absorption by Uncondensed Gases," in D.E. Kerr (ed.), *Propagation of Short Radio Waves*, Boston Technical Publishers, pp. 646-664, MA, 1964.

Weisberg, L.R., "Millimeter Waves - The Coming of Age," 6th DARPA Tri-Service Millimeter Wave Conference, pp. 4-9, 1977.

Williams, H.S., "Proximity Sensing with Microwave Technology," *Sensors*, pp. 6-15, June, 1989.

Williams, H.S., "The Basic Principles of Microwave Sensing," *Sensors*, pp. 26-28, May, 1991.

Wolfe, W.L., Zissis, G.J., eds., *The Infrared Handbook*, Environmental Research Institute of Michigan, Ann Arbor, MI, 1985.

10

Collision Avoidance

Truly autonomous control implies the ability of a free-roaming platform to travel anywhere so desired, subject to nominal considerations of terrain traversibility. Many potential applications await an indoor mobile robot that could move in a purposeful fashion from room to room, with the intelligence to avoid objects and if necessary choose alternative routes of its own planning. *Navigational referencing sensors* typically require high angular and/or range resolution over fairly long distances and will be extensively treated in several follow-on chapters. *Collision avoidance sensors*, on the other hand, usually operate over shorter ranges with less resolution required. The field of view should provide sufficient coverage for a turning vehicle, and allow enough time to stop or alter course.

The various proximity and non-contact ranging techniques presented in Chapters 3 through 9 all represent potential candidate solutions for meeting the collision avoidance needs of a moving platform. We shall examine in this chapter some of the issues involved in incorporating such sensors on indoor robotic systems, and then consider a few special problems that arise in more complex outdoor scenarios.

10.1 Navigational Control Strategies

A number of different navigational control strategies have been adopted by various parties, and lumping them together here under a common heading is a bit like mixing apples and oranges. Accordingly, it is perhaps advantageous to first make a distinction between high-level *global navigation* (i.e., planning an optimal path to some desired goal location in world coordinates), and *localized navigation* (i.e., piloting the robot around unexpected obstructions). We shall address only the latter category in this chapter, from the dual perspective of: 1) the required sensors and 2) the interpretation of data collected by those sensors for purposes of collision avoidance.

Six general approaches for such localized collision avoidance will be discussed:

- "Wander" routine.
- Circumnavigation.
- Potential fields.
- Certainty grids.
- Motor schema.
- Vector field histogram.

The above candidates can be subdivided into three basic categories: 1) *reactive control*, 2) *representational world modeling*, and 3) some combination of both.

10.1.1 Reactive Control

Reactive control for our purposes refers to a behavior-based strategy that directly couples real-time sensory information to motor actions, without the use of intervening symbolic representations that attempt to model in an absolute sense all or part of the robot's operating environment. Arkin (1992a) lists the following general characteristics of reactive control:

- It is typically manifested by a decomposition into primitive behaviors.
- Global representations are avoided.
- Sensor decoupling is preferred over sensor fusion.
- It is well suited for dynamically changing environments.

Wander

The most simplistic *reactive* collision avoidance capability for an autonomous mobile robot is perhaps illustrated by the basic *wander routine* implemented by several research groups (Everett, 1982; Brooks, 1986; Arkin, 1987; Anderson & Donath, 1988). The term *wander* is used here to describe a behavioral primitive that involves traveling more or less in a straight line until an obstacle is encountered, altering course to avoid impact, then resuming straight-line motion. Such a capability can be simply hard-coded, rule-based, or inherent in a more sophisticated layered *subsumption architecture* (Brooks, 1986).

By way of illustration, the *wander routine* employed on ROBART I was based on a six-level scheme of proximity and impact detection using the following sensor inputs (see Figure 10-1):

- A positionable near-infrared proximity scanner mounted on the head.
- A forward-looking LM-1812 sonar mounted 20 inches above the floor.
- Ten near-infrared proximity detectors to sense close (< 18 inches) obstructions.
- Projecting "cat-whisker" tactile sensors to detect pending (< 6 inches) collisions.
- Contact bumpers to detect actual impact.

- Drive motor current sensors to monitor for overload condition indicative of a stall.

The first two categories were loosely classified as non-contact ranging sensors that looked out ahead of the robot for planning purposes, while the next three were considered close-in proximity and tactile sensors requiring immediate action. Drive motor overload was a last-resort internal sensor in the event contact with an object was for whatever reason not detected by any of the above.

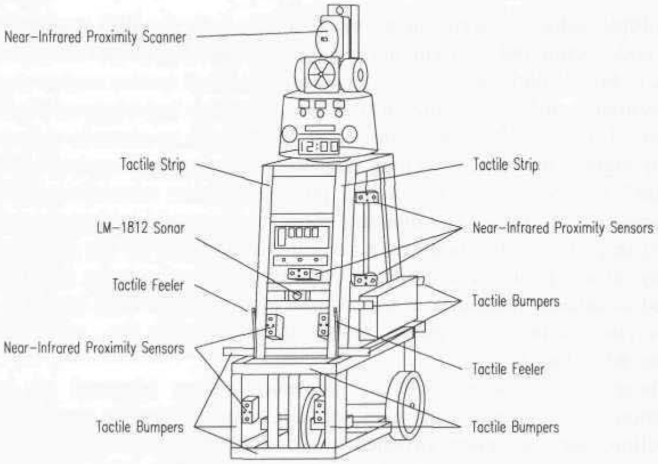

Figure 10-1. Location of collision avoidance sensors employed on ROBART I.

In some ways the software implementation was similar to Brooks' subsumption approach (1986), in that it was a bottom-up design with two distinctly separate hierarchical layers: 1) a *low-level interrupt-driven layer* and 2) an *intermediate-level polling layer* in the main program loop. This layering was basically just an algorithmic differentiation of software categories running on a single processor, however, and limited in actual embodiment to only two layers, although a future higher-level expansion was suggested (Everett, 1982). Brooks, on the other hand, developed a much more powerful and versatile *subsumption architecture,* wherein multiple layers are implemented as additional finite-state machines to support progressively intelligent control.

Those sensors monitoring ROBART's close-in environment (i.e., proximity detectors, feeler probes, bumpers, drive current overload) were considered high priority and consequently read by a maskable interrupt request (IRQ) routine. Unless deactivated by the main program loop, the IRQ routine continuously monitored the sensor output states, and would redirect the motion of the robot in accordance with preprogrammed responses specifically tailored to the individual sensors in alarm. By way of illustration, a preprogrammed response for a right-front bumper impact would consist of the following steps:

- Stop all forward travel.
- Turn steering full right.
- Back up for *x* number of seconds while monitoring rear bumper.
- Stop and center steering.
- Resume forward travel.

Multiple behaviors were incorporated within both the IRQ routines and the main code, arbitrated in accordance with preassigned priorities in the event of conflict. For example, the collision avoidance interrupt service routine would poll all potential inputs to determine which specific device had triggered the interrupt request. Those inputs representing actual impact with an obstacle were naturally ranked higher in terms of polling sequence than inputs associated with "cat-whisker" probes, which in turn had precedence over near-infrared proximity detectors, and so forth. The interrupt service code would initiate the associated canned response for the first active condition discovered by the polling routine, thereby ensuring the higher concern situations received priority attention. The canned avoidance responses in turn would also poll the other inputs to ensure appropriate reaction as depicted in the above example (i.e., monitoring rear bumper while backing).

Whenever a close-in collision avoidance sensor triggered an interrupt, execution of the intermediate level software was temporarily suspended as the controlling microprocessor switched to the interrupt service routine. The low-level avoidance maneuvers would then dominate until the robot was clear of the detected obstruction (quite unlike the subsumption architecture approach). The intermediate level software, on the other hand, had the ability to disable IRQ interrupts associated with collision avoidance sensors, or otherwise suppress or inhibit the lower level behaviors.

A typical example involving suppression as well as complete disabling is illustrated for the case where the robot is homing in on its battery recharging station. The intermediate-level docking software would set a flag that the low-level interrupt service routine always checked before executing canned avoidance maneuvers. If a potential obstacle were encountered during a docking procedure, the normal avoidance response would be suppressed. The robot would instead back up a short distance, then move forward for a predetermined interval at fixed offset from the charger. This action effectively caused the robot to travel along a circular arc to a new relative position before reattempting to dock, thus clearing the intervening obstruction (Everett, 1982). Obstacles are still avoided in this fashion, but without losing sight of the higher-level goal to recharge. Once the robot closed to within a predetermined distance from the charger, all collision avoidance behavior is deactivated to permit physical contact with the charger contacts.

To facilitate somewhat more intelligent movement than the purely reactionary "bump-and-recover" IRQ routines, the intermediate-level software would repeatedly poll the sonar and head-mounted near-infrared scanner on each pass through the main loop. These longer-range sensors were tasked with monitoring the area in front of the robot out to a distance of approximately 5 feet and storing a suitable representation of detected targets in a relative world model as illustrated in Figure 10-2. The *wander algorithm* reacted to the information in the model by choosing the least obstructed direction for continued transit in the event the forward path became blocked. Since all zones were equally weighted in a binary fashion (i.e., either blocked or clear), the least obstructed direction was taken to be that opening defined by the largest number of adjacent clear zones. The inherent simplicity of this modeling scheme enabled real-time on-the-fly response, without the robot having to "stop and think" before continuing on its way.

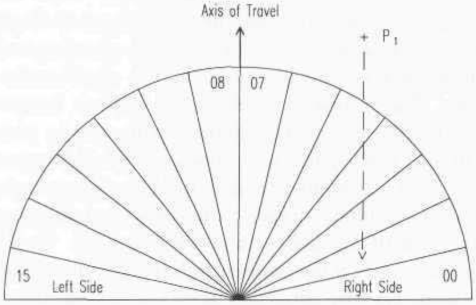

Figure 10-2. The world model employed on ROBART I consisted of sixteen wedge-shaped zones relative to the direction of travel (Everett, 1982).

A fundamental deficiency with this simplistic world representation arises from the fact that a polar model cannot be continuously updated over time to develop a set of certainty values reflecting an increasingly more accurate probability of zone occupancy. For example, a detected obstacle located at point P_1 in the above diagram would transition due to robot motion from *Zone 05* to *Zone 00*, crossing through all zones in between. Repetitive sightings would likely not be associated with the same zone number. As a result, reactions are always made to "snapshot" sensor information subject to numerous sources of potential error, usually resulting in jerky or erratic vehicle movement. Borenstein and Koren (1990a; 1990b) solve this problem by deriving the polar model in real time from a certainty grid representation, as will be discussed in a later section.

Circumnavigation

The term *circumnavigation* describes a collision avoidance behavior in which the robot deflects laterally to "sidestep" an obstruction, while still attempting to move in the general direction of the goal. When the onboard sensors indicate the perceived object is no longer a threat, normal transit is resumed along the desired path. In a sense, *circumnavigation* can be thought of as a *wander* behavior that reverts to a *goal-seeking* behavior when clear, instead of simply resuming straight-line motion.

A good example of a *circumnavigation* collision avoidance behavior is that implemented by Cybermotion, Inc. on their *K2A Navmaster* autonomous vehicle. In normal operation, the *K2A* drive controller calculates a motion vector from its perceived position to some downloaded *X-Y* goal location. This vector is recomputed on the fly as the robot moves so that vehicle heading is continuously reset in accordance with the vector orientation. If a threatening obstacle is detected in the forward path of the robot, speed of advance is reduced and a fixed bias is added to the heading command. The sign of the bias is such that the platform veers away in the direction of free space. Once the obstruction is cleared, the steering bias is removed, and the robot closes on the goal location.

The *circumnavigation* approach has the obvious advantages of simplicity and speed of execution without any requirement for significantly more complex processing power. The technique is obviously limited, however, in that only relatively minor incursions into the intended path can be surmounted in this fashion. Any obstacle that significantly blocks the desired route can potentially deflect the vehicle too far from its intended trajectory for normal path resumption to occur. In addition, circumnavigation techniques must always progress forward in the general direction of the goal without backtracking. No provision is made for choosing alternate routes if the original path is completely blocked.

In addition to being specific behaviors, both *wander* and *circumnavigation* can also be considered as stand-alone *collision avoidance control strategies*. (We're touching now upon that apples and oranges problem I mentioned earlier.) The following sections deal with additional examples of control strategies for collision avoidance (and other purposes) that are capable of implementing not only *wander* and *circumnavigation* but various other behaviors as well.

Potential Fields

The concept of *potential fields* was introduced by Krogh (1984) for simulations of localized mobile robot control, and by Khatib (1985) for manipulator control using Cartesian as opposed to joint coordinates (Tilove, 1990). The classical approach involves an artificial force acting upon the robot, derived from the vector summation of an attractive force representing the goal and a number of repulsive forces associated with the individual known obstacles (Tilove, 1990):

$$\vec{F}_t(x) = \vec{F}_o(x) + \vec{F}_g(x)$$

where:

$\vec{F}_t(x)$ = resultant artificial force vector

$\vec{F}_o(x)$ = resultant of repulsive obstacle forces

$\vec{F}_g(x)$ = attractive goal force.

The attractive goal force can be classically represented as (Tilove, 1990):

$$\vec{F}_g(x) = Q_{goal}\,\frac{\vec{x}-\vec{x}_g}{\left|\vec{x}-\vec{x}_g\right|}$$

where:

Q_{goal} = a positive constant (i.e., the "charge" of the goal).

The *classical potential field* is the summation of the attractive goal force and the repulsive force contributions from those directions defined by the various fields of view of the obstacle detection sensors. The individual repulsive forces are aligned away from their respective obstacles and towards the robot, falling off with the *k-th* power of separation distance (Tilove, 1990). For example, an early MIT implementation on the robot depicted in Figure 10-3 treated each detected sonar target as the origin of a repulsive force decaying as the square of the indicated range (Brooks, 1986). The desired vehicle heading was represented as an attractive force. The resultant of all such virtual forces acting on the robot, if greater than a predetermined threshold, was used to compute the instantaneous drivemotor commands for steering and velocity, effectively moving the platform away from obstacles and in the general direction of the goal.

Figure 10-3. Brooks (1986) applied the *classical potential field* method to an autonomous mobile robot equipped with a ring of 12 Polaroid ranging sensors (courtesy MIT AI Lab).

Alternatively, Arkin (1992b) uses a repulsive force magnitude that is a linear function of obstacle range:

$$O_m = G \, \frac{S-d}{S-R} \quad \text{for } R < d \le S$$

where:

O_m = magnitude of repulsive force associated with obstacle
G = gain constant
S = sphere of influence from center of obstacle (i.e., $O_m = 0$ for $d > S$)
d = distance from center of obstacle to robot
R = radius of obstacle.

In both of the preceding examples, since the resulting field depends only upon the relative positions of nearby obstacles with respect to the robot, it is possible for repulsive forces to be generated by objects that in fact do not lie along the intended path of travel. Such a situation is illustrated in Figure 10-4A below.

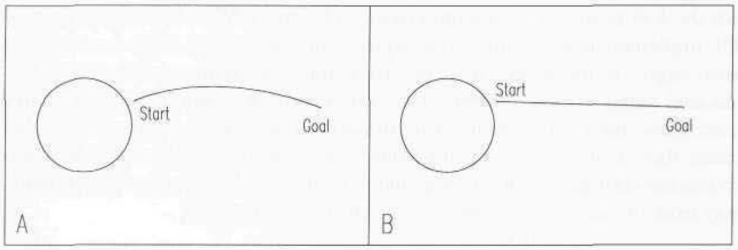

Figure 10-4. (A) The *classical potential field* method considers only separation distance, causing the robot to deviate from a straight-line path segment even though moving away from the circular obstacle; (B) the *generalized potential field* method (see below) considers relative velocity in addition to separation distance (adapted from Tilove, 1990, © IEEE).

In recognition of the above concerns, Krogh (1984) had introduced the concept of *generalized potential fields,* wherein the potential field intensity is a function of not only relative position with respect to obstacles but also the robot's instantaneous velocity vector at that position. The *generalized potential* is the inverse of what Krogh calls the *reserve avoidance time.* Consider a robot approaching a stationary object at some constant velocity V_o as illustrated in Figure 10-5. There is some maximum allowable deceleration rate a_{max} that will bring the robot to a halt in the shortest possible length of time t_1. Similarly there is some minimum deceleration rate a_{min} that will cause the robot to stop just before impact over some longer time interval t_2. *Reserve avoidance time* is simply the difference in time required to stop for the two cases of maximum-allowed versus minimum-required decelerations, (i.e., t_2 - t_1). The *generalized potential* *field* is thus sensitive to time to impact as opposed to separation distance (Figure 10-4B), and approaches infinity as the *reserve avoidance time* approaches zero (Tilove, 1990).

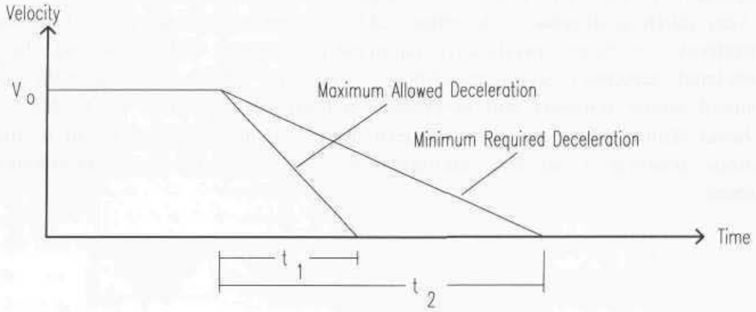

Figure 10-5. Krogh (1984) defines the *generalized potential* as the inverse of *reserve avoidance time*, which is the difference in stopping times associated with *maximum-allowed* and *minimum-required* decelerations (i.e., $t_2 - t_1$).

The principle limitation of the potential field approach is its vulnerability to becoming boxed in or "trapped" by intervening obstacles as illustrated in Figure 10-6. This problem was predicted by Culbertson (1963) for the more general case of "memoryless robots" that react to current stimuli in a deterministic fashion without taking into consideration the results of previous behavior under similar conditions. The likely occurrence of cyclic behavior as well as local maxima and minima make any system that relies solely on the potential-field navigation approach somewhat unreliable (Arkin, 1992a). To get around this problem, Thorpe (1984a; 1984b) employed a grid-based search to find a good low-cost path to the goal, adjusted the path off grid to further minimize costs, then executed the path with a variant of potential fields to keep the vehicle on the path. Krogh & Thorpe (1986) discuss the integration of a generalized potential field collision avoidance scheme with a global path planner based on *certainty grids* (to be discussed in a later section) for optimal route planning and trap recovery.

