DOCTORAL THESIS

1993:117 D

DIVISION OF INDUSTRIAL ELECTRONICS

ISSN 0348 – 8373

On a laser anglemeter for mobile robot navigation

Kalevi Hyyppä

TEKNISKA HÖGSKOLAN I LULEÅ
LULEÅ UNIVERSITY OF TECHNOLOGY

DOCTORAL THESIS
DIVISION OF INDUSTRIAL ELECTRONICS

1993:117 D
ISSN 0348-8373

On a laser anglemeter for mobile robot navigation

Kalevi Hyyppä

AKADEMISK AVHANDLING

som med vederbörligt tillstånd av Tekniska fakultetsnämnden vid Tekniska högskolan i Luleå för avläggande av teknisk doktorsexamen kommer att offentligen försvaras i sal A109, Alfahuset, Tekniska högskolan i Luleå, tisdag den 1 juni 1993 kl 11.15.

Fakultetsopponent är Professor Risto Myllylä, University of Oulu.

On a laser anglemeter for mobile robot navigation

Kalevi Hyyppä
Division of Industrial Electronics

April 1993

Abstract

A laser anglemeter for use in a navigation system for mobile robots has been developed. The anglemeter measures heading angles to beacons made of vertical stripes of retroreflective tape. The anglemeter uses an optical system with a rotating mirror. It scans a laser beam which illuminates the beacons and receives pulses of reflected light from them. The pulses are amplified by a low-noise amplifier and then fed to a comparator. When the comparator trips it is assumed that the laser beam has hit a beacon. Then the angle is measured with a counter which is fed with pulses from an incremental encoder fixed on the axis of the mirror.

A navigation system using the anglemeter has been developed. The system includes an algorithm that associates measured angles with beacon identities. The algorithm has been patented by the author. The navigation system has been implemented on a test vehicle.

The optical and electronic parts of the anglemeter have been thoroughly analysed and simulated. The performance limits of the anglemeter are stated and design criteria for the optical and electronic parts are proposed. Measurements support the simulations and the theoretical analysis.

Some results of the analysis of the optical part of the anglemeter are:

- A Gaussian expression describing the bidirectional reflection distribution function (BRDF) of the beacons is proposed.
- The shape and power of the received light pulses from the beacons are found to be determined by the (Gaussian) intensity distribution and the divergence of the laser beam, the scan velocity, the width of the beacons and their BRDFs.
- No signal is received from beacons at close range.
- At long range the laser beam diameter is much larger than the width of the beacons which causes the range-dependence of the width of the received pulses to disappear. The shape of the pulses is approximately Gaussian in this region due to the intensity distribution of the laser beam. The width of the pulses is determined by the laser beam divergence and the scan velocity.

Optical misalignments in the anglemeter cause the dominating systematic error in measured angles. Another important error source is the range dependence of the received signal. Furthermore a non-circular cross-section of the laser beam is a potential source of large systematic errors while random errors due to electronic noise are surprisingly small.

Some results of the analysis of the electronic part of the anglemeter are:

- A low-noise photodiode-amplifier circuit with the photodiode in the feedback path has been developed. No noise generating resistor is needed to provide a DC-path to ground for signal and bias currents at the amplifier input.
- It is possible to design the amplifier as a matched filter for square shaped pulses.
- An implementation of the amplifier was found to have a NEP $\approx$ 3 fW/$\sqrt{\text{Hz}}$ at frequencies below 10 kHz.

To Gunhild and Maria

Contents

Preface

My graduate studies begun in 1967 at the Royal Institute of Technology. The pace was adequate in the beginning. Eventually other topics became more interesting than the studies. There were too many inventions, not invented yet, in instrumentation for research on electron accelerators and for space physics research.

The status of my graduate studies did not change significantly when I moved to Luleå in 1975 and became a research engineer at Luleå University of Technology. I did engineering work in different industrial projects and lectured to undergraduate students.

The development work and research leading to the present thesis begun in 1982. In 1991 I received the "The Engineer of the Year" reward in Sweden as a recognition of my work on the navigation system described in this thesis. That reward gave me enough self-confidence to restart my graduate studies.

Many thanks to my supervisors Per-Ola Börjesson, Lennart Gustafsson and Klas Ericson who took confidence in my ability to finally bring the many years of graduate studies to a successful end. A special thanks to Klas who conducted the transition of an engineer to a scientist with an iron-hand. Without the countless hours he spent on me this thesis would not exist.

Many former and present colleagues at the Department of Systems Engineering have contributed to the thesis through valuable discussions and comments. I would like to specially mention Ulf Andersson, Lars Bergström, Mats Blomqvist, Anders Grennberg, Kent Mrozek, Åke Wernersson and Urban Wiklund.

My family Gunhild and Maria suffered a lot during this last year. Thanks for still standing by me.

Financial support for this research has been given by Carl Tryggers Foundation for Scientific Research and by The Swedish Board for Technical Development.

Abstracts and publication information of parts I–V

Part I
Method of navigating an automated guided vehicle

Kalevi Hyyppä

Method of navigation a vehicle (11) moving generally in one plane in which the actual position and direction (L) of the vehicle is determined by measuring the angles in said plane from a reference point of the vehicle (X,Y) to fixed points in relation to a reference direction on the vehicle. A plurality of points (X_1Y_1, X_2Y_2, X_3Y_3) anonymous but fixed as to the position thereof are arranged for the navigation, and the position of all the fixed points together with existing sight obstacles between the reference point of the vehicle and the fixed points are registered in an electronic memory. The movement of the vehicle in said plane is initiated in a known position and with a known direction for identifying the fixed points visible from the reference point in said position. In each position and direction of the vehicle during movement thereof the latest known position and direction are utilized in determining the identity of the fixed points visible at that moment.

Published as *United States Patent 4811228* in March 1989.

Part II
AGV navigation by angle measurements

Urban Wiklund, Ulf Anderson and Kalevi Hyyppä

We describe an optical navigation system for the navigation and control of an autonomous guided vehicle (AGV). The navigation system consists of a low-power laser, a rotating mirror and the necessary optics. It is used to measure the angles to several identical reflective beacons. The position and heading of the AGV is recursively updated each time a valid angle is measured. It is easy to define and change the drive path which is a list of coordinates. The AGV follows straight lines between these coordinates. The system has been tested on an AGV prototype.

Contribution by Kalevi Hyyppä: The navigation principle, all hardware and

the analysis of error sources in the anglemeter.

Presented at *The 6th International Conference on Automated Guided Vehicle Systems* in Brussels in October 1988.

Part III
A low-noise photodiode-amplifier circuit

Kalevi Hyyppä and Klas Ericson

A photodiode-amplifier circuit with the photodiode in the feedback path is presented. It is named the PIF-circuit. No resistor is needed at the amplifier input to provide a path to ground for the signal and leakage currents from the photodiode and the amplifier input bias current. Thereby one potentially dominating noise source is eliminated. Small-signal and noise properties of the circuit are analysed. A first order approximation of a matched filter for square pulses can be implemented directly with the PIF-circuit and a cascaded first order low-pass whitening filter. Simulations and measurements on an implementation of the PIF-circuit show good agreement with the analysis. At frequencies below 10 kHz the implemented PIF-circuit has a NEP ≈ 3 fW/$\sqrt{\text{Hz}}$. The noise model of a photodiode is discussed.

Contribution by Kalevi Hyyppä: All work but the theoretical analysis which is made in collaboration with Klas Ericson.

Research Report TULEA 1993:11 April 1993. A condensed version will be submitted to *IEEE Journal of Solid State Circuits* in May 1993.

Part IV
Range dependence of the shape and amplitude of the received pulse for a laser anglemeter used for mobile robot navigation

Kalevi Hyyppä

A laser scanner, used as an anglemeter in a navigation system for mobile robots, has been developed. It measures heading angles to beacons made of vertical stripes of retroreflective tape. Expressions giving the received optical power and energy from a retroreflecting beacon, illuminated by the laser beam in the anglemeter, are derived. A Gaussian expression describing the bidirectional reflection distribution function (BRDF) of the beacon is proposed. A design rule for the optimum field of view of the receiving optics is proposed.

The shape and width of the received pulse are functions of the range R to the beacon. The range dependence of the received optical power is divided

into five intervals with different characteristics. Measurements support the derived expressions.

At long range the dependence of the received power amplitude on range is R^{-3} and the pulse shape and width are independent of range. At long range the pulse shape is determined by the Gaussian intensity distribution in the laser beam. The pulse width is governed by the laser beam divergence and the scan velocity.

A design rule for a filter which maximizes the signal-to-noise ratio for the anglemeter is proposed. A function giving a conservative estimate of the signal-to-noise ratio is derived.

Keywords: BRDF, Gaussian beam, ladar, anglemeter, laser scanner, range, retroreflecting beacon, signal-to-noise ratio.

Research Report TULEA 1993:12 April 1993. A portion of this part will be presented at *Robotikdagar-93* in Linköping in June 1993. A condensed version will be submitted to *Optical Engineering* in May 1993.

Part V
Error sources in a laser anglemeter used for mobile robot navigation

Kalevi Hyyppä

A laser scanner, used as an anglemeter in a navigation system for mobile robots, has been developed. It measures heading angles to beacons made of vertical stripes of retroreflective tape. A number of systematic and random error sources in the anglemeter are identified and analysed. The error from optical misalignments is shown to dominate the systematic error. Another important error source is the range dependence of the received signal from a beacon. Experiments verify the derived model of the misalignment induced error and they also show that the error can be efficiently compensated through calibration. A non-circular cross-section of the laser beam is shown to be a potential source of large systematic errors. Random angular errors due to electronic noise are surprisingly small.

Research Report TULEA 1993:13 in April 1993. A condensed version will be submitted to *The Third International Symposium on Measurement and Control in Robotics (ISMCR '93)* in Torino in May 1993.

Introduction, summary and epilogue

1 Introduction

1.1 Historical background

This thesis deals with an angle measurement system. The angles are used for navigational purposes in connection with autonomous vehicles and mobile robots. The origin of this research is the development of some navigation apparatus for sounding rockets performed by the author several years ago.

Sounding rockets are used for measurements of several physical quantities in the ionosphere. Many of the quantities are vectorial. The attitude of the rocket must therefore be accurately known to enable the transformation of the measured quantities from the coordinate system of the rocket to a more suitable coordinate system.

The development work started with a moon sensor. The first version of the moon sensor was successfully flown on a sounding rocket in 1974 [1], [2]. To stabilize the flight of sounding rockets, normal practice is to let them spin around their longitudinal symmetry axes with a few revolutions per second. The sensor was mounted with its optical axes pointing radially out from the rocket spin axes. Two pulses were produced by the moon sensor for each spin period of the rocket. The pulses were generated by moon-light striking a photodiode through two inclined slits. The elevation angle to the moon could be found from the time interval between the pulses and from the spin period. With some modifications the moon sensor was also flown as a sun sensor.

A star sensor for sounding rockets was then developed. A principal drawing of the star sensor optics is shown in Figure 1. Due to the rocket spin, images of stars cross two slit-shaped photodiodes which are mounted in the focal plane of a lens. When the image of a bright star passes a photodiode a detectable current pulse is generated. The pulse amplitude and the detection time are recorded by a microprocessor and subsequently transmitted to ground through the telemetry system of the rocket. Synchronizing signals from the telemetry system are used as time reference. Four units of the first version of the star sensor was successfully flown in 1976. The used attitude evaluation method is presented in [3]. The moon sensor, the sun sensor and the star sensor have been developed further into commercial products in a company founded by the author.

After some years it was realized that some of the methods and part of the hardware used in the attitude reconstruction of sounding rockets could be useful in navigation of Automatically Guided Vehicles (AGVs) and mobile robots. Angle measurements to vertically mounted stripes of retroreflective tape was chosen as the navigation method. The tape stripes act as "stars" when they are hit by a laser beam from a specially developed laser scanner — an anglemeter.

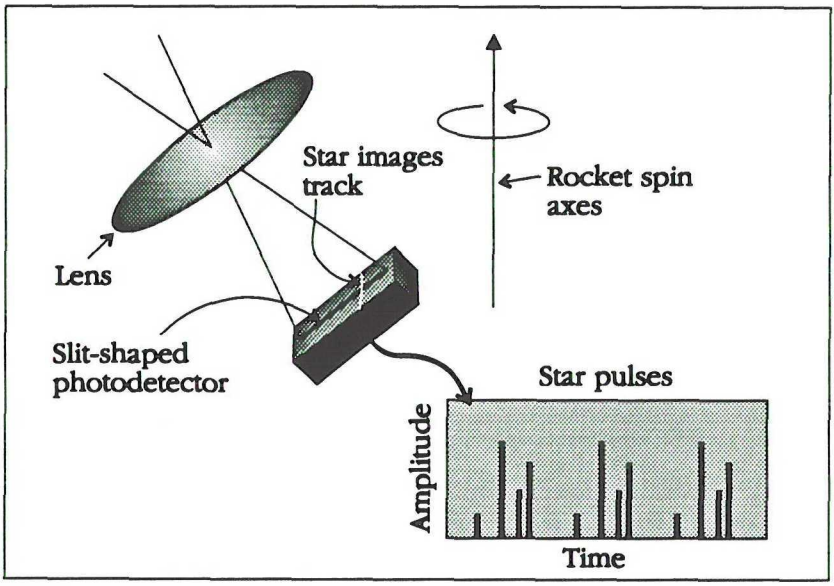


Figure 1: Principal drawing of star sensor optics. The shown pulse-pattern is generated by four stars during three spin-periods.

A test vehicle for navigational experiments was also developed. The vehicle was named Luleå Turbo Turtle (LTT). The LTT anglemeter is the main theme of this thesis.

1.2 Prior work on navigation apparatus for AGVs and mobile robots

AGVs have been in use for some time in warehouses, car plants etc. They are still too primitive to be called mobile robots. The prime ability for a mobile robot is to navigate. Most AGVs of today can't do that; they typically follow the magnetic field from a guide wire imbedded in the floor.

A number of significant papers in the field of autonomous robots have been collected in a recent book [4]. Overviews of different navigation techniques for AGVs are given in [5], [6], [7] and [8]. The last two papers are written by Toshihiro Tsumura. He has also suggested several navigation methods which uses one [9], [10] or three [11] lasers for angular measurements. A few other laser systems are described in [12]. A positioning system using two stationary scanning lasers is proposed in [13]. Two systems using a position sensitive optical detector — either a 4-quadrant photodiode or a lateral effect photodiode — are reported in [14] and [15].

The first commercially produced advanced AGV navigation system uses a scanning laser which measures directions to bar-coded reflectors [16]. The bar-codes are used for identification of the reflectors. Another navigation system using only one coded target is described in [17]. The code is not used for identification but to extract all necessary navigation information from the single target. The code pattern read by the navigation system is a function of distance to the target and orientation between the target and the navigation system.

The use of an optical range finder for navigation of an autonomous mobile robot is described in [18]. Another optical range finder using an array of LEDs and a lateral effect photodiode is described in [19]. An interesting range measurement method is presented in [20]. It is based on measurements of the mode hop time interval in a modulated laser diode.

The acoustic counterpart to the optical range finder is the sonar. Several groups are using an inexpensive sonic range finder originally designed for an automatic camera. The interpretation of sonar signals is fairly complex [21], [22]. Video cameras are also used for navigation. Powerful picture processing computers are needed to get reasonable interpretation times [23]. The computational burden can be reduced if the environment is partly structured and if special visual sensors are used. An example of such a system is described in [24]. There an omnidirectional image sensor (COPIS) is presented.

Almost all navigation systems include some kind of dead-reckoning which can give proper accuracy for short runs [25], [26]. A detailed analysis of inertial navigation, both for space and vehicle applications is given in [27].

Researchers tend to give names to their robot vehicles. As they build several generations many of the names often carry a generation tag. A few name examples are given in Table 1.

1.3 Overview

Section 2 starts with some geometrical considerations. It is followed by a discussion on measurement media. Then it is shown that motion restrictions of the robot can be used to advantage in the measurement task. Section 3 is the main body where technical solutions and performance limitations for a navigation system based on angle measurements are discussed. References to the individual parts constituting the rest of this thesis are given. In section 4 some results achieved with the LTT anglemeter are presented. Section 5 contains a summary of the contributions to the navigation technique of mobile robots by this thesis. Some concluding remarks and a few areas of future research are given in section 6.

Name	Interpretation
Blanche	Unknown [18]
CARMEL	Computer Aided Robotics for Maintenance, Emergency and Life support [22]
CYCLOPION	From Cyclops (a one-eyed giant in the Greek myth) [28]
Harunobu	A guide dog robot which can follow a person [29]
HelpMate	Commercial product name [30], [31]
HERMIES	Hostile Environment Robotic Machine Intelligence Experiment Series [32], [33], [34]
HILARE	Heuristiques Intéggrées au Logiciel et aux Automatismes dans un Robot Evolutif [35], [36]
ICAGV	Imperial College free ranging Automated Guided Vehicle [37]
IPAMAR	IPA Mobile Autonomous Robot [38]
LTT	Luleå Turbo Turtle [39]
ROBART	Unknown [40]
SIR	Sandia Interior Robot [41]

Table 1: A few examples of names on mobile robots.

2 Measurement variables and methods for use in navigation

2.1 Geometrical considerations

Distance and direction measurements to a set of reference points are two common navigation methods. A combination of distance and direction measurements is often the most appropriate method. Most distance measurement apparatus inherently give some information on direction. Direction measurement apparatus seldom give any distance information.

The pose of a body is defined to be the 6-dimensional vector describing the position and direction of the body. All six degrees of freedom are resolved if the coordinates (in R^3) of three points in the body are determined. Using distance measurements each points position can be determined by measurements of distances to three reference points with known positions. Neither the points in the body nor the reference points may *lie on a line.* The proposed method would give nine measurements to resolve six unknowns. By using the three additional constraints coming from the knowledge of the positions of the points in the rigid body, only six distance measurements need to be done. An example of this method is the Stewart platform [42] from the robotics field. The measured distances should be of the same order of magnitude or less than the separation of the points in the body to enable

good accuracy in the calculation of the body attitude.

When measuring directions only, three external reference points are needed to resolve the pose of the body. Two angles are measured to each reference point. The reference points may not *lie on a circle.* In this case the attitude accuracy is not dependant on the distances to the reference points. The error in the calculated position of the rigid body is on the other hand approximately proportional to the distances.

The highest accuracy is achieved in both cases when the reference points are symmetrically distributed around the rigid body. The accuracy improves if there are more reference points than the required three.

Two conditions have to be fulfilled to make the discussions above valid.

1. The association between the measurements and the reference objects is correct, i.e. the identities of the reference objects are known or can be determined.

2. All measurements for a position update must be taken with the body in the same pose.

The normal state of AGVs or mobile robots are of course that they are moving. The motion makes the second condition difficult to reach as most distance or direction measuring apparatus measures to one object at a time. The condition can be approximately fulfilled if the movement is small during the measurements. This was in fact the approach taken in the first runs with the LTT.

A more fruitful approach is to make use of the fact that all bodies have mass and therefore inertia. A state space model of the motion is made. The body pose is updated by integrating the appropriate variables in the model and when a measurement is available it is used to correct the pose errors due to unavoidable errors in the model. Pictorially this can be described as follows [43].

Assume that at time t_0 all the components of the 6-dimensional pose vector have the same uncertainty. Think of it as a point in R^6 surrounded by a 6-dimensional uncertainty sphere. (Some definition should of course be given to enable the comparison of errors in displacement and attitude. It is not important for the discussion at hand so it is omitted.) At time t_1 when the next pose calculation is done the errors have propagated with different gains so the uncertainty sphere has grown to a bigger and probably non-spheric volume. This process goes on until time t_k when a measurement to a reference point is available. The single measurement does not uniquely define the pose of the body but it restricts it in some sense. If for instance the distance from a point on the body to a reference point was measured, the body-point must lie on a sphere with the measured distance as the sphere radii and the

reference point as the sphere center. In the real world all measurements have errors so the body-point is only known to be in the neighborhood of the sphere surface.

Anyway, this knowledge can be used to shrink the uncertainty volume in one direction. As time elapses the uncertainty volume growths in all directions when only integration of the state space model is used for pose updates. The uncertainty volume is squeezed in one direction when a measurement is at hand. The errors in the positions of the reference points, the errors in the measurements, the time between measurements and the separation between the actual reference points governs the maximum size of the uncertainty volume i.e. the maximum pose error.

2.2 Measurement media for directional and distance measurements

Electromagnetic radiation such as IR and light is a suitable media for directional measurements. Its short wavelength gives reasonably sized antennas with small directional uncertainties due to diffraction. The high propagation velocity of electromagnetic radiation makes it an expensive method for distance measurements in this context. Typical distances for an in-door navigation system are less than 10 m and desired errors are less than 1 cm. As an example, the roundtrip time for light to travel 1 cm is 67 ps.

A cheeper choice could be to use ultrasound with its much lower propagation velocity. Its roundtrip time for 1 cm is 61 μs. A drawback with ultrasound is that the wavelength must be at least a few mm to give sufficiently low attenuation in air. The long wavelength gives poor directional resolution as a result of diffraction. The lobe-width of a common sonar system is 14 degree (half-angle).

Most ultrasonic systems use a simple envelope detector to detect the reflected ultrasound. The distance measurement resolution of ultrasonic systems can be considerably improved by synchronous demodulation [44].

From this point on, distance measurements will not be discussed any more. The attention will be turned to the chosen measurement method; Angle measurements to reference points using electromagnetic radiation. The term beacon will often be used for a reference point.

2.3 Measurement benefits from motion restrictions

Often the motion of a robot can be approximated to be constrained to a plane. An example of this situation is a mobile robot with wheels driving on a floor. (The kinematics of the robot can give further restrictions.) A natural choice of reference frame is then a Cartesian frame with its z-axes normal to

the motion plane. The only unknown components of the robots pose vector are then the x- and y-displacements and the rotation angle around the z-axes.

The motion restrictions make the navigation problem a planar problem. Accordingly the beacons should be placed in a plane parallel to the motion plane. The measurement space is then essentially one-dimensional; the important information is contained in the measured angle to a beacon. More dimensions could be added to the measurement space by measuring e.g. the amplitude and width of a received pulse from a beacon. These additional dimensions could aid in finding the identity of the beacon. That information is not used in the navigation system described in this thesis.

For obvious reasons a robot is not spinning like a sounding rocket. Therefore there is a need for some kind of scanning mechanism to cover the omnidirectional field of view.

The advantage in making use of the planar motion and measuring only the longitude of a reference point and not both the longitude and the latitude can be illustrated as follows. Denote the allowed inaccuracy in measured angle by $\Delta\gamma$ for the planar case. The measurement space can be divided into n_1 pixels each with the width $2\,\Delta\,\gamma$. The task of measuring the longitude of one reference point is then the same as to determine which pixel it occupies. The number of pixels are

$$n_1 = \frac{2\pi}{2\,\Delta\,\gamma} \tag{1}$$

Let the allowed inaccuracy be equal to $\Delta\gamma$ both in longitude and latitude for the general case and let furthermore $\Delta\gamma \ll \pi$. The number of pixels in the general case then becomes

$$n_2 = \frac{4\pi}{(2\,\Delta\,\gamma)^2} \tag{2}$$

A reasonable assumption is that the time to determine the position of one reference point in angular space is proportional to the number of pixels in both cases. The speed advantage of only measuring the longitude then becomes

$$\frac{n_2}{n_1} = \frac{1}{\Delta\gamma} \tag{3}$$

A typical value of $\Delta\gamma$ could be 1 mrad which would give a measurement speed advantage of 1000 times, for the planar case.

3 Technical solutions and performance limitations for angle measurements

3.1 Navigation principle

Vertical reflector tape stripes are used as beacons in the proposed navigation system. Bearings to visible beacons are measured by an anglemeter using a horizontally scanning laser onboard the mobile robot.

One design objective is to have many beacons in the work-area of the mobile robot. This gives increased accuracy and tolerance against occasionally hidden beacons due to obstacles and people moving around. To make the navigation system commercially interesting the beacons should be inexpensive. To enable this the beacons should be passive. The price of an illuminating laser, onboard the mobile robot, is much lower than the cost of installing maybe hundreds of active beacons of some kind in the operational area of the mobile robot.

Each beacon could be made up of a number of tape stripes arranged as a barcode [16]. It would solve the association problem because the barcode could contain the identity of the beacon. The drawback would be decreased range for the system due to the more complex signal read from the coded beacon. The receiver of a navigation system using coded beacons must have a wider bandwidth — to be able to read the code — than the receiver in a system using non-coded beacons. The wider bandwidth gives more noise in the receiver and therefore shorter range than in the non-coded system.

The tape stripes in the proposed navigation system carry no code. The association problem is solved by the following algorithm: When the anglemeter measures a bearing to a beacon the navigation system on the robot has an estimate of the robot pose from earlier measurements and from integration of the state variables. The known positions of all beacons and the estimated pose is used to predict bearings to beacons visible from the actual position. If the measured bearing is close to one of the predicted bearings the navigation system makes the conclusion that the reflected signal came from a valid beacon and it also deduces the beacon identity [45].

This algorithm has been patented by the author in several countries [46], [47], [48], [49], [50]. The algorithm is described in more detail in **Part I.**

The navigation principle and the LTT anglemeter are presented in **Part II.** A discussion on some of the error sources in the anglemeter is given. The main topic of **Part II** is the estimation and control algorithms which uses angles as input data for navigation.

3.2 A low noise optoelectronic receiver

An important part of the navigation apparatus is the front end of the receiver electronics i.e. the photodetector-amplifier combination. Two alternatives for the photodetector are the semiconductor photodiode and the photomultiplier. The semiconductor photodiode was chosen due to its small size, its low cost, its ruggedness and the fact that it does not need a high voltage supply as is needed for the photomultiplier. One drawback with this choice is that a low-noise amplifier for the photodiode must be designed. The photomultiplier has it built-in.

The one important design factor for the front end is noise. Several sources contribute to the total noise. The most important are:

- background radiation
- the photodiode
- the first active element of the amplifier
- a resistor at the amplifier input providing a DC-path to ground for signal and bias currents.

Background radiation generates a current in the photodiode. If the background radiation is uniform the current is a DC-current. The current is added to the leakage current in the photodiode. The resulting DC-current generates shot-noise. Both the leakage current and the photodiode capacitance are dependent on the bias voltage of the photodiode. The leakage current increases and the capacitance decreases with increasing bias voltage. The photodiode capacitance couples the voltage noise of the amplifier to the circuit and the capacitance should therefore be small. The optimum bias voltage is dependent on actual component values but is typically less than 1 V. The first active element of the amplifier should have low values on both voltage noise and current noise to give low-noise properties to the circuit. A FET with high transconductance and low bias current is a good choice. The resistor at the amplifier input should have a high resistance to minimize the influence of the thermal noise in the resistor.

In **Part III** a circuit is presented and analysed which eliminates the resistor. By using the photodiode itself as the feedback element the resistor is not needed. The circuit is named the PIF-circuit from **P**hotodiode **I**n **F**eedback path. Comparisons are made with two commonly used photodiode-amplifier circuits — the voltage amplifier circuit and the transimpedance amplifier circuit. It is shown that above a certain frequency the noise is dominated by the amplifier voltage noise in all three circuits. Below this break-frequency the PIF-circuit has lower noise than the other two due to the absence of the mentioned resistor.

It is also shown that an approximately matched filter can be easily implemented with the PIF-circuit. Only a first order low-pass whitening filter has to be added.

3.3 Range performance

In the field of radar there is the radar equation which gives the obtainable range as a function of a number of system parameters. Something similar can be derived for optical wavelengths. A complicating fact in the optical case is that the transmitting and receiving apertures are not identical as for the radar. An investigation is presented in **Part IV**.

A short description of the anglemeter is given. Expressions for the received power and energy from a retroreflective beacon are derived. A Gaussian expression is proposed for the bidirectional reflection distribution function of the used beacons.

It is shown that radiation which is reflected from a point on the beacon and enters the receiver aperture must have a direction almost parallel with the receiver optical axis to reach the photodetector. An unexpected result of this condition is that the detected radiation vanishes at close range. At short range only part of the receiver aperture gives directions which makes the incoming radiation reach the photodetector. This part of the receiver aperture is named the effective aperture of the optical receiver.

The effective aperture is a function of several receiver parameters, of the range and of the distance between the reflecting point on the beacon and the receiver optical axis. The dependence on range and the mentioned distance changes character at certain values of range, distance and receiver parameters. Nine different cases are identified and corresponding expressions for the effective aperture are derived. It is shown that beyond a certain range the expressions for the received power and energy can be simplified to closed form expressions.

Some numerical results are given with parameters taken from the implemented anglemeter. The range can be divided into five regions with different dependencies of the received signal strength on range. The amplitude of the received signal becomes maximal at a range value of 3 m. Beyond that range the amplitude is proportional to R^{-2} until the long range region is reached where the amplitude becomes proportional to R^{-3}.

With a couple of insignificant exceptions experimental data confirm the derived expressions. The measured received signal have a larger amplitude than predicted and is amplitude modulated at short range. The main causes of the discrepancies are scattering in the photodetector plastic casing and the non-uniform reflection properties of the beacons. Neither are modeled in the derived expressions.

An expression describing the noise from ambient light is derived. That

expression combined with expressions giving the noise from internal sources give the total noise of the anglemeter. Finally using the closed form expression for received power an expression for the signal-to-noise ratio at long range is derived.

3.4 Errors in measured angle

A number of sources which could influence the accuracy of a measured angle in the LTT anglemeter have been identified. They are:

- the incremental encoder measuring the scan angle
- missalignments between the anglemeter optics and the scan axes
- range-dependent signal strength
- elliptic laser beam cross-section
- noise in the received signal
- partly blocked beacons.

A first analysis of errors in measured angle is given in **Part II**. A more detailed error-analysis including measurements of some of the errors is presented in **Part V**.

The results of the error-analysis indicate that the misalignment between the anglemeter optics and the scan axes is the dominating error source. The misalignment generates a systematic angle-error with a sinusoidal dependence of the scan angle. The error can be efficiently compensated through calibration.

If a semiconductor laser is used instead of the HeNe-laser in the LTT anglemeter the elliptic cross-section of the laser beam could be the dominating error source.

The error from the range-dependent signal strength is in principal systematic but as the range information is not available for the anglemeter when the measurement is done the error can not be compensated in the anglemeter. The error could however be compensated in the navigation algorithm where the range information is available. It is not done in the present implementation of the navigation system.

The errors due to partly blocked beacons have other statistical properties than the rest of the errors. For a special case, discussed in the paper, their frequency function is rectangular over an angle interval from zero to a positive angle with an additional impulse at zero angle. This means that the most probable error from partly blocked beacons is no error at all. When the errors are non-zero they can be large compared to the other errors. The

errors from partly blocked beacons can be characterized as outliers which have a negligible influence on the navigation system.

With proper choice of components the errors due the incremental encoder and noise are small compared to the discussed errors.

4 Navigational experiments using the anglemeter

4.1 The first successful run

The first successful run with the LTT was reported in 1986 [45]. The tests were made in a rectangular lab (6x12 m^2) with 9 beacons glued to the lab walls. A photograph of the LTT is shown in Figure 2 and a photograph of the anglemeter is shown in Figure 3.

