SENSORS FOR AUTONOMOUS NAVIGATION AND HAZARD AVOIDANCE ON A PLANETARY MICRO-ROVER

by
William N. Kaliardos

B.S. Mechanical Engineering
B.S. Aerospace Engineering
University of Michigan (1990)

Submitted to the Department of
Aeronautics and Astronautics In Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 1993

Signature of Author _____
Department of Aeronautics and Astronautics
May 7, 1993

Approved by _____
Dr. David S. Kang
Technical Supervisor

Certified by _____
Professor Joseph F. Shea
Thesis Supervisor

Accepted by _____
Professor Harold Y. Wachman
Chairman, Department Graduate Committee

SENSORS FOR AUTONOMOUS NAVIGATION AND HAZARD AVOIDANCE ON A PLANETARY MICRO-ROVER

by
William N. Kaliardos

Submitted to the Department of Aeronautics and Astronautics on May 18, 1993 in partial fulfillment of the requirements for the degree of Master of Science at the Massachusetts Institute of Technology.

ABSTRACT

A sensor selection and design process for a small planetary rover is presented for the purpose of autonomous navigation and hazard avoidance. The selection process includes an overview of the options in methods and hardware, followed by the choice/design of sensors based on the performance requirements, vehicle constraints, and other factors. Emphasis is on the design methodology from a systems engineering perspective. Because of the need for prototype microrover demonstrations, real issues of cost and availability were strong drivers in the design. These issues, along with the platform limitations, have led to a *practical* sensor selection for an autonomous planetary microrover.

Thesis Supervisor: Dr. David S. Kang

Thesis Advisor: Dr. Joseph F. Shea
Adjunct Professor, Department of Aeronautics and Astronautics

ACKNOWLEDGEMENTS

I would like to thank the members of the micro-rover development team who have worked with me on this project: Sean Adams - Electronics, Shane Farritor - Geometric Simulation, Matt Fredette - Electronics, Rob Kim - Design & Manufacturing, Brenda Kraft - Image Processing, Giang Lam - Manufacturing, Kent Lietzau - Systems Testing, Anthony Lorusso - Electronics, Steve Lynn - Power & Electronics, Calvin Ma - Dynamics Simulation, Eric Malafeew - Controls, and Kimball Thurston - Animation. Special recognition goes to Tony and Sean, whom I worked with a great deal on the hardware, often late into the night. In addition to the student members, I would like to thank CSDL employeees Greg Capiello and Steve Finberg for providing technical help on occasions.

Special thanks to my thesis advisors Dr. David Kang and Professor Joseph Shea for providing guidance and support for my efforts, and most importantly for giving me the chance to work on the microrover project.

Finally, I would like to thank my family for all the usual things, and for allowing me to keep my car on the side of the house for the past ten years (although they wouldn't have used that space anyway).

This thesis was supported by The Charles Stark Draper Laboratory, Inc. (Lunar/Mars Micro-Rover Project.) Publication of this thesis does not constitute approval by The Charles Stark Draper Laboratory, Inc., of the findings or conclusions contained herein. It is published for the exchange and stimulation of ideas.

_____

William N. Kaliardos

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
INTRODUCTION

1.1. WHY A MICROROVER

Many rover studies have been done in the past, resulting in suggested designs of rovers that can perform a large variety of tasks. The ability to accomplish many scientific tasks is important especially for Mars, since there is an enormous cost and risk involved in just getting there. However, space missions of today have to operate within a budget that is nowhere near that of the Apollo program. By accepting this economic reality, it is impossible to plan for large complex missions. In fact, national studies have suggested scaling down any proposed space mission for the purpose of maintaining a reasonable cost. Therefore, it is not likely that any large rover will be launched in the next few decades.

Despite the recent space cutbacks, plans for future exploration of the surfaces of Mars and the Moon still exist. The long term (possibly very long term) goal is to send humans to Mars, in which case characterization of the Martian surface is needed in much more detail. Even without the plan for human exploration, there are many scientific motivations to collect data from Mars. Although the Viking landers (1976) successfully transmitted surface and atmosphere data, there have been no US. landers since then.

In order to continue Martian surface exploration within today's budgets, very small autonomous rovers have been proposed. These 'microrovers' are typically under a meter in length, compared to car or truck-size rovers that have been proposed in the past, fresh after the Apollo program. Due to their small size, microrovers do not have the ability to perform large complex tasks. However, they can carry payloads of the order of a few kilograms to various locations near the lander. The more complicated tasks can be accomplished by the lander, which is immobile. The lack of mobility of the lander involves a large risk if the interesting points of interest are not 'within reach.' Therefore, rovers are important for providing the ability to carry payloads to interesting locations near the lander, increasing the exploration capabilities of the mission.

Because microrovers are small and light, there is a much lower cost involved with transportation to a planet. Even so, future missions may utilize multiple microrovers opposed to a single large rover. The mission risk can be reduced significantly by operating in this way, since one or more rover failures can be tolerated. Therefore, microrovers offer some important advantages over large rovers, and can significantly improve the capabilities of the lander itself.

1.2. MISSION REQUIREMENTS

The microrover must be able to carry a small payload to various points in the vicinity of the lander. Currently, the plan is to limit the payload mass to 1 kg. Although some plans call for tethered rovers, these are limited to very short distances. Therefore, freedom in mobility requires an untethered rover. Depending on the method of operation, these rovers can conceivably travel very large distances. However, the rover needs to stay relatively close to the lander for communications purposes. Therefore, the microrover will need to operate likely within 1 km of the lander.

Because Mars is so far from earth, human rover control is not practical. In fact, radio signals can take up to 45 minutes round trip, depending on the orbital phases of the two planets. Therefore, it is necessary for the rover to have some degree of autonomous behavior. The limited processing capabilities of the microrover limit its ability to perform high-level path planning. Therefore, the accepted scenario is to send commands periodically from earth, so that the rover then is required to produce low-level path planning over small distances. By having a human in the loop periodically, the autonomous requirements of the rover are greatly reduced. In short, the rover is required to accept a command from earth, and travel to a nearby goal before the next communication. Therefore, the rover needs sufficient autonomy to arrive at that goal safely. The basic autonomous behaviors needed are 1) navigation, and 2) hazard avoidance.

The microrover needs a camera on board to transmit images to the lander, and then to earth. The transmitted image will serve as the basis for ground support to select the next interesting location. Although the command uplink will be preceded by an optimal path analysis, it is difficult command an exact path in an unstructured environment. Therefore, the rover needs to sense the environment sufficiently so that it can avoid

hazardous situations such as collisions with rocks, falling in craters, overturning, etc. It will have to circumvent these local obstructions, and resume travel towards its goal.

When the goal is reached, a scientific experiment may be performed with the payload. If not, the rover can send a video image, and then 'sleep' until it receives its next command. As an example, a typical day in the life of a microrover will involve the following:

1. Wake up and receive the latest command from the earth "ground support" station, via lander relay.
2. Travel to the goal, avoiding hazards along the way.
3. Perform experiment if needed, and transmit data along with a video image of the local terrain.
4. Shut down to conserve power, while still charging from solar panel.
5. Ground support receives data via lander relay, and after analysis chooses the next target of interest based on the video image.
6. Ground support chooses the next rover target, or series of targets, based on computer simulations of the "optimal" path. Command is sent to the rover.
7. Rover receives command, the process starts over.

This incremental operation is necessary due to the long radio signal travel time. Despite the slow process, the microrover may be able to continue for weeks, or even months, due to a small solar panel. Currently, the goal is to travel 100 meters per day, which is about the length of a football field.

1.3. SPACE ENVIRONMENTS

Although there are many difficulties associated with any mobile robot, the microrover must additionally be able to survive the harsh environments of space. Structural loading from launch and landing will require rugged components and possibly shock isolation. Also, radiation protection will be required to a certain degree. Thermal issues, however, are the most important

The microrover is considered to be Lunar or Martian, since it is being designed with enough flexibility so that it can be used in both environments with minor changes. The Lunar environment is different from Mars in many ways, some of the most important being the lack of an atmosphere. In addition, the terrain composition does not consist of rock fields that are common to both of the Viking landing sites. In fact, hazards are not considered to be rocks, but instead are craters and soft soil or dust.

There has been a great deal of effort spent in characterizing the Martian environment, and the details of this will not be presented in this paper. However, it is of interest to discuss some of the relevant Martian parameters. The atmosphere is thin, with a pressure of about 1% of earth's. The composition is over 95% carbon dioxide. The average solar energy is 44% of terrestrial values, and fluctuates strongly over the year due to the high eccentricity of the Martian orbit. The total diffuse light contribution has been measured to be roughly equal to the direct sunlight, still allowing distinct shadows. The spin rate is very nearly that of earth's, so that there are still nearly 24 hours in a night/day cycle. Temperature is on the cold side, varying from about -125°C in the winter to 22°C in the summer.

Gravity on Mars is about 1/3 of that on earth. The surface composition is divided into four classes of materials: drift, crusty to cloddy, blocky, and rock. Drift is fine enough to be moved by winds, resulting in frequent dust storms. Rock distributions from the Viking sites show up to 14% field coverage, which is large.

1.4. BASIC MICROROVER CONFIGURATION

The current proposed microrover is an articulated frame design, consisting of three platforms connected by spring steel wire. The spring steel suspension allows the individual platforms to pitch and roll relative to the others, but not yaw. This allows a high degree of mobility in a passive manner. Each platform has two drive wheels. Each drive wheel is individually powered by a DC motor located inside the wheel hub. Velocity feedback from each motor is provided. Figure 7.16 shows the MITy-2 prototype, which is representative of the basic configuration. One important change that will occur is the addition of conical wheels to increase mobility.

A microprocessor on the central platform controls the rover's operations. Nickel-cadmium batteries are located on the rear platform, along with a solar panel. The rear platform also holds the payload, as well as a small robotic arm for deployment. The rear platform is designed to be modular, to facilitate quick changes to other payloads. Therefore, the majority of the rover's autonomous sensors are located on the front and middle sections.

The front platform holds most of the hazard avoidance sensors, since vehicle travel is mainly in the forward direction. The video camera is also located in front, along with the communications hardware. The navigation sensors are located mainly on the middle platform.

Because of the large number of electronic components, it is important that all of them have a low power consumption. It is expected that the solar panel output will be only a few watts, so that every milliwatt counts. Of course, solar charging the batteries can take place throughout the day, whereas travel should be completed within 30 minutes. Although the solar panel does not have to provide all of the power during travel, the rover must be designed as efficiently as possible. Therefore, it is imperative that every sensor not only meet the size and mass constraints, but also the power consumption limitations. In addition, the rover's power regulation supplies DC voltages to the various components, so that the sensors should also be DC operated.

1.5. OUTLINE

This thesis is divided into two main sections: navigation and hazard avoidance. Chapters 2-4 discuss navigation, starting with an overview of the various methods, which then leads into the hardware selection. Chapter 4 then presents the prototype hardware that was constructed during this study. The second part of this thesis, Chapters 5-7, focuses on hazard avoidance. This begins with an overview of the sensing needs based on hazard definitions, followed by the general sensor choices, including a range finder. Chapter 6 then is dedicated to range finding entirely. The hazard-avoidance section ends also with a discussion of the prototype hardware. Finally, Chapter 8 suggests future work for both navigation and hazard avoidance.

CHAPTER TWO
NAVIGATION METHODS

2.1. NAVIGATION REQUIREMENTS

2.1.1. Overview

For any autonomous vehicle to travel towards a goal, navigation is an essential part of its system. The term 'navigation' is defined as "the science of getting ships, aircraft, or spacecraft from place to place" [10]. An autonomous vehicle needs to determine, from sensors, its location relative to some other reference. This behavior is a basis for all travel.

Navigation is needed on an autonomous vehicle so it can travel towards a goal, or target. In the case of the rover, the goal is a location in the camera's field of view that is interesting to human operators on earth, based upon visual information from the on board camera. Hence, the navigation system must present enough information to let the rover's motion controller achieve that goal. Therefore, it is not so much important to navigate along a certain path, or navigate to a certain inertial coordinate, but instead to arrive at a location that is initially observed from the camera image.

Since the target is chosen visually from a sensor (the camera), it would seem fitting to arrive at the target based on visual information. For example, process the visual information during travel, and lock on to that target until it is achieved. After all, this is what animals do, and it works easily (Figure 2.1). Unfortunately, mimicking this seemingly simple process is very troublesome for a microprocessor, and is a subject of much work in the robotics field.

This visually-based navigation is known as *reference* navigation, since a recognizable object (in this case the target itself) is used as the primary position reference for all position calculations [5]. Any motion of the rover is irrelevant, unless it is relative to the particular reference. If the target is not the reference itself, then the target position is known relative to the reference, and so achieving the target is possible by measuring the rover position relative to that same reference.

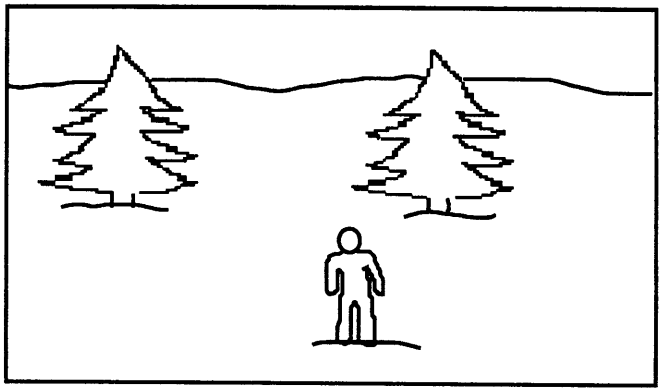

Figure 2.1: Navigation by reference to fixed objects.

Another type of navigation is *dead reckoning*. This is not based on a position reference, but position is estimated by integrating motion parameters on board the rover. This is usually velocity and heading. Although the measurements of these motion parameters are relative to something (e.g. the ground, an inertial frame), the measurements are local, and thus need to be accumulated to calculate a position change from a particular starting location. Dead reckoning, then, requires continuous measurements of motion parameters, opposed to reference navigation, which allows position calculations regardless of the vehicle's motion history (Figure 2.2).

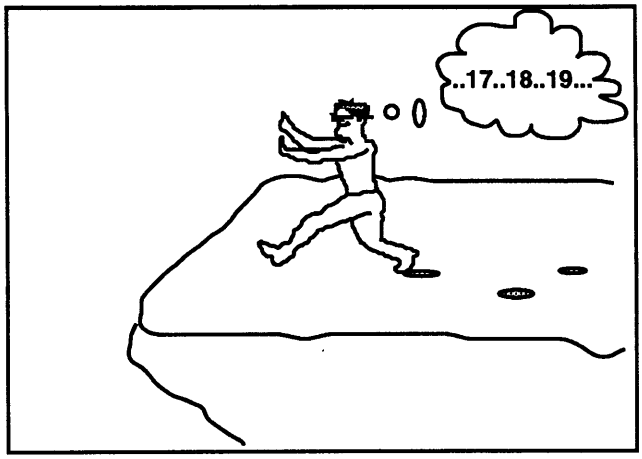

Figure 2.2: Dead reckoning requires accumulation of previous measurements.

Both types of navigation systems have inherent advantages and disadvantages. In choosing a method, there are many considerations needed. Before discussing these, it is important to clearly define the requirements of the navigation system, based on the mission requirements.

2.1.2. Mission Requirements

As stated earlier, the goal of the mission is to carry scientific payloads to interesting locations on the Lunar or Martian terrain that are in the vicinity of the lander. Daily traverses of 100 m are expected, with position accuracy of 10 m (10% of the total distance). The mobile platform must accept daily commands from earth, and travel autonomously towards the target, avoiding hazards on the way. The lander will provide a means for deployment, and a communication system to earth, but will not provide other tools.

2.1.3. Navigation Requirements

In order to achieve the mission goal, the navigation subsystem must sufficiently perform its own part. The most relevant mission requirement is the 10 % position accuracy in traveling to a target [3]. This number, although loosely defined, was based upon the fact that the error is small enough to recognize the location from the previous image (when the target was selected), yet large enough to allow for a practical navigation system to be implemented on such a small platform. By being able to recognize the new location from the previous image, the cumulative position error is reduced significantly.

The other important mission requirement is that the lander is not expected to contribute any navigational tools, such as external beacons. This forces the rover to have to navigate with its own on board sensors.

Also, the rover is expected to travel 100 meters per day, with the 10 % accuracy described. At this distance, the rover should take anywhere from 20 minutes to 2 hours, depending on the terrain. The elapsed time can be an important factor for inertial navigation systems. It should also be mentioned that travel will take place during the day, so there will be sufficient lighting for passive sensing.

2.2. NAVIGATION METHODS

2.2.1. Overview

As discussed in the previous section, navigation for autonomous vehicles can be classified as either reference or dead reckoning. Indoor mobile robots have the advantage of using a known, structured environment from which references can be taken rather easily, and also compared to a known map. In these cases, sensing the reference objects can be performed, for instance, with laser range finders by recognizing the object distribution, or by using additional aids such as retroreflectors or bar codes. Navigation is greatly facilitated in such structured environments, especially when the safe paths such as hallways are easily distinguished from the remaining area.

In an unstructured, outdoor environment, autonomous navigation is much more difficult. There are two main reasons for this:

1. Naturally occurring landmarks are not easily recognizable
2. The terrain is not smooth, and hence more difficult to measure motion from

The first point is relevant to reference-based navigation. Consider a random field of rocks. For the rover to measure its position relative to these rocks, it must first be able to recognize them, which can be difficult due to their similar features. Figure 2.3(a) shows a typical scene in a structured and unstructured environment. Recognizing one object from the other is obviously much simpler in the first case. Also, in order to even recognize the

rocks, the rocks must be in the rover's line of sight, which can be difficult unless the reference rocks stand well above the rest.

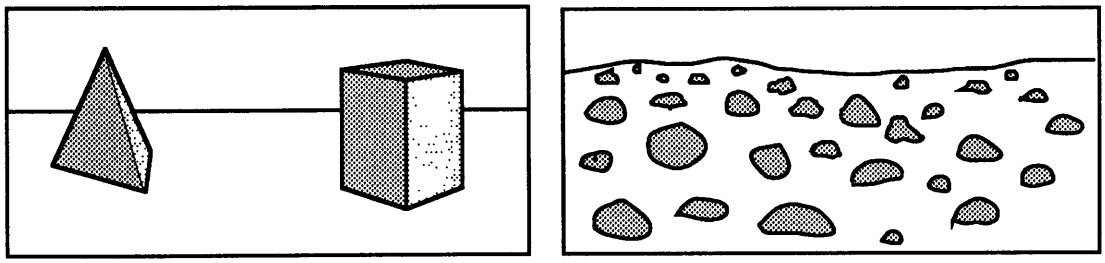

Figure 2.3a: Recognition of objects, for reference navigation, is more difficult in an unstructured environment.

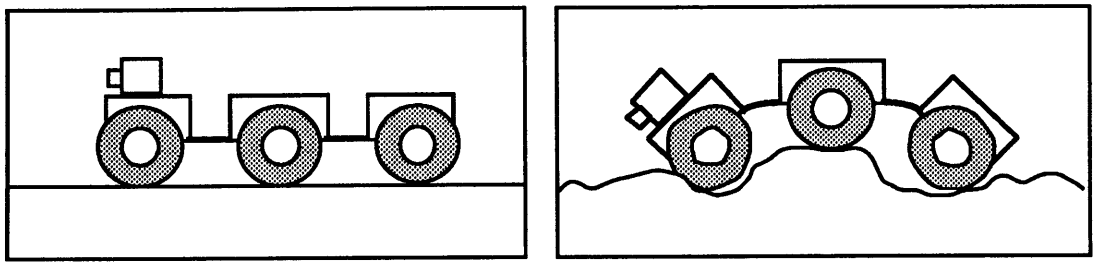

Figure 2.3b: Motion in the ground plane, for dead reckoning navigation, is more difficult to measure when the terrain is irregular.

The second point is relevant to dead reckoning. Meaningful motion parameters can be difficult to measure, whether they be inertial or relative to the ground. Integrating inertial data for position is possible, but errors can grow without bound easily unless very high-quality instruments and electronics are used, especially over rough terrain. Inertial sensors can be both very costly, and unfit for packaging on small platforms. Ground-referenced sensors, such as odometers, tend to be rather inaccurate over irregular terrain (Figure 2.3b), so that position errors can continue to grow without bound also, but as an explicit function of distance, and not time.

Nevertheless, many navigation options exist under the two categories of reference and dead reckoning. These are shown in Figure 2.4. Note that there are few options for measuring translation directly, which is the fundamental problem. Heading, on the other hand, is available from many more sources.

Reference navigation has the advantage of allowing direct position calculation with bounded errors, but can suffer from periods of time where referencing is not possible.

Dead reckoning, being self-sufficient, can offer position calculation at any time, but suffers from unbounded error growth. The two methods are thus often combined to yield a complete navigation system that offers continuous positioning with bounded errors. However, packaging limitations can prevent this.

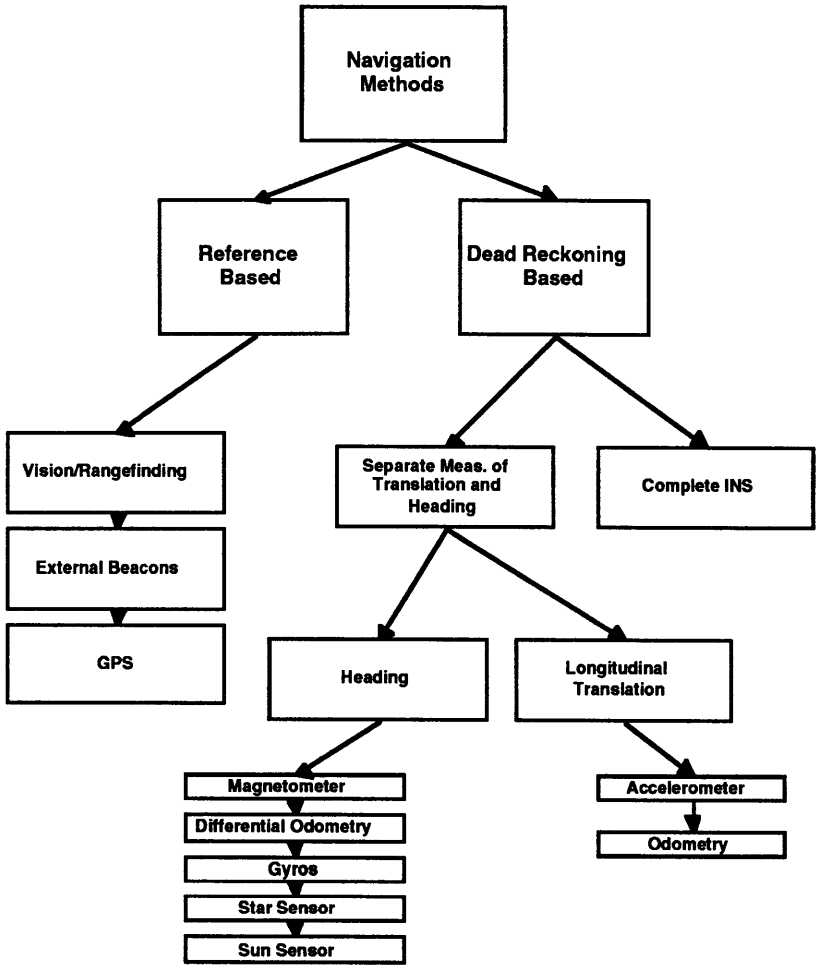

Figure 2.4: Applicable navigation techniques and sensors

2.2.2. Reference Navigation Techniques

Vision/Range Finding

As stated earlier, navigation by reference to known objects is rather easy indoors, where objects can purposely be made distinguishable through bar codes or retroreflectors. In an outdoor, unstructured environment, it may be difficult to even keep an object in the line of sight, let alone recognize it. However, this technique is still feasible, and is the most

direct navigation solution to arriving at a goal that is initially defined through a visual image (the video transmission to earth).

A scanning range finder can yield a depth map directly, which is often the goal of a passive vision system. However, a scanning range finder is complex, large, and requires high speed sampling to build a map with good resolution. In addition, its active nature results in high power consumption. Despite these drawbacks, the direct depth map that results can make navigation (and hazard avoidance) much easier. Of course, this places high demands on the processor.

On the other hand, a passive vision system has the advantage of a large instantaneous field of view, so that an entire scene can captured at once. One disadvantage of such passive systems is that the illumination of the scene in not controlled, and is hence dependent on the ambient light, the sun. In most daytime situations, the intensity of the illumination should not be a problem. However, the contrast may be low, depending on the object's signature, which can make visual interpretation difficult.

The most difficult part of any vision system is the correlation of successive images, or in some cases the characterization of optical flow. Images need to be processed constantly, at a rate that depends on the rover's speed. For navigation, additional motion information is often needed along with the visual data. Although visual navigation occurs rather easily with our eyes, interpretation of successive images from a camera often results in the need for powerful computers, leaving little dedication for other processing.

In short, both passive vision and active range finding requires a large amount of processing, although both can offer depth mapping capabilities. High local accuracies are possible, especially if the goal is within sight and is recognizable. Indoor mobile robots can use these methods effectively, but the process is more difficult in an unstructured outdoor environment.

External Beacons

One way to get around the difficulty of recognizing reference objects in an outdoor environment is to introduce artificial ones. In fact, taking this one step further by making the artificial objects active, makes position calculations rather easy. Active 'landmarks' are known as beacons, and are a common navigation method for aircraft and marine use.

This method is actually not very different from visual referencing, except that recognition and ranging of the reference landmark is simplified because the beacons transmit a known waveform. In order to calculate the two-dimensional position, a single or dual beacon system is needed. Single beacons are essentially range finders that can locate the angle and range of the rover, yielding a polar coordinate position relative to it. However, since radar is not very accurate in azimuth, or more importantly, in order to have omnidirectional capability, a single beacon system is not used. More common, then, is a dual-beacon system, as shown in Figure 2.5. This system provides two range values, allowing position calculations to be made through the use of triangulation, with only one position ambiguity (which should not be a problem). For either case, the positions of the beacons are assumed to be known. Although the single beacon system may seem simpler, the dual-beacon system is the system of choice, mainly because of its omnidirectional capabilities.

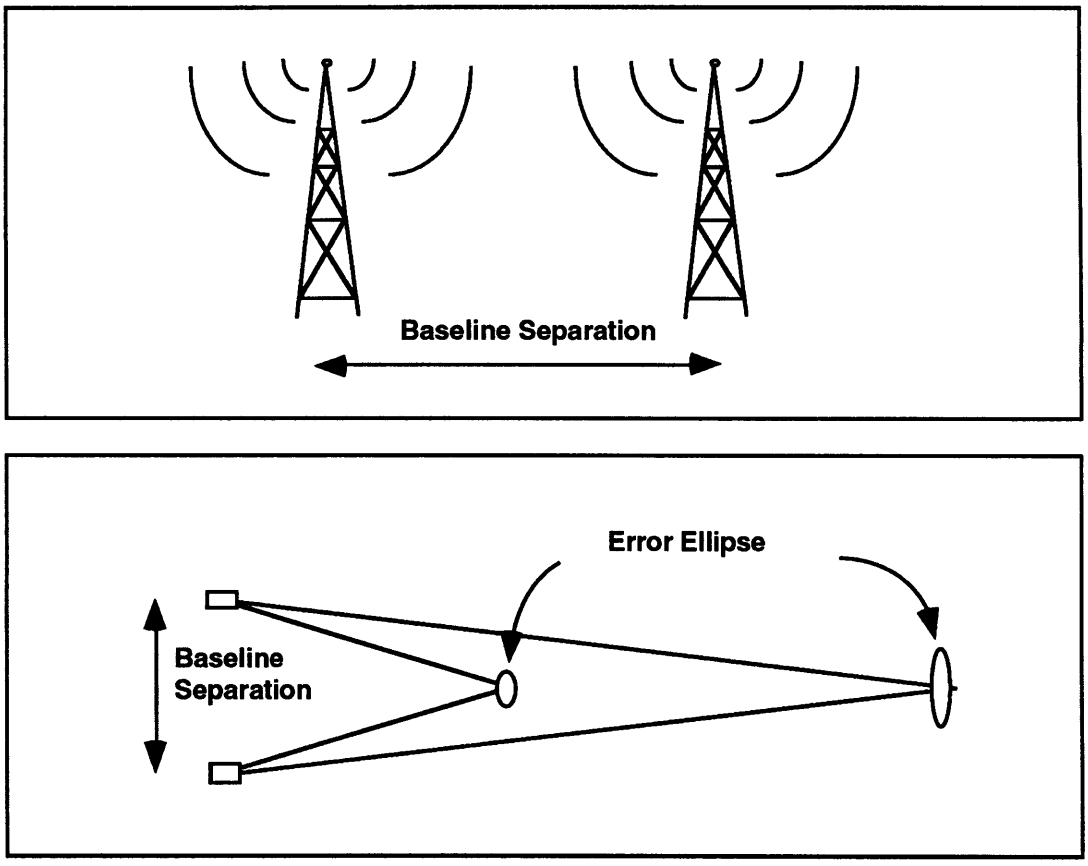

Figure 2.5: Dual beacon navigation. Lower picture shows how position errors due to individual range errors are magnified at distances that are large relative to the baseline separation.

For a dual-beacon system, the rover would have to carry radio receivers in order to collect the beacon signals. If the transmitting power of the beacons is large, the rover receiver can be relatively small. In order to extract range, the rover will need some electronics to detect the beacon signals, using either phase shift or pulsed time-of-flight techniques (see Ranging Techniques, section 6.2). Hence, the rover will still be required to carry ranging electronics, as well as the receiver itself.

The advantage of a beacon system is that position can be calculated at any given time, as long as the rover is within the line-of-sight. Of course, this may be a problem over irregular terrain, and hence may limit the allowable distance from the lander. If the beacons are mounted very high, longer ranges are possible. However, the support tower for such a beacon would have to be collapsible and very light for the long space flight, which has its obvious limitations. Also, the dual-beacon system needs to have a large baseline separation, so that position accuracy is maintained at larger distances, as is the case with all triangulation systems. At distances that are very long relative to the baseline separation, the range errors of each beacon combine to yield a larger rover position error, due to geometrical dilution of precision (GDOP), which is also shown in Figure 2.5.

It should be noted that the majority of all aircraft navigation over the U.S. is performed by similar positioning systems. There are many of these beacons around the country, and so GDOP is not much of a problem. Each beacon broadcasts all around it, and so the capacity of aircraft that can be supported at any time is essentially infinite. This is one of the main advantages of an omnidirectional beacon network. Unfortunately, however, having such an infrastructure in place is difficult on other planets, especially when the physical size and mass of each is considered. This is true for even a simple dual-beacon system.

Global Positioning System (GPS)

Similar to land beacons, a GPS system uses an orbiting satellite network to calculate position using triangulation methods, as shown in Figure 2.6. Typically 3 satellites are needed to provide range information in 3D. However, at least four satellites are usually used so that the additional information can significantly reduce the range bias error due to the inaccurate receiver clock.

For each measurement, the distance to the user can be expressed as

$$| x_i - x_o | = c\Delta t \tag{2.1}$$

where x_i denotes a satellite position (known), c is the speed of light, and Δt is the measured time of flight, which is the actual time dt_i offset by the measurement error dt.

$$\Delta t = dt_i - dt \tag{2.2}$$

By squaring each term, the governing equation for each measurement in scalar notation is:

$$(x_i - x_o)^2 + (y_i - y_o)^2 + (z_i - z_o)^2 = c^2(dt_i - dt)^2 \tag{2.3}$$

By taking four measurements of t_i (i=1,2,3,4), each from a known position (x_i, y_i, z_i), the system can be solved for the clock error t, and hence the rover coordinate (x_o, y_o, z_o) can be calculated.

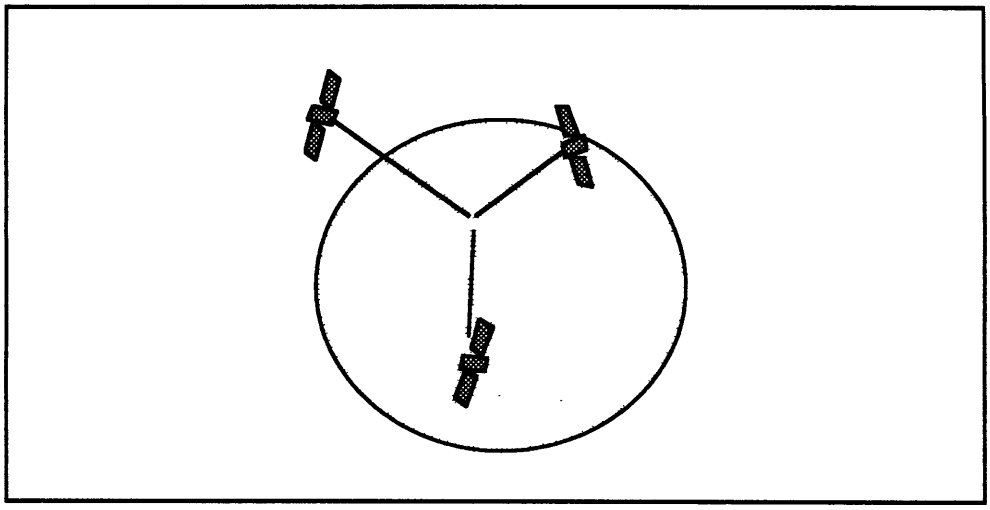

Figure 2.6: GPS also requires multiple range measurements from various positions

The bias error is due to the time-of-flight ranging method. A GPS satellite transmits a signal at a known time, at which point the satellite is at a known position (orbits can be determined to very high precision). The receiver must also be able to time this signal at the instant of reception. The time of transmission can either be pre-determined, or the signal can contain this information. In either case, the clocks on the transmitters and the receiver, must be synchronized. Because of the high speed of light, though, errors of only 1 μs correspond to range errors of about 300 meters. This is the typical maximum

accuracy of a quartz clock. But, as mentioned, an additional measurement can remove much of this bias error, assuming it is the same for each. This greatly relaxes the timing requirement on the receiving platform. However, timing errors still exist, so that ranging is typically accurate to within 10 meters, although 1 meter accuracy is possible. Accuracy of the system can be improved using differential GPS, which involves using another receiver at a known location that is relatively close to the rover, such as at the lander. In this case, GPS position errors at the lander can be assumed to be applicable to the rover too, and hence subtracted.

Inaccuracies in GPS can occur from many sources. These include:

1. Clock Errors
2. Satellite Orbit Uncertainty
3. Speed of light (through atmosphere)
4. Multipath errors
5. Satellite geometry (GDOP)

Under most circumstances, the orbits are well-known, and the errors due to speed of light and reflections (multipath) are not a factor. Clock errors can contribute the most, which is a direct ranging error. In addition, the satellite geometry is very important for the same reasons as with land beacons. This dilution of precision can result in much larger errors. Hence, multiple satellites are needed, or the same satellite can be used at different positions in the orbit. If only one satellite is used, then position calculations can be made only after the satellite has moved a sufficient amount within its orbit, which will result in slow measurement updates. A large satellite network can allow continual coverage, so that positioning can be made at any time, nearly instantaneously. However, the cost of such a network is great, and can only be justified if there are many users. Hence, a single satellite GPS system is the most likely, but positioning will be very infrequent. This may be acceptable for monitoring the position over long periods of time, but is insufficient for real-time navigation.

2.2.3. Dead Reckoning Techniques

Complete Inertial Navigation

Inertial navigation allows position calculations by measuring motion relative to an inertial frame of reference. The basic element is the accelerometer, which measures linear accelerations along its axis. Another main element is the gyroscope, which, depending on the form, measures angular displacement or rate. In any case, linear displacement is calculated by integrating the accelerometer twice with respect to time.

$$x = \iint a_x dt^2 \qquad y = \iint a_y dt^2 \qquad z = \iint a_z dt^2 \tag{2.4}$$

Theoretically, there are no inaccuracy problems with this, as long as the integration is performed well. However, since accelerometers have errors, position errors can grow without bound, with the square of time. Therefore, the time of integration is a very large factor in controlling the error growth. Even extremely small bias errors in the accelerometer can result in large position errors over time. This is the most difficult problem with an inertial navigation system.

If the platform stays oriented with respect to the same inertial coordinate frame throughout the travel, accelerometers would be the only sensors needed. However, since the platform typically needs to rotate during travel, these angular displacements need to be measured. The standard sensor for measuring such rotations is the gyroscope. A true gyroscope consists of a spinning disk, whose angular momentum vector is fixed in inertial space in the absence of external moments. However, the term 'gyro' has now (inappropriately) become a name for any inertial angular rate sensor, not limited to the spinning disk design.

With accelerometers, gyros, and the proper integrating electronics, a complete INS is possible. There are two configurations for this system:

- Stabilized Platforms
- Strapdown

In the stabilized system, the entire accelerometer package rotates relative to the platform such that they are oriented the same way throughout the travel. Hence, the accelerometer's orientation with respect to inertial space does change. Rotation is usually

done actively, by sensing any rotations and applying the proper torque to the accelerometer system through actuators. In short, the purpose of the gyros are only to stabilize the accelerometer platform, so that integration of the accelerometers yield the correct translation directly. The stabilized platform method is highly accurate, but also complex.

The strapdown system is more common, due to its robust design. In this case, the entire gyro-accelerometer package is not gimbaled, but directly mounted to the vehicle platform. When the vehicle rotates, the angles are monitored, and corrections are applied to any measured accelerations. Hence, instead of sensing rotations and physically rotating the accelerometers, rotation is performed mathematically. This results also in a smaller and lighter system, although it is less-accurate than a stabilized system.

In order to package a system on the rover, a strap-down system is the better choice. A full INS would require at least three accelerometers and three gyros to sense the six degrees of freedom of possible motion. However, it would likely be sufficient to sense only motion in the ground plane, so that a vertical accelerometer would not be necessary. Although a full INS system is feasible, there are associated problems that would make it difficult to use effectively:

- The presence of gravity can introduce very large errors in the horizontal axis accelerometers due to mounting errors, or if the body is tilted and not measured accurately (Figure 2.7). Since gravity is a large acceleration compared to the rover's motion, even very small tilt angles can be a problem.
- The size and power requirements of any currently available INS is too large for the rover's platforms, as well as overly expensive.

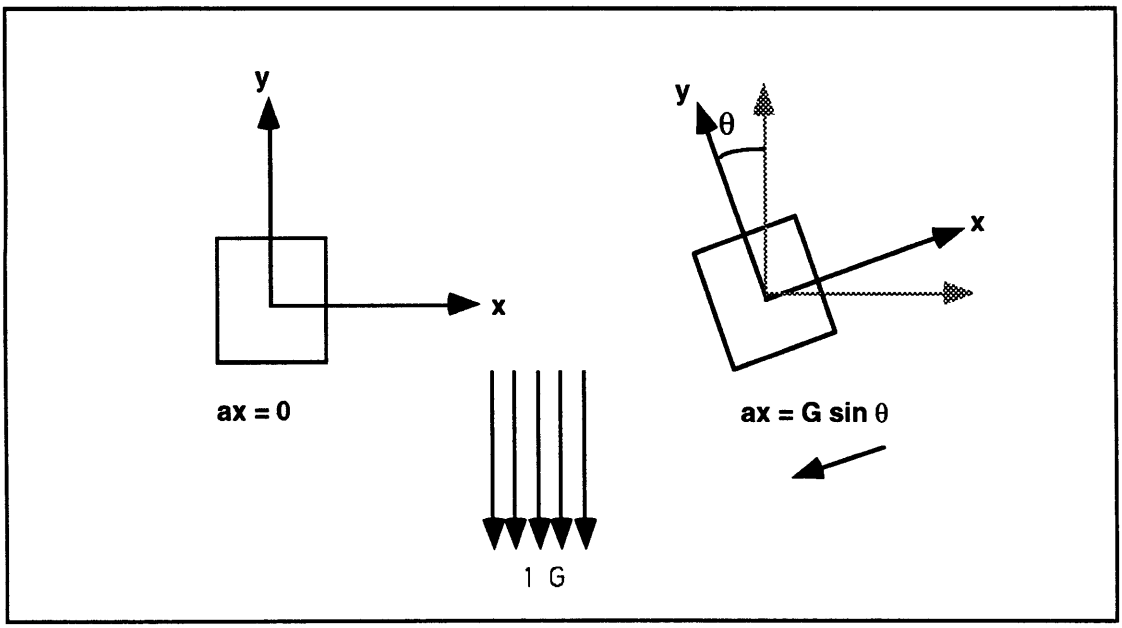

Figure 2.7: Small angular offsets can be detected as large horizontal motion in the presence of a gravitational field.

ZUP

It should be noted that any inertial system is best used when updates from a known reference can be used to periodically calibrate it [6]. Although there is no way to compare position, stopping the rover can guarantee zero velocity and acceleration, and zero rotation rates, except for the small contribution due to the planet's rotation. By taking a 'zero update' or ZUP, any drift rates can be measured and subtracted from future readings before the next update. The rover has the advantage of being able to update often. This is one way to help contain the error growth due to sensor drifts and mounting errors.

Longitudinal Translation

Since the rover is designed to move along one axis, measurements of translational motion can be reduced to only along this axis. In this case, it is necessary to know the orientation of the axis for any translation measurements, which will be discussed shortly. For the microrover, the axis along which translation occurs is called the longitudinal axis. The configuration of the longitudinal translation sensors is shown in Figure 2.8.

Accelerometer

Forward and backwards motion can be sensed inertially, using only an accelerometer. Of course, the accelerometer is the core element of an IMU, and the problems with IMUs are mainly due to the accelerometer errors, which grow unbounded in the absence of position updates. Motion in the ground plane is especially difficult due to the gravitational components when tilted. Also, large longitudinal accelerations from the rover's interaction with obstacles can make position calculations even more difficult.

Drive Wheel

Like in a car, forward distance can be measured using the drive wheel as an odometer. In this case, the displacement is a linear function of the effective wheel radius and the wheel angular displacement. Hence, it is necessary to have some way of measuring wheel rotations, either through angular position sensors such as an encoder, or by integrating a tachometer. Fortunately, some sensor will be available on each drive motor for the purpose of speed control, so a redundant use as a navigation sensor requires no additional hardware.

Using the drive wheel as the odometer is the most accurate when the terrain is flat, and the wheels are stiff and do not slip. This, of course, will not be the case; with the rover. The rover will have very deformable tires, and the terrain will likely be very irregular. However, empirical data will likely give results that can be used for navigation purposes, with adequate testing. The main problem will arise on drift material, where the wheel slip can be large. When the wheel slips significantly, large odometry errors will result, unless corrected through the use of other measurements. Even if data from all six wheels are used, there is still a chance that all of them will be slipping significantly, especially if the rover is inclined. The reason that the drive wheels slip is due to the large applied torque from the motors. This leads to the need for an unpowered drag wheel.

Drag Wheel

Unlike a drive wheel, a drag wheel has no applied torque. In fact, the only torque is a resistive torque that opposes the wheel motion, due to friction. Therefore, a drag wheel eliminates the problem of slipping substantially under low-cohesion conditions.

An unpowered drag wheel is, however, subject to the same difficulty over rough terrain as the drive wheels. This error can be reduced if the wheel size is very large relative to the size of the irregularities, but this is not possible. Packaging is a very important factor, since the drag wheel can easily interfere with mobility if not placed correctly. *Drive* wheels have the advantage in that they provide the rover's mobility, and so odometry would just be another use for them. On the other hand, drag wheels do not suffer from slip, but require additional hardware and packaging space.

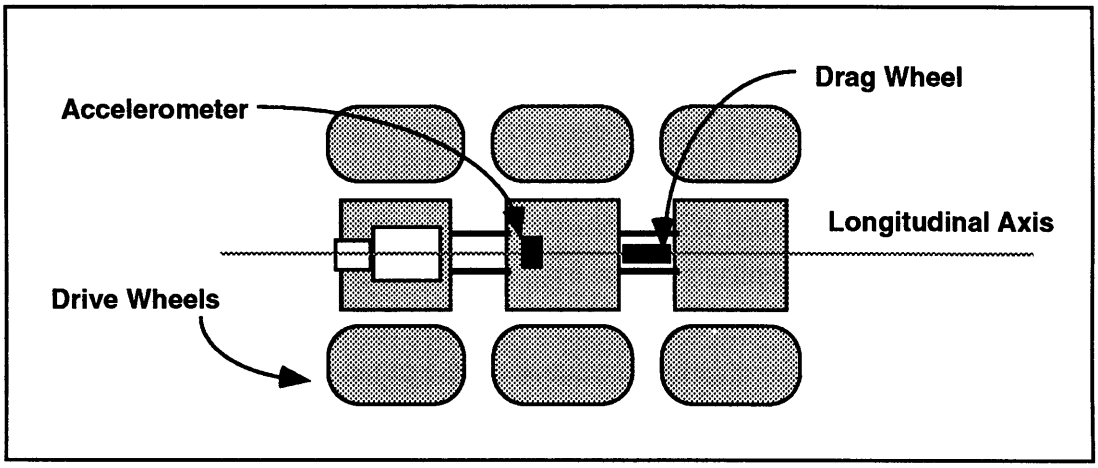

Figure 2.8: Sensors for measuring longitudinal translation.

In order to maintain contact with the terrain, the drag wheel should have the ability to flex up and down. This requires space, as well as some kind of suspension system. The placement of the drag wheel must take this into account. In addition, the drag wheel should fall near the longitudinal centerline, so that it is representative of the vehicle during turns. Also, since the three platforms flex negligibly in yaw, there is a significant lateral motion of the front and rear platforms during turning. Hence, the best drag wheel position would fall under the center platform, for kinematic reasons. Unfortunately, this is not good for packaging, since vertical space is needed. The only place there is vertical space is in the two areas between the front and middle platforms. These areas not only offer drag wheel packaging, but are also 'dead zones' where mobility is not provided. Because of this, a drag wheel can possibly help mobility by providing a passive way of preventing hang-ups on rocks.

Heading

For navigation, rotation in yaw, or heading, must be measured. This is the only relevent rotation, since navigation is essentially a two dimensional problem in the ground plane. Because the vehicle body does not flex in yaw, only one rotation sensor is needed.

Magnetometer

The heading angle can be determined by measuring the vehicle heading relative to the ambient magnetic field lines. The direction of the field can be found from magnetometers. Magnetometers are of the quantum or induction type. For navigation purposes, induction magnetometers are most often used, of which there are two types: search coil, and fluxgate. Search coil magnetometers are usually used on spinning satellites, since the induced voltage in the sensor's coil is proportional to the time rate of change of the field normal to the coil.

The more common *fluxgate* magnetometer detects the ambient magnetic field by using primary and secondary coils [8]. The primary coil is electrically driven into saturation, typically with an applied triangle wave. The resulting flux time history is altered by the ambient field that is along the axis of the core. This alteration is detected by secondary coils that produce induced pulses whose phase or amplitude is proportional to the ambient field.

The accuracy of magnetometers is usually under 1°, if the field is strong enough. Of course, it should be understood that the field lines do not necessarily follow lines of constant longitude, so that the field lines can have lateral components as well. Earth, for example, can be modeled as a magnetic dipole that is tilted about 12 ° from its spin axis, as shown in Figure 2.9. The field lines can thus give a false indication of north, depending on where the measurement is taken. Within a local area, however, the orientation of the field is constant, so that navigation with respect to the field lines does not present a problem, whether the heading is relative or absolute. Magnetometers work well under static conditions, when the primary coil is oriented correctly with respect to the earth. Many are thus passively gimbaled, allowing gravity to perform the alignment. Because of this, there is a considerable amount of filtering that is required under dynamic situations. The result is a slow response, with time constants of the order of one second. Therefore, magnetometers are most often used as an absolute reference for calibration of

other sensors. Calibration is performed during relatively static conditions, when the magnetometer is most accurate.

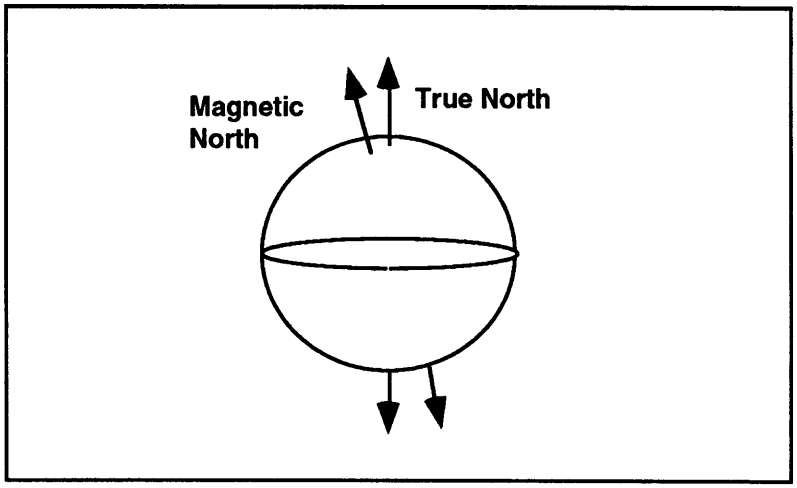

Figure 2.9: On earth, the magnetic field is offset from true north by about 12°

Differential Odometry

Heading information can also be provided by differencing two drag wheels that are left and right of each other. Ideally, during straight motion the difference in wheel velocities is zero. During a turn, the inner wheel rotates slower than the outer wheel, indicating the rate of turn. Figure 2.10 shows a set of wheels turning at a radius, r. The outer wheel has a velocity V_1, and the inner wheel V_2, so that the differential velocity is $\Delta V=V_1-V_2$.

The best sensitivity to rotation is when the wheels are separated far apart, which is limited by the rover width. For a given rover center velocity, V, at a turn radius r, the differential velocity is ideally

$$\Delta V = \frac{d}{2}\frac{V}{r} \tag{2.5}$$

Again, best results are obtained on flat terrain. When the terrain is irregular, and different for each wheel, it is easy to cause a differential velocity even when traveling straight. In all cases, the differential velocity error ΔV_{err} is introduced so that

$$\Delta V = \frac{d}{2}\frac{V}{r} + \Delta V_{err} \tag{2.6}$$

When the terrain is irregular on the scale of the wheel diameter, the differential error can be a significant portion of the measurement. If the separation d is large, the error contribution can be reduced. However, since d is limited, the error contribution will not be insignificant.

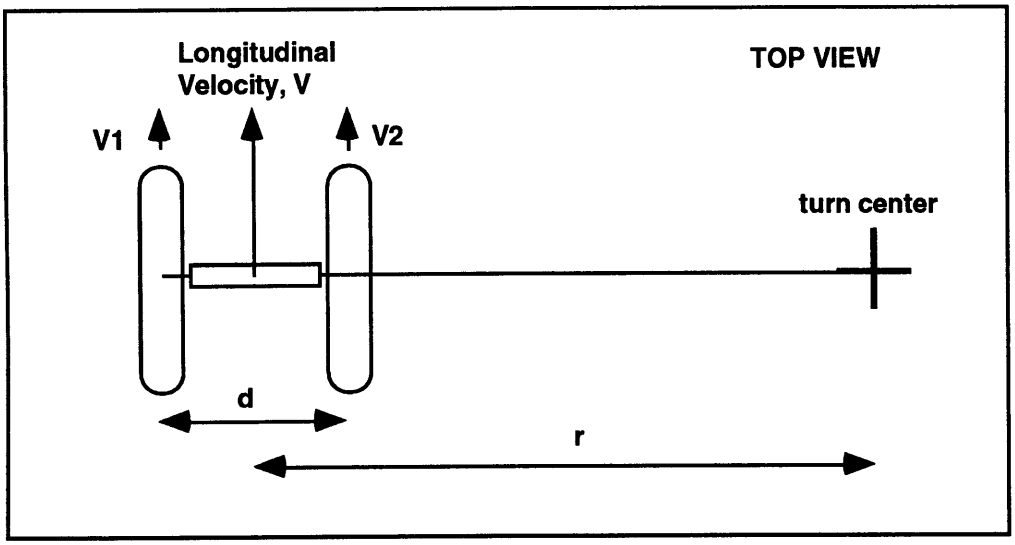

Figure 2.10: Differential Odometry

Gyroscopes and Angular Rate Sensors

The true gyroscope consists of a spinning disk, or rotor, whose angular momentum vector is fixed in inertial space in the absence of applied torques. In fact, the word gyroscope means 'to view rotation.' The term 'gyro' has now become a name for any inertial angular rate sensor, not limited to the spinning disk design.

The classical gyroscope can measure rotations that are not in line with the spin axis. Because of this, gyroscopes are at most two-axis. There are many different types of gyroscope configurations, most of which have been developed for aircraft navigation. These are:

- free
- vertical
- directional
- floatation
- rate

Gyroscopes are governed by the equation:

$$\mathbf{T} = \omega \times \mathbf{H} \tag{2.7}$$

where **T** is the torque normal to the spin axis, ω is the spin axis precession rate, and **H** is the angular momentum of the rotor (Figure 2.11). The governing equation can be considered 'reversible', in that :

1. An applied torque will result in a precession of the spin axis, normal to the applied torque.
2. An applied angular rate (precession) will result in a torque, normal to the applied rate.

In either case, the angular momentum determines the magnitude of the sensitivity between the two quantities.

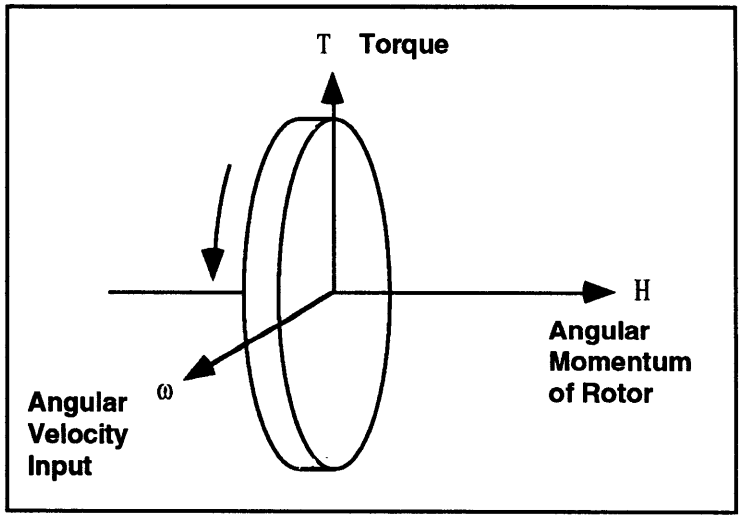

Figure 2.11: Classical Gyroscope

In addition to the rate gyroscopes, there are a wide variety of other angular rate sensors. These include:

- fluid stream
- fiber optic
- coriolis

Angular rate sensors offer the advantage of less mechanical complexity, although the electronic complexity is increased.

The following sections will discuss the method of operation of all of the above 'gyros.'

Free Gyro

By gimbaling twice, a gyro can measure rotations about two axes. Since the gyro is then free to rotate in any direction, these are called 'free' gyros. An illustration is shown in Figure 2.12. The only limitation to movement is when the inner and outer gimbals become nearly coplanar (called gyro lock), in which case rotation can result in an applied torque. Otherwise, the gyro is torque-free, other than friction in the gimbal pivots. This small friction, however, is the cause sensing errors known as gyro drift.

Gyro drift can be reduced by lowering the torque due to gimbal pivot friction, or by increasing the angular momentum of the disk. Because of the physical limitations of these, all free gyros have an associated drift. Good gyros then are heavy and rotate fast, typically about 24,000 rpm. It is thus difficult to obtain a high quality free gyro that is compact and low power.

Free gyros have complete 360° freedom about the spin axis, and are limited to about ±85° due to gyro lock. When the gyro is spun up, it is done so with the spin axis aligned in a particular direction, allowing the two other orthogonal axes to be sensed. Over time, drift may cause the spin axis to wander from the original position, so that the inner gimbal will have to be 'caged' back to the desired position. Caging can be performed mechanically, or by applying a torque in a sense that causes the gyro to precess to the desired position. Hence, free gyros can be rather complex, mechanically. Nevertheless, they provide angle measurements directly, without the need for integration.

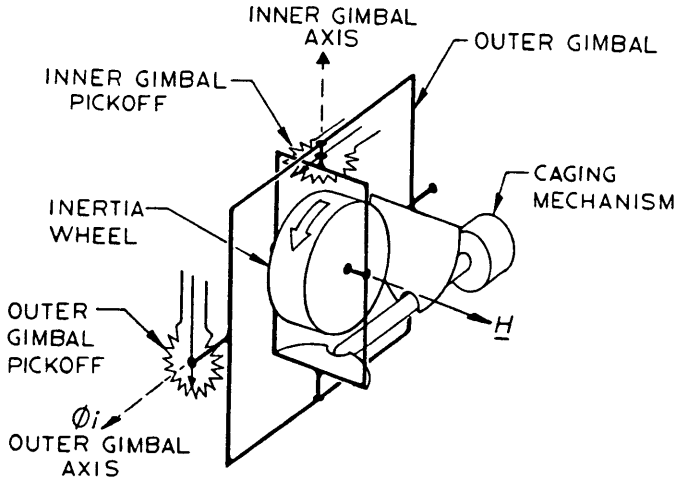

Figure 2.12: Free Gyro [7]

Vertical Gyro

The vertical gyro, shown in Figure 2.13, is a special case of a free gyro that is designed to keep the H vector vertical, usually by sensing the gravity gradient and applying the necessary torque. Hence, yaw is not measured, but roll and pitch are. Although it is possible to design a gyro to measure angles in any two of three orthogonal axes, the vertical gyro has the advantage of using gravity as a reference for alignment.
Like the free gyro, ± 85° deflections are possible about two axes (pitch and roll in this case), while there is no limitation about the other (yaw) axis. For pitch and roll angles that exceed 85°, gyro lock can occur, so this condition is avoided. Maintaining a vertical spin axis, as well as limiting the geometrical problems with drift, is typically performed by actively torquing the gimbals to precess the gyro to the desired orientation.

In most cases, the drift is small enough so that vertical alignment is not needed often. Thus, the vertical gyro can measure pitch and roll under dynamic conditions, relying on a calibration with gravity periodically.

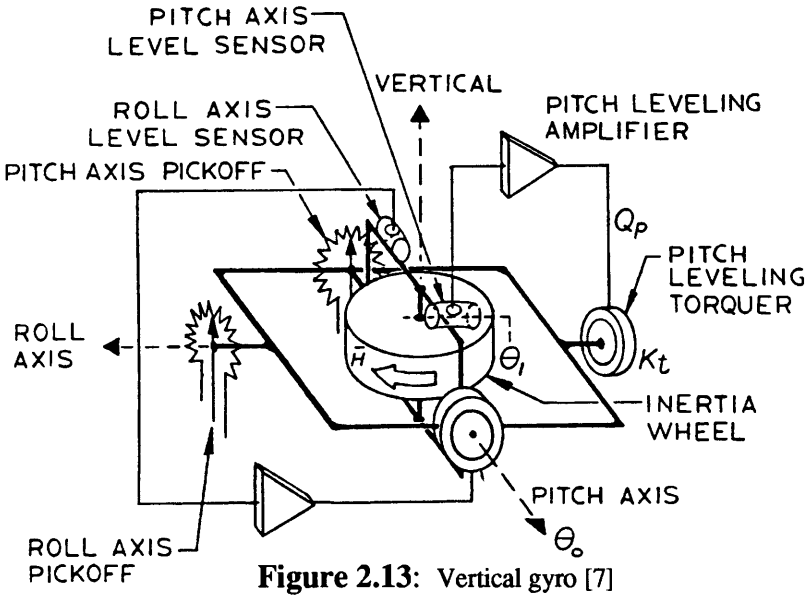

Figure 2.13: Vertical gyro [7]

Directional Gyro

Another special case of the free gyro, the directional gyro has its spin axis aligned in the horizontal plane, so that yaw motion can be detected (see Figure 2.14). As with the vertical gyro, the directional gyro uses gravity as a reference for the horizontal, so that the spin axis can be maintained in that plane. A sensor on the inner gimbal is provided for that purpose. Hence, any pitch or roll motion is decoupled from the outer gimbal, which indicates yaw.

Because of its ability to sense yaw motion, the directional gyro can provide relative heading angle directly, similar to a compass. The main difference is that it is not an absolute heading, so that calibration from an absolute reference is needed periodically due to drift. The directional gyro has an advantage over a magnetometer or other reference-based heading sensors, due to its ability to provide accurate information during dynamic maneuvers. In fact, aircraft often use a magnetometer as an integral part of a directional gyro. This device offers the long-term accuracy of a magnetometer, combined with the dynamic response of a gyro.

