CHAPTER 2
HEADING SENSORS

Heading sensors are of particular importance to mobile robot positioning because they can help compensate for the foremost weakness of odometry: in an odometry-based positioning method, any small *momentary* orientation error will cause a *constantly* growing lateral position error. For this reason it would be of great benefit if orientation errors could be detected and corrected immediately. In this chapter we discuss gyroscopes and compasses, the two most widely employed sensors for determining the heading of a mobile robot (besides, of course, odometry). Gyroscopes can be classified into two broad categories: (a) mechanical gyroscopes and (b) optical gyroscopes.

2.1 Mechanical Gyroscopes

The mechanical gyroscope, a well-known and reliable rotation sensor based on the inertial properties of a rapidly spinning rotor, has been around since the early 1800s. The first known gyroscope was built in 1810 by G.C. Bohnenberger of Germany. In 1852, the French physicist Leon Foucault showed that a gyroscope could detect the rotation of the earth [Carter, 1966]. In the following sections we discuss the principle of operation of various gyroscopes.

Anyone who has ever ridden a bicycle has experienced (perhaps unknowingly) an interesting characteristic of the mechanical gyroscope known as *gyroscopic precession*. If the rider leans the bike over to the left around its own horizontal axis, the front wheel responds by turning left around the vertical axis. The effect is much more noticeable if the wheel is removed from the bike, and held by both ends of its axle while rapidly spinning. If the person holding the wheel attempts to yaw it left or right about the vertical axis, a surprisingly violent reaction will be felt as the axle instead twists about the horizontal roll axis. This is due to the angular momentum associated with a spinning flywheel, which displaces the applied force by 90 degrees in the direction of spin. The rate of precession Ω is proportional to the applied torque T [Fraden, 1993]:

Apparent Drift Calculation
(Reproduced with permission from [Sammarco, 1990].)

Apparent drift is a change in the output of the gyroscope as a result of the Earth's rotation. This change in output is at a constant rate; however, this rate depends on the location of the gyroscope on the Earth. At the North Pole, a gyroscope encounters a rotation of 360° per 24-h period or 15°/h. The apparent drift will vary as a sine function of the latitude as a directional gyroscope moves southward. The direction of the apparent drift will change once in the southern hemisphere. The equations for Northern and Southern Hemisphere apparent drift follow. Counterclockwise (ccw) drifts are considered positive and clockwise (cw) drifts are considered negative.

  Northern Hemisphere: 15°/h [sin (latitude)] ccw.
  Southern Hemisphere: 15°/h [sin (latitude,)] cw.

The apparent drift for Pittsburgh, PA (40.443° latitude) is calculated as follows: 15°/h [sin (40.443)] = 9.73°/h CCW or apparent drift = 0.162°/min. Therefore, a gyroscope reading of 52° at a time period of 1 minute would be corrected for apparent drift where

corrected reading = 52° - (0.162°/min)(1 min) = 51.838°.

Small changes in latitude generally do not require changes in the correction factor. For example, a 0.2° change in latitude (7 miles) gives an additional apparent drift of only 0.00067°/min.

$$T = I\,\omega \qquad\qquad (2.1)$$

where
T = applied input torque
I = rotational inertia of rotor
ω = rotor spin rate
Ω = rate of precession.

Gyroscopic precession is a key factor involved in the concept of operation for the *north-seeking gyrocompass*, as will be discussed later.

Friction in the support bearings, external influences, and small imbalances inherent in the construction of the rotor cause even the best mechanical gyros to drift with time. Typical systems employed in inertial navigation packages by the commercial airline industry may drift about 0.1° during a 6-hour flight [Martin, 1986].

2.1.1 Space-Stable Gyroscopes

The earth's rotational velocity at any given point on the globe can be broken into two components: one that acts around an imaginary vertical axis normal to the surface, and another that acts around an imaginary horizontal axis tangent to the surface. These two components are known as the *vertical earth rate* and the *horizontal earth rate*, respectively. At the North Pole, for example, the component acting around the local vertical axis (vertical earth rate) would be precisely equal to the rotation rate of the earth, or 15°/hr. The horizontal earth rate at the pole would be zero.

As the point of interest moves down a meridian toward the equator, the vertical earth rate at that particular location decreases proportionally to a value of zero at the equator. Meanwhile, the horizontal earth rate, (i.e., that component acting around a horizontal axis tangent to the earth's surface) increases from zero at the pole to a maximum value of 15°/hr at the equator.

There are two basic classes of rotational sensing gyros: 1) rate gyros, which provide a voltage or frequency output signal proportional to the turning rate, and 2) rate integrating gyros, which indicate the actual turn angle [Udd, 1991]. Unlike the magnetic compass, however, rate integrating gyros can only measure relative as opposed to absolute angular position, and must be initially referenced to a known orientation by some external means.

A typical gyroscope configuration is shown in Figure 2.1. The electrically driven *rotor* is suspended in a pair of precision low-friction bearings at either end of the rotor axle. The *rotor* bearings are in turn supported by a circular ring, known as the *inner gimbal ring*; this inner gimbal ring pivots on a second set of bearings that attach it to the *outer gimbal ring*. This pivoting action of the inner gimbal defines the horizontal axis of the gyro, which is perpendicular to the spin axis of the rotor as shown in Figure 2.1. The outer gimbal ring is attached to the instrument frame by a third set of bearings that define the vertical axis of the gyro. The vertical axis is perpendicular to both the horizontal axis and the spin axis.

Notice that if this configuration is oriented such that the spin axis points east-west, the horizontal axis is aligned with the north-south meridian. Since the gyro is space-stable (i.e., fixed in the inertial reference frame), the horizontal axis thus reads the horizontal earth rate component of the planet's rotation, while the vertical axis reads the vertical earth rate component. If the spin axis is rotated 90 degrees to a north-south alignment, the earth's rotation does not affect the gyro's horizontal axis, since that axis is now orthogonal to the horizontal earth rate component.
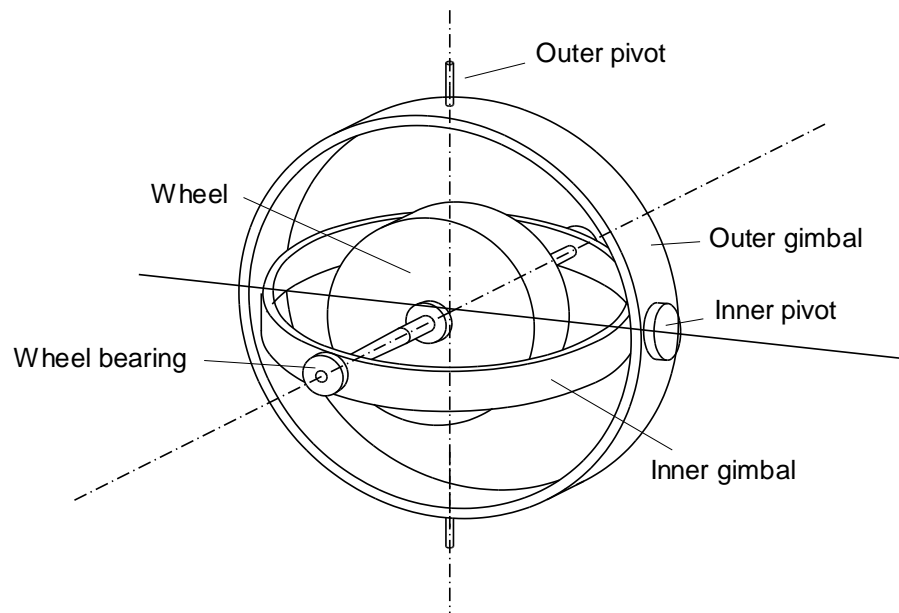
Figure 2.1: Typical two-axis mechanical gyroscope configuration [Everett, 1995].

2.1.2 Gyrocompasses

The gyrocompass is a special configuration of the rate integrating gyroscope, employing a gravity reference to implement a north-seeking function that can be used as a true-north navigation reference. This phenomenon, first demonstrated in the early 1800s by Leon Foucault, was patented in Germany by Herman Anschutz-Kaempfe in 1903, and in the U.S. by Elmer Sperry in 1908 [Carter, 1966]. The U.S. and German navies had both introduced gyrocompasses into their fleets by 1911 [Martin, 1986].

The north-seeking capability of the gyrocompass is directly tied to the horizontal earth rate component measured by the horizontal axis. As mentioned earlier, when the gyro spin axis is oriented in a north-south direction, it is insensitive to the earth's rotation, and no tilting occurs. From this it follows that if tilting is observed, the spin axis is no longer aligned with the meridian. The direction and magnitude of the measured tilt are directly related to the direction and magnitude of the misalignment between the spin axis and true north.

2.1.3 Commercially Available Mechanical Gyroscopes

Numerous mechanical gyroscopes are available on the market. Typically, these precision machined gyros can cost between $10,000 and $100,000. Lower cost mechanical gyros are usually of lesser quality in terms of drift rate and accuracy. Mechanical gyroscopes are rapidly being replaced by modern high-precision — and recently — low-cost fiber-optic gyroscopes. For this reason we will discuss only a few low-cost mechanical gyros, specifically those that may appeal to mobile robotics hobbyists.

2.1.3.1 Futaba Model Helicopter Gyro

The Futaba FP-G154 [FUTABA] is a low-cost low-accuracy mechanical rate gyro designed for use in radio-controlled model helicopters and model airplanes. The Futaba FP-G154 costs less than $150 and is available at hobby stores, for example [TOWER]. The unit comprises of the mechanical gyroscope (shown in Figure 2.2 with the cover removed) and a small control amplifier. Designed for weight-sensitive model helicopters, the system weighs only 102 grams (3.6 oz). Motor and amplifier run off a 5 V DC supply and consume only 120 mA. However, sensitivity and accuracy are orders of magnitude lower than "professional"
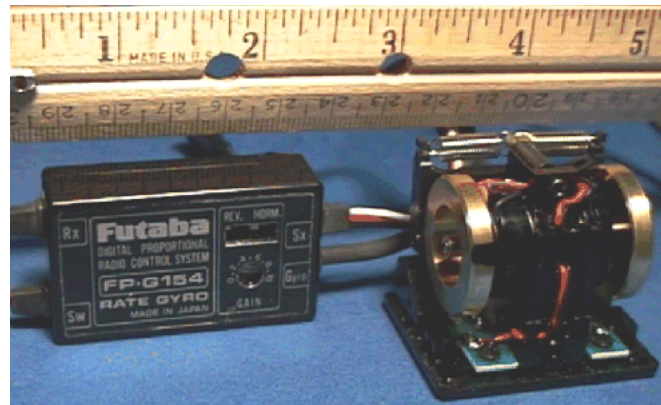
Figure 2.2: The Futaba FP-G154 miniature mechanical gyroscope for radio-controlled helicopters. The unit costs less than $150 and weighs only 102 g (3.6 oz).

mechanical gyroscopes. The drift of radio-control type gyroscopes is on the order of tens of degrees per minute.

2.1.3.2 Gyration, Inc.

The *GyroEngine* made by Gyration, Inc. [GYRATION], Saratoga, CA, is a low-cost mechanical gyroscope that measures changes in rotation around two independent axes. One of the original applications for which the GyroEngine was designed is the *GyroPoint*, a three-dimensional pointing device for manipulating a cursor in three-dimensional computer graphics. The *GyroEngine* model GE9300-C has a typical drift rate of about 9°/min. It weighs only 40 grams (1.5 oz) and compares in size with that of a roll of 35 millimeter film (see Figure 2.3). The sensor can be powered with 5 to 15 VDC and draws only 65

Figure 2.3: The Gyration *GyroEngine* compares in size favorably with a roll of 35 mm film (courtesy Gyration, Inc.).

to 85 mA during operation. The open collector outputs can be readily interfaced with digital circuits. A single *GyroEngine* unit costs $295.

2.2 Piezoelectric Gyroscopes

Piezoelectric vibrating gyroscopes use Coriolis forces to measure rate of rotation. in one typical design three piezoelectric transducers are mounted on the three sides of a triangular prism. If one of the transducers is excited at the transducer's resonance frequency (in the Gyrostar it is 8 kHz),

the vibrations are picked up by the two other transducers at equal intensity. When the prism is rotated around its longitudinal axis, the resulting Coriolis force will cause a slight difference in the intensity of vibration of the two measuring transducers. The resulting analog voltage difference is an output that varies linearly with the measured rate of rotation.

Figure 2.4: The Murata *Gyrostar* ENV-05H is a piezoelectric vibrating gyroscope. (Courtesy of [Murata]).