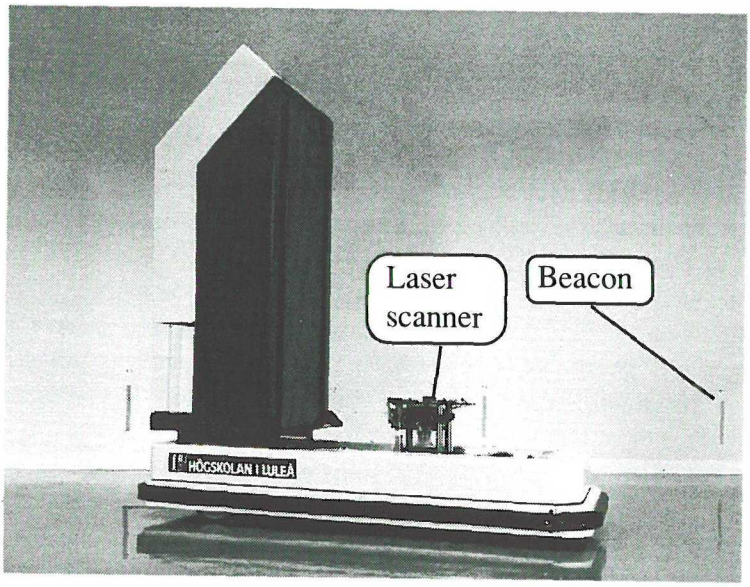


Figure 2: A photograph of the test vehicle Luleå Turbo Turtle (LTT). The length of the vehicle is about 1.2 m.

The used navigation method assumes that all angles are measured with the vehicle stationary. The vehicle is not stationary so a pose error is introduced which is dependent on the rotational speed of the anglemeter mirror and the speed of the vehicle. The maximum rotational speed of the anglemeter is dependent on laser power, maximum range and the anglemeter noise properties. The laser output power was 1 mW and the rotational speed was 1 rev/s. Only by letting the vehicle drive very slowly (10 mm/s) it could follow the set path. The goal was not to drive the LTT with high speed but

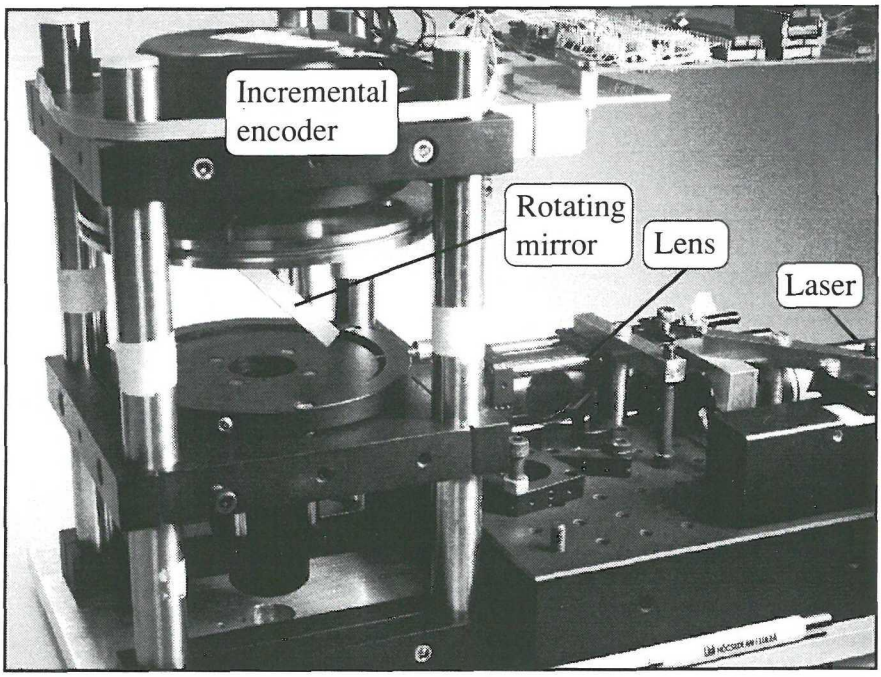


Figure 3: A photograph of the LTT anglemeter.

to show the possibility of using identical beacons in a navigation system for mobile robots.

The set path for the LTT was a "laying eight". The result of the run is reproduced in Figure 4 which is taken from [45]. It shows a copy of the picture drawn on the screen of the PC used as supervisory computer for the runs. The rectangle represents the walls of our lab. The small triangles indicate the beacons. The exact position of a beacon is marked by the triangle corner facing the interior of the lab. One of the beacons was attached to a pillar which caused that beacon to be slightly displaced from the nearby wall.

The inner path represents the pre-programmed set path and the outer zaggy curve is the vehicles recording of its track during the test run. The three short straight lines mark the pose of the LTT at the end of the run. The noise in the recorded track is mainly due to the simplifying assumption that the vehicle is stationary during angle measurements. The separation between the two tracks is due to the simple control algorithm used in the LTT. It was a pure proportional control algorithm which gives a remaining pose error.

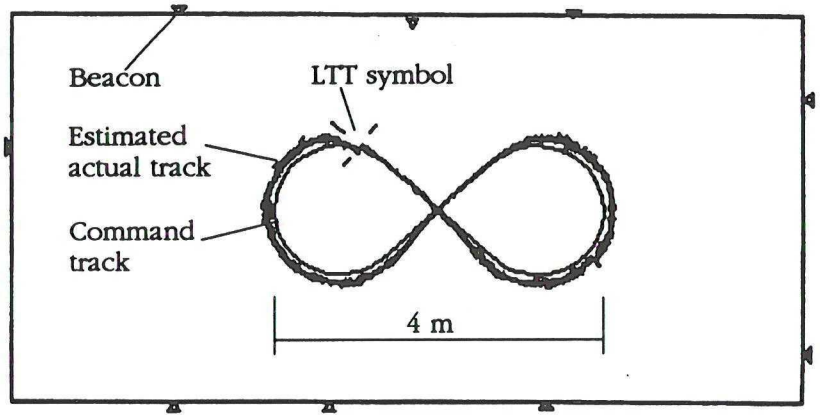


Figure 4: First succesful run with the LTT. The rectangle represent our lab (6x12 m^2) and the triangles indicate beacons. The inner path is the set-path and the outer zaggy curve is the vehicles recording of its track.

4.2 Further developments of the navigation algorithms and the anglemeter

The experience from the first run showed that a motion model was essential to increase the speed and drive accuracy of the LTT. A motion model and a new control algorithm were implemented in 1988 by co-workers to the author. The model and the control algorithm are reported in **Part II.** The design of an extended Kalman filter used in the navigation algorithm is also presented. More details of the navigation algorithm and the control algorithm can be found in [51] and [52] respectively.

The first runs with the new navigation and control algorithms were done with the algorithms implemented in the supervisory computer and not onboard the LTT. The serial channel connecting the LTT with the supervisory computer was a speed bottleneck. In spite of that the LTT could run safely with a speed of 0.3 m/s. Test runs was made in a lab (5.5x9 m^2) which had 11 beacons glued to its walls.

Later a new version of the anglemeter was developed with a semiconductor laser with a maximum output power of 12.5 mW and with a rotational speed of 6 rev/s. All navigation and control software were also rewritten and implemented onboard the LTT by co-workers to the author. Some more details on the new anglemeter and the software are given in [39]. The maximum speed of the LTT was now limited by its drive motor to about 1 m/s. The repeatability of the track was estimated to be better than 2 mm (2σ-value) when the LTT drove along a straight path at 0.3 m/s.

5 Contributions to the navigation technique of mobile robots

A navigation system for mobile robots which uses an anglemeter with a scanning laser beam on the robot to measure heading angles to beacons has been presented. The beacons consist of a stripe of reflective tape vertically attached to walls in the surroundings of the robot. The navigation system associates measured angles with beacon identities by an algorithm patented by the author in several countries.

A low-noise photodiode-amplifier circuit — the PIF-circuit — has been presented and analysed. It is a circuit which does not use a bias-resistor. The essential feature is that the photodiode itself constitute the feedback element in the circuit. It has been shown that the PIF-circuit has better noise performance than two commonly used photodiode-amplifier combinations. It was also shown that an approximation to a matched filter can easily be implemented with the PIF-circuit.

The range performance of the anglemeter has been investigated both theoretically and experimentally. It has been shown that the range can be divided into five regions with different dependencies of the received signal strength on range. An expression giving the anglemeter signal-to-noise ratio as a function of range and a number of system parameters has been derived. The results should be applicable on other similar laser scanners.

The accuracy of the navigation system is strongly dependent on the accuracy of the measured angles to the beacons. Five sources of angle errors in the anglemeter have been discussed. Measurements confirm that the dominating error source is optical misalignment in the anglemeter. It has been shown that the error from misalignment can be compensated through calibration.

6 Concluding remarks and future areas of investigation

In 1990 a company was founded by the group working with the LTT and a local investor. The company concept was to commercialize the results from the research done on the anglemeter into a navigation system for advanced AGVs. The company has been successful on the market so far. A picture of an example of the commercial version of the anglemeter at work in a factory is shown at the front cover of this thesis.

The LTT and the anglemeter have had a considerable influence on the research profile of the Department of Systems Technology at Luleå University of Technology. Apart from this thesis two students have written their licentiate theses in connection with the anglemeter and the LTT. Eight students

have fulfilled their diploma work in this field. More than ten papers have been presented at conferences and a number of project reports have been written by undergraduate students dealing with the LTT and the anglemeter.

This research has produced some new ideas and answers to questions in connection with laser scanners used for angle measurements. Several of the ideas have not been investigated and new questions have also emerged. A few examples of topics for further research are:

- Design principles for a step motor with high angular resolution and high power efficiency for the scan mirror in an anglemeter. Founding for this research has been granted from The Research Council of Norrbotten.
- A study of the idea of using time interpolation on the pulses from the incremental encoder in the anglemeter as a method of increasing the resolution and accuracy in angle measurements.
- Investigation of the feasibility of using common objects in a building as natural beacons and letting the navigation system find them by itself, i.e. map-building.

References

[1] K. Hyyppä, "En enkel månsensor för sondraketer", *Rakettinstrumenteringsmøte nr. 3*, Hurdalssjøen, Norway, May 1974.

[2] K. Hyyppä, "En enkel månsensor för sondraketer", *RVK 75*, Linköping, Sweden, April 1975.

[3] B. Schmidtbauer, "High-accuracy sounding rocket attitude estimation using star sensor data", *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES–14, no. 6, November 1978.

[4] I. J. Cox and G. T. Wilfong (editors), "Autonomous robot vehicles", New York, Springer-Verlag, 1990.

[5] S. K. Premi and C. B. Besant, "A review of various vehicle guidance techniques that can be used by mobile robots or AGVs", *Proc. AGVS-2*, pp. 195–209, Stuttgart, June 1983.

[6] P. Boegli, "A comparative evaluation of AGV navigation techniques", *Proc. AGVS-3*, pp. 169–198, Stockholm, October 1985.

[7] T. Tsumura, "Survey of automated guided vehicle in Japanese factory", *Proc. of IEEE International Conf. on Robotics and Automation*, pp. 1329–1334, April 1986.

[8] T. Tsumura, "Recent development of automated guided vehicles in Japan", *Robotersysteme*, vol. 2, pp. 91–97, 1986.

[9] T. Tsumura and M. Hashimoto, "Positioning and guidance of ground vehicle by use of laser and corner cube", *Proc. of IEEE International Conf. on Robotics and Automation*, pp. 1335–1342, April 1986.

[10] T. Tsumura, N. Fujiwara, M. Hashimoto and T. Tang, "A new method of vehicle position measurement by use of laser beam tracking", *Advanced Robotics*, vol. 2, no. 2, pp. 121–135, 1987.

[11] T. Tsumura and N. Komatsu, "Proposed position and heading measurement system using laser beams on the vehicle and corner cubes", *Proc. IEEE/RSJ International Workshop on Intelligent Robots and Systems (IROS '91)*, pp. 1436–1441, Osaka, November 1991.

[12] K. Nishide, M. Hanawa and T. Kondo, "Automatic position findings of vehicle", *Proc. of IEEE International Conf. on Robotics and Automation*, pp. 1343–1348, April 1986.

[13] T. Ochi, "A positioning system for mobile robots using symmetrical rotating laser beams", *Advanced Robotics*, vol. 4, no. 3, pp. 217–222, 1990.

[14] A. J. Mäkynen, J. T. Kostamovara and R. A. Myllylä, "Position-sensitive detector applications based on active illumination of a cooperative target", Engineering Systems with Intelligence, S. G. Tzafestas (ed.), pp. 265–274, Kluwer Academic Publishers, the Netherlands, 1991.

[15] J. P. Simon, M. Betemps and A. Jutard, "An optoelectronic sensor for an autonomous mobile robot using passive reflectors", *Sensors and Actuators A*, vol. 31, pp. 46–50, 1992.

[16] M. P. Robins, "Free-ranging automatic guided-vehicle system", *GEC Review*, vol. 2, no. 2, 1986.

[17] J. H. Kim and H. S. Cho, "Real-time determination of a mobile robot's position by linear scanning of a landmark", *Robotica*, vol. 10, pp. 309–319, 1992.

[18] I. J. Cox, "Blanche: Position estimation for an autonomous robot vehicle", *Proc. IEEE/RSJ International Workshop on Robots and Systems (IROS '89)*, pp. 432–439, Tsukuba, September 1989.

[19] J. A. Marszalec and H. M. Keränen, "A photoelectric range scanner using an array of LED chips", *Proc. IEEE International Conf. on Robotics and Automation*, pp. 593–598, Nice, May 1992.

[20] S. Shinohara et al., "Compact and high-precision range finder with wide dynamic range using one sensor head", *Proc. of IEEE International Conf. on Robotics and Automation*, pp. 126–130, 1991.

[21] A. Elfes, "Sonar-based real-world mapping and navigation", *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 3, pp. 249–265, 1987.

[22] J. Borenstein and Y. Koren, "Histogramic in-motion mapping for mobile robot obstacle avoidance", *IEEE Trans. on Robotics and Automation*, vol. 7, no. 4, pp. 535–539, 1991.

[23] H. P. Moravec, "The Stanford Cart and the CMU Rover", *Proc. IEEE*, vol. 71, no. 7, pp. 872–884, 1983.

[24] Y. Yagi, Y. Nishizawa and M. Yachida, "Map based navigation of the mobile robot using omnidirectional image sensor COPIS", *Proc. IEEE International Conf. on Robotics and Automation*, pp. 47–52, Nice, May 1992.

[25] T. Hongo et al., "An automatic guidance system of a self-controlled vehicle", *IEEE Trans. Industrial Electronics*, vol. IE-34, no. 1, pp. 5–10, 1987.

[26] W. L. Nelson and I. J. Cox, "Local path control for an autonomous vehicle", *Proc. IEEE International Conf. on Robotics and Automation*, pp. 1504–1510, Philadelphia, April 1988.

[27] M. M. Kuritsky and M. S. Goldstein, "Inertial navigation", *Proc. IEEE*, vol. 71, no. 10, pp. 1156–1176, 1983.

[28] N. Eberhardt and M. D. Wagh, "Cyclopion, an autonomous guided vehicle for factory use", *Proc. SPIE*, vol. 635, pp. 536–544, 1986.

[29] H. Mori and M. Sano, "A guide dog robot Harunobu-5 — Following a Person —", *Proc. IEEE/RSJ International Workshop on Intelligent Robots and Systems (IROS '91)*, pp. 397–402, Osaka, November 1991.

[30] J. Evans et al., "HelpMate™: A robotic Materials transport system", *Robotics and Autonomous Systems*, vol. 5, pp. 251–256, 1989.

[31] J. Evans et al., "Handling real-world motion planning: A hospital transport robot", *IEEE Control Systems*, vol. 12, no. 1, pp. 15–19, February 1992.

[32] G. de Saussure, C. R. Weisbin and P. F. Spelt, "Navigation and learning experiments by an autonomous robot", *Robotics & Computer-Integrated Manufacturing*, vol. 6, no. 4, pp. 295–301, 1989.

[33] C. R. Weisbin et al., "HERMIES-III: A step toward autonomous mobility, manipulation and perception", *Robotica*, vol. 8, pp. 7–12, 1990.

[34] D. B. Reister, "A new wheel control system for the omnidirectional HERMIES-III robot", *Robotica*, vol. 10, pp. 351–360, 1992.

[35] G. Bauzil, M. Briot and P. Ribes, "A navigation sub-system using ultrasonic sensors", *Proc. 1st Int. Conf. Robot Vision and Sensory Controls*, pp. 47–58, Stratford-upon-Avon, April 1981.

[36] M. Julliere, H. Place, E. Bazin and J. F. Radenac, "A relative localization system of a mobile robot", *Journal of Intelligent and Robotic Systems*, vol. 1, pp. 243–257, 1988.

[37] A. Sen, C. S. Wang, M. Ristic and C. B. Besant, "The Imperial College free ranging automated guidance vehicle system", Computer-Aided Production Engineering, V. C. Venkatesh and J. A. McGeough (ed.), Elsevier, 1991.

[38] G. Drunk, "A free ranging AGV", *Int. J. Mach. Tools Manufact.*, vol. 28, no. 3, pp. 263–272, 1988.

[39] K. Hyyppä, "Luleå Turbo Turtle (LTT)", *Proc. IEEE/RSJ International Workshop on Robots and Systems (IROS '89)*, pp. 620–623, Tsukuba, September 1989.

[40] H. R. Everett and G. A. Gilbreath, "A supervised autonomous security robot", *Robotics and Autonomous Systems*, vol. 4, no. 3, pp. 209–232, November 1988.

[41] P. R. Klarer and J. J. Harrington, "Development of a self-navigating mobile interior robot application", *Proc. 27th annual meeting INMM*, pp. 402–408, New Orleans, June 1986.

[42] D. Stewart, "A platform with six degrees of freedom", *Proc. Inst. Mech. Eng.*, vol. 180, pp. 371–386, 1965.

[43] Å. Wernersson, U. Wiklund, U. Andersson and K. Hyyppä, "Vehicle navigation using image information: on association errors", *Proc. Intelligent Autonomous Systems 2*, pp. 814–822, Amsterdam, December 1989.

[44] T. Jonsson, P. Klöör, B. Salomonsson, Å. Wernersson, "Coherent tracking using pulse doppler SODAR ('Ultrasonic Radar') - some robotics applications", *Proc. EUSIPCO-86*, pp. 1181–1184, The Hauge, August 1986.

[45] K. Hyyppä, "Optical navigation system using passive, identical beacons", *Proc. Intelligent Autonomous Systems*, pp. 737–741, Amsterdam, December 1986.

[46] K. Hyyppä, "Sätt för navigering av en i ett plan rörlig farkost, t ex en truck, samt truck för utövning av sättet", Swedish patent 451770, Stockholm, February 1988.

[47] K. Hyyppä, "METHOD OF NAVIGATING AN AUTOMATED GUIDED VEHICLE", US patent 4811228, March 1989.

[48] K. Hyyppä, "METHOD OF NAVIGATING AN AUTOMATED GUIDED VEHICLE", European patent 0238615, Munich, July 1991.

[49] K. Hyyppä, "Fremgangsmåte for å navigere en automatisert, ledet farkost", Norwegian patent 172012, Oslo, February 1993.

[50] K. Hyyppä, "Menetelmä ohjaajattoman kulkuneuvon ohjaamiseksi sekä siihen soveltuva ajoneuvo", Finnish patent 88655, Helsinki, February 1993.

[51] U. Wiklund, "Algorithms for navigation of autonomous guided vehicles based on measurements of directions to identical beacons", *Licentiate Thesis 1988:13L*, part 1, pp. 8-10, Luleå University of Technology, 1988.

[52] U. Andersson, "Trajectory estimation and control of autonomous guided vehicles", *Licentiate Thesis 1989:10L*, Luleå University of Technology, 1989.

Part I

Method of navigating an automated guided vehicle

The United States of America

The Commissioner of Patents and Trademarks

Has received an application for a patent for a new and useful invention. The title and description of the invention are enclosed. The requirements of law have been complied with, and it has been determined that a patent on the invention shall be granted under the law.

Therefore, this

United States Patent

Grants to the person or persons having title to this patent the right to exclude others from making, using or selling the invention throughout the United States of America for the term of seventeen years from the date of this patent, subject to the payment of maintenance fees as provided by law.

Commissioner of Patents and Trademarks

Attest Melvinia Gary

NOTICE

If the application for this patent was filed on or after December 12, 1980, maintenance fees are due three years and six months, seven years and six months, and eleven years and six months after the date of this grant, or within a grace period of six months thereafter upon payment of a surcharge as provided by law. The amount, number, and timing of the maintenance fees required may be changed by law or regulation.

United States Patent [19]

Hyyppä

[11] Patent Number: 4,811,228

[45] Date of Patent: Mar. 7, 1989

[54] METHOD OF NAVIGATING AN AUTOMATED GUIDED VEHICLE

[75] Inventor: Kalevi Hyyppä, Luleå, Sweden

[73] Assignee: Inik Instrument Och Elektronik, Luleå, Sweden

[21] Appl. No.: 72,257

[22] PCT Filed: Sep. 16, 1986

[86] PCT No.: PCT/SE86/00413

§ 371 Date: Jun. 29, 1987

§ 102(e) Date: Jun. 29, 1987

[87] PCT Pub. No.: WO87/01814

PCT Pub. Date: Mar. 26, 1987

[30] Foreign Application Priority Data

Sep. 17, 1985 [SE] Sweden 8504299

[51] Int. Cl.[4] G06F 15/14

[52] U.S. Cl. 364/424.02; 180/169

[58] Field of Search 364/424, 436, 443; 180/167, 168, 169; 356/1, 152, 4, 375

[56] References Cited

U.S. PATENT DOCUMENTS

4,225,226	9/1980	Davidson et al.	356/1
4,533,918	8/1985	Virnot	343/451
4,729,660	3/1988	Tsumura et al.	364/424 X

FOREIGN PATENT DOCUMENTS

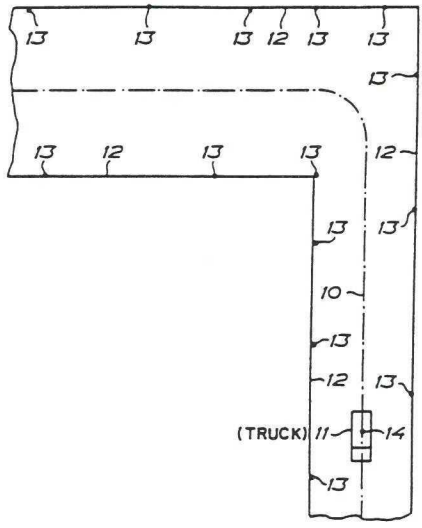
1146428	6/1985	European Pat. Off. .
3305119	8/1984	Fed. Rep. of Germany .
2526181	4/1983	France .
1181162	2/1970	United Kingdom .
2143395	2/1985	United Kingdom .

Primary Examiner—Eugene R. Laroche
Assistant Examiner—David Mis
Attorney, Agent, or Firm—Merchant, Gould, Smith, Edell, Welter & Schmidt

[57] ABSTRACT

Method of navigating a vehicle (11) moving generally in one plane in which the actual position and direction (L) of the vehicle is determined by measuring the angles in said plane from a reference point of the vehicle (X, Y) to fixed points in relation to a reference direction on the vehicle. A plurality of points (X_1Y_1, X_2Y_2, X_3Y_3) anonymous but fixed as to the position thereof are arranged for the navigation, and the position of all the fixed points together with existing sight obstacles between the reference point of the vehicle and the fixed points are registered in an electronic memory. The movement of the vehicle in said plane is initiated in a known position and with a known direction for identifying the fixed points visible from the reference point in said position. In each position and direction of the vehicle during movement thereof the latest known position and direction are utilized in determining the identify of the fixed points visible at that moment.

10 Claims, 7 Drawing Sheets



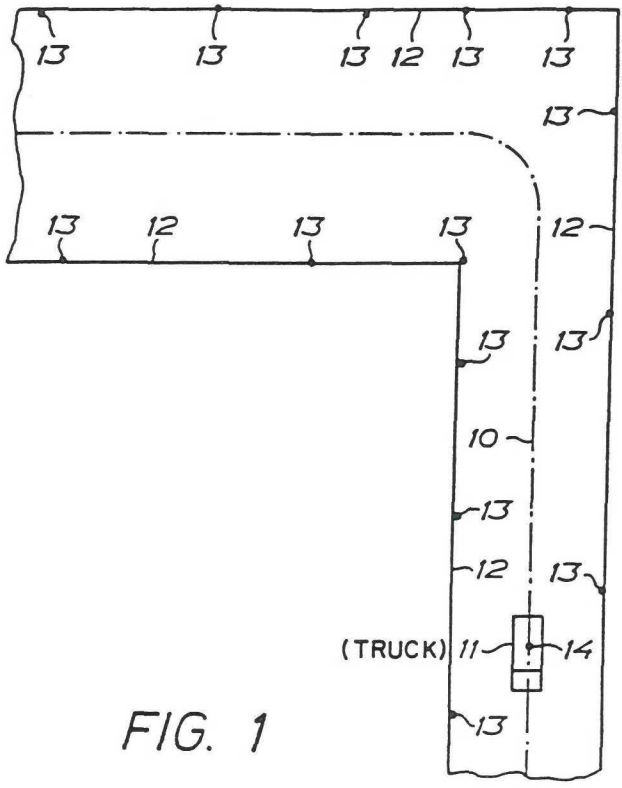


FIG. 1

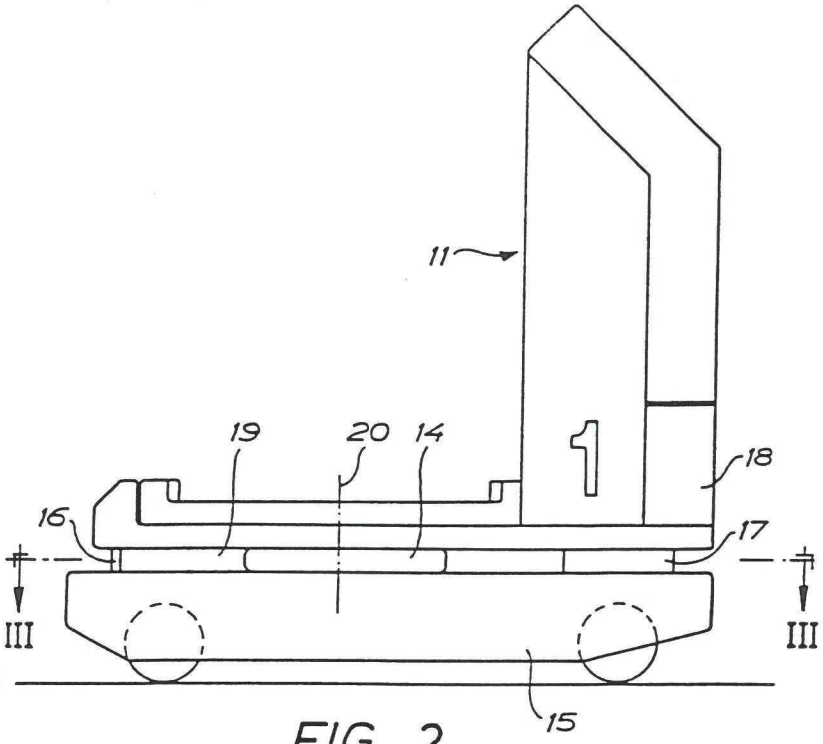


FIG. 2

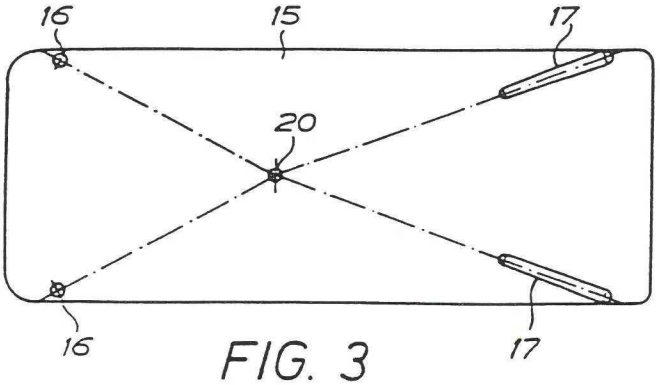


FIG. 3

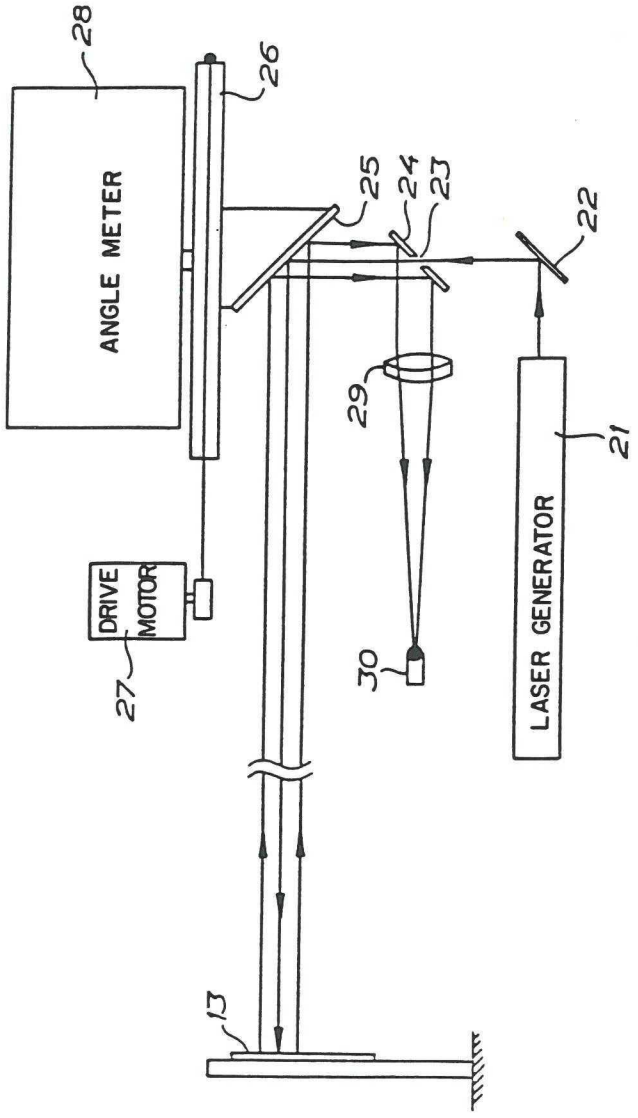

28
ANGLE METER
26
25
24
23
22
29
21
LASER GENERATOR
DRIVE MOTOR
27
30
13

FIG. 4

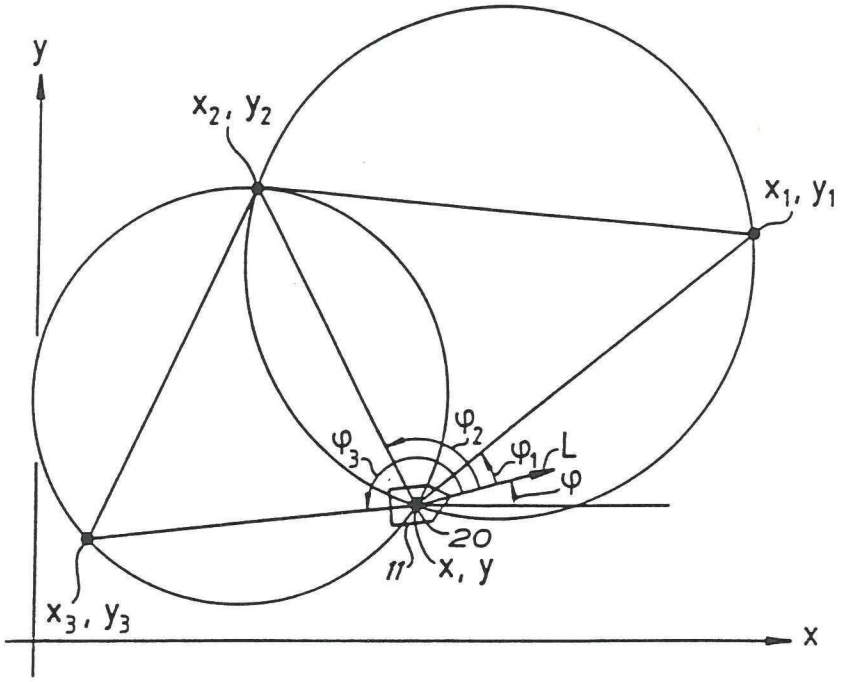


FIG. 5

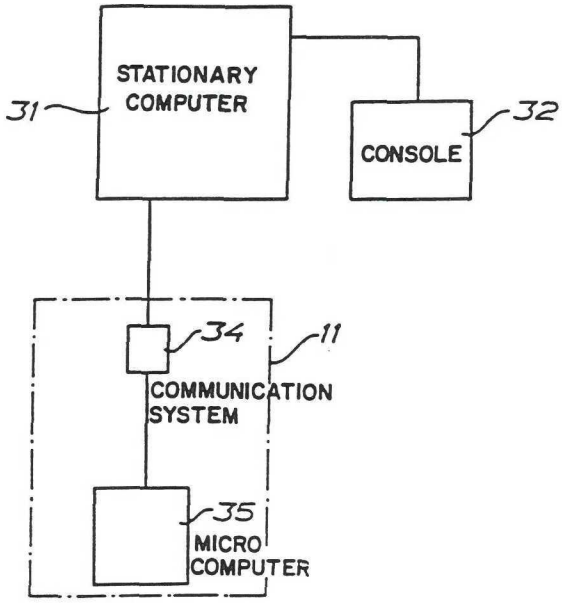

STATIONARY
COMPUTER
31
CONSOLE
32
34
11
COMMUNICATION
SYSTEM
35
MICRO
COMPUTER

FIG. 6

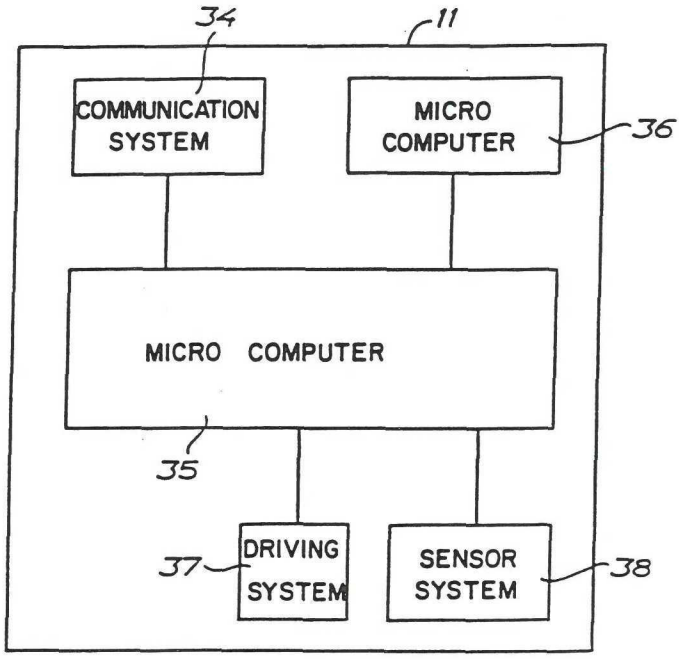

34
11
COMMUNICATION SYSTEM
MICRO COMPUTER
36
MICRO COMPUTER
35
37
DRIVING SYSTEM
SENSOR SYSTEM
38

FIG. 7

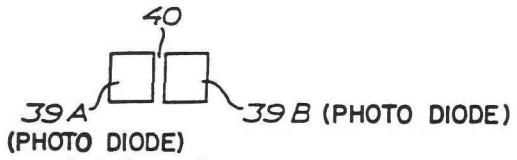


FIG. 8

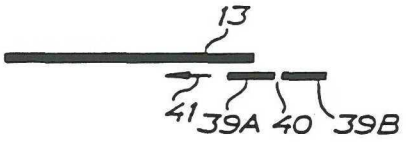


FIG. 9

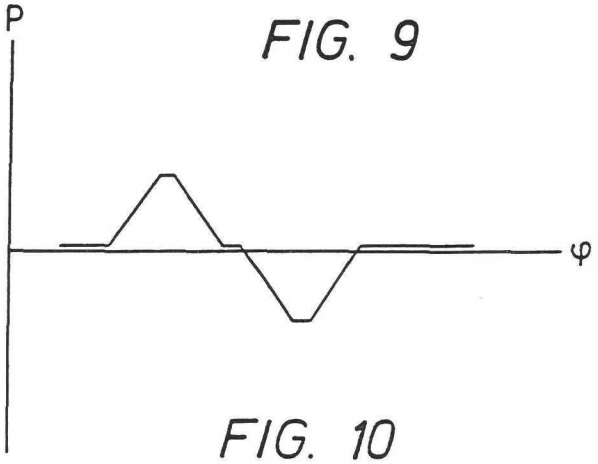


FIG. 10

METHOD OF NAVIGATING AN AUTOMATED GUIDED VEHICLE

The invention is related to a method of navigating an automated guided vehicle.