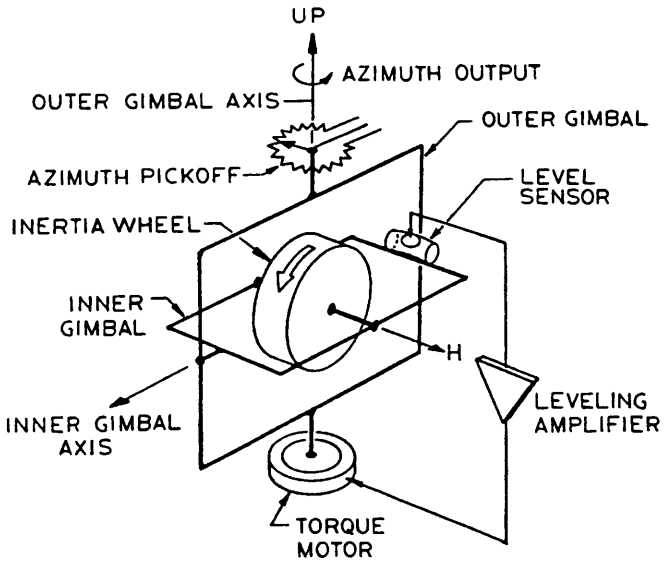

Figure 2.14: Directional gyro [7]

Floatation Gyro

This gyro was developed by Professor Charles Stark Draper, in response to a demand for gyroscopes which required drifts that were much less than those of free gyros, for the purpose of inertial platform stabilization [9]. Like the free gyro, the floatation gyro also indicates angular displacement directly, although the gyro is not ideally torque-free. A spinning disk is typically housed in a sealed cylinder. The cylinder is then submersed in a larger cylinder that is filled with a fluid. The inner cylinder is weighted such that it is neutrally buoyant, making the unit nearly independent of accelerations. A diagram is shown in Figure 2.15.

When an angular velocity is applied as shown, the resultant torque causes the cylinder to rotate at a rate that is proportional to the torque. This is due to the viscosity of the fluid, such that

$$\mathbf{T} = k\dot{\theta} \tag{2.8}$$

The governing gyro equation is then

$$k\dot{\theta} = \dot{\beta} \times \mathbf{H} \tag{2.9}$$

where $\dot{\beta}$ is the angular velocity input, usually shown as ω. Integration of this equation shows that the angular displacement θ is proportional to the input β. The accuracy of this system is dependent upon the fluid property, opposed to the gimbal bearings as with free gyros.

As mentioned, the high potential accuracy of floatation gyros make them well-suited to the stabilization of inertial platforms. Also, their neutral buoyancy makes them very insensitive to accelerations. However, floatation gyros are not widely used when size is limited, and are generally not applicable to small strapdown systems.

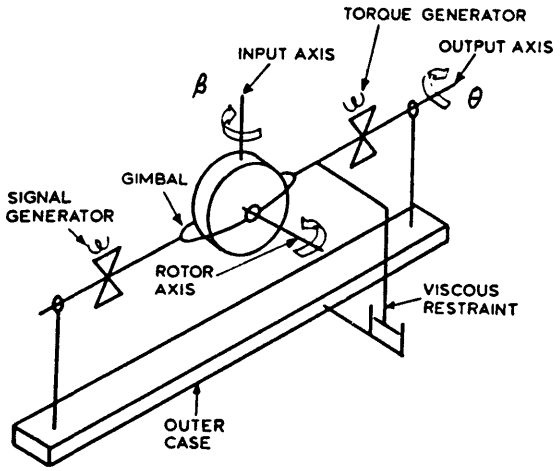

Figure 2.15: Floatation gyro [9]

Rate Gyro

This type of gyroscope outputs a signal that is proportional to the angular rate of the spin axis. It thus measures the projection of angular velocity along one axis. Again, the governing equation for a gyroscope is

$$\mathbf{T} = \boldsymbol{\omega} \times \mathbf{H} \tag{2.7}$$

where ω is the applied angular rate, $\mathbf{H}$ is the angular momentum of the spinning wheel, and $\mathbf{T}$ is the applied torque. Hence, for a given disk speed, the projected angular velocity can be calculated by measuring the resultant torque.

By limiting the precession of the rotor axis about only one normal axis, torque can be found by measuring the steady state deflection (see Figure 2.16). Hence, the angular rate is directly proportional to the spring (or piezoelectric) deflection.

Another way of measuring torque is by using an active nulling system. By sensing the rotor displacement, the signal can be fed into a torquer system, which can null the displacement by passing a current through magnetic coils. Rate can be found by measuring this feedback current. This method is preferred, resulting in better dynamic characteristics and higher accuracies than spring restrained gyro systems.

Common uses of a rate gyro are feedback sensors for control systems that require rate directly (e.g. for damping), but there are increasingly more uses as an input to an integration circuit to find angular displacement. With integration, there are always associated errors, similar to the case with accelerometers. Hence there is seemingly no real advantage to using a rate gyro for this purpose, since there is more electronic complexity involved. However, other angular rate sensors have been developed that are more compact and robust, making integration for angular position an attractive alternative to gimbaled or floatation gyros.

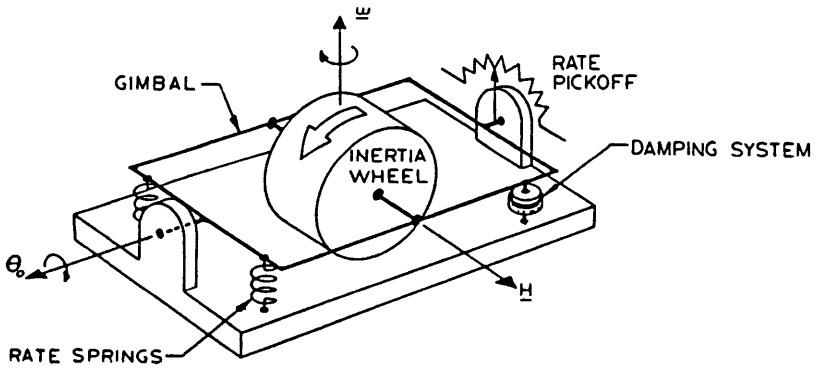

Figure 2.16: Rate gyro, spring restrained

Other Angular Rate Sensors

There any many other types of angular rate sensors that utilize an inertial frame of reference. Some of these have the advantage of being robust in design, so that they are often used with integration electronics as an angular displacement sensor. In addition to

being robust, rate sensors typically do not have hysteresis problems, are lighter, lower-power, and usually much smaller than gyroscopes.

Fluid Stream

This angular rate sensor works on the principle that a particle of fluid will tend to stay in a straight line. The stream of fluid (gas) impinges on a resistive element, similar to a hot-wire anemometer (Figure 2.17). The resistive element is arranged in a bridge circuit, and is adjusted to zero differential output when the flow is in the middle, which is at zero rate [7].

When an angular rate is applied, the fluid stream in the sensor axis is deflected to one side, causing the fluid to cool that side more. The temperature offset changes the resistance of the bridge elements, creating a differential voltage that is proportional to the angular rate. Both single- and dual-axis rate sensors are available using this method.

This rate sensor is a closed system filled with an inert gas, connecting to the outside only through electrical wires. Inside, an oscillator drives a pump, forcing the gas through a nozzle. The resultant laminar stream travels a fixed distance to the hot wire sensing elements, and is recirculated. It should be noted that the pump does not require bearings, but only flexes, resulting in very low wear. Therefore, long lifetimes are possible.

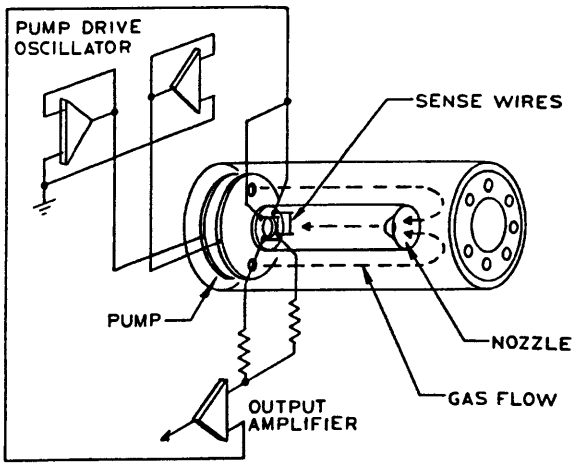

Figure 2.17: Fluid stream angular rate sensor [7]

Fiber Optic Rate Sensor

Another method of sensing angular rate is by measuring the time difference between two equivalent signals that travel the same closed loop path in opposite directions; or equivalently by measuring the difference in traveled path lengths during the same time interval. The latter technique is the basis for the fiber optic rate sensor, or FORS.

The closed loop path is provided by a spool of optical fiber. Multiple windings are used to increase the path length, to amplify the effect. An optical waveform is generated, which is split into a clockwise (CW) and counter clockwise (CCW) signal, as shown in Figure 2.18. If there is no inertial rotation rate, the two signals will travel equal distances in the path. If there is a rotation rate in the CW direction, the CW signal will travel a shorter distance than the CCW signal. When these two signals are combined properly, the small path difference can be detected with high accuracy using interferometry. Hence, rotation rate can be calculated from the interferometry data.

These gyros have yielded extremely accurate results as a rate sensor. In addition to their high accuracy, FORS are very immune to shock and vibration, unlike most other rotation sensors. It is expected, then, that FORS will eventually dominate the gyro market, due to their robust nature and high accuracy.

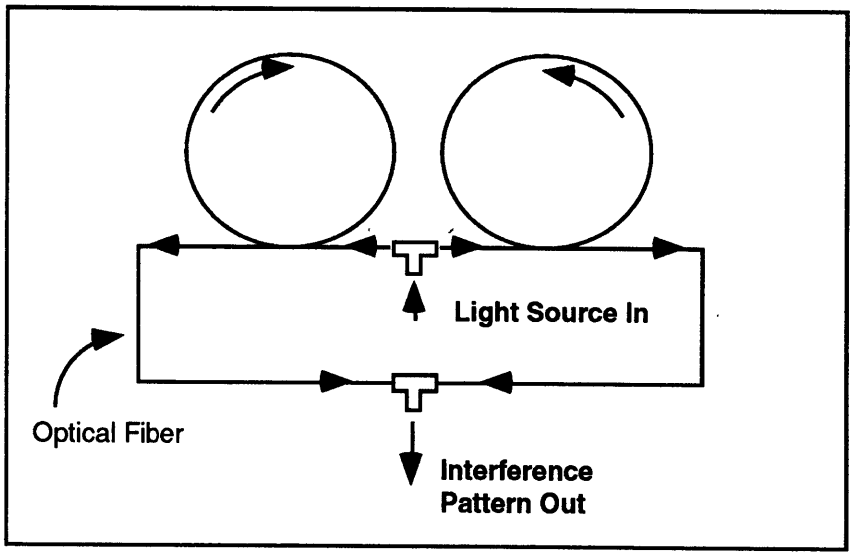

Figure 2.18: Fiber optic rotation sensor.

Coriolis Rate Sensor

The Coriolis-type rate sensor uses vibrating masses to sense angular rate through measuring coriolis accelerations. The coriolis acceleration is the cross product of a radial (linear) and tangential (angular) velocity. In this case, the radial velocity is supplied by the oscillating mass, and the tangential velocity is the rotation input, as shown in Figure 2.19. The radial velocity is sinusoidal, due to the vibrational motion. The instantaneous Coriolis acceleration is proportional to the instantaneous radial velocity of the beam, and the angular rate input. For high vibration frequencies, rotation rates appear constant over a cycle, so that the Coriolis acceleration is also sinusoidal, and in-phase with the driving frequency. Therefore, demodulation of the detected signal is facilitated by referencing the driving frequency. The result is a DC output signal that is proportional to the input rotation rate.

The mass oscillations are typically driven either electrostatically or piezoelectrically at a high frequency. There are different beam geometries and detection concepts. One of the most popular is a tuning fork configuration, in which two beams vibrate towards and away from each other. Recent advances in electronics have allowed construction on a single chip.

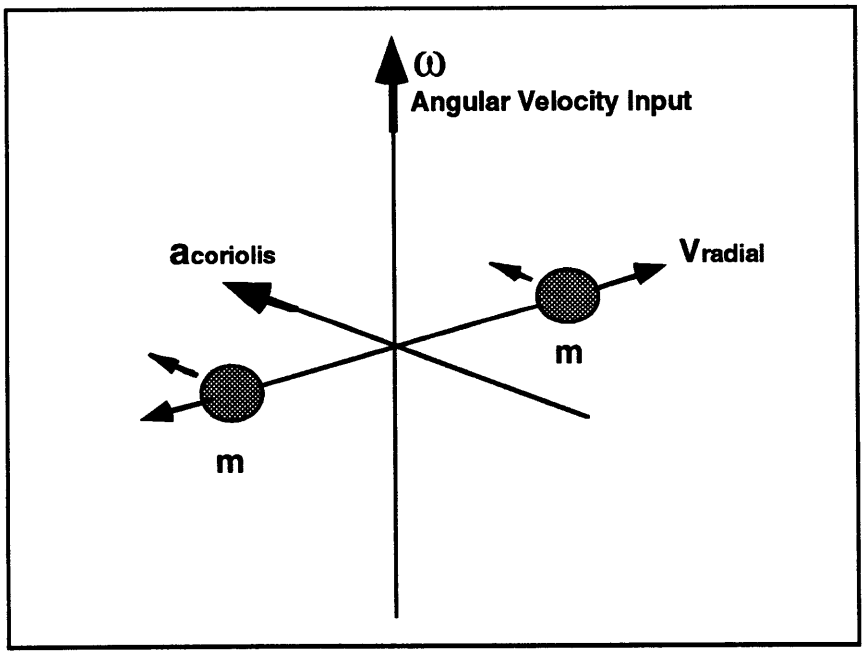

Figure 2.19: Coriolis rate sensor

Gyrocompass

The gyrocompass is a specialty device that uses at least two gyros to sense the planetary rotation. By ensuring that both spin axes are in the horizontal plane, their individual measurements can be added vectorially to yield true north. With this two-axis sensing, the latitude must be known to correct for the actual spin vector projection in the horizontal plane. Both earth and Mars rotate 360 degrees in about 24 hours, or at 15 °/hour. Although this rate is slow, a good gyro system can detect it, if other yaw rates are negligible. A ground vehicle can therefore stop, assuring there are no other angular velocity inputs. This method is even commonly used in ships and aircraft.

Unlike a magnetometer, a gyrocompass locates true north. Therefore, a gyrocompass is an absolute heading sensor, even though it is gyro operated. Its main drawback is that it is much slower than a magnetometer, and not suitable for dynamic maneuvers. In addition, like a magnetometer, the gyrocompass suffers from an inability to provide heading information near the poles of a planet. This should not be of worry, however, since the poles of Mars will not likely be explored for some time. In fact, the first mission is expected to be near the equator, at about 10° latitude.

A gyrocompass needs to be very accurate to sense the small inputs. Hence, these devices are not as miniaturized as other inertial components. In addition, at least one minute is required for an accurate sensing, so that it is useful statically only, and not as a dynamic device. Therefore, its primary use can be as a calibration source for a dynamic sensor, such as a directional gyro.

Star Sensor

Another reference-based method of obtaining heading is by using celestial bodies as the reference. The most accurate means for this is by using the stars. By detecting and recognizing the star pattern in a certain field of view, heading accuracies can be obtained to nearly 0.001° [8]. To obtain absolute heading information on another planet, it is required that the latitude, longitude, elevation of the field of view, and time be known.

The basic sensor package includes sun shade, optics, an image definition device, and a detector with associated electronics. The detector can use photomultiplier tubes, image dissectors, solid state detectors, and charge coupled devices. The recent trend has been to

use CCDs, due to their improved quality over the years, and their inherent ability to detect low light levels.

Because there are so many stars, the top level search is difficult. Stars are identified not on the basis of intensity, since detector sensitivities change with time. Hence, pattern recognition is necessary. There are 9100 stars that are used for direction sensing, due to their sufficient flux density F. A magnitude sensitivity, M, is defined such that:

$$M = -2.5 \log(F) + M_0 \tag{2.10}$$

There are 9100 stars with $M < 7.0$. Some use weaker stars ($M < 8.0$) of which there are 260,000.

There are different methods for star identification, using direct match, angular separation, and phase match criteria. All work well, once a general orientation is established, but not for the initial search. Typically, the sun is used for that, if it is within view.

Although star sensors are very accurate, they are restricted to use during night time (when viewed through an atmosphere) due to the scattering of sunlight. Also, they are generally complex in design, as well as large and heavy.

Sun Sensor

Like a star sensor, a sun sensor uses a celestial body to find heading, given the sensor's elevation, geographical location, and time. Of course, the sun is much brighter, and hence easier to detect. Also, since there is only one source, pattern recognition is not needed, making the entire process simpler. However, there is a rare chance of a singularity when the sun is directly overhead, making it impossible to extract any heading information. Nevertheless, the sun offers a relatively simple way of determining heading angle, allowing accuracies as high as 0.001° for active, and 0.1° for passive devices.

Sun sensors are commonly used on satellites, even more than star sensors. There are a wide variety of configurations, including both analog and digital methods. All of them have one thing in common: the narrower the field of view, the higher the accuracy. It should be understood that the desired information is heading, not elevation. Therefore,

only one of two axes are needed. Also, the sensor should have the ability to provide information in a hemispherical field of view (2π steradians), in order to be mechanically passive.

Analog Photocell Sun Sensor

The simplest type is the analog photocell arrangement. This uses the principle that the photocurrent generated from a given cell is proportional to the cosine of the angle of incidence. Figure 2.20 shows a solar cell at an angle of incidence, θ. For a given cell area, A, the current generated is proportional to the projected area, $A\cos\theta$. These cosine detectors provide good accuracies for large angles of incidence. When the angle of incidence is low, the cosine is inherently insensitive, so large errors can occur.

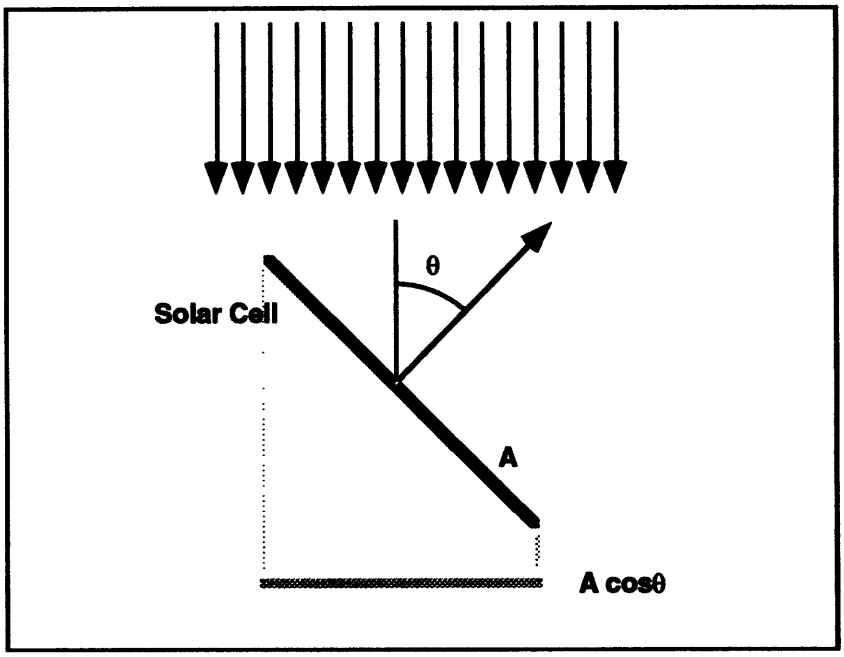

Figure 2.20: Cosine angle detection

Because the photocell efficiency changes with time and temperature, it is customary to use pairs of photocells in a near-parallel configuration, differencing their output. A diagram of this is shown in Figure 2.21. In this illustration, the sensor axis is offset by a small angle ϕ. Because the cells are nearly parallel, the angle of incidence for each *cell* is large. For large angles of incidence, or small ϕ, the result is a near-linear relationship of current vs angle. It is assumed that a change in cell characteristics will equally affect each, making the pair relatively insensitive to time and temperature, when near the null

point. In other words, the bias stability is improved using pairs, although the gain can still be affected.

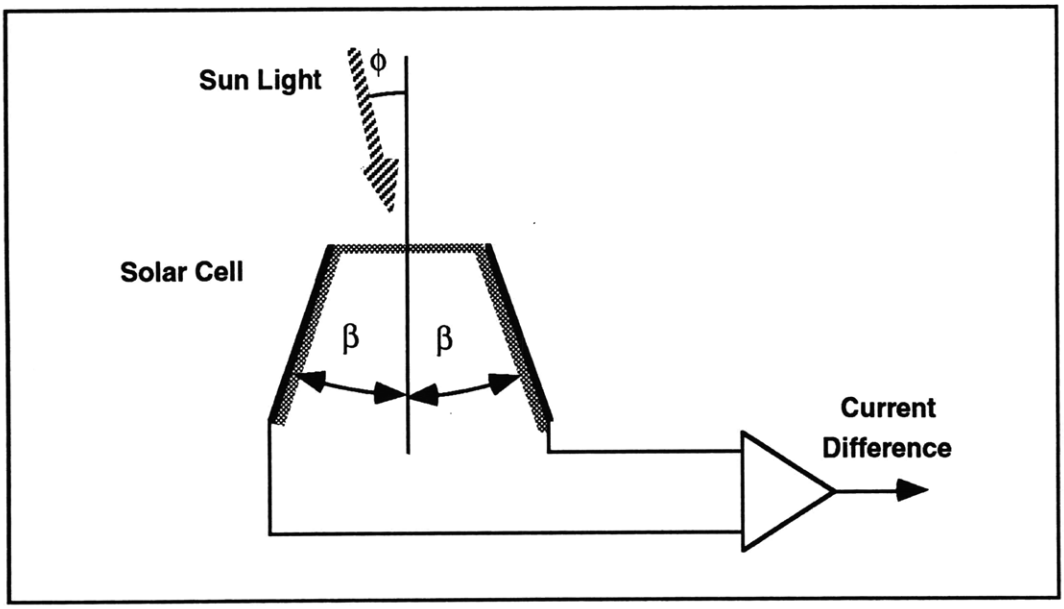

Figure 2.21: Cosine detector using a pair of solar cells. This arrangement can have very high sensitivity at the cost of a narrow field of view.

In order to achieve a large instantaneous field of view, a large number of pairs are used, each pair having a small field of view, but with good accuracy. The result, however, is a complex arrangement of many detectors.

However, since only one axis in the field of view is needed for the rover, the system can be arranged such that gravity is used to passively orient the sensor platform, so that the elevation component is decoupled. This can be done by gimballing the platform from the top. The result is a hemispherical field of view (except for the singularity condition overhead), with accuracy limited by the number of sides. For any given sun angle, there will exist a pair of photocells with a maximum angle of incidence, which is the most sensitive pair.

One drawback to this is that the dynamics of leveling will require the sensor to be used under static conditions, or dynamically with filtering (large time constant). Even more importantly however, is that the operation is still intensity-dependent, so that asymmetrical attenuation from dust, for example, can result in large bias errors.

Digital Position Sun Sensor

A better method of sun angle detection is through position sensing of the sun's rays on an array of small photocells. One method is to pass sunlight through a slit, which is refracted onto the photocells. The position of the projected line of light is then representative of the sun angle.

Rather than constructing an array of very finely spaced cells and scanning each, the cells are divided into separate rows, each with twice as many cells, until the last row has the desired resolution. By doing this, the required number of channels to read is much less. In fact, every row has one bit associated with it, with every other cell connected to the same bus. Therefore, only one channel per row is needed, even in the high-resolution rows. Every sun angle reading will thus have a digital word associated with it, with the number of bits dependent on the accuracy and field of view.

Other Sun Sensors

There are a wide variety of configurations for sun sensors, both analog and digital. Terrestrial applications of sun sensors are typically for solar panel tracking, and also employ active tracking. Active sensors use feedback to mechanically position the detectors, which are usually analog photocells. Hence, high accuracies are obtainable, since fine positioning is performed with a very narrow field of view.

To minimize complexity, a passive sun sensor is desired with an instantaneous field of view of about 2π steradians. Space qualified sun sensors, typically for use in satellites, are available with a few of them offering such a large field of view. The main drawback to these is the high cost, and their irradiance-based operation.

A sun sensor can also be made out of some simple lenses and a position detector. Today, both discrete arrays (CCDs) and continuous position sensitive detectors (PSDs) are available in small sizes, with low cost, and good accuracy. A prototype sun sensor was constructed out of a fisheye lens and a PSD, offering a hemispherical field of view, with high-speed position sensing. The basic arrangement is shown in Figure 2.22.

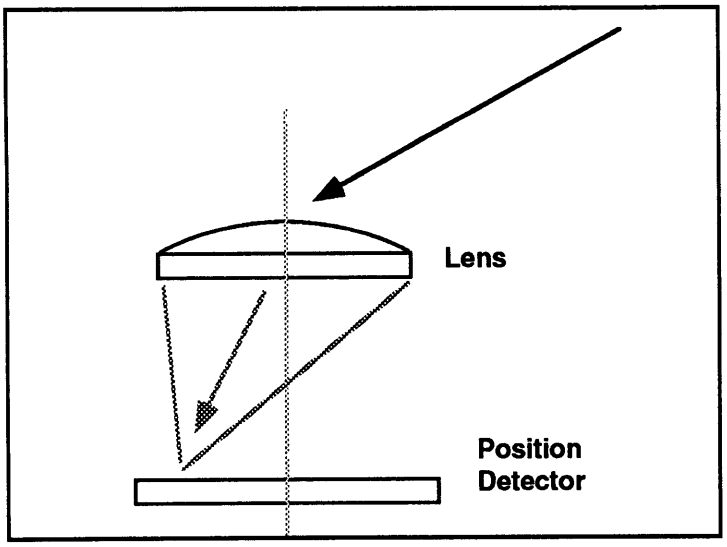

Figure 2.22: Sun angle measurement by focusing light on a detector that can sense the focus spot position.

Sunlight enters the fisheye lens, and, with the addition of an additional convex lens, is focused to a spot on the PSD. As the sun angle changes, the spot traverses the PSD. When vertically mounted, the sun elevation is represented by the focus spot distance from the center, or radial distance. As the sensor is rotated in azimuth, the spot traverses a circle at the same radius. Therefore, heading is indicated by the location on the circle. The operation of this is discussed in chapter 4.

2.2.4. Summary

The rover needs real time position estimation, so it can maneuver as needed. This calls for a self-contained system. However, dead-reckoning hardware may not be sufficient to keep the daily position errors within bound. Hence, a position calibration may be needed periodically. One important point is that knowledge of position is not necessarily needed on the large scale, since it is mainly important to explore new areas, and not map them.

There are many hardware options that fall under the defined categories of dead reckoning and reference-based navigation. The methods of operation of the sensors can be very different, yet they all are possible options for autonomous rover navigation. It remains, then, to choose the appropriate sensors that will best satisfy the navigation requirements within the limitations of the microrover packaging.

CHAPTER THREE
CHOICE OF NAVIGATION SENSORS

3.1. REFERENCE OR DEAD-RECKONING

3.1.1. Introduction

The allowable errors in position are still defined as 10% of the traveled distance that occurs in between every uplink. For complete inertial navigation, this directly applies. When measuring traveled distance and heading separately, it is less clear to apply the 10% requirement to the individual sensors.

One important difference between heading and translation sensors is that heading can be calibrated with an absolute reference (magnetic field, celestial bodies) whereas traveled distance can not be updated as easily. Because of this, one of the two dead reckoning parameters is unbounded, which is enough to fall short of the requirements. Hence, if the translation sensor is not accurate to at least within 10%, then position calibration will be required during the course of travel. The problem with position calibration is that it requires an existing external system that is functional, except for the case with vision/rangefinding referencing, which is also self-contained.

This chapter will review the navigation options that were discussed in Chapter 2, and present the selection of the final microrover navigation hardware.

3.1.2. The Problem with Reference-Based Navigation

The reference-based navigational methods described in the previous section are fundamental to navigation on earth. However, all except vision/rangefinding require an existing infrastructure that is operational. However simple that may seem, there is no guarantee that a GPS satellite or beacon network will be available. Hence, it would be optimistic to design a rover that is dependent on such a system. If one is available, it can certainly be used with the proper communications installed on the rover, but there should be no dependency at this point. One can say that a collapsible beacon system can be deployed by the rover itself, but even that is rather complex.

Therefore, it is left to vision/rangefinding as the remaining reference-based navigation. This method has many advantages:

1. Target identification can be performed through the same sensor for both man and rover (from an image)
2. Referencing is done locally, which is what is needed. Therefore, local navigation accuracy is much better than that of beacons or GPS.
3. Hazard-avoidance can be performed with the same sensors.
4. Self-contained system
5. Improved performance over irregular terrain, for recognition reasons. This is contrary to dead reckoning translation sensors such as accelerometers or drag wheels.

An image-based reference navigation system would greatly benefit a rover's operation, in the same way our eyes can help us navigate. Some of the problems with this system are chances of losing sight of the target, or inability to correlate images (section 2.2.2). For these reasons, it is suggested that this is used along with a dead-reckoning system. In fact, single camera systems usually require additional translation sensors to measure the distance traveled between each sampled image.

More importantly, however, are the computational requirements of such a system. Even in a structured environment, a vision system is very difficult to operate successfully. On Mars, correlation will be more difficult, and the computer resources will may not be available. However, if such a system is developed, it is suggested to be integrated into the design, as long as its individual requirements are acceptable (size, power, computations, etc.). In fact, a vision-based navigation system is currently being studied by Lynn [22], using the same camera as for video transmission.

3.2. CHOOSING DEAD-RECKONING NAVIGATION HARDWARE

Since no reference-based navigation hardware can be planned on, it remains to choose the hardware for dead reckoning. These will be discussed in the same order as in the last section, which describes the operation principles.

3.2.1. Complete Inertial Navigation System

A complete INS is possible, but with the limitations of size and power on the rover, the commercially available units can not be used. Most of these units are large and heavy, since they include as many as 3 accelerometers and 3 gyros, as well as supporting electronics.

As mentioned in 2.2.3, the use of IMUs for ground vehicle navigation is difficult due to the presence of gravity and a rough terrain. Sophisticated electronics and data processing can account for these difficulties, but at a cost of power, processor dedication, size, etc. Current systems operate on a minimum of 10 watts, and are not only large, but extremely expensive. Any system that could be packaged on the rover would likely suffer from position errors that grow rapidly.

Inertial navigation systems are fundamental to many vehicles, but there are only rare occasions when they are used as the sole navigation system for long periods of time. They are complemented well with a reference-based positioning system such as land beacons or GPS. In fact, an important trend for the Department of Defense has been to embed GPS with INS for many aircraft. This allows periodic calibration, keeping drift errors to a minimum, which exemplifies the best use of an inertial navigation system.

3.2.2. Longitudinal Translation Sensors

Accelerometers

As previously described, a single accelerometer along the longitudinal axis of the vehicle can provide only forward/backward translation information. This is essentially a simplified INS, yet the same principles of operation apply.

The problem, again, lies in the fact that there is no method of updating translation, but only rotation. Therefore, an accelerometer will suffer the same problems of an INS, of which accelerometers are the core. Therefore, it is not likely that translational motion can

be accurately monitored over time with only an accelerometer. However, in some instances an accelerometer may supplement other translation sensors.

Drive Wheel and Drag Wheel Odometry

There are few remaining choices for sensing translational motion, both of which are based on wheel rotations. The drive wheels have the advantage of being already needed for mobility, including their respective encoders or tachometers for speed control. Hence, all six wheel rotations will be available for odometry as well.

As previously mentioned, drive wheels have the advantage of averaging, since there are six of them. This can especially be advantageous over rough terrain. The main drawback is their potential for large slip, since they apply a high torque. This can falsely indicate translation.

An unpowered drag wheel, however, does not suffer the problem of slip since its resistive torque can be negligible. This will indicate vehicle translation much more accurately than the drive wheels under high-slip conditions. The main problem is that packaging such a device is rather limited. The result is a small wheel diameter, which is less accurate over irregular terrain.

The drive wheel and drag wheel odometer are also complementary, and so it is suggested that at least one drag wheel is mounted, and that the available data from the existing drive wheel rotations is combined with the drag wheel data. Therefore, sensor fusion of drive wheel and drag wheel data is important for the best sensing of translational motion. It should be noted that lateral motion, from sliding down a hill for example, will not be detected by these odometers. However, this blindness should be acceptable.

In short, the options for measuring longitudinal translation is limited, and odometry is the suggested sensing method. The rover has the advantage of using the ground as a reference for measuring, which is an option that aircraft or spacecraft do not have. However, there will likely be difficulties in maintaining good accuracy over rough terrain, which may call for other sensor data to improve performance. Design and integration will be presented in later sections.

3.2.3. Heading Sensors

Magnetometer

These sensors offer the rare opportunity to extract heading from an external reference easily, and so are valuable to many earth navigation systems. In fact, MITy-1 uses such a sensor for heading in its dead reckoning navigation system (section 4.2.1). It proved to be accurate, although its response was slow due to the required filtering. Unfortunately, its sensitivity to magnetic fields limited operation to outdoors, far away from any magnetic disturbances. This only made testing more difficult, but in fact has little relevance to remote applications.

Many aircraft use magnetometers for steady state calibration of inertial systems, since it is an accurate external reference. Again, it should be reminded that the magnetic field does not necessarily indicate true north. This is not usually a problem, since the location of the magnetic pole is known with high accuracy, and is updated as it moves over time. In fact, even when it does move, the local effect at most areas of the planet are negligibly affected. It is more important to have relatively stable heading reference, regardless of its orientation with respect to a common frame.

The main problem with using a magnetometer on Mars is that the magnetic field is significantly weaker; about 1/5000 of that on earth. This prevents use of such a sensor on Mars, particularly in the presence of electromagnetic disturbances from the drive motors and other rover electronics. Therefore, it is not practical to use such a sensor on the microrover.

Differential Odometry

Since drag wheel odometry will be used, turning information can be found using a pair of laterally-separated wheels. This method, as mentioned in section 2.2.3, is subject to very large error growth over irregular terrain, mainly due to the limited wheel spacing available. There are other methods that can yield much higher accuracies in heading, with similar size and weight.

Gyros

The numerous types of gyroscopes (and angular rate sensors) discussed in the previous section are very different in operation, yet all are similar in that they yield angle

information relative to an inertial reference frame. Gyros have the advantage of a high bandwidth, allowing continuous, real-time use. Additionally, unlike accelerometers, gyros can be easily calibrated with an external heading reference. This is an important advantage, especially when considering the higher drift rates associated with miniaturized packaging.

Many of the gyros available today have been designed to be 'inertial quality', with resulting drifts better than 0.1 deg/hour. This is far better than what is needed for the maximum two-hour runs that are expected. Of course, use of such a high quality gyro package has its drawbacks. In addition to the extremely high price, ($10 K and up), the gyros require significantly higher power and space than what is available on the rover. Care must be taken in descriptions that are relative, since the military's version of 'miniature' is usually the size of one complete rover platform

Despite the large variety of gyros available, there is a limited choice that meet the needs of the microrover. Most notably the size, power consumption (and format), and cost. Fortunately, there recently has been a trend to truly miniaturize gyros, and sell them at a more reasonable cost. These are mainly rate gyros, but there is also one directional gyro.

The manufacturers of the leading miniature/micro gyro candidates are:

- Systron Donner
- Gyrations
- Watson
- Murata
- Allied Signal
- Litton
- Humphrey
- Textron
- Northrop
- C.S. Draper Laboratory (prototype)

Again, there are many other gyros on the market, but all fall short of the rover's requirements, usually on the basis of size and weight. In fact, most of the following sensors also are not well-suited for use on the rover, but are presented as a gage of what is currently available. A summary of their specifications is presented in table 3.1.

Rate or Directional

Fundamentally, the type of gyro that is needed is a directional gyro. This sensor yields angle directly, without the need for external integration electronics. Additionally, the directional gyro yields not just angle about its one sensitive axis (which is confined by the mounting), but rather the angle relative to the gravity vector, or the vertical. Therefore, at inclinations, the navigation system does not have to correct the heading output, since this inherently yields heading directly. A rate gyro, on the other hand, measures rotation about its axis regardless of the orientation of its sensing axis. Since *heading* is the desired measurement, corrections need to be made in software to correct for the inclinations of the rover at the time of rotation. Therefore, a directional gyro offers two main advantages over a rate sensor:

1. Output is angular displacement directly. Thus, no additional integration electronics are needed.
2. Because of the referencing to the gravity vector, directional gyros yield heading changes, so further corrections are not needed.

The second point may not be important, since full inclines of more than 30 degrees will not likely be traversed often, so the cosine corrections that are needed are likely to be negligible. Also, calibrations can be made more often during inclined travel, limiting the projection error. Nevertheless, a directional gyro would facilitate the processing of the data, and would save the added trouble of rate integration.

Directional Gyro

The Gyrations 'Gyroengine' is a very new product on the market. The company offers both vertical and directional gyros, 90% of each is plastic. Their goal is to manufacture these gyros cheap enough to use in household computer and video equipment, such as for a remote mouse. However, they also realize a need for such gyros in navigation, and plan to introduce 'navigation quality' upgrades sometime in 1994.

The directional gyro satisfies many of the requirements of the microrover. Some of its advantages are its light weight, small enclosure, low power consumption, and digital output. The company likes to refer to it as 'the size of a film cannister.' Some of its

interesting features include an optical pickoff, which is friction free. This is done by reflecting two LED signals off of a ring of mirrors, which result in a dual phase quadrature output.

The performance is listed at 2°/min, which is about the maximum acceptable drift for the microrover. This is generally considered to be poor, but is tolerable considering the packaging. This performance is quoted at an operation voltage of 5 volts DC. At the maximum input of 12 V, the rpm of the disk is much higher, and so better performance is expected due to the higher angular momentum. The 'navigational quality' directional gyro that is expected to be introduced later will have active torquing instead of a pendulous reference for maintaining the rotor axis horizontal.

The current directional Gyroengine was ordered for use on MITy-2, as a demonstration of using a small gyro for navigation. However, Gyration's initial estimates of delivery dates were pushed farther and farther back, and it was finally decided to cancel the order. The gyro would have probably been suitable for MITy-2 navigation, but would not have been applicable for the final design. The main reason is that the gyro is not designed for aerospace uses, but instead for low-cost manufacturing and commercial applications. In addition, the company was founded in 1989, and has only recently delivered its first production model. For these reasons, the Gyration directional gyro is not suitable for use on the final microrover.

Rate Gyros

With the absence of the directional gyro, only rate gyros remain as possible options for the microrover. Rate gyros initially had use when angular rates were needed directly, such as for damping in a missile control system. These days, the improved quality of rate gyros have led to their use as inputs to an integrating circuit, so that angular displacement can be calculated with a high degree of accuracy.

Because of the need for an integrating circuit, additional error sources are introduced. With proper design, however, the errors due to integration itself can be negligible. However, drifts in the rate gyro output can result in large integrated errors, since the angular error due to a constant rate bias error increases linearly with time. Therefore, temperature stabilization or characterization of the entire gyro/electronics assembly is very important. Despite the added complexity of the integrating electronics, rate gyros

are commonly used to measure angular displacement, since electronic complexity is often preferred to mechanical complexity.

Systron Donner

Two gyros are made by Systron Donner with essentially the same characteristics: The GyroChip and the QRS-11. For about $500 more, the QRS-11 offers better temperature and g sensitivity, a metal housing, and other features that make it appropriate for aerospace applications. Any further discussion about the Systron Donner gyro can be assumed to be the QRS-11, and not the GyroChip.

The QRS-11 sensor is a small, rugged device which utilizes a micromachined 'tuning fork' to sense coriolis accelerations, indicating angular rate. The sensor is solid state, using piezoelectrics to drive the oscillations, and is thus more robust than traditional gyroscopes.

Solid state gyros such as these are not as accurate as actual gyroscopes, but they can be completely packaged as a small DC-DC device. The QRS-11 has the standard features that are desired (see table 3.1), including a low ±5 VDC power supply. Additionally, the environmental rating is particularly good for sensors of this type, and the operating life is rated at over 100,000 hours (10 years) MTBF. Most importantly, though, is that it has proven success in many applications, including the Hughes Maverick missile program, which required 12,000 units. The proven history is very important in limiting the risk associated with any piece of hardware.

The performance of this sensor is acceptable, with a 0.3°/min short term bias drift and a temperature coefficient of 0.3°/min/°C. Therefore, it is very important to maintain a stable temperature, as is the case with all rate gyros, particularly solid state. Zero updates by the rover (stopping) can be used to subtract this bias, though, if temperature cannot be controlled. Systron Donner has achieved better temperature characteristics than other coriolis-based rate sensors, due to the fact that the active components are micromachined from quartz, instead of bonding dissimilar materials. In general, although solid state gyros are more temperature sensitive than mechanical gyros, they have the advantage of less mechanical complexity and smaller packaging.

In summary, the Systron Donner QRS-11 has good performance, proven reliability, and satisfies the hardware requirements of the microrover at a reasonable cost.

Watson

A similar tuning fork design is made by Watson Industries. This device is similar in size to the Systron Donner sensor. It also satisfies the standard packaging requirements of the microrover. The gyro has a relatively low cost, especially for the models with lower rate sensitivity.

The main drawback to the Watson gyro is that the output is not well-suited for integration, as the noise level is rather high, as well as the bias instability, especially when subjected to vibrations. In fact, Watson strongly suggests a special vibration isolation mounting to help alleviate this problem. Another problem is the higher sensitivity to temperature changes, which is in part due to the larger vibrating element (which is metal), and also due to the dissimilar materials in these elements.

Although the listed specifications do not make it obvious, the performance of the Watson sensor is rather poor.

Murata

The Murata rate sensor is again similar in design to the Watson and Systron Donner products. One important difference is the vibrating beam shape, and the excitation and detection arrangement. The result is a very sensitive device that has a much better immunity to vibrations than Watson's, making the integration process more accurate. In fact, one of its main applications is in car navigation, so not only is the Murata rate sensor applicable for integration, but it is also the least expensive sensor available, with a single unit cost of only $300 ($125 in large quantities). It is also one of the lightest sensors available.

The performance of this is mainly limited by the stability of the zero point, which can lead to drifts in excess of 5°/min, even within a small temperature range. The sensitivity of the scale factor is also much higher than the Systron Donner gyro, resulting in a need for a high amount of thermal control, or much more frequent updates. This sensor works best when the integration times are under 20 seconds or so.

Despite these points, the Murata gyro is similar to the Gyrations directional gyro in that it is targeted for eventual low cost applications, and is not designed for aerospace use. Even if it was, the performance is not near that of the Systron Donner model. Hence, this gyro is not recommended for use on the final rover. However, its low-cost has allowed it to be used on the MITy-2 prototype, for proof-of-concept purposes (section 4.2.2).

Humphrey

Although Humphrey has one of the largest selections of gyros, the majority of these are much too large and heavy for use on the microrover. In addition, very few operate on a DC power supply. Nevertheless, a couple of their smallest packages can be considered.

The RG78-0300 is a spring-restrained rate gyroscope that also contains internal circuitry to make it a DC-DC device. This is similar to the much smaller -0200 model, which is an AC-AC gyro. Hence, it can be assumed that the larger packaging is due to the additional electronics.

The RG78-0300 is usually used directly for rate sensing, so the performance as an input to an integrating circuit is not clear. Because it is spring restrained, there is a zero offset of 0.1 % of full scale, which represents 6°/min (for a 100°/sec range) due to hysteresis alone. In addition, the gyro is big and heavy, and draws about 10 watts, which makes it hard to justify for use on the rover.

Another rate sensor made by Humphrey that is DC-DC is the RT15-0102. This is distinguished from the RG model above since it is considered a rate transducer (RT), and not a rate gyroscope (RG). This 'gyro' works on the principle of a fluid stream impinging on a set of 'hot wires'. Humphrey makes a large variety of transducers of this type, most of which require the usual 4 kHz supply (for the fluid pump). However, some are made with internal electronics, so that they can take a DC power supply. The RT15-0102 is such a device, containing additional electronics to amplify the millivolt sensor output. Like the Humphrey rate gyroscope above, the cost of the additional internal electronics is a larger and heavier package, but the power consumption increase is not as drastic, resulting in a draw of 1.5 watts. In addition, there is no hysteresis, and the linearity is improved. The main drawback to the RT15 sensor is that the null voltage drifts, so that DC rate measurements can be inaccurate. In fact, the performance of this sensor is excellent for measuring AC inputs. These sensors have had success as antenna stabilizers

on large ships, where the inputs are low frequency, but not DC. Hence, a highpass filter with a very low bandwidth (< 1 Hz) solved the null drift problem. The rover can also use this filtering technique, since absolute DC inputs are not likely due to the irregular terrain. In addition, zup calibrations can reduce the effect of the null drift.

In short, both the RG78 and RT15 gyro can be considered candidates for the rover, although they are relatively big and heavy. The RT15 is clearly a better choice, however, mainly due to its lower power consumption, longer lifetime, and higher accuracy. Humphrey gyros also have a long history of proven flight worthiness, and their cost is comparable to other flight-qualified gyros.

Litton

One of two dynamically tuned gyros discussed, the Litton G2000 offers some very good performance in an extremely small package. It is one of the smallest and lightest gyros available, no bigger than one inch on any side. On top of that, its random drift (uncompensatable) is less than 0.02°/min, although 'compensatable' drift can be as high as 5°/min (this can be measured periodically and subtracted).

The big issue with this gyro is that it is not ready to interface to a system. A large amount of additional electronics need to be supplied in order to:

1. Supply different AC signals for the motor and the variable reluctance pickoff.
2. Sense the pickoff signal and deliver a current to the gimbal torquers, for nulling.
3. Measure the torquer current and convert it to a usable voltage

Therefore, a significant amount of additional work is required to make the gyro ready to use. Although the gyro itself is miniaturized, the resulting electronics are much larger. Another manufacturer quoted the electronics card as 4 x 6 inches, and requiring 3 watts. Litton presently does not supply such a card, but will in the future.

If the supporting electronics can be made small and low-power, this gyro would likely be a leading candidate. Its random drift far exceeds what is needed for the rover, and the individual power consumption is low. Imbalances in the rotor result in a drift as the result of accelerations, but that is well known for each gyro, and can be measured (with accelerometers) and corrected in software. Bearing failure is a possibility, but this gyro

has been used in many military applications, and maintains a MTBF of about 50,000 hours. Most of its uses are in missiles, torpedoes, and antenna stabilization.

To summarize, the Litton G2000 gyro offers excellent performance, especially considering its moderate $4500 price tag. The main drawback is the large amount of supporting electronics that are needed, which increase the system size and power requirements.

Textron

Textron also manufactures a dynamically tuned gyro that is very similar to the Litton model. It works on the same principles as the Litton Gyro, and unfortunately also requires a supporting electronics card for the same reasons. The Textron gyro has about the same random drift as the Litton, at a similar size.

Textron also has similar models that have better drift characteristics at the cost of higher power consumption. Among other uses these have been demonstrated as a gyrocompass with very high accuracy. However, the rover requirements result in a need for smaller complete size and power, and not high accuracy.

Textron's gyros in general have been used in many military applications. However, this model has not, and is still considered as pre-production. Because of this, the price is quite large, at $18,000, although this is expected to decrease by about $10,000 in the future [11].

Since the Textron gyro is so similar to Litton's, its immaturity on the market, and higher price results in a low overall rating.

Allied Signal

Like Humphrey, Allied has a large variety of flight-qualified gyroscopes, most of which are much larger than what the rover can afford. One of their divisions specializes in short-lived applications, which have spring-initialized rotors that coast at high speed for typically two minutes. Although these can be rewound and used again, this is a difficult task to accomplish autonomously.

Another division specializes in rate gyros, which are much smaller. Their model is a spring-restrained rate gyro which is the only single-axis gyro offered that has a dedicated electronics box to make the system DC in DC out. The entire unit is comparable to that of the Humphrey gyros, and is lighter weight, but at a higher cost. Drift is of the order of 2°/min, which is mediocre for the $4000 price. One of the reasons for the 'drift' is because of its zero offset stability, similar to the Humphrey RG gyro.

In addition, Allied offers a DC tuning fork rate sensor, similar to the Systron Donner device, in limited quantities. Model TF-200 is not in full production yet, but should be by 1995 [12]. The tuning fork is steel, like the Watson model, but has much better performance. The weight and power consumption is large for a coriolis sensor, and the cost at this point is also large, mainly due to the fact that it is preproduction. Future improvements, including a quartz element, are expected to make this sensor compete with the Systron Donner QRS-11.

The Allied Signal rate gyroscope, when used with the converting electronics, is a relatively large, heavy system that has only moderate performance. The TF-200 has the advantage of being solid state, but its power consumption is large compared to other similar gyros. In addition, the TF-200 is not in full production yet.

Northrop

Like Allied Signal, Northrop also makes a large variety of AC powered gyros for many applications. Although they make a rather small spring restrained rate gyro, the dedicated electronics are still necessary. However, Northrop is in the process of developing a Coriolis-based rate sensor to compete with the Systron Donner sensor. This device uses two oscillating proof masses instead of a tuning fork design. The active elements are micromachined out of silicon, opposed to quartz. This unit will additionally provide translational acceleration in a direction that is 90° from the angular rate axis. The micromachined accelerometer/gyro (called MAG) is therefore a 'multisensor', and is intended for later use on a 6 degree of freedom INS, requiring only 3 total sensors instead of 6.

The MAG will be a DC-DC device, requiring ±15 V supply, and providing an analog signal that is proportional to rate. Initial sensors will be 1.8" x 1.8" x 0.25", and later reduced to 0.75 x 0.75 x 0.25". Cost and performance is expected to be similar to that of

the Systron Donner sensor, although the performance data is not available at this time. Prototypes are expected to be available by 1994 [13].

C.S. Draper Fiber Optic Gyro

As mentioned, fiber optic gyros are expected to eventually take over the mechanical gyro market. There have been many papers on these gyros, and it continues to be a strong area of research. The main problem is that although they are popular on paper, fiber optic gyros remain very expensive, preventing widespread use. In addition, there are limited manufacturers of these sensors. All of their products have very high accuracies, but again are very expensive.

The Charles Stark Draper Laboratory, in a combined effort with the Jet Propulsion Laboratory, has been developing a gyro known as FORS (fiber optic rotation sensor) since 1989. The purpose was to develop it for use on an upcoming interplanetary mission, where long lifetime, low weight, and low-power consumption are imperative. Of two devices made, the smaller, 3-inch diameter gyro is feasible for use on the microrover. The main limitation is the supporting electronics, which currently is spaciously packaged in the size of a shoe box. However, the relatively 'simple' electronics should be able to be packaged small enough so that the entire gyro/electronics assembly fits within a 3 inch cube [14]. In addition, the power consumption can be reduced to under one watt. It should be noted that this includes integration electronics. If funding is available, the electronics can be further reduced using CMOS circuitry.

The FORS has an advantage of very high accuracy. Although its drift is under 0.05 °/hr, the drift is expected to increase to about 1°/hr in order to lower the power requirements below one watt . This performance would currently exceed any other sensor of the same power. In addition, fiber optic gyros are robust, with no moving parts, and very insensitive to vibrations. The effects of temperature are well known, so that software corrections are possible.

In short, the Fiber Optic Rotation Sensor is a high performance device that may be feasible to package on the rover, with some repackaging of the electronics. The main issue is the cost of this effort, and the uncertainty of the final size and weight.

Choice of Gyro

Most of the gyros discussed are summarized in Table 3.1. The majority of these are coriolis rate transducers and rate gyroscopes. Although the listed specifications are not in depth, it should facilitate the comparison of the various gyros.

Manuf.	Model	Type	Operating Voltage	Power (watts)	Weight (grams)	Size (cm)	Cost $	Overall Rating
Gyrations	GyroEngine	Directional	+ 5	0.15	34	4.5x3.3D	1500	poor
Systron Donner	QRS-11	Coriolis rate sensor	± 5	0.8	60	1.4x3.8D	1500	excellent
Watson	ARS	Coriolis rate sensor	± 15	0.3	85	3x3x8	600	poor
Murata	GyroStar	Coriolis rate sensor	+ 12	0.2	45	2.5x2.5x5.3	300	poor
Humphrey	RG78-0300	Rate gyroscope	+ 28	10.0	340	5x5x7.5	2945	poor
Humphrey	RT15-0102	Fluid rate sensor	+ 12	1.5	255	5.6x7.6x 3.8	2170	good
Litton [1]	G2000	Rate gyroscope	13 rms 800 Hz	1.0	25	2.4x1.9D	4500	good
Textron [1]	XXI	Rate gyroscope	12 rms 1200 Hz	1.0	19	2.5x1.57D	18000	poor
Allied Signal		Rate gyroscope	+ 28	-	200	6.5x2.8D	4000	poor
Allied Signal [2]	TF-200	Coriolis rate sensor	± 15	2.7	150	3.2x4.4x4.2	3500	good
Northrop [2]	MAG	Coriolis rate sensor	± 15	-	-	4.6x4.6x0.7	2000	good

[1] Specifications are for sensor only. Additional electronics are required.

[2] Not in full production at this time

Table 3.1: Summary of Gyro Specifications

Out of the gyros and rate sensors discussed, there is only one currently available that combines good performance, low power and mass, small size, complete packaging (ready to use), and is acceptable by military standards. This is the Systron Donner QRS-11. All the other flight-ready hardware falls short either due to size and/or power requirements, or because a complete system is not available. The other solid state devices are not suitable for non-terrestrial applications.

The next closest choices would likely be the Humphrey, the C.S. Draper FORS, the Litton G2000, and the Northrop MAG. The Humphrey rate transducer is at least acceptable, since it is relatively low-power and reliable. Although no integration data is

available, the linearity is good and there is no hysteresis. The main drawback is its size and weight.

Unfortunately, FORS requires a moderate amount of effort in order to repackage the electronics. Ideally, CMOS circuits would be made, which would drastically reduce the size of the assembly down from the predicted 3 x 3 x 3 inch package (which is relatively large for the rover). However, the final size and weight can only be predicted at this point. It should also be mentioned that Draper Laboratory is developing a tuning fork gyro similar to the QRS-11, although performance data is not available at this time.

The Litton gyro has excellent performance in a small size, but it is unclear at this point how much work would be involved in building the supporting electronics. If the electronics can be made small and low-power, there is little question that this system would exceed the performance of most of the others, at about the same cost as most mil-spec gyros.

The Northrop MAG appears to be the best alternative to the Systron Donner gyro, although it is important to realize that it is not available yet. Northrop has a long history of inertial products, so the MAG is promising for the near future. Although power consumption and performance is not available yet, the product is designed to compete with the Systron Donner gyro, and so it is expected to have similar or better performance. Although the MAG will lack the time on the market, it has the advantage of acting as two sensors in one: a gyro and accelerometer.

In any case, the Systron Donner gyro at this point is the leading choice for a dynamic heading sensor on the microrover. Some of the qualities of the Systron Donner rate Gyro are:

- It is purely DC in - DC out, so easily interfaced with the rover electronics
- Solid state construction, with a specified 10 year MTBF
- Demonstrated in high-volume military application
- Relatively low cost
- Integration data is available

The main drawbacks to this gyro is that it has been on the market only a few years (although this is considered a long time for a coriolis-based gyro). However, it is being

used in large quantities on some important military hardware. As with all solid state gyros, there may be problems with temperature and vibration sensitivity, but these can be somewhat corrected for in software if measurements of these quantities are available. In any case, the QRS-11 is, at this point, the clear gyro choice for use on the microrover.

3.2.4. Heading Calibration

The gyro will require periodic calibration, which can be from either a:

- Gyrocompass
- Star Sensor
- Sun Sensor

The frequency of calibration depends on the quality of the gyro, and the accuracy of the update. The purpose of calibration, or zero update (ZUP), is two-fold:

1. To update the integrated angle from the actual angle as measured from the calibration source, which is assumed to be more accurate.
2. To measure the gyro drift at a known rate input of zero (actually the planet's spin rate), so that corrections can be made in hardware or software for future readings, until the next update.

The following describes some of the selection issues of the three calibration sources.

Gyrocompass

Although potentially very accurate in finding true north, a gyrocompass requires at least one minute to make a measurement, using small gyros. This would be acceptable if an accurate heading measurement was not required during travel, but this is not the case. In fact, it is desirable to have a relatively fast reading, so that the calibration source can be used in real time, during travel, in the event of a gyro failure. Although a gyrocompass is not dependent on visibility, it is dependent on the rover stopping for relatively long periods of time, which is not acceptable during its short, 100 m daily run.

In addition, although Mars spins at a rate that is very close to earth's, a gyrocompass could not be used for Lunar navigation, since the moon's spins at only about once per month.

Hence, despite the gyrocompass' advantage in finding true north, its slow speed and dependency on the planet's spin rate make it unsuitable for use on the microrover. Besides, at least two high quality gyros are required, which can result in a large package.

Star Sensor

The available star sensors today are much too large for use on the rover. Star sensors are not only bulky, but are delicate and sensitive, since they must detect such low light levels. The dusty environment of Mars could therefore easily degrade its operation.

In addition, a star sensor would require night time operation on Mars, due to the large amount of atmospheric scattering, and would thus fall short of real time operation in the same way of the gyrocompass. Although a star sensor can provide a very accurate calibration, and does not have a singularity problem at the poles or at high elevations (like the gyrocompass and sun sensor, respectively), a sensor that can supply real time heading data is much more beneficial.

Sun Sensor

Like a gyrocompass and a star sensor, a sun sensor can provide an accurate heading update. Accuracy can be near that of a star sensor, despite the sun's larger apparent size. In addition, it can also supply this information almost continuously during travel. This allows a sun sensor to be used not only for calibration of the gyro, but also in place of the gyro in the event of a failure.

This redundancy does not come at large cost either, since it is possible to make a sun sensor extremely small, yet still sufficiently accurate. This is helped by the fact that the sun is a very strong source, even as far away as Mars (minimum irradiance of 36% of terrestrial values). The high irradiance is important also due to the anticipated dust problems on Mars, which can substantially attenuate the sunlight.

In addition to dust, the sun sensor can also suffer from problems of shading, depending on the local terrain. If the terrain is such that shading can be a problem, then commands

should take this into account, so that the path and time-of-day will minimize the chance of shading. However, if a long (>30 seconds) period of shading is detected by the rover, it may be appropriate to retrace its path, or use other methods to identify and move towards an illuminated area.

Another minor problem with a sun sensor is that its ability to provide heading degrades as the sun elevation increases, with an extreme case being a singularity at an elevation of 90° (directly overhead). The problem is similar to that of the gyrocompass when in the polar regions of a planet. However, the singularity case can be shown to be rare, even when the rover is located near the equator (section 4.1).

As a calibration source, a sun sensor should provide an accurate update within a few seconds, so that the rover can minimize the time required for a daily run. When used as the only heading sensor during travel, inclinometer data needs to be used, which will result in time constants of at least a few tenths of a second. Or, if a passive leveling device is used instead of inclinometers, this will also result in a slower response. Therefore, the sun sensor will likely have similar dynamics of a magnetometer, which should be accounted for in the control software to minimize position errors from the time lag [2]. Nevertheless, a sun sensor has many advantages that make it suitable for heading calibration.

Choice of Sun Sensor

There is one dedicated manufacturer of sun sensors for spacecraft. Adcole makes a wide variety of sensors, mainly for satellites, which are designed to operate in the vacuum of space. The sensors offered by Adcole are much too costly and bulky, although a few are relatively small. The small devices with the large field of view, however, are irradiance-dependent, and are thus subject to errors from the dusty environment. The microrover's unique requirement of a hemispherical field of view is the biggest limitation. However, in order to have a mechanically passive sensor, this must be the case.

Hence, a few concepts for sun sensors were discussed, of which one was constructed and demonstrated on MITy-2. As shown in section 4.1, this sensor was constructed out of a small fish-eye lens, and a two-dimensional PSD, along with supporting electronics. This assembly is not passively leveled, and thus requires two single-axis inclinometers to provide tilt information.

It should be noted that this sun sensor has the ability to measure two axes: azimuth and elevation. Although only heading (azimuth) is required, the second axis is available as an inherent part of the design, and is required in this case since the sensor is not leveled. With this additional data, azimuth can be determined for a given measurement by measuring the relative heading and elevation (with respect to the sensor), and correcting for these by using the tilt angle (rotation of the sensor with respect to the vertical).

3.2.5. Summary of Hardware Choice

The navigation hardware chosen for the microrover utilizes separate dead reckoning sensors for longitudinal translation, and heading. If a vision system becomes available in the future, some modifications of the following hardware will be necessary. But for now, position updates will not be available from an external source, and so the accuracies of these sensors must be such that daily position errors are within 10 % of the travel. Calibration of the heading sensor will be possible, though, so that angular errors will be bounded. Because of this, it is expected that the majority of the position errors will be from the translation sensors.

The sensors chosen for microrover navigation are:

Translation
- Drive wheel motor odometry
- Unpowered drag wheel odometry
- Inclinometers

Heading
- Rate Gyro (integrated for angular displacement)
- Sun Angle Sensor
- Inclinometers

The same inclinometers can be used for each application. Details of the navigation sensors, as well as their integration into the rover, are discussed for the prototype rovers in Chapter 4, and for the final design in Chapter 8.