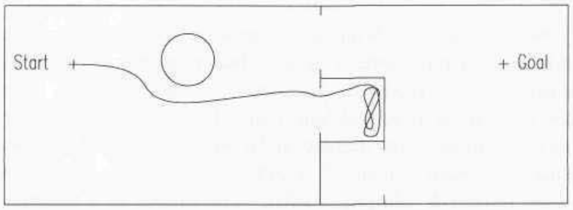

Figure 10-6. The robot successfully negotiated the first obstruction but has become trapped by the U-shaped structure of the closet and is unable to reach the goal in the next room.

Motor Schema

The *motor-schema* navigational approach (Arkin, 1989) developed at the Georgia Institute of Technology, Atlanta, GA, employs a collision avoidance strategy that is very much analogous to potential fields. *Schemas* are basically behavioral primitives, or more specifically, parameterized motor behaviors and their associated perceptual strategies (Arkin & Murphy, 1990), which in turn are denoted *motor schemas* and *perceptual schemas*, respectively. The *Motor-Schema Manager* (Figure 10-7) orchestrates the appropriate interaction of the various *schemas* to achieve intelligent goal-driven actions in a coordinated fashion.

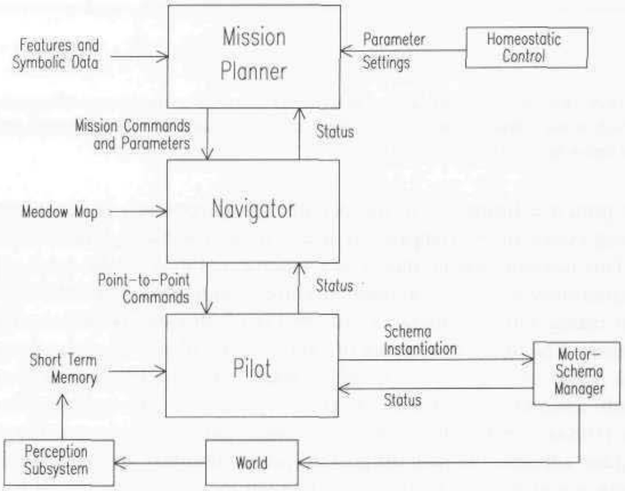

Figure 10-7. The Planning Subsystem of the *Autonomous Robot Architecture (AuRA)* developed by the Georgia Institute of Technology incorporates both a hierarchical planner (Mission Planner, Navigator, and Pilot) and a distributed control plan executor known as the Motor-Schema Manager (Arkin & Murphy, 1990).

Lyons (1986) defines a *schema* as "a generic specification of a computing agent." Each schema represents a general behavior that is instantiated when a copy of the generic specification is parameterized and activated. A collection of such schemas provide the potential family of actions for control of the robot. For example, initial schemas implemented at Georgia Tech on a *Denning Research Vehicle* included: 1) *move-ahead*, 2) *avoid-static-obstacle*, 3) *move-to-goal*, and 4) *stay-on-path* (Arkin & Murphy, 1990). The output of a schema is a single velocity vector reflecting the resolution of all potential field influences experienced by the robot at any given location, and this single vector is used to compute the desired trajectory in real time.

Referring again to Figure 10-7, the *Pilot* implements the desired path generated by the *Navigator* in a piecewise fashion by passing the appropriate selections of both sensing strategies and motor behaviors to the *Motor-Schema Manager* for instantiation. As the robot executes the path, the cartographer builds up a model of surrounding obstacles as perceived by the assigned sensors. If any detected obstacle threatens traversal of a specified path segment, the *Pilot* and *Schema Manager* coordinate in an attempt to avoid the obstruction and resume safe transit along the desired route. Should the resulting path trajectory deviate substantially from the originally specified path, the *Navigator* will be reinvoked to compute a new global path that takes into account the recently acquired sensor data and associated updates to the world model. This fall-back feature also eliminates the common tendency for conventional potential-field approaches to become boxed in or cyclically unstable.

When a suspected obstruction is detected, the instantiated motor schema associated with that particular obstacle begins to produce a repulsive force that deflects the robot away. The magnitude of the deflection vector is a direct function of the certainty of the obstacle's existence. The commanded velocity and direction of the robot is derived from the normalized vector addition of the individual output vectors for all of the active schema instantiations. No arbitration between competing behaviors is involved (Arkin, 1992a). The simplicity of this summation approach and the fact that each schema is a distributed computing agent (preferably running on separate processors on a parallel machine) combine to enable real-time robust performance while the platform is in motion (Arkin, 1989). A learning capability is realized by allowing the system to determine its own schema gains and parameter values, within preset bounds (Clark, et al, 1992).

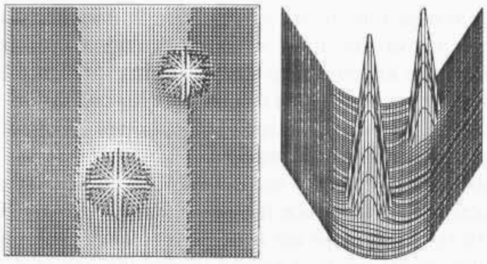

Figure 10-8. Two *avoid-obstacle* plus one *stay-on-path* plus one *move-to-goal* schema instantiations are depicted (left) in 2-D vector representation and (right) in 3-D analog (Arkin, 1989). The entire potential field (shown here for clarity) is not calculated, only each behavior's individual contributions.

10.1.2 Representational World Modeling

Representational world modeling involves the coupling of appropriate non-contact ranging sensors with some type of absolute world modeling capability. Traditional efforts reflected in the literature can generally be decomposed into the following subelements:

- Collecting sensor data on surrounding objects.
- Building an appropriate world model.
- Planning the desired path.
- Avoiding obstacles while en route.

The principle advantage of such an approach over purely reactive strategies is that finding a clear path to the goal is guaranteed, provided such a path exists.

One of the simplest absolute world model representations is a two-dimensional array of cells, where each cell in the array corresponds to a square of fixed size in the region being mapped. Free space is indicated with a cell value of zero; a non-zero cell value indicates an object. The most compact form of a cell map consists of one bit per cell, and thus indicates only the binary presence or absence of an object. By using multiple bits per cell, additional descriptive information can be represented in the map, including the probability of a given square being occupied. This feature is useful when the precise location of an object is unknown, and will be discussed in more detail in the following subsection. Memory usage is independent of map content, so cluttered surroundings are not a problem. However, the resolution of the map is only as good as the square size, and doubling the resolution quadruples the memory requirements.

A slightly more sophisticated and elegant approach is to use a quadtree representation (Fryxell, 1988). Each map begins as a square that is subdivided into four smaller squares. Each of these squares is in turn recursively subdivided (down to the map resolution if necessary) until the region occupied by the square is homogeneous (all object or all free space). For an uncluttered environment, a substantial savings in memory usage is achieved, with a decrease in find-path execution time, since the effective map size is smaller. In highly inhomogeneous environments however, memory usage can increase beyond that of the simple cell map, thus negating the primary advantage of the quadtree. Octrees can be used if a three-dimensional representation is required.

A third technique uses polyhedra and curved surfaces or geometric primitives to represent objects in the workspace (Lozano-Perez & Wesley, 1979; Brooks & Lozano-Perez, 1983). Such maps are quite compact, and with no inherent *grid*, the locations of the objects can be precisely entered into the model. These maps are also easily extended into three dimensions, in contrast to the *cell map* where memory cost would be prohibitive. However, updating the map with real-world data is difficult, as it is hard to accurately glean polygonal information from inexpensive sensors mounted on a mobile robot (Gilbreath & Everett, 1988). Statistical uncertainty of the existence of objects is difficult to implement as well.

Regardless of the particular map representation employed, target distance information must be acquired and entered into the world model as the robot is moving. This seemingly trivial operation turns out to be somewhat difficult due to problems associated with the operation of ultrasonic ranging systems in air. These problems include temperature dependence, which has a significant impact on range accuracy, and beam dispersion, which contributes to angular uncertainty. Specular reflections from target surfaces can cause additional problems. Adjacent sensor interaction requires the transducers in the array be individually fired in sequence rather than simultaneously. Finally, the slow speed of sound in air yields marginal update rates, resulting in significant displacements due to robot motion during the sequential firing of all transducers in the array. Consequently, effective interpretation of range data is critical to achieve a reasonably accurate representation of surrounding obstacles.

Certainty Grids

Moravec and Elfes (1985) of CMU describe a scheme for mapping imprecise sonar range returns into certainty grids using probability distribution functions. For each sensor reading, the assigned probability of an object being at the exact indicated range and bearing decreases radially from a maximum value at that point, according to a specified distribution function. In addition, a second distribution function characterizes the "emptiness" of cells between the sensor and the returned range. Points near the sensor have a high probability of being unoccupied, with decreasing probability for those points closer to the indicated range value or displaced from the beam centerline. The CMU technique is applied to a map where the state of occupancy for all cells is initially marked as unknown. The robot is moved to various vantage points in the room; several sonar readings are taken at each point and averaged to create the probability map. The robot thus creates its own map in an exploratory fashion but must stop periodically to take sensor readings.

Fryxell (1988) also uses a probability scheme for mapping sonar data, modeling the sonars as rays and taking several range readings from different vantage points in the robot's operating environment. Two arrays are constructed in memory, one observing the number of times each cell was "hit" and the other observing each time a cell was "missed." A voting procedure combining both maps is then used to create the final representation, where each cell is marked as either occupied or unoccupied.

Beckerman and Oblow (1988) use a similar method but model the sonar beam as a cone subtending an angle of 18 degrees. The reduced effective beamwidth (18 degrees versus 30 degrees) is achieved by employing the four-transducer phased-array system developed by Kilough and Hamel (1989) for use on the robot HERMIES-IIB. The head-mounted array is sequentially repositioned in azimuth to achieve the desired coverage. As with Fryxell (1988), the robot is moved to various points in the room to make sonar observations. These data are saved in auxiliary buffers and used to update a cumulative map, with each cell labeled as conflicting, unknown, occupied, or empty. (A conflicting cell occurs when one or more sonar readings intersect such that one marks the cell as occupied while the other marks it as empty.) After all the non-conflicting data has been integrated into the cumulative map, the original saved data for each observation point are used to resolve the status of the remaining conflicting cells through pattern analysis. This technique generates maps similar to those created by Fryxell's method, but with better resolution even though fewer sonar readings are taken.

A faster and less computationally expensive variation of these procedures was implemented by Gilbreath (1988) on ROBART II. By using a simplified probability scheme and range-gating a fixed array of sonar sensors, the mapping process can take place in real time while the robot is in motion. When entering range data into the model during actual execution of a path segment, only the center seven transducers in the lower sonar array are activated. If a given sensor reading indicates an object is within 5 feet, the cell at the indicated location is twice incremented (up to a specified maximum). The probability value assigned to each of the eight neighboring cells is incremented once, to partially take into account uncertainties arising from the 30-degree dispersion angle of the ultrasonic beam. (Borenstein and Koren (1990b) carry this simplification one step further by eliminating the probability distribution altogether and incrementing only a single cell on the beam centerline at the indicated range.)

In addition, each time a sonar range value is processed, all the cells within a cone 10-degrees wide and 4-feet long (or less if an object appears within 4 feet) have their assigned values decremented by one. This procedure erodes objects no longer present and also serves to refine their representation as the robot approaches. If the probability is reduced to zero, the cell is again regarded as free space. *Transient objects* are erased from the map at a slightly slower rate than they are entered, so the system tends to err on the side of avoiding obstructions. As with object addition, *permanent objects* and *growth* are left untouched.

Early bit-mapped collision avoidance approaches involved the development of a second localized *relative* map which represented the relative locations of objects detected in front of the robot by onboard sensors while traversing a path segment (Crowley, 1985; Harrington & Klarer, 1987; Everett, et al., 1988). When range to an obstacle fell below a critical threshold, robot motion was halted and a path around the obstacle was planned, using the smaller relative map. In this approach, however, the relative map is very transitory in nature, created at the beginning of each move and discarded at the end. The only information in the map is that obtained from range sensors while the robot is in motion. Since there is no memory of previously encountered obstacles, no learning curve exists, and several avoidance maneuvers may be required to complete the path if the area is congested.

The real-time mapping procedure employed on ROBART II, however, encodes the position of newly detected transient objects into the original *absolute* map while a path is being executed. This scheme has the advantage that all previous information about the environment is also available. All collision avoidance sensor information is statistically represented, based on the number of times something was seen at a given cell location. Figure 10-9 shows a three-dimensional bar-chart depiction of such a map created by a second-generation version of the collision avoidance software that was ported over to the MDARS Interior robot (Everett, et al., 1994). The height of each vertical bar is proportional to the probability that the given cell is occupied.

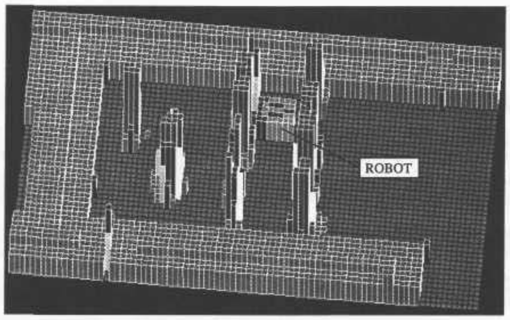

Figure 10-9. Three-dimensional probability distribution resulting from maze traversal during formal collision avoidance technical feasibility testing of the MDARS Interior robot (courtesy Naval Command Control and Ocean Surveillance Center).

The distinction between *permanent* and *transient* objects is an important feature largely responsible for the robust nature of the modeling scheme. *Permanent* objects remain in the model as a baseline from which to restart if the model for some reason becomes overly congested and must be flushed; only the *transient* objects are deleted. Only relatively immobile objects such as walls, desks, filing cabinets, etc. are recorded during the initial map generation procedure. *Permanent objects* are created under human supervision and cannot be later erased by the robot during path execution. *Transient objects* (i.e., chairs, trash cans, carts) are not recorded during the original map-building evolution and present a problem during actual path execution (hence the need for an effective collision avoidance capability).

Each object in the map is automatically *grown* by half the width of the robot in order to model the robot as a dimensionless point during subsequent find-path operations (Lozano-Perez & Wesley, 1979). The path planner will always avoid permanent objects and their associated growth, whereas the algorithm can "eat through" temporary growth surrounding *transient* objects in an attempt to find a path. This ability was found to be necessary, as in congested environments the growth operation often closes off feasible paths due to inaccuracies inherent in the range data. The cost of traversing *transient* growth increases linearly in the direction of the associated object to minimize chances of a collision.

In the original implementation of this collision avoidance strategy on ROBART II, sonar range data was transmitted in real time over an RF datalink to the world modeling software running on an *80386*-based desktop PC. The MDARS Interior program requirement to monitor twelve or more robots from a common host console (Everett, et al., 1993) made this scheme somewhat impractical. As a consequence, the concept of operation was modified slightly to eliminate continuous transfer of sonar data from the various remote platforms to the host. Instead, position-stamped range and bearing information collected over the last 10 feet of travel are stored in a circular buffer on board each MDARS platform.

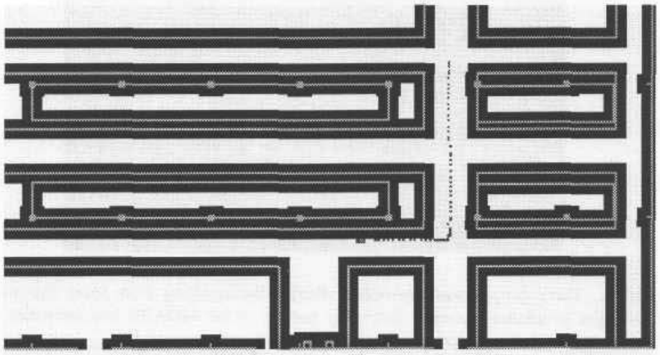

Figure 10-10. Screen dump from the MDARS Planner after an upload of historical collision avoidance data from the robot's circular buffer. Dots corresponding to position stamps for collected data can be used to recreate the path of approach (courtesy Naval Command Control and Ocean Surveillance Center).

If the on-board software determines a threatening object has entered the protected envelope in the direction of travel, the robot is halted and a "blocked" status is reported to the host control architecture. A planner resource is then assigned by the host to resolve the problem. The historical sonar and proximity sensor data are uploaded from the platform after the fact and used to update the world model for path planning purposes. The resultant avoidance maneuver is then downloaded to the robot for execution. This approach allows for a limited number of planner resources to be shared by a large number of robots but is somewhat less than optimal in the sense that the robot must stop and wait while an avoidance maneuver is being generated. One method for overcoming this inconvenience is discussed in the next section.

10.1.3 Combined Approach

In addition to the *reactive control* and *representational world modeling* schemes discussed in the preceding sections, there are also some interesting and innovative implementations employing combinations of both methodologies to achieve more robust operation in dynamic environments.

Vector Field Histogram

As was implied earlier, the *vector field histogram (VFH)* technique developed at the University of Michigan (Borenstein & Koren, 1990a; 1990b) is a combination of both Cartesian and polar representations. This hybrid approach allows for accumulated versus "snapshot" data interpretation in building the model, while exploiting the "on-the-fly" response capability of a polar representation. The VFH method does not reduce the perceived obstacle field to a single resultant vector, and thus allows object distribution to be taken into account in choosing the appropriate avoidance maneuver (Wolfensberger & Wright, 1993). This feature allows the robot to enter narrow passages that would otherwise be inaccessible with a *potential field* collision avoidance strategy.

Raschke and Borenstein (1990) claim more accurate representation of surrounding obstructions is achieved using a *histogramic probability* scheme as opposed to alternative methods employing an assumed probability function. Only a single cell (on the beam centerline at the indicated range) is incremented for each sonar reading (i.e., no probability distribution function is applied to neighboring cells). The rapid sampling of each sonar sensor while the vehicle is moving, however, creates a temporal or *histogramic probability* distribution, wherein recurring sightings cause occupied cells to achieve high certainty values over some finite time frame (Borenstein & Koren, 1990b). The individual *certainty values* represent an assigned *level of evidence* for the existence of an obstacle at that particular grid location.