In industry there is an increased use of automated guided vehicles in transporting goods and material between different stations in storing premises and workshops. Then, the trucks are controlled automatically along predetermined paths between the various stations. In a reliable method of the truck steering the truck control system detects a cable loop arranged within or on the floor, or a painted loop on the floor. Such a control system is very reliable and simple to use but the conditions in the premises where the trucks are being used, are not always such that it is suitable to use a control loop within or on the floor, and in addition to this the system suffers from the no little drawback that the truck movement is limited strictly to the paths defined by the control loop, and that these paths cannot easily be altered because they are permanent installations.

Accordingly, systems of navigating automated guided trucks have been developed wherein the truck movement is not restricted to a permanently arranged control loop, the truck being provided with apparatus continuously defining its current position in relation to fixed objects in the premises through which the truck is moving. The present invention is related to this type of navigation system and accordingly concerns a method of navigating a vehicle moving generally in one plane in which a calculation of the current position and direction of the vehicle is effected by measuring the angles in said plane from a reference point on the vehicle to fixed positions in relation to a reference direction on the vehicle.

In a prior art systems for free navigation of automated guided vehicles a laser beam is used detecting reflectors fixedly arranged in the premises where the truck is moving. This system is called TURTLE and is described in the FMS Magazine, July 1983, pp 232–236. It is based on the fact that the truck for each position to be determined "sees" a number of fixedly arranged reflectors in the premises and that these reflectors can be identified, which is effected by the optical signal emitted from each of the reflectors, being coded. This is done by the laser beam emitted from the truck scanning over 360° and, when passing a reflector, generating a bar coded reflected signal.

Another prior art system described in Journal of Dynamic Systems, Measurement, and Control Vol. 105, September 1983, pp 152–155 is based on a type of position calculation similar to the TURTLE-system, but using instead of laser light ultra sound for locating and identifying the fixed positions in the space. From each fixed position a sound signal is then emitted having a predetermined frequency unique to the actual position. The latter system is similar to the system described in Robotics Age, March/April 1983, pp 31–33 and which is also subject matter of U.S. Pat. No. 4,328,545. However, in that case beacons are provided in the fixed positions for emitting light individually coded such that a receiver on the truck can identify the beacon emitting the light signal.

As far as the inventor of the present invention knows, prior art systems for free navigation of automated guided trucks are based on the fact that the signals received by the truck to be used for calculation of the current position of the truck, are specific to the points from which they originate, and can be used to identify these points. This is a limitation of the prior art systems, and the object of the invention is to make possible the use of simple and inexpensive reflectors providing the signals from the fixed points in the space without the necessity of coding these signals for identifying purposes. At the same time it will be possible to determine the position even if some of the fixed points are temporarily hidden or there are reflectors not associated with the system.

For this object the method according to the invention has been given the characteristics appearing from claim 1.

In order to explain the invention in more detail reference is made to the accompanying drawings, in which:

FIG. 1 is a plan view of premises through which an automated guided truck is to navigate by applying the method according to the invention,

FIG. 2 is a side view of the truck,

FIG. 3is a horizontal sectional view of the truck in FIG. 2 taken along line III—III,

FIG. 4 is diagrammaticside view of the optical system for emitting a scanning laser beam,

FIG. 5 is a geometrical figure for calculating the position and direction of the vehicle from the angles to three fixed reference points,

FIG. 6 is a block diagram of an electronic system for working the method according to the invention,

FIG. 7 is a more detailed block diagram of the electronic system arranged on the truck,

FIG. 8 is a diagrammatic front view of a detector suitable for use in the system,

FIG. 9 is a diagrammatic plan view illustrating the scanning of a reflector by use of the detector in FIG. 8, and

FIG. 10 is a graph illustrating the signal obtained from the detector from the scanning according to FIG. 9.

With reference to FIG. 1 a possible path 10 is shown to be followed by an automated guided truck 11 by applying the method according to the invention. The path extends through a free space in the premises, limited by the walls of the premises and/or objects existing therein, such as machines, storage racks or the like, here generally indicated with 12. The free space through which the truck can move does not actually have to be limited by fixed objects or be enclosed by walls; the space for the truck movement may be limited to a certain area having a specific shape and extension in the horizontal plane for other reasons. When utilizing the method according to the invention a plurality of reflectors 13 shall be mounted in the motion space, the reflectors being of the most simple and inexpensive type and are suitably made of reflecting tape, for instance of the type marketed by 3M Svenska AB under trademark SCOTCH LITE. The tape may be attached directly to walls or objects or may be mounted on located vertically standing pins and having a diameter of about 20 mm. The tape should have some extension in the vertical direction. Behind the reflectors a dark background should be arranged to make the reflectors sharply outlined against the background.

On the truck an optical system 14 is arranged for emitting a laser beam scanning 360°. The truck is constructed in a special way for the mounting of the optical system, FIG. 2 and FIG. 3. Thus, it comprises a chassis

15 supported on two uprights or columns 16 and two flanges or webs 17 the body 18 of the truck, leaving a gap 19 between the chassis and the body. In this gap the optical system 14 is mounted in such a way that the flanges or webs 17 extend radially to the vertical scan movement axis 20. In that way there is an almost unbroken area for the scanning laser beam around the axis 20 at a level close to the floor or ground. Though the beam per se is harmless, a lot of people are of the opinion that every laser beam, though being weak, is dangerous to the eyes. In the described arrangement the beam is far below eye-level, this placing of the optical system providing also the advantage that the body and the load carried thereon at no occasion screens the laser beam during the scanning movement thereof, which is not the least important for a secure and reliable navigation.

The body 18 of the truck may comprise suitable lifting means or other load handling means and space for electronic equipment, batteries etc.

The optical system may in principle be arranged as is diagrammatically shown in FIG. 4. A laser generator 21 which may be a gas laser (He-Ne) for visible light or a semiconductor laser for IR-radiation, possibly pulsed, directs its beam towards a mirror 22 which in turn directs the beam through an opening 23 in a mirror 24 to a mirror 25. The latter mirror is supported on a rotatably mounted balance wheel 26 connected over a drive belt to a drive motor 27 to be rotated at a constant rotational speed of the order of 2 or 3 revolutions per sec. The balance wheel is connected to an angle meter 28 of very high accuracy, or to a device for measuring the period used by the balance wheel to rotate at a constant speed from a reference position to a given position, the angle position of which is to be indicated. Such time measurement can be performed at very high reliability and by utilizing a simpler and less expensive device than an angle sensor of high accuracy.

When the laser beam hits a reflector 13 it is reflected back to the mirror 25 transmitting the reflected beam onto the mirror 24 from where the reflected beam by suitable optics 29 is directed onto a detector 30 to be focused thereon.

In FIG. 5 the truck is shown in a co-ordinate system, and in order to identify the position of the truck in a given moment the co-ordinates x and y for the axis 20 and the angle ϕ have to be determined. The angle ϕ is the angle between the x-axis and a vehicle reference line L, usually the longitudinal axis of the vehicle. This determination can be done by utilizing three fixed points, i.e. by using three reflectors, which the truck "sees" from the actual position. On the basis of the angles ϕ_1, ϕ_2 and ϕ_3 between the sight lines from the reference point 20 of the truck to the three reflectors having the co-ordinates (x_1y_1), (x_2y_2) and (x_3y_3), and the reference line L, the position and direction of the truck can be determined by applying conventional trigonometry as previously known per se and as described in the above mentioned publication Journal of Dynamic Systems, Measurement, and Control, September 1983, Vol. 105/153. Thus, by utilization of the optical system on the truck the angles ϕ_1, ϕ_2, ϕ_3 are measured and on the basis of these angles the equation is determined for those two circles which in FIG. 5 have been drawn through the three points and the position of the truck for calculating the position and direction of the truck. Since this technique is previously known per se these trigonometrical calculations are not presented here. There is, however, a substantial difference between the system described here and those systems previously known and described in the references initially referred to: the three points needed for the calculations have not identified themselves by a coding. The truck can see more fixed points than three and, thus, it is necessary to identify these points in one way or the other.

The motion area with reflectors and path are stored in an electronic memory. The storing in the electronic memory can be done by using known technique. A method which can be applied is the utilization of a CAD system, for instance AutoCAD, which is a designing program by means of which any drawing can be reproduced in a microcomputer. By using program the plan of the truck movement area of FIG. 1 can be programmed into a micro computer. To store the decided truck path the truck as an alternative to storing via the AutoCAD program can be driven along said path and its position in relation to the reflectors be continuously stored in the program. The positions of the reflectors are determined by utilization of conventional land surveying technique, and the measured positions are stored in the program. As an alternative the reflector positions can be located by driving the truck along a known path and be determined by utilizing the ptical system and a program loaded in the computer system of the truck.

Then, when the truck shall automatically follow the defined path, a truck position and direction must first be defined so as to serve later as a base for identifying subsequent positions and directions. This can be done in different ways. According to one method the truck is placed in a predetermined position and in a predetermined direction initially defined. According to another method the position and direction are defined from three fixed points at the time, assuming that said points have certain identities, and this is repeated with reference to all fixed points visible from the truck in a given position. This results in a plurality of positions and directions for each three points, which form the basis for the calculation. The variance of the obtained result is calculated, and then the calculations are repeated with another assumption as to the identity of the points. When the assumption is found, which giving the least variance, the points are identified.

As the truckis moving, consecutive calculations of the truck position and direction are made in each individual position on the basis of the angle measurements made by the apparatus 14 to the reflectors visible from the truck in the actual position by guidance of the obstacles programmed. Thus every new positioning of the truck will be based on information previously obtained with regard to the truck position and moving direction.

It should be noted that in each truck position several determinations of the truck position and direction may be made depending on how many reflectors reaction may be made depending on how many reflectors that are seen by the truck from the actual position. since the reflectors are of a simple and inexpensive nature and do not have to transmit a coded reflected signal, the detectors can be set out generously whereby a high accuracy in the calculation of position and direction can be obtained.

To work the method according to the invention by using the above mentioned calculations a computer system constructed in principle according to FIG. 6 is used. A controlling stationary computer 31 which can be common to many trucks, is provided for programming the motion area and the desired paths in this area, for instance by using the above mentioned AutoCAD

program. The computer 31 is also connected to a operator's console 32 with a keyboard to allow temporary manual operations in the program loaded. The computer 31 is connected through a communication system 34, FIG. 6, for instance a wireless link (radio or IR) to a micro computer 35 on each truck.

The computer 31 operates as a coordinator of several trucks in a system and gives through the link the necessary information to the computer 35 on the truck. The computer 35 can be a micro computer of the type Motorola MC 68010, and it is connected to another micro computer 36 on the truck which also can be of the type Motorola MC 68010 receiving angle measurement signals from the optical system 14 and processing these signals so as to give to the computor 35 information on the momentary position and direction of the truck by applying the above mentioned method. Then in the computer 35 a comparison with the programmed data is made, and depending on the result of the comparison control signals are sent to the driving system 37 of the truck to provide the correction of the truck movement necessary to make the truck move along the predetermined path. Also connected to the computer 35 is an appropriate sensor system 38 to stop the truck movement when a risk of collision with persons or objects arises.

The accuracy in the control of the truck can be substantially improved by utilizing a detector of the type described in FIG. 8. This detector is of a known design and comprises two silicon photodiodes indicated at 39A and 39B. Each of the diodes has a width of the order of 100 μm, and the diodes are separated by a gap 40 of the order of 10 μm. The scanning of the reflectors 13 in principle can be regarded as a projection of the detector, that is the two silicon diodes, on the surface of the reflector. This is illustrated in FIG. 9. When the projected image of the detector is entering the reflector in the direction of the arrow 41, one silicon diode 39A will first be in the area of the reflector and then the other one. By adding the signals from the silicon diodes with reversed polarity the detector will provide a signal as illustrated in FIG. 10 where the horizontal axis defines the scanning angle and the vertical axis defines the signal from the detector. When the image of the diode 39A only is projected on the reflector, the signal from the detector will rise to a largest positive value so as to decrease again to zero later when the second diode starts to enter the reflector, as the signal provided by the second diode is subtracted from the signal from the first diode. after that no signal from the detector will be obtained until the projection of said first diode moves from the surface of the reflector, the signal from said second diode being emitted from the detector. In other words, by this arrangement there is obtained a distinct signal easy to define, which means that the determination of the angle can be done at high accuracy with a corresponding high accuracy in the determination of the position and direction of the truck.

The determination of the position and direction by applying the above described method can be completed with odometry, which means that sensors on the wheels and steering mechanism of the truck provide signals representing the distance run and the direction of the existing movement, respectively, and that these signals then are used for calculating the change of the position and direction of the truck, to that it is possible to obtain, on the basis thereof by calculations executed in the computer 36, information on the actual position and direction of the truck. The quantities hereby obtained then may be processed together with those obtained by applying the method according to the invention.

By using quantities representing the speed of the truck and easily calculated by known methods, it is also possible to compensate for the truck movement between the measurements of the angles to the fixed points. In other words, the truck is not immobilized during the measurements, and even if there is a slight movement it does effect the calculations of the position.

Means can be provided for manually controlling the truck, for instance a control panel connected to the truck through a cable.

Some minor lateral deviation between the path followed by the truck on one occasion and the path followed in the same part of the path on another occasion can be programmed providing an advantage as the truck in that case does not wear wheel tracks in the support (the floor) during movement along an unchanged path for a long time.

Essential of the invention is the method in which the position and direction of movement of the truck is determined by utilizing a number of reflectors providing anonymous reflection signals. The practical embodiment of the mechanical and electronical means through which this method is realized has been illustrated and described in the embodiment preferred at present but can be modified or replaced by other means than the one described here without departing from the scope of the appended claims.

I claim:

1. Method of navigating a vehicle for instance an industrial truck moving generally in one plane by which a determination of the actual position and direction of the vehicle is provided by measuring the angles in said plane from a reference point on the vehicle to fixed points in relation to a reference direction of the vehicle at least three angles being used for a trigonometric determination by calculation means of position and direction, characterized in that a plurality of fixed points anonymous but determined as to the position thereof are arranged for the navigation, that the position for each of the points as well as existing sight obstacles between the reference point of the vehicle and the fixed points is stored in an electronic memory, that the movement of the vehicle in said plane is initiated in a known position and with a known direction so as to identify the fixed points visible from the reference point in said position, and that in each position and direction of the vehicle during the movement thereof, the latest known position and direction are utilized to determine with the aid of the stored position of the sight obstacles the identity of the fixed points visible at the moment.

2. Method according to claim 1, characterized in that the bearing of the fixed points are taken by emitting a light beam from the vehicle, scanning horizontally, and by reflection of the light beam from reflectors in the fixed points and reception of the reflected light beam on the vehicle.

3. Method according to claim 2, characterized in that the storing of the fixed points is effected by reproducing said plane with sight obstacles and the fixed points in said electronic memory.

4. Method according to claim 3, characterized in that the reproducing is effected in a CAD system for visual representation of said plane with sight obstacles and the fixed points, the path to be followed by the vehicle being included in the visual representation.

5. Method according to claim 2, characterized in that the determination of the position of the fixed points is effected by driving the vehicle along a known path while measuring the angles to the fixed points by utilizing the optical system and the program controlled computing and storing of the positions of the fixed points in the electronic memory.

6. Method according to claim 2, characterized in that the angles are determined by measuring the period for the movement of the light beam scanning horizontally at constant speed, from a reference point to a position where it hits a reflector.

7. Method according to claim 2, characterized in that the reflected light beam is detected by utilizing two photo detectors arranged side by side, the outputs of which are added with reversed polarity and which are hit, one after the other, by the reflected light beam.

8. Method according to claim 2, characterized in that the light beam is emitted close to the plane of the floor or ground.

9. Method according to claim 1, characterized in that the vehicle during movement thereof is brought to a position in relation to a desired path, which is displaced more or less towards one side or the other of the nominal path on successive tours along said path.

10. Truck for carrying out the method according to claim 1 comprising a chassis (15) and a body (18), characterized in that the body (18) is supported on the chassis (15) at a limited number of positions leaving a gap (19) between the body and the chassis, and that means (14) for emitting a light beam scanning horizontally is placed in the gap, existing means (16, 17) for supporting the body on the chassis of said positions having a limited extension in the path of the scanning movement of the light beam.

* * * * *

Part II

AGV navigation by angle measurements

Proc. 6th Int. Conf. Automated Guided Vehicle Systems, 199-212
October 1988, © IFS Ltd and authors, ISBN 1-85423-024-7

AGV navigation by angle measurements

U Wiklund, U Andersson and K Hyyppä
Luleå University of Technology, Sweden

We describe an optical navigation system for the navigation and control of an autonomous guided vehicle (AGV). The navigation system consists of a low-power laser, a rotating mirror and the necessary optics. It is used to measure the angles to several identical reflective beacons. The position and heading of the AGV is recursively updated each time a valid angle is measured. It is easy to define and change the drive path which is a list of coordinates. The AGV follows straight lines between these coordinates. The system has been tested on an AGV prototype.

1 INTRODUCTION

In this paper we describe how an optical directional measuring system is used for the navigation of an autonomous guided vehicle (AGV). The directional measuring system consists of a low-power laser, a rotating mirror and the necessary optics. The measuring system is called anglemeter throughout this paper. The anglemeter is used to determine the direction to reference points, consisting of identical stripes of retro-reflective tape. The position and heading of the vehicle is determined recursively by using the angles measured to these reflectors. The vehicle used has three wheels where the single front wheel is used both for steering and traction.

For a non-moving vehicle, at least three reflectors have to be visible in order to determine the position. This requires that we know from which reflector each measured angle originates. Since the reflectors have no identity marks, we have to associate each angle with a reflector. There might also be false reflections from other objects. These false detections have to be detected and discarded.

If the initial position is unknown, there is no way that the association problem can be solved if only three reflectors are used. Therefore we have to use more reflectors than three to get redundancy in the pattern of measured angles. Our previous work concerned this topic - a method of how to find the initial position. We also made a prototype of the anglemeter and tested the system on an AGV prototype [1,2].

When the vehicle is operating in a production environment, only a fraction of the reflectors will be visible from its current position. Therefore many reflectors are needed for the determination of the position with high accuracy anywhere in the area of operation.

The main advantages of having passive reflectors of the type we use, are the flexibility and the economy. It is easy to change the transportation routes in a factory, compared with wires in the floor or painted lines. A change in the routes can be caused by a change in production or, simply, the planned route is obstructed. Since every reflector strip is inexpensive, the cost of the system is essentially proportional to the number of AGV's - not the present or future working area. The same navigation system can also be used to operate, say, a cleaning machine. The present system can be used outdoors if the surface it travels on is reasonably flat, or if the beacons are not too far away. Applications might be in mines, in agriculture, when cleaning airfields, for the positioning of a dredging boat etc.

In our current work we have taken the dynamics of the system into consideration. We have restricted our study to a single vehicle and describe the navigation system and guidance law that we use. A complete system will consist of several vehicles coordinated by a supervisory computer via radio communication.

1.1 Outline of the paper

The vehicle and the anglemeter are briefly described and analyzed in section two. We also give some suggestions on how to locate the reflectors.

Section three deals mainly with estimating the position when the vehicle is moving while angle measurements are made. In [1,2] we assumed that the speed of the vehicle was very low so three subsequent angles could be considered to be measured from the same point. The rotational speed of the anglemeter is 1 Hz. We now present a method (based on Kalman filtering) for updating the position each time an angle is measured. It is possible to use a high speed vehicle, typically 0.5-2 m/s. The association problem is solved preliminary using a windowing technique. To determine the initial position we use the method from [2]. We then give an overview of the different components in the navigation and guidance system.

In section four we turn to the control problem. The drive path of the vehicle consists of straight line segments between given points. The list of points can for instance be generated with a CAD-program on a personal computer. It is then simple to change the drive path that the vehicle is to follow.

Simulation results are presented in section 5.

2 SYSTEM DESCRIPTION

2.1 The vehicle

A picture of our vehicle is shown in fig 1. The upper part, which we call the "tower" is connected to the "frame" through thin legs leaving a 40 mm wide slit between them. This allows the anglemeter which is mounted in the center of the frame to have almost 360° of horizontal field of view. The dimensions of the vehicle are (w,l,h) 0.5m, 1.3m, 1.4m. The combined traction and steer unit is placed in front of the anglemeter in the frame. Both motors in it are of the permanent magnet 24 V DC type.

The tower contains, from the top: some push-buttons and control lights, a frame with a VME-back-plane containing all the electronics and at the bottom a battery box with two sets of lead-acid batteries. One set is used for the motors and the other for the electronics. The dangling cables seen in the figure goes to a personal computer which is used to test and simulate the system. More details of the electronics can be found in [1,2].

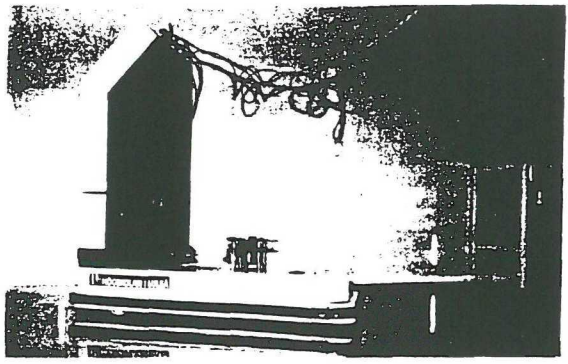


Fig 1. The test vehicle in our lab photographed with a flashlight to "light up" some of the beacons.

2.2 The anglemeter

There are many methods to measure angles to reflective beacons, our method take advantage of the known constraints in the measuring situation. It will gain orders of magnitude in receiver signal to noise ratio compared to other systems which use a vertical fan-shaped illuminating beam and a corresponding receiver field of view.

A picture of our prototype is shown in fig 2. A rotating mirror deflects a 1 mW laser beam to sweep in a horizontal plane. Before the laser beam hits the rotating mirror it travels through a hole in another deflecting mirror which is part of the optical receiver. The receiver is optically coaxial with the outgoing laser beam. The lens in the receiver has a focal length of 100 mm and a maximum aperture of 18 mm. The detector, which is placed on the optical axis in the focal plane of the lens, is a silicon photodiode with a diameter of 0.1 mm.

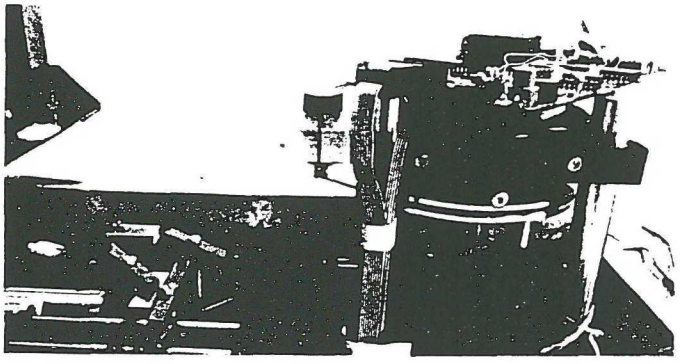


Fig 2 The prototype anglemeter. The rotating mirror is mounted under the flywheel in the right part of the photograph. The laser and part of the optical receiver can be seen in the lower left corner.

Our vehicle is moving on a reasonably flat floor, which means that we do not need to waste our illuminating laser power by spreading it out in a fan. We know at what height the laser beam will hit the wall and put our beacons at that height. This means that we can restrict the vertical field of view of the optical receiver in the anglemeter to the same order as the divergence of the illuminating laser beam. In this way we minimize the background noise and the probability of catching false beacons from strong external light sources. The divergence of the laser is 1 mrad. The discussion above is of course valid only if the scanning beam plane is parallel with the floor. The vertical length of the reflective stripes makes the anglemeter tolerate non-ideal i.e. real floors.

The improved signal to noise ratio will influence four parameters positively:

Longer range
Higher rotational speed of the mirror
Narrower reflective strips
Smaller error in measured angle

The reason to have mechanically moving parts in the anglemeter, which naturally makes it more fragile than e.g. a system built around CCD-cameras, is the precision level we want to reach. Our demand on almost 360° of horizontal field of view takes several CCD-cameras. The resulting adjustments and calibration of them and their non-ideal imaging optics will be very difficult.

The most important part of the anglemeter is the precision incremental encoder. The rotating mirror is attached to its axis. The error in measured angle to a beacon originates from:

Errors in the incremental encoder
Alignment errors
Noise in the receiver
Partly blocked beacons

The incremental encoder outputs two 90° shifted pulsetrains and a zero pulse. The pulsetrains contain 9600 pulses/revolution which gives a resolution of better than 0.2 mrad after electronic processing. The manufacturer does not state any absolute non-accuracy but we believe it is of the same order as the resolution. Its contribution to the total error is therefore negligible.

There are several alignments to be made in the anglemeter which can cause errors in the measured angle. The misalignment between the mirror rotational axis and the laser beam is the dominant one and will be discussed in some detail.

Fig 3 shows the rotating mirror, the laser beam, an attached coordinate system and the symbols which will be used. The mirror rotates around the vertical Z-axis. Its rotational angle is γ, which also is the angle of the reflected beam with no alignment error. The incoming laser beam can be assumed to lie in the XZ-plane and to hit the mirror in the origin of the coordinate system without loss of generality. The alignment error is β and the error in γ is $\Delta\gamma$. Observe that the reflected laser beam does not lie in the XY-plane and that $\Delta\gamma$ is the angle between the nominally reflected beam and the actual beams projection in the XY-plane.

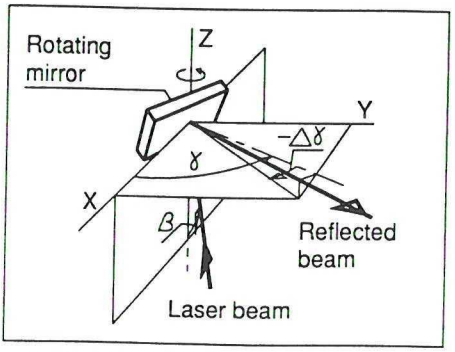


Fig 3 Coordinate system and symbols used to analyze the influence of the alignment error β on the measured angle error $\Delta\gamma$

Let n_0 be a unit vector normal to the mirror surface when $\gamma=0$.

$$n_0 = \frac{1}{\sqrt{2}} \cdot (1, \quad 0, \quad -1)^T \tag{2.1}$$

This can be transformed to the rotating vector $n(\gamma)$ by premultiplying with the rotational matrix $R(\gamma)$.

$$n(\gamma) = R(\gamma) \cdot n_0 \tag{2.2}$$

The rotational matrix is given by

$$R(\gamma) = \begin{pmatrix} c\gamma & -s\gamma & 0 \\ s\gamma & c\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.3}$$

where we have used $c\gamma$ and $s\gamma$ as shorts for $\cos(\gamma)$ and $\sin(\gamma)$.
This gives

$$n(\gamma) = \frac{1}{\sqrt{2}} \cdot (c\gamma, \quad s\gamma, \quad -1) \tag{2.4}$$

The reflection of the laser beam can be modeled with a rotational matrix A, which is defined by

$$A(\gamma) = I - 2n(\gamma) \cdot n^T(\gamma) \tag{2.5}$$

This gives

$$A(\gamma) = \begin{pmatrix} s^2\gamma & -c\gamma \cdot s\gamma & c\gamma \\ -s\gamma \cdot c\gamma & c^2\gamma & s\gamma \\ c\gamma & s\gamma & 0 \end{pmatrix} \tag{2.6}$$

A unit vector p_{in} parallel with the incoming beam can be found by inspection

$$p_{in} = (s\beta, \quad 0, \quad c\beta) \tag{2.7}$$

A unit vector p_{out} parallel with the reflected beam is formed by premultiplying p_{in} with A.

$$p_{out} = A \cdot p_{in} \tag{2.8}$$

Which becomes

$$p_{out} = (s\gamma \cdot s\gamma \cdot s\beta + c\gamma \cdot c\beta, \quad -s\gamma \cdot c\gamma \cdot s\beta + s\gamma \cdot c\beta, \quad c\gamma \cdot s\beta)^T \tag{2.9}$$

Finally the projection of p_{out} in the XY-plane is p_{outp}

$$p_{outp} = (s\gamma \cdot s\gamma \cdot s\beta + c\gamma \cdot c\beta, \quad -s\gamma \cdot c\gamma \cdot s\beta + s\gamma \cdot c\beta, \quad 0)^T \tag{2.10}$$

this can also be written as

$$p_{outp} = k \cdot (\cos(\gamma + \Delta\gamma), \quad \sin(\gamma + \Delta\gamma), \quad 0)^T \tag{2.11}$$

After some manipulations where we have assumed that β and $\Delta\gamma$ are small we arrive at the result

$$\Delta\gamma = -\sin(\gamma) \cdot \beta \tag{2.12}$$

This is a systematic contribution to the total error in the measured angle and might also be the dominating one if the alignment is not done with outmost care.