CHAPTER FOUR
PROTOTYPE NAVIGATION HARDWARE

4.1. PROTOTYPE SUN SENSOR

4.1.1. Requirements

In order to quickly construct a sun sensor for the MITy-2 prototype microrover, some of the components were chosen for their availability. However, the goal was still to satisfy the system needs. These are:

1. Provide heading angle to within about 2° accuracy
2. Provide an approximately hemispherical field of view (2π steradians), without mechanically scanning
3. Provide accurate heading data fast enough to use while the rover is moving

As mentioned in section 3.2.4, the concept chosen was to use a fisheye lens, pointed upwards. By focusing the collected light to a point, the sun position can be found by measuring the position of the point in the focal plane relative to the optical axis.

4.1.2. Choice of Detector

To sense the position of the spot, a two-dimensional position sensor is required. There are a variety of choices for position sensing, all of which can be categorized as discrete or continuous. The most common detectors of these types are the charged coupled device (CCD) and the position sensitive detector (PSD). The last decade has allowed both types to mature on the market, and they are now available in miniaturized packages, and at a low cost.

The CCD is a discrete element sensor, and so it has the advantage of scanning for the intensity (and color, if desired) of each individual pixel. The resolution is excellent for their size, and are thus commonly used for imaging, such as in video cameras. However, accuracy and resolution are somewhat limited by their discrete nature. But for most position applications this is not a problem, since today's CCD can easily have over a million pixels. In addition, scanning can be very fast, up to several hundred megapixels

per second, and can be controlled by the user. CCDs also have a high quantum efficiency, making them ideal for low light level applications, such as part of a star sensor. However, this is not as important with the high intensities of the sun. The main advantage of the CCD is the ability to scan, so that the actual profile, or distribution of intensity can be measured. This is particularly important in the presence of other light sources, such as atmospheric scattering, and reflections off objects. In these cases, scanning the CCD can isolate the peak, so that other light sources than the direct sunlight will not affect the accuracy.

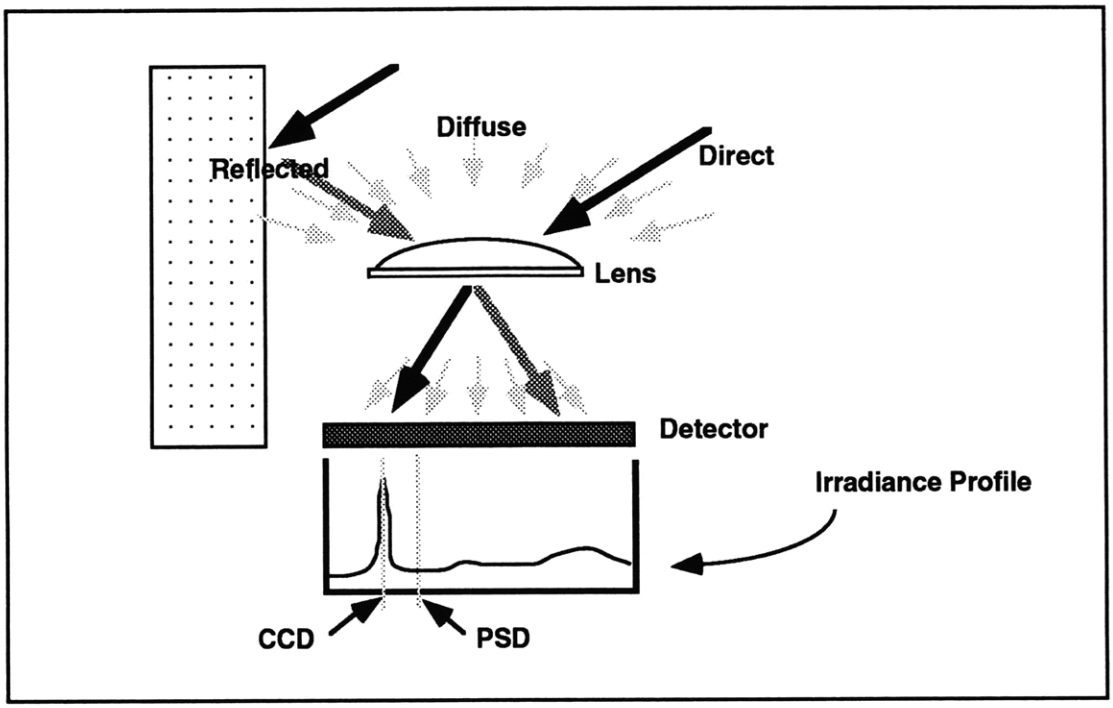

Figure 4.1: Irradiance profile across a detector. Contributions from indirect sunlight can yield an erroneous sun angle measurement with a PSD, while a CCD can isolate the peak.

Continuous position sensing is possible with a PSD. Although a PSD does not have the advantage of providing the intensity *distribution*, it does offer higher speed and simpler electronics. The theory of operatation of a PSD is explained in section 7.1.5. Continuous operation is performed simply by reading the four anode currents from the sensor. The main drawback to a PSD is the inability to distinguish the direct sunlight from other sources, due to the inability to scan. All of the light striking the PSD contributes to the anode currents; and from these currents, a light spot location is calculated, based on the assumption of a single, small incident light spot. Therefore, additional light sources will

degrade the accuracy of the direct sunlight measurements, unless the sum of these results in an equivalent 'centroid' that coincides with that of the direct sunlight. However, the effect of other sources is usually small on a clear day, so that a PSD should suffice for the prototype, anyway. Also, some degree of optical filtering can be used to help reject secondary sources of known spectral distribution, such as a blue sky.

Therefore, a PSD was chosen for the prototype sun sensor, mainly for the simpler electronics and output signal. Since PSDs were also being ordered for the laser rangefinder, a two-dimensional sensor was also ordered from the same manufacturer (UDT Sensors, Inc). These sensors have very good linearity and accuracy, and are relatively low-cost. The choice was mainly on the basis of size, since it was found that with a 36 mm focal length lens attached to the fisheye lens, the area encompassed by the focal point was bounded by a 9 mm diameter circle. Therefore, a 10 mm active surface was chosen. The specifications for this PSD are shown in table 4.1.

Model	DL 10
Sensing Area	10 x 10 mm
Position Detection Error [1]	20 μm
Responsivity [2]	0.2 A/W @ 550 nm
Dark Current [3]	200 nA
Rise Time [3]	2 μs

1 Typical value. Maximum of 50 μm
2 Spectral responsivity curve shown in Figure 4.2
3 Typical value, at a bias voltage of 15 V

Table 4.1: UDT Position Sensitive Detector Specifications

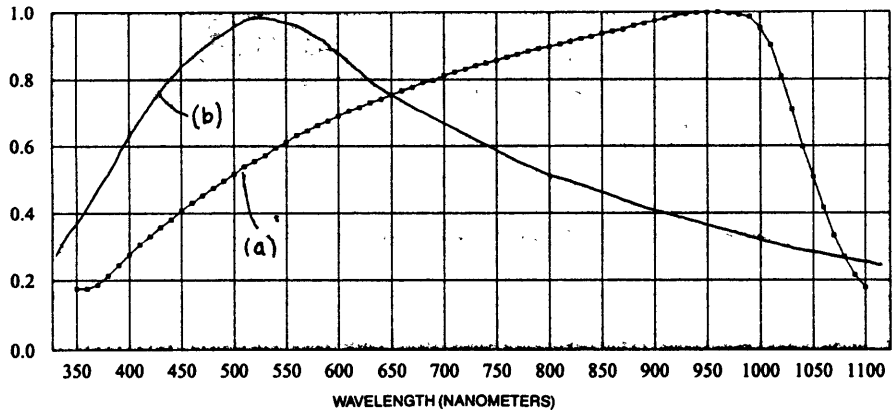

Figure 4.2: (a) PSD spectral response, (b) solar spectral emission (normalized)

Note that the responsivity of the detector peaks near 900 nm, so the detector is less sensitive in the visible part of the spectrum. This is acceptable, however, since sunlight has a high irradiance.

In choosing the fisheye lens for the prototype sun sensor, size and cost were the main issues. A miniature fisheye lens was sought after, and the choice was actually a door peephole, bought from a local hardware store [1]. This was slightly modified to reduce internal reflections. As mentioned, a plano-convex lens of 36 mm focal length was attached to the end of the fisheye to focus the light to a small spot on the PSD. The entire optics head was very small, with a price tag of about $25.

Other than the PSD electronics, the sun sensor consists of three components:

1. Fisheye lens
2. Focusing lens
3. Position sensitive detector

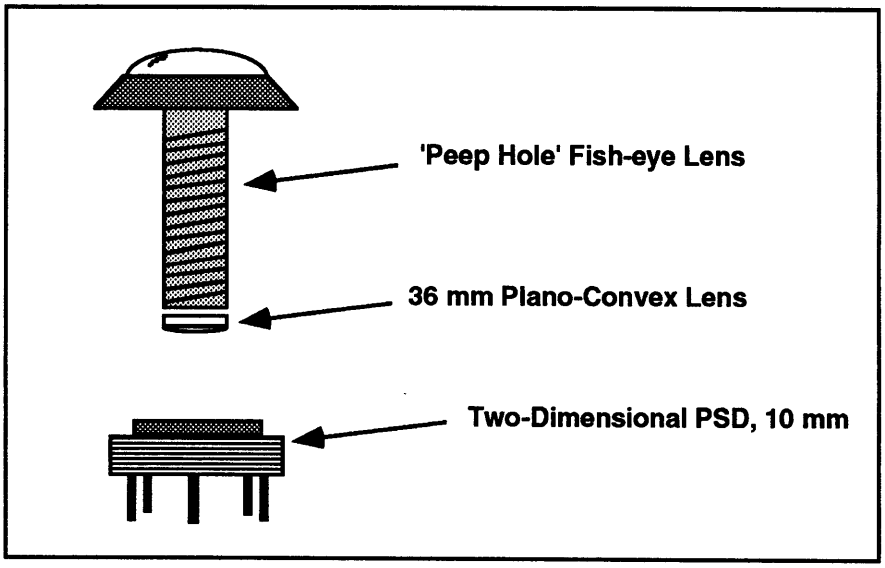

Figure 4.3: Prototype sun sensor

4.1.3. Calculating Heading from the Sensor Data

The traverse of the focused spot on the PSD is a function of the relative elevation and azimuth of the sun. First assuming a level case, with the optical axis vertical, the relative elevation and azimuth is then equal to the true elevation and azimuth.

For a given azimuth β, as the elevation ε changes from 0° to 90°, the focus spot traverses radially towards the center of the PSD, which is located along the optical axis. The two spherical coordinates (β,ε) are transformed into a two-dimensional spot postion described conveniently in polar coordinates (r,θ), and measured in cartesian coordinates (x,y). In the level case, then, ε is transformed into r, and β into θ. This is shown in Figure 4.4.

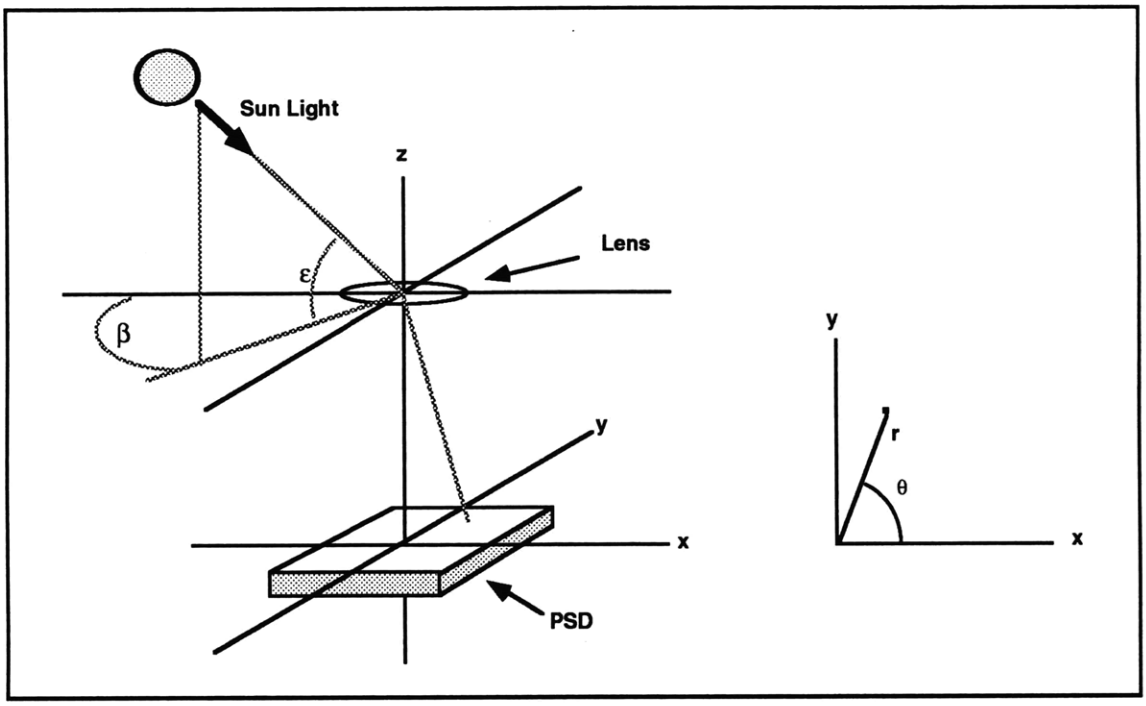

Figure 4.4: Azimuth and elevation from PSD measurements. The two spherical angles are transformed to a polar coordinate representation on the PSD, which are directly measured as x and y.

The transformation of β to θ is direct. However, the transformation of ε to r is a function of the optics. In this case, it is a nonlinear function, as shown in figure 4.5. Note that the curve is linear at high elevations (near center of PSD), and is thus most sensitive there.

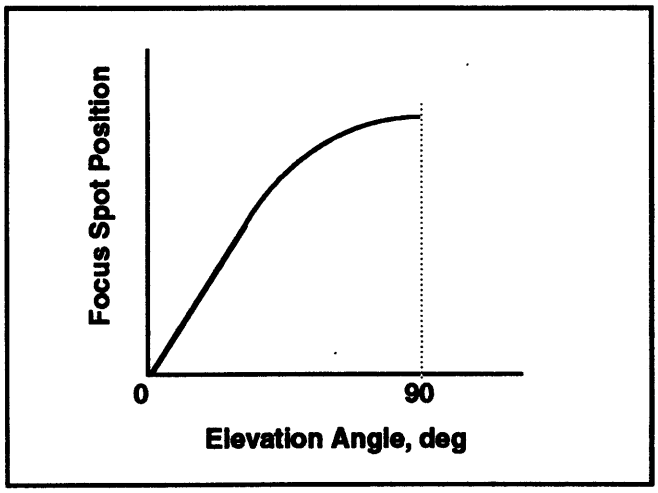

Figure 4.5: Relationship between focus spot displacement and elevation angle.

The fact that the elevation transformation is nonlinear is not important for the level case, since azimuth is what is needed. For any given elevation angle, the azimuth can be determined by simply calculating θ. Since the PSD yields x and y values, θ is calculated from:

$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \tag{4.1}$$

This is the heading of the sun relative to the rover's longitudinal axis. By knowing the time of day, geographical location, and time of year, the absolute heading of the sun can be calculated, so that the rover's absolute heading can also be determined.

4.1.4. Heading Errors

For a perfectly accurate sensor, the heading accuracy would be nearly independent of the elevation. In reality, the detector has position sensing errors, so that when r is small the resultant heading angle error becomes large. This occurs at high elevations, with the extreme case at 90°. In this case, the sun sensor fails to provide heading. However, for this situation to occur, the following conditions must apply simultaneously:

1. Latitude is near or lower than the planet's declination angle (d=24.8° for Mars).
2. Orbital phase is near the equinoxes
3. Time of day is near solar noon

Therefore, even when the geographical location and the time of year is supportive of high elevation angles, only a fraction of the day will make this achievable. In short, the morning and afternoon always provides the opportunity to use the sun as a source for heading.

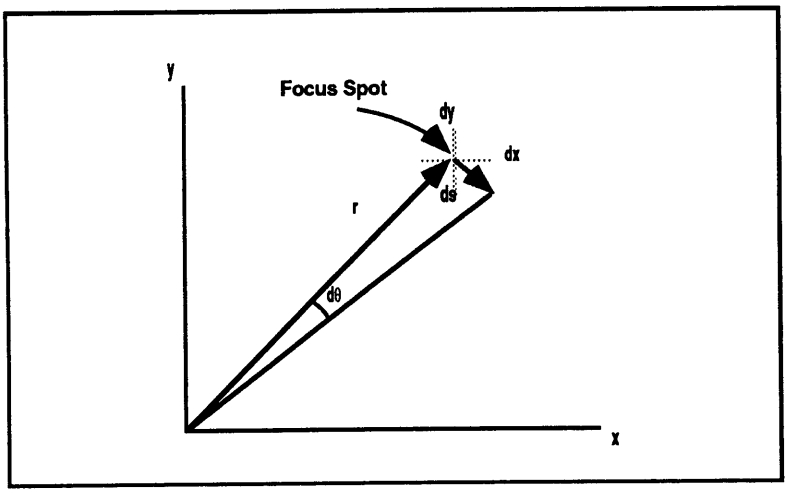

Figure 4.6: Heading error due to the PSD measurement error, ds.

As mentioned, the detector has associated measurement errors. For a given r, and position detection errors dx and dy, the tangential position error ds is largest when |x| = |y|, so that θ is 45° from an axis. In this case:

$$ds = \sqrt{(dx)^2 + (dy)^2} \tag{4.2}$$

Assuming dx = dy, the tangential position error is:

$$ds = \sqrt{2}\, dx = \sqrt{2}\, dy \tag{4.3}$$

When this error is small relative to r, the resulting heading error is:

$$d\theta = \frac{ds}{r} \text{ (radians)} \tag{4.4}$$

Heading error can then be found using the PSD maximum error of dx = dy = 0.05 mm. In this case the maximum tangential error is 0.07 mm. For this position error, the resultant heading error exceeds 2° when r < 2 mm.

This error is from only one source: position sensing. Although the error calculation was conservative in that the maximum detector error was used, there will be many other sources that can potentially degrade this accuracy further. These include:

- alignment errors
- electronic noise
- degraded optics

The electronic noise lowers the signal-to-noise ratio (S/N), which results in the need for a stronger signal. Fortunately, when the geometry is such that the position sensing error is large (high elevation), the S/N is also large, due to the low attenutation of the fisheye optics. At low angles of incidence, there are a number of cosine losses that greatly attenuate the sunlight. However, the geometrical errors are improved at these angles.

4.1.5. The Tilted Sun Sensor

When the sun sensor is tilted, corrections have to be made based on the direction of tilt. First, in order to measure this, two inclinometers are required. Of course, these inclinometers are also required for position corrections when travelling over inclined surfaces, and thus serve a dual role.

For the prototype sun sensor, two inclinometers were purchased (see section 4.2.3). These were mounted along the pitch and roll axes of the rover, directly to the sun sensor body.

When the rover is tilted, the azimuth and elevation measurements are relative to the sensor. In order to obtain azimuth in the fixed frame, a coordinate transformation is necessary, based on the measured rotations of pitch and roll. Two coordinate frames need to be defined: the body and the fixed frame. The body frame, denoted by the subscript b, is centered along the optical axis of the sun sensor. The fixed frame, denoted by the subscript f, is the same as the body frame when the pitch and roll are each zero. Therefore, the fixed frame is partially defined by the local gravity gradient, which is the vertical.

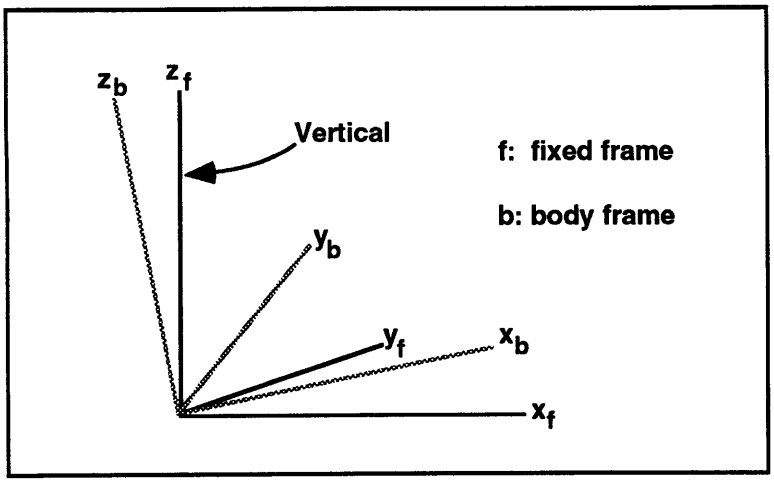

Figure 4.7: The fixed and body coordinate frames. The body frame is alligned with the sun sensor optical axis, while the fixed frame is alligned with the vertical.

For a given pitch, θ, and roll, ϕ, the coordinate transformation involves use of the Euler angles. The process of calculating heading, β_f, involves the following steps:

1. The PSD spot position (x,y) is measured, along with the two inclination angles
2. The position of the sun relative to the sensor is calculated (body frame)
3. The body frame angles are transformed to the fixed frame angles, yielding heading

4.1.6. Position Sensing Electronics

For the sun sensor, a standard two-dimensional PSD circuit was used to detect the position of the focus spot in figure 4.8. For each of the two axes, there are 4 stages:

1. Each anode current is transformed to a voltage using a transimpedence amplifier, with the necessary gain.
2. The two signals are differenced.
3. The same two signals are summed.
4. Both the summed and differenced values are read by the rover's microprocessor and processed for PSD position calculations, which is discussed in section 7.1.5.

This circuit does not bias the PSD, so that the listed dark currents in table 4.1 do not apply. Also, position detection requires a division at some point, which is done in software in this case. Division may also be performed in the electronics, with an analog

chip. Although the latter method is more expensive and electrically complex, it saves a step in the software, and automatically normalizes the signals being read, which improves the accuracy in many cases. Because of the circuit's analog nature, heading calculations are limited by the A/D and the processing, and sometimes the inclinometers. Nevertheless, heading information can be provided in real time. This circuit was fabricated and debugged by Lorusso [26].

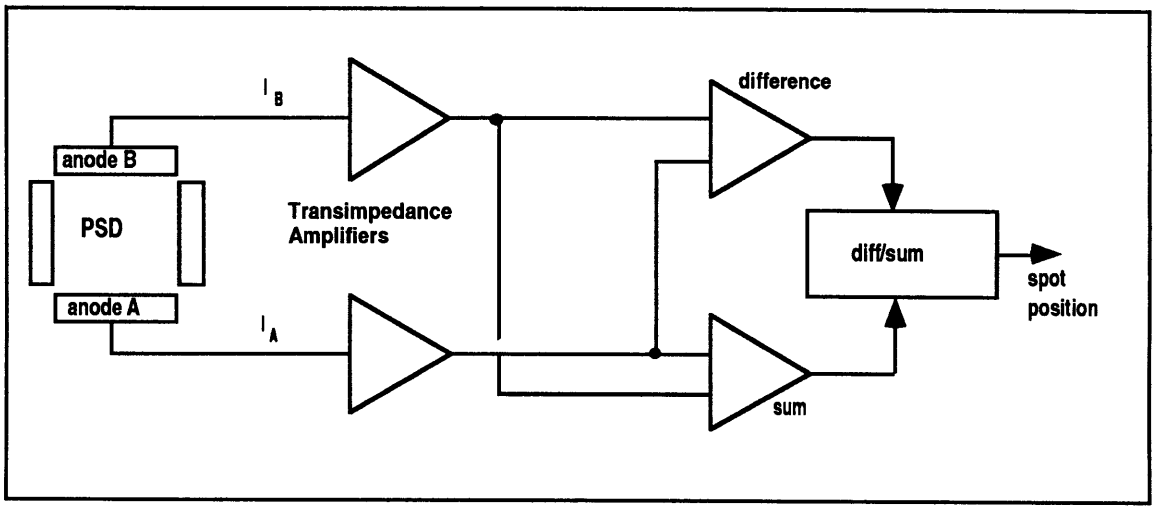

Figure 4.8: PSD position detection electronics.

4.1.7. The Complete Sun Sensor Design

To mount the lens assembly relative to the PSD, a supporting block was made out of aluminum, with the lens mounted at the top. A housing was made for the PSD, so that adjustments in the x and y directions could be made. This was done through four adjustment screws. An exploded view of the sun sensor is shown in Figure 4.9.

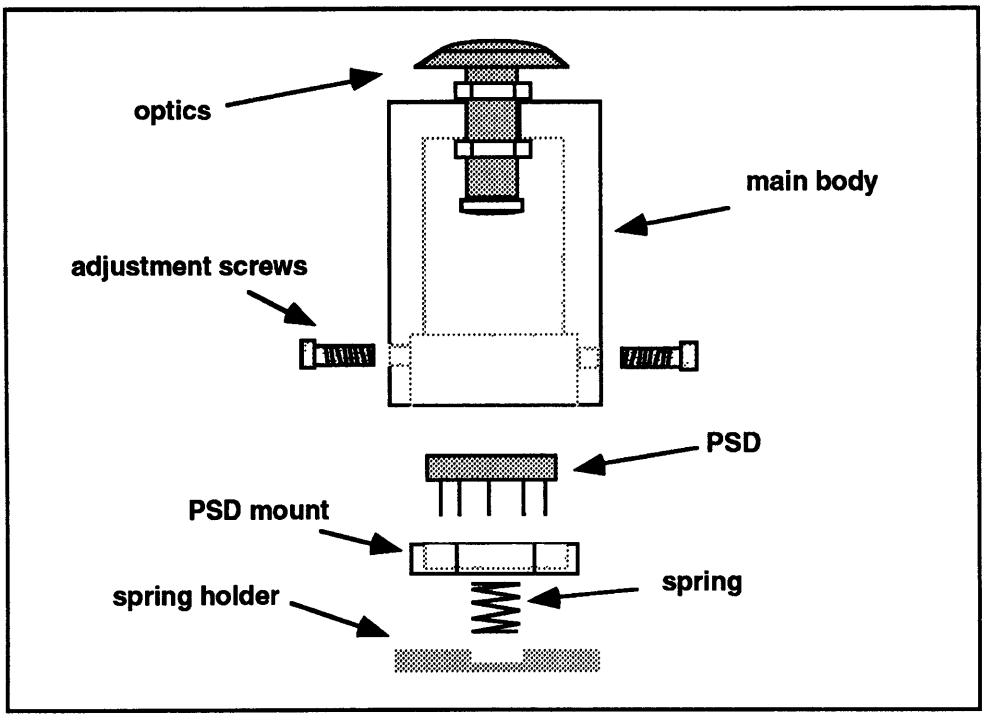

Figure 4.9: Prototype sun sensor for MITy-2: exploded view.

To minimize mounting errors, the sun sensor formed the hub of the 'heading block', which holds the two inclinometers, one accelerometer (also for measuring tilt), and the angular rate sensor. This heading block in turn mounts to a platform that attaches to the top of the rover's center body. The associated electronics also mount to this platform, allowing the entire package to be removed from the rover. This unit is shown in Figure 4.16.

It is desirable to have the PSD leads as short as possible, since they carry very small currents. When the wires are long, additional currents (noise) can be induced from electromagnetic disturbances, mainly from the rover itself. Although the sun sensor circuitry is currently within a few centimeters of the PSD, future designs should locate the transimpedance amplifiers within millimeters if possible.

4.1.8. Preliminary Sun Sensor Test Results

After the PSD was centered with the adjustment screws, the sun sensor was ready for testing. Gains were adjusted so that the amplifiers and the microprocessor A/D did not saturate.

Testing of the sun sensor involved placing the rover on a rotation table, and rotating in azimuth. During the testing interval, it was assumed that the sun was at a fixed position, although it moved about 0.1°.

The sampling of the data does utilize some filtering, due to the randomness of the A/D, which fluctuates about 0.50 %. This was performed with the sun sensor levelled. Although the absolute accuracy was not defined, the relative accuracy was measured to be 1° after filtering. The preliminary test was performed in an artificially-low noise environment, which suggests that future on-the-fly accuracy will be lower. However, accuracy is expected to improve with better optics and detection circuitry, as well as better filtering.

4.2. PROTOTYPE MICROROVERS

4.2.1. Introduction

Two microrover prototypes were designed and built in order to characterize performance, demonstrate their abilitites, and provide feedback for future design iterations. The hardware work, although time consuming, did indeed serve as valuable aids for the design process. Although iterations of entire systems result in slow progress, the end result is a better understanding of what parameters are important. Proof-of-concept methodology also provides more credibility to any suggested design.

The following sections describe the navigation hardware for MITy-1 and MITy-2, which are, respectively, the first and second prototype microrovers.

4.2.2. MITy-1 Navigation Sensors

As the first prototype, MITy-1 was designed to obtain feedback from the environment for the purpose of navigation and hazard avoidance. Although these sensors are not representative of MITy-2 or the final proposed design, they have demonstrated their effectiveness in providing navigation data through a variety of successful runs.

MITy-1 has a very limited processing ability, with most of its dedication applied towards the microrover control code [1]. Since data processing is limited, the amount of sensors also is. Hence, navigation sensors comprised of only:

1. Drag wheel
2. Flux-gate magnetometer

This is shown in Figure 4.10.

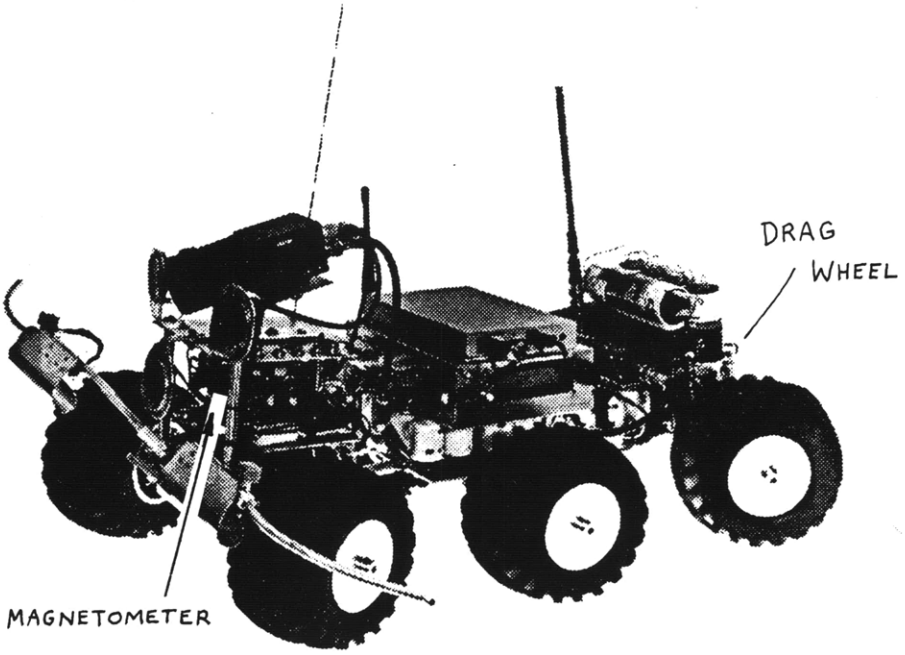

Figure 4.10: MITy-1 prototype micro-rover. Navigation sensors are labeled.

Drag Wheel

The drag wheel was suspended on a trailing arm at the rear. Due to the instability of this configuration, data was used only when travelling forward. Rotation was measured with an infrared optical encoder. In addition to odometry, the drag wheel was also used to provide heading information during short periods when the magnetometer could not function well. This was accomplished by allowing the trailing arm to also pivot in yaw, and measuring its angular displacement with a potentiometer. The deflection of the straight line position was approximately a linear function of the turning rate. For relatively flat surfaces, this was sufficient for short periods, but certainly not for durations of more than 30 seconds or so. Although somewhat crude, the drag wheel provided good distance measurements on uniform surfaces.

Magnetometer

The main method of finding heading on MITy-1 is with the onboard magnetometer, made by KVH Industries. Earth provides a strong enough magnetic field to make these sensors invaluable in providing low-bandwidth heading data. Since measurements are referenced to an external source, there is no accumulated heading error (unlike the drag wheel). However, the time constant is about 0.5 seconds, making it slow. When the rover is in a full turn, the slow behavior results in significant position errors (reference Calvin). Nevertheless, the magnetometer was very successful in providing accurate heading data, as long as runs were made outdoors, where magnetic disturbances were minimized [15].

Test Results

A moderate amount of testing was performed with MITy-1 during the summer of 1992. The control code for MITy-1 (and MITy-2), which used the sensor data for autonomous control, was developed by Malafeew. A successful autonomous demonstration with this simple sensor arrangement was performed at a rover exposition in Washington DC, September 1992. This, however, was on a very smooth surface. In fact, all of MITy-1 testing was performed on a smooth surface, so a direct application of the test data to a real unstructured environment is not suggested. Nevertheless, the results of MITy-1 testing provided important feedback for the design of MITy-2. Simulation and experimental test results are available in reference [1,2].

Summary

MITy-1 could successfully demonstrate autonomous navigation on flat terrain, using data from only a magnetometer and a drag wheel. However, this was possible under very limited conditions. The goal of future prototypes and the final design is to use sensors that can be used in space applications, which includes the ability to autonomously navigate in unstructured environments.

4.2.3. MITy-2 Navigation Sensors

The first prototype, MITy-1 was designed to get feedback from the environment and react to it. The important point was that the behavior was 'closed-loop', even though the sensors were not of the types that could be used on a planetary mission. MITy-2, on the other hand, was designed so that its autonomous sensors were of the type that could indeed be used on the Lunar or Martian surface.

As discussed in 3.2, the suggested navigation hardware for the final design is:

- Drive wheel and drag wheel odometry
- Sun sensor
- Angular rate sensor
- Inclinometers

The actual hardware chosen, or designed, for MITy-2 was based on performance, cost, and ability to implement quickly. Again, none of these sensors are space qualified, but other hardware of the same type are, or can be, space qualified, and their principles of operation will still apply.

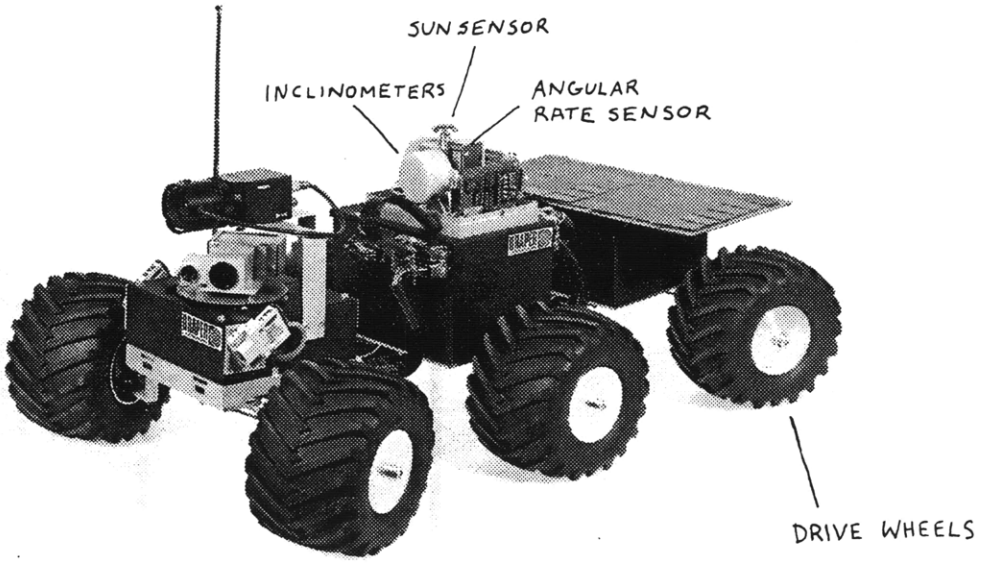

Figure 4.11: MITy-2 prototype micro-rover. Navigation sensors are labeled.

Odometry

There are six DC motors for mobility, located within each wheel hub. Each of these motors has the same gear ratio, so that motor speed is proportional to the wheel speed. On the back of each motor is a tachometer, which outputs a voltage that is proportional to the motor speed. Currently these are used for speed control, but they are also used to estimate the distance traveled, since the drag wheel has not been installed at this time.

In the current state, two tachometer signals from the middle wheels are averaged, and this combined signal has been calibrated with vehicle speed on relatively flat and smooth surfaces. The main assumption in this method is that the wheel speed is directly proportional to the vehicle longitudinal speed, so that no slipping is occurring. Therefore, this method is not expected to be accurate on soft soil, which is why a drag wheel is proposed. The drag wheel can be used along with the tachometer readings, and testing and simulation results will help to guide the way for the best sensor fusion algorithms. Hence, it may be necessary to use all six tachometers, possibly even with inclinometer data. The important point is that the all six tachometers are available for use, although data processing may not be able to accommodate this.

Proposed Drag Wheel Design

The drag wheel, as mentioned, requires space to flex up and down, so that it maintains contact with the ground, but does not interfere with mobility. Since the rover will travel in reverse at times, the drag wheel cannot use a trailing arm suspension. Vertical travel needs to be about the same as the rover's ground clearance, which is about 6 cm on MITy-2.

To accurately measure longitudinal travel when turning, the drag wheel should be placed along the longitudinal axis of the rover. If not, the wheel will spin too fast or slow, depending on the direction of turn, resulting in an asymmetrical output. Also, the drag wheel should be placed longitudinally near the center platform, so that the lateral translation is minimized during turns.

Unfortunately, the very center of the rover is crowded with components as it is. Thus, it would not be appropriate to raise them up to make vertical space, since this would raise the cg. on the already-tall center platform. The next closest locations are the two areas in between the three platforms. These have about 10 cm of separation, and are connected by two spring steel wires.

It is suggested that a drag wheel be placed in between the 2nd and 3rd platform. Although other openings provides an equal option, there is the possibility of placing a second wheel there for the purpose of smoothing out soft terrain, so that the drag wheel will encounter a smoother surface. This idea remains to be tested.

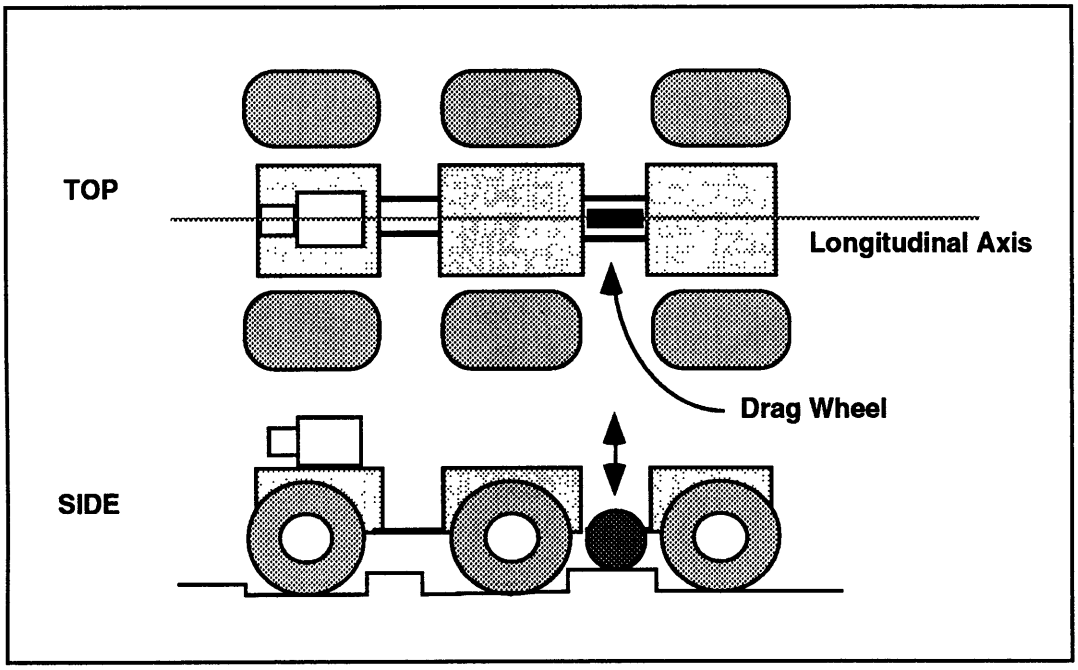

Figure 4.12: Drag wheel on MITy-2.

For accuracy over irregular terrain, a large wheel is desired. However, the wheel diameter is limited by the platform spacing. The drag wheel ideally should be able to extend below the drive wheel. However, for the given small diameter, this could result in some mobility interference if an object was suddenly encountered. Therefore, the drag wheel should extend at least down to the level of the drive wheels, and should be able to detect when the lowest position is reached. In the case that the drag wheel is fully extended, it may have lost contact with the ground, so that the drive wheels should be used for odometry during this brief period.

The drag wheel needs some kind of suspension that will keep the travel nearly vertical as the rover moves forward and backward. However, this is rather difficult to do without joints. A passive, flexing suspension would be desirable to prevent the frequent sticking problem in space, but a satisfactory design has not yet been achieved. Hence, for now, sliding surfaces will be considered acceptable.

A possible design is shown in figure 4.13. The strut will have to take some bending load, but the vertical spring load should be low, so that the wheel can easily retract upwards when encountered by most obstacles. It is desirable to have a nearly constant vertical force on the ground, which should be small to prevent sinking, yet large enough to ensure

that the wheel does not slip. This force can be made constant if the vertical travel (jounce and rebound) is small compared to the spring length.

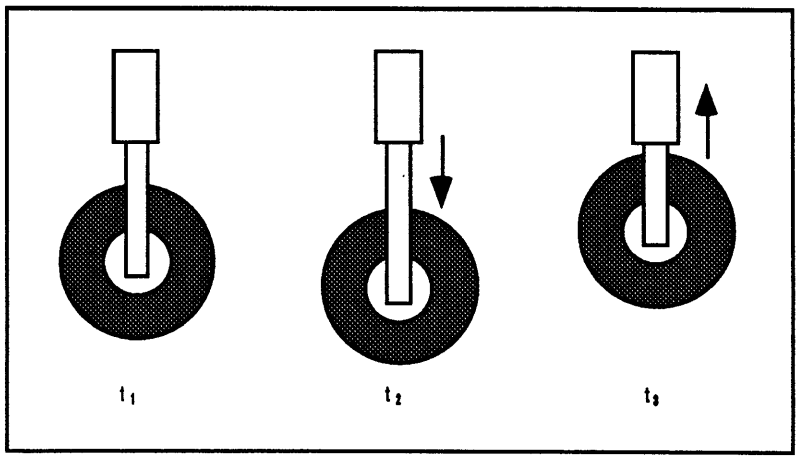

Figure 4.13: Vertical motion of the drag wheel over irregular terrain at different times (illustrative example).

In order have higher odometer accuracy, information about the vertical motion of the drag wheel can be useful. Note that if the vertical motion is zero, then it can be implied that the rover's drive wheels are traveling at the same local ground height as the drag wheel. If not, it means that there is different terrain at the drive and drag wheel. In this case, then, corrections should be applied to the drag wheel readings, since it will measure a surface distance that is larger than what the rover is traveling. The same can be said about the drive wheels over rough terrain too, but the changes in the inclinometer readings can be used for corrections to the drive wheels. So, it is suggested that a sensor that provides a signal proportional to the vertical *velocity* be used to correct the drag wheel data.

The drag wheel itself should have a low rolling resistance to minimize slip on the soft soil. The design should not be based on those for indoor mobile robots, since a hard floor ensures no sinkage. For soft soil, the wheel needs to be wide enough to keep the contact patch pressure low. Detailed wheel design, including the tread pattern, is left for future work. However, the ideas just discussed are expected to be implemented on MITy-2 during the summer of '93.

Sun Sensor

The sun sensor design for MITy-2 has been described in section 4.1.

Angular Rate Sensor

As previously discussed, the final design requires a space-qualified gyro, which was chosen to be the Systron Donner QRS-11. For MITy-2, however, the Murata Gyrostar angular rate sensor was chosen, mainly due to its availability and low cost.

This Murata gyro is very light (45 g) and miniaturized. For yaw detection, the sensor is mounted with its long dimension vertical. A five-pin header is provided, with the following terminal output:

Terminal Number	Description
1	+12 V power supply
2	Output, 2.5 ± 2.0 V
3	Reference voltage (2.5 V)
4	Diagnostic output (4.5 V min)
5	Ground

Table 4.2: Pin Out for Murata GyroStar

Other specifications for this gyro are:

Model	ENX-0011
Max Angular Velocity	90 deg/sec
Scale Factor	22.3 mV/deg/sec
Zero Rate Input	2.5 V ± 100 mV
Linearity	± 0.5 % full scale (max)
Hysteresis	none
Bandwidth	7 Hz
Offset drift	200 mv p-p (max)
Zero Point Stability	20 mV/10 minutes (max, at 25 ± 3°C)
Operating Temperature	-20 to +70°C

Table 4.3: Murata GyroStar Specifications

As discussed in section 3.2.3, the Murata gyro has a high temperature sensitivity, which can easily be the main source of drift. In addition, it is sensitive to vibrations along its sensitive axis. The low bandwidth of 7 Hz should not affect the navigation performance, since any frequencies above this should be due to noise. The offset drift is a variance from gyro to gyro, but should be fixed for any particular device. Hence, the listed zero point stability is the main performance indication, of which temperature variation has a large influence.

Rate Integration

To calculate angle, this gyro was connected to an analog integrating circuit. A standard integrating circuit is shown in figure 4.14. There are a variety of integration techniques, most of them digital. Digital integration can have errors if the sampling rate is not high enough, relative to the rate inputs. However, digital integration has the advantage of stability over time. This makes it well-suited to navigation systems, which often requires long term stability. Analog integration, however, has the advantage of simplicity, and does not rely on software to keep track of rate history. However, analog integration can be a problem when the thermal environment is not well-controlled, and thus can lead to significant errors over time. Hence, digital integration is suggested for future prototypes. However, analog integration was chosen for MITy-2, mainly because of the limited processing available, and the familiarity with the analog circuit.

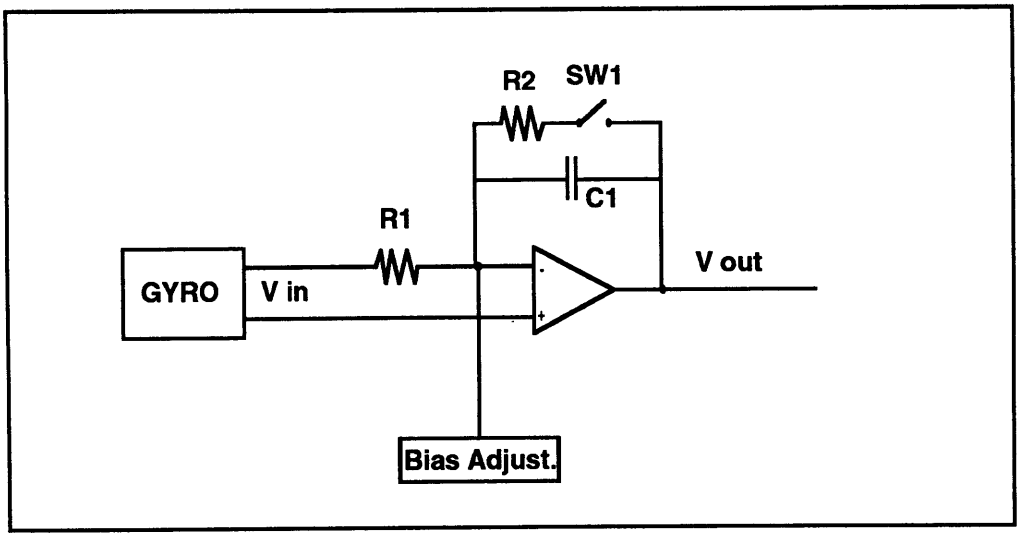

Figure 4.14: Analog integration circuit, with a manual bias adjustment for zero rate nulling.

The output of the circuit is:

$$V_{out} = -\frac{1}{R_1C_1}\int_0^t V_{in}dt$$

Therefore, the integration scaling factor is determined by the R_1C_1 constant. R_2 is not relevant to the integration, but is there to dissipate the energy stored in C_1 upon reset. One of the key factors in choosing R_1C_1 is the frequency of the reset, or discharge, of the capacitor. Higher sensitivity is achieved when R_1C_1 is small, but resets need to occur more often.

The gyro signal was connected to the inverting input of the op-amp, and the 2.5 V reference pin was connected to the noninverting input. By doing this, the differential voltage input to the circuit is zero at zero rate. Since every gyro has a slightly different bias, a bias adjustment circuit was also constructed to null the input to the integration circuit at zero input rate. This circuit is designed to be adjusted by the user once. Of course, changes in the bias may occur, so that the null may not exist for long. An active nulling system may be designed, so that when the rover stops for a ZUP (zero update), a high resolution D/A can control the bias adjustment current. Another simpler method is to stop and measure the drift, and subtract this from future readings (in software), until the next ZUP. The latter method will be used on MITy-2.

For simplicity, R_1C_1 was chosen on the large side, so that resets were not needed until the rover turned about 180°, at which point the capacitor saturated. When this level was approached, a small relay (SW1) closed, and discharged the capacitor in about 10 milliseconds.

Drift In the Integration Circuit

One drawback of analog integration is that current leakage can occur at various points in the circuit. Because of this, the output voltage of the circuit, which represents angular displacement, tends to decrease to zero with time. For these cases, it is important to use a low leakage capacitor (teflon, polycarbonate, etc) and low-leakage op-amp. In addition, distances from the op-amp to R_1 and C_1 should be kept small, and the board should have a very high resistance, and be maintained clean. There are a number of other procedures that help to minimize current leakage. Future designs may utilize digital integration to circumvent this problem. However, the *gyro* drift itself would remain a separate issue, which can be controlled somewhat by stabilizing the temperature. MITy-2 will not use thermal control, so periodic ZUPs will be necessary to compensate the gyro drift, and to obtain absolute heading information from the sun sensor.

Inclinometers

For navigation, inclinometers serve two purposes:

1. Correct the sun sensor readings
2. Correct odometer readings

As discussed earlier, inclinometers that are referenced to the gravity gradient are subject to noise from rover-induced accelerations. However, this is the simplest method, and hence the most commonly used. Filtering acceleration data is needed, so that the DC term (gravity) can be isolated. The result is a slow inclinometer response.

It is proposed that the final design use accelerometers as the means for finding inclination angle, mainly due to their smaller package. For MITy-2, however, actual *inclinometers* were purchased, although an accelerometer was also purchased for a direct comparison.

The inclinometers are manufactured by Lucas Sensing Systems, Inc. These sensors are DC devices that measure a bubble position (similar to a standard level) by measuring the

capacitance across two electrodes. Although they are rather large and heavy, the power consumption is very low and the accuracy is good. Specifications are given in table 4.4.

Range	± 60°
Linear range	± 45°
Resolution	0.001°
Linearity Null to 10°	± 0.1°
10° to 45°	± 1% angle
45° to 60°	± Monotonic
Null Repeatability	0.05°
Cross Axis Error	<1% up to 45° cross axis angle
Time Constant	0.3 second
Frequency Response (-3dB)	0.5 Hz
Voltage Supply	Regulated +5 to +16 VDC
Operating Temperature	-40° to +65°
Temp Coeff of Null	0.008° per °C
Temp Coeff of Scale Factor	0.1% per °C

Table 4.4: Lucas Accustar Inclinometer Specifications

Because of the slow response, these inclinometers can be used for sun sensor corrections when traveling smooth terrain, or when stopped for a ZUP. The inclinometers are mounted directly to the sun sensor, to minimize offset errors (Figure 4.16). Each inclinometer is aligned with an axis on the sun sensor PSD, which has a square active area. Leveling can be performed with two mounting screws.

Accelerometer

As mentioned, the accelerometer is mounted not for the purpose of integrating for position, but to use as an inclinometer. This will effectively measure the *sine* of the inclination angle along its axis, and not the angle itself. Although subject to the same rover-induced acceleration noise as the Lucas inclinometers, the accelerometer has a

much faster response. Testing will allow a direct comparison of the two sensors, since they are mounted to sense the same axis.

The accelerometer purchased for MITy-2 is a single-axis signal-conditioned device. It is not inertial grade, allowing it to be low-cost, and packaged with its electronics in a very small housing. This device is manufactured by IC Sensors, which is a large supplier of silicon accelerometers, pressure sensors, valves and other silicon microstructures. ICS utilizes silicon micromachining to create a piezoresistive operation. The piezoresistors are used as strain gages on tiny beams that support a mass. When the device is accelerated along its axis, the beams deform and their strain is measured, indicating force, or acceleration. A Wheatstone bridge configuration is often used to provide high sensitivity, cancel cross-axis inputs, and minimize the effects of temperature changes. The entire active assembly occupies only about 2 mm length.

A cross section of a typical ICS piezoresistive accelerometer is shown in figure 4.15. The piezoresistive strain gages are located on each of the four beams, and motion of the mass causes the beams to deform. Overforce stops are machined to limit the travel in the case of shock loads, in order to reduce the chance of a beam failure. Because of the spring-mass arrangement, the mass would oscillate significantly if not damped, so the entire cavity is filled with air. By providing only a narrow gap in between the mass and the caps, air is displaced dissipatively, slowing the motion of the mass. This allows damping to be well controlled, usually with a value of $\zeta = 0.7$. Resonant frequencies for piezoresistive accelerometers are between 500 and 5000 Hz.

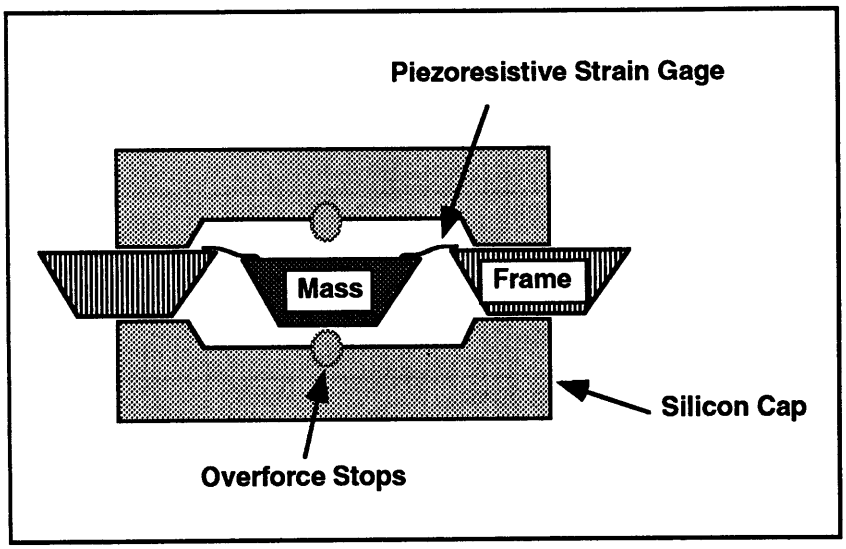

Figure 4.15: Diagram of ICS accelerometer. Small strain gages detect beam deflections, indicating the force along the sensor's axis, hence acceleration.

Unlike piezoelectric accelerometers, which are only AC devices, piezoresistive accelerometers can measure DC accelerations as well, allowing them to measure gravity. The entire enclosed package can be very small, less than one centimeter, with an unamplified output. Accelerometers can also be purchased with internal electronics, resulting in a slightly larger package. This was the type used on MITy-2.

The accelerometer purchased for MITy-2 is an ICS model 3145. It is a DC-DC device, with an amplified output. The specifications are shown in table 4.5.

Range	± 2g
Useful Bandwidth	1000 Hz
Supply Voltage	8 to 30 V
Supply Current	5 mA
Reference Voltage	2.5 V
Output	2.5 ± 2.0 V
Full Scale Output Span	4.0 V
Zero Acceleration Output	2.5 V
Accuracy	± 0.2 (typ), ±1.0 (max) % of span
Operating Temperature	-20 to 80 °C
Temperature Coeff - span	2.0 % of span (typ)
Weight with Cable	13 g

Table 4.5: ICS Model 3145 Accelerometer Specifications

This device can also be purchased with temperature compensation for a slightly higher price, which reduces the temperature sensitivity by a factor of 2. Biases that develop in the null voltage or changes in the scale factor are especially important when using an accelerometer as an inclinometer. This is because angle is sensed by the *magnitude* of the signal relative to the zero-input voltage. Other inclinometers measure angle by sensing the position of the bubble, which will always be in the same place regardless of the magnitude of the DC acceleration. The zero stability is thus particularly important. Because of this, it may be necessary to control or monitor the temperature of the accelerometer for future designs.

The Heading Unit

The entire heading unit is basically the sun sensor core, with 4 additional sensors mounted to the mutually orthogonal surfaces (Figure 4.16). These sensors are:

- inclinometers (2)
- accelerometer
- gyroscope

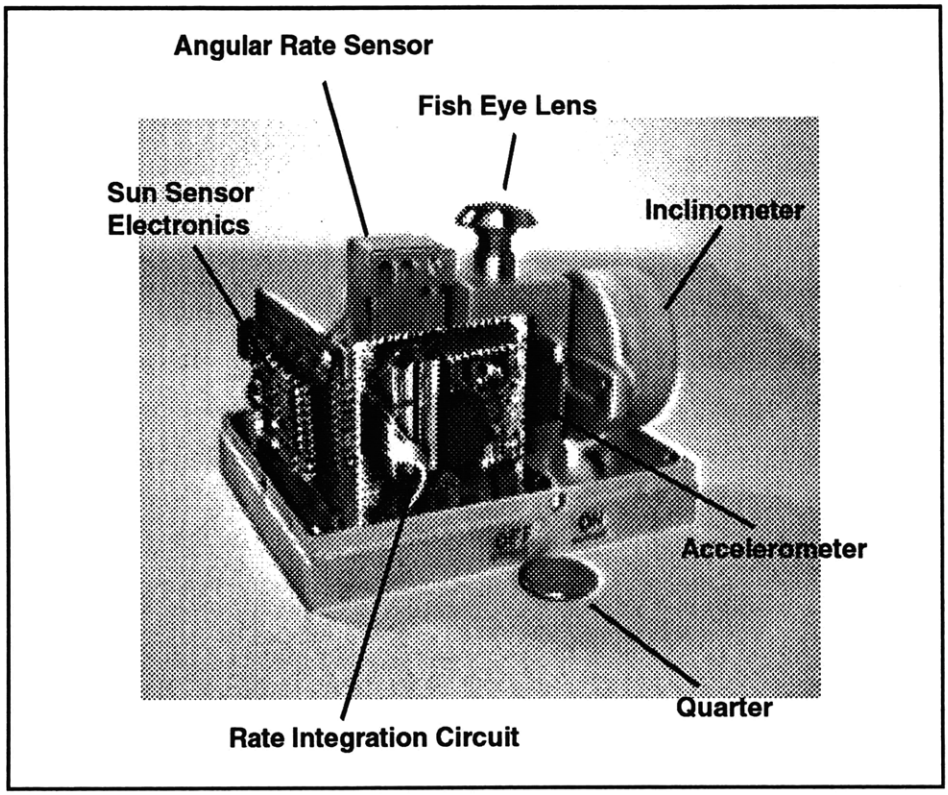

Figure 4.16: Heading block. This module, consisting of various heading sensors, mounts to the center platform on MITy-2.

This sensor block is further mounted to a tray, which currently holds the electronics for the sun sensor and the gyroscope. This heading unit is removable from the vehicle, to allow for external testing. It is desirable to have the sun sensor at one of the highest positions on the rover, to minimize the possibility of shading from other parts of the rover. The same idea applies to the solar panel.

By having all of the sensors located together, it can be assured that they are all subject to the same environmental inputs.

Summary

The data from all of the navigation sensors is used for the purpose of calculating the position of the rover relative to the last position, as determined from the video image. Translation along the longitudinal axis of the rover is measured with the drive and drag wheels, while heading is measured using an angular rate sensor and a sun sensor. Two inclinometers and an accelerometer are used to find the tilt of the rover, so that corrections can be applied to the sun sensor, as well as the gyro and odometer data.

Corrections to the sun sensor are imperative for heading calculations, while corrections to the drag wheel and gyroscope are important mainly at the higher angles (above about 20°), since they are mainly for finding the projections of the measured parameters into the ground plane, which vary as the cosine of the inclination. Sensor data processing is discussed by [1].

At this stage, MITy-2 is in the final stages of construction, and should be available for full testing during the summer of '93. Plans include a semi-autonomous 12-hour demonstration. Testing will be very important for improving the data processing algorithms, especially when the rover is subjected to rough terrain. Chapter 8 discusses future work in this area.