The *certainty values* from the X-Y *histogram grid* (which is absolute and does not translate with robot motion) are repeatedly mapped over into a *polar histogram* that subsequently reflects the probability of obstruction as a function of relative bearing (Figure 10-11). To simplify this procedure, a smaller window of n-by-n cells symmetrically overlays the vehicle centerpoint, defining a dynamic subset of the *histogram grid* known as the *active region*. Only those cells contained within this *active region* (appropriately denoted as *active cells*) can influence the avoidance behavior of the robot.

A 360-degree *polar histogram* comprised of an integral number of identical pie-shaped sectors is generated around the vehicle's current location, with each sector assigned a variable representing the *polar obstacle density* in that particular direction. To establish the individual values of this circular array, the *certainty value* associated with each *active cell* is treated as an *obstacle vector* oriented along an imaginary line running from the midpoint of the cell to the vehicle centerpoint. The magnitude of this *obstacle vector* is calculated using the equation (Borenstein & Koren, 1990a):

$$m_{i,j} = (c^*_{i,j})^2 \left[a - b\,d_{i,j}\right]$$

where:

$m_{i,j}$ = magnitude of obstacle vector at cell (i,j)
$c^*_{i,j}$ = certainty value of active cell (i,j)
$d_{i,j}$ = distance between active cell (i,j) and vehicle centerpoint
a, b = positive constants.

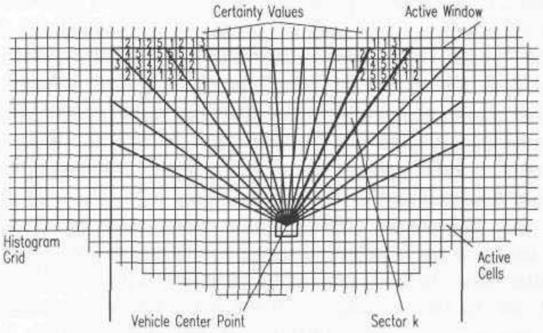

Figure 10-11. *Active cells* from a 33-by-33-cell *active region* of the Cartesian *histogram grid* are mapped into the *polar histogram* to facilitate real-time response (adapted from Borenstein & Koren, 1990a, © IEEE)

The intensity of the avoidance reaction is thus tailored to the square of the magnitude of the *level of evidence* that a perceived object is actually present (Borenstein & Koren, 1990b) and falls off linearly with increasing distance *d*. Clusters of neighboring non-zero cell values are summed together to yield a composite *obstacle cluster strength* that presents an even stronger measure of occupancy for multicell representations indicative of real targets (Borenstein & Koren, 1990a):

$$h_k = \sum_{i,j} m_{i,j}$$

where:

h_k = polar obstacle density for sector k.

Any random noise appearing as unreinforced sightings (i.e., a single unclustered cell) is basically ignored by the avoidance algorithm. Those sectors for which the associated *polar obstacle densities* fall below an assigned threshold are suitable candidate headings for robot motion, with the logical choice being the heading most closely matching the direction of the desired goal. The magnitude of *polar obstacle density* in the direction of forward travel is indicative of anticipated congestion and can be used accordingly to establish an appropriate speed of advance.

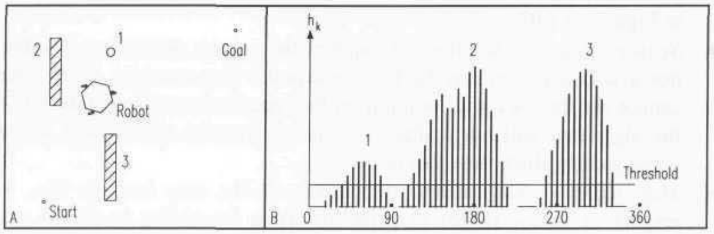

Figure 10-12. Three detected obstacles (A) consisting of two partitions and a vertical pole produce the histogramic probability distribution shown in (B) plotted as a function of relative bearing (adapted from Borenstein and Koren, 1990a).

The *VFH* algorithms were initially implemented at the University of Michigan on a modified Cybermotion *K2A* platform equipped with an onboard *80386*-based PC-compatible computer. An add-on ring of 24 Polaroid ultrasonic transducers (Figure 10-13) provides 360-degree coverage out to 200 centimeters, with a system update rate of around 6 Hz (i.e., 160 milliseconds required to update all sensors). The following steps occur on each pass through the control loop (Borenstein & Koren, 1990a):

- The most recent sonar data is read from the ranging modules.
- The *Histogramic Certainty Grid* is updated.
- A new *Polar Histogram* is created.
- The free sectors and steering direction are determined.
- The maximum allowable speed command is calculated.
- Speed and steering commands are passed to the low-level drive controller.
- Vehicle navigation parameters (X, Y, θ) are received from drive controller.

The above actions repeat every 27 milliseconds, enabling robust real-time operation. Consistent results were achieved in successfully avoiding obstructions in cluttered environments at platform speeds of up to 0.78 meters per second (Borenstein and Koren, 1990a; 1990b).

The success of the *VFH* approach in moving the robot to the desired goal position, however, is situationally dependent. Wolfensberger and Wright (1993) cite four sensitivity factors that influence robust performance:

- *Goal position* — The likelihood of becoming trapped when a valid path exists is sensitive to the relative location of the goal with respect to the robot and any obstructions. In other words, the attractive goal force can in certain situations draw the robot into an intermediate trap as illustrated in Figure 10-14A.

- *Threshold level* — The value assigned to the polar histogram threshold can influence the tendency to enter a trap situation (Figure 10-14B), or potentially preclude the robot from reaching the desired goal as illustrated in Figure 10-14C.
- *Sensor range* — The effective range of the sensors determines whether or not a trap situation can be fully assessed. For example, if the sensor cannot see the back wall of a trap of the type illustrated in Figure 10-14A, the algorithm will not realize there is an eventual obstruction, and will subsequently allow the robot to enter.
- *Map scale* — The relative dimensions of the map features (i.e., with respect to sensor range) likewise influence the ability to assess a trap situation. If the trap illustrated in Figure 10-14 is large enough in terms of depth, the sensors cannot perceive the back wall prior to entry.

Figure 10-13. CARMEL, developed by the University of Michigan, incorporates a Cybermotion *K2A Navmaster* base outfitted with a ring of 24 Polaroid ultrasonic rangefinders (courtesy University of Michigan).

As was the case with potential fields, a limitation of the *vector field histogram* approach is its inherent vulnerability to becoming caught in dead-end trap scenarios. In a series of 100 simulation runs conducted by Wolfensberger & Wright (1993), the *VFH* method became trapped and was unable to reach the goal 30 percent of the time. (It should be noted the potential-field method failed in 79 percent of the tests.) This susceptibility to trapping is not an issue if the *VFH* collision avoidance algorithm is used in conjunction with a higher level global planner as was in fact intended (Borenstein & Koren, 1991). The problem becomes a simple matter of detecting that a trap situation has occurred. (Zhao and BeMent (1990) define a *VFH* trap situation as any time the robot's heading exceeds an angle of 90 degrees with respect to an imaginary line drawn from the current position to the desired goal.) The *VHF* algorithm does not attempt to resolve the problem, but instead automatically passes control to the global path planning, which in turn is better suited to the task of optimal trap recovery. This integrated approach provides a very powerful and robust method for real-time response while ensuring an optimal path for goal attainment.

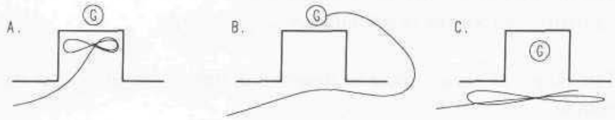

Figure 10-14. The location of the goal in (A) is such that the robot is drawn into an intermediate trap situation, while in (B) the *polar histogram threshold* has been decreased so the proximity of the back wall is above the threshold value. This lower threshold can preclude the robot from reaching the relocated goal in (C) (adapted from Wolfensberger & Wright, 1993).

10.2 Exterior Application Considerations

The collision avoidance problem for exterior applications is much more complex than for interior, even in the case of relatively structured scenarios (i.e., as opposed to open-terrain cross-country operations in unfamiliar territory). An example of such a lower risk category is illustrated by the MDARS Exterior vehicle, which operates almost exclusively on existing facility roadways, for which the associated collision avoidance needs can be subdivided into a number of specific scenarios:

- Fixed obstructions blocking part or all of the roadway.
- Moving obstructions on the roadway.
- Moving obstructions at an intersection (crossroads or train track).
- Potholes or washouts in the roadway.
- Obstructions along the sides of the roadway.
- Hazards (open ditches, lakes, mud bogs, etc.) along the sides of the roadway.

One of the most obvious concerns in fielding an autonomous exterior navigation capability at an industrial site is the need to deal with railroad crossings and roadway intersections. The remote platform must anticipate arrival at such locations in order to "stop and look" before proceeding. The most practical means for detection of oncoming traffic would seem to be Doppler radar and video motion detection.

The added variable which significantly complicates matters in outdoor settings (relative to indoor) is terrain traversability. Wilcox (1994) uses the terminology *non-geometric hazards* to describe pitfalls that cannot be characterized solely by shape, but rather by their properties (such as friction and density, for example, that in turn could adversely impact tire slippage or sinkage). In indoor environments, the floor surface is known in advance and permanent in nature, with the only significant hazards being drop-offs along loading docks and stairwells. Outdoors, this is not the case. Road surfaces can undergo day-to-day as well as seasonal variations in drivability, and hard to detect but potentially hazardous conditions can coexist in close proximity along either side.

This situation introduces two fundamental problems:

- The potential hazard must be detected in time to suitably alter the vehicle's course.
- Some representation of terrain traversability must be encoded within the world model for consideration by the path planning algorithms.

These issues are of little concern in indoor warehouse environments, where it is generally assumed that any areas of potential danger will be readily detected by onboard sensors before an accident can occur. Even if the remote platform is sufficiently disoriented with respect to its true absolute position and orientation, it is generally physically bounded by some type of structure. If the robot wanders too far from its intended location, it will eventually encounter an easily detectable wall or shelf and be halted by the onboard collision avoidance system.

In outdoor environments, however, there is no such bounding structure. Accumulated dead reckoning errors could result in a large enough offset between actual and perceived platform position and heading to where the vehicle could stray off the roadway. Detection of roadway limits is extremely difficult under all weather conditions likely to be encountered, and there is a very real possibility the platform could wind up in a ditch.

Automatic execution of any avoidance maneuver must also consider the fact that other vehicles may be operating on the road section. It is highly probable that conditions will be encountered where an obstacle blocks all or part of the right side of the road, and the required avoidance maneuver by necessity crosses the roadway centerline. Some reliable means of checking for oncoming traffic must precede any automatic execution of the unrestricted path. As in the case of railroad crossings and roadway intersections above, Doppler radar and image processing are strong contenders for this technological need. It must be realized, however, that humans address this issue with the most sophisticated sensors (eyes) and processing resources (brain) in existence, coupled with an extensive database of learned experiences, yet still on occasion make fatal mistakes.

Possible candidates for broad-area first-alert coverage include stereo vision and microwave radar, and a mix of the two is highly desirable from the standpoint of increased likelihood of target detection. Microwave radar is the preferred choice for *intelligent vehicle highway systems* (Siuru, 1994), due to the high reflectivity associated with metal targets (i.e., other vehicles), and the ability to see through obscurants such as fog, rain, or snow. Higher-resolution mapping of target location can also be addressed by stereo vision, with complimentary back-up from narrow-beam ultrasonic sensors (assuming vehicle speed is sufficiently reduced upon initial detection of a potential problem).

In the automotive industry, BMW incorporates four ultrasonic transducers on both front and rear bumpers in its *Park Distance Control* system, but maximum range in this parking-assist application is limited to about 5 feet (Siuru, 1994). This system uses piezoceramic transducers sealed in a membrane for environmental protection, which very likely means ranges much in excess of 5 feet are impractical. Environmental-grade Polaroid transducers have been incorporated in some exterior applications, however, with supposedly good survivability even under adverse conditions, although signal quality and hence performance degrade somewhat in the presence of wind and rain. Alternate possibilities worthy of consideration for the high-resolution mapping task are scanning-laser and structured-light range-finding approaches.

10.3 Navigational Re-referencing

Although reactionary control strategies are unsurpassed in terms of real-time localized control for collision avoidance, purely reactionary control schemes have somewhat limited appeal outside of the laboratory environment. Practical applications generally require some type of global representation to ensure logical and timely attainment of goals. A coupling of global and localized schemes has been suggested and in some cases demonstrated by a number of research groups (Krogh & Thorpe, 1986; Arkin, 1992a; Borenstein & Koren, 1991), and probably represents the optimal approach for most real-world situations for a number of reasons.

The integrity of any world model constructed and refined as the robot moves about its workspace is unfortunately directly dependent upon the accuracy of the robot's perceived location and orientation. Accumulated dead-reckoning errors soon render the information entered into the model invalid, in that the associated geographical reference point for the acquired data is incorrect. As the accuracy of the model degrades, the ability of the robot to successfully navigate and avoid collisions diminishes rapidly, until it fails altogether. For largely this very reason, only a very small handful of autonomous mobile robots have been successfully fielded in real-world applications, despite millions of dollars in developmental efforts over the last several decades.

In a nutshell, except for the case of highly structured environments (such as hospital corridors, for example), it is very difficult to keep from getting lost when operating in the real world. Some reliable and routine means of periodically resetting the navigational parameters (X, Y, and θ) is therefore required if continuous unattended operation is to be sustained. Such methods and their associated sensor needs are addressed next in Chapters 11 through 16.

10.4 References

Anderson, T.L., Donath, M., "Synthesis of Reflexive Behavior for a Mobile Robot Based Upon a Stimulus-Response Paradigm," SPIE Mobile Robots III, Vol 1007, W. Wolfe, Ed., Cambridge, MA, pp. 198-211, November, 1988.

Arkin, R.C., "Motor-Schema-Based Navigation for a Mobile Robot: An Approach to Programming by Behavior," IEEE International Conference on Robotics and Automation, Raleigh, NC, 1987.

Arkin, R.C., Murphy, R.R., "Autonomous Navigation in a Manufacturing Environment," *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 4, pp. 445-454, August, 1990.

Arkin, R.C., "Motor-Schema-Based Mobile Robot Navigation," *International Journal of Robotics Research*, Vol. 8., No. 4, pp. 92-112, August, 1989.

Arkin, R.C., "Behavior-Based Robot Navigation for Extended Domains," *Adaptive Behavior*, Vol. 1, No. 2, MIT, Cambridge, MA, pp. 201-225, 1992a.

Arkin, R.C., "Homeostatic Control for a Mobile Robot: Dynamic Replanning in Hazardous Environments," *Journal of Robotic Systems*, Vol. 9, No. 2, pp. 197-214, 1992b.

Beckerman, M., Oblow, E.M., "Treatment of Systematic Errors in the Processing of Wide Angle Sonar Sensor Data for Robotic Navigation," Oak Ridge National Laboratory Technical Memo, CESAR-88/07, February, 1988.

Borenstein, J., Koren, Y., "High-Speed Obstacle Avoidance for Mobile Robots," IEEE Symposium on Intelligent Control, Arlington, VA, pp. 382-384, August, 1988

Borenstein, J., Koren, Y., "Real-Time Obstacle Avoidance for Fast Autonomous and Semi-autonomous Mobile Robots," American Nuclear Society, Third Topical Meeting on Robotics and Remote Systems, Charleston, SC, CONF-890304, Section 4-4, pp. 1-6, March, 1989a.

Borenstein, J., Koren, Y., "Real-Time Obstacle Avoidance for Fast Mobile Robots," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 5, pp. 1179-1187, September/October, 1989b.

Borenstein, J., Koren, Y., "Real-Time Obstacle Avoidance for Fast Mobile Robots in Cluttered Environments," IEEE International Conference on Robotics and Automation, Vol. CH2876-1, Cincinnati, OH, pp. 572-577, May, 1990a.

Borenstein, J., Koren, Y., "Real-Time Map Building for Fast Mobile Robot Obstacle Avoidance," SPIE Vol. 1388, Mobile Robots V, Boston, MA, November, 1990b.

Borenstein, J., Koren, Y., "The Vector Field Histogram - Fast Obstacle Avoidance for Mobile Robots," *IEEE Journal of Robotics and Automation*, Vol. 7, No. 3, pp. 278-288, June, 1991.

Brooks, R.A., "A Robust Layered Control System for a Mobile Robot," *IEEE Journal of Robotics and Automation*, Vol. RA-2, No. 1, pp. 14-20, 1986.

Brooks, R.A., Lozano-Perez, T., "A Subdivision Algorithm in Configuration Space for Findpath with Rotation," International Joint Conference on Artificial Intelligence, Karlsruhe, Germany, 1983.

Clark, R.J., Arkin, R.C., Ram, A., "Learning Momentum: On-Line Performance Enhancement for Reactive Systems," IEEE International Conference on Robotics and Automation, Nice, France, pp. 111-116, May, 1992.

Crowley, J.L., "Navigation for an Intelligent Mobile Robot," *IEEE Journal of Robotics and Automation*, Vol. RA-1, No. 1, March, 1985.

Culbertson, J., *The Minds of Robots: Sense Data, Memory Images, and Behavior in Conscious Automata*, University of Illinois Press, p. 50, Chicago, IL, 1963.

Everett, H.R., "A Microprocessor Controlled Autonomous Sentry Robot", Masters Thesis, Naval Postgraduate School, Monterey, CA, October, 1982.

Everett, H.R., "A Multi-Element Ultrasonic Ranging Array", *Robotics Age*, pp. 13-20, July, 1985.

Everett, H.R., Gilbreath, G.A., Bianchini, G.L., "Environmental Modeling for a Mobile Sentry Robot", NOSC Technical Document 1230, Naval Ocean Systems Center, San Diego, CA, January, 1988.

Everett, H.R., Gilbreath, G.A., Heath-Pastore, T.A., Laird, R.T., "Coordinated Control of Multiple Security Robots," SPIE Mobile Robots VIII, Vol. 2058, Boston, MA, pp. 292-305, September, 1993.

Everett, H.R., Gilbreath, G.A., Heath-Pastore, T.A., Laird, R.T., "Controliing Multiple Security Robots in a Warehouse Environment," NASA Conference on Intelligent Robotics in Field, Factory, Service, and Space, Vol. 1, Houston, TX, pp. 93-102, March, 1994.

Fryxell, R.C., "Navigation Planning Using Quadtrees," SPIE Mobile Robots II, Cambridge, MA, pp. 256-261, November, 1987.