Our present beacons have a horizontal width of 20 mm and one can ask the following question. To what part of the beacons do we measure our angles? The mirror rotates in the positive sense so that the laser beam crosses the beacon from right to left. Due to our high signal to noise ratio we can have a rather low threshold level in the receiver without introducing false detections as a result of electronic noise. Experiments have shown that the angle is measured to a point less than 1 mm from the right edge of the reflective tape. This is at a range of 10 m and at a rotational speed of 1 Hz. Thus the contribution to the error angle from the electronic noise is negligible.

From the discussion above it follows that if the right side of a beacon is blocked due to some obstacle we will get an error which could be significant. We would be better off if the beacon were completely blocked! The error is random with a positive mean.

2.3 Reflector maps

A large number of reflectors are needed in a production environment. It would be almost an impossible task to find the coordinates of all the reflectors with high accuracy in a global coordinate system. Therefore it is wiser to divide the area of operation into smaller areas. Different rooms in a factory is a simple example of this division. The area near a docking station is another. The approximate location of the local rooms in the global system can easily be found. The reflectors are then localized within these local coordinate systems with high accuracy. We use the anglemeter to do the localization of the reflectors.

If the reflectors are located in a local room with too much symmetry, there can be several solutions when the initial position is determined. Therefore the reflectors should be located randomly. We can however allow a symmetrical location of the reflectors in for instance a narrow corridor which the AGV only drives through, without docking at any specific points. Without knowing the actual position of the beacons, the symmetry in the measured angles could be used for navigation when the vehicle is passing the corridor.

3 ESTIMATING THE TRAJECTORY OF THE AGV

The position is updated each time a valid angle is measured. This means that all angles will be measured from different positions of the AGV. We therefore have to model the vehicle's motion between the measurements. In this section we start by deriving a motion model which will be used in the estimator design. We have chosen a first-order model instead of the more common second order models where acceleration is taken into account. To compensate for this model simplification, we assume that the values of the speed and steer angle are contaminated with additional noise. The control laws are also designed such that the changes in the set values to the servos between two sampling instants are small. This helps to justify our reduced order model.

The motion model can be used to update the position even if there are no angles available. This will however lead to large accumulated errors. We therefore need the angles to make corrections. But the measured angles have to be associated with reflectors or discarded as false. An association method is presented where we use the estimated position from the filter.

Before the filter can be used, the initial position has to be found. The method used at present will be described at the end of this section.

3.1 Motion model

We will study planar motion for a vehicle that moves with the translational velocity u_1 and the angular velocity u_2. The location of the AGV is described by the coordinates x,y and its orientation by the angle θ to the x-axis. In the description we have chosen the midpoint of the rear axis as reference point on the vehicle, cf. fig 4. The angle θ is also the heading of the AGV.

The differential equations for the motion are

$$\dot{x} = u_1 \cos\theta \qquad \dot{y} = u_1 \sin\theta \qquad \dot{\theta} = u_2 \tag{3.1}$$

On the vehicle we can control the speed of the front wheel, v, and the steer angle, α. In fig 4 we can see that there will be an instantaneous center of rotation for the motion of the vehicle.

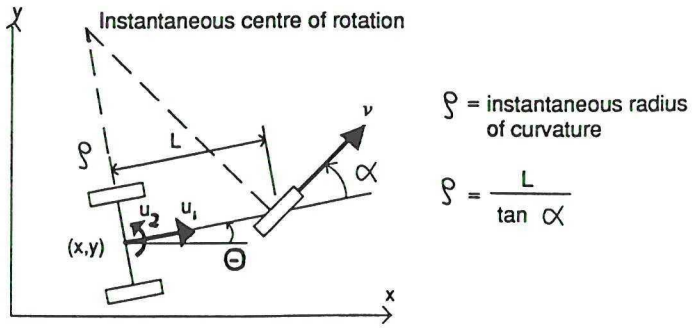


Fig 4. When the steer angle is constant, the motion of a three wheeled vehicle is along a segment of a circle.

Since the vehicle is a tricycle, we will get the following relations for our input signals u_1 and u_2

$$u_1 = v\cos\alpha \qquad\qquad u_2 = \frac{v}{L}\sin\alpha \tag{3.2}$$

where L is the distance between the steer axis and the rear axis. It is also possible to derive expressions for the input signals when the rear wheel movements are measured. In the future we will use incremental encoders to measure the wheel movements. This will increase the accuracy of the update of the motion model between the sampling instants. At present we do the assumptions that the input signals u_1 and u_2 are constant during the sampling interval. This means that we have assumed that the speed and steer angle only changes at the sampling instants and remain constant during the sampling interval, T. The equations of motion can be integrated between the sampling instants with constant input signals. This gives the non-linear discrete time model

$$x(k+1) = x(k) + \frac{u_1}{u_2}(\sin(\theta(k) + u_2 T) - \sin\theta(k)) \tag{3.3a}$$

$$y(k+1) = y(k) + \frac{u_1}{u_2}(-\cos(\theta(k) + u_2 T) + \cos\theta(k)) \tag{3.3b}$$

$$\theta(k+1) = \theta(k) + u_2 T \tag{3.3c}$$

Throughout this paper we will use k to denote the value of a variable at the discrete time instants t_k when sampling is made. For convenience, this will be omitted for the input signals.

The geometrical interpretation of (3.3), is that the motion is along a segment of a circle. For small changes $u_2 T$ in the heading, we can approximate the trajectory with a straight line.

$$x(k+1) = x(k) + u_1 T \cos\theta(k) \tag{3.4a}$$

$$y(k+1) = y(k) + u_1 T \sin\theta(k) \tag{3.4b}$$

$$\theta(k+1) = \theta(k) + u_2 T \tag{3.4c}$$

With a proper control law, the steer angle often will have a small variation around a nominal value needed to follow the current segment. The steer servo also has a short time constant. The neglected dynamics in the servos will introduce some uncertainties in the input signals. We will take these model errors into account by including noise signals in the discrete time model, i.e. we replace the idealization in (3.2) with the additive noise model

$$u_1 = (v + w_v)\cos(\alpha + w_\alpha) \qquad u_2 = \frac{(v + w_v)}{L}\sin(\alpha + w_\alpha) \tag{3.5}$$

This can now be inserted in (3.3) or (3.4) where we, after straightforward identification of terms, can find the coefficients for the noise terms. Only the first order terms are used. For sampling intervals of 0.2 s or shorter, the numerical values of the coefficients in the two cases are practically the same. We therefore use the coefficients from the second linearized case. We will now have a model of the type

$$x(k+1) = f_1(x(k), y(k), \theta(k), u_1) + g_{11} w_v + g_{12} w_\alpha \tag{3.6a}$$

$$y(k+1) = f_2(x(k), y(k), \theta(k), u_1) + g_{21} w_v + g_{22} w_\alpha \tag{3.6b}$$

$$\theta(k+1) = f_3(x(k), y(k), \theta(k), u_2) + g_{31} w_v + g_{32} w_\alpha \tag{3.6c}$$

where for instance

$$g_{11} = T\cos\alpha(k)\cos\theta(k) \tag{3.7}$$

The noise signals w_v and w_α are assumed to be independent zero mean white noise processes with covariance matrix Q.

We can write this motion model in a more compact way by introducing the state vector $X = (x, y, \theta)^T$ and noise vector $w = (w_v, w_\alpha)^T$.

$$X(k+1) = f(X(k), u_1, u_2) + G(k) w(k) \tag{3.8}$$

3.2 Measurement model

Reflector i is located at (x_i,y_i). The index i is used to denote the current reflector. The position of the midpoint of the rear axis is $(x(k),y(k),\theta(k))$. Using fig 5 we see that the angle to this reflector is

$$\gamma_i(k) = -\theta(k) + \arctan\left(\frac{y_i - y(k) - d\sin\theta(k)}{x_i - x(k) - d\cos\theta(k)}\right) + \upsilon(k) \tag{3.9}$$

where d is the distance between the anglemeter and the rear axis. We can write this as

$$\gamma_i(k) = h_i(X(k)) + \upsilon(k) \tag{3.10}$$

We assume that the sequence $\{\upsilon(k)\}$ is white noise with zero mean and covariance R and uncorrelated with the process noise. The real measurement noise was analyzed in section 2.2. There are also errors in the reflector coordinates. The white noise assumption on $\{\upsilon(k)\}$ has been chosen to simplify the estimator design.

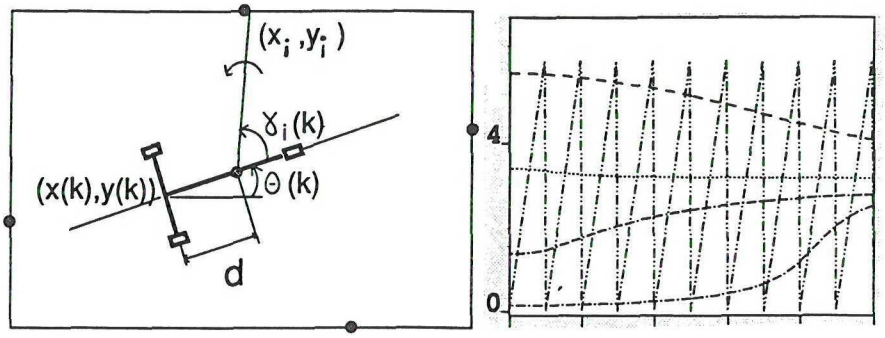


Fig 5. Principle of angular measurements and variation in the angles to four reflectors when the vehicle follows a straight line with the speed 0.5 m/s.

The mirror in the anglemeter rotates with one revolution per second. Suppose that the vehicle is moving straight forward in a room with four reflectors. In fig 5 we show the variation in the angles to the reflectors. The instantaneous angle $\gamma(t)$ of the anglemeter is also shown in the same figure. It is only possible to measure an angle to a reflector at those time instants when we have a crossing between the anglemeter curve and the corresponding reflector curve. If we get a detection or not, depends on if the reflector is obstructed or not. There can also be other objects that occasionally give false reflections. The detections will occur at a non-uniform sampling rate.

3.3 Estimator design

The position is updated with a discrete extended Kalman filter. Using the motion model (3.6) and measurement model (3.10), the design is straightforward. We will not rederive the filter equations here. They can be found in for instance [3] or [4] and in our forthcoming report [5]. In this paper we only describe the principles of the filter.

Suppose we want to estimate the state $(x,y,\theta)^T$ at time t_{k+1}. The previous state estimate was made at time t_k and determined by the measurements up to and including that time. We denote this estimate by $\hat{X}(k \mid k)$. This estimate and the motion model is used to predict the state at the time when the next sampling is made. The predicted state is notated $\hat{X}(k+1 \mid k)$ and is given by (3.3), or for short sampling intervals or small steer angles by (3.4). In general we can write this as

$$\hat{X}(k+1 \mid k) = f(\hat{X}(k \mid k), u_1, u_2) \tag{3.11}$$

At this stage we have to determine from which reflector the angle originates. At present we use a preliminary association method which is described in the next section. Let us assume that we have associated the angle γ_i with reflector j. The association problem is to make sure that $j=i$, where i is the true reflector number. No matter if the association was successful or not, the predicted measurement at time t_{k+1} will be

$$\hat{\gamma}_j(k+1 \mid k) = h_j(\hat{X}(k+1 \mid k) \tag{3.12}$$

We then correct the prediction of the state according to

$$\hat{X}(k+1 \mid k+1) = \hat{X}(k+1 \mid k) + K(k+1)(\gamma_i(k+1) - \hat{\gamma}_j(k+1 \mid k)) \tag{3.13}$$

where $K(k) = (K_1(k), \ K_2(k), \ K_3(k))^T$. If no angle is measured at time t_{k+1} then the gain vector is $K(k) = 0$, otherwise it is determined by linearizing the nonlinear system, given by (3.6) and (3.10), around a nominal value. In our case we choose the estimated state as nominal value. A Kalman filter is then designed for the linearized system. We then use the calculated filter gain on our nonlinear process.

3.4 Association of measured angles with reflectors

When an angle is measured, it has to be associated with a reflector or discarded. Only the angles that can be associated with reflectors are used for the update of the position. At present we use an association method based on knowledge of an approximate value of the position and orientation. This is determined using the motion model (3.11) with the set values of the control signals. The predicted angles to a number of reflectors is then determined with (3.12). If the measured angle falls into an angular tolerance band around the predicted angle to one reflector, we assume that this is the one.

This association routine will work satisfactory if we only have 6-7 reflectors in the room. The corresponding window to reflectors close to the vehicle can be allowed to be rather large, cf. fig 5. But if it is made too large, we will have problems with false detections.

3.5 Estimating the initial position

In this version of the system, the initial position is determined with a method proposed in [2]. From the unknown position the vehicle will only be able to measure the angle to perhaps 60-70% of the reflectors in the room. The rest will be obstructed by different objects. The position can be calculated using triangulation if we know the angles to three reflectors. So we start with assuming that the first three angles comes from the first three reflectors. This gives a position in the room. If this position is correct, there should be reflectors located in the directions given by the rest of the angles. In this way we try all possible combinations of angles and reflectors, until we find a position where we can associate all angles with reflectors.

To reduce the number of possible combinations, we assume that the number of reflectors in a room is limited to 15. All reflectors are located on the walls so the order between them never change. We also assume that there are no false detections among the measured angles. If there is a false value, the proposed method will fail. The vehicle then has to be moved and the procedure repeated. However, when we have made tests on our experimental setup, this problem seldom occurs. When the vehicle has found the initial position, false detections cause little problem.

3.6 System overview

An overview of the navigation and the control system is given in fig 6. We have already described the hardware and the navigation system. We will now briefly summarize each component in the system. In the next section we will turn to the control of the AGV, i.e. how the set values of the speed and steering angle are determined. These set values are used in the position estimator to predict the position of the AGV at the time instant when an angle is measured. The angle originates from reflector i. The predicted position and the map of reflector coordinates are then used to associate the measured angle with reflector number j and correct the estimated position. The estimated position is compared with the desired position and new set values of the control signals are determined. In this figure we can note that we do not need any measurements of the rear wheels movements. This will however be implemented in a future version of the system to increase the accuracy further.

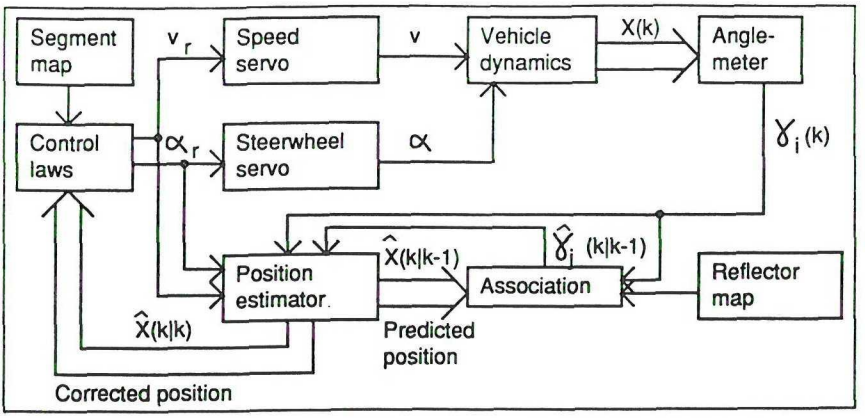


Fig 6. Block scheme over the different components in the navigation and control system.

4 CONTROL LAWS

The variables that are controlled, are the speed and the steer angle of the single front wheel. The control laws are designed to take the following requirements into account:

A The vehicle should follow the drive path.
B The speed should be as high as possible.
C The change of the controlled variables from one sampling to another should lie within certain limits.

The first and second points are obvious. One reason for the third requirement is that the acceleration of the vehicle can cause slipping of some wheel. The estimator does not model the dynamics in the servos or slipping. If there are large changes in the set values to the servos or slipping occurs, this may result in poor estimates of the position and heading. Another reason is that large changes of the set values to the servos cause wear of the mechanical components of the vehicle.

4.1 Guidance control

The drive path is given in the memory of the onboard computer as a list of coordinates (x_s, y_s). It is defined by drawing straight lines between neighboring coordinates [6]. One such line is referred to as a segment. Since the vehicle is supposed to follow the segments, it is natural to transform the position and heading received from the estimator to a position (x_s^*, y_s^*) and heading θ_s^* in a local coordinate system around the segment s which the vehicle is following. $x_s^*(k)$ is the orthogonal projection of the position of the front wheel onto segment s, $y_s^*(k)$ is the perpendicular distance from the segment to the front wheel and $\theta_s^*(k)$ is the heading of the vehicle relative to the segment. This information is used in the guidance controller. The set value to the steer servo is given by

$$\alpha_r(k) = Atan2(-y_s^*(k), d_s) - \theta_s^*(k) \quad , \quad -\alpha_{max} \le \alpha_r(k) \le \alpha_{max} \tag{4.1}$$

The interpretation of (4.1) is that the steer wheel is aimed at a point on the segment which is the distance d_s away from the projection x_s^*. The aiming point is chosen such that large changes in the set value of the steer angle are avoided. α_{max} is the maximum allowed steer angle.

When the vehicle starts to follow the drive path, the distance d_s is small if the vehicle is far away from the segment so that the vehicle will travel almost perpendicular to it. When the vehicle is close to the segment and almost parallel to it, d_s can not be too small. If that is the case, or the speed compared to the sampling time is large, there will be oscillatory behavior in the control signal since the control is done in discrete time.

The way we define the drive path has the advantage that it is easy to design and that it will occupy a small amount of memory in the computer. The disadvantage is the sharp corners between the segments. It is not desirable that the vehicle makes such sharp turns. The vehicle starts to follow the new segment before the end of the old is reached to prevent overshoots in the path the vehicle is taking. If we call the distance from the end of the segment where the vehicle starts to follow the new segment l_s^* and the length of the segment l_s, we have the condition for switching between segments

$$x_s^*(k) \geq l_s - l_s^* \tag{4.2}$$

The angle between segment s and segment $s+1$ is λ_s. l_s^* is calculated as

$$l_s^* = c \cdot |\lambda_s| \quad , \quad -\pi \leq \lambda_s \leq \pi \tag{4.3}$$

where c is chosen such that the vehicle makes a smooth turn when it starts to follow the new segment, $s+1$. To prevent a large change of the steer angle when a switch to a new segment is made, d_s depends on how the vehicle is located relative to the new segment and also on the steer angle at the time instant when the segment change is made. Fig 7 visualizes (4.1) and shows the aiming distance as a function of y_s^*.

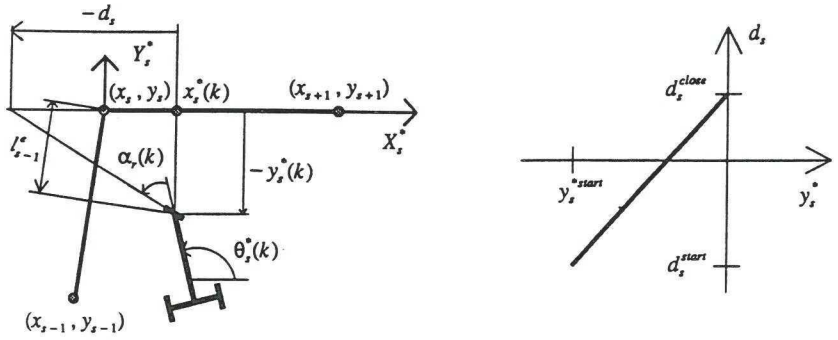


Fig 7. Here equation (4.1) is visualized. The guidance controller calculates at each sampling instant a set value to the steer servo such that the steer wheel is aimed the distance d, away from the projection of the vehicle's position on the segment. In this figure we see the vehicle at the sampling instant when it leaves segment s-1 and starts to follow segment s. The segments are defined in the global coordinate system by the coordinates (x_{s-1}, y_{s-1}), (x_s, y_s) and (x_{s+1}, y_{s+1}). To the right we can see the aiming distance d_s on segment s as a function of the front wheel's perpendicular distance to it.

In fig 7 we see the vehicle at the sampling instant when it leaves segment s-1 and starts to follow segment s. The aiming distance is chosen at this sampling instant such that it is the distance between the projection $x_s^*(k)$ on the new segment, (in this case segment s), and the point where the tangent of the front wheel crosses the x-axis of the new segment before the set value has been changed. This value of d_s is called d_s^{start} in the right figure, where we can see the aiming distance d_s on segment s as a function of the front wheel's perpendicular distance to it. y_s^{*start} is the value $y_s^*(k)$ when the vehicle starts to follow the new segment. d_s can be positive or negative depending on how the vehicle is located relative to the segment and the steer angle when a switch is made to a new segment. d_s^{close} is the aiming distance when the perpendicular distance to the segment is zero. When the steer angle and heading to the segment are small one can linearize and discretize the closed loop system, assuming no dynamics in the servos. It turns out that $1/d_s$ and 1 are the feedback parameters in a state-feedback regulator. d_s^{close} is the value of d_s that gives the linearized closed loop system proper poles. Choosing d_s as in the right figure prevents large changes in the steer angle at switching time and at any other sampling instant if the sampling time compared to the speed is low.

4.2 Speed control

The linear and the angular accelerations of the vehicle depend on the steer angle, the speed of the front wheel and their derivatives with respect to time. One possible control strategy is to limit the acceleration to prevent slipping by choosing suitable set values to the servos. We have chosen to calculate the set value to the steer servo according to (4.1), which leaves the set value to the speed servo as the only

remaining variable left for limiting the acceleration. The guidance controller is designed such that the derivative of the steer angle is low. This property cooperates with the ambition of the speed controller to limit the acceleration, and also allows a higher speed since the acceleration partly depends on the derivative of the steer angle.

One can use models of the vehicle and the servos to predict the acceleration that a certain set value would give rise to and chose a set value to the speed servo that will not cause slipping. However, estimating the acceleration is not an easy task. Therefore we limit the acceleration using the rule of thumb.
The set value to the speed servo is calculated as a function of the set value to the steer servo as shown in fig 8. The angular velocity of the vehicle gives rise to forces acting perpendicular to the wheels. If these forces exceeds certain values, slipping occurs. Therefore the speed has to be reduced when the steer angle is large, at least when the maximum allowed speed is high.

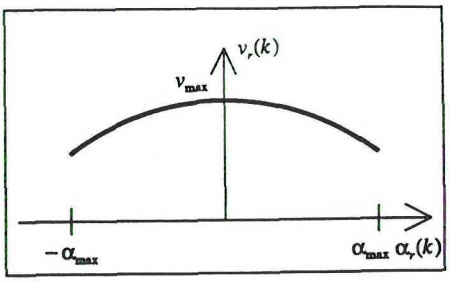


Fig 8. The set value to the speed servo as a function of the set value to the steer servo.

There is also a restriction in the change of the set value to the speed servo from the previous sampling instant to the current sampling instant according to

$$|v_r(k) - v_r(k-1)| \leq \Delta v_{lim} \tag{4.4}$$

This is to limit the acceleration of the vehicle. This function is mainly active at start-up. At that time the speed is 0 m/s, and if the steer angle is 0°, the set value to the speed servo will be maximum speed according to fig 8. This can lead to slipping of the front wheel.

Calculation of the set value to the speed servo according to fig 8 and (4.4) guarantees that the speed control is smooth if the guidance control is smooth.

5 SIMULATION RESULTS

The navigation and the control system have been simulated and tested on our AGV prototype with the simulation and control program REGSIM, which is developed at Luleå University of Technology [7]. When simulating the system, we assumed that the speed and steerwheel servos in fig 6 were first order systems. The time constant was set to 0.2 s for the speed servo and 0.1 s for the steerwheel servo. The set value of the speed was set to 0.5 m/s. Equations (3.1) and (3.2) were used to simulate the vehicle dynamics.

The room has the dimensions 9x5.5 m^2. There are eleven reflectors on the walls, marked as dash-dotted lines in the fig 9, and five tables that the AGV should avoid to move. Since the mirror rotates with 1 Hz, the mean time between angle measurements would be approximately 0.1 s if only a few of the reflectors are obstructed. To simplify the simulation routines we assumed that angles were measured with the sampling interval 0.1 s. In tests on the prototype we have a non-uniform sampling interval.

In fig 9 and fig 10 we see the result from two different simulations. We assumed that the initial position had 50 mm error in the x- and y-coordinates and 0.05 rad error in the heading. In reality the initial position will be determined with high accuracy since it is determined using a lot of angles measured from one point.

In fig 9 we see the result when no angles were used for the update of the position. Neither the initial errors nor the accumulated errors due to model errors were removed. The estimated trajectory of the AGV is shown with a dashed line. We can also see the function of the guidance controller. The initial position is not on the first segment. The guidance controller will make the estimated position follow a smooth path until it reaches the segment. A change to the next segment is made before the end of a segment is reached. The control laws for guidance control and speed control are further described in [8].

In fig 10 we see the result when measured angles were used to correct the estimated position. We assumed that the measured angles had been associated with the correct reflectors. Errors of the same magnitude as the assumed initial errors can however occur when the AGV is moving if an erroneous association is made. In this simulation the initial errors are reduced. This indicates that the system can handle up to around 10% false detections and erroneous associations. The errors in position is less than 50 mm during the 30 s simulation. The errors can be reduced by increasing the rotational speed of the anglemeter, i.e. have a shorter sampling interval, and measuring the rear wheel movements.

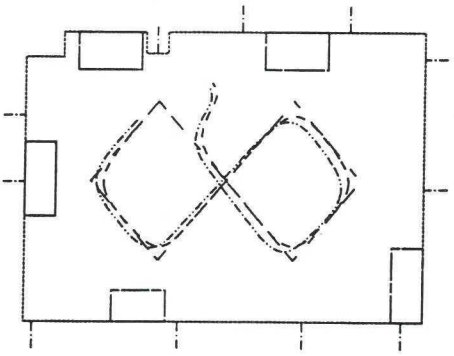


Fig 9.The estimated and simulated trajectory differs when no angles are used for the update of the position.

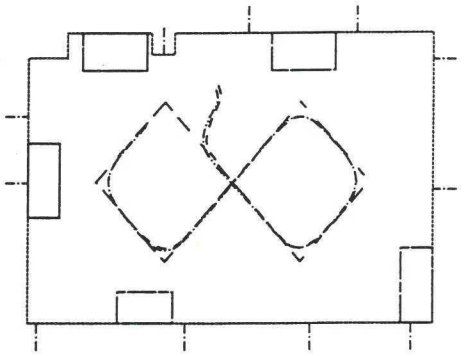


Fig 10. In this figure angles are used. We can also see how the vehicle follows and changes between segments.

6 SUMMARY

We have described an optical navigation system for an AGV based on directional measurements to several identical beacons. The position and heading of the AGV is updated recursively each time a valid angle is measured. The drive path that the AGV should follow consists of straight line segments between given points. This gives a very flexible system where it is easy to change the transportation routes. We also described the control laws that are used for segment following, segment change and speed control.

The system has been tested on an AGV prototype in the room shown in fig 9 and fig 10. It is however difficult to present any results from these tests, since we do not have any way to measure the position of the AGV when it is moving, except using the position estimator. At present we use a personal computer for all the calculations. This increases the computational times. Therefore we have only tested the system with the speed 0.3 m/s. In a future version we will use a Motorola 68020 with a 68881 floating-point coprocessor.

Our present research concerns a thorough error analysis of the position estimator, a new method to determine the initial position and a robust association method. The anglemeter will be redesigned to enable a rotational speed of at least 5 Hz.

7 ACKNOWLEDGMENTS

We want to thank Professor Åke Wernersson and Professor Jan Sternby, Luleå University of Technology, for helpful discussions.

This project is supported by the National Swedish Board for Technical Development.

8 REFERENCES

[1] Hyyppä, K., "Optical Navigation System Using Passive Identical Beacons", Proc. Intelligent Autonomous Systems, Amsterdam, 8-11 December, 1986, pp 737-741, North-Holland (1987).

[2] Hyyppä, K. and Wiklund, U., "Optical Navigation System Using Passive Identical Beacons", Technical Report 1987:11T, Luleå University of Technology, S-951 87 Luleå, Sweden (1987).

[3] Jazwinski, H.J., "Stochastic Processes and Filtering Theory", Academic Press (1970).

[4] Anderson, B.D.O. and Moore, J.B., "Optimal filtering", Prentice-Hall (1979).

[5] Wiklund, U., "Navigation by Angle Measurements to Several Identical Beacons", Report, Luleå University of Technology, S-951 87 Luleå, Sweden, (to be published).

[6] Julliere, M., et al, "A guidance system for a vehicle which has to follow a memorized path", Proc. Automated Guided Vehicle Systems, Stuttgart, W.Germany, 7-9 June, 1983.

[7] Gustavsson, T., "REGSIM - User's manual", Report, Division of Automatic Control, Luleå University of Technology, S-951 87 Luleå, Sweden.

[8] Andersson, U., "A Control Strategy for Autonomous Guided Vehicles", Report, Luleå University of Technology, S-951 87 Luleå, Sweden, (to be published).

Part III

A low-noise photodiode-amplifier circuit

A low-noise photodiode-amplifier circuit

Kalevi Hyyppä and Klas Ericson
Luleå University of Technology
S-971 87 LULEÅ, Sweden

April 1993

Abstract

A photodiode-amplifier circuit with the photodiode in the feedback path is presented. It is named the PIF-circuit. No resistor is needed at the amplifier input to provide a path to ground for the signal and leakage currents from the photodiode and the amplifier input bias current. Thereby one potentially dominating noise source is eliminated. Small-signal and noise properties of the circuit are analysed. A first order approximation of a matched filter for square pulses can be implemented directly with the PIF-circuit and a cascaded first order low-pass whitening filter. Simulations and measurements on an implementation of the PIF-circuit show good agreement with the analysis. At frequencies below 10 kHz the implemented PIF-circuit has a NEP ≈ 3 fW/$\sqrt{\text{Hz}}$. The noise model of a photodiode is discussed.

1 Introduction

Several papers and books have been written about low-noise amplifiers in general. Two often cited references are [1] and [2]. Low-noise amplifiers are often used to amplify the signal from a sensor of some kind. The sensor can have a strong influence on the noise properties of the sensor-amplifier combination. A photodiode presents a high source impedance to an amplifier and this must be taken into account to ensure good results. Another high impedance transducer which motivates similar design considerations is the semiconductor radiation detector.

Veljko Radeka at Brookhaven National Laboratory has published several papers on low-noise amplifiers for radiation detectors [3]-[7]. In paper [6] he gives an in-depth review of low-noise techniques for semiconductor radiation detectors. Reviews by other authors are given in [8] and [9]. Optimum coupling transformer ratio in conjunction with particle detectors is analysed in [10].