CHAPTER FIVE
HAZARD AVOIDANCE SENSING NEEDS

5.1. HAZARD AVOIDANCE REQUIREMENTS

5.1.1. Introduction

When a man is walking through a forest, sensory data from the eyes, ears, nose, and skin are used to gain information about the environment, so that it can be traversed safely. When a tree is encountered, he does not try to climb it, but walks around it instead. When a small rock is encountered, he will likely step right over it, instead of walking around it. When a hill is encountered, he may or may not climb it, depending on the slope and the surface characteristics. When a cliff is encountered, he must stop and try another route.

A person has the ability to recognize these hazards and avoid them continuously, and *seemingly* without much thought. However, there is an enormous database that is actually being used to make these decisions, along with a large amount of real-time input from the many sensors. It is difficult to contemplate what is truly happening, and is hard to appreciate until efforts are made to mimic the same processes with a mobile robot.

In the case of a microrover, the microprocessor 'brain' (control software) has a limited ability due to the size and power constraints. However, even if a large processor were available, it would fall short of mimicking the human brain in many ways. The main reason is its difficulty or inability to learn significantly from previous experience. Because of this, although a processor can do repeated tasks with high accuracy and efficiency, it lacks the ability to react to a new environmental situation.

Even if a processor was so capable, it is difficult to gather information about the environment with the efficiency of a human, or any animal. A person has the extraordinary ability to continually feel temperature and pressure at nearly every point on the body's surface, hear sounds of sufficient amplitude from almost any direction, and see in a near-hemispherical instantaneous field of view. More importantly, most of this data is ignored unless it is considered important. On a rover, every single sensor typically

needs to be read and processed with the same dedication whether it is important or not, requiring a significant amount of computation. Because of this, the amount of useful data that can be handled is greatly reduced. The result is that sensors present only a simplified, crude representation of the environment. On top of that, this data is filled with noise, making it difficult to accomplish even simple tasks autonomously.

Even so, one of the key goals of the microrover is to sense its environment *sufficiently* so that it can react to certain situations. When traveling towards a goal, just like a person walking through a forest, the microrover will encounter certain situations that can prevent it from reaching that goal, resulting in a mission failure. Hence, the microrover needs to be able to detect these 'hazards', and react to them so that the goal can be reached. This process is called hazard avoidance.

The first step in hazard avoidance is then detection. This is accomplished with sensors, and the proper interpretation. These sensors need to provide a large amount of information from the environment, but cannot require large amounts space and power. It is then important to define the hazard avoidance requirements of the microrover, so that the mission goals can be met efficiently.

5.1.2. Mission Requirements

As stated earlier, the goal of the mission is to carry scientific payloads to interesting locations on the Lunar or Martian terrain that are in the vicinity of the lander. Daily traverses of 100 m are expected, with position accuracy of 10 m (10% of the total distance). The mobile platform must accept daily commands from earth, and travel autonomously towards the target, avoiding hazards on the way. The lander will provide a means for deployment, and a communication system to earth, but will not provide other tools.

5.1.3. Hazard Avoidance Requirements

In order to achieve the mission goal, the hazard avoidance subsystem must sufficiently perform its own part. The rover will be able to know its position to within 10 %, and this is accomplished through the navigation subsystem. This is important, since hazard avoidance requires some degree of navigation, so that the two processes are codependent. As stated, during the course of travel hazards must be detected and avoided. Detection is therefore the first part of the process, and avoidance is the end result. The latter is the

responsibility of the control system, and is mainly algorithm-dependent. Detection is primarily a function of the sensors only, although interpreting the data is dependent on the software. In any case, the sensor hardware and their corresponding algorithms must provide enough sensing of the environment so that the rover control system can use this data to circumvent the hazard and resume travel towards its goal.

Because of the microrover's unique constraints, it will not be feasible to gather as much data as possible. Therefore, sensing is needed just to the point that the rover can sufficiently achieve its goal. It is then important to define the expected hazards, and design a sensing scheme accordingly.

5.2. HAZARD DEFINITION

For the microrover, there are many ways such a complex system can fail. Radiation degradation, the impact of landing, thermal cycling, lack of power, etc. These are relatively independent of the way the rover 'thinks' and of the way it moves. However, there are many situations that the rover may encounter when traveling towards a target that can prevent it from reaching its goal, and are indeed very dependent on the rover's mobility and its ability to sense the environment. These are *mobility* hazards, which are hereforward referred to as just hazards. A hazard can present a failure condition, and the rover must be properly equipped to deal with these.

The high degree of mobility of the rover platform reduces the amount of hazardous conditions, and so the definition of what is hazardous is closely tied to mobility. Dynamic simulations by Ma [2] have been developed to characterize the performance of the rover, and to define its envelope of operation. These take into account many parameters, including the soil properties, gravity, frame stiffness, motor torque, and wheel-ground interaction.

Currently, simulation data and experimental data show that, on earth, the rover can:

- Climb objects that are about 2 times the wheel diameter
- Traverse up a 30° slope

- Roll as much as 45°

These performance specifications are by no means fixed, for they are very dependent on the shape and mechanical properties of the terrain. However, they can serve as an estimate of the rover's mobility.

If the environment presents situations that are known to exceed these limitations, then provisions must be made to detect and avoid them. If they *are* exceeded, the result could be a loss of mobility due to the following scenarios:

1. Overturning
2. Tall obstacles in Front
3. Bounded on All Sides
4. Loss of Wheel Traction
5. Vertical Drops
6. Sinking

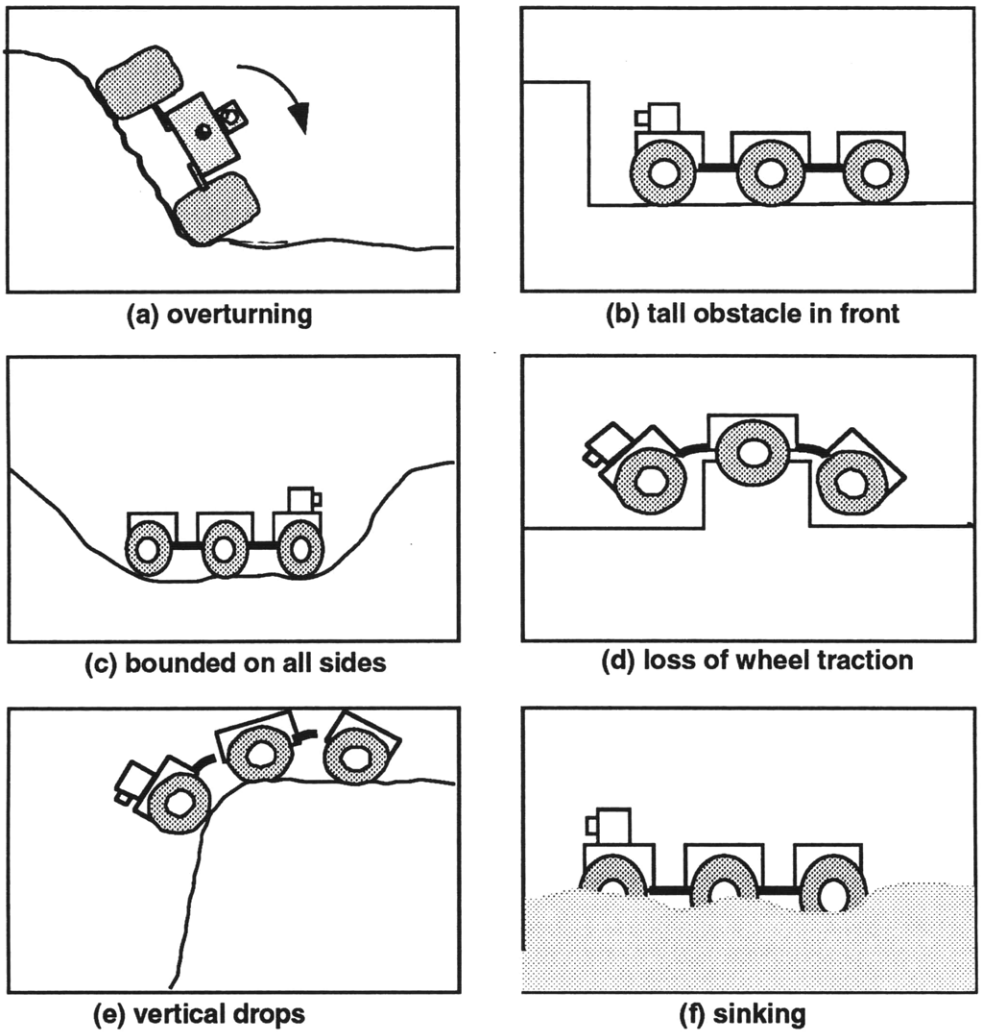

Figure 5.1: The various types of mobility hazards

In addition to exceeding the mobility envelope, a hazard may also be defined as operation close to the envelope, where the efficiency of mobility is so low that it severely degrades the rover's performance. For instance, traveling up a large slope can cause a large degree of wheel slip, and although the rover can still move, it is at the expense of large amounts of energy. The lack of energy can thus prevent the rover from reaching its goal for that day. Hence, very inefficient mobility can be considered hazardous too.

The rover can negotiate different terrain through its mobility, which can be categorized as

- In the ground plane
- Normal to the ground plane

Motion *in* the ground plane is basically forward motion combined with steering in yaw. Motion *normal* to the ground plane can be considered traversing up and down irregular terrain, that is the result of forward motion combined with pitch and roll angles imposed on the rover by the terrain. The rover can make use of these motions to negotiate its environment.

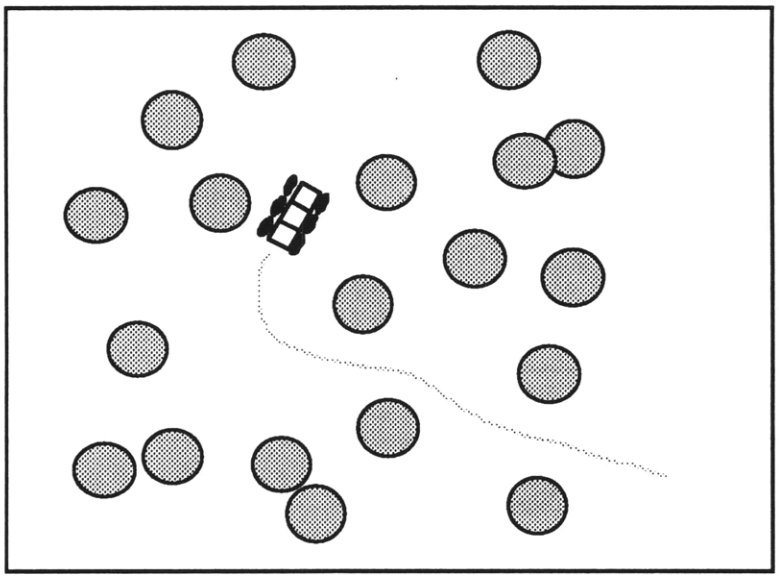

Figure 5.2: Two dimensional obstacle avoidance in the ground plane. Obstacles are sensed at a distance.

5.2.1. Overturning

Some proposed rover designs do not worry about overturning, since they are designed to operate on either side, and/or have provisions for returning to the optimal position. However, this is at a large cost in overall performance, since a vehicle can be much better optimized if only operated in one standard configuration. It would be wasteful to have twice the amount of solar cells, sun sensor, antennas, etc, such that half would never be used at any given time.

Hence, vehicles that are designed with a defined top and bottom inherently are better optimized, but are subject to the hazard of overturning. Overturning can occur from a variety of conditions, most of them from negotiating very irregular terrain, or from exceeding slopes.

Slopes can present a problem especially when the rover center of gravity (cg) is located high above the ground. For similar reasons that race cars are low to the ground to deal with lateral accelerations, a rover needs to have a sufficiently low cg so that the gravity component tangential to the slope ($G\sin\theta$) does not create a moment about the wheel-ground contact that is high enough to overturn. Due to the large length of the rover relative to its width, an overturning hazard on a slope will first occur when the rover is sideways on the slope (Figure 5.1a). It is therefore important for the rover to detect this 'roll' angle.

Irregular terrain can also present the same overturning hazard for the rover. These, however, can occur more quickly, due to the fast-changing slopes of the terrain. The most common example of this is a densely scattered rock field. If the rover is traversing over rocks, its pitch and roll angle will be continually and randomly varying with time. If the angle combination is too high, the rover can overturn rather quickly. Detection of roll and pitch is again necessary, at a sufficiently high bandwidth based on the rover's speed.

5.2.2. Tall Obstacles in Front

The most obvious hazard is the presence of an object that is too large to climb. Any object, no matter how narrow, must be avoided, as there is no other alternative. Hence, motion in the ground plane will be needed to avoid these objects.

A two-dimensional environment is often used to analyze this situation. Consider a flat floor with tall cylindrical objects. Figure 5.2 is representative of such a field. In this environment, the rover has the possibility of traversing this obstacle field if the field coverage is low, based on the rover's size and kinematic constraints (turning ability). In fact, this kind of environment is the model used for the control code on both MITy-1 and MITy-2. Future development is on a three-dimensional model, that accounts for object heights and pitch and roll angles [28].

For the two-dimensional analysis, the rover needs to detect these objects *at least* upon contact. However, if only this level of detection is achieved, then the rover motion will have to involve a large amount of reverse motion, especially due to its kinematic constraints. In this case, mobility will be reduced to an inefficient trial and error process. Hence, it is important to sense the objects at a distance, before actually contacting them.

By sensing obstacles at a distance, it allows the rover to:

1. Choose an efficient path
2. Reduce the chance of having to stop
3. Avoid contact with tall obstacles

In any case, sensing objects that are too large to climb must be performed to some degree, whether it be at a distance, or through contact. When these obstacles are detected, the rover must use its mobility to circumvent the hazard through motion in the ground plane.

5.2.3. Bounded on All Sides

There are some situations that can allow the rover to travel into a certain region, and not be able to travel out. This irreversibility of travel can occur easily when traveling down, and not being able to travel back up, due to gravity. This is shown in Figure 5.1c.

This may be avoided by not traversing down slopes that are greater in magnitude than the maximum positive slope. But when traversing irregular terrain, large local slopes that greatly exceed the maximum uniform slope will inevitably be encountered, due to the articulated frame design. Without some kind of complicated vision system to characterize the shape of the local terrain, it may be very difficult to avoid a crater situation.

It should also be mentioned that reversibility is also possible in a cul-de-sac situation, but this is an primarily a control code issue as long as the azimuthal sensing is adequate.

5.2.4. Loss of Wheel Traction

All motion of the rover, like any wheeled vehicle, is due to the forces exerted by the terrain on the wheel. If this force is reduced, the rover's ability to move is also reduced, like a car on a snowy road. This is especially true in low-gravity environments, where normal and tangential forces at the wheel are already small.

Reduction of the wheel-ground force can occur mainly for two reasons:

1. The ground consists of a low-cohesion material, such as dust or drift

2. The weight of the rover is supported at other portions of the rover other than the wheels.

Much of the terrain of both the moon and Mars can be covered by such material, which can introduce hazards that are non-geometrical, and thus more difficult to detect.

The second situation is possible when traversing irregular terrain, such as over rocks. If there is not proper wheel coverage on the underside of the rover, then objects may support the rover at these points, reducing the forces at the wheels. In the extreme case, the wheels may not even contact the ground, and the rover will have as much mobility as a turtle on its back.

For MITy-2, there is a large amount of room on the underside that can invite this situation. The next prototype rover will have a larger degree of wheel coverage underneath by using conical wheels. Even in this case, losing wheel traction is still possible, but is significantly reduced. Drag wheels at the suggested locations (section 4.2.3) may also help this situation.

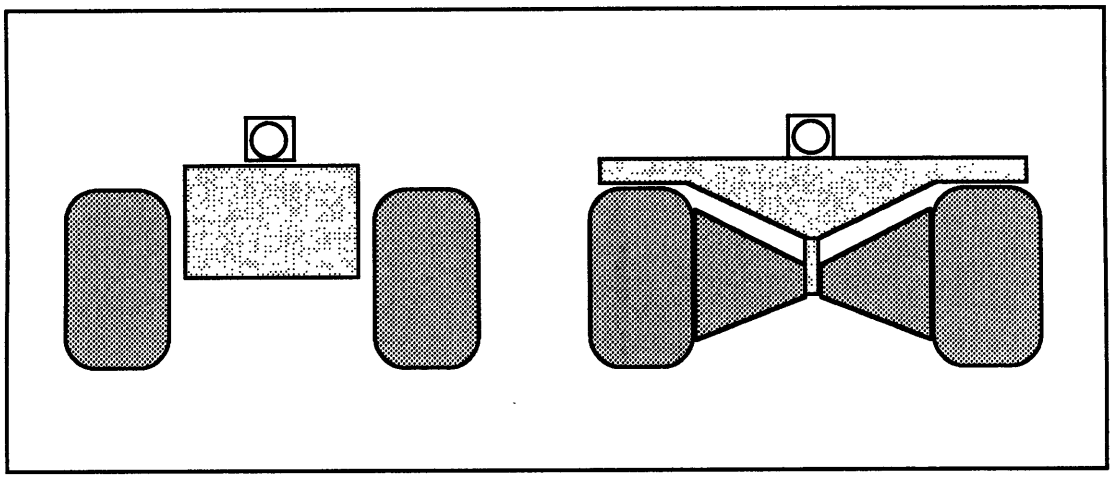

Figure 5.3: 'Dead zones' underneath the rover can unload the wheels, causing a loss of traction. The next prototype, MITy-2a, will use conical wheels for this reason.

Preventing a reduction in wheel traction with sensors is difficult, but detecting when this situation has occurred may still be helpful, since wheel reversal or manipulator arm motion may help to free the rover. This type of hazard can best be avoided in the mechanical design through increased underside wheel coverage, or larger ground clearance. The latter solution can create overturn problems, however, due to the higher

cg. In any case, a way of monitoring wheel forces and motion should be provided, so that a loss of traction can be detected. Monitoring this along with the unpowered drag wheel can be fused to provide more meaningful data for both hazards and efficiency of mobility.

5.2.5. Vertical Drops

A catastrophic failure may occur if the rover falls off of a tall obstacle, or into a steep crevice. This can occur when the slope suddenly changes from near horizontal to near vertical. If the vertical distance is large, the rover can be damaged from the impact of falling, in addition to the possibility of overturning.

Local steep slopes will often be encountered when traversing rocks, which is acceptable as long as the vertical distance is short enough. The problem then arises when the slope is both steep, *and* the vertical distance is large. A large vertical distance is acceptable if the terrain slope is small, and a steep slope is acceptable if the vertical distance is small. It is the combination of large slope and distance that comprises a hazardous situation. Both of these parameters should then be sensed at certain locations on the rover to avoid falling.

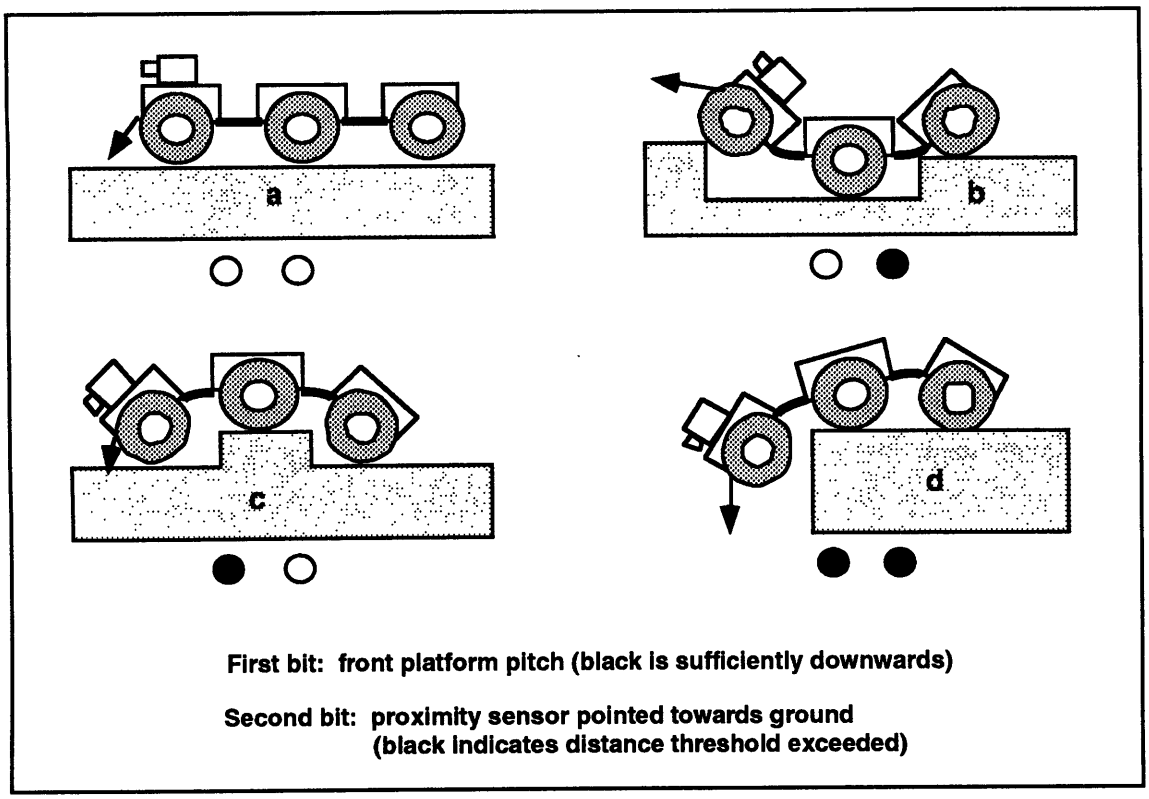

Figure 5.4: Use of pitch and proximity sensors for vertical drop sensing (d). A hazardous situation is defined by two pieces of information. Other cases (a,b,c) are shown to illustrate the need for both sensors.

5.2.6. Sinking

A rather obscure hazard can be introduced by the soft powdery dust or drift material that is common on both Mars and the moon. This can not only cause the wheels to lose traction at the ground, but also may cause the rover to go through, or sink into this material.

For instance, if the rover encounters a hill that is made of drift, then instead climbing over it, the rover may just bury itself. Similarly, if the ground consists of drift, the spinning of the wheels may cause the rover to essentially dig its own grave. This can impose a difficult situation on many of the other sensors as well.

The types of hazards which cannot be identified by their shape, are called non-geometric, and are more difficult to sense. Other obstacles just are assumed to be of a strong solid material, and it is not important to identify one type of rock from another, but only to detect where they are. Non-geometric hazards, on the other hand, require identification,

either at a distance, or upon contact. Fortunately there are not a large variety of non-geometric hazards on the moon or Mars, from what is known anyway. Earth, however, has an abundance of non-geometric hazards (water, tall grass, etc) which can make hazard-avoidance rather difficult. So it is fortunate that at least the amount of these hazards are limited on other planets.

The degree of danger involved with such drift material remains to be determined. Depending on the difficulty in identifying such material, provisions for detection may not be practical. especially when considering the advantage of having man in the loop on at least a daily basis.

5.3. SENSING NEEDS

In the previous section, different hazards were defined, and their corresponding detection was sometimes suggested. The general sensing scheme needs to be determined explicitly in order to satisfy the requirements for hazard-avoidance. The hazards and their sensing strategies are:

1. Overturning: pitch and roll sensors on at least one platform, ideally on all three.
2. Tall obstacles: contact sensors and range finding
3. Bounded on all sides: pitch and roll sensors on one platform, vision system
4. Loss of wheel traction: wheel speed and torque, along with drag wheel odometry
5. Falling: pitch sensor on all three platforms, and short range sensing downward
6. Sinking: distinguishing different terrain materials

It may not be feasible to use every sensor, since the platform size is small, and power requirements are so stringent. However, if the individual sensors can be integrated into the rover design easily, or can be used redundantly for other purposes (i.e. navigation), then use of every sensing strategy can be justified. It should be restated that the design philosophy of the rover is to do the most with the least information, so keeping things simple is the desirable. However, satisfying the mission requirements is the goal.

5.3.1. Pitch and Roll Sensors

Isolating the Gravity Gradient

These sensors are already required for navigation, and are needed for avoiding various hazards too. Data from these sensors will be used to determine the angle of a certain platform of the rover relative to the vertical. This is done by measuring the angle relative to a known reference. This reference is usually the local gravity gradient, which is itself the vertical. The reference can also be from other sources, such as celestial bodies, but this can be more difficult. Hence, most 'tilt sensors' simply use gravity as a reference.

Unfortunately, since gravity is an acceleration, identifying the location of the gravity gradient in the presence of other accelerations can be difficult. Other accelerations from vehicle motion over rough terrain are randomly varying, and can thus be filtered from the gravitational acceleration, which is constant. This, however, requires filtering time which is dependent on the frequencies of the rover-induced accelerations. Since these frequencies are relatively low (due to the slow rover speed), the addition of a low-pass filter can cause accurate tilt measurements to be sluggish.

One way to get around this is to use inertial angle sensors, such as a vertical gyroscope, to give a very fast response to any pitch or roll motion. These sensors do not use accelerations to measure angle, and thus are not affected by the random accelerations of the rover. However, since gyros have drift, they would have to be updated by a true vertical reference periodically. This system would provide both high accuracy from the updates, and fast response from the gyros. Aircraft and spacecraft use this methodology often, so that accuracy is maintained during dynamic maneuvers.

Fortunately, the hazard avoidance control does not require high accuracy during its motion, since it only has to know if the rover is pushing its stability limits. Accuracy within 5° may even be sufficient. Hence, a complex gyroscope system is unnecessary. However, the problem of detecting the local gravity gradient in the midst of other random accelerations still persists. As mentioned, low-pass filtering can solve this problem, at the cost of sluggish sensor response. Time constants of one second or so are considered large for many sensors, but, as all things are relative, the slow rover speed can make this acceptable, depending on the intelligence of the software. Another important point is that the 6-axle design can allow one platform to exceed its inclination limits, and the rover can then stop and back up using the traction of the remaining wheels. In short,

acceleration-based pitch and roll sensors should provide an adequate way of detecting the angle relative to the local gravity gradient.

Inclinometer Techniques

Sensors that measure the angle from the vertical are often called tilt sensors or inclinometers. These typically operate by sensing the position of a bubble in a curved tube of viscous fluid. The bubble position is sensed using one of two methods:

1. Measuring the capacitance of a dielectric fluid
2. Measuring the resistance of an electrolytic fluid

In either method, the capacitance/resistance changes with the position of the bubble within the fluid. As the sensor is tilted, the bubble position changes with respect to the electrodes, due to gravity and other accelerations. Due to the fluid's viscosity, these sensors are inherently low-frequency, usually with a bandwidth under 10 Hz. However, very high accuracies are possible with these devices, especially when the angular range is small.

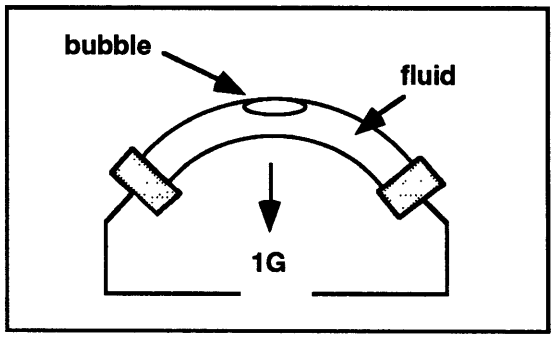

Figure 5.5: Many inclinometers electronically sense the position of a bubble in a fluid.

The *electrolytic* tilt sensors have been used in the aerospace industry regularly, and have been designed to be small, accurate, and often hermetically-sealed. They are therefore a better choice of the two types. A slight drawback to these sensors is that they need to operate on an AC power supply (about 1 kHz), and yield an AC output signal. Therefore, more complicated electronics are needed for both power supply and output signal demodulation and rectification. However, the manufacturers supply these additional electronics in small packages. Unfortunately, electrolytic tilt sensors are more expensive

than capacitive sensors, and therefore MITy-2 uses the latter type. This sensor, made by Lucas Sensing Systems, is described in section 4.2.3.

Some drawbacks to these sensors are their method of detection. Because of the strong dependence of their performance on the fluid properties (viscosity, conductivity, etc), they are not well-fortified against harsh temperature environments, especially on the cold side. In addition, their operation principle involves a somewhat complex design. Although the detector itself can be small, the complete inclinometer package ends up being relatively large. This is especially important when as many as six are considered.

Accelerometer as a Tilt Sensor

Another way to measure the tilt angle is indirectly through accelerometers. Although the capacitive inclinometers can also be considered accelerometers, since the bubble position is dependent on the total acceleration field, they are non-linear when used in this way. Accelerometers are typically high-bandwidth, solid state devices that yield an output that is linear with the acceleration component along its axis. An accelerometer can be used to measure the inclination angle by mounting its sensitive axis horizontally. At level conditions, there is no gravitational component along the accelerometer axis. As it is tilted, the measured acceleration will vary as the *sine* of the angle.

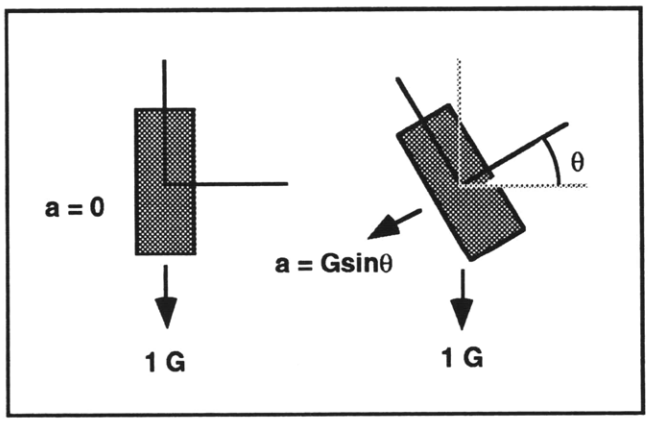

Figure 5.6: An accelerometer can be used as a tilt sensor. By mounting the sensitive axis horizontally, the sine of the tilt angle is obtained directly.

At first, this nonlinearity may appear to be a nuisance, but this is not be the case. One reason is because the sine function can be treated as linear for the majority of the tilt range, mainly since high accuracy is not required. More importantly, however, is that the actual tilt angle itself is rarely used for any computations. The end result of any angle

measurement in the case of the rover control system (for hazard-avoidance or navigation) is the sine or cosine of that angle, so that obtaining the sine of the angle directly is no disadvantage. This is true when operating near the zero portion opposed to the 90 degree portion of the angle range, for sensitivity reasons.

Some of the advantages of accelerometers are:

- Compact and low power
- Solid state construction
- High bandwidth
- Space-qualified
- Self-contained electronics

Because of their compact size, it is not out of the question to place a pair (pitch and roll) of accelerometers on each of the three rover platforms. Even though accelerometers draw little power, further power savings can be accomplished by operating at a low duty cycle. That is, turning on, reading, and turning off until the next sample.

One drawback to using accelerometers is that tilt angle is calculated by the *magnitude* of the measured acceleration. Therefore, any bias error will result in essentially an angular bias error of the same proportion. The main problem with this is that compensation of the bias is difficult, since a known reference is required.

The MITy-2 prototype (section 4.2.3) has both capacitive tilt sensors and an accelerometer so that a direct comparison is possible. Electrolytic tilt sensors were not purchased due to their high cost.

5.3.2. Contact Sensors

Introduction

Contact sensors, are a simple form of tactile sensing that indicate when the vehicle contacts a rigid object with sufficient force. They are generally needed to indicate when a collision has occurred, and offer a simple, reliable method of detecting obstacles in the event of a non-contact sensor failure. The term 'collision' is somewhat relative, since any contact can be considered this. It is the amount of force generated, and the flexibility of the contact area that defines the kind of contact. For example, a cat's whiskers can detect

the presence of an object through contact, yet due to their flexibility, it is not considered a collision, and the cat is not necessarily required to stop. On the other hand, if the cat's head hits the object, a collision has occurred and the cat is forced to stop. Similarly, the rover may have contact sensors that are flexible, allowing deviations without having to stop, and more rigid sensors that indicate when the main body has collided.

Problems with Protruding Sensors

The MITy-1 prototype (7.2.1) uses semi-flexible feelers, similar to a cat's whiskers. These protrude out about 1 foot, enough so that the rover can react to objects on the sides while turning. This was mainly required since the non-contact ultrasonic sensors were not useful under 1.5 feet, and the three of them had a combined field of view of only about 60 degrees. The feelers sensed not only contact above a threshold, but also the degree of deflection through a potentiometer located the pivot. Hence, the location of obstacles could be roughly determined by measuring the potentiometer rotation. The feelers supplemented the ultrasonic sensors sufficiently, but were sometimes troublesome because of their tendency to get snagged on objects and interfere with mobility. Part of this was due to their rigid nature (only one degree of freedom). However, other more flexible designs, such as bend sensors (large-deflection strain gages), can yield noisy data, and still interfere due to their protrusion. In fact, any portion that protrudes can present a mobility problem, and will also be very exposed to the harsh temperatures, unable to be protected near in the rover's main platforms. If the rover has a larger field of view with *non-contact* sensors, there is not a good reason to keep feelers. Hence, the contact sensing should be restricted to the main platforms of the rover.

Ideally a bumper is needed around the perimeter of the vehicle, and any other protruding portion that may collide with obstacles. However, this can also severely interfere with mobility. The MITy-1 prototype actually used retractable bumpers in front of each of the wheels. Although this worked sufficiently, the goal is to minimize as many moving parts as possible on future designs. Much thought went into determining a good solution to bump sensing in front of the front wheels, so that distinguishing obstacle heights is possible. Instead of this, it was decided to have only one bumper at the front, and one at the rear, in between the wheels as shown in MITy-2 (section 7.2.2). Other than these, collision avoidance will mainly take place through the use of non-contact sensing, above the wheel heights.

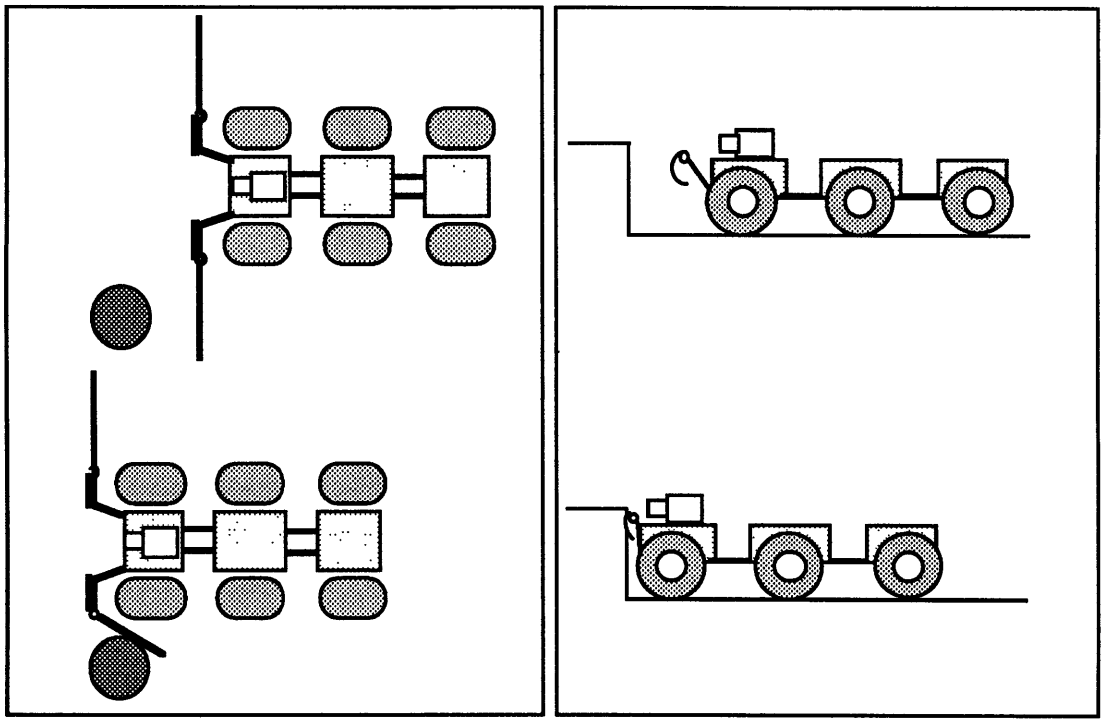

Figure 5.7: Protruding contact sensors on MITy-1. On the left, the feelers can allow a limited reaction before a hard collision. On the right, the bumpers can distinguish object heights in front of the wheels, but needed to retract.

Placement

The placement of the bumpers is very dependent on the wheel and steering design. MITy-2 has a dead zone underneath, in between the right and left wheels, making it vulnerable to rocks. To detect collisions, the bumpers were placed at the bottom edge, at each end of the rover. The next prototype, MITy-2a, will have conical wheels and a much different type of steering. The bumpers still should be placed at the front and rear of the rover (both are leading edges when traveling forward and backwards, respectively), but may have to be higher to clear the wheels when turning. Also, it may be appropriate to place bumpers at other points on the rover where protrusions exist, such as edges of solar panels, the front of the camera, etc.

Mechanical vs Piezoelectric

Bumpers can be passive, making them simple, reliable, and low-power. The design should strive for as little motion as possible, especially if joints are needed. Two ways to achieve this are through piezoelectrics and small mechanical switches.

Piezoelectric material creates a voltage when strained. When a thin piezoelectric film is applied to a side of a 'beam', the strain is magnified and a substantial voltage can be produced. Although motion is needed, it is all in flexing of the beam, so no joints are needed. The result is a robust design that not only indicates contact, but also the degree of force (strain).

Mechanical switches indicate contact discretely, with one bit of information. For instance, a normally open switch can be wired into a single digital port, so that upon contact the switch closes and the voltage is pulled down from logic high, to low. Multiple switches may be used in parallel for the normally open case, or in series for the normally closed case, so that any one or more switch change will yield indicate a digital signal. Although mechanical switches offer only one bit of information, opposed to continuous piezoelectrics, this is really all that is needed. The bumper design can offer a force threshold through springs, so that contact will not be detected unless sufficient force is exerted, thereby discriminating against weak sources of contact.

Although mechanical switches have moving parts, which are to be avoided whenever possible, their complexity is confined to a small, well-protected volume. They have proven their reliability in space, since switches are used in so many space vehicle applications. Hence, mechanical switches are the suggested contact sensor for use on the rover. The integration of the actual switch into the bumper design, and the integration of the bumper into the rover, is discussed in section 7.2.2.

5.3.3. Rangefinder

For efficient and safe mobility around obstacles, some degree of non-contact detection is needed. This is especially important for vehicles with kinematic constraints since forward space is needed for turning. For obstacles that are to large to climb, obstacles must be avoided through steering and forward (or rearward) motion; that is, motion in the ground plane.

Because of this motion, it is important to provide obstacle information mainly in the azimuthal direction. For the two-dimensional case an azimuthal scan at any given height will provide the necessary information for characterizing the obstacle locations. In order to achieve at least 180° scanning, which has been found to be the minimum required for efficient obstacle avoidance, the range finder must be located above the wheel height.

Otherwise, smaller scans can be placed at a lower height, depending on the steering method.

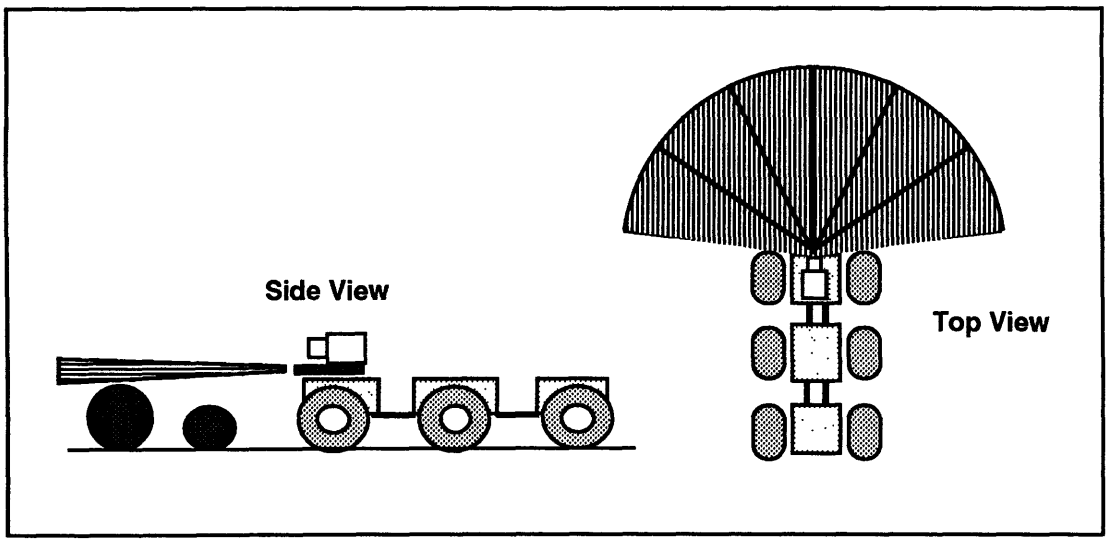

Figure 5.8: Range finding with a 180° azimuthal field of view, from scanning. For a 'tight' beam, objects below the sensing plane are undetected.

A large azimuthal scan, or field of view (FOV), is required mainly for path optimization. If the FOV is too narrow, it is easy to steer into objects, since they cannot be detected in advance. This would result in many collisions, as demonstrated in simulations by Malafeew [1].

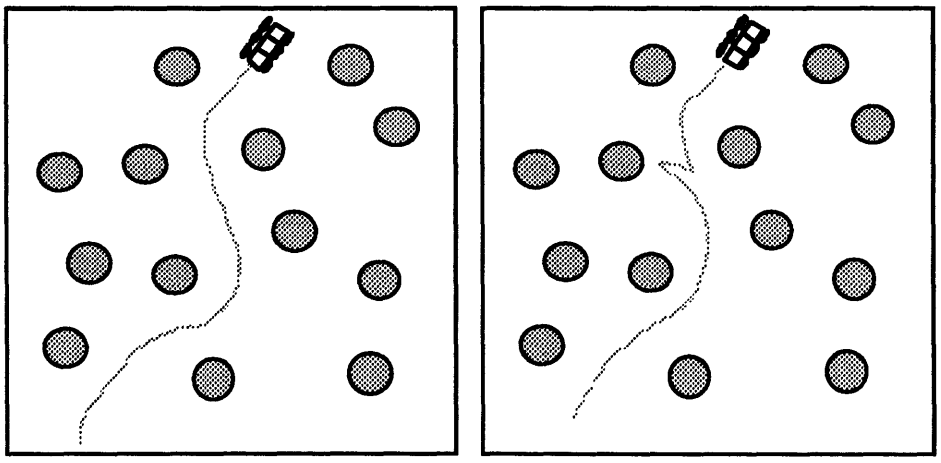

Figure 5.9: Example of collisions that occur while turning, due to a narrow range finder field of view.

A scan in elevation would be helpful, but may not be necessary. In fact, there are many complexities that can arise from the elevation-scanned range data, which will require

intensive computations to be useful. Although range finding in both elevation and azimuth can create a detailed depth map of the environment, such systems are generally practical on much larger platforms. The methods for range finding will be discussed in section 6.2.

Ideally, a 360° FOV in azimuth would be provided, so that the obstacle environment can be characterized all around the rover. However, the most important FOV is in front, since the rover's kinematic constraints. Although a small portion of the time may be spent traveling backwards (after a collision), it does not justify placing an equal amount of sensing in the rear. Therefore, a 180° FOV in front should suffice. The accuracy the range finding is most important up close, when there is little space to react. These and many other range finding issues are discussed in chapter 6.

5.3.4. Short Range Sensing

There are some occasions where motion in the ground plane will benefit from non-contact sensing at distances that are much shorter than those needed for large obstacle avoidance. In order to detect vertical drops before it is too late, it has been established that a short range sensor is needed. In addition, general detection of obstacles in the near field can be an important supplement to contact sensors, which are limited to placement on the vehicle body due to their interference with mobility. Hence, short range sensing can provide important data for both vertical drops and short-range obstacle detection.

Due to the frequent pitching and rolling that the rover will encounter when traversing rocks, the chance of false alarms is high with vertical drop detection, if only ranging is used (see Figure 5.4). In fact, this kind of vehicle motion will make any range data more difficult to interpret unless the direction of the field of view is known at a given instance. Also, when a tall obstacle is approached, only at close range can its height be known accurately. Since it may be costly to traverse all obstacles below the plane of view of the long range sensors, the addition of a few close range sensors that are directed in front of the wheels may be appropriate. Also, it is suggested that vertical drops make use of the pitch of each platform to first determine if there is a potential of a vertical drop, which can then be verified by rangefinding.

In theory, short range sensors should be no different than the range finder that is used for longer-range obstacle detection for optimization of the path. However, the fact that the

range is shorter may offer some important advantages that are not available to longer-range sensing. In addition, larger instantaneous field of views are desired for the short range applications, because at this level of sensing it is more important to have proper coverage than to have the angular resolution. These are some of the reasons that short-range non-contact sensing is separately discussed.

For these kinds of sensing situations, it may be satisfactory to have a *proximity* sensor, opposed to a rangefinder. Proximity sensors, like a switch, indicate whether or not a parameter is above or below a certain threshold. In this case, a proximity switch would indicate if an object is within a certain predetermined distance. This is not rangefinding, since only 1 bit of information is available. Although seemingly crude, this type of simple operation can fit well into an autonomous vehicle with limited processing capabilities.

5.3.5. Measuring Material Properties of the Terrain

In order to distinguish different material properties, sensing may be required either through contact, or from a distance. Because of the restricted variety of terrain materials on the moon or Mars (compared to that on earth), this may seem easy. However, it is generally much more difficult to describe an object, opposed to just sensing where it is. Actually, the goal is to only determine if the material is capable of supporting the rover, so that the rover does not sink or become stuck like a car in mud. Since the types of materials that may be encountered are known, this may be done with non-contact sensing. However, in doing so, certain assumptions will be made, such as the dielectric constant of the material (see 6.1). Hence, it is suggested that physical contact is used to directly determine if the terrain is supportive.

Physical contact sensing can be done in a probing manner, so that analysis is possible before the rover actually traverses the area. This can be similar to a blind person probing the area ahead with a stick. However, for the same argument as feelers, a probing device can introduce some problems with mobility, and can make the device very prone to damage.

If the contact sensor is located inboard, the rover can potentially detect the dangerous material only when it is already upon it. This may be the only practical alternative though. If this is indeed the case, individual sampling of the ground may prove

meaningless anyhow, since the vehicle will not be in danger unless a large portion of the vehicle is in contact. Even so, the best solution to determining soil properties is through the rover's permanent contact sensors: the drive wheels.

Just as in the case for determining slip, the wheel speed and torque, along with measurement of the drag wheel rotation, can provide meaningful data on the material properties of the ground. This will likely be easier than detecting sinking.

Ground measurements will likely be beneficial for scientific reasons too, so that the need seems clear to provide some means of measuring torque and speed on at least one drive wheel. Speed should not be a problem, since this feedback will be used for rover speed control anyway. Torque can be sensed indirectly by measuring the current through the motor, which is still relatively simple. The methods to obtain ground strength from these parameters will likely be based on empirical data from tests with the same wheel. Torque measurements can also aid in the detection of collisions.

5.3.6. Summary

The many types of hazards have resulted in a need for various sensors, some of which have multiple uses. Although much more elaborate sensing schemes would provide better detection, there is a large cost associated with such complexity. The detection of hazardous situations needs to only be 'sufficient' enough to allow the rover to achieve its goal.

The suggested sensors for hazard avoidance are:

- accelerometers (tilt sensor)
- contact switches
- 180° FOV range finder
- proximity sensors
- odometry and drive wheel torque sensor

From the above list, the odometry sensors and accelerometers are also needed for navigation, and the range finder can also supplement navigation. Hence, a significant portion of the hazard avoidance sensors have multiple uses.

CHAPTER SIX
RANGE FINDING

6.1. NON-CONTACT SENSING

6.1.1. Introduction

Many different methods have been developed for sensing the presence of materials without actually contacting them. The most common case is sensing boundaries of greatly different mediums, such as a rock in a gaseous atmosphere. These mediums are distinguishable due to their differences in material properties, most important of which are their electromagnetic properties. For non-contact sensing, the signal that is received is dependent on these properties. For instance, the letters on the page that you are reading can be distinguished from the paper mainly because of the difference in reflectivities of the paper and the ink under uniform illumination. A hot aluminum block that is polished will radiate heat less than one that is dull, making the two distinguishable because of their different emissivities.

In order to detect these boundaries without contacting them, electromagnetic (e-m) waves are observed. In general, the signals do not have to be electromagnetic in nature, but e-m waves are unique in that they require no medium for propagation, allowing use in the vacuum of space. Any non-contact sensing then requires information to be transmitted by means of e-m waves. The source of these waves are:

1. Reflection from ambient illumination sources like the sun
2. Blackbody emission due to the surface's finite temperature,
3. Reflection of waves from an externally supplied active source, such as a laser

The former two of these is called *passive* sensing, while the latter is referred to as *active* sensing.

Ambient Illumination

The most common example of passive e-m wave sensing is vision. Our eyes sense the environment passively. We depend on ambient light to illuminate objects, and if the reflected light is strong enough, the signature (intensity and color) of the world around us can be detected. Eyes are most sensitive at the peak spectral emission of the sun, which is near 550 nm. Far from the visible portion of the e-m spectrum however, our eyes can not detect e-m energy, no matter what the intensity. Similarly, if the ambient light is below a certain amount, the eyes cannot function. These same situations apply to robotic vision systems.

Blackbody Emission

Detectors exist that can sense radiation over the entire e-m spectrum, from gamma rays to radio waves. It has been shown that all bodies at a finite temperature radiate e-m energy according to Planck's blackbody distribution. The peak intensity of this radiation falls at a *frequency* that varies linearly with the absolute temperature of the body. This is known as Wien's Law ($\lambda_{peak}T$=constant). Detectors are typically chosen so that they are most sensitive near the same frequency (with wavelength λ_{peak}). Just like our eyes are most sensitive near 550 nm, a typical thermal infrared detector is sensitive around the peak emission of a 300°K body, which is near λ=10,000 nm. It should be noted that the power of this radiation increases as T^4, so detecting low-temperature bodies in this way can be difficult due to the weak emitted signal.

Active Illumination

In some cases, it may be more advantageous to illuminate a surface selectively, and hence detect the emitted signal after a round trip from the emitter (which is usually located close to the detector) to the target, and back to the detector. When the signals are radio waves, this becomes the building block of what is known as radar. By filtering the detected signal based on the emitted signal, it is possible to reject ambient noise well, yielding a good signal-to-noise ratio. In active systems, ambient illumination is undesirable, unlike in a passive system.

6.1.2. Electromagnetic Wave Background

Electromagnetic (e-m) waves exist all around us. They are not only generated by such common devices as radio transmitters, but also by natural means like the sun, and by all other finite-temperature bodies.

Producing E-M Waves

The name 'electromagnetic' refers to the two types of fields that coexist with any disturbance of an electric (E) or magnetic (B) field. Maxwell first showed that a changing electric field produces a magnetic field, and vice-versa. *Radiant energy,* or waves, are produced by *nonuniformly moving charges.* The case of constant current flowing through a long straight wire represents uniform motion of charge, producing fields that are constant with time, but producing no radiation.

Now, if the current in the wire oscillates back and forth sinusoidally with time, we have the case of accelerating, or nonuniform charge motion, and the wire will radiate. E and B fields are in phase in time, and orthogonal in space, and travel outward at speed c. This forms the basis for a basic dipole antenna, which produces e-m waves that propagate normal to the E and B fields. There are of course other ways to produce e-m waves, but all methods have the accelerating charge in common. Even a radiating blackbody radiates in this way, from thermal agitation.

Speed in a Dielectric Medium

The speed of an e-m wave depends on both the permeability, μ, and permittivity, ε, of the propagation medium [17]. However, these parameters are nearly that of free space for thin mediums like the earth's atmosphere. The speed is maximum in free space, where c is 3.00×10^8 m/s. This is about one million times faster than the speed of sound in air. Any medium will decrease the speed of an e-m wave, as well as its wavelength, while the frequency is preserved.

The permittivity and permeability of any propagation medium are the factors that determine most e-m interactions with it. For clarity, it is customary to present both parameters relative to their free space values. Hence, we have the relative permittivity, $K_e = \frac{\varepsilon}{\varepsilon_0}$, also known as the dielectric constant, and relative permeability $K_m = \frac{\mu}{\mu_0}$.

The speed of e-m radiation in free space, as predicted by Maxwell, is

$$c = \sqrt{\frac{1}{\mu_o \varepsilon_o}} \tag{6.1}$$

In a dielectric medium, the speed becomes

$$c = \sqrt{\frac{1}{\mu\varepsilon}}. \tag{6.2}$$

The ratio of the e-m wave velocity in free space to that in the material is known as the index of refraction,

$$n = \sqrt{\frac{\mu\varepsilon}{\mu_o \varepsilon_o}} = \sqrt{K_e K_m} = \frac{c_o}{c} \tag{6.3}$$

Most dielectric materials have no magnetic properties, and hence $K_m=1$. Therefore,

$$n = \sqrt{K_e} \quad \text{(Maxwell's relation)} \tag{6.4}$$

showing only a dependence on the permittivity ratio. Since $n > 1$, the e-m speed in materials is simply $c = \frac{c_o}{\sqrt{K_e}}$.

It should be noted that although Maxwell's relation holds for all dielectrics, K_e, and thus n, are functions of the wavelength of the propagating disturbance. In most of the e-m spectrum these parameters are relatively constant, but in certain bands can vary strongly. This phenomena, known as dispersion, is due to resonant interactions with atoms in the medium. For a given medium, many variations can exist, due to the different atomic or molecular oscillatory modes. In these bands, e-m absorption is large, and energy is largely removed from the wave and thermally dissipated, especially in dense materials. The main point here is to be aware that because of dispersion, Ke, and hence n and speed, can be strongly affected by the wavelength, resulting in widely varying material properties that are important to sensing.

Reflection from a Boundary

When an e-m wave encounters a boundary of two mediums, say air and glass, reflection and refraction take place due to the larger dielectric of the glass. A portion of the signal gets reflected back, and a portion of it is transmitted. The transmitted portion undergoes

refraction, obeying Snell's law, which forms the basis of lens theory. This portion of the signal will sometimes appreciably decay with penetration distance due to absorption. The reflected signal, which is of main interest, will typically be both diffuse (uniformly reflected in all directions) and specular (mirror like, at the same reflectance angle as the angle of incidence), depending primarily on the surface roughness relative to the size of the wavelength.

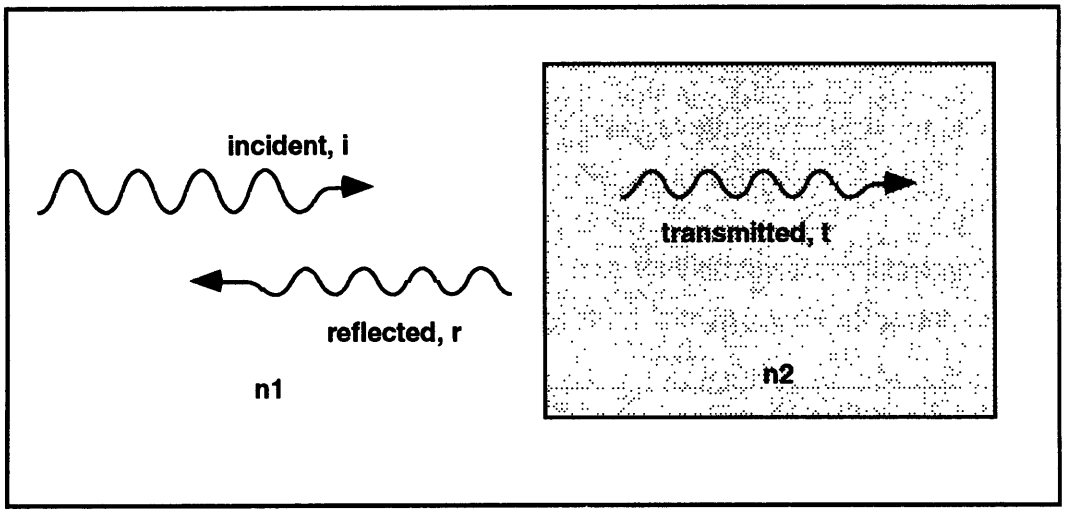

Figure 6.1: Reflection of an electromagnetic wave at the boundary of two dielectric media.

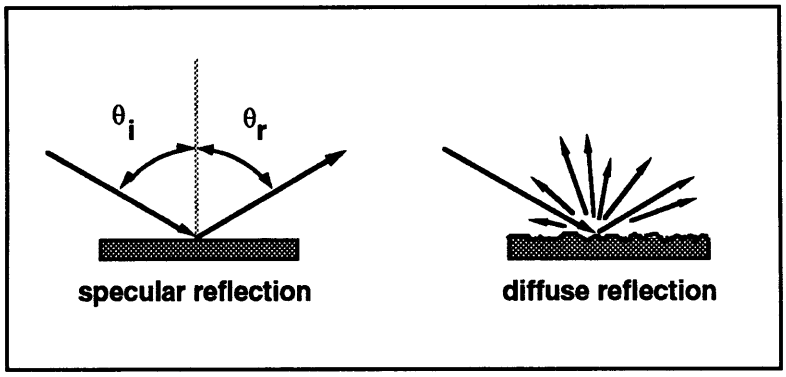

Figure 6.2: The two basic types of reflection: specular and diffuse.

The strength of the reflected signal is for the most part due to the different n's of the two mediums, but is also a strong function of the angle of incidence. First, take for example a wave incident normal on a piece of glass that has no absorption. Since the glass has a higher index, the permittivity and hence the magnitude of the E field of the wave in the glass is lower. Since no absorption takes place, the 'missing' energy is in the reflected wave.

The equations governing transmission and specular reflection of e-m waves are known as the Fresnel equations. These generally apply to optics, and are constrained to smooth planar surfaces between two isotropic media. However, they still can apply to the rest of the e-m spectrum, especially the longer wavelengths where most reflections are specular. The Fresnel equations present the reflection coefficient, which is the ratio of the amplitude of the electric field of the reflected wave to that of the incident wave. These are distinguished by the orientation of the electric field relative to the plane of incidence, yielding a coefficient for a perpendicular and parallel reflections:

$$r_{\perp} = \frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t} \qquad r_{\parallel} = \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_i \cos\theta_t + n_t \cos\theta_i} \qquad \text{(6.5 a and b)}$$

where i denotes the incident medium, and t denotes the transmission medium, θ is the angle from the surface normal, and n is the index of refraction.

For randomly polarized waves, or large beamwidths relative to the surface features, these equations may be more meaningful if the calculated values are interpreted as the envelope of possible values. Nonetheless, they offer insight into the effect of the angle of incidence and the dielectric index. One important characteristic of each coefficient is that the reflection approaches 100% near grazing angles. A typical plot of reflectance vs angle of incidence is shown in Figure 6.3. Note that both coefficients have the same value at 0° and 90°. Although the parallel coefficient actually dips (due to Brewster's law, [16]), it can generally be stated that as the angle of incidence is increased, e-m waves are reflected more. This is especially true at the larger angles.

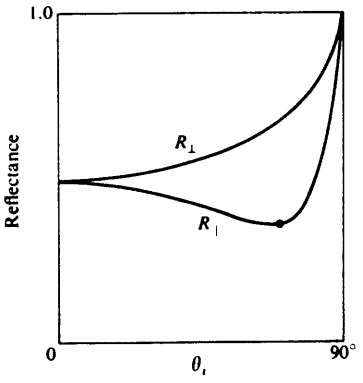

Figure 6.3: Typical reflectance for a linearly polarized beam of white light on an absorbing medium [16].

Although the Fresnel equations offer insight into the strength of the reflected wave, they do not apply to diffuse reflections. As mentioned earlier, the shape of a reflected wave depends on the surface roughness relative to the wavelength. The reflection can be considered specular or diffuse, or anywhere in between. Using light as an example, a mirror yields a specular reflection, so the angle of reflection equals the angle of incidence. A mirror is smooth not just to the touch, but also relative to even the small wavelengths of light. However, to X-rays, a mirror may not appear very smooth, and so the reflection will be more diffuse in that case. For longer wavelengths such as microwaves or radio waves, many surfaces appear mirror-like, making it difficult to get a strong return at oblique angles. Anyone who has worked with common ultrasonic sensors (the same idea applies to acoustic waves too) has probably experienced specular reflection problems due to their long (millimeters) wavelength.