Gilbreath, G.A., Everett, H.R., "Path Planning and Collision Avoidance for an Indoor Security Robot," SPIE Mobile Robots III, Cambridge, MA, pp. 19-27, Novemeber, 1988.

Harrington, J.J., Klarer, P.R., "SIR-1: An Autonomous Mobile Sentry Robot," Technical Report SAND87-1128, UC-15, Sandia National Laboratories, May, 1987.

Khatib, O., "Real-Time Obstacle Avoidance for Manipulators and Mobile Robots," IEEE Conference on Robotics and Automation, pp. 500-505, March, 1985.

Kilough, S.M., Hamel, W.R., "Sensor Capabilities for the HERMIES Experimental Robot," American Nuclear Society, Third Topical Meeting on Robotics and Remote Systems, Charleston, SC, CONF-890304, Section 4-1, pp. 1-7, March, 1989.

Krogh, B.H., "A Generalized Potential Field Approach to Obstacle Avoidance Control," Proceedings, SME Conference, *Robotics Research: The Next Five Years and Beyond*, Bethlehem, PA, August, 1984.

Krogh, B.H., Thorpe, C.E., "Integrated Path Planning and Dynamic Steering Control for Autonomous Vehicles," IEEE International Conference on Robotics and Automation, San Francisco, CA, pp. 1664-1669, April, 1986.

Lozano-Perez, T., Wesley, M.A., "An Algorithm for Planning Collision-Free Paths Among Polyhedral Obstacles," *Communications of the ACM*, Vol. 22, No. 10, pp. 560-570, 1979.

Lyons, D., "RS: A Formal Model of Distributed Computation for Sensory-Based Robot Control," Ph.D. thesis, COINS Technical Report 86-43, University of Massachusetts, Amherst, MA, 1986.

Moravec, H. P., Elfes, A., "High Resolution Maps from Wide Angle Sonar," Proceedings of the 1985 IEEE International Conference on Robotics and Automation, St. Louis, MO, pp. 116-121, March, 1985.

Raschke, U., Borentstein, J., "A Comparison of Grid-Type Map Building Techniques by Index of Performance," IEEE International Conference on Robotics and Auutomation, Cincinnati, OH, May, 1990.

Siuru, B., "The Smart Vehicles Are Here," *Popular Electronics*, Vol. 11, No. 1, pp. 41-45, January, 1994.

Thorpe, C., "FIDO: Vision and Navigation for a Mobile Robot," PhD Thesis, Carnegie Mellon University, Pittsburgh, PA, November, 1984a.

Thorpe, C. "Path Relaxation: Path Planning for a Mobile Robot," Proceedings, AAAI-84, National Conference on Artificial Intelligence, University of Texas, Austin, TX, August, 1984b.

Tilove, R.B., "Local Obstacle Avoidance for Mobile Robots Based on the Method of Artificial Potentials," IEEE International Conference on Robotics and Automation, Vol. 1, Cincinnati, OH, pp. 566-571, May, 1990.

Wilcox, B.H., "Non-Geometric Hazard Detection for a Mars Microrover," Conference on Intelligent Robotics in Field, Factory, Service, and Space, NASA Conference Publication 3251, Houston, TX, Vol. 1, pp. 675-684, March, 1994.

Wolfensberger, M., Wright, D., "Synthesis of Reflexive Algorithms with Intelligence for Effective Robot Path Planning in Unknown Environments," SPIE Vol. 2058, Mobile Robots VIII, Boston, MA, pp. 70-81, September, 1993.

Zhao, Y., BeMent, S.L., "A Heuristic Search Approach for Mobile Robot Trap Recovery," SPIE Vol. 1388, Mobile Robots V, Boston, MA, pp. 122-130, November, 1990.

11

Guidepath Following

One of the simplest forms of autonomous platform control is *guidepath following*, where the vehicle reacts to the sensed position of some external continuous-path reference track. *Automated guided vehicles (AGVs)* have found extensive use in factories and warehouses for material transfer, in modern office scenarios for supplies and mail delivery, and in hospitals for distribution of meals and medication to nursing stations. In 1989 about 500 AGVs were sold in the United States, compared to approximately 3000 vehicles purchased by European companies, and an estimated 5000 more by Japanese corporations in the same time period (Cahners, 1990).

Advantages of guidepath control in material handling applications are seen primarily in the improved efficiency and reduction of manpower that arise from the fact that an operator is no longer required to perform the delivery function. AGV systems offer a much more flexible and cheaper alternative to conventional hard automation. Large numbers of vehicles can operate simultaneously without getting lost or disoriented, scheduled and controlled by a central computer that monitors overall system operation and remote vehicle flow. Communication with individual vehicles can be over RF links, modulated near-infrared light beams, or other means.

From a manufacturing or industrial point of view, guidepath following represents a tremendous improvement over hard automation through significantly increased flexibility, in terms of both route modifications and floor space usage. From an autonomous mobile robot perspective, however, the fundamental disadvantage of guidepath control is the lack of flexibility in the system: a vehicle cannot be commanded to go to a previously unserviced location unless the guidepath is first modified. While this requirement is certainly an inconvenience for any desired changes to product flow lines, it represents a significant drawback in the case of a security robot, for example, that must investigate a potential break-in at some arbitrary remote location.

The most common guidepath following schemes in use today involve some type of stripe or wire guidepath permanently installed on the floor of the operating area. Specialized sensors mounted on the front of the platform are used to servocontrol the steering mechanism, causing the vehicle to follow the intended route. For purposes of this discussion, such guidance schemes can be divided into three general categories: 1) those which sense and follow the audio or RF field from a closed-loop wire embedded in the floor, 2) those which optically sense and follow some type of stripe affixed to the floor surface, and 3) those which sense a permanent-magnet pathway. A very general comparison of relevant features is provided in Table 11-1 below.

Table 11-1. General comparison (there will be some exceptions) of common guidepath schemes (adapted with changes from Kamewaka & Uemura, 1987).

Salient Feature	Embedded Wire	Optical Stripe	Magnetic Tape
Relative installation cost	High	Low	Low - medium
Maintenance cost	Low	High	Low - medium
Flexibility	None	High	Medium
Communication channel	Yes	No	No
Longitudinal markers	Intersections only	Yes	Yes
Branching	Yes	Yes	Yes
Passing	No	No	No
Breakage tolerant	No	Yes	Yes
Works outdoors	Yes	No	Yes

11.1 Wire Guided

Wire-guided systems represent a practical and proven technology for those AGV applications that enjoy long-term stability in terms of both product flow and equipment location, with higher positioning accuracies (< 0.25 inches typical) than generally achievable using other methods. An inherent safety characteristic is the fact the vehicle will stop automatically should the guidance signal disappear, a feature often used to ensure AGVs do not enter a particular area or load handling station until the appropriate time. By combining off-wire dead-reckoning capabilities with fixed wire guidepaths, a high degree of flexibility can be achieved. Figure 11-1 shows a representative *wire-guided* system installed by Control Engineering, Harbor Springs, MI, for delivery of food and medical supplies in hospital environments.

The principle of operation is illustrated in Figure 11-2 below. A pair of inductive pick-up coils is arranged on the bottom of the vehicle chassis on either side of centerline. The signal amplitude induced in these sense coils due to the alternating magnetic field generated by current flow through the wire guidepath is an inverse function of the coil-to-wire separation distance squared. Any difference in amplitude between the left and right sensor signals can thus be used to servo vehicle heading in order to maintain a balanced output condition. Multiple adjacent wire runs excited at different operating frequencies (generally within the 2- to 10-KHz range) are sometimes employed to facilitate path branches. AGV Products, for example, uses 2575, 3433, 4390, and 5240 Hz in their four-frequency path installations (AGV, 1993). No inherent provisions are readily available for resolving longitudinal position along the route from the guidewire itself, other than detecting cross wires at intersections, and so most systems rely on complementary optical or magnetic sensors for this purpose if required.

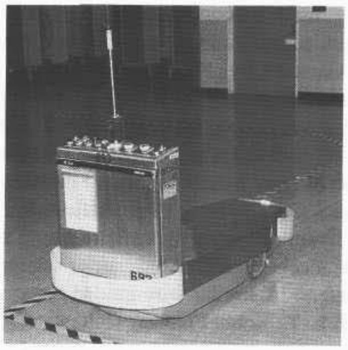

Figure 11-1. This wire-guided system is used to deliver meals and medical supplies to nursing stations at Balboa Naval Hospital (courtesy Naval Medical Center, San Diego, CA).

As an added benefit, RF-modulated information and vehicle commands can be relayed to the AGV over the wire itself. (Since no realistic capability exists to similarly support AGV-to-host communications, this technique is not generally employed in the more sophisticated multivehicle installations, which rely instead on full-duplex RF or optical datalinks.)

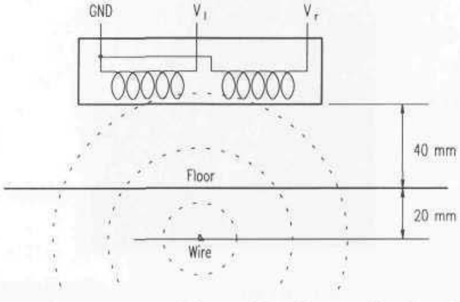

Figure 11-2. Use of a pair of inductive pick-up coils is illustrated in this schematic drawing of the *ANT10* guidepath antenna manufactured by AGV Products, Inc. (courtesy AGV Products, Inc.).

For repetitive AGV operations that involve few if any changes to the guidepath, wire-guided technology is hard to beat from the standpoint of reliability. Initial installation is expensive and then difficult to alter, but the individual vehicle costs are relatively low due to the simplicity of required onboard equipment. Cracks in the flooring due to settling has been known to cause wire breakage, however, which is both disruptive and expensive to repair (Guidoni, 1992). Current installation techniques usually incorporate a styrofoam rope between the wire and epoxy or grout sealer to allow freedom of motion in the event of slab expansion or settling, thus eliminating this problem.

Though traditionally used indoors, there have been some very successful applications of wire-guided technology in rather demanding exterior settings. Barrier Systems, Inc., Carson City, NV, has successfully applied the concept in the control of a large automated vehicle that repositions concrete highway barriers along the 1.7-mile bridge connecting San Diego with Coronado Island (Figure 11-3). Moving at speeds up to 5 miles per hour, the patented system picks up and laterally displaces 2,500 of these barrier sections a distance of 6 feet (four times a day) to optimize lane usage during rush-hour traffic (Murray, 1994). The 25-ton front- and rear-steered hydrostatic-drive vehicle relies on wire guidance to ensure proper placement of the mechanically linked barrier sections, with a path-tracking accuracy of 2.5 centimeters. A similar wire-guided system is in operation along a 10-mile section of the R.L. Thornton Freeway (Interstate 30) in Dallas, TX.

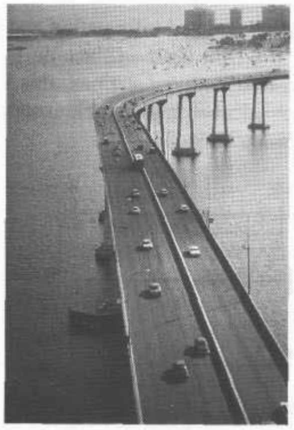

Figure 11-3. The *Quickchange® Moveable Barrier* system uses buried-wire guidepath control to automatically relocate concrete barrier sections on San Diego's Coronado Bridge for optimal lane usage during rush-hour traffic (courtesy Barrier Systems, Inc.).

11.2 Optical Stripe

Optical stripe-following schemes have evolved over the past several decades as favorable alternatives to wire-guided schemes for those less structured applications requiring more flexibility in terms of path additions and/or modifications. Various implementations of the concept exist, including the most elementary case of a high-contrast (dark-on-light, light-on-dark) line, a retroreflective tape illuminated by an onboard light source, and a chemical stripe that glows when irradiated by ultraviolet energy. A number of marker and barcode schemes are employed to yield localized longitudinal position and branching information along the path of travel. Petriu (1991) describes a unique although somewhat impractical implementation employing three parallel tracks (Figure 11-4) to provide continuous absolute position encoding over the full length of the path.

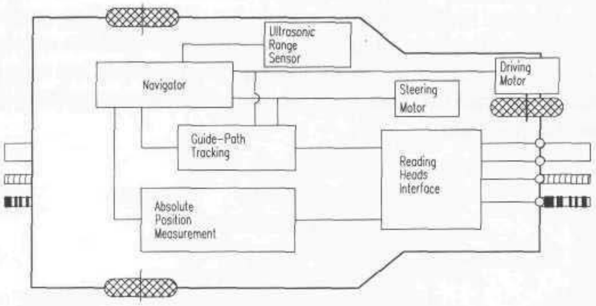

Figure 11-4. Four optical sensors on this experimental AGV follow a three-stripe guidepath that yields continuous absolute position information along the route (adapted from Petriu, 1991, © IEEE).

Most optically based systems have the ability to follow either the left edge or the right edge of the stripe on command, a capability that facilitates branches down alternate paths. Installation problems arise in the case of curved pathways when conventional retroreflective tape is employed as the guidepath stripe. (Retroreflective tape comes on rolls and has little lateral flexibility.) Tsumura (1986) proposed the requirement for curved tape sections of varying radii as a solution in place of piecewise approximation. This problem is of course avoided in systems that follow a painted guidepath.

Optical guidepaths are much easier to install and more flexible in terms of subsequent changes to the route layouts than embedded-wire schemes. Since the guidepath is passive, there is essentially no limitation on route length. Unlike wire-guided systems, any small breaks or discontinuities in an optical path are of little significance, but the tradeoff is recurring installation and maintenance costs due to the exposed nature of the stripe. Some users claim the inherent visibility of the stripe is an advantage from the standpoint of alerting humans to the route, and as a consequence fewer obstructions are inadvertently left to block the vehicle. The same effect could be achieved, of course, by painting lane markings on the floor for embedded-wire runs, but it involves an extra expense. Another disadvantage associated with optical stripe followers (relative to wire-based schemes) is seen in the potential for path occlusion by minor debris inadvertently left on the floor (i.e., packing materials, trash, paint spills, metal shavings).

11.2.1 ModBot Optical Stripe Tracker

The prototype stripe follower developed for use on the ModBot is based on a near-infrared analog sensor module (P/N C5-1DN05) manufactured by Banner Engineering (Banner, 1993a, 1993b), tracking a 1-inch-wide retroreflective tape (P/N BRT-THG-1-100). The sensor head consists of four of these modules arranged in an array to yield a 4-inch footprint of illumination on the floor as shown in the block diagram of Figure 11-5. All sensor modules are active when the system is in the acquisition mode looking for the stripe, but only the center two are used in the tracking mode once the stripe has been located and the robot is centered on the guidepath.

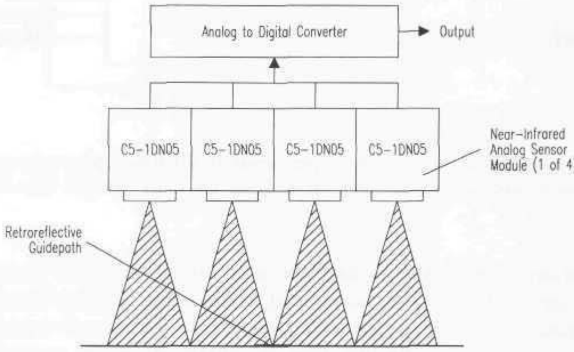

Figure 11-5. The prototype stripe following module used on the ModBot tracks a 1-inch-wide retroreflective tape affixed to the floor.

The initial design called for two additional sensors with fan-shaped beam patterns to provide a binary indication of stripe presence out to 4 inches on either side of the coverage area depicted in Figure 11-5 above. This additional capability turned out to be unnecessary, however, as the fast update rate associated with the four analog sensors proved more than sufficient for real-time tracking, and platform motion was incorporated to expand the window of acceptance during initial stripe acquisition. In the event a stripe segment was not acquired within 6 inches of its expected position (see Section 11.5), the robot would stop and pivot in place, turning 90 degrees to the left and right of its original heading while monitoring for stripe detection. With the particular geometry of the ModBot base (Figure 11-6), this action results in a 14-inch radius of curvature for the search arc.

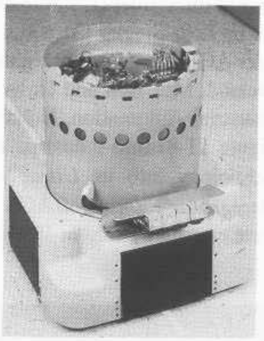

Figure 11-6. The prototype retroreflective guidepath tracking sensor employed on the ModBot was based on four Banner Engineering analog retroreflective sensors (courtesy Naval Command Control and Ocean Surveillance Center).

To assist in making a conventional straight-line transition from off-path to guidepath control, the binary stripe-status information (i.e., present, not present) is monitored for each of the four sensor heads to establish the intercept angle with respect to the path (Figure 11-7). For example, if the robot's search trajectory crosses the guidepath in a perpendicular fashion, all four channels will detect the tape at approximately the same time. If the intercept angle is obtuse, the sensors will respond in sequential fashion from left to right, and vice versa if the angle is acute. The rate of sequential detection determines the magnitude of the angle. This information is used by the control software to determine how far forward the robot should continue after initial acquisition and in which direction it should then turn to center itself over the stripe and shift to tracking mode.

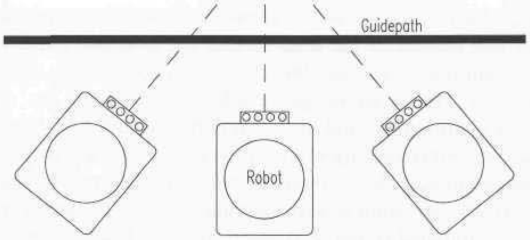

Figure 11-7. The rate and direction of individual sensor detection determine the magnitude and direction of the guidepath intercept angle.

11.2.2 U/V Stimulated Emission

Probably the most versatile optical stripe following system in widespread use today is based on the principle of *stimulated emission* of an essentially invisible chemical stripe when exposed to ultraviolet (UV) radiation. This technology was developed in the early 1970s by Lear-Siegler Corporation, Grand Rapids, MI, for use on their *Mailmobile-Series* AGVs, with patent rights subsequently sold to Bell and Howell Corporation, Zeeland, MI. Bell and Howell sold the patent rights in the early 1980s to Litton Industrial Automation, Holland, MI, in a bid to expand from office applications (Figure 11-8) into industrial material handling markets. Litton implemented the technology into its *UltraFlex®* guidepath scheme in support of the company's *Integrator®* series of AGVs (Litton, 1989). Litton's Holland-based operations were eventually sold in 1994 to Saurer Automation.