A general problem with high impedance sources is low bandwidth due to source and stray capacitances. The transimpedance amplifier presents a low input impedance to the source and thereby lessens the influence of the source capacitance. Noise considerations dictate a high value of the feedback resistor in a transimpedance amplifier and then stray capacitance associated with the feedback resistor can have a detrimental effect on bandwidth. A solution to the stray capacitance problem is given in [11]. Two review-type papers on low-noise photodiode-amplifiers are [12] and [13].

A rule for good low-noise design was presented as early as 1939 in [14]:

> " — the process consists of first obtaining a good signal-to-noise ratio, with little regard to frequency characteristics, and then, after the signal is well above the noise level, correcting the frequency response to conform with the desired characteristic."

In several applications the input signal is in the form of a pulse and the task is to detect the pulse and possibly to measure its amplitude. In these applications low noise should be the primary design objective and not the reproduction of the actual pulse-form.

A common method to achieve low noise properties in an amplifier is to connect many input devices in parallel. A design using 8 pairs of the JFET J309 and the BJT 2SB737 in the input stage is reported in [15]. The measured performance was a voltage noise of 0.42 $nV/\sqrt{Hz}$ and a current noise of 2.8 $fA/\sqrt{Hz}$ at 1 MHz. A differential amplifier, using 3 JFET:s of the type 2SK162 at each input is described in [16]. The authors claim a voltage noise of 0.5 $nV/\sqrt{Hz}$ and a current noise of 35 $fA/\sqrt{Hz}$ at 70 kHz. These results concern only the amplifier. When the amplifier is combined with a sensor, additional noise in the sensor and in interface components can dominate over the amplifier noise.

Common photodiode-amplifier circuits have a resistor connected to the amplifier input. The resistor functions as a load element for the photodiode or as a feedback element for the amplifier. The resistor also provides the necessary DC-path to ground for signal and leakage currents from the photodiode and bias currents from the amplifier input. The resistor produces thermal noise. The noise contribution decreases with increasing resistance. The upper limit of resistance is set by the allowed amplifier output offset voltage produced by the DC-current at the amplifier input.

Several circuits have been proposed which eliminate the resistor. Optoelectronic feedback is used in [17] and [18]. The accumulated charge at the gate of the FET in the input stage can be injected to the FET channel by a very short pulse which momentarily forward-biases the gate [19]. The fact that the gate leakage current in a FET is somewhat dependant on the gate-drain voltage is used in [20] and [21]. A special purpose FET with an extra

PN-junction to the channel is used in [22]. Minority carriers can be injected into the channel through the extra PN-junction, thereby controlling the gate leakage current.

In the circuit proposed in this paper the resistor is eliminated by placing the photodiode in the feedback path. The circuit has been given the acronym PIF from: **P**hotodiode **I**n **F**eedback. Only one paper [23] has been found which describes a similar connection of the photodiode. The main topic in that paper, however, is bandwidth and the low-noise benefit of the photodiode placement is not treated.

An implementation of the PIF-circuit, discussed below, is designed to detect pulses with pulse lengths in the interval 0.1–1 ms. Most of the power of such pulses is within the frequency band 100 Hz to 10 kHz. The circuit models to be presented have been chosen to be valid in the frequency band 10 Hz to 100 kHz.

In Section 2 the PIF-circuit is presented. In Section 3 small-signal and noise analyses of the PIF-circuit are presented. An implementation of the PIF-circuit is described in Section 4. Details of the implemented PIF-circuit are given in an appendix. Simulation and measurement results on the implemented PIF-circuit are given in Sections 5 and 6. In Section 7 the properties of the PIF-circuit are discussed. Comparisons are made with two common photodiode-amplifier circuits. It is also shown how an approximation of a matched filter can be implemented. Finally in Section 8 some conclusions are made.

2 The PIF-circuit

The gate current of a FET and the dark current of a photodiode are of the same nature — they are both leakage currents in reverse biased PN-junctions. With proper selection of devices and bias voltages the currents can be made equal. The correct bias can be achieved automatically by letting the photodiode itself constitute the feedback path for an operational amplifier with FETs in the input stage. This circuit is shown in Figure 1. The bias point

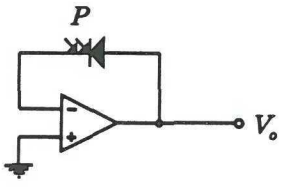


Figure 1: A circuit which have the photodiode in the feedback path.

of the input FET is only weakly dependent of the output voltage if the gain of the amplifier is high. The FET gate current is then almost independent of the output voltage. The photodiode leakage current is controllable with the reverse voltage over the photodiode, i.e. the output voltage. At a certain output voltage the two currents are equal. This voltage becomes the quiescent output voltage of the circuit.

The capacitance and resistance of the photodiode creates a pole in the transfer function for this circuit. The corresponding break frequency is typically below 1 Hz. The output signal level, above 1 Hz, might then be too low to be above the noise level of the following stages. A useful modification of the circuit is to feed only a small portion of the output voltage back to the photodiode at frequencies above the break frequency given by the mentioned pole. The DC-voltage at the amplifier output which controls the leakage current through the photodiode is fed back unaltered. The modified circuit is the PIF-circuit. It is shown in Figure 2. The requirement to feed only a small portion of the output voltage back to the photodiode at frequencies, above the break frequency results in the condition $R_1 \gg R_2$ on the feedback resistors. An additional requirement on the resistors is that they should be chosen with sufficiently low resistances to make their contributions to the circuit noise insignificant.

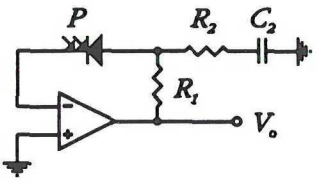


Figure 2: The PIF-circuit. The feedback network (R_1, R_2 and C_2) enhances the high frequency gain of the PIF-circuit.

A radiation pulse generates a current pulse in the photodiode. The photocurrent pulse charges the junction capacitance in the photodiode provided the photodiode is reverse biased. A stronger and longer photocurrent pulse will reduce the reverse voltage and eventually some of the photocurrent will be wasted in the decreasing dynamic resistance of the photodiode. The photodiode can even be forward biased. This fact can be used to advantage. When the diode becomes forward biased its dynamic resistance decreases exponentially and therefore efficiently shunts the photocurrent away from the photodiode capacitance. Through this non-linear behavior the PIF-circuit can handle incoming optical pulses with a large dynamic range without saturating. We shall, however, not pursue this matter any further but concentrate on the linear behavior of the PIF-circuit.

3 Small-signal analysis

3.1 Transfer function

A linear small-signal model of the PIF-circuit is shown in Figure 3. The photodiode is represented by the photocurrent generator I_s, the capacitor C_d the parallel resistor R_d and the series resistor R_s. The capacitor C_d represents the photodiode junction capacitance. It decreases with increasing reverse bias of the photodiode. The resistor R_d represents the dynamic resistance of the photodiode. The dynamic resistance is composed of the differential resistance of an ideal diode, described by the Schockley equation, and of some other phenomena which will be treated in a later section. The dynamic resistance increases with increasing reverse bias of the photodiode. The series resistor R_s is mainly composed of the bulk resistance in the silicon between the junction and the top-side contact to the photodiode chip. The series resistance is only weakly dependent of the photodiode bias.

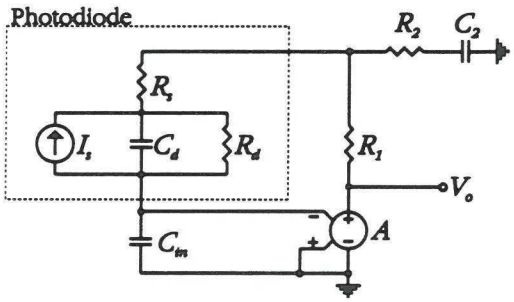


Figure 3: Small-signal equivalent model of the PIF-circuit. The photodiode is represented by the photocurrent generator I_s, the capacitor C_d and the resistors R_d and R_s. The input capacitance of the operational amplifier is represented by C_{in}.

The operational amplifier in the PIF-circuit is represented by the input capacitor C_{in} and a controlled voltage generator. A first order approximation is used for the transfer function of the controlled voltage generator.

$$A(s) = \frac{A_0}{1 + s\tau_0} \tag{1}$$

where A_0 denotes the low-frequency voltage gain of the operational amplifier, τ_0 denotes the time constant of the transfer function and s denotes complex angular frequency. The real part of the input conductance of typical operational amplifiers with FETs in the input stage is small enough to be omitted

in the current frequency band. The output impedance of the operational amplifier have also been omitted because the magnitude of the open-loop output impedance is much smaller than the magnitude of the total loading impedance at the output node of the operational amplifier.

The transfer function from photocurrent to output voltage for the PIF-circuit becomes

$$H_{PIF}(s) = \frac{R_d\,(1 + s\,\tau_1)}{\tau_t \tau_1 \tau_d' s^3 + \tau_2 \tau_d\, s^2 + (\tau_2 + \tau_d)\, s + 1} \tag{2}$$

where

$$\tau_t = \frac{\tau_0}{A_0} \tag{3}$$

$$\tau_1 = R_1 C_2 \tag{4}$$

$$\tau_2 = R_2 C_2 \tag{5}$$

$$\tau_d = R_d C_d \tag{6}$$

$$\tau_d' = \tau_d \left(1 + \frac{C_{in}}{C_d}\right) \tag{7}$$

In the derivation of (2) a number of terms have been cancelled due to the conditions below. They are valid for a correctly designed PIF-circuit.

$$A_0 \gg 1 \tag{8}$$

$$\tau_t \ll \tau_1, \tau_2, \tau_d \tag{9}$$

$$R_s C_d \ll \tau_1, \tau_2, \tau_d \tag{10}$$

$$R_1 \ll R_d \tag{11}$$

$$R_2 \ll R_1 \tag{12}$$

The symbol R_s for the series resistor is not represented in (2) because of the condition (10). The resistor is included in the equivalent circuit because of the noise associated with it. It will be discussed in the section on noise.

A general expression for the denominator in (2) is $(1 - s/p_1)(1 - s/p_2)(1 - s/p_3)$ where p_1, p_2 and p_3 denote the poles of (2). By identifying the two expressions for the denominator it can be shown that $p_1 \approx -1/\tau_d$, $p_2 \approx -1/\tau_2$ and $p_3 \approx -1/\tau_3$ where

$$\tau_3 = \tau_t \left(1 + \frac{C_{in}}{C_d}\right) \frac{R_1}{R_2} \tag{13}$$

These approximate values of the poles of the PIF-circuit transfer function are valid if $\tau_3 \ll \tau_1, \tau_2$ which hold for a correctly designed PIF-circuit. An approximate expression for the transfer function then becomes

$$H_{PIF}(s) \approx \frac{R_d\,(1 + s\,\tau_1)}{(1 + s\,\tau_d)(1 + s\,\tau_2)(1 + s\tau_3)} \tag{14}$$

3.2 Noise analysis

A noise equivalent circuit of the PIF-circuit is shown in Figure 4. In the common noise model of a photodiode the noise is dominated by the shot noise from carriers that cross the PN-junction [1], [13]. The shot noise is represented by a current noise source I_{nd} connected in parallel with the photocurrent source. The series resistor R_s generates thermal noise. The noise is represented by a voltage noise source V_{nR_s} connected in series with the resistor. The issue of possible thermal noise in the dynamic resistance R_d will be discussed later. The amplifier noise is represented by a voltage noise source V_{na} and a current noise source I_{na} at the amplifier input. The thermal noise associated with the resistors in the feedback network are represented by the voltage noise sources V_{nR_1} and V_{nR_2}.

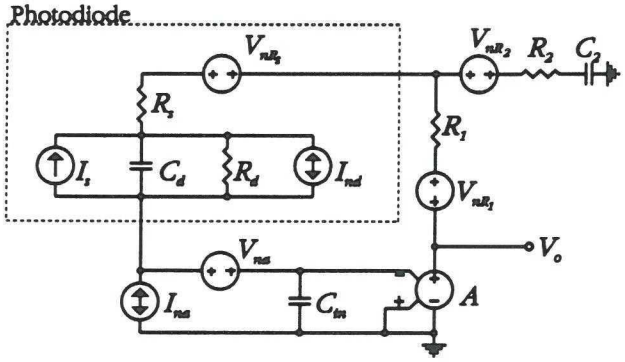


Figure 4: Noise equivalent circuit of the PIF-circuit. The symbols I_{nd}, I_{na}, V_{nR_s}, V_{nR_1}, V_{nR_2} and V_{na} denotes the noise generators in the PIF-circuit.

The symbols I_{nd}, I_{na}, V_{nR_s}, V_{nR_1}, V_{nR_2} and V_{na} should be interpreted as the square roots of the spectral densities of the currents and voltages of the respective noise generators. The spectral density of the shot noise current in the photodiode is given by [1]

$$I_{nd}^2 = 2e(|I_D| + 2I_0) \tag{15}$$

where I_D denotes the DC-current through the photodiode, I_0 denotes the saturation current of the photodiode and e denotes the electron charge. The spectral densities of the thermal noise voltages in the resistors R_s, R_1 and R_2 are given by [1]

$$V_{nR_s}^2 = 4kTR_s \tag{16}$$

$$V_{nR_1}^2 = 4kTR_1 \tag{17}$$

$$V_{nR_2}^2 = 4kTR_2 \tag{18}$$

To enable a simple comparison between the signal power and the noise power, the total noise of the PIF-circuit is represented by a current noise source connected in parallel with the photocurrent source. With the usual assumption that all noise sources are uncorrelated the spectral density of the total noise of the PIF-circuit becomes

$$I_{nPIF}^2 = I_{nd}^2 + I_{na}^2 + \left(V_{na}^2 + V_{nR_s}^2 + k_1 V_{nR_1}^2 + k_2 V_{nR_2}^2\right)\left(\frac{1}{R_d^2} + \omega^2 C_d^2\right) \tag{19}$$

where

$$k_1 = \frac{1+\omega^2\tau_2^2}{1+\omega^2\tau_1^2} \tag{20}$$

and

$$k_2 = \frac{R_1^2}{R_2^2} \cdot \frac{\omega^2\tau_2^2}{1+\omega^2\tau_1^2} \tag{21}$$

The function k_2 is approximately equal to one for frequencies above $1/(2\pi\tau_1)$. This frequency is typically less than 1 Hz in a correctly designed PIF-circuit. Due to the condition (12) the noise contribution from R_1 is much smaller than the contribution from R_2 throughout the frequency range defined in the introduction. The expression for the total noise of the PIF-circuit then becomes

$$I_{nPIF}^2 = I_{nd}^2 + I_{na}^2 + \left(V_{nR_s}^2 + V_{nR_2}^2 + V_{na}^2\right)\left(\frac{1}{R_d^2} + \omega^2 C_d^2\right) \tag{22}$$

4 An implementation of the PIF-circuit

A photograph and the schematics of an implementation of the PIF-circuit are shown in Figures 5 and 6 respectively. The high impedance level at the gate of the input FET creates a need for efficient electrical shielding. The photodiode and the FET are surrounded by a separate shielding visible in the left part of the photograph in Figure 5. The whole circuit is inserted in a light-tight metal box. The light inlet to the photodiode is through a metal tube. The tube length is many tube diameters to ensure good efficiency of the shielding. No special measures were taken to protect the circuit from magnetic interference.

The current generator *Is* in Figure 6 represents the photocurrent of the photodiode *D1*. The feedback components of the circuit are denoted by *R1*, *R2* and *C2*. The PIF-circuit amplifier consists of the cascode *J1-Q1* and the operational amplifier *X1*. The purpose of using a cascode is twofold; it gives low-noise properties to the amplifier and it increases the gain-bandwidth product of the amplifier. Note that the combination of the cascode and

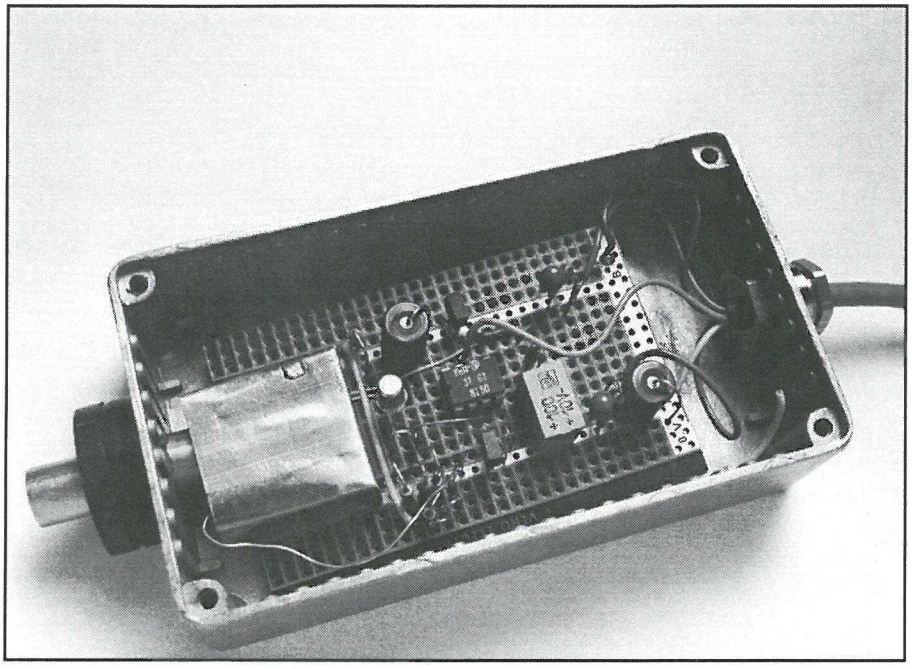


Figure 5: Photograph of the implemented PIF-circuit.

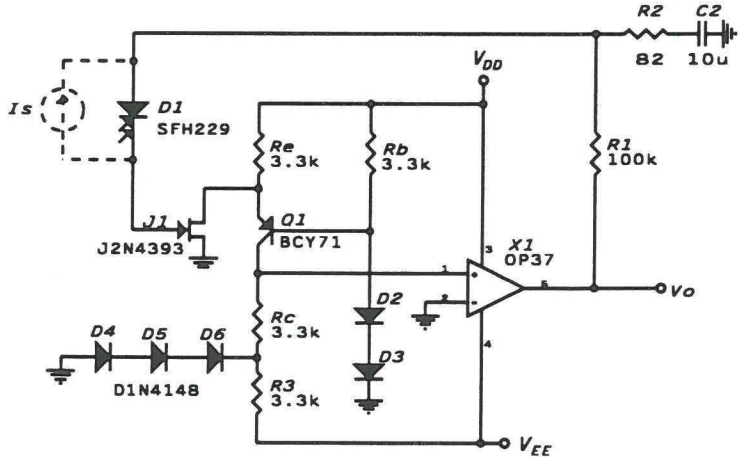


Figure 6: Circuit diagram of the implemented PIF-circuit. The current generator *Is* represents the photocurent in the photodiode.

the operational amplifier in the implemented PIF-circuit corresponds to the operational amplifier symbol in Figure 2. Details of the circuit are given in appendix A.

5 Simulations

To gain insight into the properties of the PIF-circuit a number of SPICE-simulations have been made of the implemented PIF-circuit. Both the linear model, described above, and a complete SPICE model have been used. The parameter values used for the linear model are shown in Table 1. The parameters C_{in}, A_0 and τ_0 are the equivalent input capacitance of the cascode stage and the low frequency gain and the time constant of the transfer function of the cascode stage and the operational amplifier.

R_d	R_s	R_1	R_2	C_d	C_{in}	C_2	A_0	τ_0
150GΩ	5Ω	100kΩ	82Ω	8pF	16pF	10μF	$55 \cdot 10^6$	7.5ms

Table 1: Parameters used in the simulations of the implemented PIF-circuit using the linear model. The parmeters C_{in}, A_0 and τ_0 are the equivalent input capacitance of the cascode stage and the low frequency gain and the time constant of the transfer function of the cascode stage and the operational amplifier.

5.1 Pulse response

The pulse responses of the implemented PIF-circuit, given by the linear model and the SPICE model, are shown in Figure 7. The photocurrent pulse used in the simulations has an amplitude of 10 pA. The rise and fall times of the pulse are 10 μs and the pulse width is 1 ms. The two output voltages are almost identical.

5.2 Frequency response

Even though the most important frequency interval for the implemented PIF-circuit is 10 Hz to 100 kHz, the simulations have been performed over the interval 10 mHz to 10 MHz to illustrate some properties of the PIF-circuit.

The magnitudes of the transfer functions using the linear model and the SPICE model are shown in Figure 8. To illustrate the influence of the photodiode bias point several curves corresponding to different amounts of DC-current in the photodiode — generated by background light — are shown.

Both the curve corresponding to the linear model and the curve corresponding to the SPICE model with no background generated DC-current

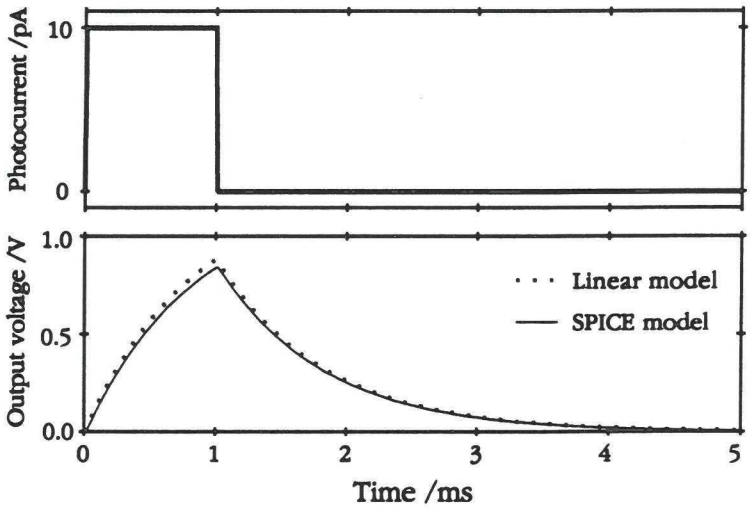


Figure 7: Pulse responses of the implemented PIF-circuit, given by the linear model and the SPICE model. Simulation results.

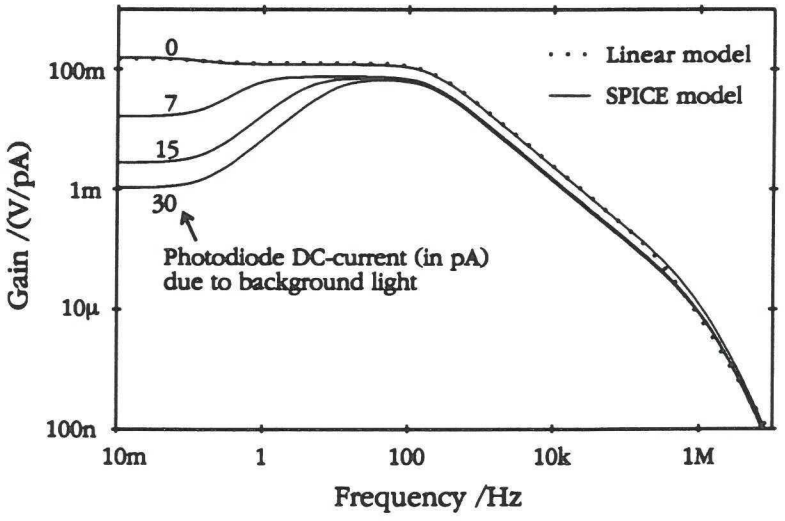


Figure 8: Magnitudes of the transfer function of the SPICE model of the implemented PIF-circuit. The pole with the lowest frequency and the gain constant of the transfer function are dependent on the bias point of the photodiode. Simulation results.

conform well with the approximate expression (14) for the transfer function. This approximation gives break frequencies at 0.16 Hz, 0.13 Hz, 200 Hz and 320 kHz corresponding to the zero n_1 at $-1/\tau_1$ and the poles p_1, p_2 and p_3 respectively. The influence of the pole p_1 is almost canceled by the influence of the zero.

The simulations indicate that the photodiode becomes forward biased for currents larger than 6.5 pA. The change in bias point affects both R_d and C_d which changes the pole p_1 and the gain constant of the transfer function. Above 100 Hz the influence of the background light is only a slight reduction of the gain constant.

5.3 Noise properties

The total noise of the implemented PIF-circuit is represented by a noise current generator connected at the input, in parallel with the photocurrent generator. The square root of the spectral density of this total noise current, given by a SPICE simulation, is shown in Figure 9. The low-frequency noise is 1.4 fA/$\sqrt{\text{Hz}}$ and the break frequency is 10 kHz. For comparison the thermal noise contributions from two fictitious resistors connected in parallel and in series with the photodiode are also shown. Note that the break frequency is not due to any pole or zero in the transfer function of the PIF-circuit. The cause of the break frequency is that above that frequency the voltage noise from the amplifier, coupled by the photodiode capacitance, dominates the total noise.

The derived noise expression (22) has been used to calculate the noise properties of the implemented PIF-circuit. The low-frequency noise became 2.0 fA/$\sqrt{\text{Hz}}$ and the break frequency became 20 kHz. The noise parameters used in the calculations were $I_{nd} = I_{na}$ =1.4 fA/$\sqrt{\text{Hz}}$ and V_{na} =2.0 nV/$\sqrt{\text{Hz}}$.

In this context a remarkable discovery was made; the shot noise in the gate current of a FET is not modeled in the SPICE model of a FET [24]. The shot noise in the gate current should be approximately equal to the shot noise in the photodiode leakage current because the currents are equal. Then both the simulated low-frequency noise and the simulated break frequency should be scaled by a factor $\sqrt{2}$ to give a correct result. Considering this and the general uncertainties in noise parameters the calculated noise conforms well with the simulated noise.

Another view of the proposed PIF-circuit noise properties is shown in Figure 10. There the square root of the spectral density of the output noise voltage is plotted.

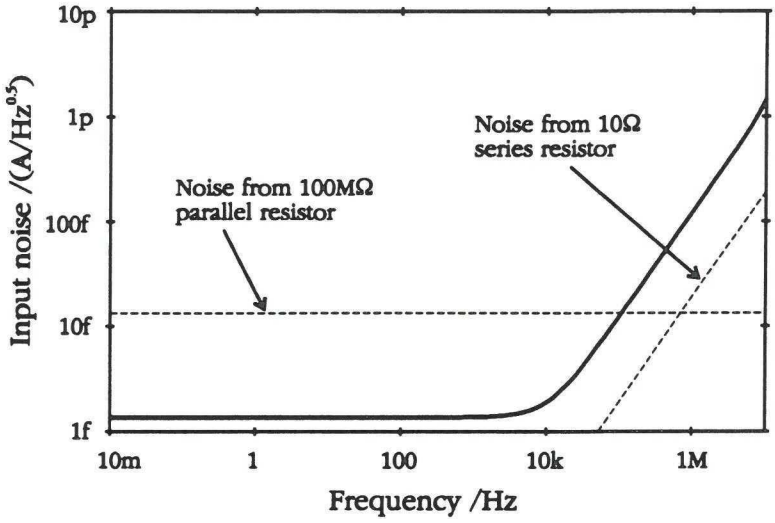


Figure 9: Input noise of the SPICE model of the implemented PIF-circuit. Noise contributions from a small resistor in series with the photodiode and from a large resistor in parallel with the photodiode are also shown for comparison. Simulation results.

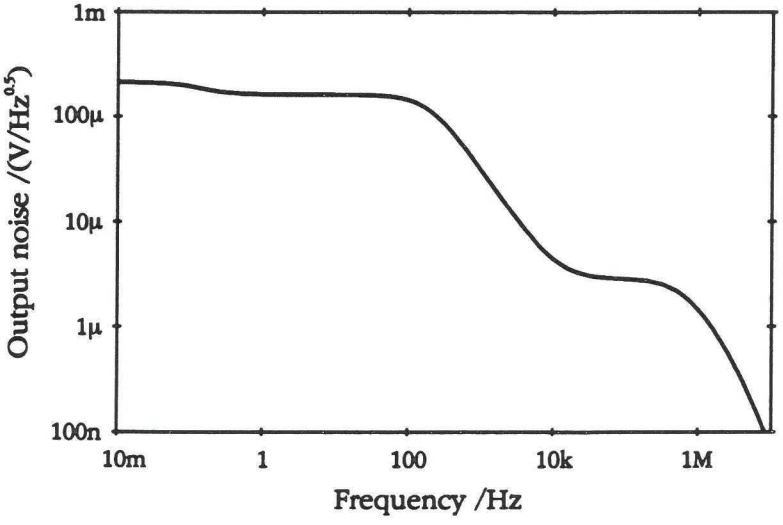


Figure 10: Output noise of the SPICE model of the implemented PIF-circuit. Simulation results.

6 Measurements

Interference through microphony is not a common problem in transistor circuits. However the high impedance level at the input of low-noise photodiode-amplifier circuits make them quite susceptible to microphony. Small vibration-induced variations in the stray capacitances at the input creates minute interference currents which can easily dominate the internally produced noise if the circuit is not rigidly built. To minimize interference through microphony the metal box with the implemented PIF-circuit was placed on a foam-rubber block during measurements to isolate it from vibrations. Even though the implemented PIF-circuit should not be particularly sensitive to magnetic interference the measurements showed that the power supply, feeding the circuit, had to be placed at some distance from the circuit to make the mains-induced interference much smaller than the noise in the output voltage.

An LED-lamp with an LED radiating at $\lambda = 650$ nm was used as a light source for the frequency and pulse response measurements. The custom-made lamp housing contains several aperture stops to ensure that the photodiode is only reached by direct light from the LED. The light coupling metal tube in the implemented PIF-circuit contains also an aperture stop to suppress indirect light.

6.1 Pulse response

The voltage pulse driving the LED was adjusted to produce a current pulse in the photodiode with similar parameters as in the simulations, i.e. an amplitude equal to 10 pA and a length equal to 1 ms. The irradiance from the LED, producing the desired current pulse in the photodiode, was calibrated with the aid of a radiometer. The error in the photodiode current amplitude is estimated to be less than 10 %. The main error source is the uncertainty in the photodiode parameters used for the calibration.

A digital oscilloscope was used to measure the pulse response of the implemented PIF-circuit. The sampling rate of the oscilloscope was 51.2 ks/s in the measurement. The result of the measurement is shown in Figure 11. The noise in the output voltage is visible. The simulated response, based on the SPICE model with no background generated DC-current, is also shown for comparison.

6.2 Frequency response

The LED was driven by a function generator with a DC-bias of 2 V and an added sinusoidal voltage. Measurements of the LED radiation due to the bias voltage, with a radiometer, indicate that the DC-bias of the LED generates approximately 7 pA photocurrent in the photodiode.

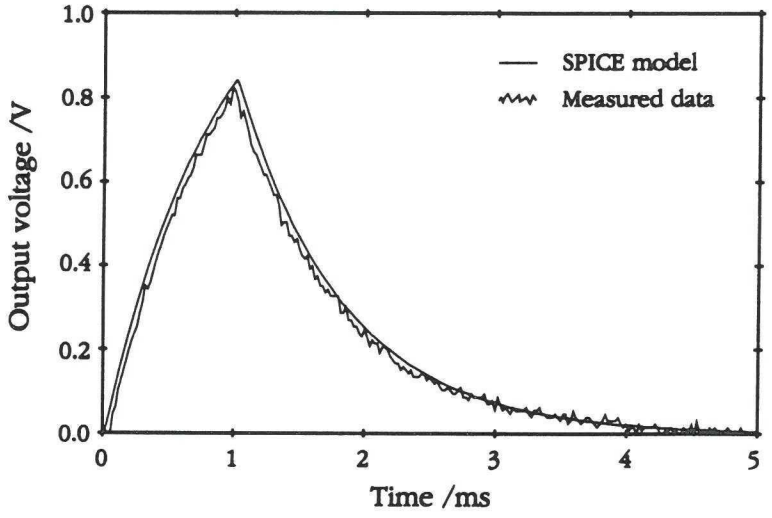


Figure 11: Pulse response of the implemented PIF-circuit. Measured and simulated data.