In order to determine whether or not diffuse reflection is possible, one must analyze the object's surface roughness, relative to the wavelength. Before stating the criterion, it should be expected that the relative smoothness of an object is dependent on both the wavelength, and the angle of incidence (even dull objects can be mirror-like at near grazing angles). For an object to be rough, thereby diffusely reflecting, the *Rayleigh roughness criterion* states that

$$\xi > 1$$

$$\xi = \frac{8\pi s}{\lambda}\cos\theta \tag{6.6}$$

Where s is the variance in height distribution for the reflecting surface, λ is the incident signal wavelength, and θ is the observation angle with respect to the surface normal [18]. Note that the equation is proportional to s/λ, and to the *cosine* of θ. Because of this angular relationship, small deviations from the normal have little effect on the apparent roughness, while very large changes in apparent roughness occur near grazing angles. From this equation it is easy to see that as θ approaches 90 degrees, $\cos\theta$ and thus ξ rapidly decrease, and the surface is no longer considered rough. This quantitatively explains why objects can easily appear mirror-like at near grazing angles. This behavior is similar to the specular reflection case.

A uniformly diffuse surface reflects equally in all directions. Although this does not happen in reality, many surfaces approximate this, such as a laser spot on a wall. A more accurate representation of a good diffuse reflection is what is known as a Lambertian reflection, which has an intensity distribution as shown in Figure 6.4. In this case, the reflection intensity is zero parallel to the surface, and maximum normal to the surface. If the Rayleigh roughness criterion is satisfied, the reflection can be considered Lambertian.

Proceeding one step further, most diffuse surfaces reflect with both a specular and a Lambertian distribution. In general it is desired that the reflection be as uniformly diffuse as possible, so that detection is still possible at even grazing incidences. This is particularly important for detecting edges of spherical objects.

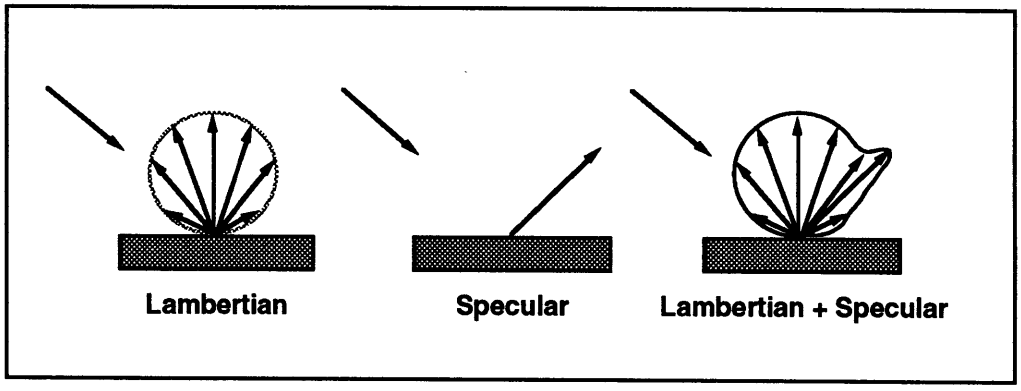

Figure 6.4: Lambertian and specular reflections, and a combination of both.

6.1.3. Choosing a Wavelength: Environmental Considerations

Overview

The electromagnetic spectrum is shown in Figure 6.5. In designing a ranging system, many factors must be taken into account in choosing which part of the spectrum to operate in. Considerations include not only the behavior of the signals in the operating environment, but also the issues governing the hardware, such as size and power constraints of both the emitter and detector.

Before proceeding, it should be noted that the portion of the spectrum that will be considered is from the visible to radio waves. Much work has been done with radar systems, and this discussion certainly applies. However, the system requirements for common radar systems are quite different from those of a small rover, so that it should not be surprising if a different portion of the spectrum is suggested.

Considering the issues of an electromagnetic wave interacting with the environment, some of the main issues are

- attenuation in the atmosphere
- penetration of solid matter
- resolution
- reflection characteristics
- ambient noise

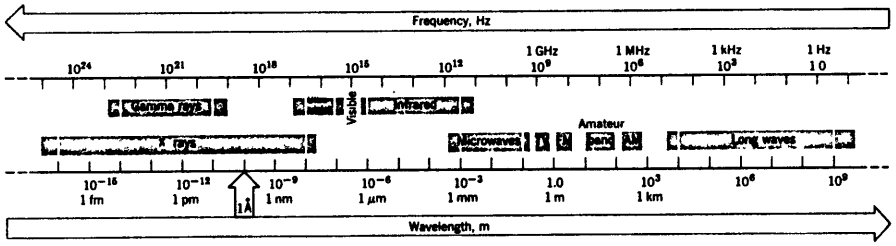

Figure 6.5: The electromagnetic spectrum [17]

Attenuation

In many sensing or communications applications, attenuation in the atmosphere is an extremely large factor governing the design of the system. For example, a communications satellite needs to transmit a signal through the entire atmosphere of the earth. For these large distances, the attenuation properties of the atmosphere on the emitted signal can strongly decrease the intensity by the time it reaches the receiver. The attenuation characteristics can vary strongly with the wavelength of the signal. This can be seen in Figure 6.6, which shows a plot of the transmission properties of the earth's atmosphere over a wide range of the e-m spectrum. To keep power requirements low, it is desirable to transmit within certain frequency 'windows' of the atmosphere, where the attenuation is relatively small.

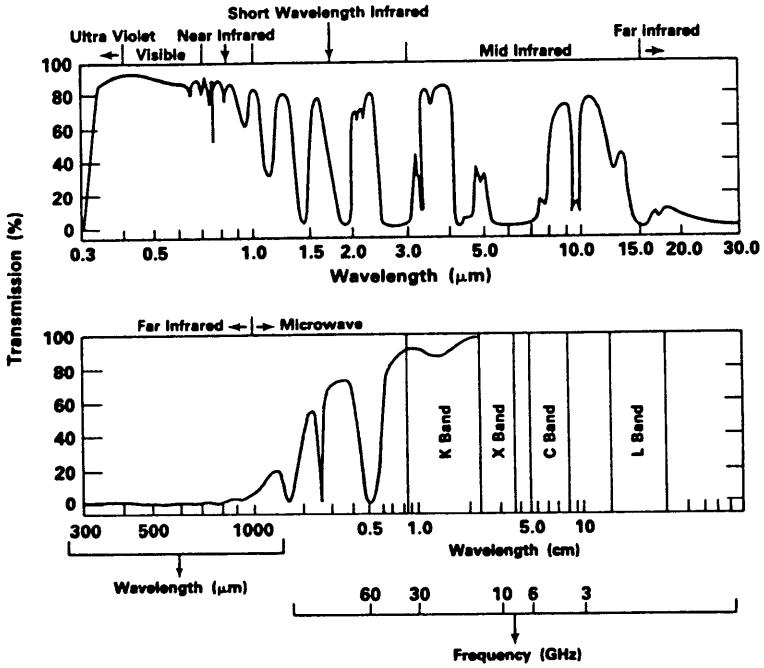

Figure 6.6: Transmission vs. wavelength through the earth's atmosphere. The peaks are known as atmospheric windows [19].

Attenuation can occur through many mechanisms. However, most atmospheric attenuation is due to the absorption and reradiation to the surrounding molecules. This occurs very strongly near the resonant frequencies of atoms or molecules, which is why there often are distinct windows of in the atmosphere. Hence, attenuation is in general very frequency dependent. This applies to both solid and gaseous media.

For close range finding, attenuation due to the thin atmosphere of Mars will not be a factor, simply because the distances are small (of the order of meters, not kilometers). In fact, thin atmospheres favor e-m wave propagation (unlike acoustic waves), the extreme of which is free space, which transmits 100%. Hence, both the moon and Mars will not pose problems regarding attenuation of the signal on the basis of the gaseous atmosphere. However, both surface present the problem of solid matter attenuation, most likely from the fine soil (dust or drift).

Penetration of Solid Matter

Another important factor to consider is the penetration of solid matter. For example, in some terrestrial cases, it may be desirable for the signal to penetrate the ground or water to a certain depth. Specifically, when analyzing the ocean floor from an orbiting satellite, the signal needs to penetrate the water depth, and reflect when the floor is reached. In the rover's case, some interesting situations may occur if soft powdery hills exist, as was discussed in section 5.2. These examples illustrate why in some cases, the wavelength may be chosen to be selectively penetrable through certain materials.

In a dust storm or under generally dirty conditions, the atmosphere or dirt on the optics may be enough to seriously degrade transmission or detection. This case presents a basic problem, since it is desirable to detect this material in the form of obstacles, yet when covering the transmitter/detector, transparency is desired. The visible spectrum falls under this category, as does most of the infrared. In the longer wavelengths such as micro- and radio waves, penetration of dirt deposits can become significant. One reason that 'all-weather' rangers use longer wavelengths is because they are not seriously impaired by dirt, rain, fog, and small amounts of solid matter in general. In fact, they can be covered by protective shields that are easily penetrated, such as the nose (radome) of an aircraft. The current development of rangers for automobiles is in fact in the millimeter/microwave regime strongly for the same reasons. Even longer wavelengths are even better suited for penetration, but, hardware then become an issue.

Resolution

The influence of wavelength on resolution is mainly due to diffraction. Diffraction is essentially the spreading of waves due to interference. The theory behind it will not be discussed here, but its effects can be explained easily.

When waves of water in a ripple tank pass through a slit that is of the order of the wavelength, the waves exiting the slit do not continue to travel in the same direction bounded by the slit width, but instead flair out. The smaller the slit, the more the flair. However, this beam divergence effect is not very noticeable unless the slit width, or 'aperture' is small enough relative to the wavelength.

For circular apertures, divergence angle is governed by (for small angles):

$$\theta = 1.22\frac{\lambda}{D} \tag{6.7}$$

where θ is the beam half-angle, and D is the aperture diameter. This applies to the far field pattern, which is basically the irradiance pattern far away, relative to the aperture size. θ then locates the angular position to the first minimum of the of electric field, and hence the first minimum of irradiance I (irradiance is proportional to the square of the electric field). This central bright spot within the first minimum is known as the Airy disk. It is important to realize that radiation still exists outside of the Airy disk, but is relatively weak. Hence when producing a beam through an aperture, pure collimation is never possible, and is at best diffraction-limited according to the above equation. The same principle applies to *collecting* energy such as focusing a camera lens.

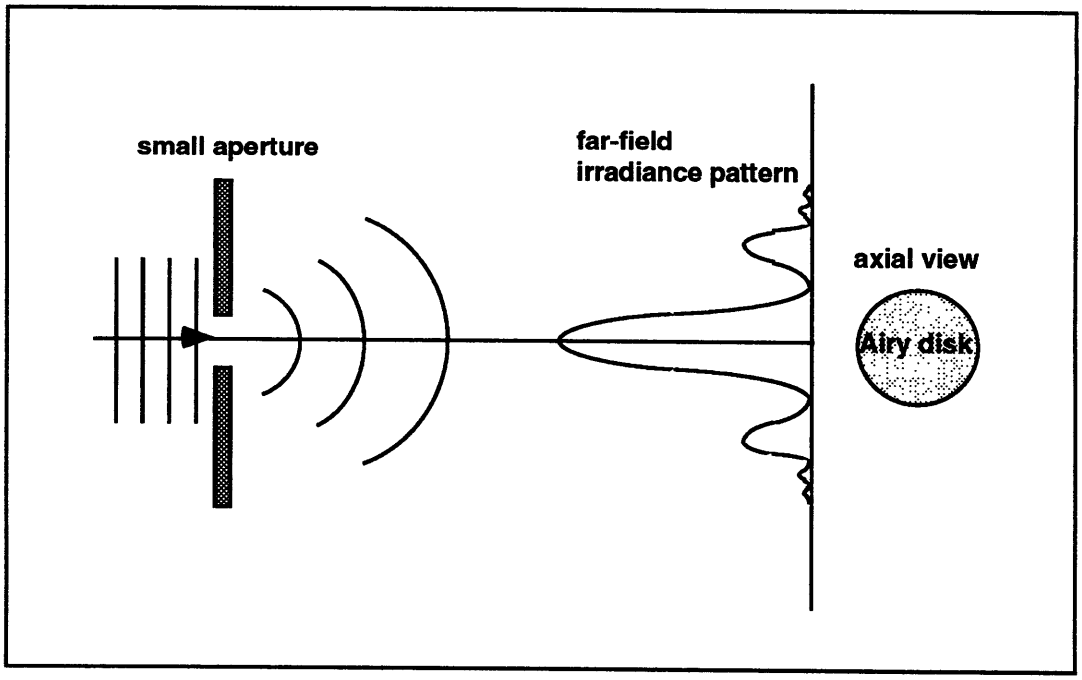

Figure 6.7: Diffraction pattern due to a small circular aperture. Most of the irradiance falls within the central peak, which is known as the Airy disk.

To have control of the geometry of e-m energy that is produced or collected is extremely important, especially for resolution. A typical photographic lens that is many millimeters in diameter is much larger than the visible light wavelengths, which are less than a micrometer. Hence, the ability to gather light and produce a high resolution image is rather easy since diffraction effects are not evident. For most practical purposes, the resolution is bottlenecked by the hardware and film, not by diffraction.

However, as the spectrum of interest falls under longer wavelengths, diffraction can quickly become the resolution bottleneck. A good example of this is infrared and microwave imagery of the earth by satellites. The fine details produced by visible light is lost, and the image is noticeably blurred. This effect can be countered by using a large collection aperture, but this has its obvious practical limitations.

In addition to imagery, resolution is important in collecting energy even for the purpose of getting a strong signal. Radar dishes for example need to be large since λ is so large. Parabolic dishes gather and focus energy to a 'point' only if the dish is large relative to λ. Therefore, resolution is important even for the purpose of increasing signal strength.

As mentioned, resolution is a major factor also in generating a beam in a certain pattern. Tight, or collimated beams, are often desired for good angular resolution in object detection. In order to have a well-collimated beam, the emitting aperture must be large relative to the wavelength. In the visible spectrum, this is not much of a problem, and lasers can produce a 0.2 milliradian divergence beam that is only a few millimeters in diameter. With larger wavelengths, the emitting aperture needs to be proportionally larger to have the same beam divergence. For instance, a microwave signal of 6 cm wavelength would require, for the same divergence, an aperture of over 300 m! This is obviously unrealistic, so it is customary to accept the penalty of larger beam divergence when in the longer wavelengths.

In these cases, when a small λ/D is not possible, it is important to realize also the effect of the emission pattern beyond the first minima. The far field irradiance pattern from a uniformly illuminated circular aperture is governed by

$$I(\theta) = I(0)\left[\frac{J_1(\frac{\pi D}{\lambda}\sin\theta)}{\frac{\pi D}{\lambda}\sin\theta}\right]^2 \tag{6.8}$$

where θ is the angle off the boresight, and J_1 is the Bessel function (of the first kind) of order 1 (see Figure 6.8). A slice of the 3-D pattern is shown in Figure 6.7. The magnitude of the second maximum is less than 2% of I(0), and the third maximum is less than 1/2 % of I(0). These are commonly known as side lobes. Fortunately, 84 % of the total power falls within the main lobe, or Airy disk. These are characteristics that are associated with any uniformly illuminated circular aperture. However, the side lobes are often forgotten.

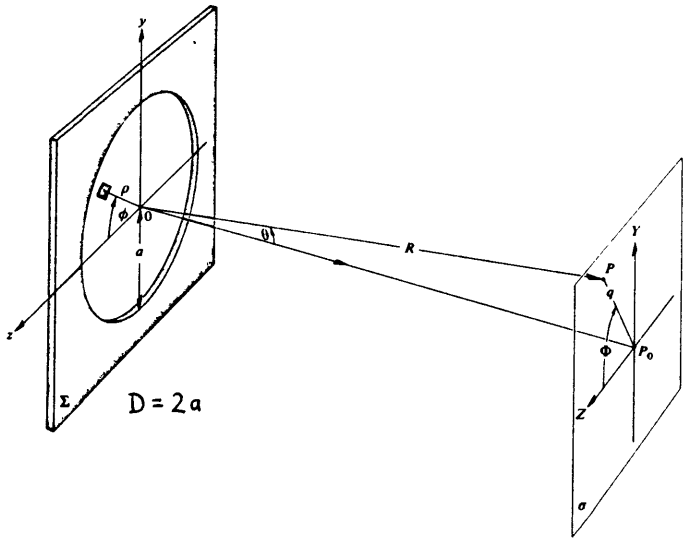

Figure 6.8: Geometry for equation 6.8 [16]

When the beam distribution does have significant sidelobes, it may be possible to reduce these by reducing the amount of illumination near the edges of the aperture. However, this only reduces but does not eliminate the side lobes, so they must still be taken into account. Therefore in either case, there is still a rather strong variation of the beam with the angle off boresight. This is commonly shown as in Figure 6.9. The beamwidth is usually designated as the boresight angle at which the power has decreased to 1/2 of the maximum , or by 3 dB. The presence of sidelobes essentially degrades the resolution of the system, since the ability to focus on a small specific area is lost. Also, although the side lobes are weaker, they can become a problem at close ranges.

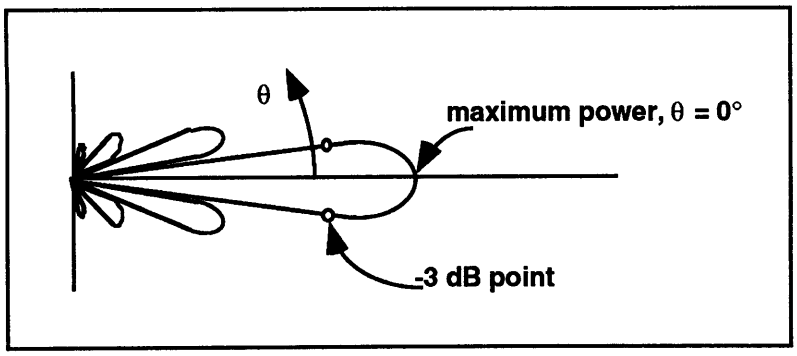

Figure 6.9: Polar plot of illumination power (not to scale). The beamwidth is typically defined as the full angle of the -3 dB points

Reflection Characteristics

One of the main reflection issues in choosing a wavelength is that of diffuse vs. specular reflection. As mentioned earlier, diffuse reflection results in a uniform reflection in all directions from the surface, while specular reflection results in mirror-like reflections, in which the incident angle equals the reflection angle. For environments where the angle of incidence cannot be controlled, diffuse reflection is preferred, since otherwise there is negligible returned signal strength when the surface is at oblique angles. Hence, active systems operate much better if the surface is rough relative to the wavelength based on the Rayleigh roughness criterion (eq. 6.2). Unfortunately, many objects appear mirror-like to wavelengths in the millimeter regime and up. In these cases, especially for small beam diameters, only normal angles of incidence will suffice. This problem can be solved by either using a smaller wavelength, or a beam that is sufficiently large so that the object within varies enough to provide local surfaces that are normal to the beam. The latter, however, results in an angular resolution loss due to the wider beam. Diffuse reflections occur on most rocky surfaces when the wavelength falls in the IR or lower.

Ambient Noise

For all non-contact sensing, detecting a weak reflected signal can be impaired by the ambient noise. Hence, the emitted wavelength should not coincide with that of the peak noise band.

Noise sources can be from emission of all objects in the sensor's field of view, either directly produced by the object's finite temperature, or reflected energy from another source.

For example, when the active source is near 10 μm, the thermally generated noise from room temperature bodies is also significant in this regime, making detection difficult unless the active source is relatively strong. A more common example is when the active source is visible light. In this case, detecting the reflected signal can be difficult on a sunny day, since visible light from the sun is also reflected from the target to the detector.

The amount of acceptable noise is dependent upon the strength of the returned signal, and the signal-to-noise ratio is thus a more meaningful parameter. Although the noise in the entire system can be from many sources, the ambient noise cannot be controlled. What can be controlled however, is the filtering of the detected signal to reject as much ambient noise as possible. It should be realized, however, that the fundamental problem is the band of operation. By a more judicious choice of the active source wavelength (and detector responsivity), the amount of ambient noise in this band may be minimized.

Summary

In closing, many issues are involved in the selection of wavelength. This section described many of the environmental considerations. These are:

1. Attenuation in the atmosphere should not be an issue for such short distances
2. The penetration of solid matter has a relevant issue concerning dirty conditions, in which the milli-, micro-, and radiowave bands have an advantage.
3. Resolution is important in both emission and detection. A high degree of resolution can be achieved only if the aperture is large relative to the wavelength, giving the visible portion of the spectrum an advantage over the longer wavelengths.
4. Reflection of the diffuse type is desired, so that a strong return signal is possible at large angles of incidence. Again, shorter wavelengths have an advantage.
5. Ambient noise should be strongly considered when choosing a wavelength, so that the active source does not fall in the same band as the noise.

6.1.4. Choosing a Wavelength: Hardware Considerations

Size

As discussed earlier, for a given resolution in the emitted or collected e-m energy, the size of the system increases proportionally to the wavelength, mainly due to diffraction effects. Hence, platforms that need a high degree of resolution, but can not afford the size or weight, will have to stay below the microwave regime.

For instance, consider a 5 mm wavelength. The resulting divergence for a given aperture diameter is shown in table 6.1. Note that even with a large aperture of 50 mm, the beam is still far from collimated.

Aperture Diameter, mm	Approx. Beamwidth (full)
10	76°
20	36°
50	14°

Table 6.1: Effect of Aperture Diameter on Beamwidth for 5 mm Wavelength

Power

In general, longer wavelength systems can accommodate higher power levels, mainly due to their larger size, which can handle the heat generated by all of these relatively inefficient components. In fact, the efficiency of some of these systems is about 30% at best, so that the remaining power ends up as heat which must be removed. For a given type of emitter, the power conversion efficiency is usually strongly dependent on the wavelength.

Large power levels are required when the transmission distances are very large, the surface reflectivity is low, the noise levels are relatively high, the apertures become dirty, or if the collection aperture is small. Power levels can always be increased by simply using emitters in parallel, or by increasing their size. However, the goal of many systems is to produce the most power in the smallest package. In that case, heat dissipation becomes important. In addition, a small source is often desired to produce a spherical wave that can then be converted into a planar wave (collimation), so that power levels are limited due to this.

Of course, the underlying reason in producing a powerful signal is to be able to detect it above the noise level. In short, it all comes down to the signal-to-noise ratio. Hence, the first step taken should be to minimize the noise levels, which has its limitations. A strong tradeoff exists in the cost/benefit effects of a more powerful emitter versus lowering the noise.

The Signal-to-Noise Ratio

To be able to detect an e-m signal of interest, whether it is initiated from an active source, or passively detected, the strength of that signal must be higher than the noise level. For active systems, increasing the signal-to-noise ratio (S/N) can be achieved by simply increasing the power of the emitted signal. However, that certainly has its limitations, especially on a platform with a tight power budget.

In order to improve S/N, the signal strength must be increased, or the noise level must be decreased.

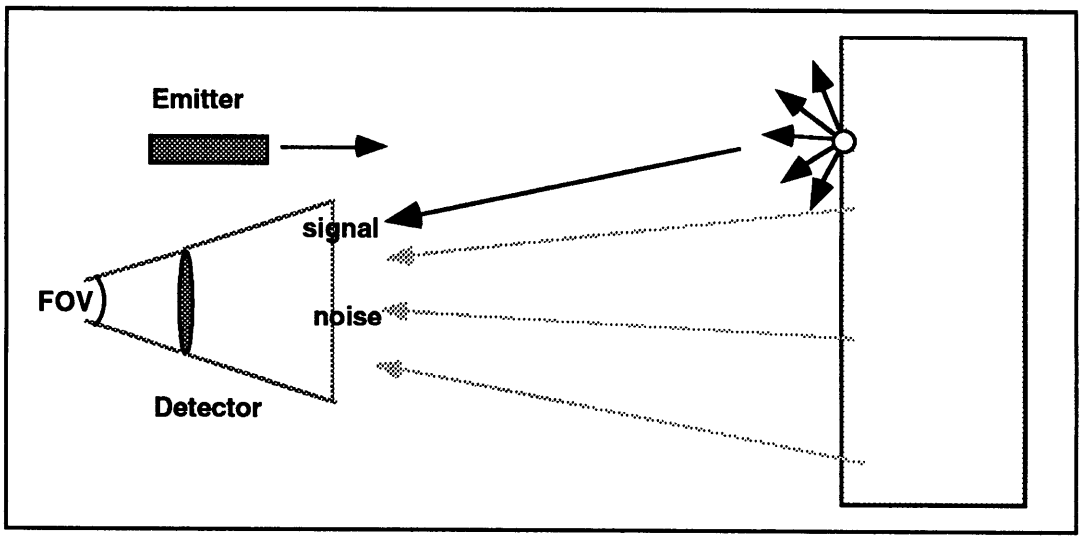

Figure 6.10: Signal and noise in an active system. For the active system shown, noise is other energy within the field of view that the detector is sensitive to.

Signal Strength

Let's assume the emitted signal is a perfectly collimated beam of small diameter. In this case, upon diffuse reflection, the returned signal intensity varies as the square of the range. Assuming that all of the reflected energy is uniformly reflected in a hemispherical

(surface area of $2\pi R^2$) pattern normal to the surface, the irradiance at a given range, R, is:

$$I = \frac{P_r}{2\pi R^2} \tag{6.9}$$

where P_r is the power after reflection. Although a reflection is better modeled as Lambertian (see 6.1.1), this model assumes no dependence on the angle of incidence, and is thus simpler. However, even a Lambertian surface reflects diffusely, but with different strengths in different directions, so that the above model can always be adjusted on an absolute scale by using a gain factor.

The reflected power, P_r, depends upon the emitted power from the aperture, P_e, corrected for the losses involved with traveling through the atmosphere and the reflectivity of the target surface. The reflectivity of many surfaces is rather low, often less than 0.1. Hence, P_e must be chosen accordingly.

Assuming a reflectivity K_{refl}, the reflected signal has a total power of $K_{refl}P_e$. This assumes there is no attenuation in the round trip travel through the atmosphere. Hence, at the collection aperture, the irradiance is

$$I = \frac{K_{refl}P_e}{2\pi R^2} \tag{6.10}$$

The irradiance varies with the inverse square of the range in this case. For other systems with wide beams, only a fraction of the beam may intercept the target [20], and therefore the irradiance may decrease as much as $1/R^4$. This is another advantage of having a well-collimated beam with a small diameter.

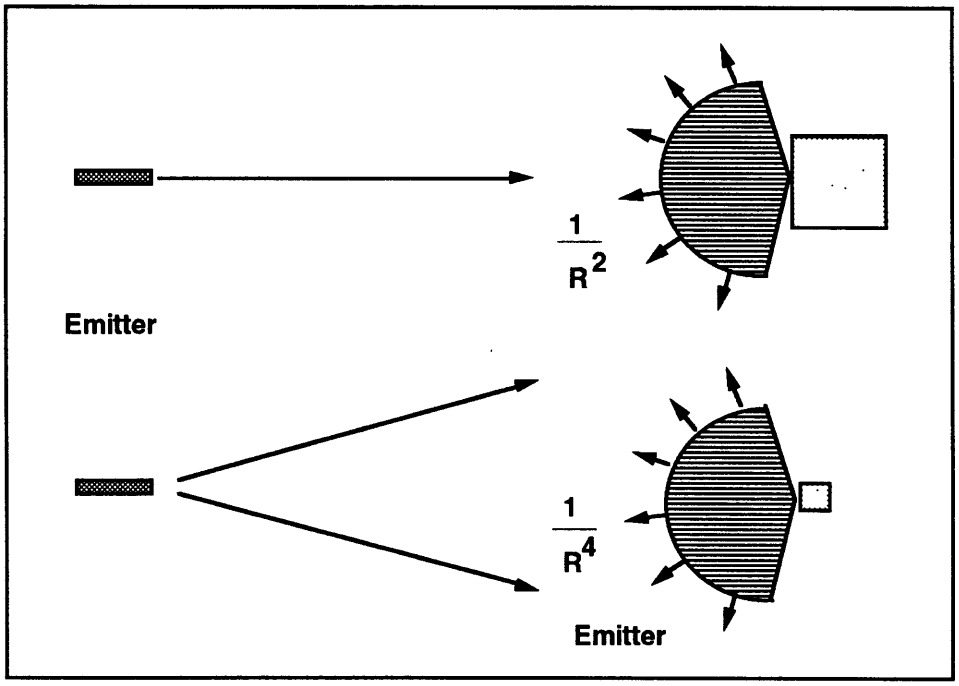

Figure 6.11: Returned signal strength. When the beam is small, the returned signal follows the inverse square law. When the beam is larger than the target there is an additional loss mechanism involved, since a portion of the beam is wasted.

Because of the inverse square relation, the strength of the reflected signal can vary significantly. Not only does the signal decrease rapidly with range, but the receiver must be able to accommodate the large *differences* in signal strength, or dynamic range. This is why the power of the received signal is often expressed in decibels.

The received power depends not only on the irradiance at the receiver aperture, but on the aperture itself. Typically the captured power is focused to a small spot by reflection (parabolic dishes or mirrors) or by refraction (lenses). This allows the actual detector itself to be miniaturized and protected. There are, of course, losses involved with the capturing of the irradiance upon the aperture. These can be from diffraction effects, attenuation from transmission through lenses, unwanted reflections, etc. Hence, every collection aperture has an associated aperture efficiency η (< 1) that represents its efficiency in capturing incident irradiance. The aperture efficiency can range from about 0.45 for parabolic dishes, to 0.99 for lenses. In any case, the efficiency is typically higher when the aperture is designed for a particular narrow band, so in general, the greater the bandwidth, the lower the efficiency.

The collected power is thus dependent on the aperture area, A, and the aperture efficiency, η. These can be lumped together and called the effective aperture area, A_{eff}.

$$A_{eff} = \eta A \tag{6.11}$$

The collected power is then the irradiance at the aperture times the effective aperture area.

$$P = A_{eff} I = \frac{A_{eff} K_{refl} P_e}{2\pi R^2} \tag{6.12}$$

Detector

Once the signal irradiance is collected, the power of that signal is transformed to a usable signal (such as a small voltage or current) by the detector itself. The stage before this is referred to as the collection optics.

Depending on the frequency band and other factors, the detector may take on many forms and concepts. In some cases, it can even be the same element that emits the signal, in which case it is called a transducer or transceiver.

Whatever the transforming concept, there is a responsivity, K_{resp} associated with the collected power. For instance, a silicon detector will produce about 0.6 amperes per watt of e-m power for certain portions of the spectrum. The responsivity is frequency dependent (and therefore is also called 'spectral response'), and thus plays a major role in signal and noise detection.

The responsivity has a maximum value, which decreases to zero eventually towards the higher and lower wavelengths. Eyes, for instance, have a peak responsivity near 555 nm, which decrease to a very small amount at wavelengths of 430 and 690 nm. Hence, these fuzzy boundaries define the visible portion of the spectrum. Similarly, other detectors have bands of operation. Figure 6.12 shows a typical responsivity curve, or spectral response, for a common silicon detector. The detector is responsive in the band from about 300 to 1100 nm, thus extending slightly into the ultraviolet and infrared portions. Note that responsivity curves do not typically have sharp cut-off bands, but instead smoothly decrease towards zero. One must be careful that the detector responsivity may

be small, but non-zero in a very large band, so that strong noise sources *may* still be detected far from the sensitive band.

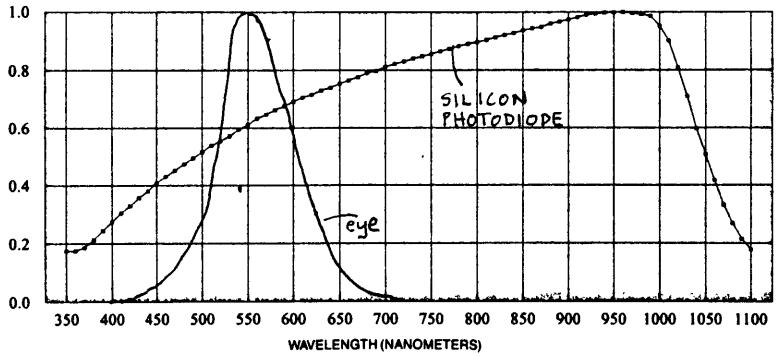

Figure 6.12: Normalized responsivity of the human eye and silicon photodiode

If possible, the detector is chosen so that its peak response falls near the wavelength of interest, and far from the wavelength of high noise levels. Any signal that is passed through the collection optics and to the detector has an associated responsivity that is wavelength dependent, so that the usable signal (including noise) is a product of both the collected power spectral distribution, and the spectral response of the detector. The power or irradiance of an e-m signal is more accurately presented normalized by its wavelength. For instance, instead of P expressed in watts, P_λ can be expressed as watts/nm. The total power incident on the detector is then

$$P = \int_0^\infty P_\lambda(\lambda)d\lambda \tag{6.13}$$

By expressing power in this form, it is possible to accurately find the collected usable signal given the responsivity:

$$S_{collected} = \int_0^\infty P_\lambda(\lambda)K_{resp}(\lambda)d\lambda \tag{6.14}$$

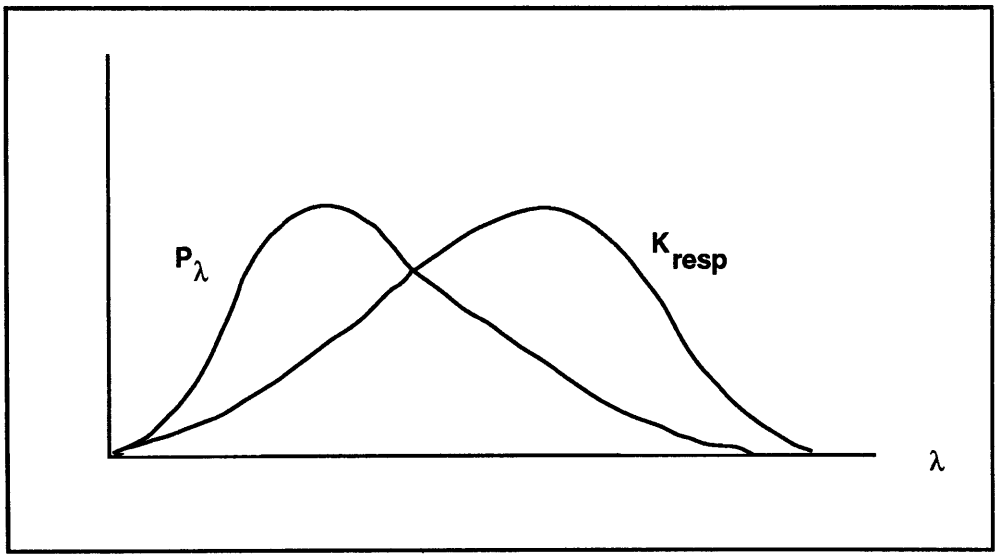

Figure 6.13: Example of spectral power and responsivity. The usable signal is the integral of the product of these two curves.

The final curve is what contributes to all the power collected, whether it be signal or noise. It should be clear that matching the responsivity with the signal of interest is very important. However, noise may also be strong in the same band. Hence, both the signal and the noise will be present in the collected energy. Passive filtering may help reduce the noise, but from this point on in the detection circuitry, the receiving electronics operate on only this 'usable' electrical signal, which can be manipulated by analog or digital methods.

Noise Strength

Noise is considered the incident irradiance that is in not in the spectrum of interest. For instance, when detecting passive infrared radiation at 10 μm, the weak signals can be impaired by reflected solar radiation, if the detector is still responsive near the visible spectrum. In fact even if the responsivity in the visible is only say 1 % of that at 10 μm, the irradiance from each mechanism can be of similar magnitudes since the solar irradiance is so much larger, even after being reflected from a 'dark' surface.

This example assumed no passive filtering. Filtering can limit the band of what passes through to the detector, so that noise outside the filter band can be significantly reduced, even if the detector is still responsive in that band. Passive filtering is thus very beneficial when the ambient noise and emitting signal bands do not overlap much, and

when the detector is still responsive in both bands. However, when the signal and noise bands do overlap significantly, the advantages of passive filtering is greatly reduced.

With the addition of a passive filter, the usable collected signal represented in eq. 6.14 now becomes

$$S_{collected} = \int_0^{\infty} P_\lambda(\lambda) K_{filter}(\lambda) K_{resp}(\lambda) d\lambda \tag{6.15}$$

where

$$P_\lambda = P_{\lambda_{signal}} + P_{\lambda_{noise}} \tag{6.16}$$

For a constant ambient noise source, S/N will decrease as the signal strength decreases, which is as the inverse square of the range [20]. At some point, the range is large enough that the S/N approaches unity (Figure 6.16). Beyond this, it is not possible to isolate the signal of interest. The noise level in Figure 6.16 represents all noise contributions, not just ambient noise. These will also be discussed in later sections.

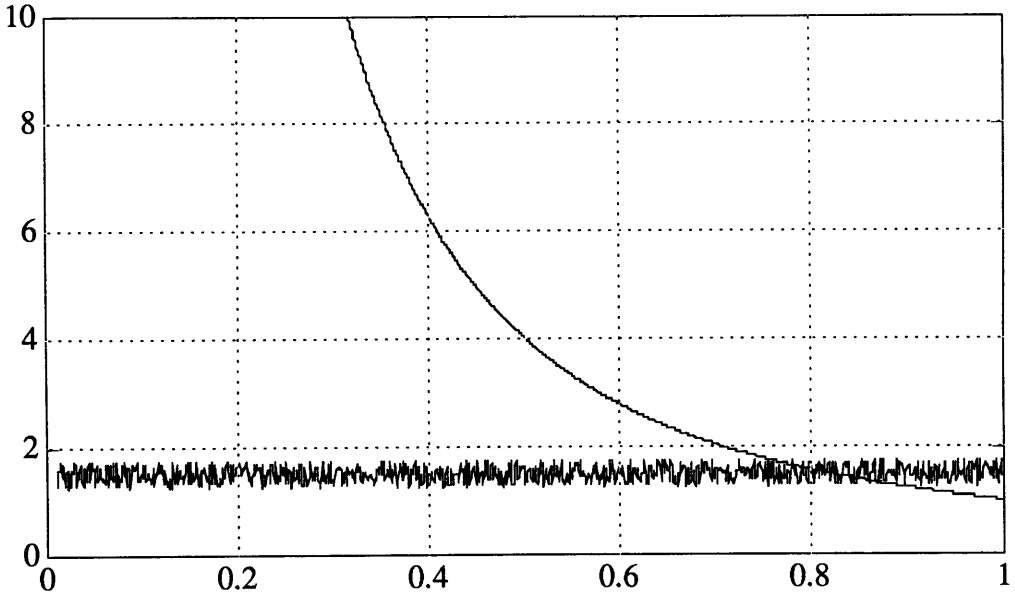

Figure 6.14: Example of signal and noise levels as a function of range. At some range, the signal is about the same as noise, making it difficult to distinguish the two. Noise is shown as partially random, which is from detection and associated electronics.

Summary

The hardware considerations when choosing a wavelength are mainly size, power, and S/N related.

1. The size increases with wavelength, mainly for resolution purposes.
2. The efficiency of all devices is only about 30% maximum, so that dissipation is a factor for high power levels.
3. The S/N is a very important factor, and is very dependent on appropriately matching the emitter (for active systems), detector, and filter, based upon the frequency band of interest and the ambient noise distribution.

The S/N is the main factor in detection of e-m energy. Improvements in detection have many benefits, including the ability to use weak emitters (for active systems) or small collection optics. In addition to the ambient noise, other noise sources influence the systems design in a slightly different manner.

6.1.5. Other Noise Sources

The relatively-constant noise level shown in Figure 6.16 represents all noise contributions. In addition to ambient noise, there are additional noise sources from the detector and its associated electronics.

Detector noise is mainly an issue at very low signal levels. These can be due to the discrete nature of the charge carriers, which yield a small randomly fluctuating current that is superimposed on any bulk charge motion (shot noise) [21]. Because of this, detection is limited by this level, so that shot noise limited detection is the best that nature allows.

Another common detector noise source is thermal induced, since charge agitation is increased in proportion to the absolute temperature. The result is a noise equivalent power that is dependent on temperature and the bandwidth of detection. It is this reason that many IR detectors are required to be cooled. Heterodyne detection is often used to get around this problem [18].

The randomness of these noise sources eliminate the possibility of simply subtracting it. Similarly, pseudo-random noise in the electronics can be introduced from a variety of sources onboard the rover, including e-m fields from the drive motors. The end result is a randomly fluctuating signal that is superimposed on any other collected signals.

The effects of the secondary noise sources has different implications on the design than that of ambient noise. If secondary noise was not present, for a given emitter and detector, the S/N from an object at a fixed distance is constant for virtually any size collection optics. A larger lens, for example, would collect more light, but would also collect more noise by the same proportion. If this were the case, collection optics could be extremely small, based only on this argument. However, with secondary noise a (randomly varying) bias is introduced, which then makes the absolute magnitude of signal and ambient noise important, not just their ratio.

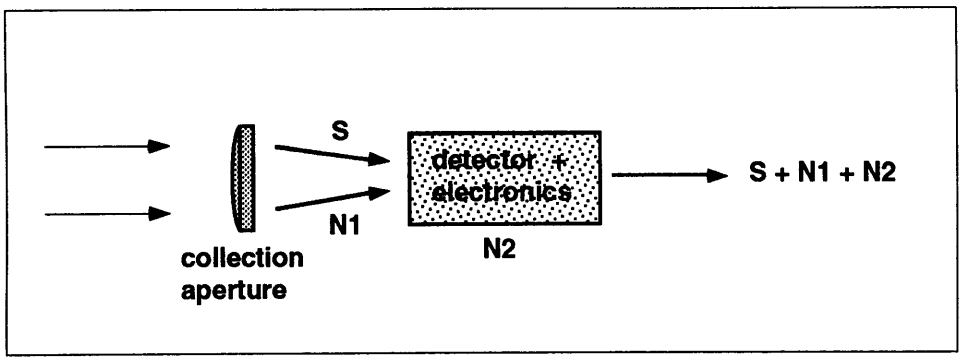

Figure 6.15: Illustration of ambient noise and secondary noise from the detection circuitry.

Hence, because of the presence of secondary noise, it is important to keep the magnitude of the received signal above a certain level. If the signal is too weak, then improvements can be made by:

1. Increasing the collection aperture
2. Increasing the emitter power (for active systems)
3. Reducing the amount of passive filtering to increase the signal magnitude

The last item may seem counterintuitive at first, since reducing the amount of filtering (no longer *matched* filter) will *decrease* the *collected* S/N. However, although the ratio of S to N decreases, their individual magnitudes increase, thereby allowing the final S/N to increase after the secondary noise is added. This is not generally the case, but may be when the signal is weak and the secondary noise is strong.

Conclusion

In summary, it is always beneficial to reduce both ambient noise and secondary noise. However, secondary noise from the detector and electronics is independent of the signal, while ambient noise reduction often results in a *signal* reduction too. If possible, S/N can always be increased by using a stronger emitter. In addition, if secondary noise is large, increasing the collection aperture can help. Finally, if the passive filter is matched to the emitter, increasing the bandpass can help the final S/N by allowing more total signal irradiance to pass through. There are additional filtering techniques other than passive optical filtering, which can help the S/N even further.

6.1.6. Active Filtering

Active filtering is performed in the electronics, opposed to at the collection optics. Instead of attenuating a signal due to material properties of the transmission medium, active filtering takes advantage of additional modulation of the active signal. For example, by modulating at a particular frequency, ambient noise can be rejected since it is typically DC. An active bandpass filter can be designed to pass only the AC component of the collected energy. In fact, by directly monitoring the modulation of the emitted signal, the filtering band can be very narrow indeed, allowing a high S/N. Typically the passed AC signal is then demodulated.

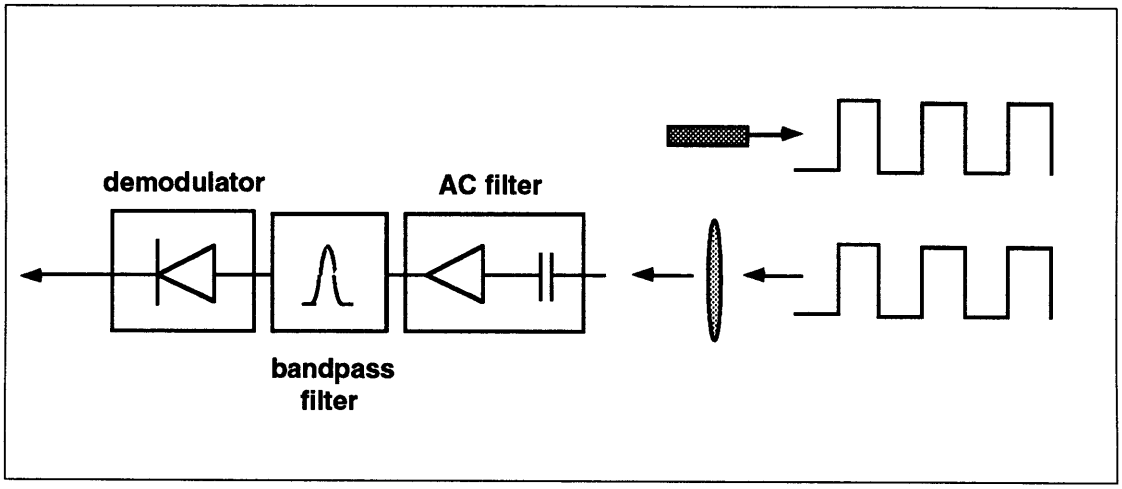

Figure 6.16: Example of active filtering by modulating the illumination source.

This is a very effective method of rejecting ambient noise. In addition, many detectors are more sensitive to AC rather than DC signals. AC detection allows signal shot noise limited operation, yielding much more sensitive systems. Although visible and near IR systems can be shot noise limited without modulation, the advantages are clear in the far IR. There is quite more complexity involved, but the technology is well-developed, and is in common use today even in common household electronics.

In summary, a modulated active source can provide an additional layer of filtering to improve the S/N. AC detection offers the possibility of other ranging benefits too (section 6.2.2), although at a price of complexity.

6.1.7. False Alarms In Active Systems

Because of secondary noise, it is often difficult to distinguish a weak reflected signal from the random noise fluctuations (see Figure 6.16). When the noise bounds are well-

known, it is customary to threshold the receiver output, so that the probability of noise exceeding the threshold is very low. By setting the threshold well above the noise level, the chance of a 'false alarm' (the perception of a returned signal) is low, but many returned signals will be lost below this. However, if the threshold is too low, weak signals can still be detected, but the chance of a false alarm is higher. Hence, the choice of the threshold is critical in detecting weak signals (S/N near 1), while maintaining a low probability of false alarms.

Depending on the use of the ranging system, false alarms can be well-tolerated or devastating to a mobile robot's performance. There are other false alarm protection methods besides thresholding that work well, but these methods depend on the type of ranging system used. In general, any particular ranging system should offer some way of determining if the range value is good or not. Averaging or filtering multiple readings is one way, but this is an inefficient process. The various ways of assigning a level of confidence is mainly dependent on the ranging method. Now that the fundamentals of non-contact sensing have been discussed, the methods of range finding can be introduced next.

6.2. RANGING METHODS

6.2.1. Overview

Having discussed the fundamentals of non-contact electromagnetic wave sensing, the applications of these to a range finder will now be reviewed. There are many different concepts involved in ranging, each with its own particular advantage. Categorization of the different methods can be confusing, but one clear separation is between active and passive systems.

Passive systems have the advantage of not requiring an active illuminating source on board, but require complex detection circuitry and processing. Active systems, the more common types, need to supply their own illuminating source, but detection can be closely matched for good detection of the weak returned signal.

The common robotic ranging techniques are:

Passive
1. Triangulation
2. Dynamic Focus

Active
1. Pulsed Time of Flight (TOF)
2. Continuous Wave (CW)
3. Triangulation
4. Returned Signal Intensity

6.2.2. Passive Ranging Techniques

Triangulation

Triangulation is one of the oldest ranging techniques, and is still actively used today by surveyors. It basically makes use of the law of sines: by knowing two angles of a triangle, and their common side, the remaining sides can be calculated.

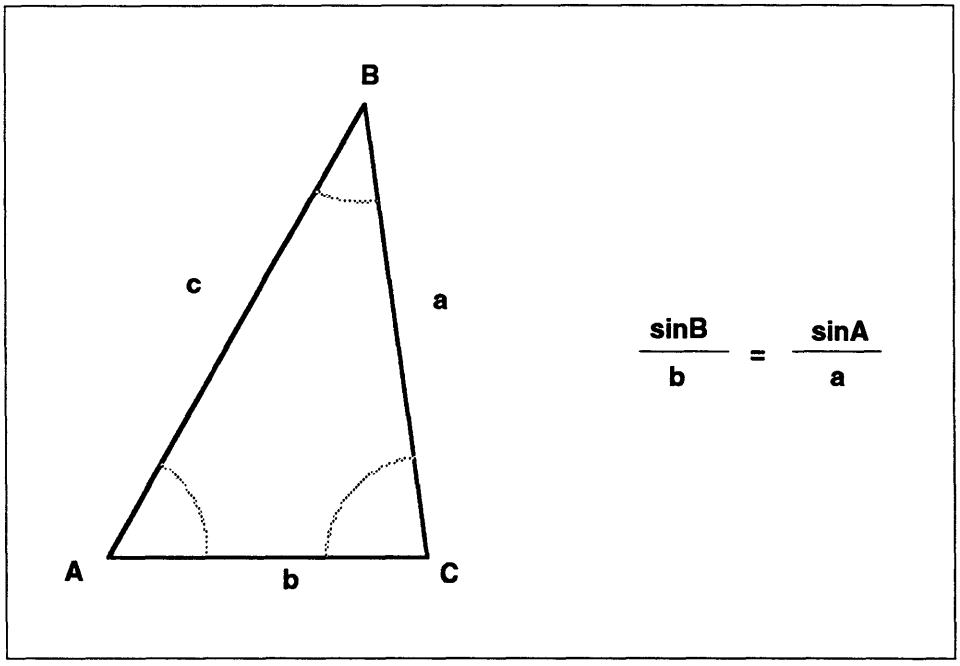

Figure 6.17: Ranging by triangulation uses the law of sines.

For range finding, corners A and C in Figure 6.17 can represent two cameras (or a light source and a detector in the active case). Since the baseline separation, b, is known, range can be calculated by measuring two angles A and B:

$$a = b \frac{\sin A}{\sin B} \tag{6.17}$$

Of course, since B cannot be measured directly, it can be indirectly measured from angles A and C:

$$a = b \frac{\sin A}{\sin (\pi - A - C)} \tag{6.18}$$

Passive triangulation works on the same principles that our eyes do. By looking at the same scene from different positions (each eye is separated by a baseline of about 6.5 cm), the differences in the scene are correlated by the brain and range is obtained.

Similarly, camera images from at least two separate locations can be correlated by a microprocessor. These are known as passive vision systems. Vision systems, whether active or passive, are computationally intensive. For a given scene, an object must be first correlated in each image, so that the object of interest in each image is the same. Hence, the target must have sufficient contrast from its surroundings, and also be unobstructed in each view. The requirements imposed are then

- The ambient illumination must be sufficiently high to view the scene
- The object must be distinguishable visually, as a result of its shape, reflectivity, sharp edges, etc.
- The distance between successive images must be sufficiently small that the chances of obstruction of the scene by another object is low.
- The distance between successive images must be sufficiently large relative to the range that the images are recognizably different, for ranging purposes.

For outdoor systems, ambient illumination is sunlight. Although passive triangulation is possible with other portions of the spectrum, artificial illumination is usually needed, and the resolution degrades at the longer wavelengths. This discussion will thus be limited to

the visible and near IR, where ambient illumination is strong during the day, resolution is high, and detectors are advanced and low cost.

The advancement of the charge coupled device, or CCD, has opened the door to fast and cheap processing of images. The CCD has a wide variety of applications, from the common household video camera, to spectrum analyzers and imaging payloads on satellites. CCD's are very sensitive to light, due to their high quantum efficiency, but are resolution limited due to their discrete nature. For visible light, resolution is limited by the spatial density of the pixels, which although high, is still discrete. Although this does not achieve the spatial resolution of photographic film, it is sufficiently high for image processing. In addition to spatial resolution, the grey level (for black and white systems) for each pixel can be 8 bits or higher, yielding at least 256 levels of brightness. An entire CCD scan takes as little as 5 ms, allowing real time processing. The speed limitation, then, is the processing of the digitized image.

Processing of a single image is not an easy task, as it seems to be for the human brain. It involves recognition based on shadows, edges, and shapes. Once an object is separated from its surroundings, there still lies the task of correlating that object with a second processed image. Only after correlation is performed is it possible to perform algorithms for ranging, based on the known relative locations at the time of imaging.

The different views can be from cameras separated by a given baseline distance, or by a single camera that has moved to a different position such as on a mobile platform. For people with good vision, range can be determined by simply looking with both eyes at an object. If one eye is closed, the ability to detect range is severely impaired. In that case, however, ranging can be restored by moving the head normal to the line of sight. For the same reason, a single camera may suffice if the camera can be moved a known amount between images.

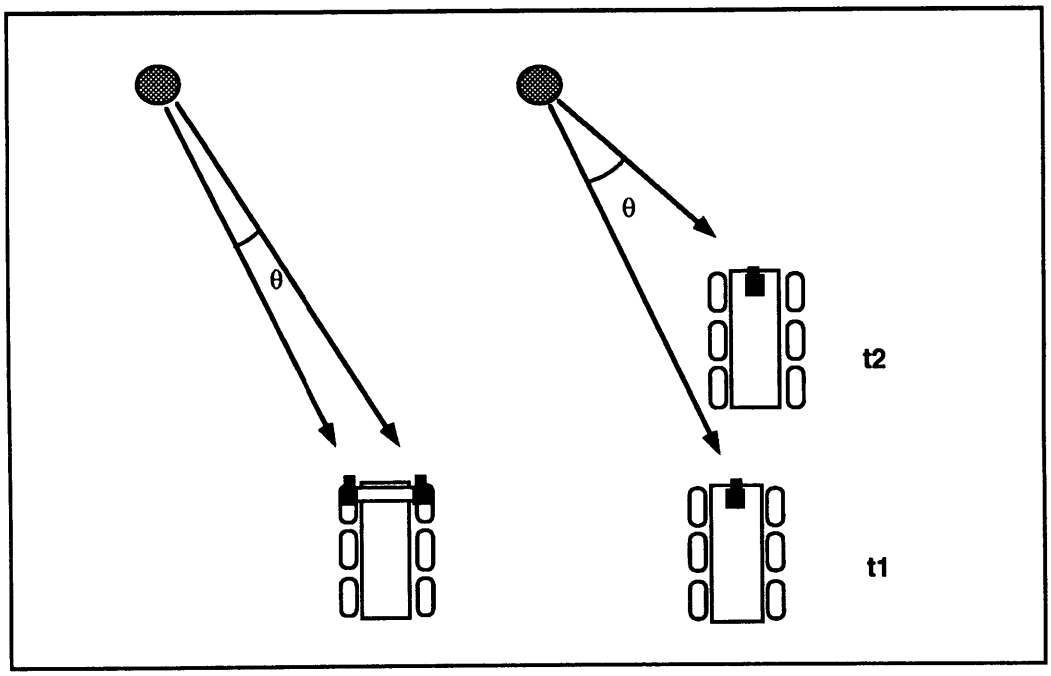

Figure 6.18: Passive triangulation by viewing an object at two different locations. On the left, two cameras are separated by a fixed baseline distance. On the right, a single camera is used, with the mobile platform providing the different camera positions.

If the scene is too uniform, without proper illumination or distinguishing features, image processing is not effective. Even with a sufficient scene, however, processing for recognition and correlation between successive images can be rather difficult. For instance, if both images involve looking at a field of pebbles, the uniformity is high enough that image correlation will be virtually impossible. However, if the field involves rocks that are large relative to the baseline distance, then the uniformity on the scale of the baseline is rather low, and the individual rocks may be distinguishable.

It turns out that the ability to distinguish and range depends on the object size and range relative to the baseline separation. If the baseline is of the order of the size of the pebble field, then each pebble would be distinguishable. Similarly, the ability to extract range information from two images of the same object is possible only if the range is within a few orders of magnitude of the baseline separation. For instance, a 6.5 cm baseline for eyes is not sufficient to determine the range to an object that is 1 km away, but is very effective at ranging at distances of the order of meters. When $B/R << 1$, the view is essentially the same for each picture, and ranging is more difficult.

Hence, a large baseline separation is beneficial for image correlation and ranging. However, there are usually packaging limitations that do not allow large baselines. More importantly, though, is the 'missing parts' problem [23] that can occur for large baselines, so that the object of interest is blocked in one of the two frames, as shown in Figure 6.19. For these reasons, the baseline is usually kept fairly small, and so the range accuracy degrades at distances that are much larger than the baseline separation.

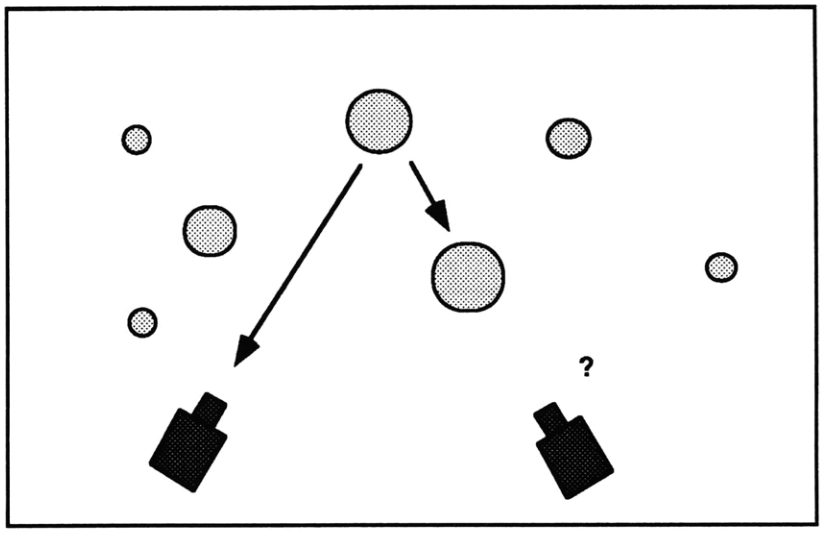

Figure 6.19: Obstruction of the scene in triangulation. This is often referred to as the missing parts problem.

It should be noted that all triangulation-based ranging systems have accuracies that degrade with distance, limited by the baseline separation and the resolution of the captured illumination. Because of this, triangulation systems are mainly effective at close range.

Passive triangulation or vision systems are limited mainly by the effectiveness and speed of the image processing. Although the idea is simple, the reality of image processing for correlation and ranging can be very difficult, and requires a significant amount of computation. When computing resources are limited, vision systems can be sluggish. However, if high speed is not important, vision systems may suffice.

Dynamic Focus

This ranging technique, although not well-known, is unknowingly used by more people than any other ranger. The reason is that the majority of today's autofocus video and still cameras use this technology.

Although it is considered a passive technique, dynamic focus requires mechanical motion, opposed to an artificial light source as with 'active' systems. By changing the distance between the lens and image plane, the image will sweep in and out of focus. Then by monitoring the spatial frequency change across the image at each lens position, the in-focus position can be determined. This method actually directly yields the best focus position, which makes it ideally suited for autofocusing. Range, however, can then be indirectly calculated from knowledge of the lens parameters, typically through the use of a lookup table.

Measuring the spatial frequency variation requires a rather simple form of video signal processing. For each lens-detector distance, the grey level of each pixel on the CCD (assume black and white CCD) is measured across a certain row. To reduce the amount of processing, only the high-frequency portion of the row scan is stored. Thus out-of-focus objects do not contribute much to this. The lens-detector distance is then changed a small amount, and another scan is performed. This is continued over hundreds of positions, storing the filtered data in each case. The strongest contribution of high frequency response is representative of the in-focus position [23].

Like passive vision systems, dynamic focusing requires some visual contrast in the scene to function properly. If the scene is a uniformly painted wall, the in-focus image will appear the same as the out-of-focus image, so there will be no spatial frequency variation as the lens-detector distance is varied. However, if the object has distinguishable visual features such as noticeable edges, shadowing, or different reflectivities of the ambient light, then the spatial frequency change will be the highest when the image is in focus. When looking at an object, the distinguishable features form the foundation that makes dynamic focusing possible.

Honeywell Corporation first developed this technology, but the Japanese camera companies have manufactured the subsystems at apparently low cost. Unfortunately, at this time there are many infringement law suits taking place by Honeywell, and companies have not been cooperative in publishing information about these products. In

fact, there is not a *ranging* product available that uses the dynamic focusing concept, although the cameras that use this technology are readily available.

Dynamic focusing has the advantage of being able to have a large instantaneous field of view, based on the lens chosen. Hence, after a given sweep, every object within the field of view can be analyzed, and range can be extracted for multiple targets. The range accuracy and the ability to distinguish closely-spaced objects is limited by the detector resolution and the lens depth of field. Better resolution can be obtained by using a longer focal length lens or a larger aperture, both of which yield a shorter depth of field. However, this is at the cost of a smaller field of view and larger size respectively. Like passive triangulation, dynamic focusing requires sufficient illumination and visually distinguishable features, but is not impaired by the missing parts problem of triangulation. Also, it does not require a large degree of mechanical scanning in azimuth, due to the relatively large instantaneous field of view. However, it should be stressed that depth scanning is still necessary, since the lens-detector distance needs to change. The amount of motion, however, is very small, and can be performed with joint-free transducers such as piezoelectrics.