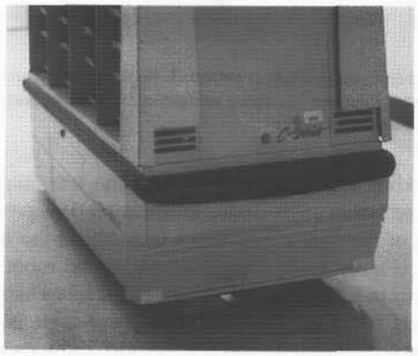

Figure 11-8. The fluorescent glow of the stimulated chemical guidepath is clearly visible beneath this Bell and Howell *Mailmobile* (see also Figure 3-27), which has been reliably picking up and delivering office mail in the main administration building at NCCOSC for over eight years (courtesy Naval Command Control and Ocean Surveillance Center).

The guidepath stripe is composed of fluorescent particles suspended in a water-based hardener that permeates the floor surface for extended durability, drying in approximately eight hours (Litton, 1992). As in the previously discussed case of retroreflective tape, the passive guidepath is unaffected by minor scuffs, abrasions, or discontinuities, and is essentially unlimited in terms of overall length. An active ultraviolet light source irradiates the guidepath, causing the embedded fluorescent particles to reradiate with a fairly narrow spectral output in the blue-green visible spectrum as seen in Figure 11-8. The lateral position sensor that detects this stimulated response is insensitive to ultraviolet wavelengths, and thus sees a fairly high signal-to-noise ratio associated with the fluorescent stripe relative to the surrounding floor surface. The Bell and Howell *Mailmobile* uses three discrete photodetectors to track the glowing guidepath, with a resultant path-tracking accuracy of ±0.25 inches (B&H, 1985), while Litton incorporated a mirror-scanned photodetector into their *UltraFlex* sensor design with a ±0.3-inch tracking accuracy (Litton, 1992). Saurer Automation has recently introduced a microprocessor-based guidepath sensor that employs a fixed array of solid-state photodetectors with a 4-inch effective field of view.

One potential drawback of the stimulated-emission guidepath scheme is the exposed nature of the path, which requires periodic maintenance upgrades to compensate for wear. Unlike the retroreflective-tape schemes discussed earlier, however, there is no need to first remove the existing chemical guidepath; the new stripe can simply be painted down right on top of the old one. The chemical solution is easily applied with a special applicator to almost any floor surface, including tile, masonry, wood, steel, and most carpeting, and dries quickly to an almost invisible finish (Litton, 1991). To minimize performance degradation due to interfering fluorescent residuals found in some carpet cleaning solutions, Bell and Howell offers a specially formulated industrial carpet shampoo free of optical brightners (B&H, 1990). There have been some isolated instances where the guidance system was confused by the glow of fluorescing salt crystals tracked in from icy sidewalks by pedestrians in cold winter climates.

While this fluorescent guidepath approach is very well suited to AGVs operating over fairly smooth floor surfaces, the additional ground clearance generally required in typical mobile robotic applications introduces some practical concerns that can adversely affect performance. One of the biggest problems is potential interference from direct sunlight. In conventional AGV applications the guidepath sensors are routinely tucked up under the vehicle chassis to minimize exposure to such effects. Bell and Howell in fact offers an optional add-on *Sun Skirt* for use with their *Mailmobile* that effectively eliminates most ambient-light access to the protected sensor region, for improved operation in areas where patches of bright sunlight are routinely encountered (B&H, 1990). With an exposed drive configuration such as the Cybermotion *K2A Navmaster* robot, however, adequate shielding of external light sources is virtually impossible. Saurer's new sensor head features improved circuitry and algorithms that minimize ambient light interference, thereby eliminating the need for physical shielding.

11.3 Magnetic Tape

Magnetic-tape guidepath schemes involve a strip of flexible magnetic material affixed to or buried beneath the floor surface, sensed by an array of fluxgate or Hall-effect sensors (see also Chapter 12) on board the vehicle. Kamewaka & Uemura (1987) describe a belt-like magnet made of barium ferrite compounded with a flexible resin to form a 50-millimeter-wide and 0.8-millimeter-thick strip. A vertical magnetic field component of 6 Gauss is generated by this material at the nominal sensor stand-off distance of 40 millimeters. Tsumara (1986) mentions the use in Japan of a ferrite magnetic powder mixture painted directly onto the floor, but no details as to sensitivity, effectiveness, or identity of the system manufacturer are provided. Advantages include extremely low maintenance, immunity to path discontinuities, tolerance of minor debris along the path, and unlimited route lengths. Primary disadvantages are relatively high initial installation costs, potential magnetic interference from surrounding sources, and the short stand off distance required between the guidepath and sensor head.

11.3.1 Macome Magnetic Stripe Follower

Macome Corporation, Tokyo, Japan has developed a magnetic guidance unit for AGVs that employs an array of saturable-core fluxgate sensors (see Chapter 12). Two analog output signals generated by a pair of coils 80 millimeters apart provide for differential servo-control of vehicle heading. The coils are driven in and out of saturation by a square wave pulse train produced by a local oscillator as shown in Figure 11-9. The excitation voltage is divided across the inductance of the coil I_s in series with a fixed resistance R_s. Since the coil inductance changes as a function of the externally applied magnetic field, the measured voltage across R_s will be indicative of field strength and hence the tape-to-coil separation distance. This signal is detected and smoothed by a half-wave rectifier and filter circuit and compared to a similar signal generated by identical circuitry associated with the other coil. The output signals from this simple interface are sufficient to support a 40-millimeter tape stand-off distance without additional amplification (Kamewaka & Uemura, 1987).

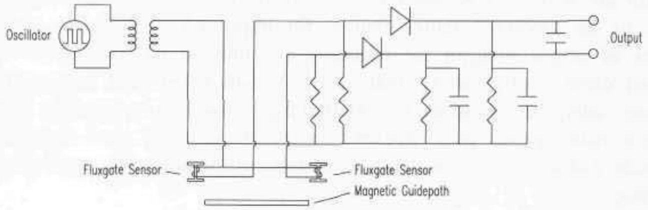

Figure 11-9. Schematic diagram of the interface circuitry associated with the saturable-core fluxgate sensors used to sense tape position (adapted from Kamewaka & Uemura, 1987).

In addition to the two analog tracking sensors, three groups of binary-output sensors furnish coarse zone (left, right, or center) and branch detection information as shown in Figure 11-10. The outputs of all binary sensors within a particular group are logically ORed together to facilitate broader coverage. The three zone signals (indicating whether the vehicle is positioned over the left side zone, right side zone, or center zone of the designated transit lane) are in turn

ORed together to produce a safety signal that stops the vehicle immediately if the guidepath is lost.

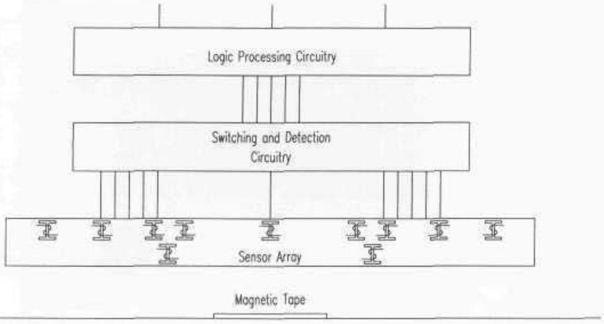

Figure 11-10. Block diagram of Macome magnetic tape sensor unit (adapted from Kamewaka & Uemura, 1987).

The left and right zone signals are logically ANDed together to generate a branch detection flag indicating the vehicle has come to a Y in the path (Figure 11-11). When a branch is detected, the vehicle is shifted in the desired direction by applying an appropriate offset voltage to the output of the steering comparator (Kamewaka & Uemura, 1987).

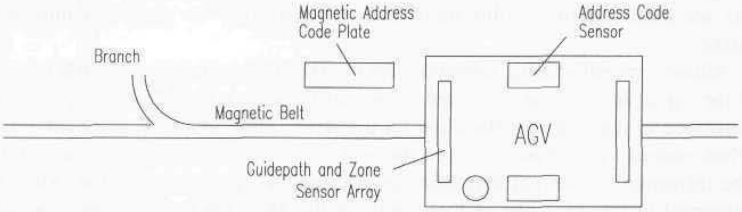

Figure 11-11. Layout of the various components of the Macome system (adapted from Kamewaka & Uemura, 1987).

11.3.2 Apogee Magnetic Stripe Follower

Apogee Robotics, Fort Collins, CO, sells a stand-alone microprocessor-based sensing head (designed for use on their *Orbitor 750* AGV) that employs an array of Hall-effect sensors for tracking a magnetic tape guidepath (Figure 11-12). Effective stand-off distance for a 0.5-inch-wide tape 0.062 inches thick is between 1 and 2 inches, with a corresponding magnetic field strength varying from 9 to 2.5 Gauss over that range. Slightly increased stand-off is possible through use of a thicker tape.

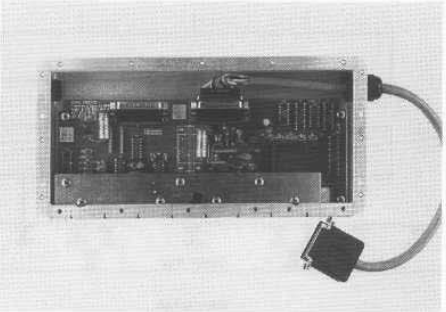

Figure 11-12. Photo of Apogee magnetic tape sensor head with cover removed (courtesy Naval Command Control and Ocean Surveillance Center).

Eight Hall-effect sensors (Microswitch *91SS12-2*) are mounted approximately 1 inch apart to yield two 4.75-inch detection zones, one centered on the vehicle axis for tracking the guidepath and the other located immediately adjacent on the right side for detecting magnetic barcode markers. Barcodes are comprised of discreet magnetic markers, with a logic "0" located 4.0 inches from the center of the guidepath, and a logic "1" displaced 5.5 inches from center. Start and stop bits are indicated by a collocation of a "0" and a "1" to yield a double-wide marker.

Allowed manufacturing tolerances for the Hall-effect devices are roughly equal to the signal level being measured. The sensor head must therefore be carefully calibrated in place after installation on the AGV, after which an automated null-offset routine is incorporated into the software to ensure maximum sensitivity. The influence of external magnetic fields caused a significant problem when we attempted to interface one of these units to the MDARS Interior robot, because unlike conventional AGVs, the rotating nature of the Cybermotion *Navmaster K2A* turret creates a constantly changing vehicle magnetic signature.

The Apogee sensor head measures 3.25 inches high by 9.5 inches wide and 1 inch deep, and requires 12 volts DC at approximately 200 milliamps. Communication with the vehicle electronics is by an RS-232 serial interface.

11.3.3 3M/Honeywell Magnetic Lateral Guidance System

The 3M Traffic Control Materials group in conjunction with the Honeywell Technology Center, Minneapolis, MN, has developed a prototype system that incorporates a magnetic guidepath into conventional retroreflective lane marking material for highway usage. Initial feasibility tests were conducted using an off-the-shelf flexible magnetic strip (*Plastiform® type B-1033*) manufactured by Arnold Engineering, Norfolk, NB. The 4-inch wide tape, consisting of barium ferrite particles suspended in a nitrile rubber binder, exhibits a remnant magnetization (B_r) of about 2500 Gauss (Stauffer, et al., 1995). Three-foot sections of this material were aligned end to end, with every other section inverted to produce an alternating (i.e., north-up, south-up, north-up) vertical field polarity, and then covered with 3M *Scotchlane™ 620-Series* pavement marking tape.

A Honeywell magnetoresistive two-axis magnetometer (see Chapter 3) in bridge configuration is used to sense the lateral position of the tape (Lenz, et al., 1990), with an effective vertical standoff of 9 inches. First-round feasibility tests using an existing three-axis sensor showed reliable marker detection with as much as 6 feet of lateral offset between the sensor and guidepath. A downward looking video camera mounted on the test vehicle adjacent to the magnetometer was used to precisely quantify actual guidepath offset for post-processing of the sensor data output. A total of 26 runs were made along a 100-foot section of marker material at speeds of 15 to 20 miles per hour, with data recorded at a 100-Hz rate. Magnetic deviation was observed to closely agree with control video deviation (in terms of repeatability) to about the same order as the video instrumentation accuracy of 0.25 inches (Stauffer, 1995).

Initially developed as a lateral position control scheme for use in *intelligent vehicle highway system* scenarios, the 3M/Honeywell approach has tremendous potential as an external guidepath for AGVs and robotic vehicles. Efforts are now underway to evaluate a prototype for navigational guidance of the MDARS Exterior program. Other near-term applications include automatic snowplow guidance and automotive safety equipment to warn drivers of impending departure from the traffic lane.

11.4 Heat and Odor Sensing

Australian researchers at the University of Monash report an unusual approach to tracking a temporary pathway similar to the technique used by ants to mark a trail back to their nest. Early work was based on a short-lived heat trail laid down by a lead robot equipped with a quartz-halogen projector bulb configured to raise the floor temperature beneath the unit during the course of transit (Russell, 1993; Kleeman & Russell, 1993). The idea was that additional robots lacking the navigational sophistication of the first could simply follow the temporary trail of the leader in convoy fashion.

A standard *pyroelectric* sensor as is commonly used for passive motion detection in security applications (see Chapter 17) was modified as shown in Figure 11-13 to sense the residual thermal energy imparted to the floor. Two thin-plate lithium-tantalate capacitor sense elements are often employed in sensors of this type for common-mode noise rejection. One capacitor generates a positive voltage across its plates when heated by incident radiation, while the other produces a negative voltage. These two voltages cancel each other for any thermal inputs that are uniform across the full field of view. A special lens is employed in security applications so the lateral motion of an intruder affects first one sensor element and then the other to produce a differential output. To eliminate this common-mode-rejection feature, a brass shim was installed as shown in the diagram to completely shield one capacitor from incoming radiation.

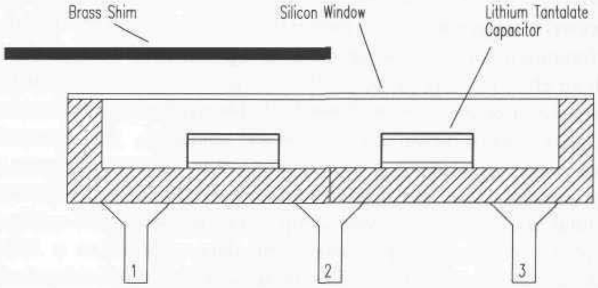

Figure 11-13. A brass shim is used to mask one of the lithium-tantalate capacitors in a conventional *pyroelectric* sensor to defeat common-mode rejection (Russell, 1993, reprinted with permission of Cambridge University Press).

Current-leakage paths in the sensor structure and associated electronics cause the voltage generated across the capacitor plates to decay with time (Russell, 1993), and so the output signal is really proportional to the magnitude of the change in incident thermal radiation. As a consequence, some means of shuttering the energy input was required between readings. In addition, two sensor readings were necessary at points 5 centimeters apart (Kleeman & Russell, 1993) to achieve the differential signal required for vehicle heading control. To meet these needs and eliminate device-specific dependencies, the same sensor was mechanically repositioned back and forth between the left and right sense apertures every 1.2 seconds by a servo-controlled linkage arm (Figure 11-14). A center "reset" position was provided to allow the sensor to settle while screened from all external radiation before reading the next aperture. The servomotor itself is mounted external to the sensor housing as illustrated for thermal isolation.

The sensor package was mounted on the front of a small tracked robot manufactured by Robotnik P/L of Melbourne, Australia. A differential drive correction δ_k was calculated as follows (Russell, 1993):

$$\delta_k = k(t_r - t_l)$$

where:

k = control system proportional gain constant
t_r = reading from right sensor position
t_l = reading from left sensor position.

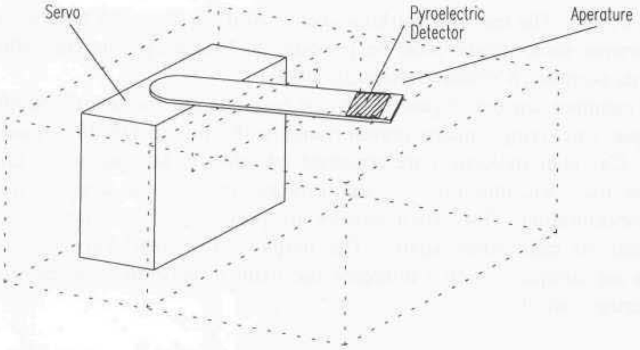

Figure 11-14. Shown inverted for clarity, a modified *pyroelectric* sensor is mechanically positioned back and forth between two sense apertures to track a heat trail on the floor (adapted from Kleeman & Russell, 1993, © IEEE).

Russell reports the optimal value of the proportional gain constant k in the above equation was found to be a function of the age of the heat trail being followed. The resultant correction factor δ_k is applied to left and right track displacements as indicated below:

$$x_l = d + \delta_k$$
$$x_r = d - \delta_k$$

where:

x_l = displacement of left track
x_r = displacement of right track
d = basic translational distance of robot along path.

The magnitude of δ_k is constrained to the range of $\pm d$ so the tracks never move backwards. An average speed of 0.3 meters per minute can be attained while following thermal trails up to 10 minutes old on vinyl floor tile (Russell, 1993). The thermal-path approach was eventually abandoned, however, due to interference problems associated with hot water and heating pipes embedded in the floor, as well as localized hot spots created by shafts of sunlight. In addition, the 70-watt halogen heater placed considerable energy demands on the limited storage capacity of the onboard battery.

The current solution involves laying down a 1-centimeter wide trail of camphor for the slaves to detect and follow (Deveza, et al, 1994). The camphor is dissolved in alcohol and applied with a felt-tip applicator; the alcohol evaporates in seconds, leaving a faint trail of camphor particles on the floor. The choice of camphor was based on its ease of detection and the fact that it is inoffensive to humans, slowly subliming over a period of several hours into a harmless vapor

(Bains, 1994). The use of a marking agent that dissipates with time is desirable, as otherwise some means must be provided to "clean up" the trail after it has served its purpose (Kleeman & Russell, 1993).