One popular piezoelectric vibrating gyroscope is the ENV-05 *Gyrostar* from [MURATA], shown in Fig. 2.4. The Gyrostar is small, lightweight, and inexpensive: the model ENV-05H measures 47×40×22 mm (1.9×1.6×0.9 inches), weighs 42 grams (1.5 oz) and costs $300. The drift rate, as quoted by the manufacturer, is very poor: 9°/s. However, we believe that this number is the worst case value, representative for extreme temperature changes in the working environment of the sensor. When we tested a Gyrostar Model ENV-05H at the University of Michigan, we measured drift rates under typical room temperatures of 0.05°/s to 0.25°/s, which equates to 3 to 15°/min (see [Borenstein and Feng, 1996]). Similar drift rates were reported by Barshan and Durrant-Whyte [1995], who tested an earlier model: the Gyrostar ENV-05S (see Section 5.4.2.1 for more details on this work). The scale factor, a measure for the useful sensitivity of the sensor, is quoted by the manufacturer as 22.2 mV/deg/sec.

2.3   Optical Gyroscopes

Optical rotation sensors have now been under development as replacements for mechanical gyros for over three decades. With little or no moving parts, such devices are virtually maintenance free and display no gravitational sensitivities, eliminating the need for gimbals. Fueled by a large

market in the automotive industry, highly linear fiber-optic versions are now evolving that have wide dynamic range and very low projected costs.

The principle of operation of the optical gyroscope, first discussed by Sagnac [1913], is conceptually very simple, although several significant engineering challenges had to be overcome before practical application was possible. In fact, it was not until the demonstration of the helium-neon laser at Bell Labs in 1960 that Sagnac's discovery took on any serious implications; the first operational ring-laser gyro was developed by Warren Macek of Sperry Corporation just two years later [Martin, 1986]. Navigation quality ring-laser gyroscopes began routine service in inertial navigation systems for the Boeing 757 and 767 in the early 1980s, and over half a million fiber-optic navigation systems have been installed in Japanese automobiles since 1987 [Reunert, 1993]. Many technological improvements since Macek's first prototype make the optical rate gyro a potentially significant influence on mobile robot navigation in the future.

The basic device consists of two laser beams traveling in opposite directions (i.e., counter propagating) around a closed-loop path. The constructive and destructive interference patterns formed by splitting off and mixing parts of the two beams can be used to determine the rate and direction of rotation of the device itself.

Schulz-DuBois [1966] idealized the ring laser as a hollow doughnut-shaped mirror in which light follows a closed circular path. Assuming an ideal 100-percent reflective mirror surface, the optical energy inside the cavity is theoretically unaffected by any rotation of the mirror itself. The counter-propagating light beams mutually reinforce each other to create a stationary standing wave of intensity peaks and nulls as depicted in Figure 2.5, regardless of whether the gyro is rotating [Martin, 1986].

A simplistic visualization based on the Schulz-DuBois idealization is perhaps helpful at this point in understanding the fundamental concept of operation before more detailed treatment of the subject is presented. The light and dark fringes of the nodes are analogous to the reflective stripes or slotted holes in the rotating disk of an incremental optical encoder, and can be theoretically counted in similar fashion by a light detector mounted on the cavity wall. (In this analogy, however, the standing-wave "disk" is fixed in the inertial reference frame, while the normally stationary detector revolves around it.) With each full rotation of the mirrored doughnut, the detector would see a number of node peaks equal to twice the optical path length of the beams divided by the wavelength of the light.
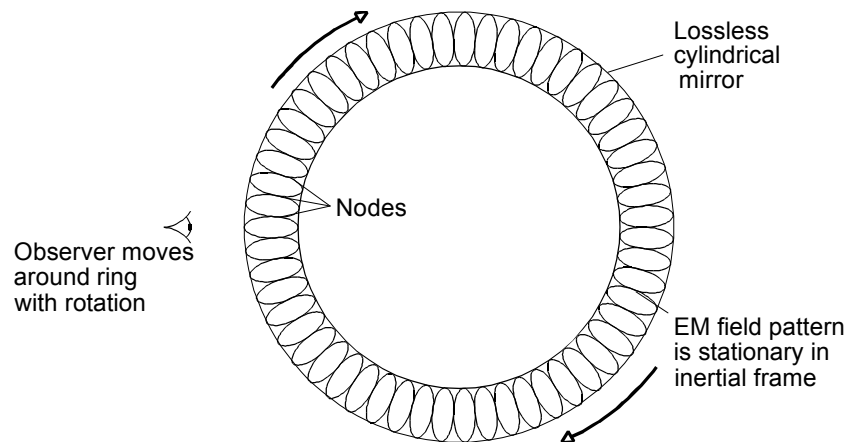

Figure 2.5: Standing wave created by counter-propagating light beams in an idealized ring-laser gyro. (Adapted from [Schulz-DuBois, 1966].)

Obviously, there is no practical way to implement this theoretical arrangement, since a *perfect* mirror cannot be realized in practice. Furthermore, the introduction of light energy into the cavity (as well as the need to observe and count the nodes on the standing wave) would interfere with the mirror's performance, should such an ideal capability even exist. However, many practical embodiments of optical rotation sensors have been developed for use as rate gyros in navigation applications. Five general configurations will be discussed in the following subsections:
- Active optical resonators (2.3.1).
- Passive optical resonators (2.3.2).
- Open-loop fiber-optic interferometers (analog) (2.3.3).
- Closed-loop fiber-optic interferometers (digital) (2.3.4).
- Fiber-optic resonators (2.3.5).

Aronowitz [1971], Menegozzi and Lamb [1973], Chow et al. [1985], Wilkinson [1987], and Udd [1991] provide in-depth discussions of the theory of the ring-laser gyro and its fiber-optic derivatives. A comprehensive treatment of the technologies and an extensive bibliography of preceding works is presented by Ezekial and Arditty [1982] in the proceedings of the First International Conference on Fiber-Optic Rotation Sensors held at MIT in November, 1981. An excellent treatment of the salient features, advantages, and disadvantages of ring laser gyros versus fiber optic gyros is presented by Udd [1985, 1991].

2.3.1 Active Ring Laser Gyros

The active optical resonator configuration, more commonly known as the ring laser gyro, solves the problem of introducing light into the doughnut by filling the cavity itself with an active *lazing* medium, typically helium-neon. There are actually two beams generated by the laser, which travel around the ring in opposite directions. If the gyro cavity is caused to physically rotate in the counterclockwise direction, the counterclockwise propagating beam will be forced to traverse a slightly longer path than under stationary conditions. Similarly, the clockwise propagating beam will see its closed-loop path shortened by an identical amount. This phenomenon, known as the *Sagnac effect*, in essence changes the length of the resonant cavity. The magnitude of this change is given by the following equation [Chow et al., 1985]:

$$\Delta L = \frac{4\pi r^2 \Omega}{c} \tag{2.2}$$

where
ΔL = change in path length
r = radius of the circular beam path
Ω = angular velocity of rotation
c = speed of light.

Note that the change in path length is directly proportional to the rotation rate Ω of the cavity. Thus, to measure gyro rotation, some convenient means must be established to measure the induced change in the optical path length.

This requirement to measure the difference in path lengths is where the invention of the laser in the early 1960s provided the needed technological breakthrough that allowed Sagnac's observations to be put to practical use. For lazing to occur in the resonant cavity, the round-trip beam path must

be precisely equal in length to an integral number of wavelengths at the resonant frequency. This means the wavelengths (and therefore the frequencies) of the two counter- propagating beams must change, as only oscillations with wavelengths satisfying the resonance condition can be sustained in the cavity. The frequency difference between the two beams is given by [Chow et al., 1985]:

$$\Delta f = \frac{2fr\Omega}{c} = \frac{2r\Omega}{\lambda} \qquad (2.3)$$

where
Δf = frequency difference
r = radius of circular beam path
Ω = angular velocity of rotation
λ = wavelength.

In practice, a doughnut-shaped ring cavity would be hard to realize. For an arbitrary cavity geometry, the expression becomes [Chow et al., 1985]:

$$\Delta f = \frac{4A\Omega}{P\lambda} \qquad (2.4)$$

where
Δf = frequency difference
A = area enclosed by the closed-loop beam path
Ω = angular velocity of rotation
P = perimeter of the beam path
λ = wavelength.

For single-axis gyros, the ring is generally formed by aligning three highly reflective mirrors to create a closed-loop triangular path as shown in Figure 2.6. (Some systems, such as Macek's early prototype, employ four mirrors to create a square path.) The mirrors are usually mounted to a monolithic glass-ceramic block with machined ports for the cavity bores and electrodes. Most modern three-axis units employ a square block cube with a total of six mirrors, each mounted to the center of a block face as shown in Figure 2.6. The most stable systems employ linearly polarized light and minimize circularly polarized components to avoid magnetic sensitivities [Martin, 1986].

The approximate quantum noise limit for the ring-laser gyro is due to spontaneous emission in the gain medium [Ezekiel and Arditty, 1982]. Yet, the ring-laser gyro represents the "best-case" scenario of the five general gyro configurations outlined above. For this reason the active ring-laser gyro offers the highest sensitivity and is perhaps the most accurate implementation to date.

The fundamental disadvantage associated with the active ring laser is a problem called *frequency lock-in*, which occurs at low rotation rates when the counter-propagating beams "lock" together in frequency [Chao et al., 1984]. This lock-in is attributed to the influence of a very small amount of backscatter from the mirror surfaces, and results in a deadband region (below a certain threshold of rotational velocity) for which there is no output signal. Above the lock-in threshold, output approaches the ideal linear response curve in a parabolic fashion.

The most obvious approach to solving the lock-in problem is to improve the quality of the mirrors to reduce the resulting backscatter. Again, however, perfect mirrors do not exist, and some finite

amount of backscatter will always be present. Martin [1986] reports a representative value as 10^{-12} of the power of the main beam; enough to induce frequency lock-in for rotational rates of several hundred degrees per hour in a typical gyro with a 20-centimeter (8-in) perimeter.
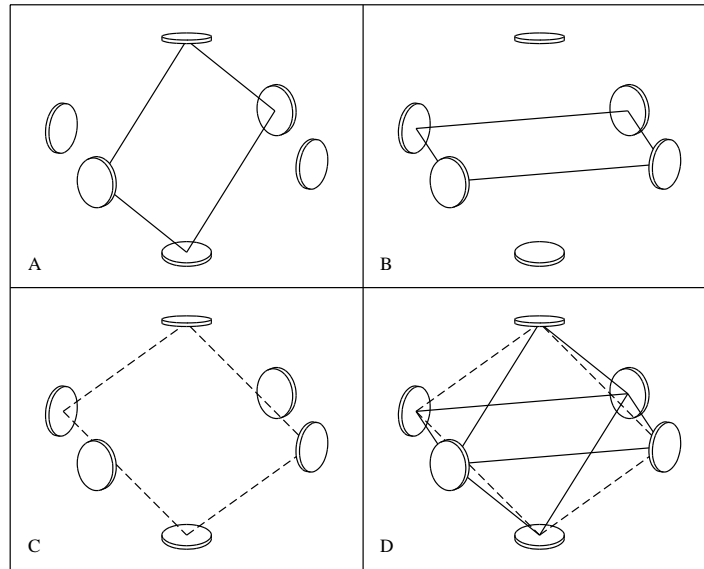

Figure 2.6: Six-mirror configuration of three-axis ring-laser gyro. (Adapted from [Koper, 1987].)

An additional technique for reducing lock-in is to incorporate some type of biasing scheme to shift the operating point away from the deadband zone. Mechanical dithering is the least elegant but most common biasing means, introducing the obvious disadvantages of increased system complexity and reduced mean time between failures due to the moving parts. The entire gyro assembly is rotated back and forth about the sensing axis in an oscillatory fashion. State-of-the-art dithered active ring laser gyros have a scale factor linearity that far surpasses the best mechanical gyros.

Dithered biasing, unfortunately, is too slow for high-performance systems (i.e., flight control), resulting in oscillatory instabilities [Martin, 1986]. Furthermore, mechanical dithering can introduce crosstalk between axes on a multi-axis system, although some unibody three-axis gyros employ a common dither axis to eliminate this possibility [Martin, 1986].