6.2.3. Active Ranging Techniques

Pulsed Time of Flight

TOF systems work on a very basic principle: the distance to an object is proportional to the elapsed time from signal emission to echo return. Since the propagation speed, v in a particular medium is known, the range is simply

$$R = 0.5\, v\, \Delta t \tag{6.19}$$

where Δt is the round trip travel time. Most sonars and radars use this principle as the basic means of ranging. Of course, the detection electronics are vastly different for each, since sound travels at about 1ft per millisecond (in room temperature air), while an e-m wave travels at about 1 ft per nanosecond; a million times faster.

Therefore, detecting e-m signals requires very fast electronics for reasonable distances. This is not much of a problem for radars that range out to many kilometers, but is difficult for short range applications.

A simple radar system is a good example of how TOF systems work. At a certain time a short pulse is emitted. At that instant, a timer is started as the pulse travels to the object and back. At a certain time Δt later, an echo is received (that is many orders of magnitude weaker than the emitted signal) and the counting stops. The elapsed time is proportional to the object range, as shown in equation 6.19. This process is illustrated in Figure 6.20.

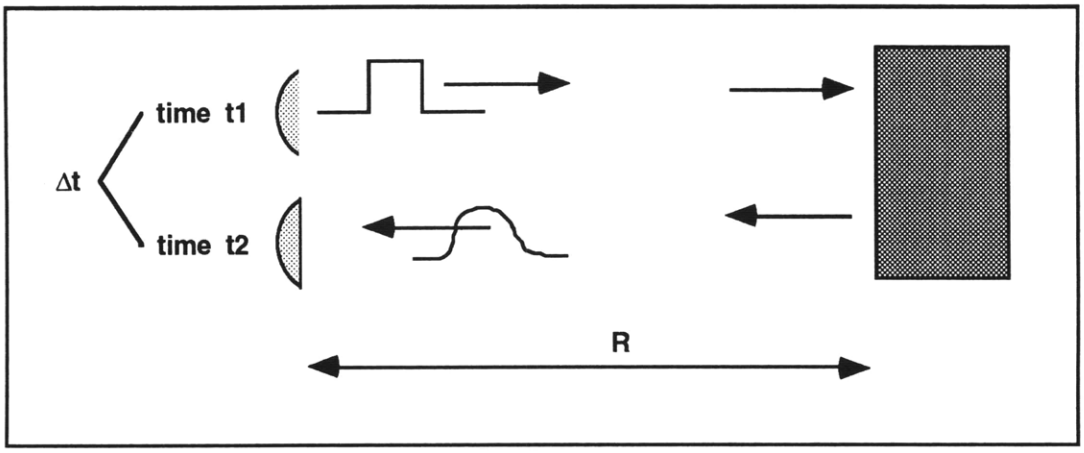

Figure 6.20: Time-of-flight range finding

The emitted pulse has a rather large amplitude, a pulse 'width' (time), a carrier frequency, and sometimes modulation within each pulse. It is desirable to have a large amplitude and a short pulse width, so that the returned signal is strong and the resolution is high. Obviously there are power limits on the amplitude, but the pulse width can be controlled easily down to a certain time. The amount of energy contained in the pulse increases with the pulse width, but the resolution in range decreases in most cases.

Pulse width is important for different reasons. For radar systems, where the beam can be wide enough to cover many targets, the length of an unmodulated pulse determines the ability to distinguish the different ranges present. If the pulse is too long, the separate echoes will blend together as one. Quantitatively, the range separation must be greater than half of the pulse width in terms of physical length.

In most short-range cases though, the object is larger than the beam, and so discrimination of different ranges is not applicable. In this case, a short pulse width is desirable, but not critical. On receiving, the leading edge of the echo is the key parameter in range accuracy. Hence, the pulse *shape* is really what is important. The leading edge

especially should rise sharply, so that leading edge of the echo is also steep, resulting in a distinct location of each for timing purposes.

The echo is generally of the same form, but usually a bit deformed due to broadening, especially for long ranges. An echo can be filtered from the background noise not only by thresholding, but also by other clever techniques. For instance, if the carrier frequency is constant, a bandpass filter can be designed to just pass this frequency. Additional modulation can occur within each pulse that can also offer another level of filtering. The pulse width itself, when transformed to the frequency domain, can additionally offer the possibility of a filter that is 'matched' to capture most of the energy within the first peak. Also, the strength of the echo is generally a known function of the elapsed time, since range is proportional to time. Hence the gain can be appropriately adjusted with time to match the expected echo strength, and echos whose strength deviates substantially from the expected can be ignored.

For radar systems, the pulsed TOF technique is particularly suitable, since the transmitter and receiver can be the same. Hence, detection of weak signals is easier since there is not a simultaneous emission [20]. There is a large transient during and shortly after the firing, so detection is usually 'blanked' for a short time until the transient ringing has decayed. The same idea applies to separate emitter/receiver pairs, although in this case the noise transient is not as critical.

Visible and IR systems do not have the luxury of varying the carrier frequency, since it is essentially fixed by the laser or LED. However, modulation within each pulse can still be easily controlled. In these systems, the detector is usually an avalanche photodiode, which has an inherent ability to amplify signals. These 'Lidar' systems are capable of very fast operation, since the distances are short compared to that of radars (range is limited due to the larger attenuation of light). However, all of the pulsed TOF systems have difficulties at close range due to the timing requirements. Objects within a few feet require sub-nanosecond timing accuracies, and detection that needs to occur only a few nanoseconds after the pulse. However, close range pulsed TOF systems have more relaxed requirements on detection sensitivity. Even so, pulsed TOF systems are not well suited to very short-range applications, but are better suited to medium and long-range.

Continuous Wave Phase Shift

Another method that is similar to the pulsed TOF is continuous wave, or CW, phase shift. Instead of firing a short distinct pulse, a continuous modulated signal is emitted. Rather than timing directly, range is calculated by simultaneously measuring the phase differences between the emitted and detected signal. Hence, CW phase shift methods are also technically time-of-flight [23], but the timing is indirectly measured, so that the actual speed of light is not a limitation at close ranges.

Like radar, the elapsed time of flight is related to range by

$$\Delta t = \frac{2R}{c} \tag{6.20}$$

where c is the speed of light, and R is the range. Figure 6.21 illustrates the transformation between the time and frequency domains, which is the basis for CW phase shift range finding. The elapsed time can be expressed in terms of phase difference of the emitted and detected signals, normalized by the period of the wave:

$$\frac{\Delta t}{T} = \frac{\Delta\phi}{2\pi} \tag{6.21}$$

where T is the wave period, $\Delta\phi$ is the phase shift, and f is the modulation frequency. This can be rewritten in terms of the modulation frequency:

$$\Delta t = \frac{1}{f}\frac{\Delta\phi}{2\pi} \tag{6.22}$$

Combining equations 6.20 and 6.22, the range is expressed as:

$$R = \frac{c}{2}\frac{1}{f}\frac{\Delta\phi}{2\pi} = \frac{\lambda}{2}\frac{\Delta\phi}{2\pi} \tag{6.23}$$

It should be clear that by modulating the emitted signal at a known frequency f, the measured $\Delta\phi$ can yield range.

For instance, if the active signal is a 10 MHz sinusoid, one period of the wave is 100 ns

At light speed, this time of one wave cycle represents a length of 30 m. Therefore, if an object is 15 meters away, the reflected wave (which has traveled a round trip distance of 30 m) will be out of phase of the emitted wave by one complete cycle, or 2π radians. Now if the object is brought in closer, the phase shift will decrease linearly with distance, theoretically to zero. Hence by simply measuring the phase difference, range can be determined easily.

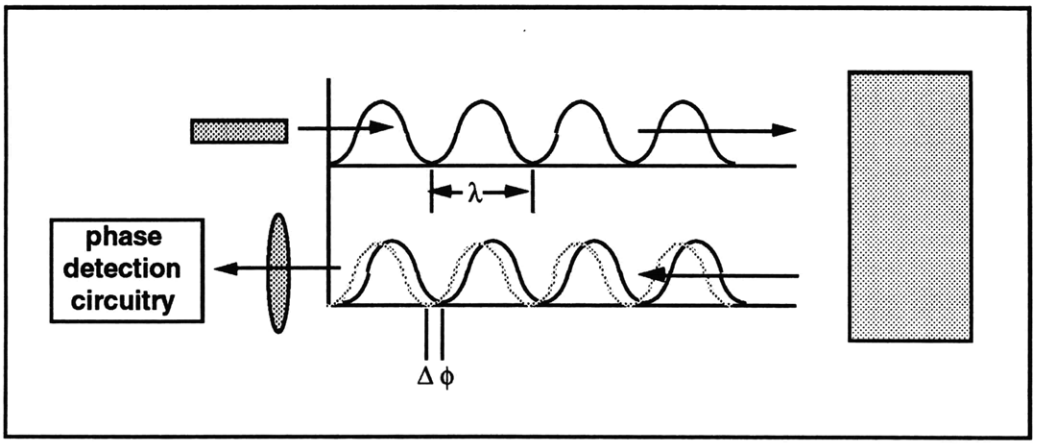

Figure 6.21: CW phase shift range finding.

One of the main problems with the phase shift method is ambiguity in the calculated range, since ranges that differ by an integral multiple of the modulated half-wavelength will yield the same phase difference. For instance, in the previous example, an object at 15 m, 30 m, 45 m, etc will each show a phase difference of 2π. $\lambda/2$ is thus referred to as the ambiguity interval [18]. To circumvent this problem, multiple frequencies can be superimposed, such that ambiguity cannot occur simultaneously in each. Or, modulation can be changed actively to yield the same result. Also, depending on range and accuracy requirements, the modulating frequency should be chosen such that the maximum range needed corresponds to 2π, so that intermediate distances cannot be ambiguous. However, longer distances may still yield the same problem, and the resolution can be degraded.

Nevertheless, CW systems offer some important advantages over TOF rangers. First, their continuous operation makes very fast operation possible, without the need to fire and wait. Second, although the detectors need to be very fast, there is not the need to measure time absolutely, but only in terms of *relative* phase differences. This lessens the requirements on the electronics somewhat, and allows practical close range detection down to centimeters. Also, filtering and detection is improved due to the AC nature, yielding the potential of much higher sensitivity and better S/N.

Of course, measuring the phase shift does require somewhat complicated electronics, even though the timing requirements are not as tight as TOF systems. Calculating range from the measured signals is not as simple as TOF systems, and ambiguity in range can be a problem if precautions are not taken. However, the continuous nature and higher sensitivity makes CW phase shift systems a good alternative to TOF systems. Using CW has a particular advantage in the ability to choose a modulation frequency that best suits the range needed, without the limitation of actual timing. This is especially advantageous at close range, where the speed of light makes timing in TOF systems difficult.

Triangulation

Unlike passive triangulation, active triangulation employs an illumination source at a known distance away from the detector, directed at a known relative angle. It thus does not require two 'pictures' at different positions, since the geometry of the illumination source relative to the detector is known.

Triangulation works because the known geometry of the emitter and receiver (active), or multiple receivers (passive). In each case by calculating the angle of the reflected light to the receiver, range can eventually be extracted. Active systems project a rather simple geometrical pattern, and in some cases their interception with an object can yield more complex patterns that are interpreted for range.

Spot Projection

The simplest case is the projection of a small collimated beam, such that its interception with an object yields a point source of reflected energy. Using a visible laser as an example, the result is a small light spot on a wall, say. Assuming a diffuse reflection, the spot is visible from anywhere with a direct line of sight to the laser spot. Now, if an angle detector is placed at a known baseline distance away from the laser, range can be determined simply from the baseline distance and the incident angle of the reflected radiation. Although light is reflected in many directions, only the portion incident on the receiver aperture is what is detected. Figure 6.22 shows a special geometry, known as the right triangle configuration. In this case, one of the two baseline angles is 90°, so that range can be expressed as:

$$R = \frac{B}{\tan \theta} \tag{6.25}$$

where θ is the incident angle of the reflected light.

In reality, it is not *angle* that is directly measured, but *displacement* of the focused energy at either the image or focal plane of the receiving lens. When θ is small, under 12°, the focal point is displaced in a plane normal to the lens axis, such that the angle on either side of the lens is preserved (see Figure 6.22). Because of this, there exists a case of similar triangles, so that

$$\frac{R}{B} = \frac{f}{d}$$

$$R = \frac{Bf}{d} \tag{6.26}$$

where f is lens focal length, and d is the linear displacement of the focal point. This applies to the parallel optical axis configuration as shown. A similar case holds with imaging systems.

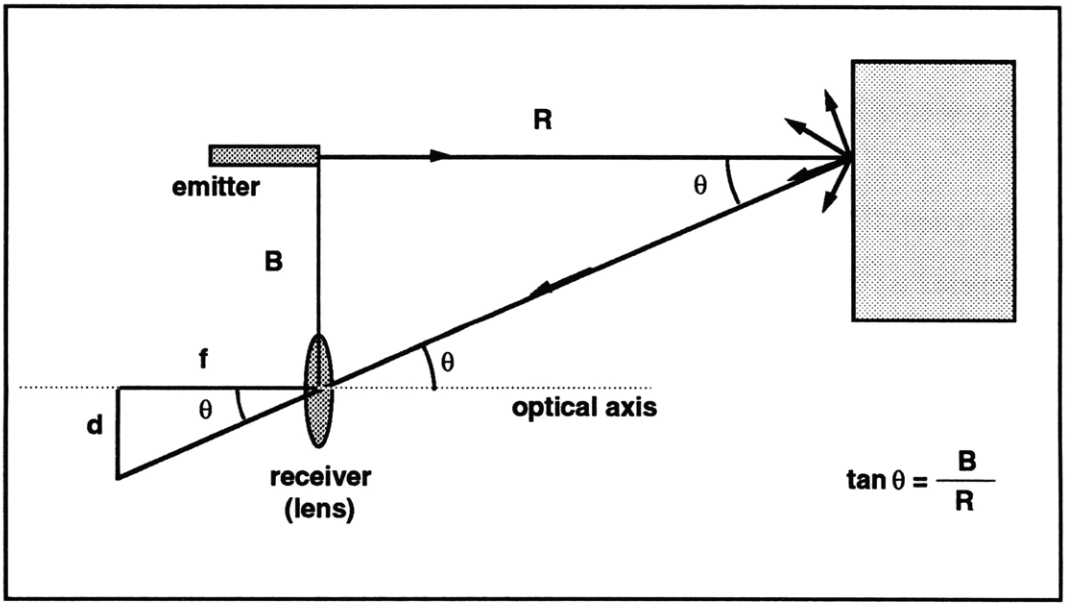

Figure 6.22: Active triangulation, right triangle configuration.

Equation 6.26 offers easy calculation of range, if the focal point displacement, d, is measured. With the advent of high resolution arrays like CCD's, and continuous

displacement sensors such as position sensitive detectors (PSD's), displacement can be measured easily and quickly, even at very low irradiance.

Although simple, this form of active triangulation still suffers from the 'missing parts' problem, which can occur if another object blocks the line of sight to the detector, similar to the passive case. This forces the baseline distance between the emitter and detector to be small. However, when B is small, the sensitivity is decreased, and both resolution and accuracy are sacrificed. In fact, another inherent property of all triangulation systems is that *accuracy significantly decreases with distance*, unlike TOF systems, where accuracy is somewhat independent of distance. Triangulation systems are much better suited, then, to close range applications.

To illustrate the decreased sensitivity with range, examine equation 6.26 a bit closer. Since range is a function of the *inverse* of d, there is a nonlinear relationship between range and displacement. Figure 6.23 shows a plot of range vs focal point displacement for the right triangle configuration. The minimum range is set by the sensor size, while there is no theoretical limit on the maximum range. However, note that the slope of the curve becomes steep as R becomes large. Hence, at large distances, small changes in d yield very large changes is R. For high accuracy, it is desired to have just the opposite. However, this latter condition then has a limited range due to the limited sensor area, so that there is a definite tradeoff between range and accuracy. For a given minimum range, triangulation systems lose accuracy and resolution at large distances due to the steep slope as shown in Figure 6.23. From these considerations, one can see the importance of position sensing accuracy and resolution, since any errors in the detection of the light spot position can yield large range errors at large distances.

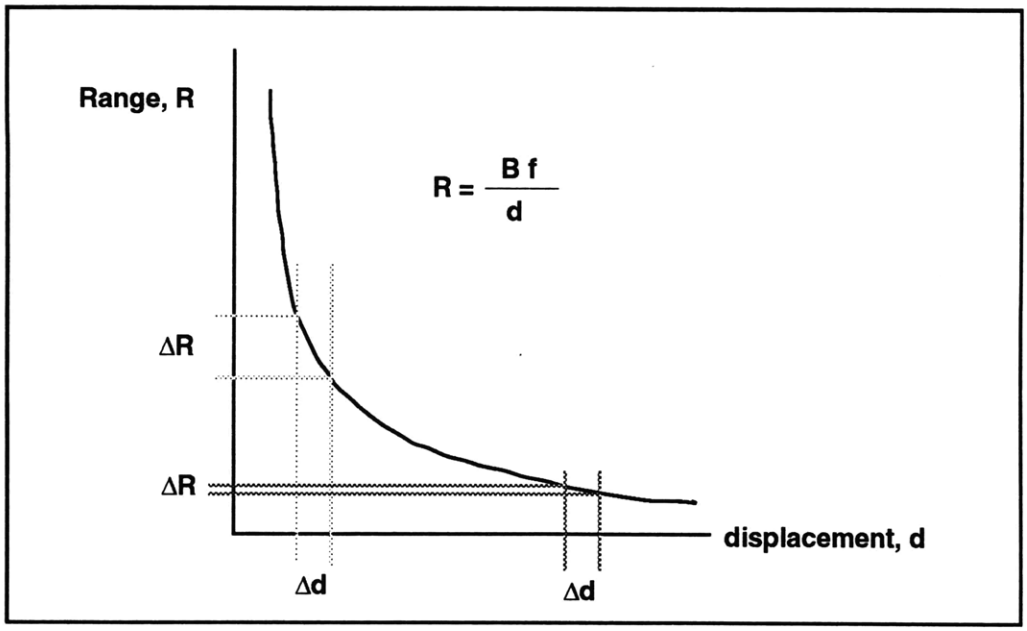

Figure 6.23: Range vs focus spot displacement (see also Fig 6.22). For a given displacement error Δd, the corresponding range error is shown.

To detect a spot accurately, the spot should be small so that the reflected radiation is essentially a planar wavefront (parallel rays) across the receiver. This allows the captured energy to be focused to a small 'point' at the position sensor, which is more accurate. If the spot size or collection aperture is too large for a given range, then the focused energy is blurred, causing a loss of accuracy. If the spot cannot be small, it should at least have an irradiance pattern that is symmetrical along the sensor axis.

Active triangulation systems of these types have relatively simple electronics. All that is needed is the detection of the light spot position. The active source may be pulsed or continuous, depending on the sensor used and the sensitivity needed. Continuous wave systems typically use a modulated laser source, providing the ability to tightly filter by passing only light of this frequency, rejecting the DC background light. Pulsed systems can compare the detected signal *before* and *during* the pulse [24] to determine the contribution from the active source only. In each case, the illumination source generally needs to be higher power (for a given range) than for other detection methods, since the position sensors are generally not as sensitive as, say, an avalanche photodiode.

In conclusion, this simple form of active triangulation, using a uniform collimated beam, can offer a limited detection range and good accuracy at close range. The associated electronics are also simple, as well as the algorithms for range calculation. The speed can

be very high, especially when used with PSD's opposed to discrete arrays for position sensing. The main drawback of active triangulation is the loss of accuracy at far distances. There exists a well-defined trade between the range interval and accuracy. Also, for large baseline distances between the emitter and detector, there is the usual 'missing parts' problem. However, for applications that can tolerate lower accuracies, a small baseline distance can virtually eliminate this problem.

Structured Lighting

This is a more advanced form of triangulation, requiring much more complex processing of the data. Instead of projecting a collimated beam and using a one-dimensional detector, a more complex illumination is projected, and detection is performed with a vision system in two dimensions.

The illumination can be of many types, such as vertical lines (from a planar source), horizontal lines, a grid consisting of both, dot patterns, etc. For any pattern, the same principles of triangulation apply as in the simple case of a single collimated beam. Knowledge of the illumination geometry and the source's position relative to the detector are the only key parameters needed for ranging.

This method, although more complex, is still simpler than the *passive* triangulation technique, which requires image correlation under the ambient lighting. Although the field of view of the receiver (camera) may be large, the only portion of the scene that is actually ranged is where the projected light intersects the targets. For a tightly spaced grid, the processed data can be interpolated with a high degree of confidence as a representation of the environment. However, this requires large amounts of computation, and begins to approach what is needed for passive triangulation.

A less complex illumination pattern can still yield a significant amount of range data, while still requiring only moderate amounts of computation. For instance, a single plane of projected light is relatively easy to analyze, relative to a passive system. Figure 6.24 shows the same reflected 'line' from the view of a camera positioned a short distance from the plane of laser light. The intersecting surfaces are random objects on the ground, and the ground itself. The reflected pattern starts to take on a complex shape. Note that the camera image simply sees a two-dimensional representation. In addition to range, relative ground slopes can be determined from the slope of the lines in the image. The

baseline separation essentially determines the scaling factor between the two domains (3D and 2D).

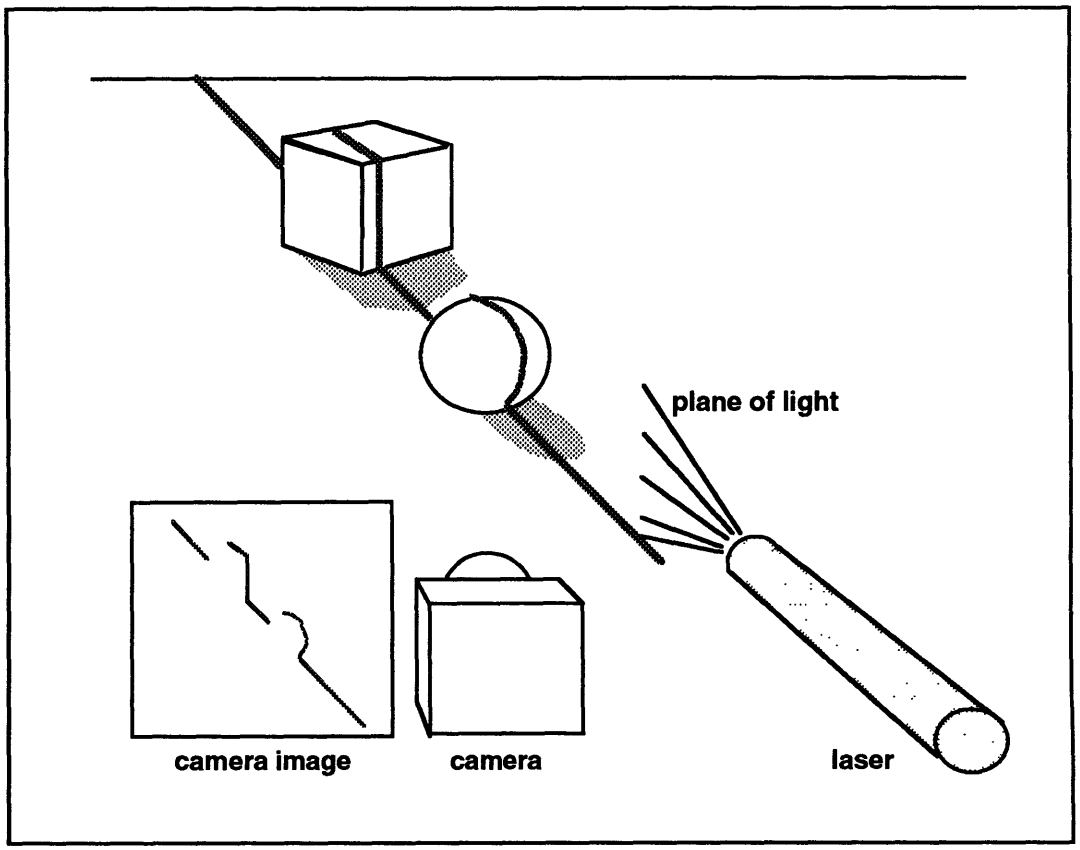

Figure 6.24: Active triangulation through structured lighting.

Detection of a structured light pattern requires a two-dimensional array of some sort. CCDs are the common choice for many reasons, especially due to their high sensitivity and good resolution. A two dimensional PSD will not suffice, since the output will yield only an effective 'centroid' spot location. However, one interesting method for line projections use a stack of one-dimensional PSDs, forming an effective 2 dimensional array [25]. Although this produces low resolution in one axis (the stacked axis), the speed and resolution in the other axis is greatly improved due to the nature of the PSD (give reference). The speed of operation has been shown to be hundreds of scenes per second continuously. CCDs operate more slowly and in a discrete frame by frame nature.

Returned Signal Intensity

This is the simplest of all ranging techniques, but is a very indirect way of measuring range. Returned signal intensity (RTI) rangers operate on the principle that the irradiance

incident on the receiver aperture is only a function of the range. It thus assumes a constant reflectivity, purely diffuse reflection, and no changes in attenuation. The simplified equation for the returned irradiance from a narrow active source is (from section 6.1.3):

$$I = \frac{K_{refl} P_e}{2\pi R^2} \tag{6.10}$$

where K_{refl} is the surface reflectivity, and P_e is the active illumination power. The irradiance, as expected, follows the inverse square law. This equation can be rearranged, so that range is

$$R = \sqrt{\frac{K_{refl} P_e}{2\pi I}} \tag{6.27}$$

In this equation, it is then assumed that the reflectivity and emitted power are constant, so that range can be calculated simply by measuring the irradiance. However, because some parameters can change, the accuracy of this system is not good at longer distances. To illustrate this, equation 6.27 can first be rewritten for clarity, lumping all of the 'constants' into one constant, K:

$$R = \sqrt{\frac{K}{I}} \tag{6.28}$$

To examine the sensitivity of range to variances in K, the first partial derivative is taken:

$$\begin{aligned} \frac{\partial R}{\partial K} &= \frac{\partial}{\partial K} \sqrt{\frac{K}{I}} \\ &= \frac{1}{2} \sqrt{\frac{1}{I K}} \end{aligned} \tag{6.29}$$

Substituting equation 6.28, this becomes:

$$\frac{\partial R}{\partial K} = \frac{1}{2} \frac{R}{K} \tag{6.30}$$

Therefore, the range error is

$$\Delta R = \frac{1}{2} \frac{\Delta K}{K} R \tag{6.31}$$

From 6.31, for a given fractional change in K, the range error is proportional to the range itself. K can represent not only the reflectivity and emitted power, but also any attenuation, from dirty optics for example. Thus, K cannot be controlled tightly. So for a given fractional change in K, it is important to express the resulting range error, which is what equation 6.31 does. From this equation, it should be clear that returned signal intensity range finders are not good for long range applications.

Because of the severe performance degradation at long range, return signal intensity rangers have only found adequate use for short range applications. Even close up, range accuracies can still be poor, though. However, these systems have still found widespread use as proximity sensors, when absolute ranging is not necessary. Since proximity sensors only have to detect the *presence* of an object within a certain short range, they can function very well in this capacity by thresholding the returned signal strength above a certain value. This can work particularly well if the system is adjusted so that the threshold falls near the knee of the curve. In addition, if the object reflectivity can be controlled, actual ranging is possible.

Unfortunately, the outdoor environment cannot offer the luxury of highly controlled parameters. Therefore, the use of these types of proximity sensors have been limited to indoor environments with objects of similar reflectivities. Product manufacturing makes extensive use of these types of sensors, which is where the majority of the proximity sensor market is. The detection process for most of these sensors is improved not by using expensive optical filters, but by modulating the emitted light and passing only this frequency band (see 6.1.5). Background noise is virtually eliminated, and the resulting AC signal is demodulated and thresholded. The entire sensor, including all of the electronics, can be low-cost and miniaturized.

In summary, the returned signal intensity method is practical only at close ranges, and works best when the environment is highly controlled. Although these types of sensors have found widespread use as short range proximity sensors, they have not been useful as rangers, especially at medium to long distances, or in unstructured environments. Any variation in the surface reflectivity, illumination intensity, or general attenuation can introduce significant errors, more so at long range. Although all of the other active systems discussed require measurement of the returned signal strength, none of them require the use of this parameter in an *absolute* sense, making other ranging methods

much more insensitive to changes that can affect the signal strength. However, their simplicity and small packaging makes RTI systems attractive for some short-range applications.

6.2.4. Choice of Sensor

In choosing the ranger to use, it is important to recognize the requirements of it, based on the rover's system and hazard-avoidance requirements. The primary issues are:

- Short range

  The requirements in range are very different from most applications, with 3 meters (10 feet) being the maximum, as determined from simulations (Malafeew) and prototype testing. Hence, the autonomous hazard avoidance acts very locally. Also important is the minimum range, which is about 15 cm (0.5 feet). The rangefinder must be able to accommodate this short absolute range, along with the large dynamic range of the return signal.

- 180 Degree Azimuthal Coverage

  The range finder must be able to cover a wide field in front of the rover, at least 180°. However, the angular resolution must be sufficiently narrow (< 5°) to meet the accuracy requirements. It may then be necessary to *scan* a range finder with a narrow instantaneous field of view.

- Moderate accuracy

  It is not important to have millimeter accuracy, but more importantly to cover the minimum and maximum distance needs. Accuracy is more important at closer distances, but not as important near 10 feet, due to the presence of dead reckoning errors. Hence, the percentage error is an appropriate definition, which is acceptable within about 6.5 %, as explained in Ch 7.

- Moderate Speed

  Although 'fast' is very relative, many rangers are limited to having to provide range data at high rates of hundreds of kHz to even MHz. One of the advantages of the slow-moving rover is the relaxed need for data rates, so that rates of 10 to 50 Hz are acceptable.

- Robustness

  Due to the harsh environments of other planets, the system should be as simple as possible, to minimize the chance of a failure. Failures need to not only be identified, but the rover should provide a way to function when a certain amount of failures occur. It would also be beneficial if the validity of a data point can be checked, since noisy measurements can be detrimental to a control system.

- Small Size, Mass, and Power

  The ranging system needs to provide the angular range necessary, yet still remain compact and light, and still allow a certain degree of redundancy. The power requirements also must be low, due to both the limited energy on board, and because of thermal issues.

Table 6.2 shows many of the available range finders, all of which use ranging methods which were discussed earlier. There are actually very few that come close to meeting the requirements of the microrover, mainly due to their size, power, etc. This is especially true for the scanning range finders, which are typically used for much larger distances, and at high data rates. Autofocus camera technology has led to the most appropriate ranging systems, due to their need for small size and low power consumption. Unfortunately, the state-of-the-art autofocus technology, dynamic focus, is not commercially available in the form of a range finder. One of the reasons for this is that the technology is rather new, although it is used in most high-quality autofocus cameras today, including video cameras.

Range Finder	Ranging Method	Range Interval (m)	Resolution/ Accuracy (m)	Sampling Rate (Hz)	Illumination	Power Supply	Comments
Hamamatsu Range Finder Chip Set	Active Triang.	4 max	continuous	700	IR LED 880 nm	3 V	Optics not provided
ESP ORS-1	Phase Shift	0.5 to 6	0.03 res 0.15 accur	-	IR LED 820 nm	± 15 V 0.4A	1 inch beam
Schwartz Electro Optics LRF 200	Pulsed TOF	1 to 200	0.3 accur.	25,000	Laser Diode (20 W)	8-24 V 5 W	Sensitive to noise
Odetics Scanning Laser Imaging System	Phase Shift	0.5 to 9.0	0.02 accur.	-	Laser Diode 820 nm	28 V 2 A	60° x 60° scanned field, 28 lbs
Cubic Precision Red Eye	Phase Shift	0.12 to 30	-	5	Laser Diode 3 mW	7.2 V 0.15 A	For human interface
Pentax Auto Focus Sensor	Active Triang.	0.75 to 4	8 steps res.	10	IR LED	3.5 V 0.05 A	-
Honeywell HVS-300 3-zone Sensor	Active Triang.	0.06 to 0.75	3 steps	100	LED	-	-
AM Sensors MSM 10500	Phase Shift	15 max	0.15 accur.	30	microwave	10 to 28 V	large beam width
Chesapeake Laser Systems	Active Triang.	0.3 to 1.0	0.001	1000	Laser Diode 30 mW	-	4 lbs
Nomadic Sensus 500 Vision System	Structured Lighting (Triang.)	0.5 to 3.0	0.12 res worst case	30 (frame rate)	Laser Diode 780 nm	12 V 1 A	-

Table 6.2: Commercially Available Range Finders

As could be expected, most of the rangers in Table 6.2 that come close to satisfying the microrover requirements are based on active triangulation methods, due to their simplicity and size. Phase shift methods are also a consideration, offering a higher accuracy towards the longer ranges. AM Sensors make the only long-wavelength range finder, but the result is a very large beamwidth (100°). Cubic Precision offers a ranging system for surveying, and is thus large, including an LCD display and a keypad. Although not suitable for the rover, the 'Red Eye' demonstrates the effectiveness of phase shift methods with a low power light source. The ESP scanning rangefinder was developed for (larger) mobile robots for indoor use, and thus has an expanded eye-safe beam. The entire unit is much too large, however, and the frequency is tuned for the longer ranges. Unlike pulsed TOF systems, phase shift detection methods are not limited by the high speed of light, although they are also considered to be TOF. In fact, if the modulation is increased to about 50 MHz, phase shift methods can yield mm accuracy in

the required microrover range interval. Unfortunately, there is not a phase-shift range finder that provides this.

The leading triangulation systems above are the Hamamatsu, Pentax, and Chesapeake range finders. The Chesapeake device is designed for high accuracy, and thus does not satisfy the range requirements. In addition, the weight is large. Pentax makes a small unit that has low power consumption, but the minimum range is too large, at 0.75 m. The best choice is the Hamamatsu range finder chip set.

For about $350, the Hamamatsu range finder evaluation board (H3559) includes the integrated circuit (H2476), supporting electronics, a high power IR LED, and a small position sensitive detector (PSD) [24]. The user must design the optics for this system. When triggered, ambient light contributions are stored by memory capacitors, after which the LED fires. After differencing, only the pulsed PSD current values are amplified. Logarithmic compression then occurs, yielding compressed voltage representations of the two PSD anode currents. These values are differenced, and compensated for temperature shift. The output is held by the sample-and-hold circuit as analog range finding output, which is nonlinear with range. Further manipulation includes evaluation by the A/D converter, as to which of the 16 zones it belongs. In addition, if the PSD signal is too weak, then the last zone is automatically assumed, by the infinite judgment circuit. All of these steps are performed on the H2476 IC.

Two output formats are available: analog voltage, and a 16 step digital signal. Due to the weak return signal, detection is difficult at ranges above about 3 m, or when the ambient light is large. However, the user is confined to pulsed DC operation, which is less sensitive than AC methods. Another problem is the collimation of the LED, which is typical when the collimating lens size is limited. The result is a large beamwidth, compared to that for lasers. Although it may be possible to replace the LED with a laser diode, this was considered dangerous practice because of the unknown internal circuitry. In addition to these problems, the H3559 did not allow the use of multiplexing other PSDs through the same electronics, since all the amplification stages were performed on the IC. Hence, any multiplexing would involve the very low (nA) current levels, which are very subject to noise. In order to multiplex, which is a possible future plan, multiple boards would be needed, which is inefficient. In short, although the H3559 offers small size, low power, and fast sampling, this range finder did not offer some other important features, mainly the ability to multiplex.

Because of the limited choices of range finders that met the microrover's requirements, it was then decided to asses the feasibility of building a custom made range finder. As mentioned, phase shift and triangulation methods offer the best solution. However, phase shift electronics are somewhat complicated, and a range finder was required within a few months. Therefore, triangulation was decided as the method to pursue.

Active triangulation can satisfy the necessary requirements for rangefinding. Some of the benefits of triangulation include:

- An inherent ability to operate at short to medium ranges
- Improved accuracy as range is reduced
- Simple electronics
- Simple ranging algorithm from the raw data
- Small optics

A triangulation system has many traits that make it well suited for a small mobile platform, as listed above. The main drawback to any triangulation system is the rapid loss of accuracy as range is increased. This, however, is acceptable for any application with dead reckoning, because of the position errors that accumulate during travel. For the microrover, good accuracy is important up close, though, and this is exactly what triangulation systems provide. The 'missing parts' problem with triangulation can be avoided on the microrover since the baseline separation can be small (a few centimeters) relative to the objects of interest. A small baseline is possible due to the limited maximum range, modest accuracy requirement, and small optics.

6.2.5. Rangefinder Error

Accuracy vs Range Interval

Before the detailed design of the rangefinder subsystem is discussed, it is important to define the acceptable errors, as well as the required range interval. These two parameters, range and error, are strongly coupled. As with most systems, the field of view and the accuracy have a clear tradeoff. In the triangulation case, the field of view essentially is the difference between the angle of incidence of the spot reflection to the collection aperture at the minimum and maximum distance (see Figure 6.22).

Field of view is mainly driven by the minimum distance. For example, if the range interval is only from, say 5 feet to 6 feet, then the sensor can be arranged so that the entire swing of the focal point is represented by this distance swing. Hence the accuracy and resolution can be very good. On the other hand, if the range of distances is from 0.1 foot to 100 feet, then the focal point swing must be very small in order to capture the entire distance swing. In addition, most of that swing is near the small distances. Hence, accuracy and resolution is severely degraded if the ratio of maximum to minimum range becomes large.

Since the range interval and accuracy are so closely tied, either parameter should not be determined completely independently. It was decided that the rover has the capability of sensing to about 10 feet. This gives the rover enough room to choose a somewhat optimized path, but is not so far that the data cannot be relied upon. This decision was based upon many items, including the rover control code, which determines how range data is used. Simulations developed by Malafeew [1] have been used as a design tool for supporting the sensing distance decision. One problem with sensing too far is that the complex pitch and roll angles of the rover can make range data very difficult to interpret 'on the fly.' At this point, it appears that a maximum distance of 10 feet is sufficient. The minimum range of 0.5 foot was determined based on the rover geometry and location of the sensors. Since the sensors are somewhat inboard, objects cannot get closer than a few inches. This minimum range is large enough so that the rover can still react to a tight obstacle field, yet small enough so that it is not blind to objects up close.

To start, then, both the minimum and maximum range are not set in stone, but are not expected to change significantly. In order to define the error budget, these ranges will be used. If it is not practical to achieve the distance swing with the error budget, then some sacrifices will have to be made with either range or accuracy.

Range Finder Error Budget

The acceptable error in range and azimuth varies with range. It is more important to have the accuracy up close for two reasons:

1. The rover needs mobility space to react to obstacles, due to its kinematic constraints. In a tight obstacle field there is not much space available, and so the location of nearby obstacles is critical for hazard avoidance.

2. Dead reckoning errors increase with distance, so that avoiding an obstacle at close range will yield good accuracy in rover position estimation. Thus ranging errors up close need to be as good as the navigation errors at close range. At larger distances, the ranging errors can be nearly the same as the navigation errors. As the rover moves towards the object, its accuracy will increase.

From the navigation section, position errors grow mainly with traveled distance, due to the odometry errors of the drag wheel and drive motor tachometers. To be conservative, it can be assumed that the mean navigation error is 5 % of the traveled distance. Simulations show that for a mean rock distribution of 10 % coverage, the total traveled distance is about 1.3 times the straight line distance. Hence, when an obstacle at 10 feet is detected, its range error should be less than its dead reckoning error from traveling 13 feet. The navigation error at 13 feet is then:

$$E_{nav} = 13 \text{ feet} \times 0.05 = 0.65 \text{ foot}$$

Hence, the range error at 10 feet should be no more than 0.65 foot. Up close, at say one foot, the navigation error is

$$E_{nav} = 1.3 \text{ feet} \times 0.05 = 0.065 \text{ foot (0.8 inch)}$$

Hence, the range error at 1 foot should be no more than 0.8 inches. In short, it can generally be said that the range error needs to be less than 6.5 % of the straight line distance. The importance of this figure is greater at close distances, where there is not much space to react. If this range accuracy is not achieved at close distances, the control code can blindly add an uncertainty to the data point, but this yields an inefficient control system. Therefore it is important, but not imperative, that this requirement is met at close ranges. On the other hand, much higher accuracy will be washed out by navigation errors, so that 0.065 % accuracy is almost just as helpful as 6.5 %. Higher accuracy still helps, but the improvement beyond the 6.5% point decreases rapidly.

6.2.6. Configuration of the Active Triangulation Rangefinder

With active triangulation, there are still many options in the method of use, as discussed previously. The simplest method is the 'spot projection' method, which measures the angle of incidence from a diffusely reflected beam of small cross section. This method

provides the least amount of data, though. But since the sensor design philosophy is to allow the rover control system to do the most with the least, it appears at this point that the spot projection method will suffice, without 2 axis scanning. The more complex methods of structured lighting can provide more information per sample, but this information has not been proven to be necessary. For instance, a complete topography of the ground within a 10 foot circle would provide enough information to improve the decision making of the path controller, but at a large cost in complexity, processing, time, and probably size and mass. As a systems issue it was decided that range data be provided, but the capability of the ranger must be weighed against its cost. At the minimum, the rover needs azimuthal coverage, since mobility is mainly in the horizontal plane. Coverage in elevation can certainly be helpful, but the extra degree of freedom, yielding a two-dimensional depth map, greatly increases the demands of the processor. One of the advantages of having man in the loop is the ability to provide high level path planning on a daily basis, so that the autonomous sensing needs are greatly reduced.

The question then remains, 'is azimuthal coverage enough?' The answer depends, of course, on most importantly the degree of mobility of the rover. Consider an obstacle field of varying heights. To negotiate this field, the rover can move *in* the ground plane by steering, and *normal* to the ground plane by climbing and descending. If all the obstacles protrude above the climbing limit, then the rover has no other choice than to steer to avoid them. At the other extreme, if all the obstacles are much smaller than the rover wheel height, then steering is not necessary, and the rover can climb over them. At intermediate-height obstacles, where climbing is still possible, it becomes a question of cost and risk, for climbing consumes large amounts of power, degrades dead reckoning accuracy, and increases the chance of a mobility failure. However, mobility in the horizontal plane is not without cost either. To properly determine the optimum path would require a very detailed analysis of the immediate area, taking into account obstacle heights and shapes, as well as other factors. An easier solution is to simply detect obstacles above a certain height, providing only one bit of data in elevation or height. For instance, if the obstacle is below the wheel diameter, then it is not even detected. This may seem simplistic, since all data below level of the sensor is lost, but this should be acceptable. The obstacle information that is mainly of concern involves that which will likely present a hazardous situation because of its height. In a way it is like saying that if the obstacles are not important enough to be considered hazardous, then it is not worth the effort to analyze them. Threshold techniques like this are not optimal, but are clearly simpler, which is important for a system like a microrover with such limited

resources. With this configuration, spot projection should suffice as the active triangulation method.

Without elevation coverage, the remaining resolution can then be concentrated in azimuth, or heading. The way in which this coverage is achieved depends on the beam geometry. Spot projection essentially gives the range of the reflection spot, which can be considered a point source for cases of small beam cross section and divergence. This allows very high angular resolution, with the drawback of low coverage. Therefore, an angular scan will require many more data points for a given coverage. This is beneficial if resolution of the scanned object is critical, and the time and computer resources are not critical issues. However, for the simpler cases without the available resources, the amount of data points taken should be as low as possible.

This may seem to be a good reason to go to a larger beamwidth. However, the spot size cannot be large since accuracy is reduced from the blurred focus. More importantly though, the spot needs to be sufficiently small so that the entire spot falls on any given object. If only a portion of the beam is intersected by the object, then the centroid of that shape is what is interpreted as the entire beam. This can cause large range errors especially at the longer distances, where range calculations are very sensitive to the spot position,. Because of this, the beam cannot have a large beamwidth or beam diameter.

In summary, the active triangulation spot projection method is chosen as the ranging method for the rover. It will suffice to have coverage in azimuth, but not in elevation, yielding planar azimuthal coverage. The height of this plane is critical, and dependent upon mobility characteristics. The active beam must be small and narrow, so that the spot size is small. Therefore, some arrangement must be made to scan the range finder, so that 180° azimuthal coverage is provided in front of the rover. The design of the prototype range finder for MITy-2 is described in Chapter 7.

CHAPTER SEVEN
PROTOTYPE HAZARD AVOIDANCE HARDWARE

7.1. PROTOTYPE RANGEFINDER DESIGN

7.1.1. Overview of Range Finder Requirements

The previous sections have led to specifications and performance requirements that are the driving factors in the design of the rangefinder. These are:

1. Ranging from 0.5 to 10 feet
2. Accuracy that is 6.5 % of the range
3. Small active beam
4. 180° coverage, only in azimuth
5. Sampling rate of about 10 to 30 Hz

The following will present the design process for an active triangulation range finder, governed by the above requirements. The configuration is of the 'spot projection' type, which is the simplest. Although the discussion in this section is specifically for a prototype range finder for use on MITy-2, the design methodology applies to active triangulation range finders in general.

7.1.2. Method of detection

Although many design aspects of the active triangulation spot projection rangefinder were discussed, the choice of the portion of the e-m spectrum was not. Detection is a big driver in this choice, since position sensors are not manufactured in a wide range of responsivities, as they are relatively new devices.

For detectors, the choice is mainly between discrete arrays and continuous sensors. Discrete arrays range from multielement photodiode arrays, to high-resolution charge coupled devices (CCD). As mentioned in section 6.2, CCDs offer the high resolution

needed to distinguish the focal point position at the longer distances, where multi-element arrays (e.g. 1024 x 1 elements) have a much smaller element density, and are prone to element failures. For low light-level applications CCDs have an advantage over photodiode arrays because of their higher S/N. Even though their unit gain is about the same as a photodiode array, the S/N of a CCD is higher due to its inherently low noise. Other advantages of CCDs include excellent dynamic range and spatially uniform response. Hence, CCDs offer many advantages over photodiode arrays. However, both of these discrete sensors require some degree of scanning. Although scanning can be advantageous in being able to analyze the *profile* of the collected energy, the process is more complex and time consuming. In active ranging systems the illumination source has a known spectral emission, so that filtering ambient noise is much easier, and it is not as important to look at the distribution across the detector (Figure 4.1).

The detector of choice, then, is the standard position sensitive detector, or PSD. These devices do not offer the possibility of measuring the profile of the collected light like discrete arrays do, but they are very simple to operate, and fast. The focal point location can be calculated by simply measuring the two currents on either side of the detector. When ambient noise is present, the linearity of the currents with spot position allows simple algorithms for isolating the contribution from the active source. This will be discussed in more detail later in 7.1.5.

7.1.3. Choice of Wavelength

One limitation to the use of PSD's are the responsivities that are offered. PSD's have been in production only since about 1983, and are finding more uses every day. Currently they are only made from standard silicon PIN photodiodes, and hence are useful only in the visible and near infrared (IR) portion of the e-m spectrum (Figure 6.13). For best detection, the active source then should be near the peak PSD responsivity of about 920 nm, which is in the near IR.

From previous discussion, there are many performance issues that are affected by the chosen wavelength. The advantages of using λ=920 nm are:

- High resolution possible with small optical components
- High probability of diffuse reflection

The first point is a very important one, since the small components can open the door for many design options, including the ability to have multiple units for redundancy. Since the wavelength is small, the resolution that is needed for both collection of the reflected energy and collimation of the active beam can be achieved with small lenses. Also, the illumination source can be from solid state devices such as LED's and laser diodes, which are compact and readily available.

The second point is important for reliability in the data collection, since specular reflections degrade the system performance. The small wavelength makes most objects appear rough, based on the Rayleigh roughness criterion (see section 6.1.3), allowing energy to be reflected at even grazing angles of incidence.

Some of the drawbacks to using $\lambda = 920$ nm are:

- Large amount of ambient noise from the sun
- Low penetration of solid matter

The solar spectral emission fortunately does not peak at 920 nm (Figure 7.1), but its overall magnitude is still large within the sensitive range of the photodiode. This should not be a problem, however, since S/N can be greatly improved with standard passive filtering techniques, especially with the narrow spectral emission of LED's and lasers. Section 7.1.10 discusses passive filtering in more detail.

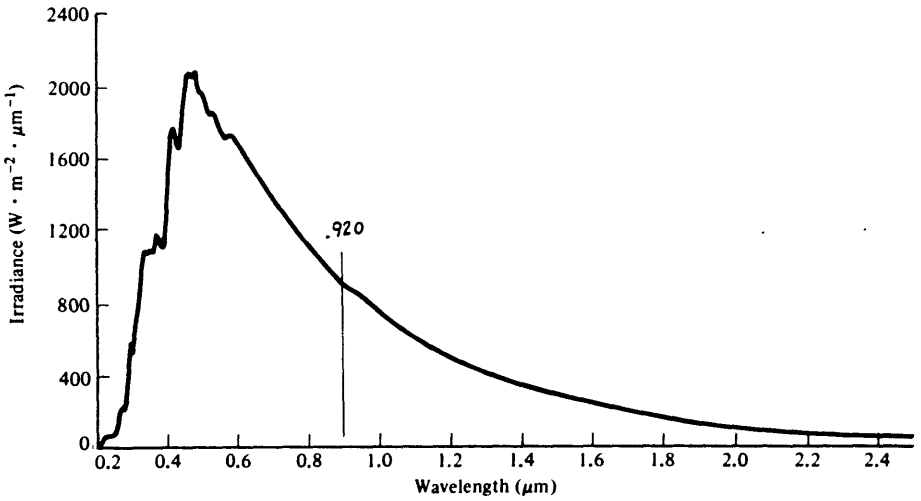

Figure 7.1: Solar spectral emission

The second item applies mainly to dust issues. Most foreign matter like dust can severely degrade the optics, since the individual particles can collect into films that are thick, attenuating the signals or ruining the resolution. Lower frequencies such as microwave have the advantage of penetrating protective coverings and small amounts of dirt easily, making sensing possible in harsh environments. However, this comes at a great expense in size and resolution. The dust issues present perhaps the most difficult issue in sensing in the near IR, and some provisions need to be made to keep the optics relatively clean.

These are the main issues involved with the wavelength of 920 nm for the active source. Despite the dust problems, the compact size and good reflection characteristics offer important advantages in the range finder design.

7.1.4. The Illumination Source

The active source needs to be near 920 nm, and should allow for sufficient collimation of the beam, so that the projected spot of energy is small. As previously mentioned, the spot needs to be small for two reasons, which are illustrated in Figure 7.2:

1. To ensure that the entire spot falls on the object, since if only portions of the spot are reflected, the triangulation geometry is distorted, which introduces range errors. Also, of course, energy is wasted in the missed portion of the beam.
2. To appear to the collection optics as an extended light source, so that the incident rays are nearly parallel. This allows the focus to be sharp, for good resolution in position sensing.

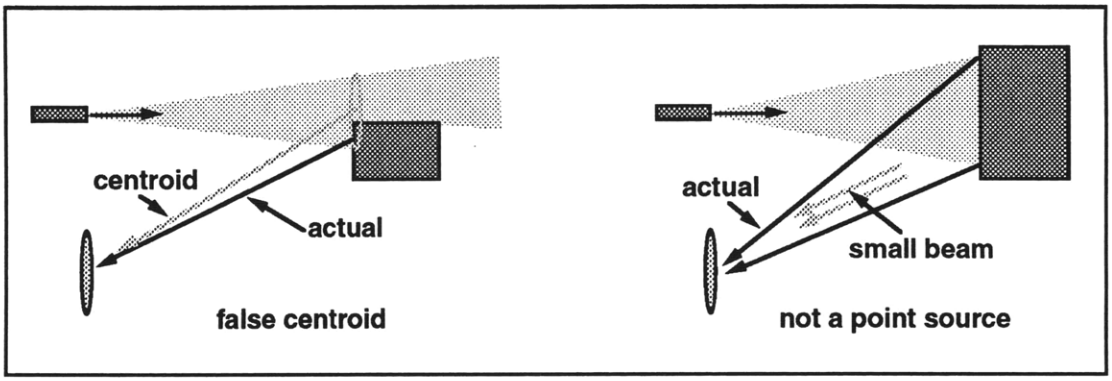

Figure 7.2: Drawbacks of a large beam diameter. Top: partial illumination results in a false interpretation of beam centroid. Bottom: large reflected light source results in a blurry focus spot on the PSD.

To achieve a small spot of illuminance, the beam must be well-collimated and also be of small diameter. Achieving both is difficult, unless the wavelength is near the visible, which is the case. The divergence angle due to diffraction is approximately:

$$\theta = 1.22 \frac{\lambda}{D} \tag{6.7}$$

At 10 feet (3 m), the spot diameter should still be less than the baseline distance, as well as be small relative to the object size. As a start, a diameter of 1 inch at 10 feet will suffice. The divergence of this is about 4 milliradians (mrad). To achieve such a low divergence, the diameter of the aperture has to be

$$D = 1.22 \frac{\lambda}{\theta}$$
$$= 1.22 \frac{(920 \text{ nm})}{(0.004 \text{ rad})}$$

$$= 0.2 \text{ mm} \tag{7.1}$$

This is indeed small! It turns out, however, that the physical limitations of producing a collimated beam prevent the use of lenses this small. The reason is due to the size of the actual light-producing area relative to the size of the collimating lens. A lens transforms a spherical wave into a planar wave, if the light source is at the focal point of the lens. This is similar to the collection of light for the PSD, where a near-planar wave is transformed into a spherical wave and focused to a point. This process is shown in Figure 7.3a.

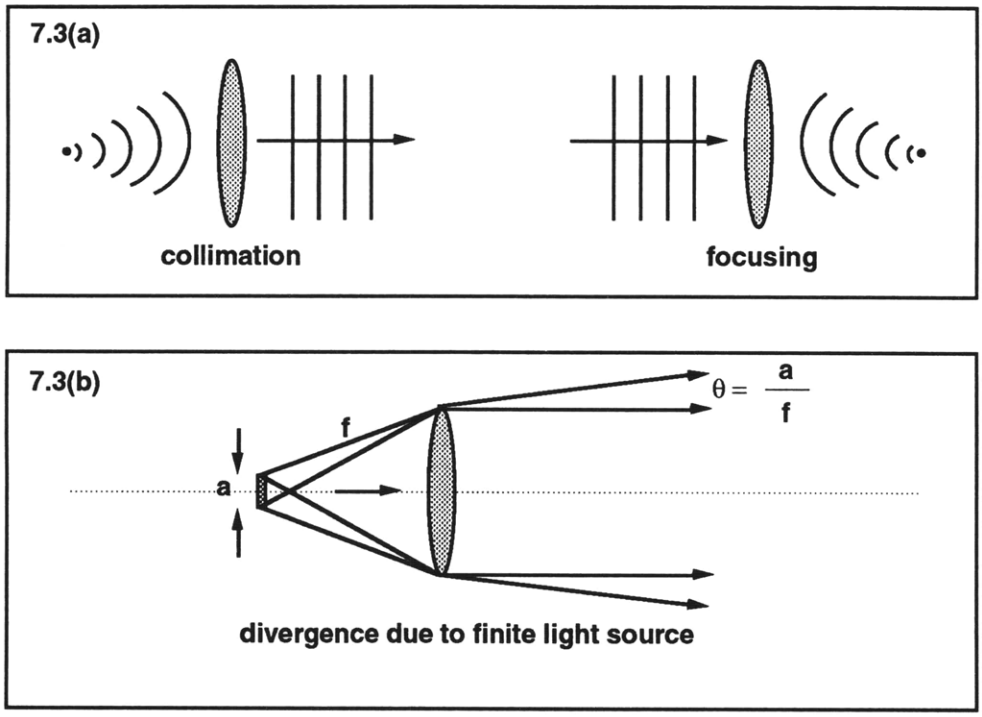

Figure 7.3: Using a lens for collimating light. a) point source of light generates spherical waves that the lens transforms into a planar wave, and vice versa. b) when the light source has a finite size, the resulting beam has an additional divergence.

If the light source at the lens focal point is from a 'point', which it never actually is, then a true spherical wave is not produced, and the resultant beam will diverge more than that from just diffraction. The amount of divergence from this finite surface is approximately just the difference in angle from each side of the active surface (that which produces light) to the edge of the lens (Figure 7.3 b). The divergence angle of a collimated beam due to the largest dimension of the active surface, a, is

$$\theta = \frac{a}{f} \tag{7.2}$$

Let's first take an LED as an illumination source. A typical high power LED has an active surface of length a = 500 μm. If this was used with a collimating lens of diameter 0.2 mm, which can be assumed to have a focal length of the same dimension f=0.2 mm, the divergence would be about:

$$\begin{aligned}\theta &= \frac{a}{f} \\ &= \frac{500\mu m}{0.2\ mm} \\ &= 2500\ mrad\ (90°)\ !\end{aligned}$$

Obviously this is not acceptable. In fact, to achieve a diffraction-limited divergence of 4 mrad, the lens focal length would need to be

$$\begin{aligned}f &= \frac{a}{\theta} \\ &= \frac{500\ \mu m}{4\ mrad} \\ f &= 125\ mm\end{aligned} \tag{7.3}$$

This is achievable, but is rather large.

Now, if a laser diode is used instead of an LED, the amount of divergence can be reduced significantly due to the much smaller active surface. The active surface dimension for laser diodes can vary from under 1 μm for low power continuous wave, to 200 μm for high power devices. Assuming this is 100 μm, in order to achieve a divergence of 4 mrad the lens focal length needs to be:

$$\begin{aligned}f &= \frac{a}{\theta} \\ &= \frac{100\ \mu m}{4\ mrad} \\ &= 25\ mm\end{aligned} \tag{7.4}$$

This is now starting to approach the small size that is sought after. Note that for small divergence angles, for a given lens the divergence is proportional to the longest active surface dimension. For the case above, if a laser diode with a=20mm is used, the focal

length of the lens can be 5 times smaller for the same divergence. If low power illumination is sufficient, a = 1μm, so that the beam divergence due to the finite active surface dimension will be of the same order as the divergence due to diffraction out of the aperture, which is the best collimation possible.

It may not be so important to have a diffraction-limited beam, but it should be clear that the size of the optics can be greatly reduced by using laser diodes as the illumination source, opposed to LEDs. Laser diodes have several other advantages over gas lasers and other sources:

- Compact package (typically 9 mm dia)
- Simple operation (current driven, like LEDs)
- Solid state construction (high reliability)
- Relatively inexpensive
- Very narrow spectral emission (about 10 nm)
- Can be well-collimated
- Efficient power conversion (30%)
- Readily available in the red and near infrared

For some uses, the coherence of laser light is also very advantageous, although this is not the case here.

Some of the drawbacks to using laser diodes include their degradation with excessive temperature. It is important to heat sink them, depending on their operation. Another feature of laser diodes is their nonuniform emission pattern, due to the noncircular active surface. Most have an active surface that is much longer in one dimension than the other, resulting in a diffraction pattern that is spread in the same manner. If it is important to have a uniform beam profile, optics can correct this. Otherwise, the beam cross section will generally be elliptical.

7.1.5. The Position Sensitive Detector

In order to detect the angle incident to the receiver for triangulation purposes, a PSD is used indirectly. Opposed to measuring angle, the PSD measures the position of the focused energy from the reflected spot on the object (see section 6.2.3). A standard geometry for the setup is shown in Figure 6.24 .

$$R = \frac{B f}{d} \tag{6.26}$$

If the incident reflected energy is in the form of a near-planar wavefront, then the focused light of the receiver lens is very small. The PSD senses the position of the light spot, so that the displacement d is measured. Once d is measured, range is then easily calculated.

Theory of Operation

PSDs are typically a PIN photodiode, with the P layer having uniform resistance across the active area. When light falls on this surface, a photocurrent is generated, and is collected by two anodes: one on either side. Because of the uniform resistance, the amount of current at each anode is proportional to distance from the incident light (see Figure 7.4), while the sum of the two currents remain constant. When light falls in the middle, the two photocurrents are equal. As the light is displaced towards anode A, more current will flow, and I_A will increase and I_B will decrease in proportion to the fractional displacement. A good analogy is a simply supported rigid beam. The amount of force (current) on each of the two supports (anodes) varies linearly with the position of the point load.