The camphor sensors (Figure 11-15) are based on the gravimetric microbalance technique, employing a quartz crystal coated with Silicone OV-17 (Russell, et al., 1994). Camphor molecules are absorbed into the crystal coating, adding to the effective mass and thus lowering the resonant frequency in direct proportion to odor concentration. Two such sensors are positioned 5 millimeters above the floor and 50 millimeters apart. The outputs from their respective oscillator circuits are compared with a reference oscillator to generate a differential signal for steering control.

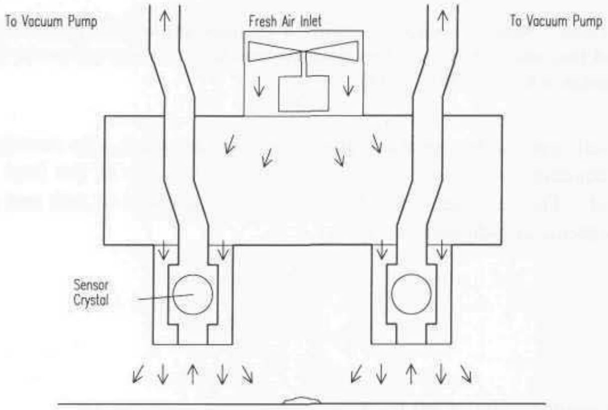

Figure 11-15. A positive pressure air curtain is used to block transient odors that may be carried by drafts into the area of the twin sensing units used to sense the trail of camphor (adapted from Russell, et al., 1994, © IEEE).

To improve tracking performance, a constant intake flow of 200 milliliters per minute is maintained by a modified aquarium pump to draw air into the sensor housing for sampling. Nevertheless, turbulence created by ventilation systems and the movement of doors and people was found to cause significant variations in sensor output, requiring time-consuming averaging for reliable operation. To overcome these effects, an over-pressure air curtain was generated by a small fan to shield extraneous inputs and focus the sensitivity on that area of the floor directly beneath the intake. At room temperature, a typical camphor/air mixture concentration of 236 PPM caused a subsequent frequency drop of 244 Hz (Russell, et al, 1994). The robot is able to follow camphor trails up to 30 minutes old.

11.5 Intermittent-Path Navigation

There is a growing trend throughout the AGV industry to move away from fixed-path guidance schemes altogether due to their inherent lack of flexibility, and more towards what has been loosely termed *off-wire* or *autonomous* guidance. (Both terms are somewhat misleading: *off-wire* overlooks optical and magnetic alternatives, and many conventional guidepath-following vehicles are already in fact fully autonomous.) Gould (1990) cites three production environments where such *free-ranging* (for lack of a better term) navigation schemes are likely to find initial application:

- Where production requirements are in a continuous state of evolution.
- In clean-room operations, where embedding a wire or painting the floor cannot be tolerated.
- In scenarios where the floor itself may be expensive and any alteration would be inappropriate.

A number of these *free-ranging* guidance schemes will be discussed later in subsequent chapters, but first let's take a quick look at a hybrid concept that incorporates some of the best features associated with both *guidepath* and *unrestricted path* navigation.

From a navigational perspective, the fundamental advantage of a fixed guidepath is the simple fact that there is essentially no way to get lost. The big disadvantage of course is the vehicle can only go where the guidepath goes. If an object temporarily obstructs a pathway, everything comes to an inconvenient halt until the obstacle is physically removed. *Unrestricted path navigation*, on the other hand, allows for free traversal around obstacles to basically any desired location for which a clear path indeed exists, but accumulated dead-reckoning errors will eventually cause the system to get lost. It seemed like a good opportunity for synergism, and so the *hybrid navigation scheme* was conceived and subsequently developed on one of the ModBots using the optical stripe follower unit previously shown in Figure 11-5.

Interestingly, while most commercial off-path navigation schemes arose from the need to momentarily leave the guidepath to maneuver around some obstacle, the ModBot implementation came about in exactly the opposite fashion. A free-ranging unrestricted path planning ability had been previously developed by Gilbreath (1988) for use on ROBART II, but suffered from eventual platform disorientation due to dead-reckoning inaccuracies. Periodic traversal of predefined tape guidepaths was therefore incorporated to routinely re-reference the platform, without giving up the ability to move freely to any desired location within the map floor plan (Everett, et al., 1990). Gilbreath's implementation of this *hybrid navigation scheme* within the existing path planner was rather elegant in its simplicity: the predefined guidepath segments were treated as "zero-cost" cells in the modified A* search algorithm, thereby giving them preferred status in the route generation process.

11.5.1 MDARS Interior Hybrid Navigation

It should be noted the *hybrid navigation scheme* is not necessarily restricted to the use of fixed guidepaths for referencing purposes. For example, the MDARS Interior robot employs a derivative of the ModBot approach that was further enhanced under a Cooperative Research and Development Agreement between Cybermotion and NCCOSC (Holland, et al., 1990). This concept combines Cybermotion's *virtual path* navigation with *unrestricted path planning* software developed by NCCOSC for improved collision avoidance. *Virtual paths* provide an effective means for correcting cumulative dead-reckoning position errors, while *unrestricted path planning* allows for transit to any desired location, with the ability to generate avoidance maneuvers around obstacles that may block a *virtual path* segment.

Figure 11-16. The MDARS Interior robot normally traverses *virtual paths* generated by the Cybermotion *Dispatcher*, switching to free-ranging off-path operation in the event of a perceived obstruction.

11.5.2 Free Ranging On Grid

As another example of the industry's desire to get away from even occasional use of guidepath segments, Frog Systems (a subsidiary of Industrial Contractors, Utrecht, The Netherlands) markets a free-ranging AGV that uses a passive floor grid to update the dead-reckoning solution in real time (Gould, 1990). An *a priori* two-dimensional map of the operating area describing the absolute location of walls, obstructions, waypoints, loading stations, and desired routes is maintained in an onboard computer, along with the precise coordinates of the grid components. Considerable flexibility is available in the *FROG (Free Ranging On Grid)* system through use of a variety of potential grid sensors, to include magnetic, optical, and RF. The vehicle's dead-reckoned position is repeatedly reset each time a known grid location is identified.

The grid itself can consist of either line segments or node points as shown in Figure 11-17. Examples of possible *magnetic line-grid* implementations include strips of metal attached to the floor with adhesive, or embedded beneath the surface as in the case of existing steel borders around removable concrete floor sections. The metal line segments are detected with an inductive proximity sensor (Section 3.2.2). An *optical line-grid* configuration has been installed at an Apple Computer factory in Singapore, using a CCD camera to detect the high-contrast intersections of a two-color checkerboard pattern in the floor tiles. Positional accuracy in this case is about 3 millimeters for the 30- by 30-centimeter grid resolution established by the chosen tile size (Van Brussel, et al., 1988).

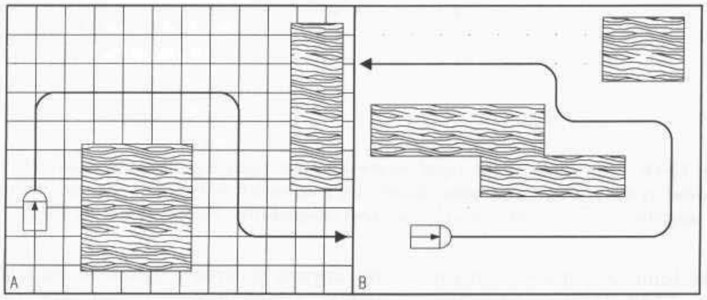

Figure 11-17. The passive grid used by FROG can consist of line segments (A) or node points (B) connected by imaginary lines (Van Brussel, et al., 1988).

Alternatively, the grid can be comprised of specific nodes in the form of metal or optical markers, or even passive RF transponder tags similar to the type used for product identification in automated inventory applications (see Chapter 17). The use of such interactive tags is attractive from the standpoint that each grid node can be individually distinguishable from all others by virtue of its own unique ID response. Such a scheme is also insensitive to adverse conditions likely to be encountered in industrial environments such as dust, dirt, and spillage, which could adversely impact the reliability of an optically based floor referencing system.

Conventional passive RF transponder tags used for personnel and product identification derive power for operation of their onboard circuitry directly from the transmitted energy emitted by the interrogator. An off-the-shelf tag of this type was modified for the *FROG* application to retransmit an easily identified "position" signal derived by dividing the incoming carrier frequency by 2^n. This intentional change in frequency facilitates detection of the weak response signal in the presence of the very powerful emissions of the vehicle's transmitting antenna (Van Brussel, et al., 1988). A "figure-eight" antenna configuration as illustrated in Figure 11-18 listens for this reply, indicative of the presence of a transponder tag at some predefined grid node location.

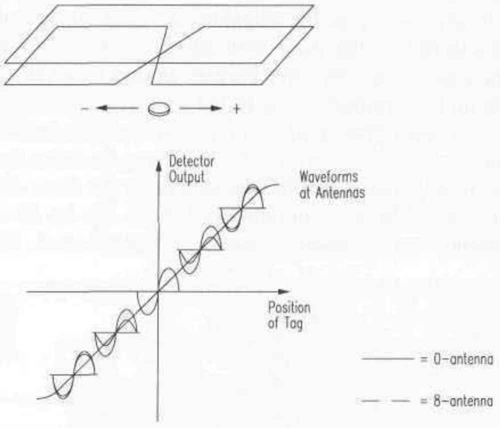

Figure 11-18. The phase of the signal received by the figure-eight antenna relative to that transmitted by the rectangular antenna indicates tag position left or right of centerline, while the signal magnitude is a linear function of lateral offset (adapted from Van Brussel, et al., 1988).

The foldover pattern in this receiving antenna generates signals in each loop that are 180 degrees out of phase, and as a consequence cancel each other completely if of equal magnitude. Such a situation would occur with the transponder tag symmetrically located with respect to the two loops (i.e., along the vehicle centerline). As the tag location shifts laterally away from symmetry, however, the loop in closer proximity begins to dominate, and a net antenna signal appears at the receiver input amplifier. The magnitude of this signal grows as a linear function of distance from centerline (within the limits of antenna coverage), while its phase in relationship to the transmitted signal determines the direction (left or right) of displacement. For a vehicle wheelbase of 1 meter, a 28- by 45-centimeter antenna situated 20 centimeters above the floor resulted in a resultant 1-centimeter positional accuracy, linear over a range of ±15 centimeters from centerline.

11.6 References

AGV, "CB20 Technical Manual," Rev. 1.0, AGV Products, Inc., Charlotte, NC, April, 1993.

Bains, S., "Robots Mark Their Territory," *SPIE OE Reports*, No. 124, pp. 1, 9, April, 1994.

Banner, Product Catalog, Banner Engineering, Minneapolis, MN, 1993a.

Banner, "Handbook of Photoelectric Sensing," Banner Engineering, Minneapolis, MN, 1993b.

B&H, "Discover Mailmobile®," Product Brochure #PM-601, Bell and Howell Mailmobile Company, Zeeland, MI, 1985.

B&H, "Mailmobile® Accessories," Product Brochure #MM-001, Bell and Howell Mailmobile Company, Zeeland, MI, 1990.

Cahners, "AGVs in America: An Inside Look," *Modern Materials Handling*, Cahners Publishing Co., a Division of Reed Elsevier, Ink., September, 1990.

Deveza, R., Russell, R.A., Thiel, D, Mackay-Sim, A., "Odour Sensing for Robot Guidance," *International Journal of Robotics Research*, 1994.

Everett, H.R., Gilbreath, G.A., Tran, T., Nieusma, J.M., "Modeling the Environment of a Mobile Security Robot," Technical Document 1835, Naval Command Control and Ocean Surveillance Center, San Diego, CA, June, 1990.

Gilbreath, G.A., Everett, H.R., "Path Planning and Collision Avoidance for an Indoor Security Robot," SPIE Mobile Robots III, Cambridge, MA, pp. 19-27, Novemeber, 1988.

Gould, L., "Is Off-Wire AGV Guidance Alive or Dead?", *Managing Automation*, pp. 38-40, May, 1990.

Guidoni, G., "SGVs Play Active Role in Chrysler's Future," *Plant, Canada's Industrial Newspaper*, Vol. 51, No. 11, 17 August, 1992.

Holland, J.M., Everett, H.R., Gilbreath, G.A., "Hybrid Navigational Control Scheme," SPIE Vol. 1388, Mobile Robots V, Boston, MA, November, 1990.

Kamewaka, S., Uemura, S., "A Magnetic Guidance Method for Automated Guided Vehicles," *IEEE Transactions on Magnetics*, Vol. MAG-23, No. 5, September, 1987.

Kleeman, L., Russell, R.A., "Thermal Path Following Robot Vehicle: Sensor Design and Motion Control," Proceedings, IEEE/RSJ International Conference on Intelligent Robots and Systems, Yokohama, Japan, July, 1993.

Lenz, J.E., et al., "A High-Sensitivity Magnetoresistive Sensor," IEEE Solid-State Sensors and Actuators Workshop, Hilton Head, SC, June, 1990.

Litton, "Litton Integrator® Automated Guided Vehicles," Product Brochure #8/89 3M M05-1303, Litton Industrial Automation, Holland, MI, August, 1989.

Litton, "Litton Integrator® Automated Guided Vehicles," Product Brochure #2/91 4M M05-1303, Litton Industrial Automation, Holland, MI, February, 1991.

Litton, "UltraFlex® Optical Guidepath from Litton," Product Brochure #1/92 5M M05-1305, Litton Industrial Automation, Holland, MI, January, 1992.

Murray, C.J., "Hydraulic Lifesaver for Highways," *Design News*, pp. 68-74, 25 April, 1994.

Petriu, E.M., "Automated Guided Vehicle with Absolute Encoded Guidepath," *IEEE Transactions on Robotics and Automation*, Vol. 7, No. 4, pp. 562-565, August, 1991.

Russell, R.A., "Mobile Robot Guidance Using a Short-Lived Heat Trail," *Robotica*, Vol. 11, Cambridge Press, pp. 427-431, 1993.

Russell, R.A., Thiel, D., Mackay-Sim, A., "Sensing Odour Trails for Mobile Robot Navigation," Proceedings, IEEE International Conference on Robotics and Automation, San Diego, CA, Vol. 3, pp. 2672-2677, May, 1994.

Stauffer, D., Lenz, J., Dahlin, T.J., "A Magnetic Lateral Guidance Concept Using Continuous Magnetic Marking," submitted for publication to *IEEE Control Systems Journal*, 1995.

Tsumura, T., "Survey of Automated Guided Vehicles in Japanese Factories," IEEE CH2282, pp. 1329-1334, 1986.

Van Brussel, H., Van Helsdingen, C.C., Machiels, K., "FROG - Free Ranging on Grid: New Perspectives in Automated Transport," Proceedings, 6th International Conference on Automated Guided Vehicle Systems, Brussels, Belgium, pp. 223-232, October, 1988.

12

Magnetic Compasses

Vehicle heading is the most significant of the navigational parameters (X, Y, and θ) in terms of its influence on accumulated dead reckoning errors. For this reason, sensors which provide a measure of absolute heading or relative angular velocity are extremely important in solving the real-world navigational needs of an autonomous platform. The most commonly known sensor of this type is probably the magnetic compass.

The terminology normally used to describe the intensity of a magnetic field is *magnetic flux density* B, measured in Gauss (G). Alternative units are the Tesla (T) and the gamma (γ), where:

$$1 \text{ Tesla} = 10^4 \text{ Gauss} = 10^9 \text{ gamma}$$

The average strength of the earth's magnetic field is 0.5 Gauss and can be represented as a dipole that fluctuates both in time and space, situated roughly 440 kilometers off center and inclined 11 degrees to the planet's axis of rotation (Fraden, 1993). This difference in location between *true north* and local *magnetic north* is referred to as *variation* (also known as *declination*), and varies with both time and geographical location. Corrective values are routinely provided in the form of *variation* or *declination tables* printed directly on the maps or charts for any given locale.

Instruments which measure magnetic fields are known as *magnetometers*. For application to mobile robot navigation, only those classes of *magnetometers* which sense the magnetic field of the earth are of interest. Such geomagnetic sensors, for purposes of this discussion, will be broken down into the following general categories:

- Mechanical magnetic compasses.
- Fluxgate compasses.
- Magnetoinductive compasses.
- Hall-effect compasses.
- Magnetoresistive compasses.
- Magnetoelastic compasses.

12.1 Mechanical Magnetic Compasses

The first recorded use of a magnetic compass was in 2634 BC, when the Chinese suspended a piece of naturally occurring magnetite from a silk thread and used it to guide a chariot over land (Carter, 1966). Much controversy surrounds the debate over whether the Chinese or the Europeans first adapted the compass for marine applications, but by the middle of the 13th century such usage was fairly widespread around the globe. William Gilbert (1600) was the first to propose that the earth itself was the source of the mysterious magnetic field that provided such a stable navigational reference for ships at sea.

The early marine compasses were little more than magnetized needles floated in water on small pieces of cork. These primitive devices evolved over the years into the reliable and time proven systems in use today, which consist of a ring magnet or pair of bar magnets attached to a graduated mica readout disk. The magnet and disk assembly floats in a mixture of water and alcohol or glycerin, such that it is free to rotate around a jeweled pivot. The fluid acts to both support the weight of the rotating assembly and to dampen its movement under rough conditions.

The sealed vessel containing the compass disk and damping fluid is typically suspended from a two-degree-of-freedom *gimbal* to decouple it from the ship's motion. This *gimbal* assembly is mounted in turn atop a floor stand or binnacle. Situated on either side of the binnacle are massive iron spheres that, along with adjustable permanent magnets in the base, are used to compensate the compass for surrounding magnetic anomalies that alter the geomagnetic lines of flux. The error resulting from such external influences (i.e., the angle between indicated and actual bearing to magnetic north) is known as compass *deviation*. A correction must be made as well for the local *variation* in order to derive true heading. *Variation* and *deviation* are usually expressed in degrees with an E or W suffix indicating which way *true north* lies from *magnetic north* or from *compass heading*. The corrections are made in sequence from left to right according to a mnemonic known to old mariners: *Can Dead Men Vote Twice*:

$$\textit{Compass} + \textit{Deviation} = \textit{Magnetic} \qquad \textit{Magnetic} + \textit{Variance} = \textit{True}.$$

Another potential source of error which must be taken into account is *magnetic dip*, a term arising from the "dipping" action observed in compass needles attributed to the vertical component of the geomagnetic field. The dip effect varies with latitude, from no impact at the equator where the flux lines are horizontal to maximum at the poles where the lines of force are entirely vertical. For this reason, many swing-needle instruments have small adjustable weights that can be moved radially to balance the needle for any given local area of operation. In addition to gimbaled mounting, marine compasses ensure alignment in the horizontal plane by floating the magnet assembly in an inert fluid.