Buholz and Chodorow [1967], Chesnoy [1989], and Christian and Rosker [1991] discuss the use of extremely short duration laser pulses (typically 1/15 of the resonator perimeter in length) to reduce the effects of frequency lock-in at low rotation rates. The basic idea is to reduce the cross-coupling between the two counter-propagating beams by limiting the regions in the cavity where the two pulses overlap. Wax and Chodorow [1972] report an improvement in performance of two orders of magnitude through the use of intracavity phase modulation. Other techniques based on non-linear optics have been proposed, including an approach by Litton that applies an external magnetic field to the cavity to create a directionally dependent phase shift for biasing [Martin, 1986]. Yet another solution to the lock-in problem is to remove the lazing medium from the ring altogether, effectively forming what is known as a passive ring resonator.
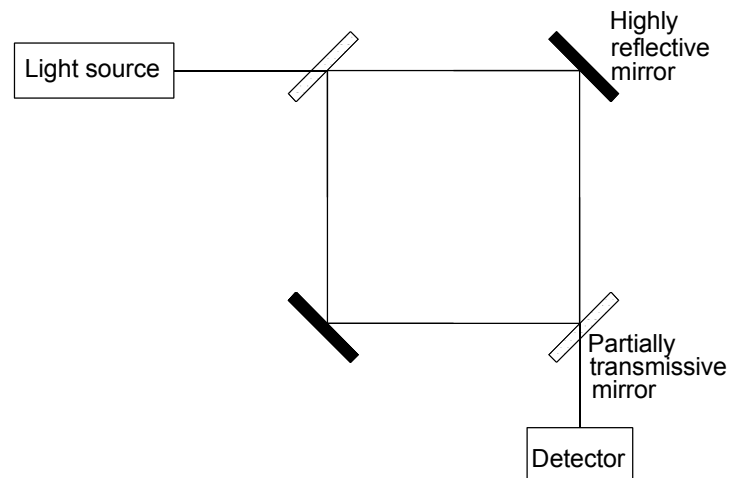
2.3.2 Passive Ring Resonator Gyros

Figure 2.7: Passive ring resonator gyro with laser source external to the ring cavity. (Adapted from [Udd, 1991].)

The passive ring resonator gyro makes use of a laser source external to the ring cavity (Figure 2.7), and thus avoids the frequency lock-in problem which arises when the gain medium is internal to the cavity itself. The passive configuration also eliminates problems arising from changes in the optical path length within the interferometer due to variations in the index of refraction of the gain medium [Chow et al., 1985]. The theoretical quantum noise limit is determined by photon shot noise and is slightly higher (i.e., worse) than the theoretical limit seen for the active ring-laser gyro [Ezekiel and Arditty, 1982].

The fact that these devices use mirrored resonators patterned after their active ring predecessors means that their packaging is inherently bulky. However, fiber-optic technology now offers a low volume alternative. The fiber-optic derivatives also allow longer length multi-turn resonators, for increased sensitivity in smaller, rugged, and less expensive packages. As a consequence, the Resonant Fiber-Optic Gyro (RFOG), to be discussed in Section 2.1.2.5, has emerged as the most popular of the resonator configurations [Sanders, 1992].

2.3.3 Open-Loop Interferometric Fiber Optic Gyros

The concurrent development of optical fiber technology, spurred mainly by the communications industry, presented a potential low-cost alternative to the high-tolerance machining and clean-room assembly required for ring-laser gyros. The glass fiber in essence forms an internally reflective waveguide for optical energy, along the lines of a small-diameter linear implementation of the doughnut-shaped mirror cavity conceptualized by Schulz-DuBois [1966].

Recall the refractive index n relates the speed of light in a particular medium to the speed of light in a vacuum as follows:

$$n = \frac{c}{c_m} \qquad (2.5)$$

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

Step-index multi-mode fiber (Figure 2.8) is made up of a core region of glass with index of refraction n_{co}, surrounded by a protective cladding with a lower index of refraction n_{cl} [Nolan and Blaszyk, 1991]. The lower refractive index in the cladding is necessary to ensure total internal reflection of the light propagating through the core region. The terminology step index refers to this "stepped" discontinuity in the refractive index that occurs at the core-cladding interface.
Referring now to Figure 2.8, as long as the entry angle (with respect to the waveguide axis) of an incoming ray is less than a certain critical angle θ_c, the ray will be guided down the fiber, virtually without loss. The numerical aperture of the fiber quantifies this parameter of acceptance (the light-collecting ability of the fiber) and is defined as follows [Nolan and Blaszyk, 1991]:

$$NA = \sin\theta_c = \sqrt{n_{co}^2 - n_{cl}^2} \qquad (2.6)$$

where
NA = numerical aperture of the fiber
θ_c = critical angle of acceptance
n_{co} = index of refraction of glass core
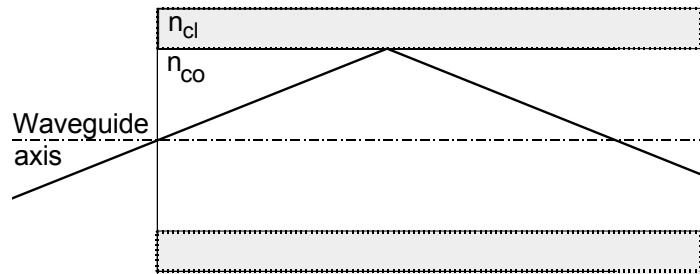n_{cl} = index of refraction of cladding.

Figure 2.8: Step-index multi-mode fiber. (Adapted from [Nolan et al., 1991].)

As illustrated in Figure 2.9, a number of rays following different-length paths can simultaneously propagate down the fiber, as long as their respective entry angles are less than the critical angle of acceptance θ_c. Multiple-path propagation of this nature occurs where the core diameter is much larger than the wavelength of the guided energy, giving rise to the term multi-mode fiber. Such multi-mode operation is clearly undesirable in gyro applications, where the objective is to eliminate all non-reciprocal conditions other than that imposed by the Sagnac effect itself. As the diameter of the core is reduced to approach the operating wavelength, a cutoff condition is reached where just a single mode is allowed to propagate, con-strained to travel only along the wave-guide axis [Nolan and Blaszyk, 1991].

Light can randomly change polariza-tion states as it propagates through stan-dard single-mode fiber. The use of special polarization-maintaining fiber, such as PRSM Corning, maintains the original polarization state of the light along the path of travel [Reunert, 1993]. This is important, since light of different polariza-tion states travels through an optical fiber at different speeds.
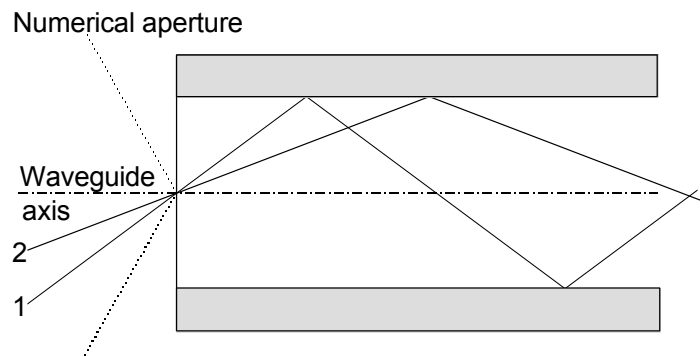

Figure 2.9: Entry angles of incoming rays 1 and 2 determine propagation paths in fiber core. (Adapted from [Nolan et al., 1991].)

A typical block diagram of the "minimum-reciprocal" IFOG configuration is presented in Figure 2.10. Polarization-maintaining single-mode fiber [Nolan and Blaszyk, 1991] is employed to ensure the two counter-propagating beams in the loop follow identical paths in the absence of rotation.

An interesting characteristic of the IFOG is the absence of any laser source [Burns et al., 1983], the enabling technology allowing the Sagnac effect to reach practical implementation in the first place. A low-coherence source, such as a super-luminescent diode (SLD), is typically employed instead to reduce the effects of noise [Tai et al., 1986], the primary source of which is backscattering within the fiber and at any interfaces. As a result, in addition to the two primary counter-propagating waves in the loop, there are also a number of parasitic waves that yield secondary interferometers [Lefevre, 1992]. The limited temporal coherence of the broadband SLD causes any interference due to backscattering to average to zero, suppressing the contrast of these spurious interferometers. The detection system becomes sensitive only to the interference between waves that followed identical paths [Ezekiel and Arditty, 1982; Lefevre, 1992].

The Sagnac phase shift introduced by rotation is given by [Ezekiel and Arditty, 1982]

$$\Delta\phi = \frac{2\pi LD}{\lambda c} \tag{2.7}$$

where
$\Delta\phi$ = measured phase shift between counter-propagating beams
L = length of fiber-optic cable in loop
D = diameter of loop
λ = wavelength of optical energy
c = speed of light in a vacuum.

The stability of the scale factor relating $\Delta\phi$ to the rotational velocity in the equation above is thus limited to the stability of L, D, and λ [Ezekiel and Arditty, 1982]. Practical implementations usually operate over plus or minus half a fringe (i.e., $\pm\pi$ rad of phase difference), with a theoretical sensitivity of 10^{-6} radians or less of phase shift [Lefevre, 1992].

IFOG sensitivity may be improved by increasing L (i.e., adding turns of fiber in the sensing loop). This effect peaks at an optimal length of several kilometers, after which the fiber attenuation (typically 1 dB/km) begins to degrade performance. This large amount of fiber represents a significant percentage of overall system cost.
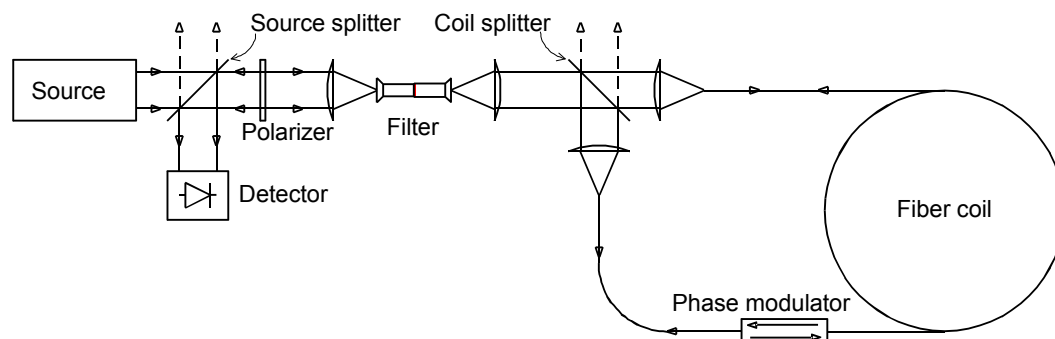

Figure 2.10: Block diagram of "minimum-reciprocal" integrated fiber-optic gyro. (Adapted from [Lefevre, 1992].)

In summary, the open-loop IFOG is attractive from the standpoint of reduced manufacturing costs. Additional advantages include high tolerance to shock and vibration, insensitivity to gravity effects, quick start-up, and good sensitivity in terms of bias drift rate and the random walk coefficient. Coil geometry is not critical, and no path length control is needed. Some disadvantages are that a long optical cable is required, dynamic range is limited with respect to active ring-laser gyros, and the scale factor is prone to vary [Adrian, 1991]. Open-loop configurations are therefore most suited to the needs of low-cost systems in applications that require relatively low accuracy (i.e., automobile navigation).

For applications demanding higher accuracy, such as aircraft navigation (0.01 to 0.001°/hr), the closed-loop IFOG to be discussed in the next section offers significant promise.

2.3.4 Closed-Loop Interferometric Fiber Optic Gyros

This new implementation of a fiber-optic gyro provides feedback to a frequency or phase shifting element. The use of feedback results in the cancellation of the rotationally induced *Sagnac phase shift*. However, closed-loop digital signal processing is considerably more complex than the analog signal processing employed on open-loop IFOG configurations [Adrian, 1991]. Nonetheless, it now seems that the additional complexity is justified by the improved stability of the gyro: closed-loop IFOGs are now under development with drifts in the 0.001 to 0.01°/hr range, and scale-factor stabilities greater than 100 ppm (parts per million) [Adrian, 1991].

2.3.5 Resonant Fiber Optic Gyros

The *resonant fiber optic gyro* (RFOG) evolved as a solid-state derivative of the passive ring resonator gyro discussed in Section 2.1.2.2. In the solid-state implementation, a passive resonant cavity is formed from a multi-turn closed loop of optical fiber. An input coupler provides a means for injecting frequency-modulated light from a laser source into the resonant loop in both the clockwise and counterclockwise directions. As the frequency of the modulated light passes through a value such that the perimeter of the loop precisely matches an integral number of wavelengths at that frequency, input energy is strongly coupled into the loop [Sanders, 1992]. In the absence of loop rotation, maximum coupling for both beam directions occurs in a sharp peak centered at this resonant frequency.