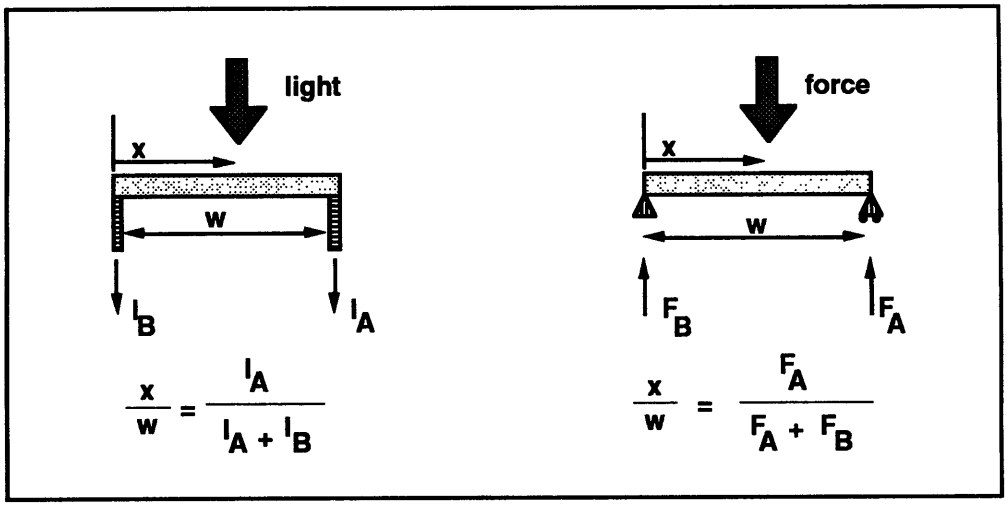

Figure 7.4: Calculating the spot position from the PSD anode currents is similar to finding the reaction forces on a beam.

Quantitatively, for a detector with an active surface width w, the position of the spot of light can be calculated from

$$\frac{x}{w} = \frac{I_A}{I_A + I_B} = \frac{I_A}{I_{sum}} \tag{7.5}$$

This is also equivalent to

$$\frac{x}{w} = 1 - \frac{I_B}{I_{sum}} \tag{7.6}$$

Note that the two fractions, I_A/I_{sum} and I_B/I_{sum} are always less than unity, and always add up to 1. Hence it is sufficient to obtain the light spot position by measuring two parameters I_A and I_B. These can be measured directly, or usually indirectly in the form of sum ($I_A + I_B$), difference ($I_A - I_B$), or normalized (I_A/I_{sum}, I_B/I_{sum}, $(I_A-I_B)/I_{sum}$).

Ambient Light

Ideally, the only light incident on the PSD is that from the active source. When ambient light is also present, such as that from the sun, the contribution from the active illumination must be distinguished. Unlike an array, PSDs do not have the ability to detect the light distribution across it, but instead they weigh all contributions to yield a net effective spot position, which can be displaced from the active contribution, as shown in Figure 4.1 (sun sensor). There are a few methods of filtering to help this problem, using both AC and DC techniques.

First, since the active source is of a known spectral emission, a passive optical filter can be placed in front of the receiving lens. In fact, the choice of a laser diode allows a very narrow bandpass filter, which is called an interference filter, so that the noise that is passed is restricted to the same band as the filter.

Another method of removing ambient light involves background light subtraction. By measuring the PSD currents immediately *before*, and then *during* the active illumination, the currents at each time interval can be subtracted, yielding only the contribution from just the active source. This is based on the assumption that the noise is constant during the time between the two measurements. This assumption is usually valid for time intervals of milliseconds, since outdoor noise is DC, and the rover moves slowly.

Background light subtraction is possible because of the linearity of the PSD. Superposition holds, and so the anode currents from multiple spots add linearly. Figure

7.5 shows a PSD with light incident at x=0.25w and x=0.75w. By removing the x=0.25w source and measuring the anode currents, subtraction can yield the position of the other source. Hence, the sensor linearity allows simple subtraction to be used in removing ambient noise from successive measurements.

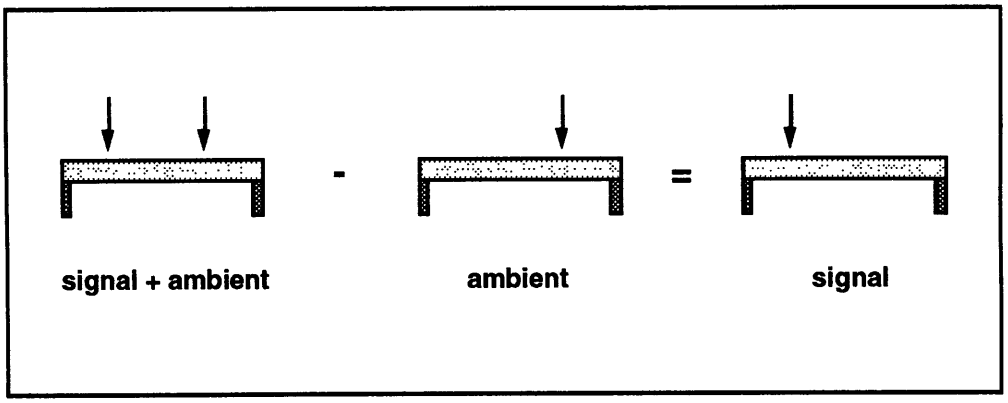

Figure 7.5: With a PSD, contributions from multiple sources add linearly. This allows subtraction of background light.

Another level of filtering can be from using a modulated light source, and using AC filtering in the detector electronics (6.1.6). This is a very powerful method of rejecting ambient DC noise. This method was not used for the prototype range finder for MITy-2, but will be used on later versions.

PSD specifications

PSDs are used in a wide variety of alignment and displacement sensing, and are available in many shapes and sizes, despite their youth in the marketplace. The one-dimensional PSDs typically have active surfaces that range in size from 2 mm to 100 mm. There are many parameters that influence the choice of a PSD other than size. Accuracy, linearity, dark current, rise time, responsivity, capacitance, and detector noise are other factors to take into account, depending on the method of operation, speed, irradiance, etc.

UDT Sensors, Inc was chosen as the PSD manufacturer for the prototype rangefinder, mainly for reasons of accuracy and cost. Some of the specifications for the UDT sensor are shown in Table 7.1, along with some smaller versions. The SL15 has a 1x15 mm active surface, which is rather large. This was chosen mainly for ease of alignment and machining of the parts. The smaller PSDs have better characteristics such as lower dark

current, lower capacitance, and faster rise time. However, for simplicity, DC operation was chosen with the active source pulsing for a few milliseconds at a time, so that the PSD response time was not an issue.

Model	Effective Sensing Area, mm	Position Detection Error (max), μm	Responsivity at 900 nm, A/W	Rise Time * (10-90%), μs
SL3-2	3 x 1	6	0.6	0.2
SL5-2	5 x 1	10	0.6	0.4
SL15	15 x 1	30	0.6	5

* Bias voltage of 15 V

Table 7.1: UDT Position Sensitive Detector specifications.

One of the main issues in accuracy of the system is the position detection error of the PSD itself. This error is proportional to the active surface dimension, w, so that there is no accuracy advantage in using larger PSDs. The position detection error is given to be 30 μm maximum, over 90% of the sensing area. Range error, however, depends on the geometry of the remaining components. This error will be largest at the maximum range (10 feet), where the slope of R vs d is largest, as shown in Figure 6.23. This slope is mainly influenced by the baseline separation, and the receiving lens focal length, which will be discussed shortly. The contribution of the position detection error to the range error is linear, so that it is easy to compare this specification with that of other PSDs.

Although the entire active area is available for sensing, it is important to know that the linearity and accuracy of the sensor degrades near each anode (5% of w on each side). Hence, the long-range light spot position, where range is very sensitive, should not fall within either of these regions.

7.1.6. Irradiance of the Reflected Signal

Before choosing the power of the laser, predictions of the returned signal strength must be made. The collected power due to the laser depends mainly upon

- illumination power
- surface reflectivity
- receiver aperture

- filter attenuation
- detector responsivity

Assuming a laser source near 900 nm, the photocurrent produced is from K_{resp}=0.6 amps/watt for the PSD. In order to keep the optics small, a 15 mm diameter receiving lens (A_{rec} = 1.77 cm^2) was chosen. Also, interference filters attenuate about 50% in their bandpass, so K_{filter} = 0.5. Hence, the resulting PSD current, I_{sum}, from the reflected irradiance I in watts/cm^2 is:

$$\begin{aligned} I_{sum} &= I\ K_{resp}\, A_{rec}\, K_{filter} \\ &= I\ (0.6\ A/W)\ (1.77\ cm^2)\ (0.5) \\ &= 0.53\ I\ (amps) \end{aligned} \tag{7.7}$$

The above equation relates the PSD output current to the incident irradiance, based on the given optics. In order to determine the required irradiance, the minimum current must be established. This was measured to be about 100 nA, due to all of the noise sources on the MITy-2 prototype rover. In theory, this should be much lower, so that future illumination sources can be weaker. However, for this prototype, the laser would have to provide enough power so that the output current of the PSD is about 100 nA. From the above equation, the required irradiance is then about 200 nW/cm^2.

$$I = 200\ nw/cm^2$$

In order to produce this irradiance at the maximum distance of 10 feet, the laser power must be relatively high. Since the laser is small and collimated, the main contributions to the weak return signal is due to the inverse square of the range R, and the object reflectivity, K_{refl} (see section 6.1.4). From equation 6.10, the simplified expression of irradiance is

$$I = \frac{K_{refl} P_e}{2\pi R^2} \tag{6.10}$$

where P_e is the emitted optical power. Again it should be noted that this is a simplified equation, neglecting any angle effects, so that the reflection is assumed to be perfectly diffuse. Assuming a reflectivity of 0.15, the irradiance due to diffuse reflection of a small laser beam from a target at 10 feet (300 cm) is

$$I = \frac{(0.15)P_e}{2\pi(300cm)^2}$$
$$= 265\ P_e\ (nw/cm^2) \qquad (7.8)$$

where P_e is in watts.

In order to achieve an irradiance of 200 nW/cm^2, the laser must produce

$$Pe = \frac{I}{265}$$
$$= 200/265 = 0.75\ W \qquad (7.9)$$

This is a relatively large output for a laser diode, but it is very feasible.

7.1.7. The Laser Diode

The choice of the laser diode is constrained to operation near 900 nm, and with a 'short term' optical power output of about 0.75 watts. The near infrared range is a very common laser diode (and LED) frequency location, since it corresponds to gallium arsenide (GaAs). GaAs emitters are often paired with silicon detectors due to the matching of the emitter's output with the detector's peak responsivity. In fact, it is easier to produce laser diodes in the near infrared range than it is in the visible, which is why shorter wavelength lasers, like blue, are still considered to be in the experimental stage. Laser diodes offer center wavelengths that range from 450 to 2500 nm, but are low-cost and readily available from about 600 to 1200 nm.

To deliver 0.75 watts, however, is beyond the capability of many laser diodes. The most important parameter is then the amount of time that this power is sustained. Although there are many classifications of laser diodes, one of the most important is their mode of operation with respect to time. There are three types:

1. Continuous wave (CW)
2. Pulsed
3. Quasi-CW

A CW laser can produce light continuously, if the diode has adequate cooling. These are generally restricted to low power applications, usually in the milliwatts. However, recently the trend has been towards higher and higher power, so that power outputs of 75 watts are possible with continuous wave laser diodes. These higher power devices have very large active surface areas (200 - 400 μm), so that collimation is limited with small lenses. Although such high power specialty lasers are possible, CW lasers are generally restricted to outputs of 200 mW.

Pulsed lasers, on the other hand, are limited to pulse lengths of typically under a microsecond. However, the power output during this time can be very large, such as 500 watts. Due to their short pulsewidth, they can operate at pulsing frequencies in the kHz, or even in the MHz. Their pulsing features make these laser diodes ideally suited for fiber optic communications, or other applications where high power in a short interval is required, and a low duty cycle is acceptable. In fact, the pulsed DC operation of active triangulation is theoretically a good application of these lasers. Unfortunately, even the small PSDs have rise times of about 1 microsecond or more, so that light pulse lengths of 5 μs are required. Additionally, for the prototype rangefinder, pulse lengths of at least 1 ms were desired to ensure adequate stabilization before sampling. Because of the long pulse length that is required, pulsed laser diodes are not adequate for now. However, future improvements in sampling can allow the use of these devices.

Fortunately there is another type of laser diode operation that falls in between pulsed and CW. These are referred to as 'quasi-CW.' These lasers offer some of the advantages of the other two types. The radiant power is high (some exceeding 250 watts), yet pulse lengths are not restricted to under 1 ms, and the active surface is not as large as with high power CW diodes. Most of these applications are for the firing of ordinance on aircraft, since accidental triggering due to electrostatic discharge is eliminated. Although they are more expensive than pulsed laser diodes, these lasers are not as costly as high power CW devices. Quasi-CW lasers are well suited for the pulsed DC operation of the prototype rangefinder.

The availability of quasi-CW laser diodes is rather limited. Two of these lasers, and an additional CW laser are shown in Table 7.2 below. The CW laser, made by Ensign Bickford Aerospace (EB) is also included due to its low cost. In fact, due to the high cost of the Spectra-Diode laser, the choice was mainly between the EG&G (quasi-CW) and the EB (CW). Both offer radiant powers above 0.75 watts, with pulse lengths above 1

ms. One advantage of the EB diode is the tolerance of the center wavelength, so that if multiple diodes are purchased, standard interference filters can be purchased at one wavelength, instead of individually matching them (a cost issue). However, this device is overrated for the application due to its CW design, and the active surface is larger, making collimation more difficult. Hence, the EG&G diode was chosen, as the specifications match the needs of the prototype rangefinder well. To summarize, the requirements that are satisfied by the EG&G C86104 quasi-CW laser diode are:

1. Greater than 0.75 watts radiant power output
2. Greater than 1 ms pulse length
3. Reasonably-sized active surface for good collimation
4. Greater than 20 Hz pulse rate
5. Wavelength near 900 nm (922), to match the PSD
6. Relatively low cost

	EG&G C86104 E	Ensign-Bickford	Spectra Diode
Power Output	1.3 watt	2 watt	1 W
Max Pulse Length	10 ms	CW	CW
Active Surface Dim.	75 μm	150 μm	-
Cost	$ 250	$ 600	$ 1200

Table 7.2: High Power Laser Diodes Specifications

7.1.8. Collimating Optics

The requirements of the laser diode optics are

1. Lens diameter that is large enough to limit divergence from diffraction, yet small enough to provide small beam diameter and allow for compact packaging
2. Large enough focal length to limit divergence from the laser's finite active surface dimension
3. Capture energy efficiently from the laser's wide angular divergence (requires large numerical aperture)

It has already been shown (eq. 7.1) that the beam divergence due to diffraction is negligible, so that the lens diameter will easily be sufficiently large. The beam diameter

is initially limited by the lens diameter, but at distances that are far relative to the lens size, divergence is the main contributor to beam size. But since divergence from diffraction is not an issue, there is not yet a reason to keep the lens diameter large.

The focal length, however was shown to be the main factor in influencing the beam divergence, from eq. 7.2. It turns out that the latter two points listed above define a design trade that relates the lens focal length to its diameter. The numerical aperture (NA) quantitatively describes the lens light gathering capability. It is defined as

$$NA = n_o \sin\theta_{max}$$

where n_o is the index of refraction of the medium between the lens and the light source, and θ_{max} is the maximum angle of acceptance, which is half of the cone angle. In other words, NA is the sine of the angle that the marginal ray makes with the optical axis. Since n_o=1 in this case, the numerical aperture is simply the sine of the acceptance angle:

$$NA = \sin\theta_{max} \tag{7.10}$$

Hence the theoretical maximum value of NA is 1, which is not physically possible. In fact, most practical lenses have NA < 0.6.

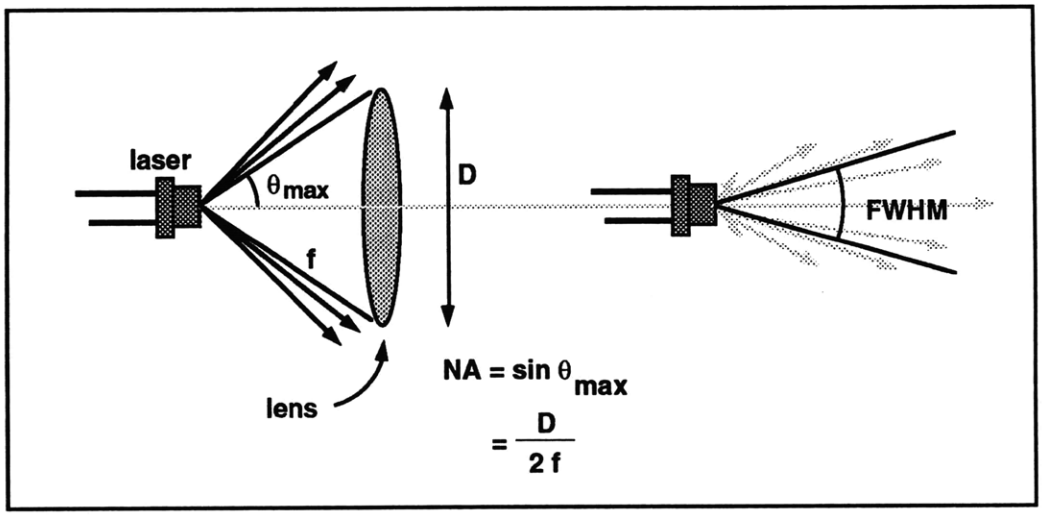

Figure 7.6: The numerical aperture of the collimating lens should be large enough to capture the angular emission of the laser.

The NA should be large enough to capture nearly all of the light that is emitted from the laser diode, so that light is not wasted. The EG&G diode, like most laser diodes, has an angular emission pattern that is not axially symmetric. One axis diverges from the emitting surface much more than the other. The full width/half max, or FWHM, angle of the EG&G diode is about 5° x 25°. FWHM is the angular envelope that defines the boundaries where the intensity is half of the maximum. Because of the angular pattern, nearly all of the light falls within an envelope that is somewhat larger than the FWHM envelope. The larger of the two angles is what is used to determine the required NA. The lens was chosen to accommodate lasers with larger angular emissions, which can be as high as 42° for single quantum well structures. Since half of 42 degrees is 21 degrees, the maximum angle of acceptance should be about 30°, to be conservative. The required numerical aperture for gathering most of the light from the laser is then:

$$\begin{aligned} NA &= \sin(30°) \\ &= 0.5 \end{aligned} \tag{7.11}$$

The remaining lens requirement is then to choose a large enough focal length so that the divergence is limited to about 4 mrad (this will yield a 1 inch spot at a distance of 10 feet). Using equation 7.3, the required focal length for an active surface size of 75 mm is:

$$\begin{aligned} f &= \frac{75\ mm}{4\ mrad} \\ &= 19\ mm \end{aligned} \tag{7.12}$$

To summarize at this point, the collimating lens should have NA=0.5, f=19mm, and a diameter that is as sufficiently small. As mentioned earlier, the focal length and diameter cannot be chosen independently, because they are related to each other through the NA. The definition of NA can be restated in more common terms. This relationship is defined as

$$NA = \frac{D}{2f} \tag{7.13}$$

Hence, for f=19mm and NA=0.5, the diameter will be about 19 mm also. In other words, for numerical apertures near 0.5, the diameter and focal length of the lens will be about the same. Hence the physical size of the collimating lens is governed by f and NA, both of which are related to the diameter.

With cost in mind, the first choice of lenses was from Optima. These are inexpensive plastic lenses that have high NAs, but are of small diameter and focal length, since they are typically used with standard CW low power laser diodes such as in laser pointers. Hence, the divergence due to the small focal length would be unacceptable. For $150, Edmund Scientific had a lens that nearly satisfied the initial requirements, with f=8.18mm, and NA=0.55. Although the focal length is only half of that specified, it was the largest focal length lens available. More importantly, the NA requirement was satisfied, so that light would not be wasted. The design requirements were again evaluated, and it was decided to accept the penalty of higher divergence of the shorter focal length rather than sacrifice the illumination power, especially since the higher cost of these lasers is due to the high power output. Hence this lens was chosen as the collimating optics for the prototype rangefinder. Its specifications are:

Description	Laser Diode Collimating Lens
Focal Length	8.18 mm
Clear Aperture Diameter	9.10 mm
Numerical Aperture	0.55
Housing Length	10.16 mm
Housing Diameter	15.24 mm

Table 7.3: Collimating Lens Specifications

When used with the EG&G laser diode, the resulting beam should have a divergence of about 9 mrad. This will result in a 2 inch long spot (major axis) at the maximum range of 10 feet.

7.1.9. Other Geometrical Considerations

With the PSD, receiving lens diameter, laser, and collimating optics determined, the remaining parameters are the receiving lens focal length and the baseline separation between the emitter and detector optics.

Before proceeding further, the equation for obtaining range from the measured focus displacement, d, is:

$$R = \frac{Bf}{d} \tag{6.26}$$

and the sensitivity of R to d is

$$\frac{\partial R}{\partial d} = -\frac{Bf}{d^2} \tag{7.14}$$

The contributions to range error can be categorized as:

1. S/N related
2. Geometrical, due to the error in measuring d

The first item has already been analyzed, and the corresponding illumination source and receiving lens diameter was chosen. Therefore, the geometrical design needs to be established. This essentially involves the baseline separation, B, and the focal length of the receiving lens, f. These must be chosen mainly to achieve the desired accuracy over the range of 0.5 - 10 feet. In addition, there are angle limitations to the receiving lens, and also to the passive optical filter.

Relative Angle of the Emitter and Detector

Previous discussions have shown the optical axis of both the emitter and detector to be parallel. Also, the PSD was shown to be oriented similar to the baseline axis. It is only for this arrangement that equation 6.26 applies.

The simple right-triangle configuration is certainly the easiest to manipulate, since some of the parameters are decoupled, and the algorithm for range is simple. Other configurations, however, may be better suited for particular applications, and are worth mentioning briefly. For instance, if the range of reflected angles to the receiving lens varies more than about 12°, the focal spot will become distorted at the PSD plane when a simple one-element lens is used. In this case, orienting the emitter and receiver axes so that the range swing creates angle swings on *each side* of the receiving lens axis will keep the spot from distorting, and will preserve the angle on each side of the lens, as shown in Figure 7.7. This configuration is also important when the range is close to the baseline separation, in which case the incident angles of the reflected light vary substantially.

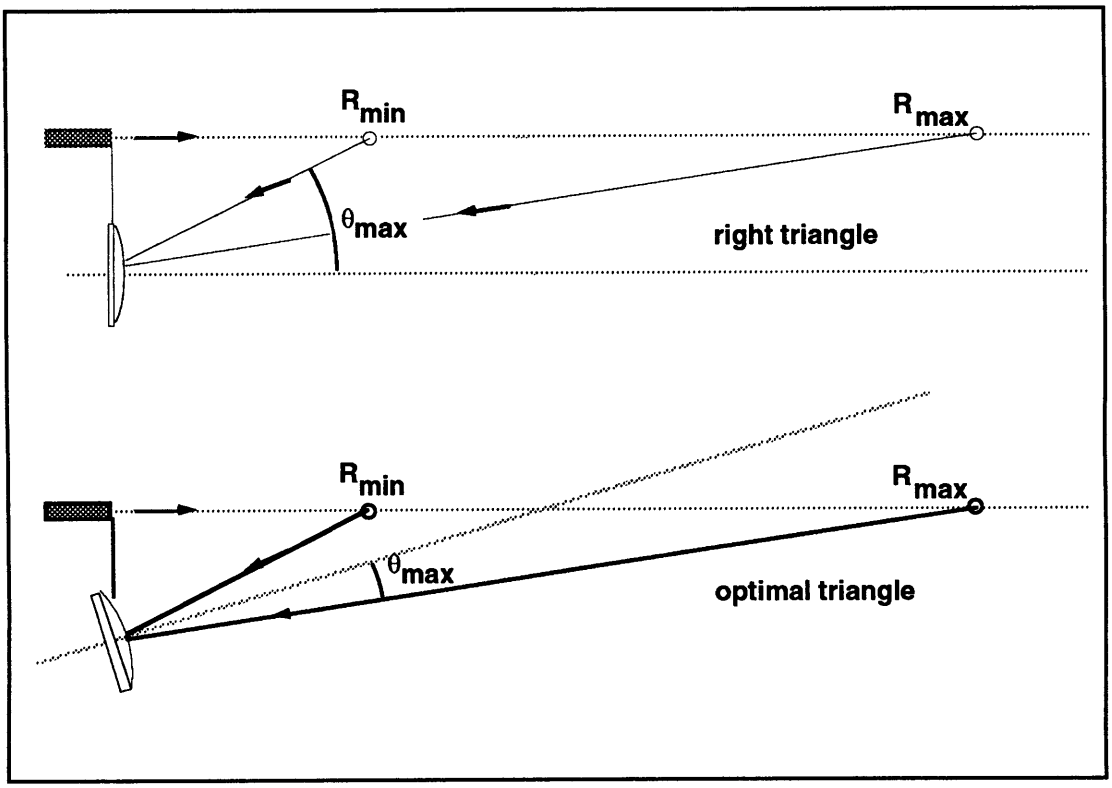

Figure 7.7: Two configurations of active triangulation. The bottom illustration shows that when the two optical axes cross, the angle of incidence can lie on both sides of the receiver optical axis, resulting in a lower angle of incidence.

The main drawbacks to a more complex configuration is that the special geometry makes:

1. Machining the parts more difficult
2. Range algorithms more complex

These are mild drawbacks for a final design, but during prototyping the machining is a big issue, for timeline purposes. Hence, it was decided to stay with the right-triangle geometry.

The Product, Bf

From equations 6.26 and 7.14, there is no distinction between the individual values of B and f. Only their product influences the range and its sensitivity. However, other geometrical constraints limit the individual values of B and f, and these will be discussed

shortly. First, however, there are two main constraints on determining the product Bf. One is governed by the range accuracy required, and the other is governed by the range interval required. It should be no surprise that there is a tradeoff between range interval, which corresponds to the receiver field of view, and accuracy. One calls for a large Bf (accuracy), and the other calls for a small Bf (range interval). The following sections analyze the effects of Bf on accuracy and range interval.

Bf: Effect on Range Error

First, the product Bf is mainly important for limiting the range error due to the PSD position detection error. Assuming a small uniform focal spot, the range error, δR, due to the error in d (δd) is (from 7.14):

$$\begin{aligned} |\delta R| &= \frac{\partial R}{\partial d}\delta d \\ &= \frac{Bf}{d^2}\delta d \end{aligned} \tag{7.15}$$

But from equation 6.26 the range error can be expressed as

$$|\delta R| = \frac{R^2}{Bf}\delta d$$

Note that the error in range due to the PSD position detection error varies *linearly* with δd, with the *square* of the range, and *inversely* with Bf. The ratio of R^2 to Bf is the key parameter, so Bf must be large enough so that the error is constrained at the maximum range.

Since the error requirements state that the range error should be less than 6.5% of the range, the above equation can be rewritten to accommodate this specification as

$$\frac{|\delta R|}{R} = \frac{R}{Bf}\delta d < 0.065 \tag{7.17}$$

Note that even when the error is expressed in a normalized form, it is still dependent on the range itself (linearly). Hence, the maximum percentage error in range will still occur

at the largest distance. Using 10 feet (3 meters) as the maximum distance, and $\delta d = 0.030$ μm (from 7.1), the minimum value of Bf is:

$$(Bf)_{min} = 1384 \text{ mm}^2 \tag{7.18}$$

To get a feel for this value, a first guess of equal B and f would yield a dimension of B = f = 37 mm. This dimension is relatively small, thus showing the possibility of a compact rangefinder. One very important point is that this prototype uses a rather large PSD for ease of mounting and alignment. It should be clear that using a smaller PSD can drastically reduce the size of the rangefinder, by as much as 80% or more. This will be discussed in more detail in chapter 8.

Bf: Matching the Range Interval to the PSD Size

Another factor in choosing the Bf product is to accommodate the full range swing that is desired. Bf must be small enough so that the focus spot does not travel off the PSD active surface throughout the entire range swing. However, Bf must be large enough to still satisfy the accuracy requirements as just described.

The swing of the focus spot on the PSD can be determined from equation 6.26. For a given range, the focus spot location is:

$$d = \frac{Bf}{R} \tag{7.19}$$

The focus spot swing, Δd, is therefore determined by the minimum and maximum range:

$$\Delta d = Bf\left[\frac{1}{R_{min}} - \frac{1}{R_{max}}\right] \tag{7.20}$$

For min and max range values of 152mm and 3048 mm respectively, the range swing is

$$\Delta d = 0.00625 \ \ Bf \quad (mm) \tag{7.21}$$

where Bf is in mm^2. Note that the focus spot swing is mainly determined by the minimum range.

For the PSD chosen, the active area is 15 mm. It is possible to use the entire active surface length, although the UDT sensor, like most manufacturer's, is linear and accurate over only 90% of the surface. To be conservative, it was decided to use about 75% of the surface. Hence the *usable* active surface length (denoted by ') is:

$$\begin{aligned} \Delta d' &= 0.75 \times 15 \text{ mm} \\ &= 11.25 \text{ mm} \end{aligned} \tag{7.22}$$

With the focus spot swing determined, the maximum Bf product can be calculated from equation 7.21:

$$\begin{aligned} (Bf)_{max} &= \frac{(\Delta d)'}{0.00625} \\ &= \frac{11.25}{0.00625} \\ &= 1800 \text{ mm}^2 \end{aligned} \tag{7.23}$$

Fortunately this does not conflict with $Bf_{min} = 1384$ mm^2, as calculated from accuracy requirements. If it did, either range swing or accuracy would have to be sacrificed.

To summarize, the product Bf has been calculated to lie within a certain range, as determined by range accuracy and range swing requirements. The result is

$$1384 < Bf < 1800 \tag{7.24}$$

in the units of mm^2. A value of Bf = 1600mm^2 was chosen as the target, which will give some room for other error sources, such as those from noise in the detection circuit.

Individual limitations of B and f

Now that the product Bf has been determined, it remains to choose their individual values.

Focal length, f

The focal length of the receiving lens has few constraints. Lenses can be made easily with long focal lengths, and can be made with relatively short focal lengths too. However, the focal length does have a minimum value which, for simple lenses, is

roughly that of its diameter, so that the numerical aperture is near 0.5 (see collimating lens). For the 15 mm *diameter* lens chosen, it can be loosely stated that the minimum *focal length* of the receiving lens is about 15 mm:

$$f_{min} = 15 \text{ mm}$$

Baseline Separation, B

The baseline separation, on the other hand, has two major limitations. First, like in all triangulation systems, the baseline separation can yield a problem in obstruction of the returned signal (see Figure 6.21). This is often called the 'missing parts' problem. Hence, the baseline separation should be small enough so that the chance of obstruction is small. Quantitatively, this is difficult to determine. For the size of the rover, which is about 50 cm wide, the baseline should be a small fraction of that, if possible.

The second constraint of the baseline separation is the limitation of the angle of incidence of the receiving lens. For a simple lens, the reflected light that is incident upon it in the form of a planar wavefront is focused to a point in the focal plane of the lens. For small angles of incidence of the wavefront to the lens axis, the focal point is displaced along the plane such that similar triangles (from which equation 6.26 is derived) exist. In reality, this is valid only for small angles, since the focal point actually swings in an arc rather than remaining in the focal plane. Since the PSD is constrained to a plane, the angle must be limited so that the focal spot in that plane remains relatively undistorted, which, as mentioned, was shown to be about 10 or 12 degrees for the plano-convex lenses tested. Above this angle, the spot in the focal plane distorted from near circular to an asymmetrical shape that was elongated in the displacement direction.

Actually, PSDs are relatively insensitive to the focus spot size, as long as it is a small fraction of the active surface (usually a diameter less than 1 mm is suggested). However, this only applies to near-circular spots. The effect of an asymmetrically-distorted spot is a loss of linearity in position vs anode current. Therefore, at positions on the PSD where the range sensitivity to spot position is high, corresponding to the farther ranges, the spot should not be distorted. At close ranges, spot distortion is more tolerable since the range sensitivity is so low. However, the maximum off-axis angle of the reflected light to the receiving lens should still be limited to about 12°.

For a right-triangle configuration, the maximum angle of incidence is governed by the minimum distance. From Figure 6.22 this angle is

$$\theta_{max} = \tan^{-1}\left[\frac{B}{R_{min}}\right] \tag{7.25}$$

From the limitation of $\theta_{max} < 12°$, the corresponding baseline is:

$$\begin{aligned} B_{max} &= (152 \text{ mm}) \tan(12°) \\ &= 32 \text{ mm} \end{aligned} \tag{7.26}$$

A larger value of B would result in an unacceptable amount of spot distortion in the focal plane.

Using the previously determined goal of $Bf=1600mm^2$, the resultant focal length with B=32 mm is:

$$f = 50 \text{ mm}$$

This value does not conflict with f_{min}=15mm, as determined from the lens diameter.

Choice of the Receiving Lens

For collection of light, the receiving lens diameter was chosen to be 15 mm. From geometrical considerations, the focal length was calculated to be 50 mm. The actual lens order was a 15 x 55 mm plano-convex lens from Edmund Scientific.

The lens-PSD distance was adjusted to about 51 mm, so that the spot distortion was limited at the larger angles. This made the far-range spot a bit blurred (but uniform), although that is acceptable for uniform spots. Another point should be mentioned about the lens-PSD distance. The focus spot is recommended to be small (under 1 mm diameter), but the power density is recommended to be less than 10 mW/cm^2. At power densities above this, nonlinearities in the generation of photocurrent will result.

In summary, the receiving lens was mounted so that its effective specifications are:

$$f = 51 \text{ mm}$$

$$D = 55 \text{ mm}$$

The Displacement Bias

In mounting the PSD relative to the receiving lens, the two triangles must lie in the same plane (Figure 6.22). However, the remaining geometrical parameter is where the displacement swing of the focus spot will lie on the PSD active surface.

As calculated earlier, the usable part of the active surface is 11.25 mm out of the 15 mm. Actually, for Bf=1600 mm^2, the displacement interval is about 10 mm, from equation 7.20. Therefore, there is some room to spare on the PSD active area (which is at a cost of accuracy). Since the sensitivity of range to displacement is very large at low angles of incidence, which corresponds to large range, the corresponding spot position should fall in the linear range of the detector. Also, with the right-triangle configuration, the spot displacement is limited on one side to reflected light from infinity (θ=0). Figure 7.8 illustrates the geometrical parameters associated with the PSD. As range is decreased from infinity, the angle increases from zero, and the spot displaces away from the emitter a total of 10 mm at the minimum range of 0.5 feet. By choosing the bounded side of the displacement swing at 3.75 mm from one anode, it is ensured that the far range focus spot will lie well into the linear portion of the detector. At close range, the focus spot is displaced away from the emitter, towards the other PSD anode. With this arrangement, the spot will traverse off the scale at ranges slightly below 0.5 feet, and some errors may be introduced from the spot being so close to the anode near the minimum range. However, this is acceptable since the accuracy at the close ranges is well within the requirements. These sacrifices were made for the purpose of obtaining the predicted accuracy at the far ranges.

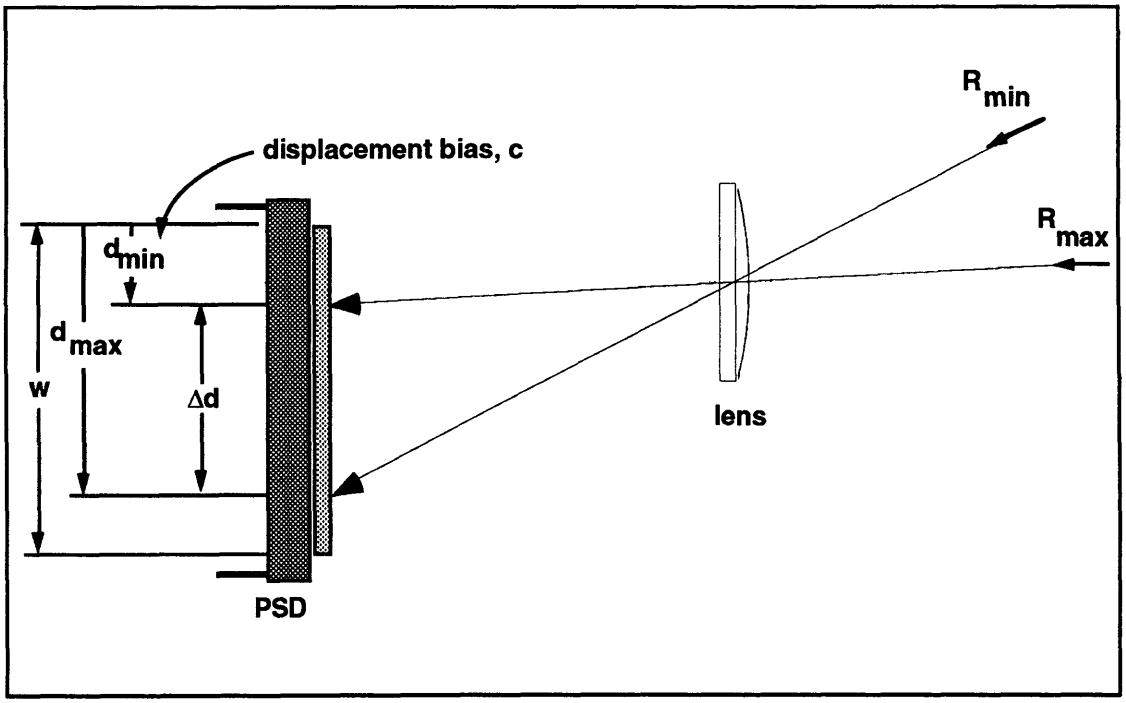

Figure 7.8: Geometric parameters associated with the PSD mounting.

The bounded side of the displacement swing is at 3.75mm from the anode that is closest to the emitter, and so the focus spot can never get closer than this. This distance is referred to as the displacement bias, c.

It should again be stated that the choice of c=3.75 mm was rather conservative, and could have been much less. In fact, to stay within the linear range of the detector active surface, c could be as low as 0.75 mm. However, due to uncertainties in the component mounting, and in machining, c was increased to give room for errors. Future designs can, and should, use lower fractional values of c so the PSD is used to its fullest, allowing a larger Bf product, and hence better accuracy.

7.1.10. Optical Filter

By using a laser as the illumination source, optical filtering with a very narrow bandwidth can greatly decrease the ambient noise incident on the detector. These passive devices can easily offer bandwidths that are of only 10 nm. Hence, the noise that passes is restricted to that which falls within the bandpass of the filter. In optics, these narrow band filters are called interference filters, and are commonly available in a large range of center wavelengths, bandwidths, and sizes.

The main drawback to these filters is that the laser light is also attenuated by at least 50 % typically. However, in a noisy environment such as in sunlight, this is an easy penalty to take, since the overall S/N is improved.

In choosing the filter, it is important to not only match the center wavelength of the laser diode, but also accommodate the laser's spectral width, so that the laser light is attenuated as little as possible. However, the filter bandwidth should not be so large that the additional noise lowers the S/N. Hence, the filter should be 'matched' to the laser, with their center frequencies and bandwidths coinciding.

One important property of these filters is that the center wavelength shifts with the angle of incidence of the reflected light. This can create a problem with narrow bandwidths, since it is then very easy to attenuate the laser significantly. However, this phenomenon can also be used to an advantage in a right-triangle configuration of the rangefinder, to reduce the signal strength at close range. This is advantageous to reduce the dynamic range of the returned signal.

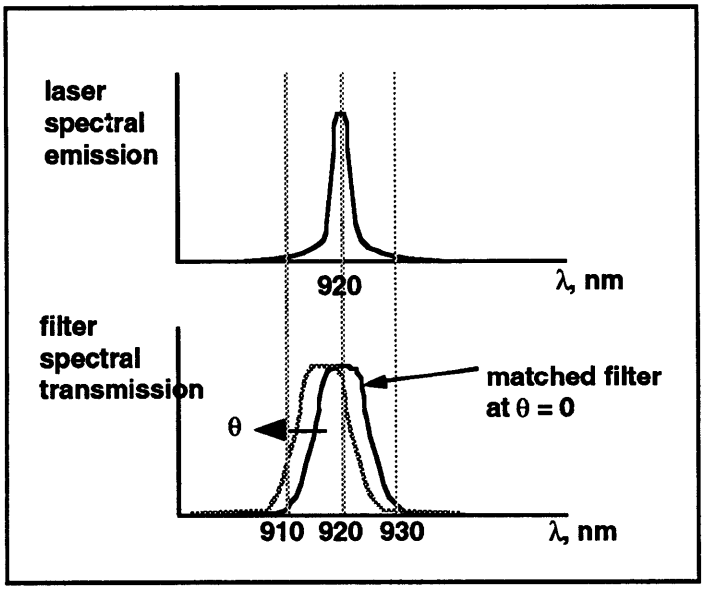

Figure 7.9: Laser spectral emission and filter spectral transmission. As the angle of incidence increases from zero, the filter transmission curve shifts towards the lower central wavelength.

An interference filter from Andover Corporation was purchased for the prototype rangefinder. This filter has a center wavelength λ_o=920.4 nm, a FWHM bandwidth of 10

nm, and an effective index of refraction of n_e=2.05. For an angle of incidence, θ, the new center wavelength is:

$$\lambda_o(\theta) = \lambda_o \sqrt{1 - \left[\frac{n}{n_e}\right]^2 \sin^2\theta} \tag{7.27}$$

Where n is typically unity. At the closest range of 0.5 feet, the angle of incidence is about 12°. so that the center wavelength of the filter is

$$\begin{aligned} \lambda_o(\theta_{max}) &= 920(0.995) \\ &= 915 \text{ nm} \end{aligned} \tag{7.28}$$

Hence, the filter center wavelength is reduced by 5 nm. In this shift, the bandwidth stays approximately the same. From the transmission profile of the filter, and the spectral emission of the laser, the resultant transmission due to this 5 nm shift is an attenuation of approximately 80%. The result is that in the very close ranges, the irradiance of the returned signal starts to deviate from the inverse square law, which prevents the curve from 'blowing up', as shown in Figure 7.10. This shows how the attenuation properties of the optical filter can be used not only to improve S/N, but also to reduce the dynamic range of the returned signal.

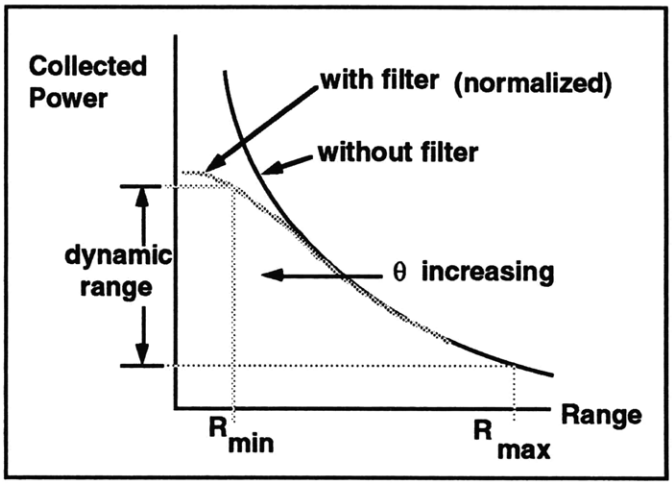

Figure 7.10: Illustration of reflected irradiance as a function of range. With the filter in place, attenuation becomes large at close distances due to the higher angle of incidence.

7.1.11. Summary of Prototype Rangefinder Design Specifications

As the previous sections have shown, there are many considerations involved in the design of the prototype range finder. However, the design process is rather straightforward. Before discussing the mounting, a summary of the chosen components is presented in Figure 7.11.

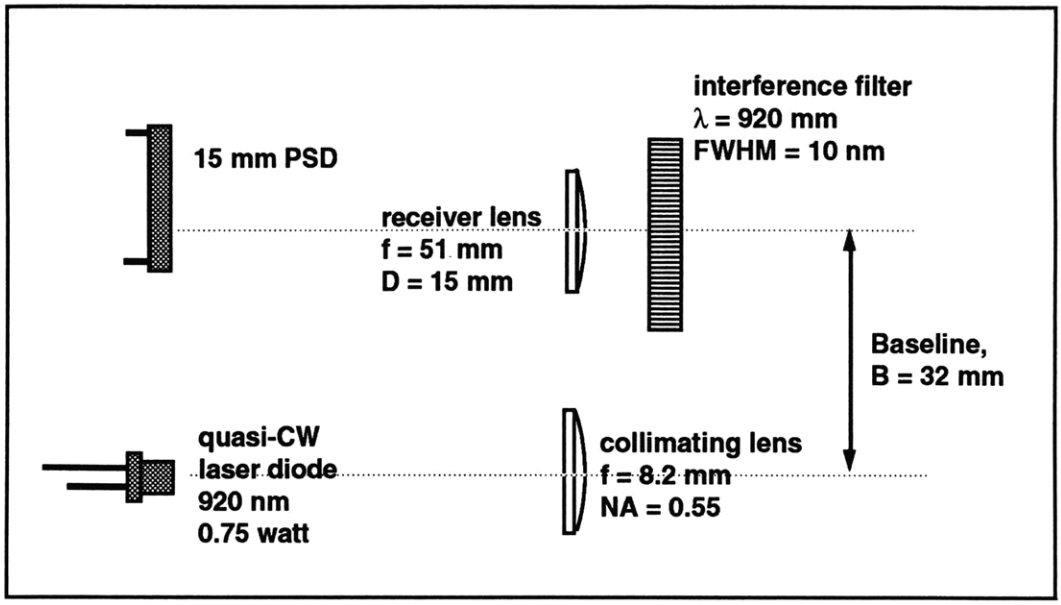

Figure 7.11: Component specifications for the prototype range finder.

The range finder was machined out of aluminum in three main pieces:

1. Main body
2. PSD mount
3. Laser mount

The receiver lens and optical filter mounted directly to the main body. One side of the receiver lens has a planar surface, and a very small recess was bored into the aluminum to ensure good alignment, without reducing the effective aperture size. The optical filter was also mounted concentrically, in a similar manner.

The PSD mount was made to hold the PSD flat against the surface of the rear main body. This not only provided proper alignment, but also a way to seal the detector from the dirt, water, etc. Adjustment slots were provided to allow the PSD to shift along its sensing direction, so that the displacement bias, c, could be adjusted.

Finally, the laser mount had a flat surface to keep the laser beam parallel to the receiver optical axis. The mounting block had a hole bored and reamed, so that the collimating

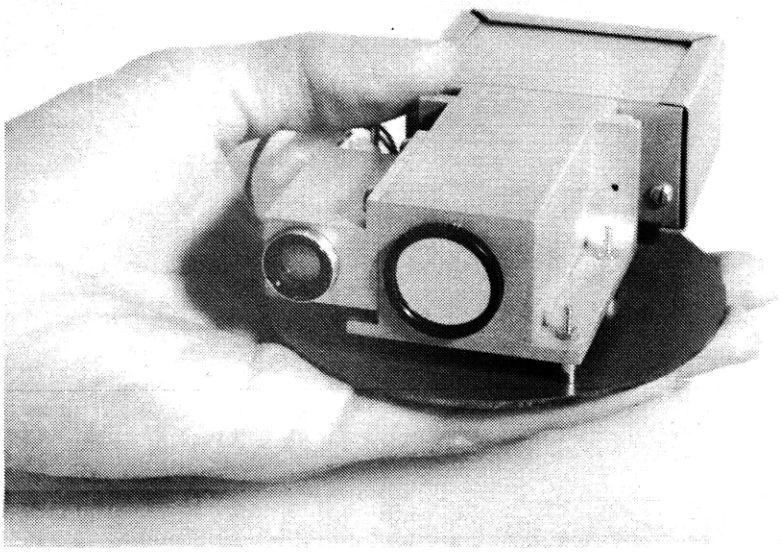

Figure 7.12: Photograph of the prototype range finder

lens had a snug fit. The laser diode actually mounted to a separate 'piston' which was made out of brass, for heat sinking. This also had a snug fit, but allowed the laser to be adjusted axially. In general, the laser should always be able to be adjustable in all directions, but mainly axially. This because beam collimation is very sensitive to the laser-lens distance. Hence, the axial position was adjusted in the lab, and set in place with epoxy.

The range finder is shown in Figure 7.12. The box in the rear contains the receiver circuit, which will be discussed next. The laser firing circuit, which fits in a similar box, is not shown. The laser is shown on the left, and the filter on the right. The receiver lens is not visible, as it is located behind the filter. As shown, the entire range finder (which was purposely designed large) fits easily in one hand. Size reduction is discussed in chapter 8 (future work).

7.1.12. Electronics

The associated electronics consists of the laser firing circuitry and the receiver electronics. For MITy-2, it was decided to sample the PSD through the A/D on the rover's central microprocessor, which is mainly dedicated to motion control. The microprocessor was therefore responsible for

1. Sampling the PSD for background light
2. Initiating the firing of the laser diode
3. Sampling the PSD during the laser pulse

Much of the electronics for these functions were designed and fabricated by Lorusso and Adam.

Laser Firing

It was suggested to use the microprocessor to initiate the fire, but not the pulse length, for reliability purposes. Therefore, an analog circuit was constructed to supply a pulse of desired length after receiving a digital signal from the microprocessor. With this circuit, any software errors that may cause the digital signal to remain high for too long will be overruled by the analog circuit, which was designed to limit the signal to 1 ms. The circuit illustration is shown in Figure 7.13.

When the flag is received, a current pulse of 2 amps flows through the laser diode. This produces an optical power of 0.77 watts. Actually, the energy conversion efficiency is higher at the maximum rating of 3 amps (1.35 watts optical power), partially due to the high threshold current of 0.8 amps. However, 0.77 watts was sufficient for the given noise levels.

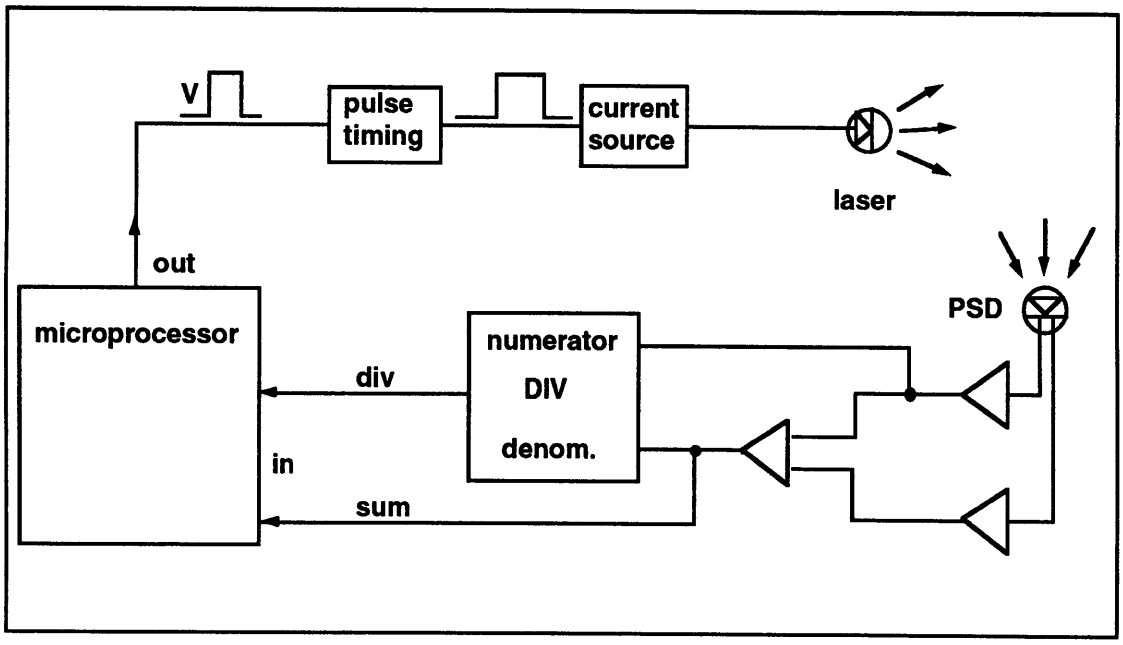

Figure 7.13: Range finder circuitry.

Optical Detection

The critical part of the range finding process is detection of the weak return signal. A slightly modified PSD circuit was used, due to the lack of A/D channels on the microprocessor. The most important part of optical detection is to keep the induced noise levels to a minimum. Therefore the weak anode currents from the PSD were confined to a very short lead length before the first stage of amplification. Also, the detection circuitry was shielded by an aluminum box and attached directly to back of the PSD (see Figure 7.12). Attention was also paid to ground loops, so that the large current pulse did not cause a false reading due to the different ground levels.

As shown in Figure 7.13, the current from each anode is changed to a voltage with a transimpedence amplifier. The gain on these was adjusted to just under saturation at close range, with some room for ambient noise. Following the first stage of amplification, the two anode values were summed. Then, a divider chip was used to normalize A by the sum (A + B). Normalization is important so that good resolution is maintained during A/D conversion. The microprocessor samples:

1. The divider output: A/(A+B)
2. The sum output: A+B

The equivalent focus spot position can be calculated using only the divider output. However, the sum was also read for two reasons. First, when background light is detected, subtraction can only take place with the individual PSD anode values, and not their normalized values. Therefore, the sum is needed to convert the divider value back to absolute numbers. In addition, the sum can also be used to detect when the divider reading is good or not, since this chip will function even when the signal is zero, since noise is also present. Hence, the sum can provide a way of assigning a level of confidence to each reading. One drawback to reading the sum values is that the dynamic range is large, resulting in either low resolution or premature saturation. Fortunately, this situation is helped by the optical filter attenuation (Figure 7.10).

As mentioned earlier, future improvements can include AC filtering, analog sampling, and CMOS circuitry. Not only can the circuitry be miniaturized, but the improved S/N can allow much weaker lasers to be used.

7.1.13. Preliminary Range Finder Test Results

Early results have demonstrated range finding ability with acceptable accuracy. These tests were performed indoors, with a surface reflectivity of about 0.1. Figure 7.14 shows the test data with linear interpolation, along with a predicted curve based on the triangulation geometry. The horizontal axis is represents the focus spot displacement, and is essentially the divider output after A/D conversion.

The predicted curve involved use of the displacement bias (c) and the Bf product, whose values were modified slightly to fit the test data (these values were not known with a high degree of accuracy anyway). This shows that the data follows the predicted curve to a certain extent. For actual range finding purposes, it doesn't matter that the test data fits any particular curve, since a look-up table can be used once calibration is performed. In this case, it is then important that the data is repeatable.

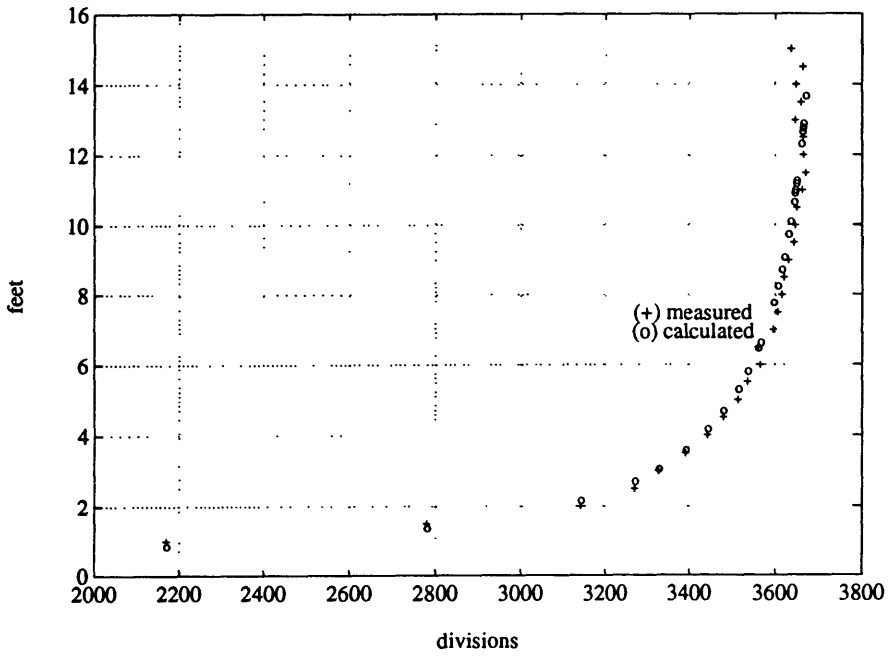

Figure 7.14: Range vs PSD focus spot displacement. The displacement is in units that are proportional to the divider output of Figure 7.13.

In summary, the range finder developed for MITy-2 has demonstrated ranging in the required range of 0.5 to 10 feet. Accuracy appears to be within specification, and will be characterized in more detail during MITy-2 tests in the summer of '93. The preliminary test data shows confirms that a small laser range finder can be made in a compact package, and that the design methodology discussed can be applied to further miniaturization and optimization of similar active triangulation systems.

7.2. PROTOTYPE MICROROVERS

7.2.1. MITy-1 Hazard Avoidance Sensors

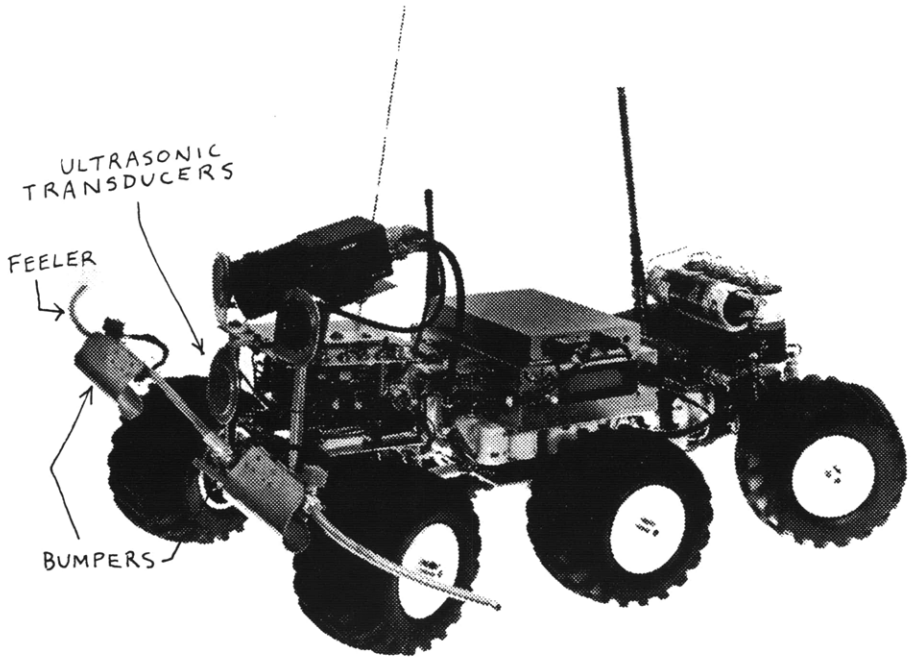

Figure 7.15: MITy-1 prototype microrover

The first prototype, MITy-1, was designed to function as an autonomous rover by sensing and reacting to the environment while traveling towards a goal. The sensors used on MITy-1 were not chosen to be applied towards a space-qualified vehicle, but were chosen mainly on the basis of cost and availability.

The MITy-1 hazard avoidance sensors are:

1. Ultrasonic range finders (3)
2. Feelers (2)
3. Contact switches (3)

The design started with a need for a low-cost range finding system. In the robotics community, that essentially limits the choices to ultrasonic rangefinders. Because of that, the tactile sensors were designed according to the field of view of the ultrasonic range finders.

It should be noted that the processor used for MITy-1 was chosen based on familiarity, even though it was under powered. Although the limited capabilities led to many headaches during testing, it forced the team to accomplish a difficult robotic task with very limited sensing abilities.

Ultrasonic Range Finder

The most common range finder in the mobile robotics community is a time-of-flight (TOF) acoustic sensor. As discussed in section 6.2.3, TOF devices require very fast electronics when the signal is electromagnetic. However, the speed of sound in air is about one million times slower than the speed of light, making acoustic TOF much more feasible.

Polaroid made this device common as an autofocus sensor on their cameras. They have also made the sensors available separately for OEM and research applications. A designers kit was purchased, along with some extra transducers and ranging modules. The transducer is the actual sensor, and is given the name because it transmits and detects the signal, using electrostatics to transform the energy between the electrical and acoustical domains. The ranging module interfaces to the transducer, providing the electrical waveform to the transducer for firing, and providing the necessary detection electronics.

The carrier frequency of the sound pulse is about 55 kHz (swept from 50 to 60 kHz). This is low enough to maintain a sufficiently low attenuation through the air, yet high enough to achieve a sufficiently-narrow beamwidth. Side lobes are relatively large, which can be troublesome at close distances. One of the reasons is due to the large λ/D ratio. Just as discussed with the laser beam, diffraction from the circular aperture causes the far field pattern to have a main lobe minimum at about:

$$\theta = 1.22 \frac{\lambda}{D} \tag{6.7}$$

The diameter of the transducer is about 35 mm, and the wavelength can be found from the frequency and the speed of sound in air, v:

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

$$= \frac{344 \text{ m/s}}{50 \text{ kHz}}$$
$$= 6.9 \text{ mm} \quad (7.29)$$

The main lobe minimum is

$$\theta = 1.22 \frac{\lambda}{D}$$
$$= 1.22 \frac{6.9 \text{ mm}}{35 \text{ mm}}$$
$$= 0.24 \text{ rad } (13.8°) \quad (7.30)$$

This is effectively the main lobe half-angle, defined by the first minimum. However, the beam angle is usually expressed as the -3dB point, which is slightly narrower:

$$\theta_{-3dB} = 12° \quad (7.31)$$

Hence the effective beamwidth is approximately 24°. Beyond this angle there is still an emission pattern due to the side lobes, although they are much weaker.