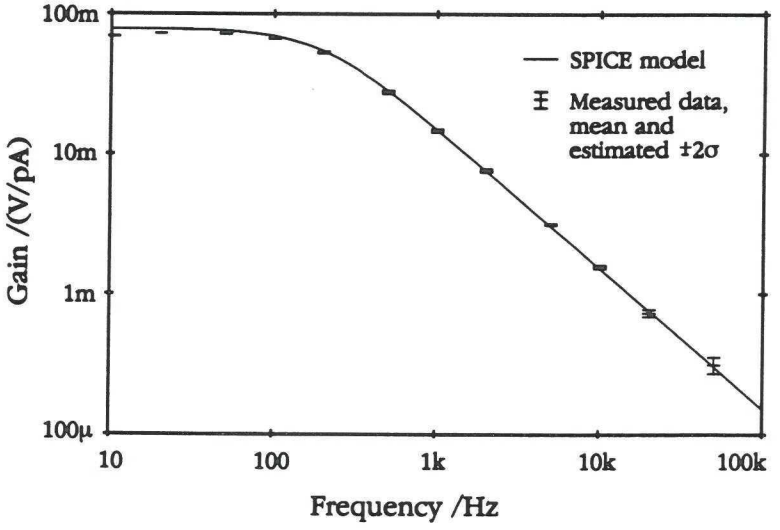


Figure 12: Magnitude of transfer function of the implemented PIF-circuit. The gain constant of the measured data has been adjusted to give a good fit with the simulated data, see the text. Measured and simulated data.

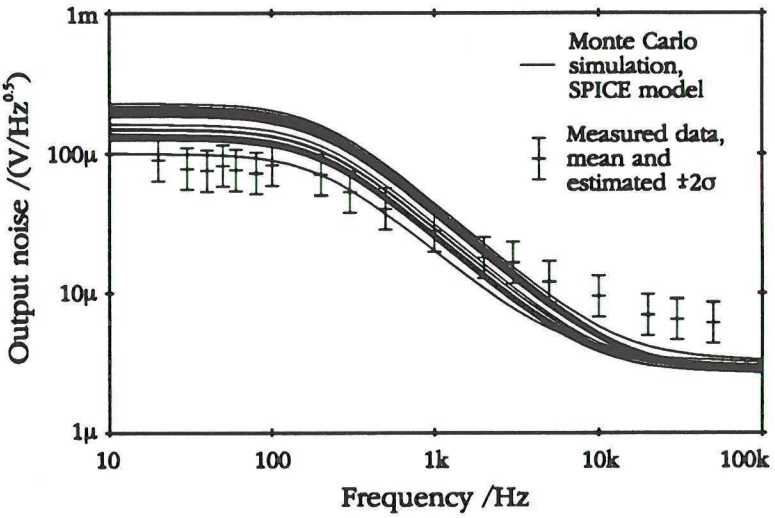


Figure 13: Output noise of the implemented PIF-circuit. Measured and simulated data.

A digital oscilloscope, with a bandwidth of 100 MHz and with the ability to make ensemble averages of repetitive waveforms, was used to measure the frequency response of the implemented PIF-circuit. The averaging capability was useful in getting an adequate signal-to-noise ratio in the measurements for frequencies above 10 kHz. The low-pass character of the implemented PIF-circuit transfer function made the unfiltered signal-to-noise ratio too low above 10 kHz. The result of the measurements is shown in Figure 12. The simulated response of the SPICE model with 7 pA DC-current in the photodiode due to ambient light is also shown for comparison. The measured pulse response indicate that the gain of the implemented PIF-circuit conforms well with the predicted gain. The gain constant of the measured data have therefore been adjusted to give a good fit with the simulated data.

6.3 Noise properties

The spectral density of the output noise voltage of the implemented PIF-circuit was measured with a narrow-band spectrum analyser. The result of the measurements, together with the result from a Monte Carlo simulation based on the SPICE model with no background generated DC-current, are shown in Figure 13.

7 Discussion

7.1 Transfer function

Simulated results using both the linear model and the SPICE model of the implemented PIF-circuit conform well with (14). Considering the uncertainty in component parameters the location of the break point in the measured data conforms surprisingly well with (14), see Figure 12.

For applications with input pulse lengths in the interval 0.1–1 ms some further simplifications of (14) can be made. In the frequency interval 10 Hz to 100 kHz the PIF-circuit transfer function can be approximated by

$$H_{PIF}(s) \approx \frac{\tau_1}{C_d(1+s\tau_2)} \tag{23}$$

An important observation is that R_d has disappeared from the transfer function (23). This is good because the value of R_d is normally poorly known and depends strongly of the photodiode bias point which is ambient light dependent. Although the photodiode junction capacitance C_d is also a function of the photodiode bias point that dependence is much weaker and is furthermore well known [25]. See Figure 8.

7.2 Noise properties

Before discussing the noise properties of the PIF-circuit, expressions will be given on the input referenced total noise of a voltage amplifier circuit and a transimpedance amplifier circuit. These circuits are commonly used with photodiodes. They are named the VA-circuit and the TIA-circuit respectively. Some details of the circuits are given in appendix B. The expressions for the input referenced total noise of the VA-circuit, the TIA-circuit and the PIF-circuit are

$$I_{nVA}^2 = I_{nd}^2 + I_{na}^2 + I_{nR_L}^2 + \left(V_{nR_s}^2 + V_{nR_2}^2 + V_{na}^2\right)\left[\left(\frac{1}{R_d}+\frac{1}{R_L}\right)^2 + \omega^2 C_d^2\right] \tag{24}$$

$$I_{nTIA}^2 = I_{nd}^2 + I_{na}^2 + I_{nR_f}^2 + \left(V_{nR_s}^2 + V_{na}^2\right)\left[\left(\frac{1}{R_d}+\frac{1}{R_f}\right)^2 + \omega^2 \left(C_d + C_f\right)^2\right] \tag{25}$$

$$I_{nPIF}^2 = I_{nd}^2 + I_{na}^2 + \left(V_{nR_s}^2 + V_{nR_2}^2 + V_{na}^2\right)\left(\frac{1}{R_d^2} + \omega^2 C_d^2\right) \tag{26}$$

The most significant deviation in the expression for the PIF-circuit noise is that it does not contain the thermal noise contribution from a resistor R_L or R_f as in the VA- and TIA-circuits respectively. Figure 9 shows the noise

contribution from a parallel resistor of 100 MΩ resistance which could be a typical value for R_L or R_f. Simulations indicate that below 10 kHz the spectral density of the noise from such a resistor is two orders of magnitude greater than the spectral density of the implemented PIF-circuit noise.

Also shown in the figure is the noise contribution from a 10 Ω series resistor. This is included to stress the importance of a low series resistance in the photodiode and a low impedance level in the feedback network. The thermal noise in R_s and R_2 contribute to the total noise in the same manner as the amplifier voltage noise. Another choice of the input FET in the amplifier giving less voltage noise makes a low series resistance in the photodiode and a low impedance level in the feedback network still more important.

The sloping part of the input noise shows that a low value of the photodiode capacitance C_d is important for low-noise properties in all three circuits.

The measured output noise of the implemented PIF-circuit is less than the predicted noise at frequencies below 1 kHz and larger than the predicted noise above 1 kHz. The discrepancy might be too large to be explained only by incorrect parameters used in the models of the implemented PIF-circuit. A cause of this discrepancy could be an erroneous model of the photodiode noise.

The simulations have shown that the noise in the implemented PIF-circuit is dominated by the shot-noise in the photodiode leakage current at frequencies below 10 kHz. When modeling the noise in a photodiode normal practice is to let the leakage current show full shot noise. Shot noise is created when charge carriers cross a PN-junction. But the leakage current of real diodes contains components which do *not* cross a PN-junction. It is well known that leakage currents of real diodes can be orders of magnitude greater than predicted by the bulk diffusion current as given by the Schockley equation [25]. In [26] a thorough investigation of the leakage current of InSb photodiodes is reported. The authors have investigated leakage current contribution from:

- the bulk diffusion current
- the surface diffusion current
- the generation-recombination current in the depletion layer
- the generation-recombination current at the surface
- the current in the surface leakage resistance
- the tunneling current.

The noise contribution from the surface leakage resistance should be treated as normal thermal noise instead of attaching shot noise to the current

passing the surface leakage resistance. Such a noise model gives less noise than the common model. To clarify this matter more research is needed.

Flicker noise has not been mentioned so far. The reason is that it can be made insignificant for frequencies above 10 Hz by proper component choices. No evidence of flicker noise is visible in the measurement results.

A common measure of the noise performance of a photodiode or a photodiode-amplifier combination is its Noise Equivalent Power NEP. It is defined as the optical input power giving a signal-to-noise ratio equal to unity. Some variations of the NEP-definition are used in the literature. A common one is [1]

$$\mathrm{NEP} = \frac{I_n}{R} \tag{27}$$

where R denotes the responsivity of the photodiode and I_n denotes the square root of the spectral density of the input referenced noise current. The responsivity is wavelength dependent. Common practice is to use the maximum value of the responsivity in NEP-calculations. The maximum responsivity of the used photodiode SFH229P is 0.62 A/W at a wavelength of 850 nm. Together with the calculated noise $I_{nPIF} = 2.0$ fA/$\sqrt{\mathrm{Hz}}$ for frequencies below 20 kHz this gives a corresponding NEP = 3.2 fW/$\sqrt{\mathrm{Hz}}$ for frequencies below 20 kHz. The measurements indicate that the implemented PIF-circuit has a somewhat lower NEP.

This compares favorably with the performance of a commercial product [27] which is claimed to have NEP = 20 fW/$\sqrt{\mathrm{Hz}}$ at 20 Hz. Another result is reported from ongoing research at the National Institute of Standards and Technology in Maryland USA [28]. The goal of the research is to develop a radiometric transfer standard. The reported noise performance is NEP = 6 fW/$\sqrt{\mathrm{Hz}}$ at 1 Hz.

The comparison with these results should not be made on a one-to-one basis because the photodiodes used in the cited papers have larger photosensitive areas than the SFH229P. The area of the photodiode SFH229P is 0.3 mm^2 while the area of the photodiode in [27] is 5.1 mm^2 and the area of the photodiode in [28] is 33 mm^2. The NEP increases with photodiode area but the dependence is not linear and other parameters also influence the NEP.

7.3 Matched filtering

The probability to detect a signal pulse, corrupted by white noise, is maximized when matched filtering is used [29]. If the noise is not white a whitening filter must be used before the matched filter. The matched filter must then be modified to be matched to the signal after the whitening filter. For linear systems the sequence of the two filters can be interchanged if it is convenient for some reason. These three cases are illustrated in Figure 14.

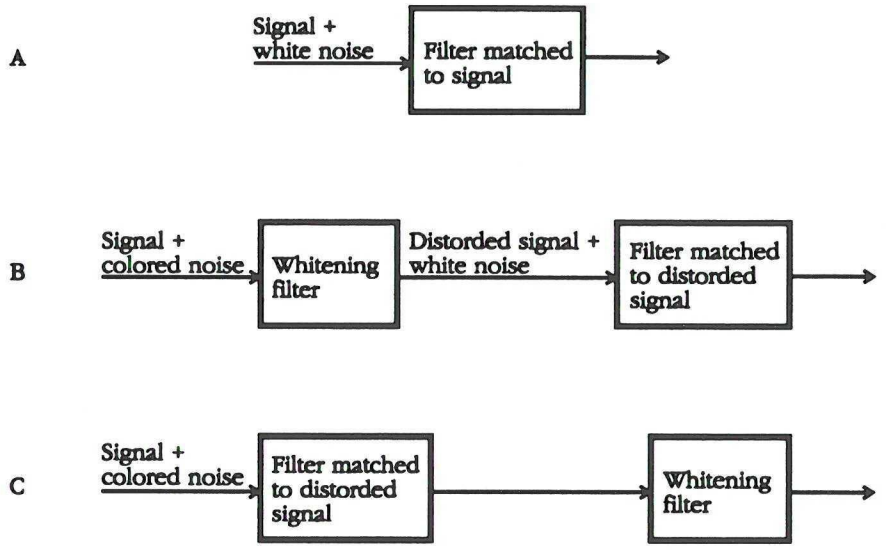


Figure 14: The signal-to-noise ratio is maximized whith a matched filter if the noise is white (A). A whitening filter is used whith colored noise (B). The sequence of the filters can be interchanged (C).

A common approximation of the matched filter when the input signal is a square pulse is a first order low-pass filter. The highest signal-to-noise ratio is achieved when the time constant of the filter is $0.8T$ where T is the length of the signal pulse [29].

The input referenced noise of the implemented PIF-circuit is colored so a whitening filter must be used. Figure 9 and the expression (22) indicates that the whitening filter should be a first order low-pass filter. The noise simulations indicate that the break frequency of the filter should be 10 kHz.

The spectral density for a square pulse of length T has the shape of the function $\left[fT\sin\left(\frac{2\pi}{T}f\right)\right]^2$. Most of the power in the square pulse is within the first lobe of this function i.e. in the frequency interval (0–1) kHz for a square pulse with a length of 1 ms. The whitening filter will thus have negligible influence on the square pulse. Simulations support this. The use of the design rule for the first order approximation of the matched filter when the input signal is a square pulse — with no regard to the whitening filter — is therefore justified.

The approximate expression for the transfer function of the implemented PIF-circuit (23) shows that the first order approximation of the matched filter is obtained directly with the implemented PIF-circuit because the time constant τ_2 has been set to 0.8 ms. The whitening filter is connected in

cascade with the implemented PIF-circuit i.e. as in case C in Figure 14. Simulations show that the whitening filter reduces the noise power at the output of the PIF-circuit with 2 dB.

8 Conclusions

A photodiode-amplifier circuit with the photodiode in the feedback path has been presented. It is named the PIF-circuit. The main benefit of the PIF-circuit is that it does not need a resistor to provide a path to ground for the signal and leakage currents from the photodiode and the bias current at the amplifier input. Thereby a strong noise source is eliminated.

Simulations indicate that the spectral density of the implemented PIF-circuit noise is two orders of magnitude below the noise spectral density from a bias resistor with a resistance of 100 MΩ for frequencies below 10 kHz.

A low series resistance in the photodiode and a low impedance level in the feedback network are shown to be important in minimizing the noise of the PIF-circuit. The photodiode junction capacitance should also be low to minimize the high-frequency noise of the PIF-circuit.

A first order approximation of the matched filter for square pulses can be implemented directly with the PIF-circuit. The colored noise of the PIF-circuit can be adjusted with a cascaded first order low-pass whitening filter.

References

[1] C. D. Motchenbacher and F. C. Fitchen, "Low-noise electronic design", New York, John Wiley & Sons, 1973.

[2] Y. Netzer, "The design of low-noise amplifiers", *Proc. of the IEEE*, vol. 69, no. 6, pp. 728–741, June 1981.

[3] V. Radeka, "Optimum signal-processing for pulse-amplitude spectrometry in the presence of high-rate effects and noise", *IEEE Trans. on Nuclear Science*, vol. 15, pp. 455-470, 1968.

[4] V. Radeka, "Field effect transistors for charge amplifiers", *IEEE Trans. on Nuclear Science*, vol. 20, no. 1, p. 182, February 1973.

[5] V. Radeka et. al., "Design of a charge sensitive preamplifier on high resistivity silicon", *IEEE Trans. on Nuclear Science*, vol. 35, no. 1, pp. 155–159, February 1988.

[6] V. Radeka, "Low-noise techniques in detectors", *Ann. Rev. Nucl. Part. Sci.*, vol. 38, pp. 217–277, 1988.

[7] V. Radeka et. al., "Monolithic, radiation hard charge sensitive preamplifier using diffused N-channel junction field effect transistors", *IEEE Trans. on Nuclear Science*, vol. 38, no. 2, pp. 83–88, April 1991.

[8] E. Gatti and P. F. Manfredi, "Present trends and realizations in readout electronics for semiconductor detectors in high energy physics", *Nuclear Instruments and Methods in Physics Research*, vol. 226, pp. 142–155, 1984.

[9] P. F. Manfredi F. Ragusa, "Low noise electronics in elementary particle physics", *Nuclear Instruments and Methods in Physics Research*, A235, pp. 345–354, 1985.

[10] E. Gatti, P. F. Manfredi and G. Ripamonti, "On the correct choice of transformer ratio in low noise preamplifiers for large detector capacitances", *Nuclear Instruments and Methods in Physics Research*, A257, pp. 331–338, 1987.

[11] I. Pelchowitch and J. J. Zaalberg van Zelst, "A wide-band electrometer amplifier", *The Review of Scientific Instruments*, vol. 23, no. 2, pp. 73–75, February 1952.

[12] R. H. Hamstra Jr. and P. Wendland, "Noise and frequency response of silicon photodiode operational amplifier combination", *Applied Optics*, vol. 11, no. 7, pp. 1539–1547, July 1972.

[13] G. H. S. Rokos, "Optical detection using photodiodes", *Opto-electronics*, vol. 5, pp. 351–366, 1973.

[14] A. A. Barco, "An iconoscope pre-amplifier", *RCA Review*, vol. 4, pp. 89–107, 1939.

[15] E. R. Brown, "Ultra-low-noise high-impedance preamp for cryogenic detectors", *Electronic Letters*, vol. 21, no. 10, pp. 417–418, May 1985.

[16] G. V. Pallottino and T.Lupi, "Low-noise FET amplifier for dc SQUID", *Rev. Sci. Instrum.*, vol. 61, no. 9, September 1990.

[17] K. Kandiah, "Unknown title", Proc. Conf. on Radiation Measurements in Nuclear Power, *Inst. of Physics and the Physical Society*, London 1966.

[18] F. S. Goulding, J. Walton and D. F. Malone, "An opto-electronic feedback preamplifier for high-resolution nuclear spectroscopy", *Nuclear Instruments and Methods*, vol. 71, pp. 273–279, 1969.

[19] V. Radeka, "Charge amplification without charge leak resistor", *IEEE Trans. on Nuclear Science*, vol. 17, no. 3, pp. 433–439, 1970.

[20] E. Elad, "Drain feedback - a novel feedback technique for low-noise cryogenic preamplifiers", *IEEE Trans. on Nuclear Science*, vol. 19, no. 1-2, pp. 403–411, 1972.

[21] G. D. Khomyakov et. al., "X-ray and soft gamma radiation spectrometer preamplifier with drain feedback", *Pribory i Tekhnika Éksperimenta*, no. 2, pp. 138–140, March-April 1975.

[22] T.Nashashibi and G. White, "A low noise FET with integrated charge restoration for radiation detectors", *IEEE Trans. on Nuclear Science*, vol. 37, no. 2, pp. 452–456, April 1990.

[23] E. J. Fjarlie, "Photodiode preamplifier systems: low-noise positive-feedback", *Applied Optics*, vol. 16, no. 2, pp. 385–392, February 1977.

[24] P. Antognetti and G. Massobrio, "Semiconductor device modeling with SPICE", McGraw-Hill Book Company, New York, 1988.

[25] S. M. Sze, "Physics of Semiconductor Devices", 2nd edition, John Wiley & Sons Inc., New York, 1981.

[26] F. K. Hopkins and J. T. Boyd, "Dark current analysis of InSb photodiodes", *Infrared Physics*, vol. 24, no. 4, pp. 391–395, 1984.

[27] "TCN-1000 Ultra-Low-Noise Photodiode/Amplifier", *Data Sheet*, D3015A-2, EG&G Electro-Optics, 1983.

[28] G. Eppeldauer and J. E. Hardis, "Fourteen-decade photocurrent measurements with large-area silicon photodiodes at room temperature", *Applied Optics*, vol. 30, no. 22, pp. 3091–3099, August 1991.

[29] M. Schwartz, "Information transmission, modulation, and noise", New York, McGraw-Hill Book Company, 1980.

A Details of the implemented PIF-circuit

The symbols used in the description are defined in Figure 6. The active components have been selected for their good noise properties. The quiescent points of *J1* and *Q1* have been chosen to get good noise performance and a voltage gain much larger than unity. The design permits the quiescent currents of *J1* and *Q1* to be set independent of each other and independent of their quiescent voltages.

The supply voltages are $V_{DD} = 12$ V and $V_{EE} = -12$ V. Any ripple on the negative supply is efficiently decoupled from the signal path by the resistor *R3* and the low impedance of the diode chain *D4-D6*. The same is true for ripple on the positive supply concerning the ripple coupled to the base of *Q1*. The ripple current injected through *Re* into the emitter of *Q1* is proportional to the conductance of *Re* which means that the resistance of *Re* should be as high as possible. Some efforts have also been made to reduced the ripple on V_{DD} by proper decoupling.

The collector potential of *Q1* is close to ground potential due to the high gain of *X1*. The emitter DC-potential of *Q1* is three diode voltages above ground (*D2, D3* and the base-emitter junction of *Q1*). The emitter-collector DC-voltage of *Q1* and the drain-source DC-voltage of *J1* are therefore fixed at approximately 2 V. The voltage should be high enough to have *J1* and *Q1* working in their saturation regions for high gain. The voltage must not be too high because the gate leakage current of the FET increases with its drain-source voltage. The chosen value is considered to be a suitable compromise.

The DC-voltage over the resistor *Rc* is also fixed at approximately 2 V by the diodes *D4, D5* and *D6*. By neglecting the input bias current of *X1* the collector current of *Q1* becomes approximately −0.6 mA. Best noise performance of the transistor BCY71 is achieved at a collector current of −0.2 mA according to the manufactures data sheet. Simulations and measurements showed that the high-frequency roll-off of the transistor current gain created too much phase shift at such a low current. Therefore the collector current of *Q1* was raised to −0.6 mA.

The current through *Re* is given by $(V_{DD} - V_E(Q1))/R_e \approx 3$ mA with $V_E(Q1)$ denoting the emitter potential of *Q1*. The current divides to the emitter current of *Q1* and the drain current of *J1*. The drain current of *J1* then becomes approximately 2.4 mA. Both the voltage noise and the transconductance of the FET have their best values at a drain current equal to the I_{DSS} of the FET. The minimum guaranteed I_{DSS} of the FET 2N4393 is 5 mA. Measurements have shown that the drain current is not critical to the noise performance. The chosen value of *Re* is a compromise to get a drain current close to I_{DSS} and to get minimum ripple current from V_{DD}. Measurements and simulations showed that the transconductance g_{fs} of *J1* is approximately 9 mS at the chosen drain current. The low-frequency voltage gain of the cascode then becomes $g_{fs}Rc \approx 30$.

The photodiode SFH229P has low dark current. The manufacturer claims that it is typically 50 pA at 10 V reverse bias. The junction capacitance is 13 pF at zero bias. The simulations indicate that the quiescent reverse bias voltage and the dark current of the photodiode in the circuit become approximately 550 mV and 6 pA respectively. The junction capacitance and the parallel resistance are approximately 8 pF and 150 GΩ respectively at this bias voltage. The series resistance is 5 Ω.

B Noise in common photodiode-amplifier circuits

B.1 The voltage amplifier circuit

In the voltage amplifier circuit the photodiode is connected in series with a load resistor R_L. This circuit is named the VA-circuit. The photodiode is reverse biased by a suitable bias voltage V_b. The voltage developed over the load resistor by the photocurrent is amplified with a voltage amplifier. The circuit diagram of this circuit and its noise-equivalent circuit diagram are shown in Figures 15 and 16 respectively.

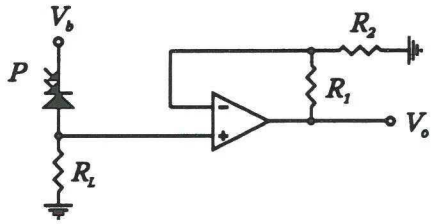


Figure 15: The VA-circuit.

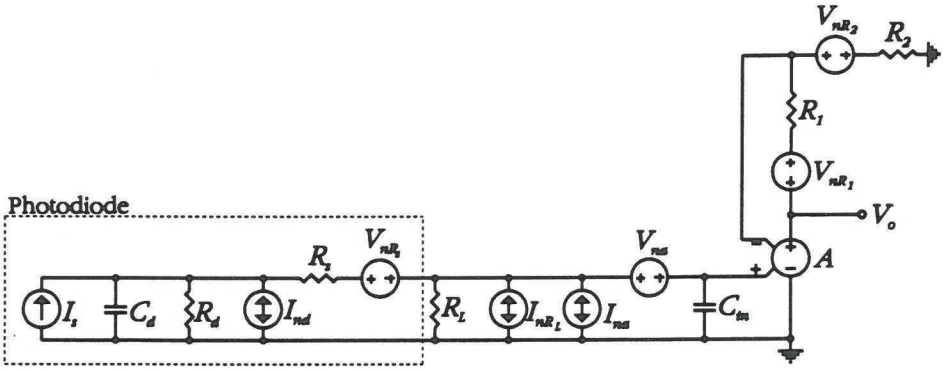


Figure 16: Noise equivalent circuit of the VA-circuit.

With straightforward methods similar to those used in the derivation of the PIF-circuit noise and with reference to Figure 16 the input referenced

noise of the VA-circuit becomes

$$I_{nVA}^2 = I_{nd}^2 + I_{na}^2 + I_{nR_L}^2 + \left(V_{nR_s}^2 + V_{nR_2}^2 + V_{na}^2\right)\left[\left(\frac{1}{R_d} + \frac{1}{R_L}\right)^2 + \omega^2 C_d^2\right] \quad (28)$$

The spectral density of the noise current in the load resistor is given by [1]

$$I_{nR_L}^2 = \frac{4kT}{R_L} \quad (29)$$

B.2 The transimpedance amplifier circuit

In the transimpedance amplifier circuit the input impedance of the amplifier constitute the load impedance for the photodiode. This circuit is named the TIA-circuit. The main benefit over the VA-circuit is a larger bandwidth due to the low input impedance of a transimpedance amplifier. The circuit diagram and the noise equivalent circuit diagram of the TIA-circuit are shown in Figures 17 and 18 respectively. The input referenced noise of the TIA-

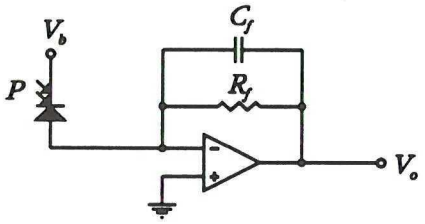


Figure 17: The TIA-circuit.

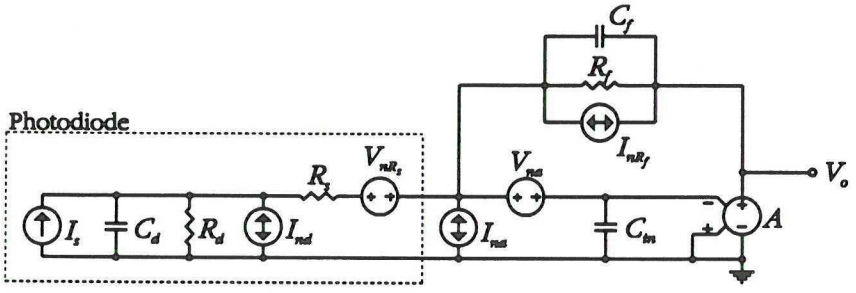


Figure 18: Noise equivalent circuit of the TIA-circuit.

circuit becomes

$$I_{nTIA}^2 = I_{nd}^2 + I_{na}^2 + I_{nR_f}^2 + \left(V_{nR_s}^2 + V_{na}^2\right)\left[\left(\frac{1}{R_d} + \frac{1}{R_f}\right)^2 + \omega^2 (C_d + C_f)^2\right] \tag{30}$$

The spectral density of the noise current in the feedback resistor is given by [1]

$$I_{nR_f}^2 = \frac{4kT}{R_f} \tag{31}$$

The expression (30) is similar to the corresponding expression for the VA-circuit. The differences are that their is no contribution from a resistor R_2, that R_L has been substituted with R_f and that the capacitor C_f has been added. The phase lead accomplished by the capacitor is needed for stability reasons to counteract the phase lag due to the photodiode capacitance C_d. The influence of C_f on the noise is small because normally $C_f \ll C_d$.

For minimum noise R_L and R_f should be as large as possible. The upper limit is given by the sum of the photodiode signal and leakage currents and the bias current at the amplifier input as mentioned in the introduction. Choosing $R_L = R_f$ makes the noise properties of the VA- and TIA-circuits essentially equal.

Part IV

Range dependence of the shape and amplitude of the received pulse for a laser anglemeter used for mobile robot navigation

Range dependence of the shape and amplitude of the received pulse for a laser anglemeter used for mobile robot navigation

Kalevi Hyyppä
Luleå University of Technology
S-951 87 LULEÅ, Sweden

April 1993

Abstract

A laser scanner, used as an anglemeter in a navigation system for mobile robots, has been developed. It measures heading angles to beacons made of vertical stripes of retroreflective tape. Expressions giving the received optical power and energy from a retroreflecting beacon, illuminated by the laser beam in the anglemeter, are derived. A Gaussian expression describing the bidirectional reflection distribution function (BRDF) of the beacon is proposed. A design rule for the optimum field of view of the receiving optics is proposed.

The shape and width of the received pulse are functions of the range R to the beacon. The range dependence of the received optical power is divided into five intervals with different characteristics. Measurements support the derived expressions.

At long range the dependence of the received power amplitude on range is R^{-3} and the pulse shape and width are independent of range. At long range the pulse shape is determined by the Gaussian intensity distribution in the laser beam. The pulse width is governed by the laser beam divergence and the scan velocity.

A design rule for a filter which maximizes the signal-to-noise ratio for the anglemeter is proposed. A function giving a conservative estimate of the signal-to-noise ratio is derived.

Keywords: BRDF, Gaussian beam, ladar, anglemeter, laser scanner, range, retroreflecting beacon, signal-to-noise ratio.

List of Symbols

A_b	area of beacon
dA_b	surface element on beacon
A_r	area of receiver aperture
dA_r	surface element on receiver aperture
A_{re}	area of receiver effective aperture
B_λ	optical bandwidth of interference filter
C_d	photodiode capacitance
E_a	irradiance on a surface by ambient light
E_b	irradiance on beacon from a source
dE_b	irradiance on beacon from a point on a source
$E_{\lambda,sun}$	spectral irradiance from the sun
G	gain of photodiode-amplifier
I_D	DC-current through photodiode
I_0	saturation current of photodiode
I_{na}^2	spectral density of amplifier noise current
I_{nd}^2	spectral density of shot noise in photodiode
I_{next}^2	spectral density of noise current due to ambient light
I_{nint}^2	spectral density of total internal noise current
I_{ntot}^2	spectral density of total noise current
L_a	radiance of a surface illuminated by ambient light
L_b	radiance of beacon
P_a	optical ambient power reaching photodiode
P_l	total power of laser beam
Q_f	signal energy at input of noise reduction filter
Q_{flr}	signal energy at input of noise reduction filter at long range
Q_r	optical energy reaching receiver photodiode
R	range i.e. distance between anglemeter and beacon
R_{fb}	feedback resistor in photodiode-amplifier
R_{ph}	responsivity of photodiode
R_{ref}	reference resistance
T	temperature of feedback resistor in photodiode-amplifier
T_p	width of a square pulse
V_{na}^2	spectral density of amplifier noise voltage
V_{nf}^2	spectral density of total noise voltage at amplifier output
W	width of beacon

e	electron charge
f	focal length of receiver lens
f_b	bidirectional reflection distribution function (BRDF) of beacon
f_w	BRDF of a surface illuminated by ambient light
f_{w0}	BRDF of a surface illuminated by ambient light for $\alpha_r = 0$
k	Boltzmanns constant
p_f	signal power at input of noise reduction filter
p_r	optical power reaching reciver photodiode
d^2p_r	optical power originating from dA_b and entering dA_r
p_{rlr}	optical power reaching reciver photodiode at long range
r	cylindrical r-coordinate
r_d	radius of receiver photodiode
r_l	local radius of laser beam (e^{-2})
r_{l0}	radius of laser beam at zero range
r_1	radius of inner border of receiver aperture
r_2	radius of outer border of receiver aperture
r_3	radius of circle given by the inverted acceptance cone
t	time
t_1	lower time limit in integral giving Q_r and in integral giving Q_f
t_2	upper time limit in integral giving Q_r and in integral giving Q_f
v_f	voltage at the input of the noise reduction filter
x	cartesian x-coordinate
x_0	cartesian x-coordinate for right edge of beacon
y	cartesian y-coordinate
α_b	angle between an incident ray and a reflected ray at beacon
α_{b0}	divergence of beacon (e^{-2})
α_l	far-field divergence of laser beam
α_r	incident angle to receiver
α_{rmax}	half-angle of receiver acceptance cone
γ	angle between z-axix and the vector $(x_0, 0, R)$
η_b	efficiency of beacon
η_r	efficeincy of receiver
λ	laser wavelength
τ_f	time constant of first order noise reduction filter
φ	cylindrical φ-coordinate
φ_1	lower integration limit given by left edge of beacon
φ_2	upper integration limit given by right edge of beacon
ψ	radiant flux density in laser beam
Ω_r	solid angle defined by receiver acceptance cone
ω	angular frequency
ω_b	angular velocity of beacon (negative)

1 Introduction

Angle measurement devices can be used for navigation of mobile robots. Several navigation systems using angle measurements have been proposed [1], [2], [3], [4], [5], [6], [7]. A laser scanner on the mobile robot measures heading angles to stationary beacons. The angles are used to determine the position and heading of the robot. The angles are often complemented with information from encoders, mounted on the robot wheels.