If the loop is caused to rotate in the clockwise direction, of course, the *Sagnac effect* causes the perceived loop perimeter to lengthen for the clockwise-traveling beam, and to shorten for the counterclockwise-traveling beam. The resonant frequencies must shift accordingly, and as a result, energy is coupled into the loop at two different frequencies and directions during each cycle of the sinusoidal FM sweep. An output coupler samples the intensity of the energy in the loop by passing a percentage of the two counter-rotating beams to their respective detectors. The demodulated output from these detectors will show resonance peaks, separated by a frequency difference *f* given by the following [Sanders, 1992]:

$$\Delta f = \frac{D}{\lambda n}\Omega \tag{2.8}$$

where
Δf = frequency difference between counter-propagating beams
D = diameter of the resonant loop

Ω = rotational velocity
λ = freespace wavelength of laser
n = refractive index of the fiber.

   Like the IFOG, the all-solid-state RFOG is attractive from the standpoint of high reliability, long life, quick start-up, and light weight. The principle advantage of the RFOG, however, is that it requires significantly less fiber (from 10 to 100 times less) in the sensing coil than the IFOG configuration, while achieving the same shot-noise-limited performance [Sanders, 1992]. Sanders attributes this to the fact that light traverses the sensing loop multiple times, as opposed to once in the IFOG counterpart. On the down side are the requirements for a highly coherent source and extremely low-loss fiber components [Adrian, 1991].

2.3.6 Commercially Available Optical Gyroscopes

Only recently have optical fiber gyros become commercially available at a price that is suitable for mobile robot applications. In this section we introduce two such systems.

2.3.6.1 The Andrew "Autogyro"

Andrew Corp. [ANDREW] offers the low-cost *Autogyro*, shown in Figure 2.11, for terrestrial navigation. It is a single-axis interferometric fiber-optic gyroscope (see Sec. 2.1.2.3) based on polarization-maintaining fiber and precision fiber-optic gyroscope technology. Model 3ARG-A ($950) comes with an analog output, while model 3ARG-D ($1,100) has an RS-232 output for connection to a computer. Technical specifications for the 3ARG-D are given in Table 2.1. Specifications for the 3ARG-A are similar. A more detailed discussion of the *Autogyro* is given

Table 2.1: Selected specifications for the Andrew *Autogyro* Model 3ARG-D. (Courtesy of [Andrew Corp].)

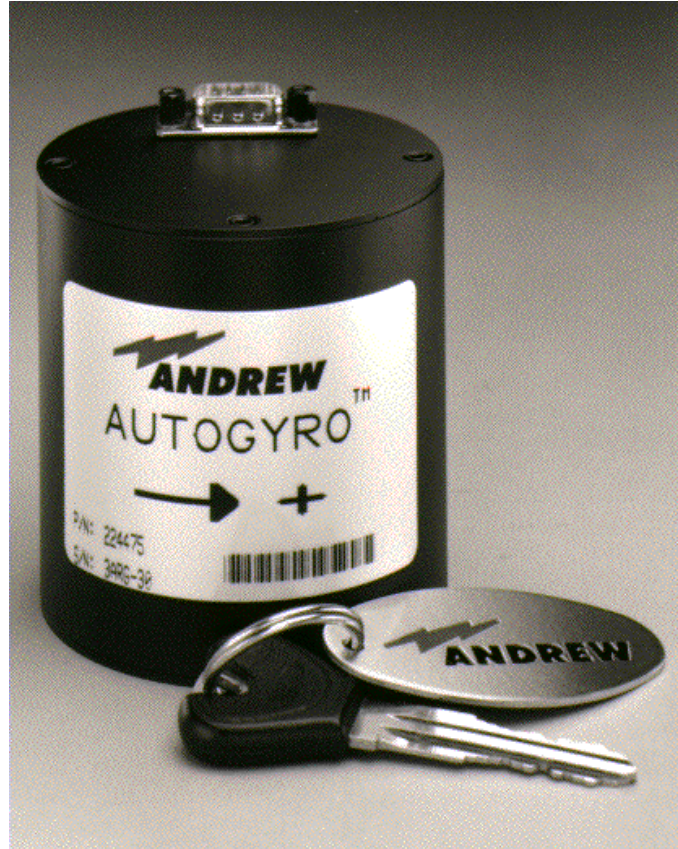| Parameter	Value	Units
Input rotation rate	±100	°/s
Minimum detectable	±0.05	°/s
rotation rate	±180	°/hr
Rate bandwidth	100	Hz
Bias drift (at stabilized	0.005	°/s rms
temperature) — RMS	18	°/hr rms
Size	77 dia × 88	mm
(excluding connector)	3.0 dia × 3.5	in
Weight (total)	0.63	kg
	1.38	lb
Power	9 to 18	VDC
	630	mA

Figure 2.11: The Andrew *Autogyro* Model 3ARG. (Courtesy of [Andrew Corp].)

Table 2.1: Selected specifications for the Andrew *Autogyro Navigator* (Courtesy of [Andrew Corp].)

Parameter	Value	Units
Input rotation rate	±100	°/s
Instantaneous bandwidth	100	Hz
Bias drift (at stabilized temperature) — RMS	0.005	°/s rms
	18	°/hr rms
Size (excluding connector)	115×90×41	mm
	4.5×3.5×1.6	in
Weight (total)	0.25	kg
	0.55	lb
Power Analog	< 2	W
Power Digital	< 3	W

in [Allen et al., 1994; Bennett and Emge, 1994].

In fall 1995 Andrew Corporation announced a newer model, called the *AUTOGYRO Navigator*. This laser gyro, shown in Fig. 2.12, is only one third the weight, consume only half the power, and cost 15% less than its predecessor, the AUTOGYRO.
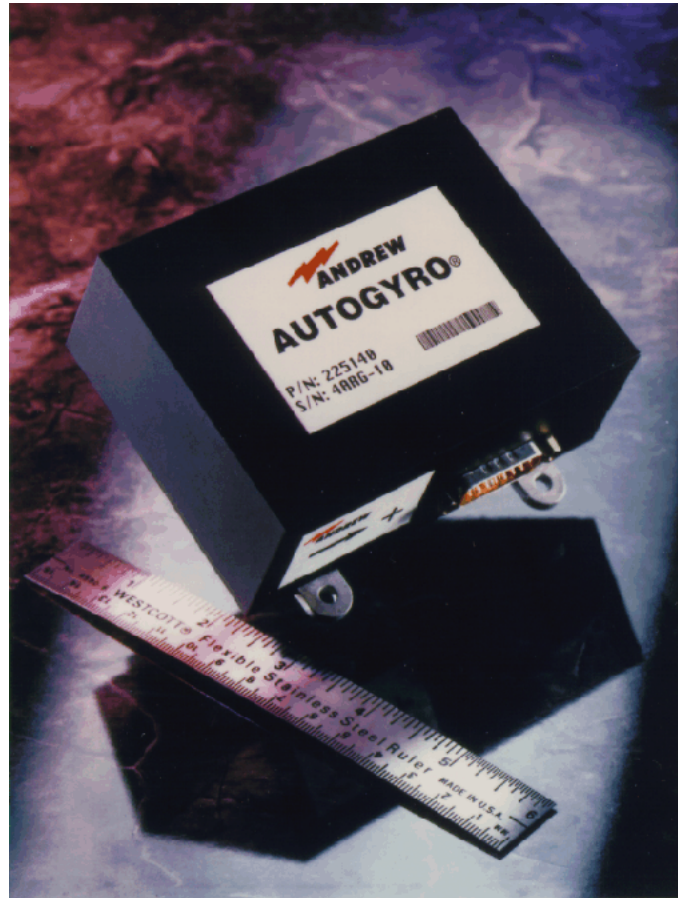
Figure 2.12: The Andrew *AUTOGYRO Navigator*. (Courtesy of [Andrew Corp].)

2.3.6.2 Hitachi Cable Ltd. OFG-3

Hitachi Cable Ltd. markets an optical fiber gyroscope called OFG-3 (see Figure 2.13). Komoriya and Oyama [1994] tested that sensor and found its drift rate to be quite linear with 0.00317°/s (11.4°/hr). This result is close to the advertised specification of 10°/hr. This low drift rate is substantially better than that provided by conventional (mechanical) gyros. Table 2.2 shows technical specifications of the OFG-3 gyro, as reported by Komoriya and Oyama [1994].

One point to keep in mind when considering the use of fiber optic gyros in mobile robot applications is the minimum detectable rotation rate. This rate happens to be the same for both the Andrew 3ARG-A and the Hitachi OFG-3 gyros: 0.05°/s. If either gyro was installed on a robot with a systematic error (e.g., due to unequal wheel diameters; see Sec. 5.1 for more details) of 1 degree per 10 meter linear travel, then neither gyro would detect this systematic error at speeds lower than 0.5 m/s.

Table 2.2: Selected specifications for the Hitachi Cable Ltd. OFG-3 fiber optic gyroscope. (Reprinted with permission from [Komoriya and Oyama, 1994].)

Parameter	Value	Units
Input rotation rate	±100	°/s
Minimum detectable rotation rate	±0.05	°/s
	±60	°/hr
Min. sampl. interval	10	ms
Zero drift (rate integration)	0.0028	°/s
	10	°/hr
Size	88(W)×88(L)×65(H)	mm
	3.5(W)×3.5(L)×2.5(H)	in
Weight (total)	0.48	kg
	1.09	lb
Power	12	VDC
	150-250	mA

Figure 2.13: The OFG-3 optical fiber gyro made by Hitachi Cable Ltd. (Courtesy of Hitachi Cable America, Inc. [HITACHI].)

2.4 Geomagnetic Sensors

Vehicle heading is the most significant of the navigation 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 navigation 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 Tesla = 10^4 Gauss = 10^9 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 magnetic north is known as *declination* and varies with both time and geographical location. Corrective values are routinely provided in the form of 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.
- Hall-effect compasses.
- Magnetoresistive compasses.
- Magnetoelastic compasses.

Before we introduce different types of compasses, a word of warning: the earth's magnetic field is often distorted near power lines or steel structures [Byrne et al., 1992]. This makes the straightforward use of geomagnetic sensors difficult for indoor applications. However, it may be possible to overcome this problem in the future by fusing data from geomagnetic compasses with data from other sensors.

2.4.1 Mechanical Magnetic Compasses

The first recorded use of a magnetic compass was in 2634 B.C., 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 navigation reference for ships at sea.

The early marine compasses were little more that 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 glycerine, 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 motion under rough conditions.

The sealed vessel containing the compass disk and damping fluid is typically suspended from a 2-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*. 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 abnormalities 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*, and along with local declination, must be added or subtracted as appropriate for true heading:

$$H_t = H_i \pm CF_{\text{dev}} \pm CF_{\text{dec}} \tag{2.9}$$

where
H_t = true heading
H_i = indicated heading
CF_{dev} = correction factor for compass deviation
CF_{dec} = correction factor for magnetic declination.

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. Marine compasses ensure alignment in the horizontal plane by floating the magnet assembly in an inert fluid.

Dinsmore *Starguide* Magnetic Compass

An extremely low-cost configuration of the mechanical magnetic compass suitable for robotic applications 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 Instrument Company, 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.

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 the adjacent cardinal-heading switches. The Dinsmore *Starguide* is not a true Hall-effect compass (see Sec. 2.4.3), 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 12.5 millimeters (0.5 in) in diameter by 16 millimeters (0.63 in) high, and is available separately from Dinsmore for around $12. Current consumption is 30 mA, and the open-collector NPN outputs can sink 25 mA 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. Both sensors may be subjected to unlimited magnetic flux without damage.

2.4.2 Fluxgate Compasses

There currently is no practical alternative to the popular fluxgate compass for portability and long missions [Fenn et al., 1992]. 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 2.3.

Permeability is the magnetic circuit analogy to electrical conductivity, and relates magnetic flux density to the magnetizing force as follows:

$$B = \mu H \qquad (2.10)$$

where
B = magnetic flux density
μ = permeability
H = magnetizing force.

Table 2.3: Permeability ranges for selected materials. Values vary with proportional make-up, heat treatment, and mechanical working of the material [Bolz and Tuve, 1979].