While gimbal and fluid suspension techniques are fairly effective for marine applications where the period of pitch-and-roll disturbances is fairly long, land-based vehicles often encounter significant acceleration effects due to rough or uneven terrain. Foster (1985) expresses the measurement error due to resulting sensor tilt in terms of the local dip angle of the earth's magnetic field:

$$\theta_{TE} = \arctan(\sin\alpha \tan\beta)$$

where:

θ_{TE} = heading error due to vehicle tilt
α = vehicle tilt displacement in north/south direction
β = dip angle of Earth's field.

12.1.1 Dinsmore *Starguide* Magnetic Compass

An extremely low-cost configuration of the mechanical magnetic compass with electronic readout is seen in a product recently announced by the Dinsmore Instrument Company, Flint, MI. The heart of the *Starguide* compass is the Dinsmore *Model 1490* digital sensor (Dinsmore, 1991), which consists of a miniaturized permanent-magnet rotor mounted in low-friction jeweled bearings. The sensor is internally damped such that if momentarily displaced 90 degrees, it will return to the indicated direction in 2.5 seconds, with no overshoot.

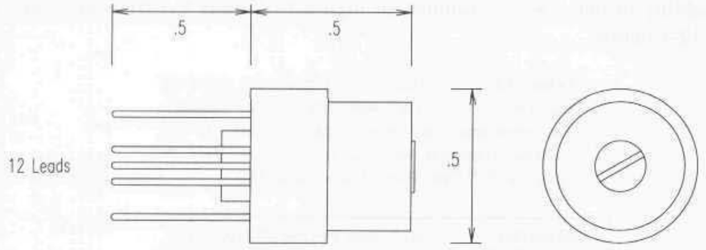

Figure 12-1. The *Model 1490* sensor used in the *Starguide* mechanical compass uses four Hall-effect sensors for electronic readout (courtesy Dinsmore Instrument Co.).

Four Hall-effect switches corresponding to the cardinal headings (N, E, W, S) are arranged around the periphery of the rotor and activated by the south pole of the magnet as the rotor aligns itself with the earth's magnetic field. Intermediate headings (NE, NW, SE, SW) are indicated through simultaneous activation of adjacent cardinal-heading switches. The Dinsmore *Starguide* is not a true *Hall-effect* compass (see Section 12.3), in that the *Hall-effect* devices are not directly sensing the geomagnetic field of the earth, but rather the angular position of a mechanical rotor.

The *Model 1490* digital sensor measures 0.5 inches in diameter by 0.63 inches high and is available separately from Dinsmore for around $12. Current consumption is 30 milliamps, and the open-collector NPN outputs can sink 25 milliamps per channel. Grenoble (1990) presents a simple circuit for interfacing the device to eight indicator LEDs. An alternative analog sensor (*Model 1525*) with a ratiometric sine-cosine output is also available for around $35. An improved analog sensor (*Model 1655*) will be introduced in the summer of 1995, identical in size and shape but with a larger output voltage swing of 1.5 volts (the output swing of the *Model 1525* is 0.8 volts). All three sensors may be subjected to unlimited magnetic flux without damage.

12.2 Fluxgate Compasses

Until most recently there was no practical alternative to the popular *fluxgate compass* for portability and long missions (Fenn, et al., 1992). (New developments in *magnetoelastic* and *magnetoinductive* magnetometers will be presented in following sections.) The term *fluxgate* is actually a trade name of Pioneer Bendix for the *saturable-core magnetometer*, derived from the gating action imposed by an AC-driven excitation coil that induces a time varying permeability in the sensor core. Before discussing the principle of operation, it is probably best to review briefly the subject of magnetic conductance, or *permeability*. The *permeability* μ of a given material is a measure of how well it serves as a path for magnetic lines of force, relative to air, which has an assigned permeability of one. Some examples of high-permeability materials are listed in Table 12-1 below

Table 12-1. Permeability ranges for selected materials; values vary with proportional makeup, heat treatment, and mechanical working of the material (reprinted with permission from Bolz & Tuve, 1979. © CRC Press, Boca Raton, FL).

Material	Relative Permeability μ/μ_o
Supermalloy	100,000 - 1,000,000
Pure iron	25,000 - 300,000
Mumetal	20,000 - 100,000
Permalloy	2,500 - 25,000
Cast iron	100 - 600

Permeability is the magnetic circuit analogy to electrical *conductivity*, and relates *magnetic flux density* to the *magnetic field intensity* as follows:

$$B = \mu H$$

where:

B = magnetic flux density
μ = relative permeability
H = magnetic field intensity.

Since the *magnetic flux* ϕ in a magnetic circuit is analogous to *current* i in an electrical circuit, it follows that *magnetic flux density B* is the parallel to *electrical current density*.

A graphical plot of the above equation is known as the ***normal magnetizing curve***, or *B-H* curve, and the *permeability* μ is the slope. An example plot is depicted in Figure 12-1 for the case of mild steel. In actuality, due to hysteresis, μ depends not only on the current value of H, but also the history of previous values and the sign of dH/dt, as will be seen later. The important thing to note at this point in the discussion is the *B-H* curve is not linear, but rather starts off with a fairly steep slope and then flattens out suddenly as H reaches a certain value. Increasing H beyond this "knee" of the *B-H* curve yields little increase in B; the material is effectively *saturated*, with a near-unity value of *permeability* (i.e., approaching that of air).

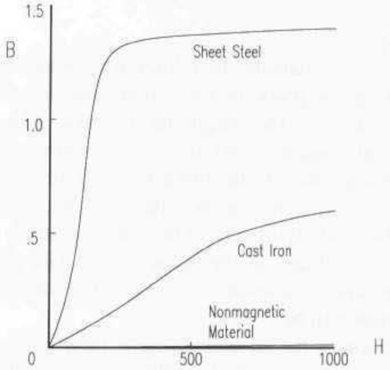

Figure 12-2. The slope of the *B-H* curve, shown here for cast iron and sheet steel, describes the *permeability* of a magnetic material, a measure of its ability (relative to air) to conduct a magnetic flux (adapted from Carlson & Gisser, 1981).

When a highly permeable material is introduced into a uniform magnetic field, the lines of force are drawn into the lower resistance path presented by the material as shown in Figure 12-3A. However, if the material is forced into saturation by some additional magnetizing force H, the lines of flux of the external field will be relatively unaffected by the presence of the saturated material, as indicated in Figure 12-3B. The fluxgate magnetometer makes use of this saturation phenomenon in order to directly measure the strength of a surrounding static magnetic field.

Various core materials have been employed in different fluxgate designs over the past 60 years, with the two most common being *permalloy* (an alloy of iron and nickel) and *mumetal* (iron, nickel, copper, and chromium). The permeable core is driven into and out of saturation by a gating signal applied to an excitation coil wound around the core. For purposes of illustration, let's assume for the moment a square-wave drive current is applied. As the core moves in and out of saturation, the flux lines from the external B field to be measured are drawn into and out of the core, alternating in turn between the two states depicted in Figure 12-3. (This is somewhat of an oversimplification, in that the B-H curve does not fully flatten out with zero slope after the knee.)

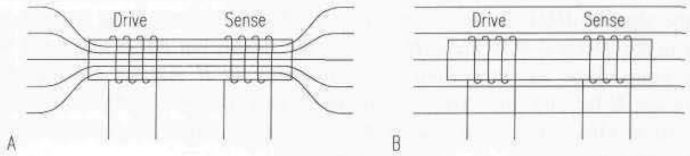

Figure 12-3. External lines of flux for: A) an unsaturated core, and B) a saturated core (adapted from Lenz, 1990, © IEEE).

These expanding and collapsing flux lines will induce positive and negative emf. surges in a sensing coil properly oriented around the core, in accordance with Faraday's law of induction. The magnitude of these surges will vary with the strength of the external magnetic field and its orientation with respect to the axis of the core and sensing coil of the fluxgate configuration. The fact that the permeability of the sensor core can be altered in a controlled fashion by the excitation coil is the underlying principle which enables the DC field being measured to induce a voltage in the sense coil. The greater the differential between the saturated and unsaturated states (i.e., the steeper the slope), the more sensitive the instrument will be.

An idealized B-H curve for an alternating H-field is shown in Figure 12-4A. The permeability (i.e., slope) is high along the section b-c of the curve, and falls to zero on either side of the saturation points H_s and $-H_s$, along segments c-d and a-b, respectively. Figure 12-4B shows a more representative situation: the difference between the left and right hand traces is due to hysteresis caused by some finite amount of permanent magnetization of the material. When a positive magnetizing force H_s is applied, the material will saturate with flux density B_s at point P_1 on the curve. When the magnetizing force is removed (i.e., $H = 0$), the flux density drops accordingly but does not return to zero. Instead, there remains some residual magnetic flux density B_r, shown at point P_2, known as the *retentivity*.

A similar effect is seen in the application of an H-field of opposite polarity. The flux density goes into saturation at point P_3, then passes through point P_4 as the field reverses. This hysteresis effect can create what is known as a *zero offset* (i.e., some DC bias is still present when the external B-field is zero) in fluxgate magnetometers. Primdahl (1970) provides an excellent mathematical analysis of the actual gating curves for fluxgate devices.

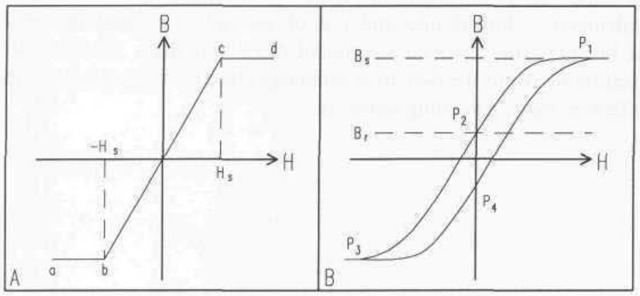

Figure 12-4. In contrast with the ideal *B-H* curve (A), minor hysteresis in the actual curve (B) results in some residual non-zero value of *B* when *H* is reduced to zero, known as the *retentivity* (adapted from Halliday & Resnick, 1974, © John Wiley and Sons, Inc.; Carlson & Gisser, 1981).

The *effective permeability* μ_a of a material is influenced to a significant extent by its geometry; Bozorth and Chapin (1942) showed how μ_a for a cylindrical rod falls off with a decrease in the length-to-diameter ratio. This relationship can be attributed to the so-called *demagnetization factor* (Hine, 1968). When a ferrous rod is coaxially aligned with the lines of flux of a magnetic field, a magnetic dipole develops in the rod itself. The associated field introduced by the north and south poles of this dipole opposes the ambient field, with a corresponding reduction of flux density through the rod. The lowered value of μ_a results in a less sensitive magnetometer, in that the "flux-gathering" capability of the core is substantially reduced.

Consider again the cylindrical rod sensor presented in Figure 12-3, now in the absence of any external magnetic field B_e. When the drive coil is energized, the lines of flux generated by the excitation current are as shown in Figure 12-5 below. Obviously there will be a strong coupling between the drive coil and the sense coil, an undesirable situation indeed since the output signal is supposed to be related to the strength of the external field only.

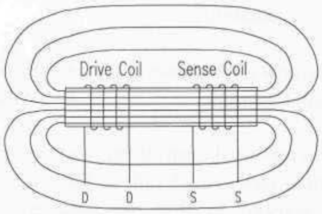

Figure 12-5. Strong coupling between the drive coil and the sense coil in this solonoidal design can interfere with the measurement of an external magnetic field.

One way around this problem is seen in the Vacquier configuration developed in the early 1940s, where two parallel rods collectively form the core, with a common sense coil as illustrated in Figure 12-6 (Primdahl, 1979). The two rods are simultaneously forced into and out of saturation, excited in antiphase by identical but oppositely wound solonoidal drive windings. In this fashion, the magnetization fluxes of the two drive windings effectively cancel each other, with no net effect on the surrounding sense coil.

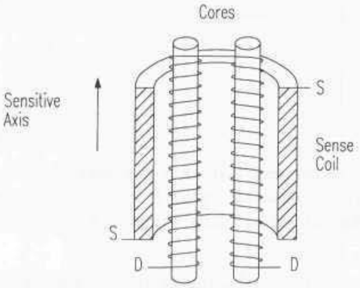

Figure 12-6. Identical but oppositely wound drive windings in the Vacquier configuration cancel the net effect of drive coupling into the surrounding sense coil, while still saturating the core material (adapted from Primdahl, 1979).

Bridges of magnetic material may be employed to couple the ends of the two rods together in a closed-loop fashion for more complete flux linkage through the core. This configuration is functionally very similar to the ring-core design first employed in 1928 by Aschenbrenner and Goubau (Geyger, 1957). In practice, there are a number of different implementations of various types of sensor cores and coil configurations as described by Stuart (1972) and Primdahl (1979). These implementations are generally divided into two classes, parallel and orthogonal, depending on whether the excitation *H*-field is parallel or perpendicular to the external *B*-field being measured. Alternative excitation strategies (sine wave, square wave, sawtooth ramp) also contribute to the variety of implementations seen in the literature. Hine (1968) outlines four different classifications of saturable inductor magnetometers based on the method of readout (i.e., how the output emf is isolated for evaluation):

- Fundamental frequency.
- Second harmonic.
- Peak output.
- Pulse-difference.

Unambiguous 360-degree resolution of the earth's geomagnetic field requires two sensing coils at right angles to each other. The ring-core geometry lends itself to such dual-axis applications in that two orthogonal pick-up coils can be configured in a symmetrical fashion around a common core. Because the drive field follows a circular path around the toroid, from the perspective of either sense coil, the flux up one side of the ring precisely cancels the flux running down the other side (Ramsden, 1994). Since there are no distinct poles in a closed-ring design, *demagnetization effects*, although still present (Stuart, 1972), are less severe. The use of a ring geometry also leads to more complete flux linkage throughout the core, implying less required drive excitation for lower power operation, and the *zero offset* can be minimized at time of manufacture by rotating the circular core (Primdahl, 1979). For these reasons, along with ease of manufacture, toroidal ring-core sensors are commonly employed in many of the low-cost fluxgate compasses available today.

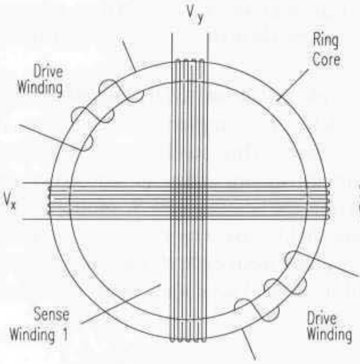

Figure 12-7. The two-channel ring core with toroidal excitation, extended by Gordon and Lundsten (1970) from the Geyger (1962) dual-drive configuration shown above, is the most popular design today in low-cost fluxgate compasses (adapted from Acuna & Pellerin, 1969).

The integrated DC output voltages V_x and V_y of the orthogonal sensing coils vary as sine and cosine functions of θ, where θ is the angle of the sensor unit relative to the earth's magnetic field. The instantaneous value of θ can be easily derived by performing two successive A/D conversions on these voltages and taking the arctangent of their quotient:

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

Another popular two-axis core design is seen in the *Flux Valve* magnetometer developed by Sperry Corporation (Figure 12-8A). This three-legged spider configuration employs three horizontal sense coils 120 degrees apart, with a common vertical excitation coil in the middle (Hine, 1968). Referring to Figure 12-8B, the upper and lower "arms" of the sense coil S are excited by the driving coil D, with a magnetizing force H_x developed as indicated by the arrows. In the absence of an external field H_e, the flux generated in the upper and lower arms by the excitation coil is equal and opposite due to symmetry.

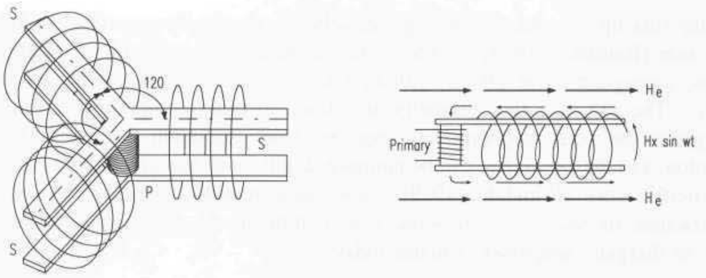

Figure 12-8. The Sperry *Flux Valve* consisted of a common drive winding P in the center of three sense windings symmetrically arranged 120 degrees apart (adapted from Hine, 1968).

When this assembly is placed in an axial magnetic field H_e, however, the instantaneous excitation field H_x complements the flow in one arm, while opposing the flow in the other. This condition is periodically reversed in the arms, of course, due to the alternating nature of the driving function. A second-harmonic output is induced in the sensing coil S, proportional to the strength and orientation of the ambient field. By observing the relationships between the magnitudes of the output signals from each of the three sense coils (Figure 12-9), the angular relationship of the *Flux Valve* with respect to the external field can be unambiguously determined.

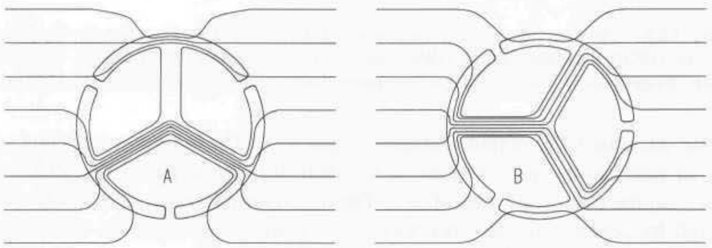

Figure 12-9. The *Flux Valve* magnetometer uses a spider-core configuration to unambiguously resolve the sensor's angular relationship to the geomagnetic field (adapted from Lenz, 1990, © IEEE).

When maintained in a level attitude, a two-axis fluxgate compass will measure the horizontal component of the earth's magnetic field, with the decided advantages of low power consumption, no moving parts, intolerance to shock and vibration, rapid start-up, and relatively low cost. If a ground vehicle is expected to operate over uneven terrain, the sensor coil is often *gimbal-mounted* and mechanically dampened to prevent serious errors introduced by the vertical component of the geomagnetic field. At latitudes encountered in the United States, a 1-degree tilt condition can result in as much as a 5-degree error in heading (Dahlin & Krantz, 1988).