Another navigation method for mobile robots is to measure distance to objects in the environment of the robot e.g. with a laser. The measurement principle can be time of flight or triangulation [8], [9], [10], [11], [12]. Systems using the time of flight principle can also be classified as laser radars or ladars.

The maximum range for an angle or distance measuring system is limited by the signal-to-noise ratio of the system. For a laser radar the received pulse shape is known and independent of range. It is not so for an angle measuring laser scanner. Some general relations dealing with the shape of the received pulse and the signal-to-noise ratio for an angle measuring laser scanner are discussed in this paper. The term anglemeter will be used to denote the laser scanner built for this investigation.

In the next section a short description of the anglemeter is given. Then expressions for the received power and energy from a retroreflective beacon are derived. It is shown that beyond a certain range approximations can be made which makes it possible to give the expressions in closed form. Then some numerical results are presented with parameters taken from the anglemeter. In a section on noise the noise originating from ambient light is treated. Results of an analysis of the noise from internal sources are given and appropriate electrical filtering is suggested. Experimental data on the anglemeter are presented and compared with the theoretically derived behavior. Finally the achieved results are discussed and some conclusions are drawn.

2 Anglemeter description

The anglemeter consists of an optical part and an electronic part. A block diagram of the anglemeter with some details of the electronic part is shown in Figure 1. The optical part is shown in Figure 2 and a photograph of the anglemeter is shown in Figure 3. A laser beam is scanned in the horizontal plane by a mirror rotating in positive direction i.e. the laser beam enters the beacons at their right edges (seen from the anglemeter). When the beam hits a beacon some of the reflected radiation will reach a photodiode in the receiving part of the anglemeter. The signal current from the photodiode is converted to a voltage and amplified. After passing a filter for noise reduc-

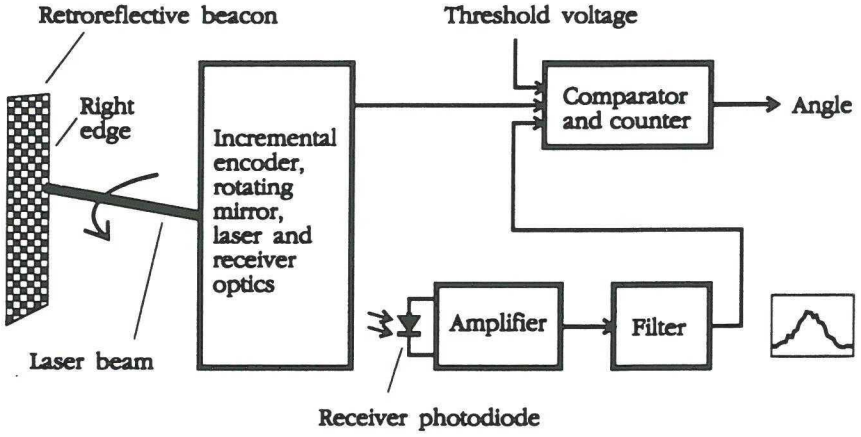


Figure 1: Block diagram of the anglemeter. Reflected radiation from the beacon hits a photodiode in the anglemeter receiver. An example of the received signal after the shaping filter is shown in the insert.

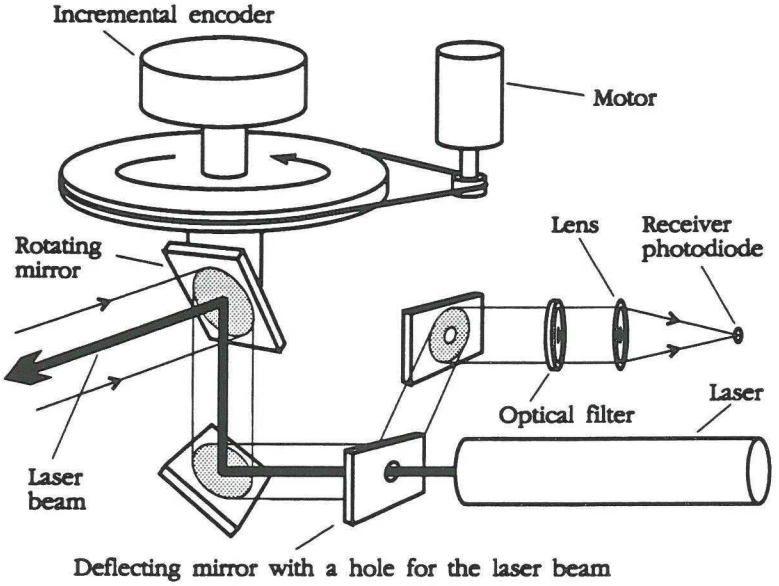


Figure 2: Optical part of the anglemeter. A small area in the center of the receiver aperture is wasted because it is occupied by the outgoing laser beam.

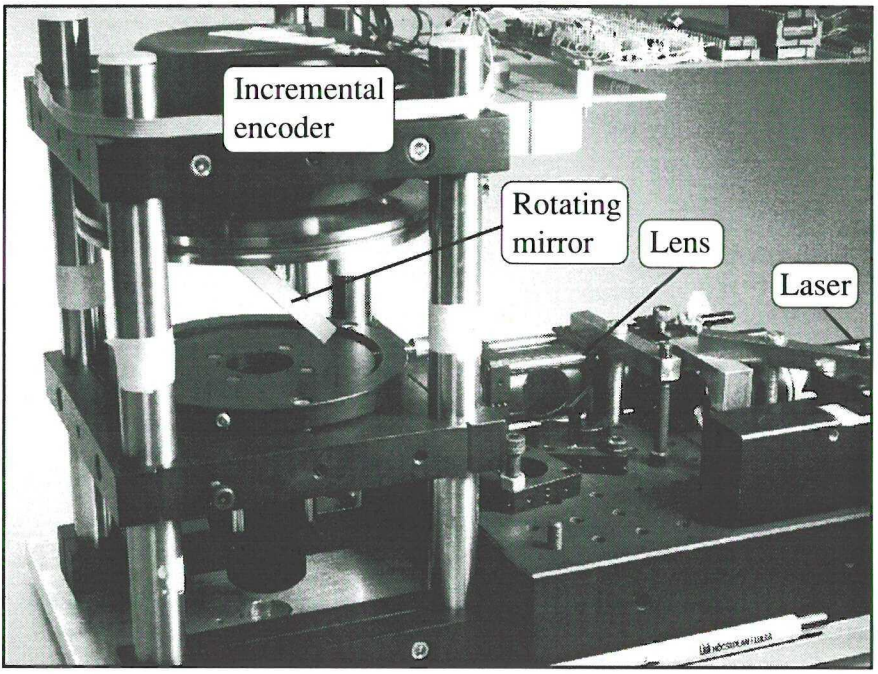


Figure 3: A photograph of the anglemeter. Note the pen in the bottom right corner for size reference.

tion, the signal voltage reaches a comparator where it is compared with a threshold voltage. When the comparator trips it is assumed that the laser beam has hit a beacon and the angle of the rotating mirror is recorded. The angle is measured with an incremental encoder attached to the axis of the rotating mirror and a counter.

The optical system is coaxial i.e. the optical axes of the laser beam and of the receiver optics are coincident. One of the deflecting mirrors has a hole on the optical axis. The outgoing illuminating beam passes the hole while part of the reflected radiation is deflected by the same mirror into the receiver optics. A small area in the center of the receiver aperture is wasted due to the hole.

3 Received power and energy

3.1 Definitions

Figure 4 shows a principal drawing of the anglemeter when the laser beam hits a beacon. A Cartesian coordinate system is attached to the rotating part of the anglemeter. The y-axis is aligned along the rotation axis of

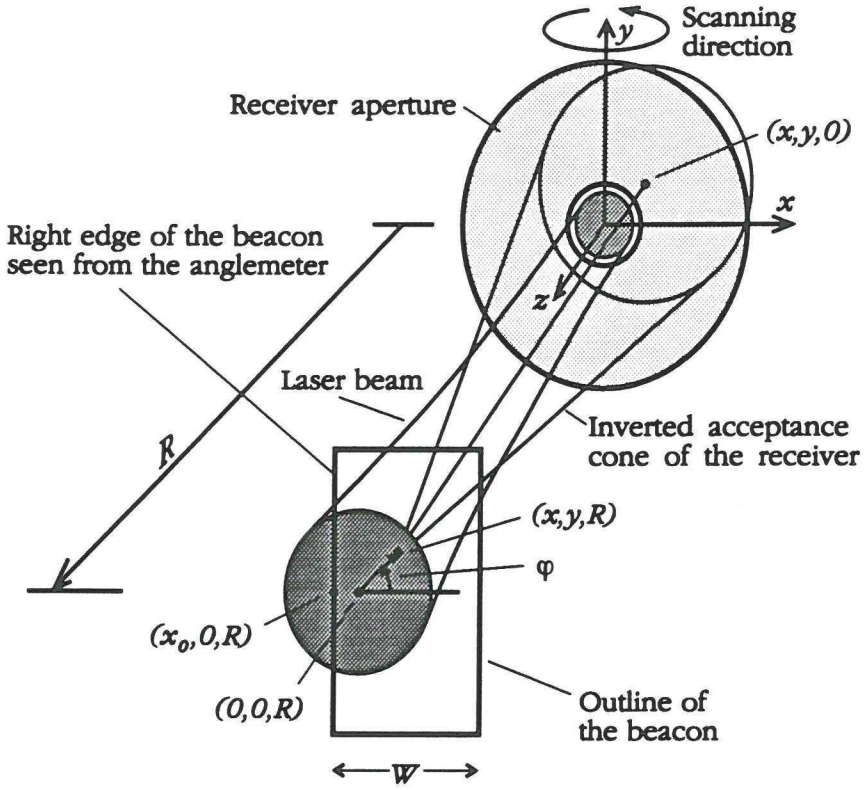


Figure 4: Definition of coordinate system and some symbols. All rays leaving the point (x,y,R) within the inverted receiver acceptance cone enters the reciever aperture with a sufficiantly small incident angle to reach the photodiode.

the anglemeter and the z-axis is aligned along the optical axis of the laser beam. The beam, which is assumed to have a Gaussian transversal intensity distribution, is indicated by the locus where the flux density has fallen to e^{-2} of the flux density on the optical axis. The retroreflective beacon is seen from its back-side and for clarity only the outline of the beacon is shown. The distance to the beacon is R. The beacon has a width W and a height which is much larger than the transmitted beam diameter.

The beacon is assumed to be orientated so that its normal and the laser beam optical axis are parallel when the optical axis hits the above defined right edge of the beacon. Because the beacon width and the laser beam diameter are much smaller than the distance between the anglemeter and the beacon, the beacon normal and the optical axis of the laser beam are approximately parallel whenever any part of the beam hits any part of the

beacon.

The retroreflective properties of the beacon are assumed to be independent of the incident angle to the beacon. This means that other beacon orientations can be treated by using the width projected onto the above assumed beacon plane instead of the real width.

The scanning direction of the laser beam is indicated in Figure 4. In the defined coordinate system the beacon has a negative angular velocity ω_b. The beam enters the beacon at the point $(x_0, 0, R)$. The x-coordinate of the entrance point is given by

$$x_0 = \gamma R \tag{1}$$

where γ denotes the instantaneous angle between the z-axis and the entrance point on the beacon. Here the small angle approximation $\sin\gamma \approx \gamma$ has been used because $x_0 \ll R$ when the beam is close to the beacon. By choosing $t = 0$ to be the time when the optical axis of the laser beam hits the right edge of the beacon the x-coordinate of the entrance point can be written

$$x_0 = R\omega_b t \tag{2}$$

since $\gamma = \omega_b t$.

3.2 The laser beam

The Gaussian beam is a consequence of the fact that the laser is supposed to be working in TEM_{00}-mode [16]. The transversal distribution of the radiant flux density in the laser beam is given by the expression

$$\psi = \frac{2P_l}{\pi r_l^2}\, e^{-2\left(\frac{r}{r_l}\right)^2} \tag{3}$$

where P_l denotes the total power of the beam, r_l denotes the local e^{-2}-radius of the beam, compare Figure 4, and r denotes the transversal distance from the optical axis [16]. For a Gaussian beam the radius increases with range according to

$$r_l = \sqrt{r_{l0}^2 + (R\alpha_l)^2} \tag{4}$$

where r_{l0} denotes the beam radius at zero range and α_l denotes the far-field divergence of the beam [16]. The expression for the far-field divergence is given by

$$\alpha_l = \frac{\lambda}{\pi r_{l0}} \tag{5}$$

where λ denotes the laser wavelength [16].

3.3 Radiance from an illuminated beacon

The reflective properties of the beacon is modeled by its bidirectional reflection distribution function (BRDF) [17]. In the general case the BRDF is dependent on four variables: elevation and azimuth angles defining the incident radiation direction, and elevation and azimuth angles defining the reflected radiation direction. Most surfaces have a BRDF dependent on only the two elevation angles and the difference between the two azimuth angles [17]. According to the assumption made above the BRDF of a retroreflective surface is considered to be independent of the incident direction. Furthermore experimental data, cited below, indicate that the reflective properties of retroreflective surfaces are symmetric around the incident direction. This implies that the BRDF of retroreflective surfaces is only dependent on the angle α between the incident direction and the reflection direction.

Nayar et al have given expressions for the BRDF of general surfaces [18], using two different reflection models — the Beckman-Spizzichino model and the Torrance-Sparrow model. Three different components of the BRDF are identified: the diffuse lobe, the specular lobe and the specular spike. The specular lobe has a Gaussian angular distribution around the specular direction. Experimental data on commercial retroreflective material are presented in [19] and [20]. In both references the data indicate that the BRDF is dominated by the Gaussian specular lobe. Therefore the following expression for the BRDF of a retroreflective beacon is proposed

$$f_b = \frac{2\eta_b}{\pi\alpha_{b0}^2} e^{-2\left(\frac{\alpha_b}{\alpha_{b0}}\right)^2} \tag{6}$$

where α_b denotes the angle between the incident direction and the reflection direction, α_{b0} denotes the divergence of the beacon and η_b denotes the efficiency of the beacon. The function f_b has a Gaussian dependence on the angle α_b and it is properly normalized to make the total reflected power equal to the incident power if the efficiency is equal to 1.

The divergence of retroreflective materials, useful in this context, is much larger than the divergence of the laser beam. Typical values are $\alpha_{b0} > 10$ mrad [19], [20] and $\alpha_l < 1$ mrad [21]. Due to this condition the incident angle to the beacon can be approximated to be zero for all parts of the beam and therefore the angle α_b above is equal to the reflection angle from the beacon. Furthermore, since the beacon and the receiver lens are supposed to be parallel

$$\alpha_r = \alpha_b \tag{7}$$

where α_r denotes the incident angle to the receiver for a ray with reflection angle α_b from the beacon.

Knowing the irradiance dE_b from a point on the source to a point on the beacon and the BRDF of the beacon, the radiance of the point on the beacon

can be expressed as

$$L_b = \int_{source} f_b \, dE_b \tag{8}$$

The TEM_{00}-mode of the beam makes all incident radiation to a point on the beacon have the same incident angle because the radiation originates from a single point on the laser aperture. Thus the expression for the radiance simplifies to the point source case

$$L_b = f_b E_b \tag{9}$$

Due to the chosen orientation of the beacon the irradiance is equal to the radiant flux density of the beam, i.e.

$$E_b = \psi \tag{10}$$

From (3), (6), (7), (9) and (10) the expression for the radiance of a point on the beacon is derived as

$$L_b = \frac{4P_l\eta_b}{(\pi\alpha_{b0}r_l)^2} \; e^{-2\left(\frac{r}{r_l}\right)^2} e^{-2\left(\frac{\alpha_r}{\alpha_{b0}}\right)^2} \tag{11}$$

3.4 Optical power reaching the photodiode

Let dA_b denote a surface element on the beacon and let dA_r denote a surface element on the receiver aperture. Let furthermore the radiation reflected from dA_b and entering dA_r have an incident angle α_r at the receiver aperture. The distance between the two surface elements then becomes $R/\cos\alpha_r$. The surface element on the receiver aperture constitutes a solid angel $dA_r \cos\alpha_r$ at the beacon surface element. The radiant intensity at dA_b is $L_b \, dA_b \cos\alpha_r$ in the direction towards dA_r. The optical radiation power which originates from dA_b and enters the receiver aperture through dA_r is then

$$d^2p_r = L_b \, dA_b \cos^4\alpha_r \frac{dA_r}{R^2} \tag{12}$$

The receiver aperture is limited by the central hole with radius r_1 — for the outgoing laser beam — and the maximum lens opening with radius r_2. The receiver aperture is shown schematically in Figure 5. A photodiode with an active circular area with radius r_d is placed at the focal point of the receiver lens. The lens focal length is f. The radius r_d and the focal length f determine a maximum incident angle $\alpha_{r\,max}$ to the receiver. Radiation entering the receiver with an incident angle greater than $\alpha_{r\,max}$ will not reach the photodiode. This maximum angle is

$$\alpha_{r\,max} = \arctan\frac{r_d}{f} \tag{13}$$

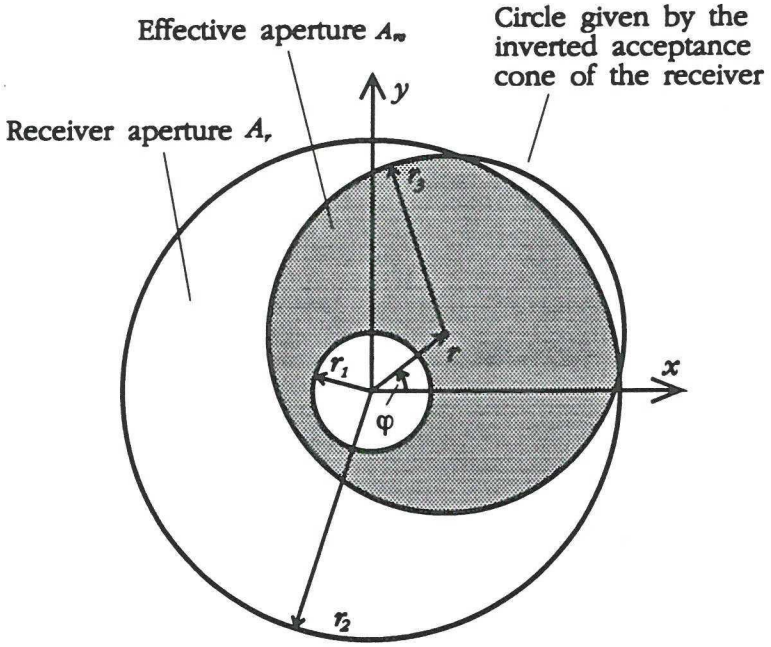


Figure 5: The receiver aperture. The shaded area marks the effective aperture for a point on the beacon. The point is at distance r from the optical axis.

The receiver acceptance cone is defined by α_{rm}. In Figure 4 an inverted receiver acceptance cone is indicated for a point (x, y, R) on the beacon. All rays leaving the point within the inverted receiver acceptance cone have an incident angle to receiver which is less than $\alpha_{r\,\mathrm{max}}$. In Figure 5 the circular area which is cut out at the receiver aperture by the inverted receiver acceptance cone is shown. The radius of the circular area is

$$r_3 = R \tan \alpha_{r\,\mathrm{max}} \tag{14}$$

or, using (13),

$$r_3 = R\frac{r_d}{f} \tag{15}$$

The shaded area in Figure 5 is defined to be the effective aperture A_{re} of the receiver for the point on the beacon which is at the distance $r = \sqrt{x^2 + y^2}$ from the optical axis. Of the radiation, which is reflected from the point (x, y, R) on the beacon, only that part which enters the receiver aperture within the effective aperture reaches the photodiode. With the beacon beyond a certain range, the effective aperture is equal to the receiver aperture

because the whole receiver aperture lies within the circle given by the acceptance cone. At shorter range the receiver effective aperture diminishes. At very short ranges the effective aperture lies within the central hole which results in an effective aperture equal to zero. This gives the unexpected property that the received optical power vanishes completely at very short ranges. Expressions giving A_{re} as a function of r_1, r_2, r_3 and r are derived in Appendix A.

Since both the beacon width and the receiver aperture diameter are supposed to be much smaller than the distance between the beacon and the receiver, the small angle approximation $\cos\alpha_r \approx 1$ can be used in (12). The power p_r reaching the detector is then given by

$$p_r = \int_{A_b}\int_{A_{re}} \eta_r L_b \frac{dA_r}{R^2}\, dA_b \tag{16}$$

where A_b denotes the area of the beacon and η_r denotes the efficiency of the receiver. Using (11) and rearranging, (16) becomes

$$p_r = \frac{4P_l\eta_b\eta_r}{(\pi\alpha_{b0}r_l R)^2} \int_{A_b} e^{-2\left(\frac{r}{r_l}\right)^2} \int_{A_{re}} e^{-2\left(\frac{\alpha_r}{\alpha_{b0}}\right)^2} dA_r\, dA_b \tag{17}$$

With cylindrical coordinates, where $dA_b = r\, d\varphi\, dr$ as indicated in Figure 6, (17) becomes

$$p_r = \frac{4P_l\eta_b\eta_r}{(\pi\alpha_{b0}r_l R)^2} \int_0^\infty 2\int_{\varphi_1}^{\varphi_2} e^{-2\left(\frac{r}{r_l}\right)^2} r\, d\varphi\, dr \int_{A_{re}} e^{-2\left(\frac{\alpha_r}{\alpha_{b0}}\right)^2} dA_r \tag{18}$$

where the limits φ_1 and φ_2 are functions of r , x_0 and W. The upper limit of the first integral have been set to infinity for simplicity because the beacon height is much larger than the laser beam radius r_l. The factor 2 after the first integral sign is introduced since, due to vertical symmetry, the integral over φ is taken only over the upper half of the beacon area. After integration over φ the expression for the received optical power is

$$p_r = \frac{8P_l\eta_b\eta_r}{(\pi\alpha_{b0}r_l R)^2} \int_0^\infty e^{-2\left(\frac{r}{r_l}\right)^2} r\,(\varphi_2 - \varphi_1)\, dr \int_{A_{re}} e^{-2\left(\frac{\alpha_r}{\alpha_{b0}}\right)^2} dA_r \tag{19}$$

By inspection of Figure 6 the angels are found to be

$$\varphi_1 = \begin{cases} 0 & \text{if} \quad x_0 + W > r \\ \arccos\left(\frac{x_0+W}{r}\right) & \text{if} \quad -r \le x_0 + W \le r \\ \pi & \text{if} \quad x_0 + W < -r \end{cases}$$

$$\varphi_2 = \begin{cases} 0 & \text{if} \quad x_0 > r \\ \arccos\left(\frac{x_0}{r}\right) & \text{if} \quad -r \le x_0 \le r \\ \pi & \text{if} \quad x_0 < -r \end{cases} \tag{20}$$

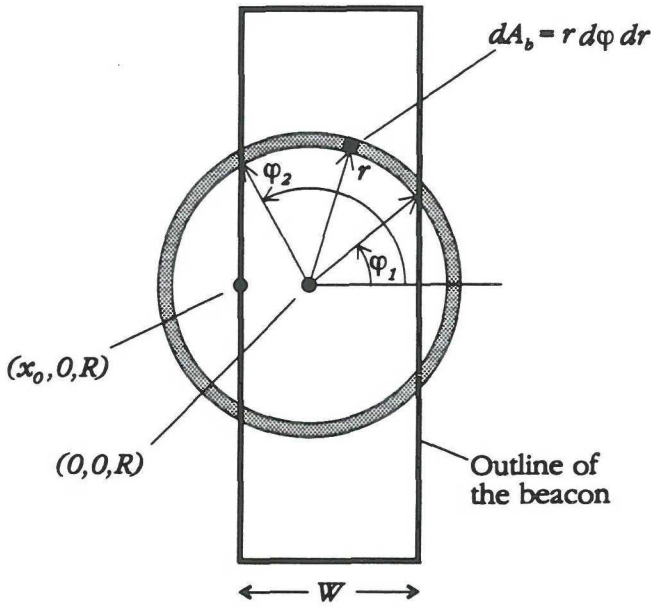


Figure 6: Area element on the beacon defining the integration limits φ_1 and φ_2.

The receiver acceptance cone is given by $\alpha_{r\,\max}$. It will be shown that it should be kept small to minimize noise from ambient light. The acceptance cone should however be chosen to be large enough to utilize most of the power of the laser beam. This is particularly important in the far-field region where the received power is lowest. In the far-field region (4) reduces to $r_l \approx R\alpha_l$. Because of the rapid fall of the flux density in the beam beyond the radius r_l the following design rule is proposed: $R\alpha_{r\,\max} \approx 2r_l$, i.e.

$$\alpha_{r\,\max} \approx 2\alpha_l \tag{21}$$

As mentioned above the laser far-field divergence is typically less than 1 mrad. The design rule then makes $\alpha_{r\,\max}$ less than 2 mrad which together with α_b typically greater than 10 mrad enables a further simplification of (19). With these values the radiance from the beacon is approximately equal to the on-axis radiance for incident angles less than $\alpha_{r\,\max}$ i.e. for the whole effective aperture A_{re} of the receiver. The expression for the received optical power can then be simplified to

$$p_r = \frac{8P_l\eta_b\eta_r}{(\pi\alpha_{b0}r_lR)^2}\int_0^\infty A_{re}\,e^{-2\left(\frac{r}{r_l}\right)^2} r\,(\varphi_2-\varphi_1)\,dr \tag{22}$$

The received optical power is time dependent. The dependence is hidden in the angle difference $(\varphi_2 - \varphi_1)$ which is time dependent through x_0.

3.5 Optical energy reaching the photodiode

The optical energy reaching the photodiode from a beacon is found by integrating the expression for p_r over time

$$Q_r = \frac{8P_l\eta_b\eta_r}{(\pi\alpha_{b0}r_l R)^2}\int_{-\infty}^{+\infty}\int_0^{\infty} A_{re}\, e^{-2\left(\frac{r}{r_l}\right)^2} r\,(\varphi_2 - \varphi_1)\; dr\; dt \tag{23}$$

A receiver designed according to the proposed design rule makes A_{re} nonzero only for $r < 2r_l$. The integration intervals in (23) can then be narrowed. The upper integration limit of the inner integral becomes $2r_l$. According to Figure 6, the condition $r < 2r_l$ can be written

$$\begin{array}{rl} x_0 & < 2r_l \\ x_0 + W & > -2r_l \end{array} \tag{24}$$

The expression (2) for x_0 and the fact that the angular velocity of the beacon is negative yield

$$\frac{-2r_l}{R\,|\omega_b|} < t < \frac{2r_l + W}{R\,|\omega_b|} \tag{25}$$

which allows the lower time limit to be raised from $-\infty$ to

$$t_1 = \frac{-2r_l}{R\,|\omega_b|} \tag{26}$$

and the upper time limit to be lowered from ∞ to

$$t_2 = \frac{2r_l + W}{R\,|\omega_b|} \tag{27}$$

3.6 Electrical power and energy at the filter input

The photodiode converts the optical power p_r to a current which is converted to a voltage and amplified by the amplifier in the receiver. With reference to Figure 1 the voltage at the filter input becomes

$$v_f = GR_{ph}p_r \tag{28}$$

where R_{ph} denotes the responsivity of the photodiode and G denotes the gain of the amplifier. With a reference resistance R_{ref} the signal power at the filter input becomes

$$p_f = \frac{(GR_{ph}p_r)^2}{R_{ref}} \tag{29}$$

The signal energy at the filter input is found by integrating the expression for p_f between t_1 and t_2

$$Q_f = \frac{(GR_{ph})^2}{R_{ref}} \int_{t_1}^{t_2} p_r^2 \, dt \tag{30}$$

Note that the signal power at the filter input is not proportional to the optical power but to the optical power squared, and that the signal energy is not proportional to neither the optical energy nor the optical energy squared.

3.7 Asymptotic expressions at long range

Long range is defined as the range where the beam diameter is larger than the beacon width, i.e.

$$r_l > \frac{W}{2} \tag{31}$$

The expression for the beam radius (4) combined with the fact that the laser diameter at zero range is much smaller than the beacon width can be used to transform the long range condition to

$$R > \frac{W}{2\alpha_l} \tag{32}$$

With the receiver acceptance cone designed according to the design rule ($\alpha_{r\,\mathrm{max}} \approx 2\alpha_l$) the radius of the circle which determines the effective aperture becomes

$$r_3 \approx 2r_l \tag{33}$$

The condition (31) then gives

$$r_3 > W \tag{34}$$

The diameter of the receiver aperture is approximately equal to the beacon width i.e.

$$r_2 \approx \frac{W}{2} \tag{35}$$

which means that

$$r_3 > 2r_2 \tag{36}$$

holds in the long range region.

Only points — on the beacon — within the distance r_l from the optical axis contribute significantly to the received power due to the Gaussian distribution in the laser beam. The combination of (33) and (36) gives for the distances of those points

$$r < r_3 - r_2 \tag{37}$$

The effective aperture is then constant and equal to its maximum value A_r for the points on the beacon which lie within the distance r_l from the

optical axis. The conclusion is then that the effective aperture is constant for all points on the beacon which contribute significantly to the received power. This conclusion and the result of the discussion in section 3.4 enable simplifications of (17) at long range. The received power at long range then becomes

$$p_{rlr} = \frac{4P_l\eta_b\eta_r A_r}{(\pi\alpha_{b0}r_l R)^2} \int_{A_b} e^{-2\left(\frac{r}{r_l}\right)^2} dA_b \tag{38}$$

By using Cartesian coordinates this can be transformed to

$$p_{rlr} = \frac{4P_l\eta_b\eta_r A_r}{(\pi\alpha_{b0}r_l R)^2} \int_{x_0}^{x_0+W} e^{-2\left(\frac{x}{r_l}\right)^2} \int_{-\infty}^{+\infty} e^{-2\left(\frac{y}{r_l}\right)^2} dy\, dx \tag{39}$$

which can be further simplified with the aid of the error function to

$$p_{rlr} = \frac{4P_l\eta_b\eta_r A_r}{(\pi\alpha_{b0}r_l R)^2} \cdot r_l\sqrt{\frac{\pi}{2}} \int_{x_0}^{x_0+W} e^{-2\left(\frac{x}{r_l}\right)^2} dx \tag{40}$$

Due to condition (31) the integrand in (40) is approximately constant over the integration interval and can therefore be approximated with its value in the middle of the interval. After some manipulation the expression for the received optical power at long range becomes

$$p_{rlr} = \frac{P_l\eta_b\eta_r A_r W}{\pi\alpha_l\alpha_{b0}^2 R^3} \cdot \sqrt{\frac{8}{\pi}}\, e^{-2\left(\frac{x_0+\frac{W}{2}}{\alpha_l R}\right)^2} \tag{41}$$

The time dependence becomes explicit if (2) is used to express x_0

$$p_{rlr} = \frac{P_l\eta_b\eta_r A_r W}{\pi\alpha_l\alpha_{b0}^2 R^3} \cdot \sqrt{\frac{8}{\pi}}\, e^{-2\left(\frac{\omega_b t+\frac{W}{2R}}{\alpha_l}\right)^2} \tag{42}$$

At long range the shape of the received pulse is thus Gaussian and the amplitude of the pulse is proportional to R^{-3}. Defining the pulse width as the time between the e^{-2}-points the pulse width at long range becomes $2\alpha_l/|\omega_b|$.