Material	Permeability μ
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

Since the magnetic flux 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 2.14 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 d*H*/d*t*, 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-zero permeability.
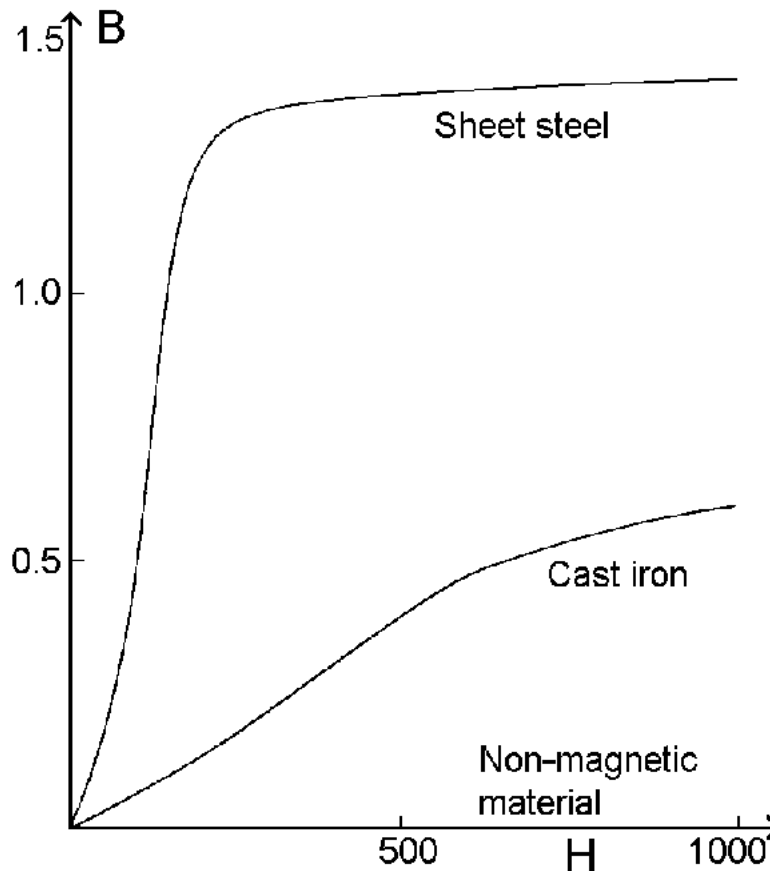
Figure 2.14: The slope of the B-H curve, shown here for cast iron and sheet steel, describes the permeability of a magnetic material, a measure of its ability (relative to air) to conduct a magnetic flux. (Adapted from [Carlson and Gisser, 1981].)

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 2.15. 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 2.15b. 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 50 years, with the two most common being permalloy (an alloy of iron and nickel) and mumetal (iron, nickel,
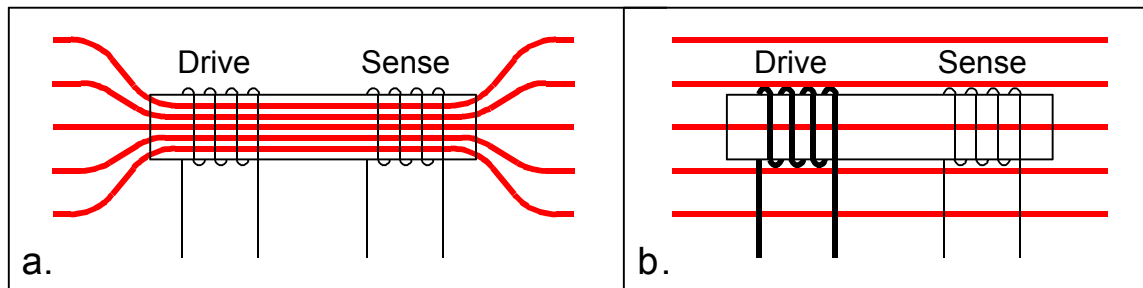
Figure 2.15: External lines of flux for: a. unsaturated core, b. saturated core. (Adapted from [Lenz, 1990].)

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 2.15. (This is somewhat of an oversimplification, in that the B-H curve does not fully flatten out with zero slope after the knee.) These   expanding   and collapsing flux lines will induce positive and negative EMF surges in a sensing coil properly oriented around the core. 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 2.16. 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 2.16 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.

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 is developed 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"
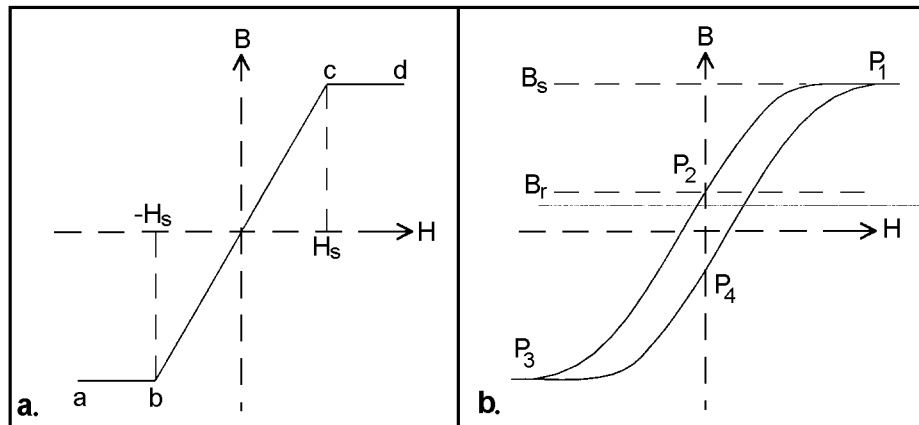
Figure 2.16: a. Ideal B-H curve.
b. Some minor hysteresis in the actual curve results in a residual non-zero
value of B when H is reduced to zero, known as the retentivity. (Adapted from
Halliday and Resnick, 1974; Carlson and Gisser, 1981).

capability of the core is substantially reduced.

Consider again the cylindrical rod sensor presented in Figure 2.17, now in the absence of any external magnetic field B_e. When the drive coil is energized, there will be a strong coupling between the drive coil and the sense coil. Obviously, this will be an undesirable situation since the output signal is supposed to be related to the strength of the external field only.

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 [Primdahl, 1979] as illustrated in Figure 2.17. The two rods are simultaneously forced into and out of saturation, excited in antiphase by identical but oppositely wound solenoidal drive windings. In this fashion, the magnetization fluxes of the two drive windings effectively cancel each other, with no net effect on the sense coil.

Bridges of magnetic material may be employed to couple the ends of the two coils 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]. An alternative technique for decoupling the pickup coil from the drive coil is to arrange the two in an orthogonal fashion. 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 are generally divided into two classes, parallel and orthogonal, depending on whether the
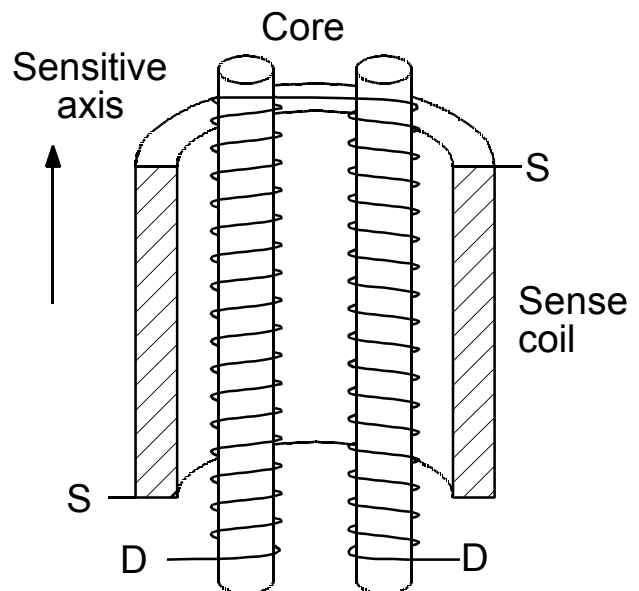

Figure 2.17: Identical but oppositely wound drive windings in the Vacquier configuration cancel the net effect of drive coupling into the surrounding sense coil, while still saturating the core material. (Adapted from [Primdahl, 1979].)

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.

Figure 2.18: Two channel ring core fluxgate with toroidal excitation. (Adapted from [Acuna and Pellerin, 1969].)

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 pickup coils can be configured in a symmetrical fashion around a common core. A follow-up version developed by Gordon and Lundsten [1970] employed a toroidal excitation winding as shown in Figure 2.19. 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 by rotating the circular core. 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.

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:
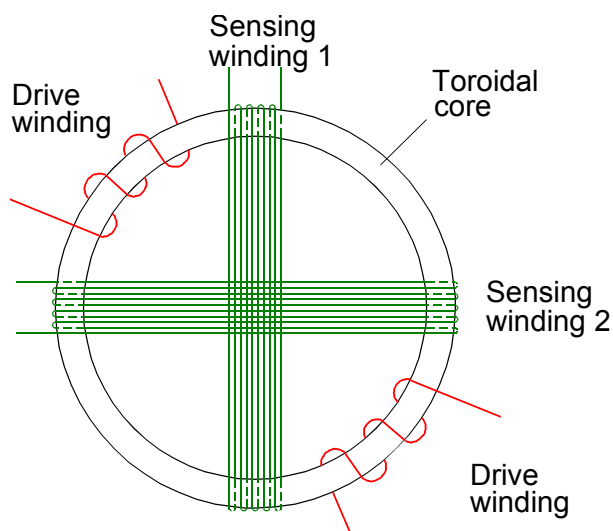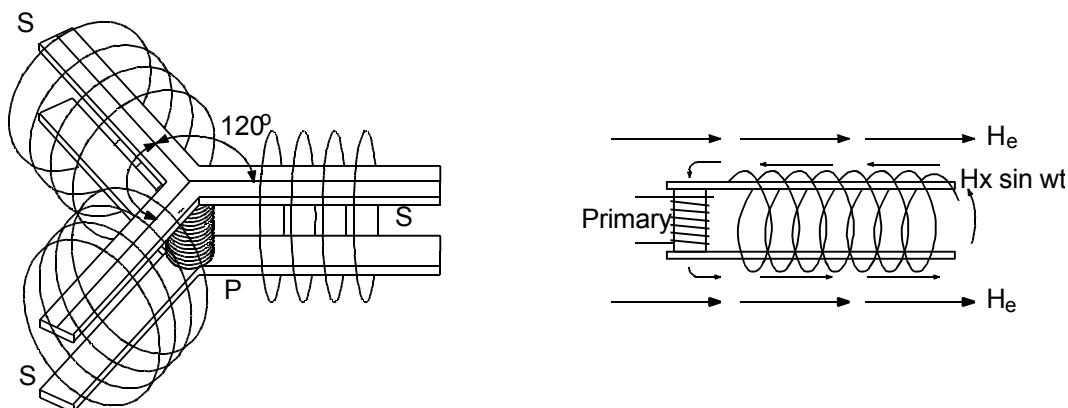

Figure 2.19: The Sperry Flux Valve consisted of a common drive winding P in the center of three sense windings S symmetrically arranged 120 ° apart. (Adapted from [Hine, 1968].)

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

(2.11)

Another popular two-axis core design is seen in the Flux Valve magnetometer developed by Sperry Corp. [SPERRY] and shown in Figure 2.19. 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 2.20, the upper and lower "arms" of the sense coil S are excited by the driving coil D, so that 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.

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 (see Figure 2.20), the angular relationship of the Flux Valve with respect to the external field can be unambiguously determined.
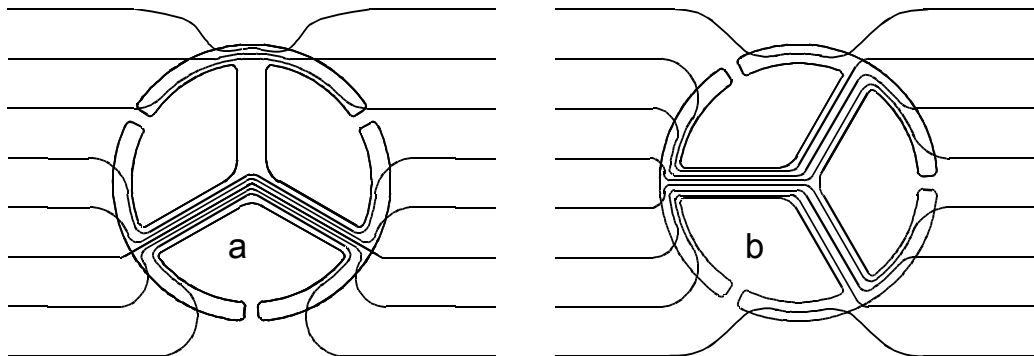

Figure 2.20: The Flux Valve magnetometer developed by Sperry Corporation uses a spider-core configuration. (Adapted from [Lenz, 1990].)