Since gimbal mounting is not always effective under high-*G* operating conditions, some manufacturers have resorted to a three-axis strap-down sensor configuration, and perform a coordinate-transform from sensor coordinates to the horizontal plane of the Earth based on measured vehicle attitude (pitch and roll). Non-gimbaled strap-down sensors are advantageous as well from the standpoint of autocalibration routines that attempt to compensate for surrounding vehicle magnetic anomalies. If the sensor orientation is allowed to change with respect to the vehicle, such automatic compensation is generally not very effective.

12.2.1 Zemco Fluxgate Compasses

One of the first sensors actually employed on ROBART II for navigational referencing purposes was a fluxgate compass manufactured by Zemco Electronics, San Ramon, CA, model number DE-700. This very low-cost (around $40) unit featured a rotating analog dial and was originally intended for 12-volt DC operation in automobiles.

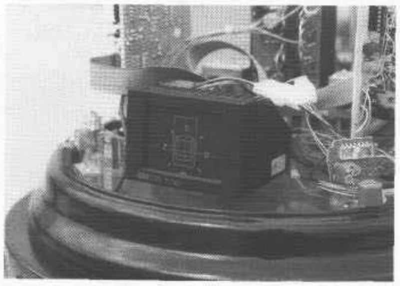

Figure 12-10. The Zemco *DE-700* fluxgate compass was used on ROBART II to resolve the 180-degree ambiguity in a ceiling-mounted heading reference (courtesy Naval Command Control and Ocean Surveillance Center).

A system block diagram is presented in Figure 12-11. The sensor consists of two orthogonal pick-up coils arranged around a toroidal excitation coil, driven in turn by a square-wave oscillator. The outputs V_x and V_y of amplifier channels A and B are applied across an air-core resolver to drive the display indicator. The standard resolver equations (ILC, 1982) for these two voltages are:

$$V_x = K_x \sin\theta \sin(\omega t + a_x)$$
$$V_y = K_y \cos\theta \sin(\omega t + a_y)$$

where:

θ = the resolver shaft angle
$\omega = 2\pi f$, where f is the excitation frequency
K_x and K_y are ideally equal transfer function constants
a_x and a_y are ideally zero time phase shifts.

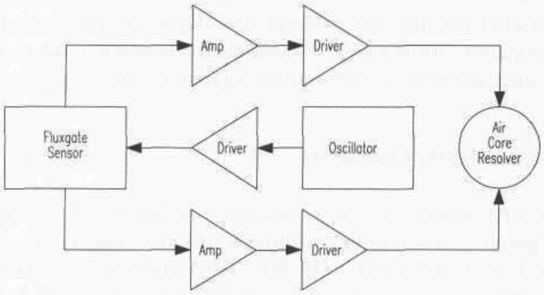

Figure 12-11. Block diagram of ZEMCO Model *DE-700* fluxgate compass (courtesy ZEMCO, Inc.).

Thus, for any static spatial angle θ, the equations reduce to:

$$V_x = K_x \sin\theta$$
$$V_y = K_y \cos\theta$$

which can be combined to yield:

$$\frac{V_x}{V_y} = \frac{\sin\theta}{\cos\theta} = \tan\theta \,.$$

Magnetic heading θ therefore is simply the arctangent of V_x over V_y, as previously stated.

Problems associated with the use of this particular fluxgate compass on ROBART II, however, included a fairly high current consumption (250 milliamps), and stiction in the resolver reflecting back as load into the drive circuitry, introducing some error for minor changes in vehicle heading. In addition, the sensor itself was affected by surrounding magnetic anomalies, some that existed on board the robot (i.e., current flow in nearby cable runs, drive and head positioning motors), and some present in the surrounding environment (metal desks, bookcases, large motors, etc.).

The most serious interference turned out to be the fluctuating magnetic fields due to power cables in close proximity (on the order of 12 inches) to the fluxgate sensor. As various auxiliary systems on the robot were powered up when needed and later deactivated to save power, the magnetic field surrounding the sensor would change accordingly. Significant errors could be introduced as well by minor changes in the position of cable runs, which occurred as a result of routine maintenance and trouble shooting. These problems were minimized by securing all cable runs with plastic tie-downs, and adopting a somewhat standardized protocol regarding which auxiliary systems would be activated when reading the compass.

There is no ready solution, however, for the interference effects of large metallic objects within the operating environment, and deviations of approximately 4 degrees were observed when passing within 12 inches of a large metal cabinet, for example. A final source of error was introduced by virtue of the fact the fluxgate compass had been mounted on the robot's head, as far away as possible from the effects of the drive motors and power distribution lines discussed above. The exact head position could only be read to within 0.82 degrees due to the limited resolution of the 8-bit A/D converter. In any event, an overall system error of ±10 degrees was typical and grossly insufficient for reliable dead-reckoning calculations, but the compass was not originally intended for this use (see Section 16.3.1).

This analog compass was later replaced by a newer digital version produced by Zemco, model *DE-710*, which cost approximately $90. The system block diagram is shown in Figure 12-12. This unit contained a built-in *ADC0834* A/D converter to read the amplified outputs of the two sensor channels, and employed its own *COP 421-MLA* microprocessor, which drove a liquid crystal display (LCD). All communication between the A/D converter, microprocessor, and display driver was serial in nature, with a resulting slow update rate of 0.25 Hz. The built-in LCD simulated an analog dial with an extremely coarse resolution of 20 degrees between display increments, but provision was made for serial output to an optional shift register and associated three-digit numerical display.

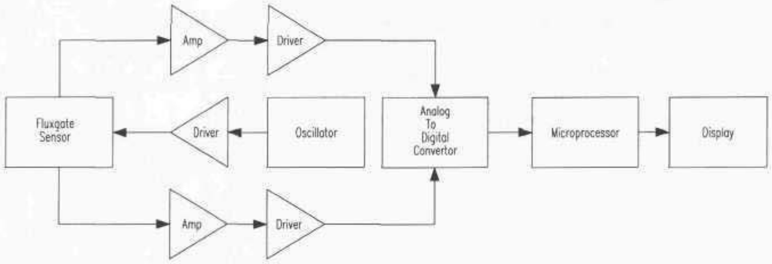

Figure 12-12. Block diagram of ZEMCO model *DE-710* fluxgate compass (courtesy ZEMCO, Inc.).

All things considered, it was determined to be more practical to discard the built-in microprocessor, A/D converter, and LCD display, and interface an external A/D converter directly to the amplifier outputs as before with the analog version. This approach resulted in a decrease in supply current from 168 to 94 milliamps. Power consumption turned out to be less of a factor when it was discovered the circuitry could be powered up for a reading, and then deactivated afterwards with no noticeable effect on accuracy.

Overall system accuracy for this configuration was typically ±6 degrees, although a valid comparison to the analog version is not possible since the digital model was mounted in a different location to minimize interference from nearby circuitry. The amount of effort put into calibration of the two systems must also be taken into account; the calibration procedure as performed was an iterative process not easily replicated from unit to unit with any quantitative measure.

12.2.2 Watson Gyro Compass

A combination fluxgate compass and solid-state rate gyro package (part number FGM-G100DHS-RS232) is available from Watson Industries, Eau Claire, WI. The system contains its own microprocessor to integrate the information from both the rate gyro and the compass for a more stable output less susceptible to interference, with an update rate of 40 Hz. The piezoelectric tuning-fork gyro (see Chapter 13) serves to filter out the effects of magnetic anomalies in the surrounding environment, while the compass counters the long-term drift of the gyro. The toroidal ring-core fluxgate sensor is internally gimbal-mounted for improved accuracy. An overall block diagram is presented in Figure 12-13.

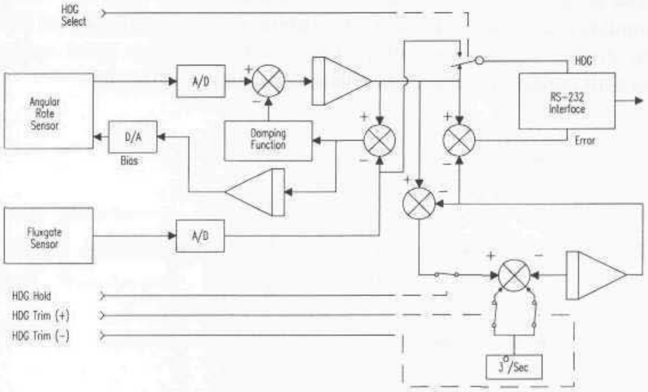

Figure 12-13. Block diagram of Watson fluxgate compass and rate gyro combination (courtesy Watson Industries, Inc.).

The Watson unit measures 2.5 by 1.75 by 3.0 inches, and weighs only 10 ounces. This integrated package is a much more expensive unit ($2500) than the low-cost Zemco fluxgate compass but is advertised to have higher accuracy (±2 degrees). Power supply requirements are 12-volts DC at 200 milliamps, and the unit provides an analog voltage output as well as a 12-bit digital output over a

2400-baud RS-232 serial link. Extensive testing of the Watson compass/gyro package on the ModBot (Figure 12-14) confirmed a fairly repeatable accuracy of about ±2 degrees in an indoor warehouse environment with planar floor surfaces.

Figure 12-14. The combination fluxgate compass and solid-state rate gyro package from Watson Industries was tested on the ModBot as a potential navigational sensor for semi-structured warehouse operations (courtesy Naval Command Control and Ocean Surveillance Center).

12.2.3 KVH Fluxgate Compasses

KVH Industries, Inc., Middletown, RI, offers a complete line of fluxgate compasses and related accessories, ranging from inexpensive units targeted for the individual consumer up through sophisticated systems intended for military applications (KVH, 1993). The *C100 Compass Engine* is a versatile low-cost (less than $700) developers kit that includes a microprocessor-controlled stand-alone fluxgate subsystem based on a two-axis toroidal ring-core sensor.

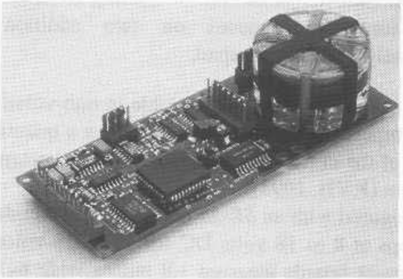

Figure 12-15. The *C-100 Compass Engine* incorporates a gimbaled two-axis toroidal fluxgate design (courtesy KVH Industries, Inc.).

Two different sensor options are offered with the *C-100*: 1) the *SE-25* sensor, recommended for applications with a tilt range of ±16 degrees and 2) the *SE-10* sensor, for applications requiring up to ±45 degrees. The *SE-25* sensor provides internal gimbaling by floating the sensor coil in an inert fluid inside the lexan housing. The *SE-10* sensor provides an additional two-degree-of-freedom pendulous gimbal in addition to the internal fluid suspension. The *SE-25* sensor mounts on top of the sensor PC board, while the *SE-10* is suspended beneath it. The sensor board can be separated as much as 48 inches from the detachable electronics board with an optional cable if so desired.

The resolution of the *C100* is ±0.1 degree, with an advertised accuracy of ±0.5 degrees (after compensation, with the sensor card level) and a repeatability of ±0.2 degrees. Accuracy claims have been substantiated in fielded robotic systems by a number of users, but problems in maintaining a steady horizontal gimbal orientation due to vehicle dynamics have been reported (Rahim, 1993). Separate ±180-degree adjustments are provided for *variation* as well as *index offset* (in the event the sensor unit cannot be mounted in perfect alignment with the vehicle's axis of travel). System damping can be user-selected, anywhere in the range of 0.1 to 24 seconds settling time to final value.

An innovative automatic compensation algorithm employed in the C100 is largely responsible for the high accuracy obtained by such a relatively low-priced system. This software routine runs on the controlling microprocessor mounted on the electronics board and continually corrects for magnetic anomalies associated with the host vehicle. Three alternative user-selectable procedures are offered:

- *Eight-Point Autocompensation* — Starting from an arbitrary heading, the platform turns full circle, pausing momentarily at approximately 45-degree intervals. No known headings are required.
- *Circular Autocompensation* — Starting from an arbitrary position, the platform turns slowly through a continuous 360-degree circle. No known headings are required.
- *Three-Point Autocompensation* — Starting from an arbitrary heading, the platform turns and pauses on two additional known headings approximately 120 degrees apart.

Correction values are stored in a look-up table in non-volatile EEPROM memory. The automatic compensation routine also provides a quantitative indicator of the estimated quality of the current compensation and the magnitude of any magnetic interference present (KVH, 1993).

The *C100* configured with an *SE-25* coil assembly weighs just 2.25 ounces and draws 40 milliamps at 8 to 18 volts DC (or 18 to 28 volts DC). The combined sensor and electronics boards measure 1.8 inches wide by 4.5 inches long. RS-232 (300-9600 baud) and NMEA-0183 digital outputs are provided, as well as linear and sine/cosine analog voltage outputs. Display and housing options are also available, and two variants of the fluxgate design can be special ordered:

- *C-100-x1* — A 3-D *Helmholtz cage* is mounted around the fluxgate sensor to neutralize hard-iron field anomalies in applications where the normal hard-iron compensation is inadequate.
- *C-100-x2* — This version provides the directions and magnitudes of sensor tilt relative to the housing.

In addition, the company also markets a low-cost sensor assembly with analog sine/cosine outputs (*Model AC75*), featuring a separate hermetically sealed fluxgate that can be remotely located in the most favorable magnetic environment.

12.2.4 Applied Physics Systems Miniature Orientation Sensor

Applied Physics Systems, Mountain View, CA, offers a miniature three-axis (pitch, roll, and yaw) angular orientation sensor in a completely self-contained package measuring 0.75 by 0.75 by 4.5 inches and weighing only 50 grams (Figure 12-16). The *Model 544* sensor contains a three-axis accelerometer along with a non-gimbaled three-axis fluxgate magnetometer to produce pitch, roll, and yaw angles that are output in digital format over a bidirectional serial datalink (TTL or RS-232 compatible).

Figure 12-16. The rounded edges of the *Model 544* angular orientation sensor allow the self-contained package to fit inside a 1-inch inside-diameter cylinder (courtesy Applied Physics Systems).

Alternatively, individual acceleration and magnetic field values can be read directly for external postprocessing, with scale factors of about 4 volts per Gauss for the magnetometer, and 2 volts per *G* for the accelerometer readings. An internal *68HC11* microprocessor converts the analog sensor inputs to digital output format with an overall device accuracy of about ±0.5 degrees for each axis (pitch, roll, yaw), providing second-order temperature compensation over a range of 0 to 70°C. Maximum update rate is about 5 Hz if all six output values are read. Power consumption is approximately 50 milliamps at 5 volts DC and 30 milliamps at -5 volts DC. Various configurations of the system are available to suit different applications.

12.3 Magnetoinductive Magnetometers

A relatively new development, *magnetoinductive* sensors employ a single solenoidal winding for each axis and consume roughly an order of magnitude less power than conventional fluxgates (PNI, 1994). The sense coil serves as the inductive element in a low-power L/R relaxation oscillator, with its effective inductance being influenced by the ambient magnetic field component running parallel to the coil axis. For such a magnetic-core solenoid, the effective inductance L can be shown to be (Kim & Hawks, 1989):

$$L = \mu_o n^2 V \frac{dB}{dH}$$

where:

L = coil inductance
μ_o = permeability of air
n = number of coil turns per unit length
V = volume of the core material
B = total magnetic flux
H = magnetizing force.

Recall that dB/dH, the slope of the B-H curve, is simply the relative permeability μ of the core material. For a typical core material such as *METGLAS 2705M*, the relative permeability μ varies as a function of the magnetizing force H as shown in Figure 12-17. (*METGLAS* is an amorphous alloy of iron, boron, silicon, and carbon, and a registered trademark of Allied Signal Corporation.) By biasing the magnetizing force through application of a static DC current flow in the solenoidal coil, the operating point can be centered at H_o in the linear region of the curve as shown. The presence of an external magnetic field H_e adds or subtracts to the bias field H_o, shifting the operating point accordingly, with an associated change in the effective permeability $\mu(H)$.

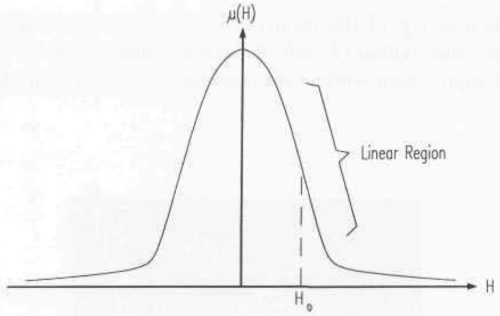

Figure 12-17. Plot of relative permeability $\mu(H)$ as a function of magnetizing force H for METGLAS core material (adapted from Kim & Hawks, 1989).

The period of the oscillator output is proportional to L/R, where L has been shown to be directly proportional to $\mu(H)$. Thus the greater contribution from the ambient magnetic field H_e, the lower the effective inductance L, resulting in an associated increase in oscillator frequency. As a result of this direct relationship, the observed frequency shift can be as much as 100 percent as the sense coil is rotated from a parallel to antiparallel orientation with respect to the geomagnetic field (Kim & Hawks, 1989). Since there is no need to drive the core into and out of saturation as in the case of fluxgate designs, the required electronics is much simpler and power consumption greatly reduced. An additional benefit is seen in the inherently digital nature of the output signal, eliminating the cost and complexities of additional signal conditioning and an analog-to-digital interface.

12.3.1 Precision Navigation *TCM Magnetoinductive Compass*

Precision Navigation, Inc., Mountain View, CA, offers an integrated electronic solution to the problems introduced by conventional mechanical gimbaling under conditions of high dynamic loading typically experienced by ground vehicles operating on uneven terrain. The *TCM* compass shown in Figure 12-18 employs a three-axis *magnetoinductive* magnetometer to measure the *X-Y-Z* components of the geomagnetic field, along with a two-axis electrolytic inclinometer to measure vehicle attitude. The integral microprocessor uses the pitch-and-roll information to automatically correct the magnetometer outputs for tilt, providing a temperature-compensated heading solution up to 16 times each second, accurate to within ±1 degree.

A block diagram illustrating the inherent simplicity of the sense coil interface is shown in Figure 12-19. The solenoidal inductor is connected in series with resistor R_2 to form a relaxation oscillator in conjunction with the *LM339* comparator. R_2 can be used to adjust the DC coil bias current to establish the desired operating point H_o, while R_3 sets the center frequency (approximately 200