Because of the large beamwidth and presence of sidelobes, the Polaroid range finder cannot achieve good angular resolution. To make matters worse, the long wavelength makes most objects appear mirror-like, resulting in specular reflections. The result is a noisy sensor that has low angular resolution. This is especially a problem when the processor is limited, so that rapid firing and filtering is not possible. Nevertheless, the ultrasonic rangefinders provide high *range* accuracy when the sample is valid. Also, the large beamwidth is desirable on many platforms just to achieve a large angular coverage, although it is at the expense of angular resolution.

The specifications for a single transducer and ranging module are shown in Table 7.4.

Range	0.26 to 10.7 m
Resolution	3 mm at a range of 3 m
Beamwidth	24° full
Voltage Supply	6 V
Current	150 mA quiescent, 2.5 A (1 mS) per pulse
Operating Temperature	-30° to 70°
Weight of Transducer	8.2 gram
Weight of Ranging Module	18.4 gram
Cost of Transducer	$18 ea, in quantities of 10
Cost of Ranging Module	$24 ea, in quantities of 10

Table 7.4: Polaroid Ultrasonic Ranging Module Specifications

Operation

When ranging is desired, a logic-high signal needs to be supplied to the ranging module. This is called the initiation, or INIT signal. When this occurs, a 16-pulse chirp is produced in discrete segments of different frequencies, from 50 to 60 kHz. After internal blanking (nominally 2.38 milliseconds), detection then begins. Because range is a function of the elapsed time, gain is automatically adjusted accordingly. Detection is also improved by using the different frequencies, which are attenuated differently. When the module has detected an appropriate echo, a receive (REC) flag is immediately sent out. Range is then a linear function of the elapsed time between the INIT and REC logic signals. By knowing the speed of sound in air, $v \approx 20\sqrt{T}$ (T in °K, v in m/s), range is simply half of the round trip time, as shown in equation 6.19.

Timing can be performed either with an analog circuit, or digitally. MITy-1 measured the elapsed time in software, since timing functions were readily available. Since the clock had a resolution of 1 ms, this led to a range resolution of only 0.5 foot. Hence, the timing was the accuracy bottleneck. However, this was sufficient for simple hazard avoidance.

Integration on MITy-1

Three transducers (center, left, and right, as shown in Fig. 7.15)) were used on MITy-1. This is a very small amount for a mobile robot, and is the result of the limited number of analog and digital ports on the processor. Because of this, the large beamwidths of the range finders were desirable. The transducers were mounted such that azimuthal coverage was increased. Although only a 60° continuous coverage was possible, a larger

field of view was created by orienting the left and right sonars farther apart, at the expense of blind gaps. The gaps were maintained small enough so that the chance of missing obstacles was small. It was necessary to have enough azimuthal coverage to react to obstacles when turning. Ideally, a ring of sonars would be used, with a full 360° of coverage. Of course, this was not possible on the microrover, although it is common on many mobile robots.

Because of its TOF nature, close ranging is difficult, resulting in a minimum range of only about 40 cm without further adjustments to the ranging module. This is because the detecting electronics are purposely shut down immediately after firing to eliminate false detection due to ringing. This process is called internal blanking by Polaroid. It is possible to adjust the circuit to decrease the internal blanking time, making closer detection possible, down to about 16 cm. In any case, the blindness up close and the limited right and left coverage created a need for additional obstacle detection. This led to the implementation of tactile sensors.

Feelers

To supplement the range finding capabilities of the sonars, especially when turning, feelers were installed in the front, extending to the right and left (see Figure 7.15). The purpose of these are mainly to fill in the blind spots to the right and left, to avoid collisions.

When in a forward turn, detection of an obstacle is from the feeler rotation. This rotation is sensed with a potentiometer, located at the feeler pivot. The proximity of an obstacle can be determined roughly by the magnitude of the deflection. By sensing an obstacle in this way, the rover may still have room to steer the other way before a 'hard' collision occurs. Ideally the feelers would be longer and extend further outwards, but this can become mechanically troublesome. The chosen length of about 30 cm allowed just enough time to avoid hard collisions in most situations. The combination of feelers and ultrasonic rangefinders created a fair amount of coverage, minimizing the chance of collisions. However, collisions still occurred, and needed to be sensed.

Contact Switches

It is very important for the rover to know when it has contacted a rigid obstacle with sufficient force. Ideally, collision detection around the entire perimeter of MITy-1 would

provide the best form of sensing. However, this could also interfere with mobility. For simplicity, three 'bumpers' were made, all located in front. The same was actually needed in the rear too, but MITy-1 was mainly designed to travel forwards, with open loop control during the brief periods of rearward motion.

The location of the bumpers was based on the mobility of the rover. Since MITy-1 was not able to easily climb obstacles that were taller than the wheel diameter, bumpers were located in front of each wheel at the appropriate height. In addition, a bumper was located at the leading edge of the front platform, since this area offered no mobility and was thus prone to collisions.

In order to prevent the left and right bumpers from interfering with mobility, these were made passively retractable. The retraction also helped the fact that there was a delay from the time of contact to the vehicle stopping, so that the bumpers automatically moved out of the way. The feelers were also mounted to the same mechanism.

Contact was sensed through miniature mechanical switches. Microswitch (a division of Honeywell) makes a large variety of switches, with many options of voltage, current, force threshold, travel, size, etc. The bumper pads were made large, so that sufficient force applied to nearly any point on the pads resulted in a switch closing. This was achieved by hinging the bumper pads, and using two switches per pad. For each pad, the switches were wired in parallel, in the normally-open configuration, so that any one closure would indicate a contact. Each pair of switches were wired to one digital port which used an internal pull-up resistor, keeping the signal at +5 V when open, and down to ground upon contact. Hence, the switches were easily interfaced to the processor.

Hazard Avoidance Performance

Using the three methods of obstacle detection (acoustic range finding, feelers, bumpers), MITy-1 was able to demonstrate hazard avoidance and path planning effectively, considering the small amount of sensing. This is explained in depth by Malafeew [1], who wrote the algorithms for the rover's control system. Sensor placement was shown to be a critical issue, as demonstrated through testing and simulations.

The ultrasonic transducers were located rather high, so that the beam would not quickly diverge into the ground and detect rough surfaces as an obstacle. For the same reason,

the transducers were angled slightly upwards, so that the lower edge of the beam was nearly parallel to the ground. Of course, the ill-defined beam pattern made alignment rather difficult. If pointed too high, many obstacles would be missed. If too low, false detection from the ground would occur. The performance was best in a two-dimensional environment, which was a flat smooth terrain, with very tall obstacles. In this case, obstacle detection was greatly simplified, due to the absence of the height problem, and the absence of ground reflections.

As previously mentioned, the large beamwidth facilitated the azimuthal *coverage*, but degraded the angular *resolution*. Hence, even tall obstacles were difficult to locate sometimes. This matter was made worse by the specular reflection problems. When the objects were smooth, detection did not occur unless a portion of the beam was incident at an angle near the surface normal. This became a problem on such obvious obstacles as traffic cones and walls! This behavior clearly demonstrated a need for range finding with a smaller wavelength (see section 6.1.3). Of course, the specular reflection problem could have been improved by using more transducers, but this was not possible on MITy-1.

The result of poor detection was a shorter obstacle clearance, and sometimes collisions. The feelers were able to provide a 'last chance' to react to obstacles before a hard contact. For instance, when the range finders missed an obstacle, the feelers could perform a mechanical version of extended sensing, allowing the rover to steer the other way. Although somewhat crude, this was an important supplement to the acoustic ranging.

Of course, the feelers do not contribute to detection in front of the rover, so that collisions could also occur just from the limited coverage (or detection errors) of the acoustic range finders. The bumpers, then, detected collisions in the three locations described. The center bumper was important for objects (or portions thereof) that were narrower than the space between the front wheels. Any object taller than the ground clearance would be detected by the center bumper. The left and right bumpers, located in front of each front wheel, could also distinguish obstacle heights. The bumper heights were adjusted based on the climbing ability of the rover, so that any object that was lower than the bumper pad was not detected at all, allowing the rover to blindly traverse it. This feature was rarely used, however, since testing was mainly in a two-dimensional environment.

The hazard avoidance worked best when the acoustic sensors properly detected the obstacles. When this occurred, the rover could maneuver through a series of obstacles smoothly, and without stopping. Feelers often supplemented these efforts, though. Occasionally, a collision would occur, which caused the rover to stop and reverse a small distance in an open loop control manner. To help with detection, the obstacle courses consisted of a flat ground and tall, nearly vertical obstacles. This two-dimensional case is the typical scenario with most indoor mobile robots, and was used for MITy-1 testing due to the limited amount of sensing.

Summary

The goal of MITy-1 was to obtain feedback from the environment, and react autonomously to that information. Despite the very limited processing available, the simple sensing scheme allowed successful demonstration of hazard avoidance in a two-dimensional obstacle environment, including a demo at a rover exposition in Washington DC in September 1992.

From the experience with MITy-1, it was clear that a higher degree of sensing was required. One of the main sensing problems was due to the detection of obstacles with the Polaroid ultrasonic range finders. These were primarily due to the large beamwidth, the presence of significant side lobes in the beam pattern, and the specular reflections due to the long wavelength. Hence, it was clear that future prototypes would require a larger azimuthal coverage, although with higher angular resolution. Also, any active ranging system should definitely have a smaller wavelength to ensure diffuse reflections.

The tactile sensors functioned properly, although the feelers and the retracting mechanism made the rover less robust. These were required on MITy-1 due to the limited number of acoustic sensors that could be used. However, it appears that a better solution to close-range sensing could be through the use of non-contact sensors, so that the only protruding portions of the rover are the wheels. Detection directly in front of each wheel may be difficult, however, and the sensing need is dependent on the mobility of the rover.

7.2.2. MITy-2 Hazard Avoidance Sensors

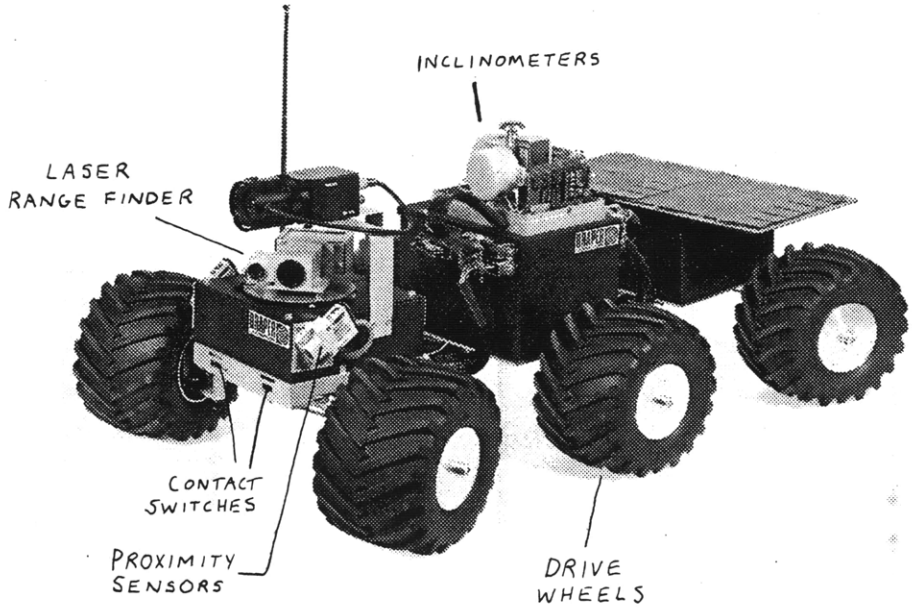

Figure 7.16: MITy-2 prototype microrover.

Much of the previous discussion of the final rover design has been focused on the range finder, which is a critical element in the overall hazard detection scheme. In fact, the prototype range finder design for MITy-2 was already discussed in detail. It remains, then, to discuss all hazard avoidance sensors for MITy-2, and their integration into the rover. To restate, the proposed hazard sensors are:

1. Range finder
2. Short-range or proximity sensors
3. Contact sensors
4. Inclinometers
5. Drive/drag wheels

Note that the last two items are also required for navigation. It should also be noted that the sensors used on MITy-2 are representative of what can be used on a planetary rover, but are not necessarily the proposed sensors for the final design. However, the arrangement of the sensors is a critical issue which can indeed apply towards the final design.

Rangefinder

As discussed in section 7.1, the prototype laser range finder has the following specifications:

Range	0.5 to 10 feet
Accuracy	within 6.5 % of range
Beamwidth	9 mrad
Optical Power	0.75 watts
Pulse length	1 ms

Initially the goal was to multiplex through a number of hard-mounted, miniaturized range finder modules. However, the lack of time led to the prototype range finder above, which can scan in azimuth to achieve the necessary field of view. For simplicity, the entire ranger is mechanically scanned by means of a small servo motor. This allows a front field of view of 180°. Actually, since the beam is tight, and the scan is only in azimuth, the scanned view can be considered a plane. The location of this plane is directly above the wheels, to prevent their obstruction.

The placement of the ranger above the wheels is an important issue. This places the 'plane of view' above the wheels, with the option of a higher mount, but not a lower (for full 180° coverage). By doing this, the ranger is essentially blind to obstacles below this level. Unless other detection is provided, the rover will proceed to traverse obstacles below the wheel height. This is one of the advantages of having a rover with a high degree of mobility, since the sensing needs are reduced. However, even though the rover has the *ability* to climb such tall obstacles, such motion imposes a greater hazard risk, and requires significantly more power. This seems to call for another plane of view, at a lower level than the main rangefinder. Unfortunately, it will be difficult to sense at this level on future designs, since the platform level will likely be above the wheel height. The simplicity of the plane of view system thus has its drawback in being blind to objects that stand shorter than the wheel diameter. However, only a depth map at various elevations can provide such information, which places large demands on processing. It may be important, then, to use other information to determine the 3-D environment below the range finder's plane of view.

Without a depth map over a *field* of view (not just a plane of view), it becomes a difficult issue to understand the terrain ahead. But, as mentioned, even if a complete depth map was available, the processor would not be able to handle such data flow. Therefore, the current range finder must be interpreted as best as possible. One of the largest difficulties in a 3-D environment is the misinterpretation of range data. For instance, an easily traversable hill can be interpreted as an obstacle, or vice versa (Figure 7.17). The same can be said in any case when the slope is changing, such as when coming onto or off of a rock. In fact, pitching and rolling motion of the front platform should be monitored so that at the time of ranging, these angles are known, and the orientation of the beam can be determined. Otherwise, the range data can be meaningless.

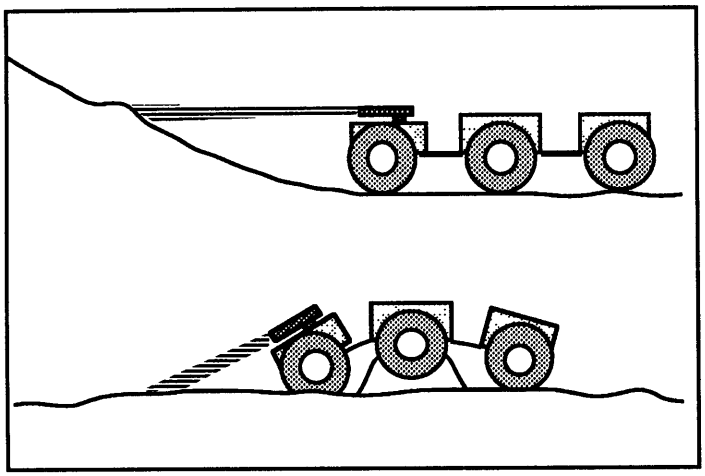

Figure 7.17: Pitching motion can make the sensor data difficult to interpret.

Because of these problems, it is desirable to keep the plane of view as high as possible to minimize the effects of pitch at close range. However, this makes the rover blind to close objects, resulting in a need for a higher degree of mobility. Hence, the plane of view should be kept as low as possible (directly above the wheel height), and the range data should fully utilize pitch and roll data of the range finder (which is not provided on MITy-2 at this time). This low position will make the range data more subject to misinterpretations, but will allow better discrimination up close.

With the range finder so low to the ground, the effect of pitch on the instantaneous view can be significant at distances that are much larger than the ranger's height above the ground. For example, at the maximum range of 10 feet, only 3° of pitch can be the difference in looking at the ground or the horizon. This is one of the main reasons that

the range finder was designed to sense only out to 10 feet. In summary, the range data is much more easy to interpret up close.

The location of the range finder on MITy-2 is basically in the center of the middle platform, directly above the top of the 6 inch diameter wheels. Because the range finder is recessed from the perimeter of the vehicle, the minimum sensing distance of 5" (as shown from experimental data) effectively results in object detection at a maximum of 3" from the outer edge of the rover. In other words, sensing objects as close as 3" from the rover can be ensured.

Although the instantaneous field of view is only a few milliradians, the scanning of the entire device creates a 180° view in front of the rover. Obstacle avoidance simulations by Malafeew show some advantage of a larger view, but the servo is limited to this range. The servo provides a scan rate of about 90°/sec, which is sufficient for the slow-moving rover. Since there is no range finder on the rear (to minimize complexity), a much simpler detection is provided, which will be discussed shortly. The primary ranging goal was to achieve a large scanned field of view in front, for efficient path planning, while still maintaining good angular resolution. This was accomplished with the scanned prototype range finder.

Proximity Sensors

In order to have some degree of non-contact sensing in other locations on the rover, without the complexity of another range finder, proximity sensors were obtained. These offer sensing at short range only, and are based on returned signal intensity methods (section 6.2.3). Proximity sensors are different from range finders in that they offer only one bit of information: whether or not an object is present within a certain range.

The need for short range sensing that was not provided by the range finder was:

1. To detect vertical drops in front of the rover
2. To offer simplified non-contact sensing when traveling backwards

As mentioned earlier, the vertical drops can be detected through a combination of both non-contact sensing and downward pitch, both on the front platform, as shown in Figure 5.4. The rearward sensing can offer obstacle detection at close range in the path of each

wheel, where there are no bumpers. In the latter case, proximity sensors can act as an extended bumper.

There are a large variety of proximity sensors on the market, primarily for use in the manufacturing industry. This has led to low-cost, rugged, and compact sensors in a large variety of ranges, depending on the type of operation. Proximity sensors use the following techniques:

1. Break-beam
2. Retroreflective
3. Diffuse reflective

Break-beam sensors use a separately located emitter and detector. When an object passes in front of the beam, the signal never reaches the detector, thus indicating its presence. This method can provide the longest range, since the beam can be well-collimated.

Retroreflective operation is similar to the break-beam technique, except that the detector can be replaced with a high reflectance surface, allowing a strong returned signal back in the direction of the emitter. The detector can then be placed adjacent to the emitter. A variation of this is to place the reflectors directly on the target object, in which case signal detection corresponds to the presence of an object.

Of course, both the break-beam and retroreflective methods require an environment that is custom designed, which is not possible when sensing in an unstructured environment. Diffuse reflectors offer this possibility, though, since no modifications to the environment are necessary. However, the reflected signal is mainly diffuse, so the received energy is small. In addition, detection is very sensitive to the target's surface reflectivity. The result is a rather limited range, since only up close is the sensitivity to reflectivity small (see 6.2.3).

Although there are many manufacturers of diffuse proximity sensors, an industry leader is Banner corporation. The chosen sensors are the Banner SM312D. Specifications for this sensor is shown in Table 7.5.

Manufacturer	Banner
Model	SM312D
Range	0.1 to 15 inch
Voltage Supply	10 - 30 VDC
Current	25 mA
Output	Source or Sink, 150 mA max
Light Source	Infrared LED, 880 nm
Operating Temperature	-20°C to 70°C
Response	1 ms minimum returned signal

Table 7.5: Proximity Sensor Specifications

These sensors have a gain adjustment screw, so that the maximum range can be adjusted, for a given reflectivity. Operation can be 'light' or 'dark', depending on if the output signal is to occur when reflected light is detected (normally dark), or when the light is not detected (normally light).

The SM312D proximity sensor provides a very high immunity to ambient light, using an AC filtering scheme (see Active Filtering, section 6.1.6). The oscillator that controls the modulation of the LED also is used in the receiver circuit, so that DC light sources (such as the sun) are not passed. In fact, preliminary testing showed no sign of false alarms even when pointed directly at the sun. Also, the sensors are sealed with O-rings, so that operation in harsh, wet environments is possible. Hence, outdoor operation should not be a problem.

Detection of Vertical Drops

Two Banner proximity sensors were mounted in the front as shown in 7.16. Under normal operation, the ground will reflect the signal back, indicating the presence of ground directly in front. When the ground suddenly drops below a certain level, a 'dark' condition will exist, at which point the rover *may* want to stop. 'May' was used because there will be many situations in a bumpy environment that can cause the sensors to point in such a way that the threshold distance can easily be exceeded. Therefore, it has been suggested to use inclinometer data on the front platform in combination with the proximity sensor data (Figure 5.4).

The inclinometer data is to detect when the rover is pitching down too severely, which will be the first sign of a vertical drop. This may seem dangerous, but the rover will should still have good traction on the remaining two axles. When the pitch is too severe in the downwards direction, the Banners can then be used to 'confirm' if the situation is indeed a vertical drop.

The bracketry for the Banner sensors were made adjustable in 10° increments about the nominal value, which is 45° from vertical. At the nominal position, the beam strikes the ground a few inches in front of the leading edge. The angle that works best will be determined from testing. In addition, testing will also be used to adjust the gain on the sensors so that the range threshold is appropriate, which is expected to be about 8 inches.

Obstacle Detection In the Rear

The Banners, as mentioned, will also be used as a simplified way of detecting obstacles when traveling backwards, which should be a small fraction of the time. Two sensors will be mounted above each of the rear wheels, pointing almost directly rearwards. These will act essentially like an extended bumper, detecting obstacles above a certain height before the wheels contact them. The middle portion will provide detection through the use of bumpers.

Contact Sensors

MITy-2 uses contact sensors in only two locations: at the leading and trailing edges. Other locations on the perimeter of the rover were not covered by tactile sensing to keep the mechanical complexity to a minimum.

Like MITy-1, the 'bumpers' were designed with a large contact surface, with electrical contact confined to within small mechanical switches. MITy-2 has a 'dead' area that is six inches wide in the front and rear (see Figure 5.3). An obstacle that fits in between the wheels can cause the rover to collide at these points, creating a need detection is necessary at the leading and trailing edges.

The bumpers are generally very simple devices, but it is important to design them to minimize the chance of sticking, which is a common problem in space vehicles. This requires the bumper mechanism to have a small amount of motion, yet still ensure

detection over a large area of contact. Instead of using any sliding mechanisms, a flexible piece of spring steel was used as a 'hinge' which held the rigid bumper. When force is applied to the bumper, it deflects slightly, constrained to motion in one axis. This force is transmitted to two small switches, which register contact after depressing a small amount. By using two switches, at least one is ensured to depress when force is applied anywhere on the bumper. This is similar to the method used in MITy-1, but the range of motion is much less.

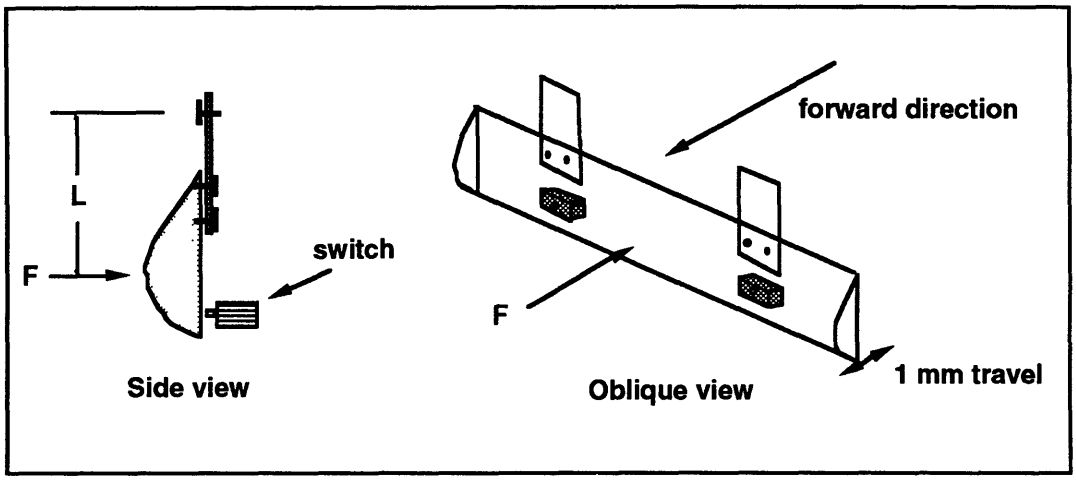

Figure 7.18: Contact sensor on MITy-2.

The force threshold is currently determined by the switch, which is made by Microswitch. This threshold can be increased by using additional elastic material. Adjustment screws are also available for this purpose. One final point is that stops should always be used with contact switches, so that the switch itself does not carry most of load in the case of excessive applied force.

Inclinometers

MITy-2 currently has Lucas inclinometers (section 4.2.3) available on the middle platform only (mainly for navigation sensors). However, if the accelerometers prove successful in providing tilt information then their small size can allow for placement on the front and rear platforms as well. The front platform needs both pitch and roll data so that the range finder beam direction is known. In addition, it would be helpful if all of the platforms had pitch and roll information. The reasons for these are:

Purpose	Platform	Required Sensing
Sun Sensor (nav)	center	pitch, roll
Rollover Detection	front, center, rear	roll
Range Finder Corrections	front	pitch, roll
Local Slope	front, center, rear	pitch, roll
Cliff Detection	front	pitch

Table 7.6: Possible uses for inclinometers

It is clear that pitch and roll information is useful on all three platforms. However, with the Lucas inclinometers, the large size and weight make additional sensors hard to justify. Therefore, pitch and roll sensors are suggested on all three platforms only if they are much more compact. The IC accelerometer (see 4.2.3) was mounted along with the Lucas inclinometers for a direct comparison. These accelerometers are not inertial grade (they certainly do not need to be), but are micromachined out of silicon and are very light and compact, as well as relatively low-cost. Specifications are also given in Table 4.5.

At this point, however, MITy-2 is equipped only with inclinometers on the center platform for navigation purposes, which can also be used to represent the entire rover for hazard avoidance. Because of this limitation, hazard detection is greatly reduced. Hence, MITy-2, or MITy-2a will have additional accelerometers installed if possible, so that pitch and roll will be available on each platform. Even though the power draw of each of the IC accelerometers is only 60 mW, this may further be reduced by operating each on a low power duty cycle. This involves powering up, reading, and powering down. The result can be a very low average power consumption, depending on the duty cycle. This same procedure can be used for many other components where a high sampling rate is not required.

Drive and Drag Wheel

The drive and drag wheels are primarily for navigation purposes. However, their information can be used for hazard avoidance also. The use of the drag and drive wheel odometry together can be used to detect two situations:

1. Loss of drive wheel traction
2. Detecting collisions that bumpers and proximity sensors missed

Loss of traction can be possible in a number of situations, as explained in section 5.2.4. Since the drag wheel will essentially not slip, this situation can be detected by differencing the drag and drive wheel speed data.

Collision detection is supposed to be accomplished by bumpers and proximity sensors, but an effective back-up for these is to monitor sudden changes in the speed of either of the front wheels. Torque sensing would also be valuable. However, MITy-2 does not have the ability to measure drive motor current (torque), but future designs may want to incorporate this sensing to know the magnitude of the wheel-ground interaction force. This data can be very important for sensing collisions, loss of traction, as well as soil characterization for scientific purposes .

MITy-2 has tachometers on all six drive motors, so that wheel speed is available. The drag wheel has not been implemented yet, but this will have an optical encoder for sensing rotation. The errors associated with each sensor can strongly affect the *navigation* performance, but hazard avoidance is much more immune to absolute accuracies in translation. Therefore, both the drag wheel and the drive wheels can offer meaningful information even over very irregular terrain. Testing of MITy-2 will be important for developing algorithms that use odometry data.

Summary

The hazard avoidance sensors used on MITy-2 are:

Scanning laser range finder (1)
Banner proximity sensors (4)
Contact sensors (2)
Inclinometers (2)
Drive wheel tachometers (6)
Drag wheel (1)

The MITy-2 testing program will help to define the algorithms for processing the sensor data, and to determine what changes need to be made. Although the above list represents what is currently on MITy-2 (or will shortly be on), the addition of inclinometers on at least the front platform will likely take place soon. A full test program of MITy-2 will

begin in June 1993. Testing at this point has involved successful demonstration of obstacle avoidance in a two-dimensional environment, with a definite improvement over MITy-1.

CHAPTER EIGHT
FUTURE WORK

8.1. FUTURE NAVIGATION WORK

8.1.1. Introduction

Computer simulations and test results from the prototype rovers will provide important feedback for the final design. The final design will have to additionally incorporate means of shock isolation, thermal control, radiation protection, and dust protection. These issues will be addressed starting in June 1993. Work needs to progress quickly, as the Mesur Pathfinder mission is currently scheduled for 1996.

The next few years will involve more prototyping, testing, and simulating. In the near future, testing on MITy-2 will be very important for determining the degree of success of the navigation system. Preliminary tests on a local football field have demonstrated the ability to use the dead reckoning information successfully, to within 10 % along a closed loop path. However, this testing was performed on a flat and relatively smooth surface. It is expected that much larger errors will be introduced when the terrain is irregular, causing the rover to continuously and randomly pitch and roll during travel. However, rough terrain will not be the only difficulty involved. The following presents some suggested future work for both the translation and heading sensors.

8.1.2. Odometry

Rough terrain will make all sensor measurements more difficult to interpret. Although stopping for zero updates (ZUP) will benefit heading measurements, odometry will always be affected by such terrain. There will have to be a large amount of simulation and empirical work dedicated to just the odometer readings, since the majority of the navigation errors will be the result of this. Research should include:

Drag Wheel
- Optimal drag wheel design
- Drag wheel loading
- Using the vertical motion information of the drag wheel

Drive Wheel
- Best use of the data from the 6 drive wheels
- Using current to measure wheel torque
- Applying corrections based on inclinometer data
- Sensor fusion of the drag wheel, drive wheel, and inclinometer data

Although odometry is one of the oldest forms of dead reckoning sensors, very little work has been done in achieving high accuracy in unstructured outdoor environments. Therefore, it is suggested that sufficient research is performed on this portion of the microrover.

Sensor fusion can be very important, since there are a variety of sensors that can be combined to yield a better estimate of the rover's state than any one type. Use of such techniques as Kalman filtering may be appropriate, to help reject poor sensor readings. In addition, one of the main advantages of the rover's operation scheme is having man in the loop on a daily basis. Therefore, it is possible to adjust the rover's control code depending on the information gained at the landing site, and from daily video transmissions. Provisions should be made to remotely modify the control code for such adjustments.

8.1.3. Gyro

Errors

For position estimation, it appears that the heading errors that are expected will affect the overall navigation less than those due to the odometer errors. However, heading is still a very important parameter, and should be maintained within 2° accuracy over the duration of travel. Detailed simulations should be developed as soon as possible, to quantify the accuracy requirements of the gyro.

The gyro is important for maintaining accuracy in between sun sensor calibrations. However, for redundancy it is desirable to have a much higher accuracy in the event of a sun sensor failure. In this case, the gyro should be able drift no more than about 5° or 10° in the typical 30 minute time of travel. With the sun sensor functioning properly, the gyro requirements are much more relaxed, to about 1 to 2° per minute, since updates can be made frequently and quickly. Therefore, a higher gyro accuracy is mainly beneficial for the times when the sun sensor data is not available. When this is the case, updates from the sun sensor are not possible, but calibrations of a known rate (zero) are possible. Hence, stopping for the calibration of heading are for two reasons:

1. To update the integrated heading angle from the actual angle as measured from the sun.
2. To measure the gyro drift at a known rate input of zero (actually the planet's spin rate), so that corrections can be made in hardware or software for future readings, until the next update.

Therefore, even in the event of a sun sensor failure, calibrations to the gyro can keep the cumulative heading error to a much lower level. If the drift of the gyro/integration circuit was constant, then only one update would be needed, and there would be no errors. However, drift varies with time, and one way to correct for it is to assume it is constant in the time in between updates. Furthermore, if the sun sensor is working, the updates can be based on an absolute reference, so that calibrations are much more beneficial. Future tests should incorporate these scenarios to determine gyro accuracy more closely.

Systron Donner Gyro

The suggested Systron Donner gyro, shown in Figure 8.1, should be purchased, and its performance characterized under a variety of vibration and thermal environments. Table 8.1 lists the QRS-11 specifications. Current information shows that this gyro should be able to achieve an accuracy of less than a degree/minute under no vibrations and a stable temperature. If the temperature is changed by 1°C, an angular drift of 0.3°/minute will occur. Therefore, temperature is a very important factor. At this point, it is not certain what the thermal capabilities of the final rover will be, so temperature monitoring may be important. In any case, frequent rover updates are still possible, so variations in the drift can have minimal effects depending on the frequency of updates, and how well the updates represent the drift during travel.

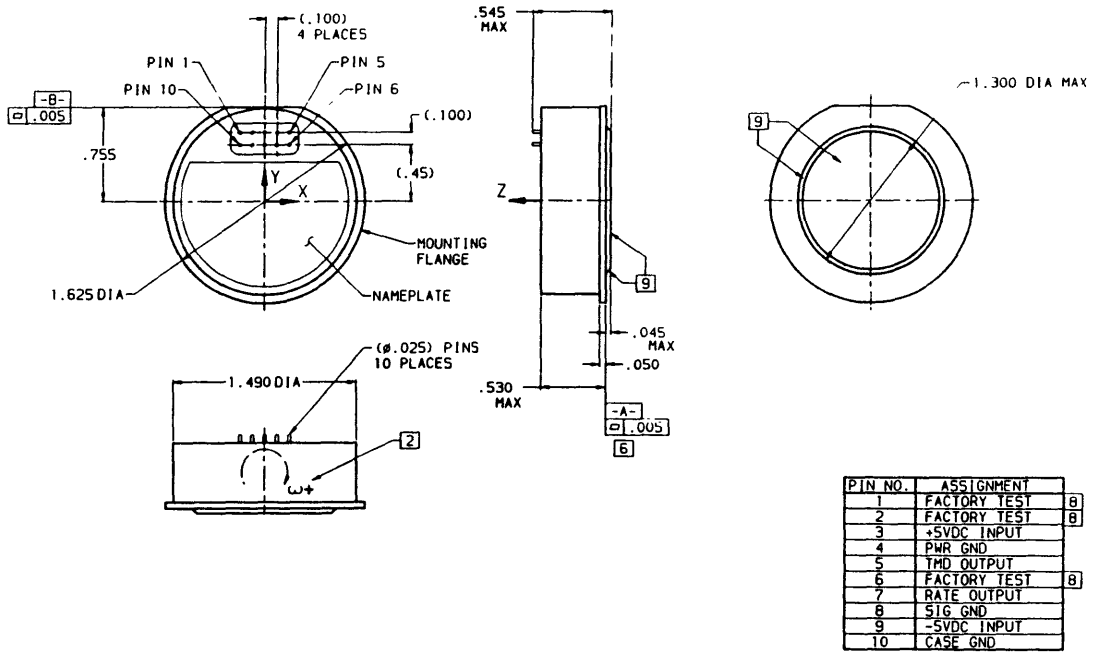

Figure 8.1: Systron Donner QRS-11 dimensions. (in)

Manufacturer	Systron Donner
Model	QRS-11
Range	100°/sec (other ranges available)
Voltage Supply	±5 VDC
Power Consumption	0.8 watts
Full Scale Output	±2.5 VDC
Bandwidth (-90°)	> 100 Hz
Scale Factor Calibration	< 1.0 % of value
- temperature Coef.	< 0.03 %/°C
Bias (short term)	< 0.3°/min
- temperature Coef.	< 0.3°/min/°C
- g sensitive	< 1.2°/min/g
Linearity	< 0.05 % of full range
Operating Life	10 years, typ
Operating Temperature	-40°C to +85°C
Vibration Survival	20 g rms, 20 to 2 kHz random
Shock	200 g, 2 milliseconds

Table 8.1: Specifications for Systron Donner QRS-11.

If tests show that the Systron Donner gyro is not satisfactory, the feasibility of using the C.S. Draper FORS, and the Litton G2000 gyro should be explored. This will likely yield a much higher performance, but the main question is the requirements of the supporting electronics; namely the size, power, and man-hours involved. Although the Humphrey sensor should be ready to interface, its large size and weight justifies spending effort on the FORS and/or G2000. Of course, by 1994 the Northrop MAG (and possibly similar sensors from other companies) should be available too. If this is the case, the MAG should be strongly compared to the QRS-11, as it is the only device that can compete in terms of size, weight, power, and cost. It should also be mentioned that the Charles Stark Draper Laboratory is currently working on a similar gyro as the QRS-11, which may be available in the near future.

Rate Integration

Although MITy-2 uses analog integration, it is strongly suggested to use digital integration for future prototypes, as well as for the final design. The reason for this is to minimize the angular errors from the integration process itself.

Digital integration has some important advantages over analog, including

- Stability over time
- Less hardware complexity

The stability is especially important when the temperature is not well-controlled, which results in scaling and bias errors with analog integration [27]. Also, hardware complexity should be minimized to increase the system reliability.

One of the main drawbacks of digital integration is the high sampling rates often required. However, the slow motion of the rover should allow low sampling rates, particularly when the turning rates are linear with time (or constant).

An alternative digital integration technique is the use of voltage-to-frequency converters. This device transforms the DC rate signal to a series of pulses whose pulse frequency is proportional to the signal voltage. Therefore, integration is simply a matter of counting the number of pulses in a given time. There are simple logic devices that take care of

digital pulse counting (such as those used with optical encoders), so that very little dedication is required from the processor.

8.1.4. Sun Sensor

Much of the discussion about the use of the gyro in case of a sun sensor failure is due to the possibility of the dust covering the optics. This issue has not been addressed at this point, although it has been shown that dust can significantly degrade the performance of any optical system. Dust is a problem on both the Lunar and Martian surfaces, but the mechanisms involved can be very different. Gilbert [1] presents possible solutions to the Lunar dust issues. These involve both active systems such as mechanical wipers and gas blasts, to passive coatings.

Another concern of future sun sensor work should involve the use of a CCD for position detection, instead of the current PSD used on MITy-2. The reason for this is to eliminate the effects of other strong light sources such as reflected light from tall objects, or diffuse light from the atmosphere. The effects of these secondary sources is not known at this point. However, if the performance is not sufficient with the PSD (even with optical filtering), then the feasibility of using a CCD should be explored. The drawbacks to a CCD include added complexity, and a potentially slower sampling rate.

In addition, the prototype sun sensor is one concept of many. The design that was described in 4.1 has the advantage of simplicity and small size, but other concepts may be explored (see section 2.2.3). For instance, it may be desirable to gimbal the sensor, so that leveling is passive, thus eliminating the need for inclinometer data for this purpose. Also, sensors can be designed to yield only heading, opposed to both elevation and heading.

It should also be mentioned that the size of the sun sensor can be reduced significantly. The aperture size only needs to be large enough to make the collected energy strong enough for a good signal to noise ratio. The high irradiance of the sunlight, then, favors the potential of a small lens diameter. The PSD or CCD are available in much smaller sizes, so that they can be matched to a smaller optics head. Therefore, it may even be feasible to have multiple sun sensors for redundancy, if they can indeed be made much smaller.

Accuracy can be increased not by increasing PSD size, since position sensing errors increase with size. The sensor area represents the entire field of view. Hence, improved accuracy is possible by using multiple sensors, each dedicated to a smaller FOV. This is not possible with the current design, but can be possible for other configurations. The drawback is a larger, more complex sensor, probably requiring multiplexing. However, accuracies can always be improved in this way.

8.2. FUTURE HAZARD AVOIDANCE WORK

8.2.1. Introduction

The Microrover Project has generated two prototypes at this time: MITy-1 and MITy-2. Mobility improvements on MITy-2 will be incorporated in the next prototype, MITy-2a. This prototype will be essentially the same as MITy-2, except for a new conical wheel design, and a new steering system (see Figure 5.3). The platforms will also be slightly different.

In order to progress rapidly, it is important that the remaining efforts be spent efficiently, so that the overall system performance can improve with the minimum amount of work. Hazard avoidance is arguably the most important aspect of a successful rover. The following sections present some of the important hazard avoidance hardware issues that should be addressed.

8.2.2. Laser Range Finder

The prototype range finder used on MITy-2 was constructed due to the need for a small ranger that used electromagnetic signals for non-contact sensing. The design process of the prototype range finder was discussed in section 7.1. The resulting active triangulation system was designed quickly for proof-of-concept purposes, and can be improved in many ways.

Size Reduction

First, the rangefinder can be reduced in size considerably. From section 7.1.9, it should be clear that much of the size of the MITy-2 range finder is due to the large focal length of the receiving lens. The entire geometry of the existing system is based on the fact that

the PSD has a large (15 mm) active surface. As seen from equation 6.26, if the PSD is reduced in size, d (the focus spot displacement) and Bf (the product of the baseline distance and the receiving lens focal length) can also be reduced the same amount while still maintaining the same accuracy as the larger unit. For instance, a 3 mm PSD is 20 % of the current PSD, so Bf can also be reduced 20 % without a loss of accuracy. The only difference is that component mounting will have to be done with a higher tolerance.

One may ask about the light collecting ability if the range finder is reduced in size so much. However, this is the reason that the prototype range finder used such a small diameter receiving lens (15 mm). Because the diameter is small, it can remain the same size in the miniaturized system, and so the same amount of light can be collected. So, assuming this lens diameter remains the same, and a 3 mm PSD is used, the Bf product can be reduced to 20 % of its original size. For example, the focal length f can be reduced from 51 mm to 20 mm, and the baseline separation can be reduced from 32 mm to 16 mm. It can also be assumed that the required laser optical power can be reduced with better detection techniques, resulting in a smaller active surface of the diode. This, in turn, can allow a smaller collimating lens. In short, the same performance can be expected out of an optics assembly that is about one-fifth of the platform area shown in Figure 7.12.

Discrete Scanning Through Multiplexing

If this is the case, small range finding modules can be placed at various locations on the rover, instead of mechanically scanning one unit (see Figure 8.3). The advantages of this method are:

1. Reduced mechanical complexity, since no motor or other moving parts are needed
2. Redundancy in the event of a component failure (such as a laser diode)
3. Scanning speed not limited by scanning servomotors

The disadvantages are:

1. Reduced short term coverage, due to discretized scan. Continuous coverage can be achieved over time by using the rover's own motion.
2. Close range obstacle avoidance more difficult due to inability to scan without moving

3. May be heavier overall, depending on the number of modules used.
4. Thermal protection may limit the placement of individual modules

Even in the scanning case, the miniaturization of the range finder would benefit the system, just due to its lower size and weight. Note that the orientation of the range finder baseline axis is not an issue as long as the beam is in the desired location. Therefore, the receiver can be placed to the left, right, above, or below the laser beam. However, since the laser beam defines the instantaneous field of view, this should be placed as low as possible (at the wheel height), so that the receiving lens is at the same level or above. For packaging, it may be advantageous to orient the module with the receiver lens above.

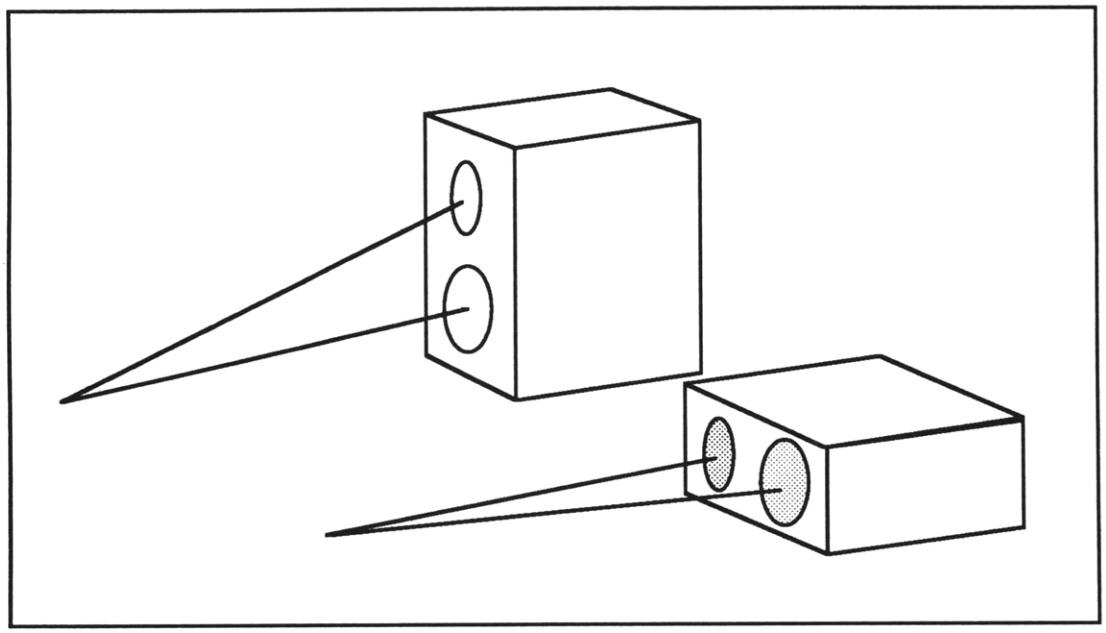

Figure 8.2: Possible orientations of the triangulation plane.

Electronics

An additional point that needs to be considered is the configuration of the firing and receiving electronics. Most of these can remain centralized, with one set used for all of the modules. This is achieved through multiplexing, which can be controlled by the processor. For instance, by using 4 digital channels, any one of 16 modules can be selected. However, some of the electronics will be needed for each module. These are mainly the transimpedance amplifiers, which transform the small PSD current to a voltage. By doing this, the PSD signal can be transferred to the centralized electronics with a much higher immunity to electromagnetic noise. Also, a capacitor and transistor

may be needed at each laser diode so that the large current pulse is confined to that area, which reduces its interference on other electronics. The remaining components should be able to be located as one unit in a centralized location.

Combined Discrete and Continuous Scanning

Another possible method of scanning can involve a combination of multiplexing and mechanical scanning. (see Figure 8.3) For instance, if the range finder on the existing rotation platform on MITy-2 is removed and replaced with, say, five ranging modules, the advantages of each method can be used. Multiplexing among the five modules can accommodate the large 180° field of view, while only a small amount of mechanical scanning (36°) can accommodate the continuous coverage that is sometimes needed for good angular resolution. Even if the scanning motor fails, sensing is then reduced to the discrete case, allowing effective sensing. Similarly, a laser diode failure can also be tolerated due to the remaining modules, and the ability of the motor to accommodate this failure through a larger amount of mechanical scanning. By reducing the amount of mechanical scanning to say 45°, versus 180° or more, the mechanical complexity can be further reduced through the use of limited rotation bearings. These flexible pivots are a frictionless, stiction-free bearing that can be used in applications with limited angular travel. Although they have torsional stiffness, these bearings permit no lubrication, and offer precise positioning and infinite cycle life. These are therefore used in a variety of space applications.

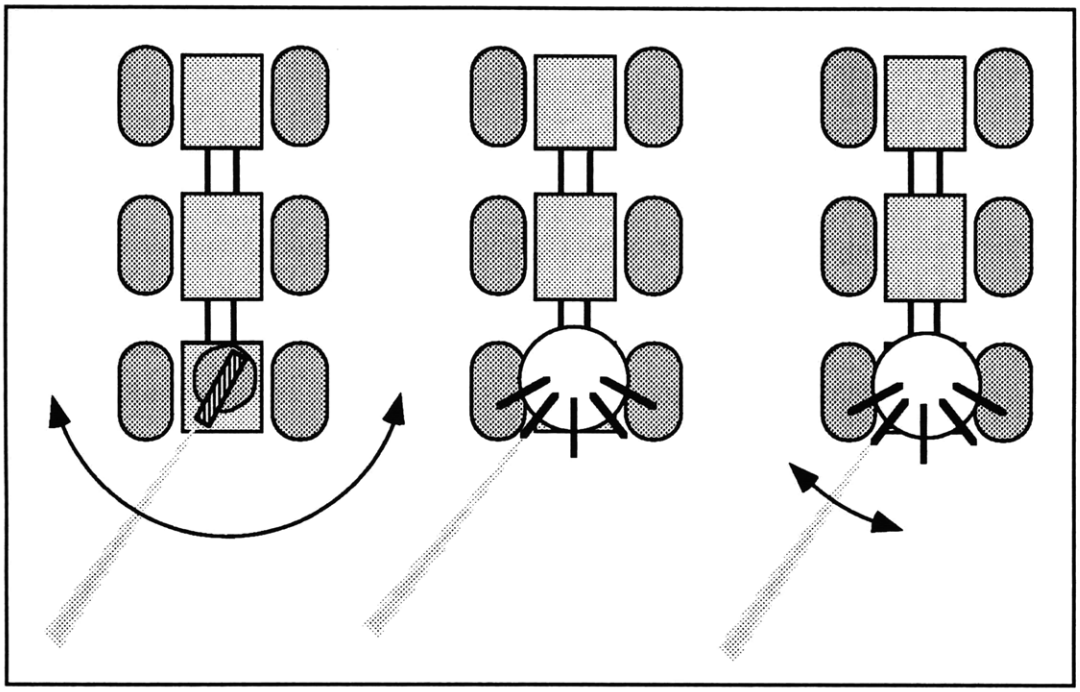

Figure 8.3: Continuous, discrete, and combination scanning configurations.

Filtering and Detection Improvements

No matter what form of scanning is used, there are other changes that can be made in future range finders to improve the system. The optical filter is an important element in reducing the amount of noise, particularly with a DC detection process in sunlight. The current optical filter, made by Andover Corp., is rather thick and heavy. If multiple range finders are used, this can be a problem. However, custom orders are available that utilize an ion plating process. These are much more rugged, and can be used in space applications. In addition, the resulting thickness is about 2.5 mm instead of 7 mm. It is suggested that optical filtering is used even with future improved detection processes, since the noise from sunlight is greatly reduced and the dynamic range of the returned signal is also reduced, both in a passive manner.

Improvements in the detection circuitry are certain, since the existing circuitry has a relatively large amount of noise, and uses no active filtering techniques (see section 6.1.6). By using AC detection methods, the signal to noise (S/N) level will be greatly improved, permitting the use of a much weaker light source. As mentioned, if the laser power can be reduced, the collimating lens can also be reduced, facilitating the miniaturization process. AC filtering may also permit direct sampling, without the need

to sample the background light first and then subtract. This will make the range finding process more efficient.

8.2.3. Proximity Sensors

The Banner proximity sensors used on MITy-2 have so far demonstrated very good performance in detecting the presence of objects, even in the presence of sunlight. Although these same sensors cannot be used in space, they have shown that a returned signal intensity method is successful at close range. Therefore, similar circuitry can be used to make a space-qualified sensor, also using AC filtering techniques.

These sensors should not be relied upon for accurate measurements of the range threshold, since changes in reflectivity or attenuation from dust can affect this. The effects of these changes are stronger at the farther end of the sensor range, but nevertheless, the returned signal intensity method is not very accurate. The strength of the emitted light should be strong enough so that even under expected attenuations or reflectivity changes, the error in the range threshold is acceptable.

If the returned signal intensity method proves to be unsuccessful for just proximity sensing, the miniaturized laser rangefinder modules can be used, which can provide range information with good accuracy. However, these are more complex. In addition to the added complexity, triangulation-based range finders are limited to a small beam diameter, which limits coverage. On the other hand, the returned-signal-intensity sensors can offer a wide beamwidth, and also a cross section of various shapes. This can help achieve the coverage needed without scanning.

Therefore, sensors that operate similarly as the Banner proximity sensors should be pursued for space qualification. These offer the advantage of simple electronics, small packaging, and the choice of beam geometries.

Testing on MITy-2 will help to define the proper mounting configurations, as well as the effects of different reflectivities and dust.

8.2.4. Bumpers

The bumper design used on MITy-2 can essentially be carried over to the final design, with small adjustments to the threshold force. Microswitch makes a hermetically-sealed,

space-qualified switch that is slightly larger, but still miniature. There may be some questions to the use of a mechanical switch, opposed to a more 'solid-state' bumper made from a piezoelectric material. However, a good design can greatly reduce the failure modes, mainly by keeping it joint-free, and with a small amount of travel. This confines the only mechanism to inside the switch, which is hermetically sealed.

The mechanism requires small motion as it is, which can probably be reduced further. The edge of the bumper pad needs to be sealed to keep foreign matter from interfering with the small motion. In addition, the shape of the bumper pad should protrude in such a way to keep the moment about its 'hinge' sufficiently large (Figure 7.18).

Since contact sensors are without a large penalty in size, weight, or power, additional locations should be considered. This especially applies to the leading areas of any protruding component, such as cameras, solar panel edges, etc. The reason for this is because certain environments may present situations that can cause collisions at these points.

8.2.5. Inclinometers

Test results will likely confirm the need for pitch and roll sensors on each platform, using accelerometers as the sensor. MITy-2 uses a small signal-conditioned accelerometer, which draws only 60 mW. Micromachining has opened the door for rugged miniaturized inertial sensors in general, so that space-qualified accelerometers of similar size should be available, especially since accelerometers are widely used in aerospace applications. Power consumption may be further reduced by operating them intermittently.

8.2.6. Drive and Drag Wheels

The use of the drive and drag wheels for odometry purposes is very important for both hazard avoidance and navigation, although the accuracy of the sensors are not as critical for hazard avoidance. As mentioned in 8.1, future work should definitely include testing of the odometry sensors over irregular terrain, since little work has been done on this.

In addition, the drive wheel torque (current) data should be available on future designs, since this can provide very important information on not only hazard avoidance, but also on soil characterization (for scientific purposes). Sensor fusion of the drag wheel and drive wheel data will be a key issue, since these sensors are typically the only portions of

the rover that contact the surface. In short, there is a large amount of data that is available from the drive and drag wheels, and the difficult part will be the interpretation of this data, especially when processing is limited.

8.2.7. Environmental Considerations

The MITy-1 and MITy-2 prototypes have not directly addressed some of the main issues of a real mission, such as shock isolation, dust issues, and thermal control. The next prototypes will deal with these issues, as they can be a large influence on the design process. Dust is especially of concern on optical components, most notably on any detector optics. Provisions need to be made to minimize the collection of dust, and to possibly use active methods to remove it.

Also, one design philosophy that has been used through the entire rover design process is the minimization of moving parts wherever possible. Tradeoffs have been presented for scanning the range finder, which has been one of the few parts possibly requiring large motion. The drag wheel also needs a stiction-free design. Shock isolation can be performed in large parts by the rover's passive suspension system (wheel compliance and spring steel wires), so that further isolation may not be necessary. Large vibration loading from launching, as well as the shock loading from landing, will have to be analyzed in detail to ensure that the sensors are properly protected. Launching and landing will be the harshest structural loading that the rover will be exposed to. However, vibrations during operation can affect the sensor performance, so that further isolation may be needed to improve the detection process.

Thermal control will also be very important, and the rover will likely need an active control of temperature to maintain the electronics within an acceptable range. This may require all of the electronics to be confined to certain prescribed areas, which may eliminate the possibility of placing sensors along the perimeter of the rover. It is likely that the temperature will have to be monitored not only for thermal control, but also to make adjustments to gains and biases that occur in most sensors. These adjustments can be made in software, assuming the sensors are well-characterized.

CHAPTER NINE
CONCLUSION

The choice of the navigation and hardware sensors was strongly governed by the small platform size of the microrover and the limited power budget. Initially, research was expected to consist of a survey of existing hardware, followed by the selection of the off-the-shelf hardware. However, it became clear early on that many of the required sensor specifications were not met by existing hardware, which usually exceeds the performance requirements at the cost of size and power. Hence, a considerable amount of effort was spent in developing some of the prototype hardware, mainly the laser range finder for hazard avoidance, and the sun sensor for navigation. Nevertheless, initial simplified testing has demonstrated successful use of the various sensors, including both off-the-shelf and prototype hardware.

Two important hurdles were overcome in the demonstration of the sun sensor and laser range finder. Although both devices became operational only recently, they show promise in their future performance. In fact, the sun sensor was made with very low-cost components, yet still allowed the rover to navigate through 3 successive shuttle run maneuvers on a local football field, 60 yards total, returning to the starting point to within one foot. The range finder used more expensive components, but was still at a cost of only $500, half of which is the high power laser diode. The range finder shows promise for future improvements for two reasons. First, better detection will be possible, allowing the use of standard low-power lasers. Secondly, the prototype range finder was intentionally made large for mounting and machining purposes, and it is expected that a transformation to a much smaller system will be possible with no loss of accuracy.

Some of the important issues that need to be addressed quickly are the drag wheel design and sensor fusion algorithms. The longitudinal translation that the drag wheel measures is expected to be the largest contributor to navigation error, especially when the terrain is rough. In fact, the rough terrain will likely introduce many difficulties in all of the sensor readings. Field tests and computer simulations should focus on algorithm development for sensor data processing, especially the fusion of different sensors.

To summarize, the suggested sensors for autonomous navigation and hazard avoidance are:

Sensor	Purpose
accelerometers	tilt sensing
angular rate sensor	integrate for heading angle
sun sensor	calibrate heading
drag wheel	longitudinal translation
drive wheels	longitudinal translation, motor torque
laser range finder	obstacle detection
contact switches	collision sensing
proximity switches	obstacle detection, crater detection

Table 9.1: Summary of microrover sensors for autonomous navigation and hazard avoidance.

It should be noted that some of these sensors have multiple use. For instance, inclinometers are required for sun sensor measurements, and also for hazard detection, correction of translational motion, and for finding the orientation of the range finder at the time of sampling. Although each sensor has an individual purpose, a better understanding of the environment can be achieved with sensor fusion techniques.

REFERENCES

1. Malafeew, Eric. An Autonomous Control System for a Planetary Micro-Rover. MS Thesis, MIT May '93, CSDL-T-1173.

2. Ma, Calvin. Dynamics, Control and System Simulation of a Planetary Micro-Rover. MS Thesis, MIT May '93, CSDL-T-1174.

3. Schondorf, Steven. Systems Engineering for a Mars Micro-Rover. MS Thesis, MIT May '92.

4. Gilbert, John. Design of a Micro-Rover for Moon/Mars Mission. MS Thesis, MIT Dec '92, CSDL-T-1163.

5. I.J. Cox and G.T. Wilfong. Autonomous Robot Vehicles. Springer-Verlag. 1990

6. Leber, Douglas E. ...Impact of Navigation Instrumentation on-board a Mars rover..., MS Thesis, MIT May '92.

7. Humphrey Inc. Gyro Guide for Systems Engineers. 1988

8. J. Hansman and S. Weiss, "Sensors for Aircraft and Spacecraft", MIT course, Fall 1992.

9. Institution of Mechanical Engineers. Gyros. Proc. 1964-65, Vol 179, Feb 1965.

10. Merriam-Webster. Webster's Ninth New Collegiate Dictionary. Merriam-Webster, Springfield, Massachusetts, 1985.

11. Textron, phone conversation on May '93.

12. Cibulskis, Ray. Allied Signal Aerospace, Cheshire operations. Phone conversation on March '93.

13. Costa, Andy. North Precision Products Division. Phone conversation on March '93.

14. Laznicka, Oldrich. CSDL employee. Personal conversation on April '93.

15. Feldgoise, Stephan. MIT ME undergraduate student, MITy team member. Personal conversation on February '92.

16. Hecht, Eugene. Optics. Addison Wesley Publishing Co, 1987.

17. Halliday and Resnick. Fundamentals of Physics. John Wiley and Sons, 1986.

18. Sampson, Robert E. Overview of Active IR Imaging Systems. ERIM, Nov '87.

19. Wertz and Larson. Space Mission Analysis and Design. Kluwer Academic Publishers, 1991.

20. Stimpson, George W. Introduction to Airborne Radar. Hughes Aircraft Co.

21. Yariv. Optical Electronics.

22. Lynn, Steve. MIT EE graduate student, MITy team member, denture-wearer. Personal conversation on February '93.

23. Everett, Hobart. Rangefinder Survey (Draft). Defense Technical Information Center, 1993.

24. Hamamatsu. H2476-01 Technical Data, 1992.

25. Araki, Kazuo. High Speed and Continuous Rangefinding System. IEICE Transactions, V E74, October '91.