When maintained in a level attitude, the 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 ad vibration, rapid start-up, and relatively low cost. If the vehicle is expected to operate over uneven terrain, the sensor coil should be gimbal-mounted and mechanically dampened to prevent serious errors introduced by the vertical component of the geomagnetic field.

2.4.2.1 Zemco Fluxgate Compasses

The Zemco fluxgate compass [ZEMCO] was used in earlier work by Everett et al. [1990] on their robot called ROBART II. The sensor 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 VDC operation in automobiles.

A system block diagram is presented in Figure 2.21. The sensor consists of two orthogonal pickup coils arranged around a toroidal excitation coil, driven in turn by a local 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 Corporation, 1982] for these two voltages are

$$V_x = K_x \sin\theta \, \sin(\omega t + a_x) \qquad\qquad\qquad (2.12a)$$

$$V_y = K_y \cos\theta \, \sin(\omega t + a_y) \qquad\qquad\qquad (2.12b)$$

where
θ = the resolver shaft angle
ω = $2\pi f$, where f is the excitation frequency.
K_x and K_y are ideally equal transfer-function constants, and a_x and a_y are ideally zero time-phase shifts.

Thus, for any static spatial angle θ, the equations reduce to

$$V_x = K_x \sin\theta \qquad\qquad\qquad\qquad\qquad (2.13a)$$

$$V_y = K_y \cos\theta \qquad\qquad\qquad\qquad\qquad (2.13b)$$

which can be combined to yield

$$\frac{V_x}{V_Y} = \frac{\sin\theta}{\cos\theta} = \tan\theta \ . \qquad (2.14)$$

The magnetic heading θ therefore is simply the arctangent of V_x over V_y.

Everett [1995] recounts his experience with two models of the Zemco fluxgate compass on *ROBART* II as follows:

Problems associated with the use of this particular fluxgate compass on ROBART, however, included a fairly high current consumption (250 mA), and stiction in the resolver reflecting back as a load into the drive circuitry, introducing some error for minor
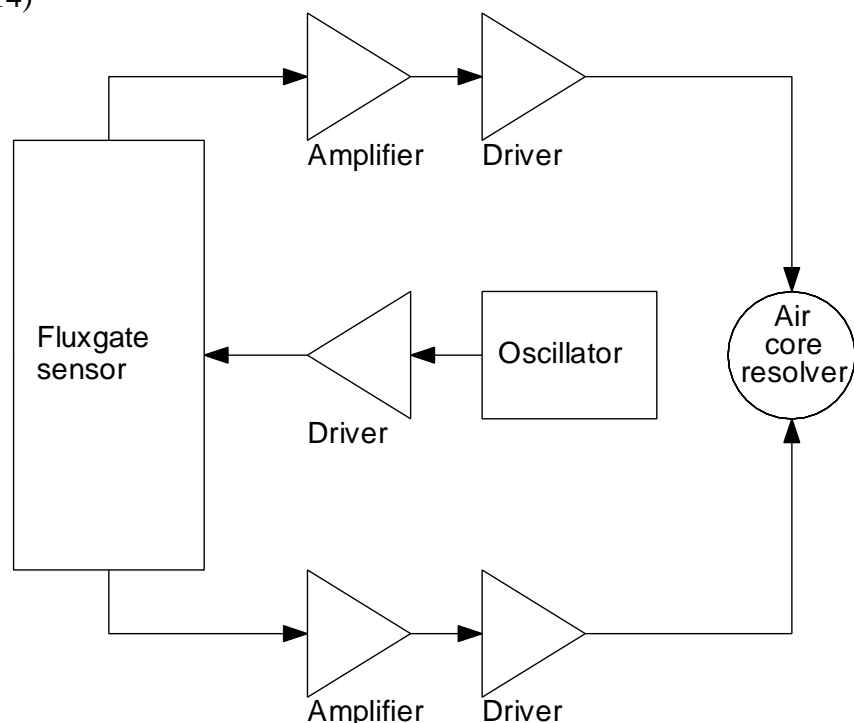

Figure 2.21: Block diagram of ZEMCO Model DE-700 fluxgate compass. (Courtesy of ZEMCO, Inc.)

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 30 centimeters (12 in) — to the fluxgate sensor. As various auxiliary systems on board the robot were turned on when needed and later deactivated to save power, the magnetic field surrounding the sensor would change accordingly. Serious 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 was no solution, however, for the interference effects of large metallic objects within the operating environment, and deviations of approximately four degrees were observed when passing within 30 centi-meters (12 in) of a large metal cabinet, for example. A final source of error was introduced by virtue of the fact that the fluxgate compass had been mounted on the robot's head, so as to be 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, which was not the original intent of the compass.

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 2.22. 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° between display increments, but provision
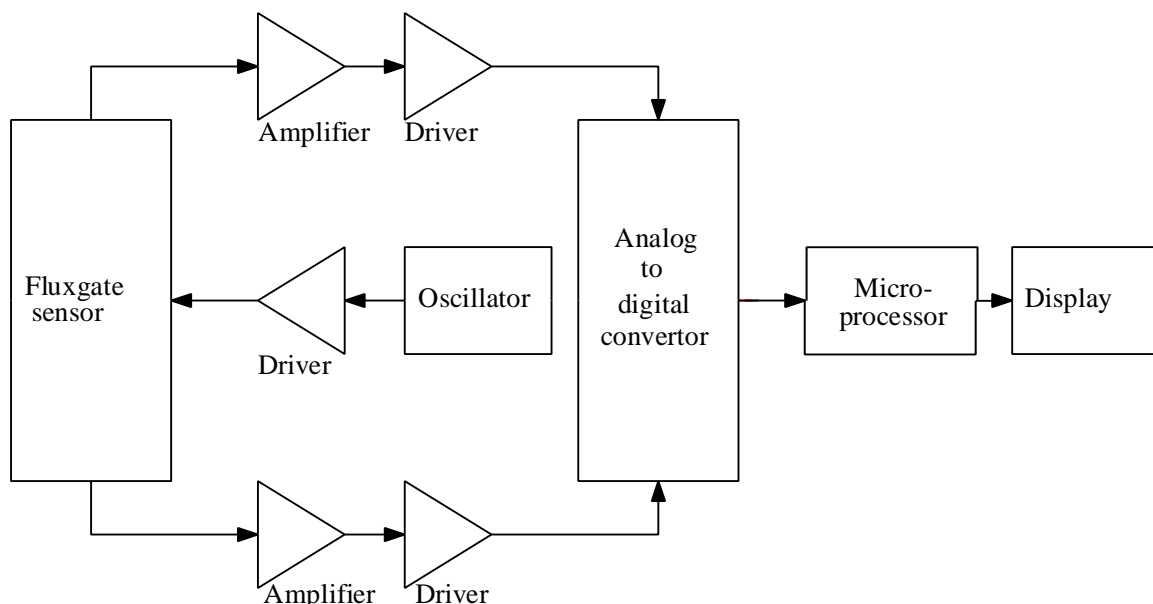

Figure 2.22: Block diagram of ZEMCO model DE-710 fluxgate compass (courtesy ZEMCO, Inc.).

was made for serial output to an optional external shift register and associated three-digit numerical display.

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 resulted in a decrease in supply current from 168 to 94 mA. 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 the 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.

2.4.2.2 Watson Gyrocompass

A combination fluxgate compass and solid-state rate gyro package (part number FGM-G100DHS-RS232) is available from Watson Industries, Eau Claire, WI [WATSON]. The system contains its own microprocessor that is intended to integrate the information from both the rate gyro and the compass to provide a more stable output less susceptible to interference, with an update rate of 40 Hz. An overall block diagram is presented in Figure 2.23.
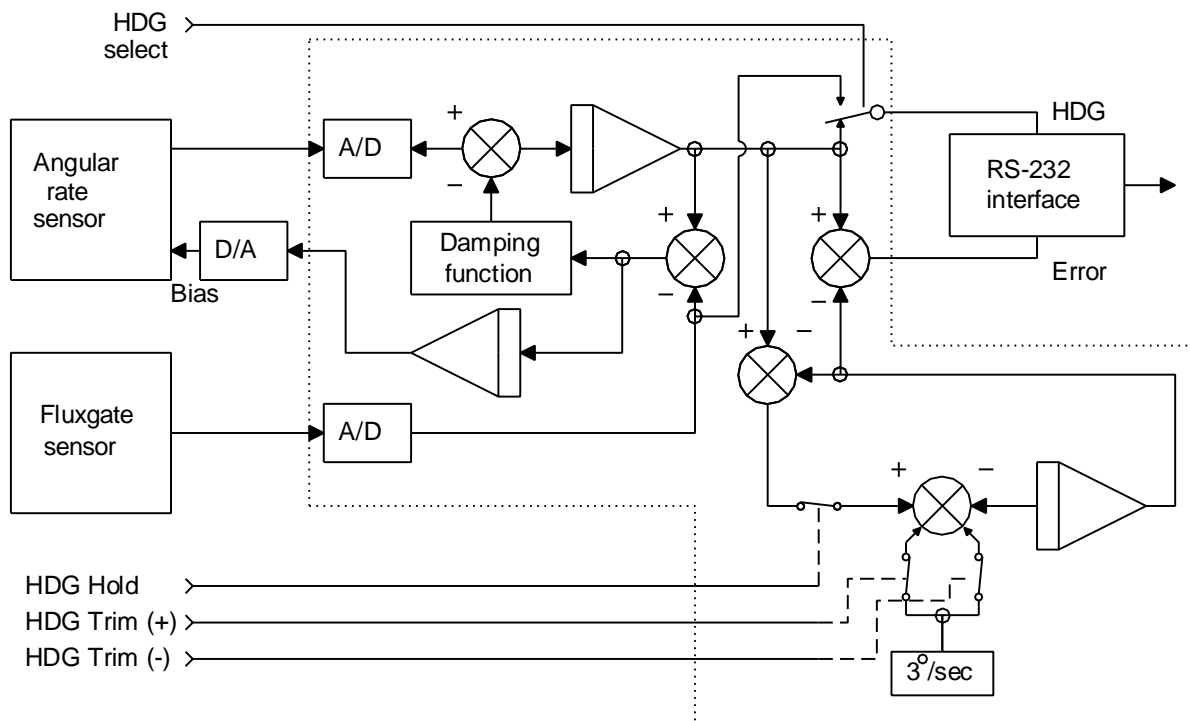

Figure 2.22: Block diagram of Watson fluxgate compass and rate gyro combination. (Courtesy of [WATSON].)

The Watson fluxgate/rate gyro combination balances the shortcomings of each type of device: the gyro serves to filter out the effects of magnetic anomalies in the surrounding environment, while the compass counters the long-term drift of the gyro. Furthermore, the toroidal ring-core fluxgate sensor is gimbal-mounted for improved accuracy.

The Watson unit measures 6.3×4.4×7.6 centimeters (2.5×1.75×3.0 in) and weighs only 275 grams (10 oz). This integrated package is a much more expensive unit ($2,500) than the low-cost Zemco fluxgate compass, but is advertised to have higher accuracy (±2°). Power supply requirements are 12 VDC at 200 mA, and the unit provides an analog voltage output as well as a 12-bit digital output over a 2400-baud RS-232 serial link.

2.4.2.3 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]. The C100 COMPASS ENGINE (see Figure 2.24) is a versatile low-cost (less than $700) developer's kit that includes a microprocessor-controlled stand-alone fluxgate sensor subsystem based on a two-axis toroidal ring-core sensor.
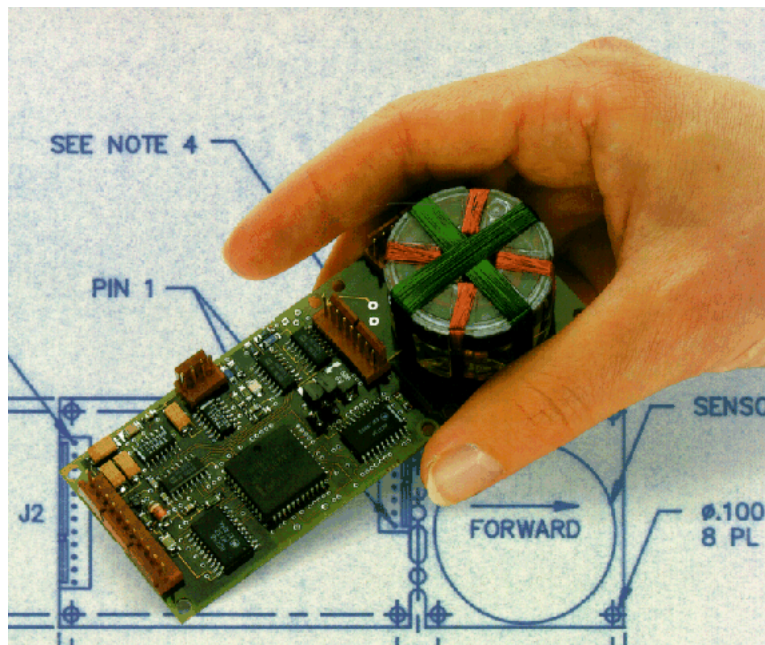

Figure 2.24: The C-100 fluxgate compass engine was tested at the University of Michigan in a *flying robot* prototype. (Courtesy of [KVH].)

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 anticipating a tilt angle of up to ±45 degrees. The SE-25 sensor provides internal gimballing by floating the sensor coil in an inert fluid inside the lexan housing. The SE-10 sensor provides an additional 2-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 PC board can be

separated as much as 122 centimeters (48 in) from the detachable electronics PC board with an optional cable if so desired.

The resolution of the C100 is ±0.1 degrees, with an advertised accuracy of ±0.5 degrees (after compensation, with the sensor card level) and a repeatability of ±0.2 degrees. Separate ±180 degree adjustments are provided for *declination* 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 corrects for magnetic anomalies associated with the host vehicle. Three alternative user-selectable procedures are offered:

- *Eight-Point Auto-Compensation* — starting from an arbitrary heading, the platform turns full circle, pausing momentarily at approximately 45-degree intervals. No known headings are required.
- *Circular Auto-Compensation* — Starting from an arbitrary position, the platform turns slowly through a continuous 360-degree circle. No known headings are required.
- *Three-Point Auto-Compensation* — 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 Industries, 1993].

The C100 configured with an SE-25 coil assembly weighs just 62 grams (2.25 oz) and draws 40 mA at 8 to 18 VDC (or 18 to 28 VDC). The combined sensor and electronics boards measure 4.6×11 centimeters (1.8×4.5 in). RS-232 (300 to 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.

2.4.3 Hall-Effect Compasses

Hall-effect sensors are based on E. H. Hall's observation (in 1879) that a DC voltage develops across a conductor or semiconductor when in the presence of an external magnetic field. One advantage of this technology (i.e., relative to the fluxgate) is the inherent ability to directly sense a static flux, resulting in much simpler readout electronics. Early Hall magnetometers could not match the sensitivity and stability of the fluxgate [Primdahl, 1979], but the sensitivity of Hall devices has improved significantly. The more recent indium-antimonide devices have a lower sensitivity limit of 10^{-3} Gauss [Lenz, 1990].

The U.S. Navy in the early 1960s showed considerable interest in a small solid-state Hall-effect compass for low-power extended operations in sonobuoys [Wiley, 1964]. A number of such prototypes were built and delivered by Motorola for evaluation. The Motorola Hall-effect compass employed two orthogonal Hall elements for temperature-nulled non-ambiguous resolution of the geomagnetic field vector. Each sensor element was fabricated from a 2×2×0.1 millimeter indium-arsenide-ferrite sandwich, and inserted between two wing-like mumetal flux concentrators as shown in Figure 2.25. It is estimated the 5 centimeter (2 in) magnetic concentrators increased the flux density through the sensing elements by two orders of magnitude [Wiley, 1964]. The output of the Motorola unit was a variable-width pulse train, the width of the pulse being proportional to the
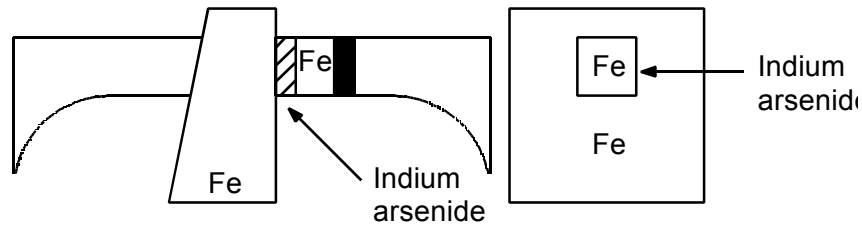
Figure 2.25: A pair of indium-arsenide-ferrite Hall-effect sensors (one shown) are positioned between flux concentrating wings of mumetal in this early Motorola prototype. (Adapted from [Wiley, 1964].)

sensed magnetic heading. Excellent response linearity was reported down to flux densities of 0.001 Gauss [Willey, 1962].

Maenaka et al. [1990] report on the development of a monolithic silicon magnetic compass at the Toyohashi University of Technology in Japan, based on two orthogonal Hall-effect sensors. Their use of the terminology "magnetic compass" is perhaps an unfortunate misnomer in that the prototype device was tested with an external field of 1,000 Gauss. Contrast this with the strength of the earth's magnetic field, which varies from only about 0.1 Gauss at the equator to about 0.9 Gauss at the poles. Silicon-based Hall-effect sensors have a lower sensitivity limit of around 10 Gauss [Lenz, 1990]. It is likely the Toyohashi University device was intended for other than geomagnetic applications, such as remote position sensing of rotating mechanical assemblies.

This prototype Hall-effect magnetometer is still of interest in that it represents a fully self-contained implementation of a two-axis magnetometer in integrated circuit form. Two vertical Hall cells [Maenaka et al., 1987] are arranged at right angles (see Figure 2.25) on a 4.7 mm² chip, with their respective outputs coupled to a companion signal processing IC of identical size. (Two separate chips were fabricated for the prototype instead of a single integrated unit to enhance production yield.) The sensor and signal processing ICs are interconnected (along with some external variable resistors for calibration purposes) on a glass-epoxy printed circuit board.

The dedicated signal-processing circuitry converts the B-field components B_x and B_y measured by the Hall sensors into an angle θ by means of the analog operation [Maenaka et al., 1990]:

$$\theta = \arctan\frac{B_x}{B_y}$$

(2.15)

where
θ = angle between B-field axis and sensor
B_x = x-component of B-field
B_y = y-component of B-field.

The analog output of the signal-processing IC is a DC voltage which varies linearly with vector orientation of the ambient magnetic field in a plane parallel to the chip surface. Reported test results show a fairly straight-line response (i.e., ± 2 percent full scale) for external field strengths ranging from 8,000 Gauss down to 500 Gauss; below this level performance begins to degrade rapidly [Maenaka et al., 1990]. A second analog output on the IC provides an indication of the absolute value of field intensity.

While the Toyohashi "magnetic compass" prototype based on silicon Hall-effect technology is incapable of detecting the earth's magnetic field, it is noteworthy nonetheless. A two-axis monolithic device of a similar nature employing the more sensitive indium-antimonide Hall devices could potentially have broad appeal for low-cost applications on mobile robotic platforms. An alternative possibility would be to use magnetoresistive sensor elements, which will be discussed in the next section.

2.4.4 Magnetoresistive Compasses

The general theory of operation for AMR and GMR magnetoresistive sensors for use in short-range proximity detection is beyond the scope of this text. However, there are three specific properties of the magnetoresistive magnetometer that make it well suited for use as a geomagnetic sensor: 1) high sensitivity, 2) directionality, and, in the case of AMR sensors, 3) the characteristic "flipping" action associated with the direction of internal magnetization.

AMR sensors have an open-loop sensitivity range of 10^{-2} Gauss to 50 Gauss (which easily covers the 0.1 to 1.0 Gauss range of the earth's horizontal magnetic field component), and limited-bandwidth closed-loop sensitivities approaching 10^{-6} Gauss [Lenz, 1990]. Excellent sensitivity, low power consumption, small package size, and decreasing cost make both AMR and GMR sensors increasingly popular alternatives to the more conventional fluxgate designs used in robotic vehicle applications.

2.4.4.1 Philips AMR Compass

One of the earliest magnetoresistive sensors to be applied to a magnetic compass application is the KMZ10B offered by Philips Semiconductors BV, The Netherlands [Dibburn and Petersen, 1983; Kwiatkowski and Tumanski, 1986; Petersen, 1989]. The limited sensitivity of this device (approximately 0.1 mV/A/m with a supply voltage of 5 VDC) in comparison to the earth's maximum horizontal magnetic field (15 A/m) means that considerable attention must be given to error-inducing effects of temperature and offset drift [Petersen, 1989].

One way around these problems is to exploit the "flipping" phenomenon by driving the device back and forth between its two possible magnetization states with square-wave excitation pulses applied to an external coil (Figure 2.26). This switching action toggles the sensor's axial magnetic field as shown in Figure 2.26a, resulting in the alternating response characteristics depicted in Figure 2.26b. Since the sensor offset remains unchanged while the signal output due to the external magnetic field H_y is inverted (Figure 2.26a), the undesirable DC offset voltages can be easily isolated from the weak AC signal.

A typical implementation of this strategy is shown in Figure 2.27. A 100 Hz square wave generator is capacitively coupled to the external excitation coil L which surrounds two orthogonally mounted magnetoresistive sensors. The sensors' output signals are amplified and AC-coupled to a
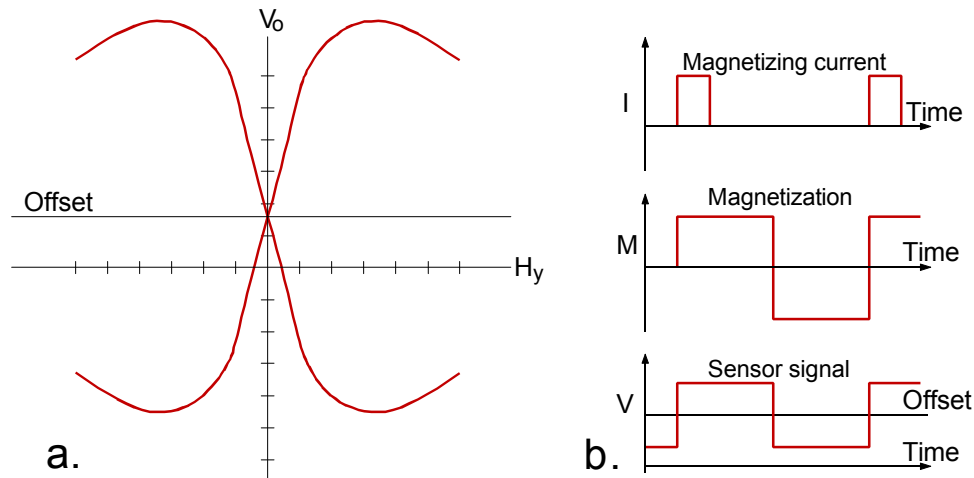
Figure 2.26: External current pulses set and reset the direction of magnetization, resulting in the "flipped" response characteristics shown by the dashed line. Note the DC offset of the device remains constant, while the signal output is inverted. (Adapted from [Petersen, 1989].)

synchronous detector driven by the same square-wave source. The rectified DC voltages V_{H1} and V_{H2} are thus proportional to the measured magnetic field components H_1 and H_2 . The applied field direction is dependant on the ratio of V to H, not their absolute values. This means that as long as the two channels are calibrated to the same sensitivity, no temperature correction is required [Fraden, 1993].

2.4.5 Magnetoelastic Compasses

A number of researchers have recently investigated the use of magnetoelastic (also known as magnetostrictive) materials as sensing elements for high-resolution magnetometers. The principle of operation is based on the changes in Young's modulus experienced by magnetic alloys when exposed to an external magnetic field. The modulus of elasticity E of a given material is basically a measure of its stiffness, and directly relates stress to strain as follows:

$$E = \frac{\sigma}{\epsilon} \tag{2.16}$$

where
E = Young's modulus of elasticity
σ = applied stress
ϵ = resulting strain.

Any ferromagnetic material will experience some finite amount of strain (expansion or shrinkage) in the direction of magnetization due to this magnetostriction phenomenon. It stands to reason that if the applied stress σ remains the same, strain ϵ will vary inversely with any change in Young's modulus E. In certain amorphous metallic alloys, this effect is very pronounced.

Barrett et al. [1973] proposed a qualitative explanation, wherein individual atoms in the crystal lattice are treated as tiny magnetic dipoles. The forces exerted by these dipoles on one another depend upon their mutual orientation within the lattice; if the dipoles are aligned end to end, the opposite
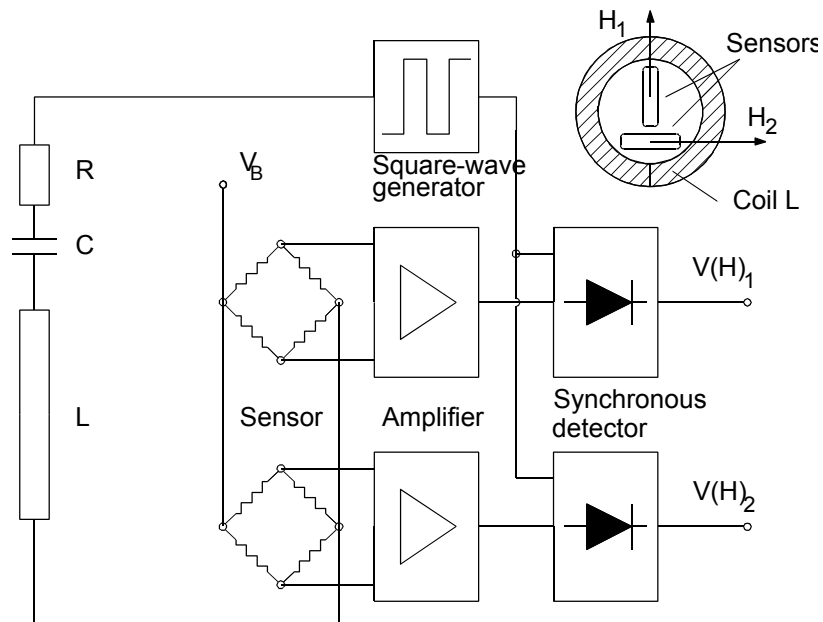
Figure 2.27: Block diagram of a two-axis magnetic compass system based on a commercially available anisotropic magnetoresistive sensor from Philips [Petersen, 1989].

poles attract, and the material shrinks ever so slightly. The crystal is said to exhibit a negative magnetostriction constant in this direction. Conversely, if the dipoles are rotated into side-by-side alignment through the influence of some external field, like poles will repel, and the result is a small expansion.

It follows that the strength of an unknown magnetic field can be accurately measured if a suitable means is employed to quantify the resulting change in length of some appropriate material displaying a high magnetostriction constant. There are currently at least two measurement technologies with the required resolution allowing the magnetoelastic magnetometer to be a realistic contender for high-sensitivity low-cost performance: 1) interferometric displacement sensing, and 2) tunneling-tip displacement sensing.

Lenz [1990] describes a magnetoelastic magnetometer which employs a Mach-Zender fiber-optic interferometer to measure the change in length of a magnetostrictive material when exposed to an external magnetic field. A laser source directs a beam of light along two optical fiber paths by way of a beam splitter as shown in Figure 2.28. One of the fibers is coated with a material (nickel iron was used) exhibiting a high magnetostrictive constant. The length of this fiber is stretched or compressed
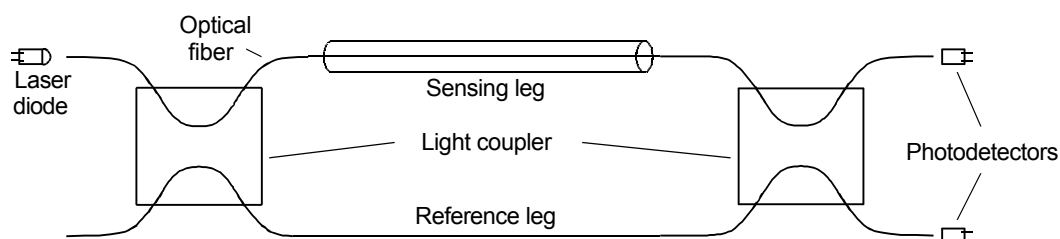

Figure 2.28: Fiber-optic magnetometers, basically a Mach-Zender interferometer with one fiber coated or attached to a magnetoelastic material, have a sensitivity range of 10^{-7} to 10 Gauss. (Adapted from [Lenz, 1990].)

in conjunction with any magnetoelastic expansion or contraction of its coating. The output beam from this fiber-optic cable is combined in a light coupler with the output beam from the uncoated reference fiber and fed to a pair of photodetectors.

Constructive and destructive interferences caused by differences in path lengths associated with the two fibers will cause the final output intensity as measured by the photodetectors to vary accordingly. This variation is directly related to the change in path length of the coated fiber, which in turn is a function of the magnetic field strength along the fiber axis. The prototype constructed by Lenz [1990] at Honeywell Corporation measured 10×2.5 centimeters (4×1 in) and was able to detect fields ranging from 10^{-7} Gauss up to 10 Gauss.

Researchers at the Naval Research Laboratory (NRL) have developed a prototype magnetoelastic magnetometer capable of detecting a field as small as $6×10^{-5}$ Gauss [Brizzolara et al., 1989] using the tunneling-tip approach. This new displacement sensing technology, invented in 1982 at IBM Zürich, is based on the measurement of current generated by quantum mechanical tunneling of electrons across a narrow gap (Figure 2.29). An analog feedback circuit compares the measured tunnel current with a desired value and outputs a drive signal to suitably adjust the distance between the tunneling electrodes with an electromechanical actuator [Kenny et al., 1991]. The instantaneous tunneling current is directly proportional to the exponential of electrode displacement. The most common actuators employed in this role are piezoelectric and electrostatic, the latter lending itself more readily to silicon micro-machining techniques.
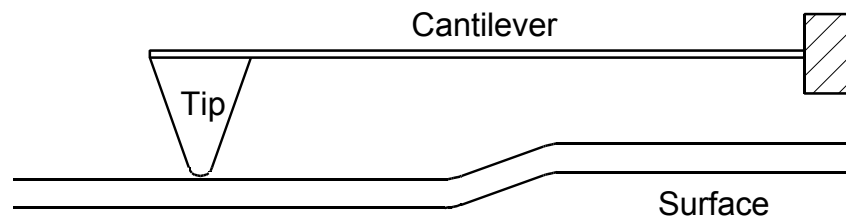

Figure 2.29: Scanning tunneling microscopy, invented at IBM Zürich in 1982, uses quantum mechanical tunneling of electrons across a barrier to measure separation distance at the gap. (Courtesy of T. W. Kenny, NASA JPL).

The active sense element in the NRL magnetometer is a 10 centimeter (4 in) metallic glass ribbon made from METGLAS 2605S2, annealed in a transverse magnetic field to yield a high magnetomechanical coupling [Brizzolara et al., 1989]. (METGLAS is an alloy of iron, boron, silicon, and carbon, and is a registered trademark of Allied Chemical.) The magnetoelastic ribbon elongates when exposed to an axial magnetic field, and the magnitude of this displacement is measured by a tunneling transducer as illustrated in Figure 2.30.

An electrochemically etched gold tip is mounted on a tubular piezoelectric actuator and positioned within about one nanometer of the free end of the METGLAS ribbon. The ribbon and tip are electrically biased with respect to each other, establishing a tunneling current that is fed back to the piezo actuator to maintain a constant gap separation. The degree of magnetically induced elongation of the ribbon can thus be inferred from the driving voltage applied to the piezoelectric actuator. The solenoidal coil shown in the diagram supplies a bias field of 0.85 oersted to shift the sensor into its region of maximum sensitivity.
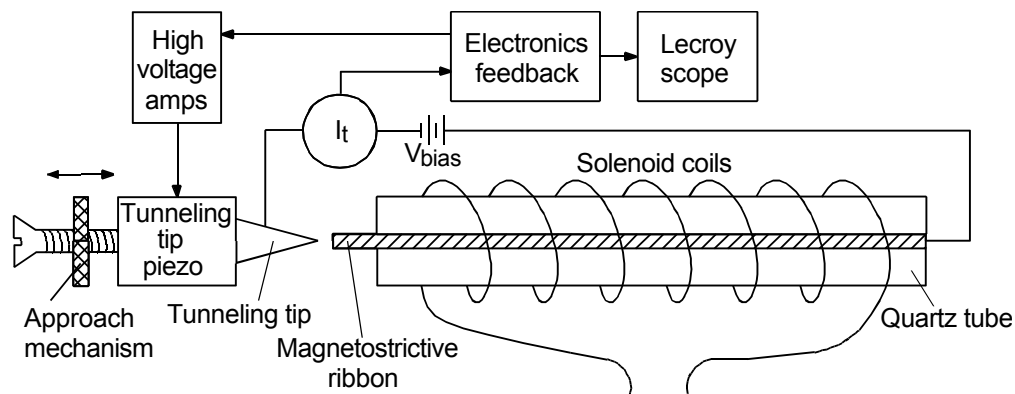
Figure 2.30: The NRL tunneling-transducer magnetometer employed a 10 cm (4 in) magnetoelastic ribbon vertically supported in a quartz tube [Brizzolara et al., 1989].

Fenn et al. [1992] propose an alternative tunneling-tip magnetoelastic configuration with a predicted sensitivity of 2×10^{-11} Gauss, along the same order of magnitude as the cryogenically cooled SQUID. A small cantilevered beam of METGLAS 2605S2, excited at its resonant frequency by a gold-film electrostatic actuator, is centered between two high-permeability magnetic flux concentrators as illustrated in Figure 2.31. Any changes in the modulus of elasticity of the beam will directly affect its natural frequency; these changes in natural frequency can then be measured and directly related to the strength of the ambient magnetic field. The effective shift in natural frequency is rather small, however (Fenn et al. [1992] report only a 6 Hz shift at saturation), again necessitating a very precise method of measurement.
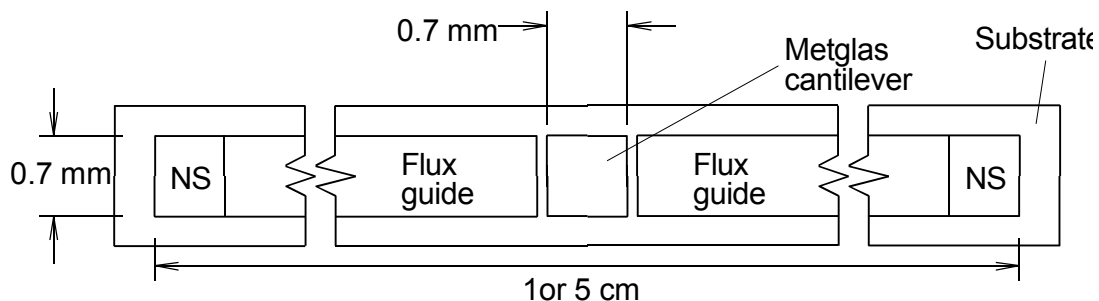

Figure 2.31: Top view of the single cantilevered design. (Adapted from [Fenn, et al., 1992].)

A second (non-magnetic) cantilever element is employed to track the displacement of the METGLAS reed with sub-angstrom resolution using tunneling-tip displacement sensing as illustrated in Figure 2.32. A pair of electrostatic actuator plates dynamically positions the reed follower to maintain a constant tunneling current in the probe gap, thus ensuring a constant lateral separation between the probe tip and the vibrating reed. The frequency of the excitation signal applied to the reed-follower actuator is therefore directly influenced by any resonant frequency changes occurring in the METGLAS reed. The magnetometer provides an analog voltage output which is proportional to this excitation frequency, and therefore indicative of the external magnetic field amplitude.
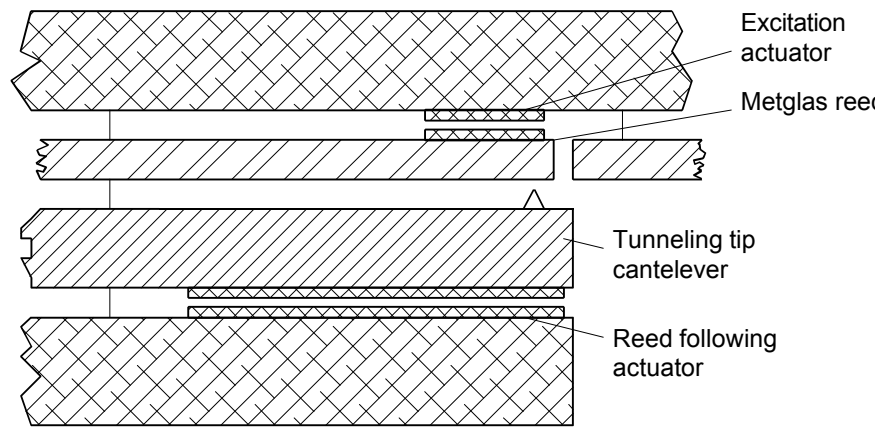
Figure 2.32: Side view of the double cantilevered design. (Adapted from [Fenn et al., 1992].)