Multiplying the expression for the received optical power above with the responsivity of the photodiode and the gain of the amplifier gives an expression for the voltage at the filter input. The signal pulse energy at the filter input is found by squaring the voltage expression, dividing the result with R_{ref} and finally integrating from $t = -\infty$ to $t = +\infty$. The integral can be evaluated with the aid of the error function and the result becomes

$$Q_{flr} = \frac{G^2 R_{ph}^2 P_l^2 \eta_b^2 \eta_r^2 A_r^2 W^2}{R_{ref}\,\pi^2\,|\omega_b|\,\alpha_l\alpha_{b0}^4 R^6} \cdot \frac{4}{\sqrt{\pi}} \tag{43}$$

Laser beam			Beacon			Receiver		
			η_b	0.70		η_r	0.20	
α_l	0.6	mrad	α_{b0}	23.0	mrad	α_{rm}	2.8	mrad
r_{l0}	0.35	mm	W	16.0	mm	r_1	2.5	mm
P_l	1.0	mW	ω_b	6.28	rad/s	r_2	8.0	mm
						R_{ph}	0.25	A/W
						G	10^7	V/A

Table 1: Anglemeter parameters.

4 Numerical results

The expressions for v_f and Q_f have been numerically evaluated for a number of different range values R. The parameter values used, valid for the anglemeter, are shown in Table 1. A representative set of plots of the voltage at the filter input for different range values is shown in Figure 7. Note the

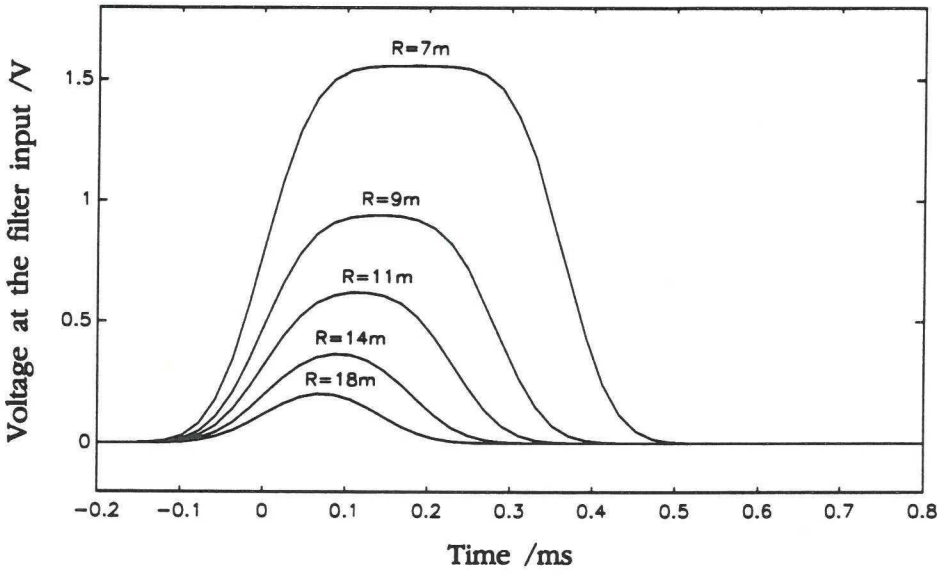


Figure 7: Numerically evaluated voltages at the filter input. The pulse shape goes gradually from the accumulated Gaussian distribution shape to the Gaussian shape when the laser beam expands with range.

change in pulse shape with range. At $R = 7$ m the beam diameter is smaller than the beacon width and therefore the rising edge of the pulse has the shape of the accumulated Gaussian distribution due to the Gaussian beam.

The rising edge has reached half the maximum pulse amplitude at $t = 0$ because then half of the beam is reflected by the beacon.

At $R = 18$ m the beam diameter is larger than the beacon width and the pulse shape is therefore approximately equal to the Gaussian beam intensity distribution.

The influence of the diminishing receiver effective aperture at close range is clearly visible in Figure 8 which shows two perspective views of the voltage at the filter input as a function of both time and range. The gradual transition of the pulse shape from the accumulated Gaussian distribution shape to the Gaussian shape with increasing range is also evident.

The peak amplitude of the signal pulse at the filter input is shown as a function of range in Figure 9, and the pulse energy at the filter input is shown as a function of range in Figure 10. The dependence of both the peak amplitude and the pulse energy on range is governed by the range dependence of the received optical power. With reference to Figures 9 and 10 the range can be divided into five regions.

1. $R < R_1$ — At very close range no optical radiation from the beacon hits the photodiode. The beam is so narrow that all reflected radiation which is within the acceptance cone of the receiver is wasted in the central hole of the receiver aperture.

2. $R_1 \leq R < R_2$ — The laser beam flux which is in the tail of the Gaussian distribution escapes the hole after reflection in the beacon and gives a steeply increasing contribution to the received power.

3. $R_2 \leq R < R_3$ — Most of the laser beam is utilized and the receiver effective aperture is still increasing, making the received power increase gradually.

4. $R_3 \leq R < R_4$ — All reflected radiation entering the receiver aperture is within the acceptance cone and the beam diameter at the beacon is smaller than the beacon width. The range dependence of the received optical power is R^{-2} since the irradiance at the receiver diminishes with range. The range dependence of the peak amplitude at the filter is then also R^{-2} and the range dependence of the pulse energy becomes R^{-5} due to the squaring of the voltage v_f and the R^{-1} dependence of the pulse width.

5. $R \geq R_4$ — The beam diameter is larger than the beacon width meaning that only part of the laser power is utilized. The signal pulse width is constant and determined by the laser beam divergence and the scan velocity. The range dependence of the received optical power is R^{-3} and therefore the range dependencies of the peak amplitude and the pulse energy at the filter become R^{-3} and R^{-6} respectively.

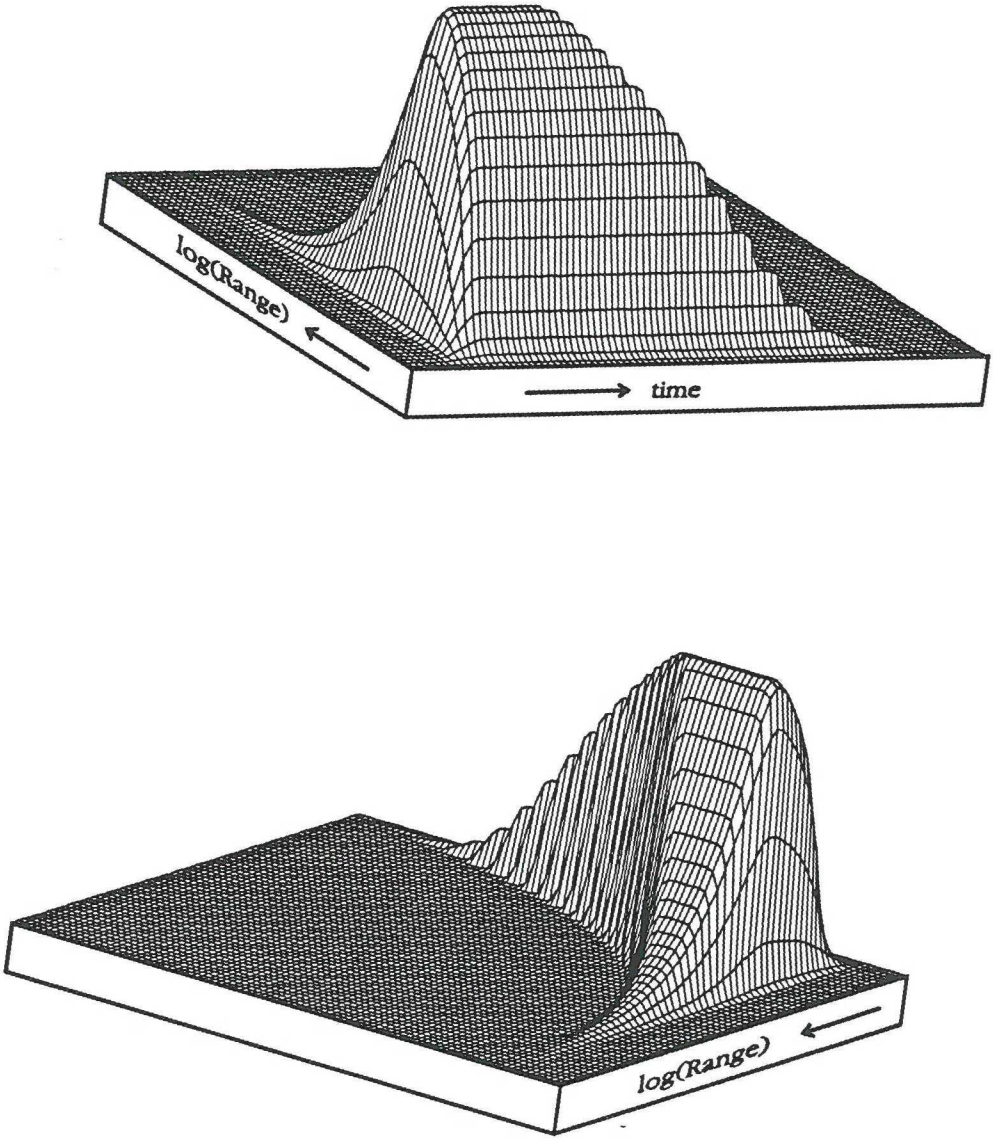


Figure 8: Two views of the voltage at the filter input as a funtion of time and range. The time interval is -0.5 ms to 4.8 ms and the range interval is 0.6 m to 60 m. The range axis is logarithmic.

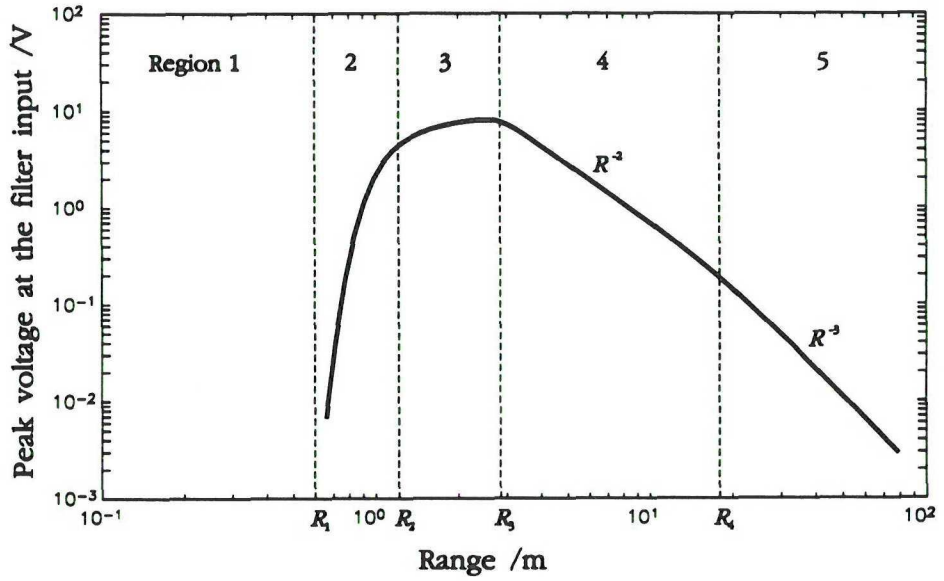


Figure 9: Numerically evaluated peak voltage at the shaping filter input. The different regions are described in the text.

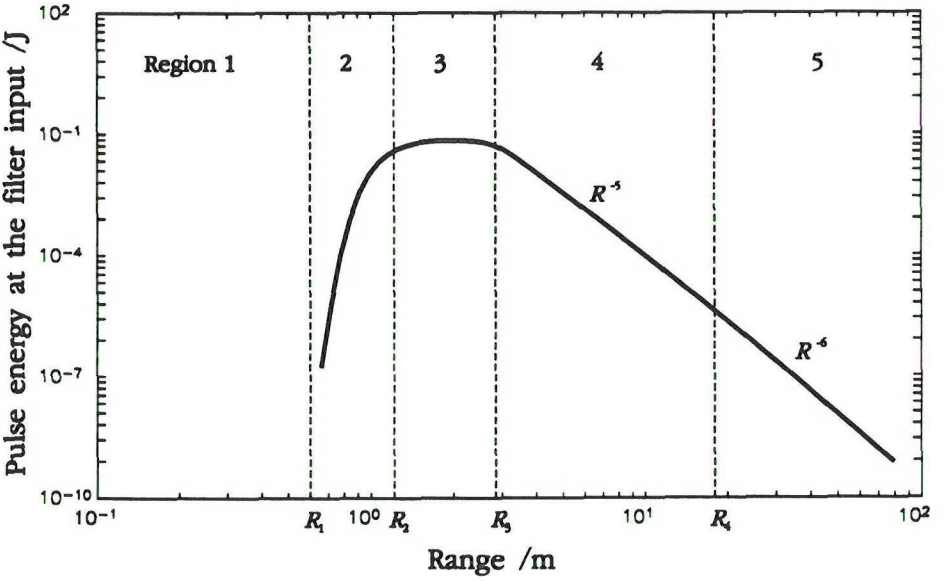


Figure 10: Numerically evaluated pulse energy at the shaping filter input. The different regions are described in the text.

5 Noise

A constant ambient illumination generates a DC current in the photodiode which gives rise to shot noise. The irradiance, from the dominating ambient light source, at a surface — e.g. a wall which is in the field of view of the anglemeter — is denoted by E_a. The surface is assumed to have the bidirectional reflection distribution function f_w and the source is approximated by a point source. The radiance L_a of the surface then becomes

$$L_a = f_w E_a \tag{44}$$

The radiance theorem states that the radiance of a ray is an invariant in free space, therefore (44) also gives the radiance at the receiver entrance aperture. Taking the worst case orientation of the surface, the surface normal parallel with the z-axis, the ambient radiation P_a reaching the photodiode becomes

$$P_a = \int_{A_r} \int_{\Omega_r} \eta_r f_w E_a \cos\alpha_r \, d\Omega_r \, dA_r \tag{45}$$

where Ω_r denotes the solid angle defined by the acceptance cone of the receiver. The acceptance cone is centered around $\alpha_r = 0$ and $\Omega_r = \alpha_{r\,\mathrm{max}}^2$ because $\alpha_{r\,\mathrm{max}}$ is small. Furthermore α_r is so small that the approximations $\cos\alpha_r = 1$ and $f_w = f_w(\alpha_r = 0) = f_{w0}$ are justified. The integrand is thus constant over the whole receiver aperture and the received optical power from the illuminated surface becomes

$$P_a = \eta_r f_{w0} E_a \alpha_{r\,\mathrm{max}}^2 A_r \tag{46}$$

If the surface is modeled to be perfectly Lambertian then $f_w = f_{w0} = 1/\pi$. The spectral density of the shot noise in a diode is given by [22]

$$I_{nd}^2 = 2e(|I_D| + 2I_0) \tag{47}$$

where I_D denotes the DC current through the diode, I_0 denotes the saturation current of the diode and e denotes the electron charge. The photo current in the photodiode becomes $R_{ph}P_a$ where R_{ph} denotes the responsivity of the photodiode. The photo current is much larger than the saturation current of the diode during normal ambient light conditions. The spectral density of the noise current generated in the photodiode by ambient light then becomes

$$I_{next}^2 = \frac{2}{\pi} e R_{ph} \eta_r E_a \alpha_{r\,\mathrm{max}}^2 A_r \tag{48}$$

In appendix B the receiver total noise is represented with a noise current generator in parallel with the photodiode. The spectral density of the total noise current is derived to

$$I_{ntot}^2 \approx \frac{2}{\pi} e R_{ph} \eta_r E_a \alpha_{r\,\mathrm{max}}^2 A_r + \frac{4kT}{R_{fb}} + I_{na}^2 \tag{49}$$

As stated earlier this shows that the acceptance cone of the receiver should be kept small to suppress the externally generated noise. The following approximate expression for the anglemeter signal-to-noise ratio at long range is also derived in appendix B.

$$S/N \approx \frac{4R_{ph}^2 P_l^2 \eta_b^2 \eta_r^2 A_r^2 W^2}{\pi^{2.5} |\omega_b| \alpha_l \alpha_{b0}^4 R^6 \left[2\pi^{-1} e R_{ph} \eta_r E_a \alpha_{r\,\max}^2 A_r + 4kTR_{fb}^{-1} + I_{na}^2\right]} \quad (50)$$

6 Experimental results

The voltage at the filter input in the anglemeter has been recorded at 40 different range values in the interval 0.2–70 m. The voltage was recorded with a digital oscilloscope. A reference pulse from the incremental encoder in the anglemeter was used as a trigger for the oscilloscope.

The retroreflective material used in the prototype beacon was "Scotchlite High Intensity" from 3M. It has an hexagonal structure with a cell size of about 3 mm. This structure gives a local variation of the reflective properties of the beacon which is clearly visible in the data at ranges where the beam diameter is small compared to the cell size. An example of that can be seen in Figure 11 which shows the measured voltage at the shaping filter input at $R = 1.4$ m. The same voltage at the end of the range interval i.e. at $R = 70$ m is shown in Figure 12. In Figure 13 some intermediate values are shown.

The measured voltages have been filtered to suppress noise and internal interference in the anglemeter at long range and the modulation due to the beacon structure at short range. The filtered results were used to produce Figures 14 and 15. They show the measured peak voltage and pulse energy from the experimental data. The numerically calculated results are shown for comparison.

7 Discussion

In the section on received power and energy the structure of the beacon material was not modeled. The theoretical pulse shape therefore lacks the modulation visible in the experimental data at short range. For longer range values Figures 7 and 13 indicate that the theory predicts the voltage at the shaping filter input well. The slight time displacements of the curves in Figure 13 are due to experimental difficulties in placing the beacon on the nominal line of sight when the range to the beacon is shifted.

The experimental peak voltage and pulse energy show a large discrepancy from theory at short range where the theory predicts much lower values than the measured. This is probably caused mainly by internal scattering in the

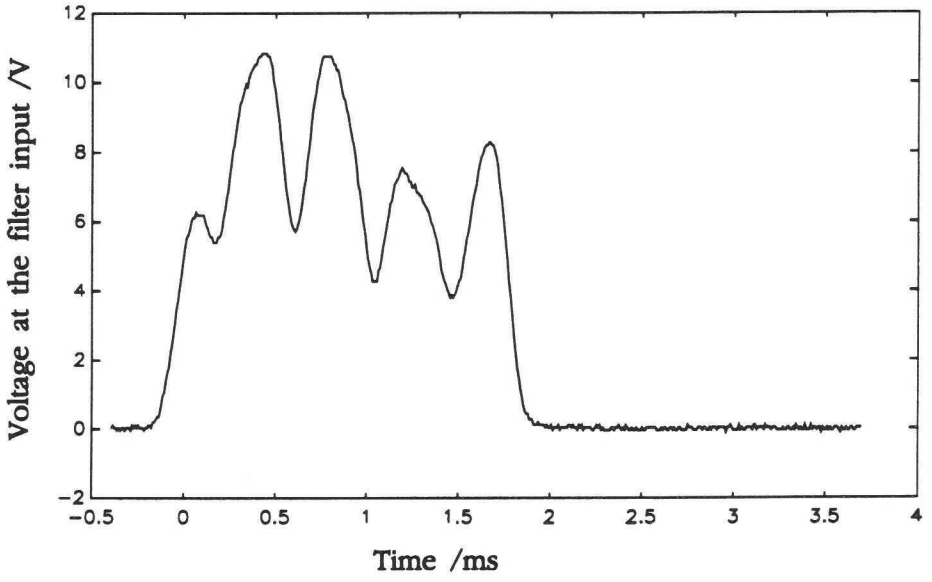


Figure 11: Measured voltage at the shaping filter input at $R = 1.4$ m. The structure of the retroreflector is clearly resolved by the narrow laser beam.

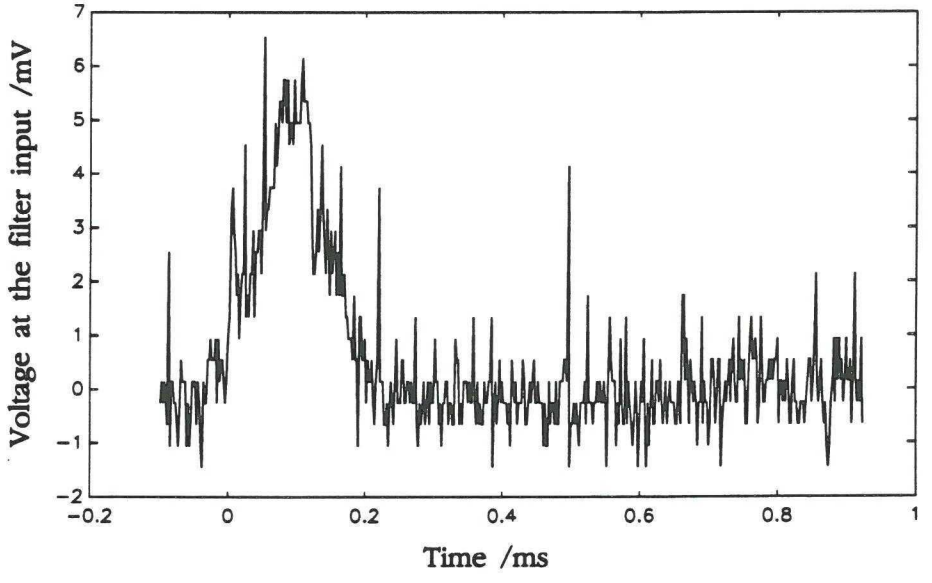


Figure 12: Measured voltage at the shaping filter input at $R = 70$ m. The beam cross section is much larger than the beacon width at this range.

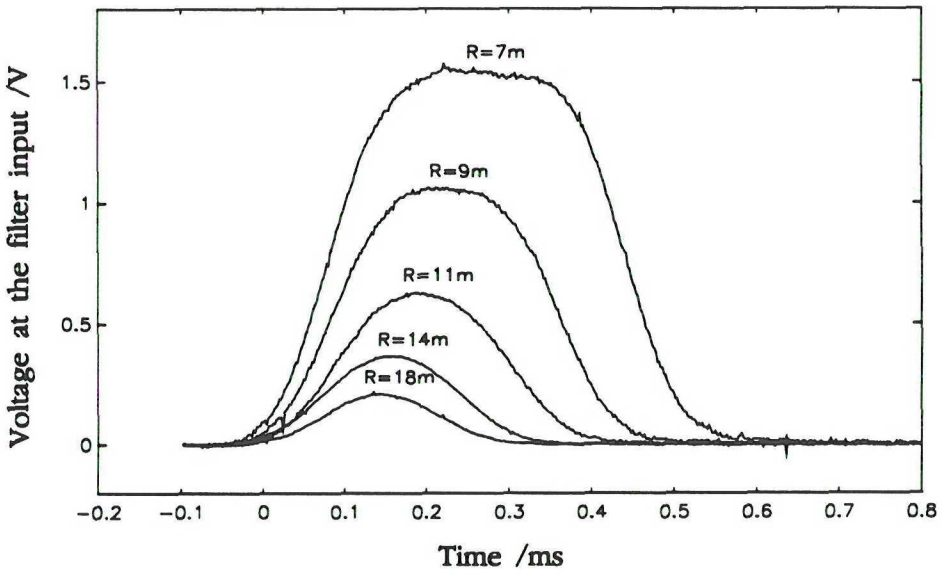


Figure 13: Measured voltage at the shaping filter input at $R = 7$, 9, 11, 14 and 18 m.

receiver. The used model — with no light reaching the active area of the photodiode if the light is outside the acceptance cone — is too conservative. The photodiode chip is encapsulated in a molded epoxy casing. The surfaces of the casing are not of optical quality and will therefore give rise to scattering. At short range the receiver aperture is flooded with light which creates regions with high radiant flux density at the front and back surfaces of the photodiode casing. Some of the light from those regions will find its way to the photodiode by scattering. Another cause of the discrepancy is missalignment between the laser beam and the central hole in the receiver aperture. It was experimentally verified that the received optical power at short range is very sensitive to displacements between the laser beam and the central hole.

It should be pointed out that when the anglemeter is used as part of a navigation system for a mobile robot this discrepancy between the presented theory and the actual behavior is of no significance.

The predicted range dependence of the peak voltage and the pulse energy in range regions 4 and 5 conform well with experimental data. These regions are of primary interest when the anglemeter is used as a part of a navigation system for mobile robots.

The noise level of the anglemeter is of greatest interest in the long range region 5 where the signal level is lowest. Fortunately the width and shape

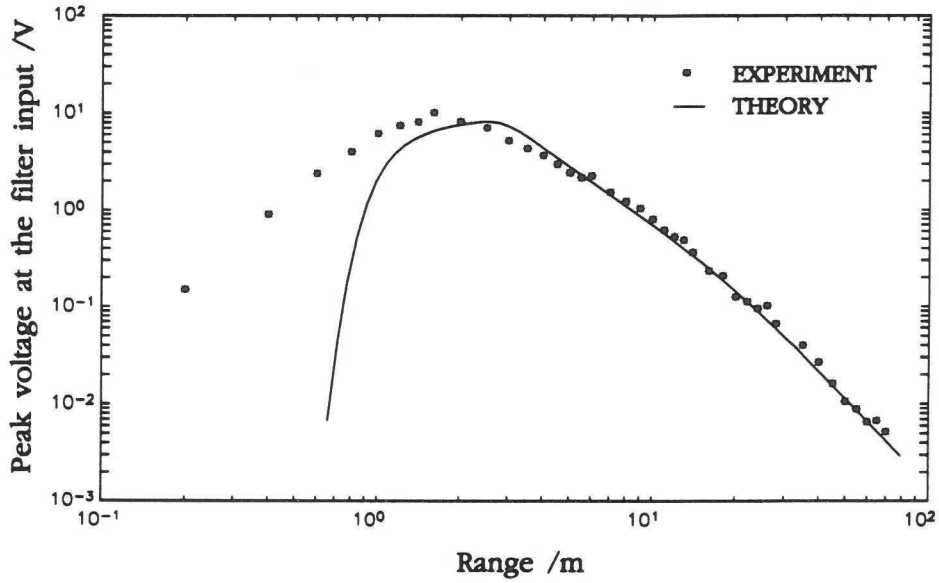


Figure 14: Experimental and theoretical peak voltage at the filter input. The descrepancy at short range is discussed in the text.

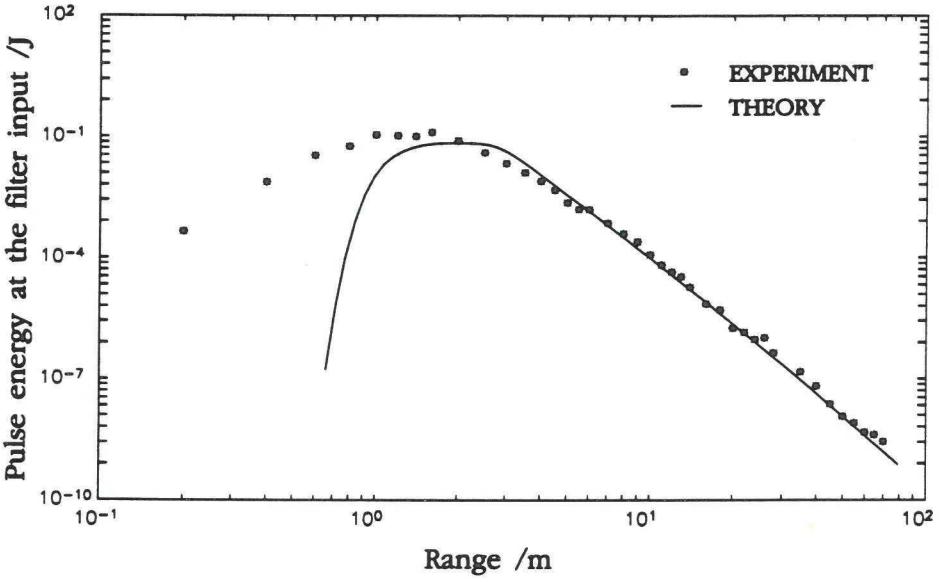


Figure 15: Experimental and theoretical pulse energy at the filter input. The descrepancy at short range is discussed in the text.

of the signal pulse becomes independent of range in the long range region. This enables the design of an approximation to the matched filter with a performance close to the matched filter in the long range region.

The analysis in appendix B shows that the thermal noise in the feedback resistor dominates the noise in the anglemeter. The actual noise was not measured due to a prohibitively high interference level in the measurement setup. The interference level in a later commercial version of the anglemeter is below the noise level.

The influence of all important design parameters on the signal-to-noise ratio at long distance is visible in (50). Note the strong influence of the beacon divergence α_{b0}. The width of the received pulse is monotonically decreasing with range and so is the amplitude in regions 4 and 5. This means that (50) gives a conservative estimate of the signal-to-noise ratio in regions 4 and 5.

8 Conclusions

An expression describing the received optical power from a retroreflecting beacon is derived. An expression describing the bidirectional reflection distribution function of a retroreflecting beacon is proposed. The received optical power dependence on range can be divided into five distinct regions (1–5). In regions 4 and 5 the dependencies are R^{-2} and R^{-3} respectively.

It is shown that the received pulse width and shape is a function of range except at long range i.e. in region 5. At long range the pulse shape is determined by the Gaussian intensity distribution in the laser beam. The pulse width is governed by the laser beam divergence and the scan velocity. An unexpected result is that the theoretical pulse amplitude vanishes at very short range i.e. in region 1. A design rule for the acceptance cone i.e. the field of view of the receiver optics is proposed.

Experimental data support the used models except at short range where the structure of the beacon material gives rise to a modulation in the received signal which is not modeled. Internal scattering in the receiver results in higher received power than predicted by the ideal receiver model.

The spectral density of the noise is shown to be essentially white. A design rule for an approximation of the ideal matched filter is proposed. An expression giving a conservative estimate of the signal-to-noise ratio is derived. All important system parameters are included in this expression .

A problem which has not been treated in this paper is pulses from false beacons. If a shiny surface is oriented such that the scanning laser beam hits the surface with a small incident angle enough optical power might be reflected to the receiver to generate a false detection.

The discussion in appendix B leading to the design rule for an approxi-

mation of the ideal matched filter has an engineering approach. It could be done more theoretically.

These topics will be addressed in future research.

References

[1] N. Eberhardt and M. D. Wagh, "Cyclopion, an autonomous guided vehicle for factory use", *SPIE Proc. Applications of Artificial Intelligence III,* Vol. 635, pp. 536–544, 1986.

[2] K. Nishide, M. Hanawa and T. Kondo, "Automatic Position Findings of Vehicle by Means of Laser", *Proc. of IEEE Int. Conf. on Robotics and Automation,* pp. 1343–1348, April 1986.

[3] T. Tsumura and M. Hashimoto, "Positioning and guidance of ground vehicle by use of laser and corner cube", *Proc. of IEEE Int. Conf. on Robotics and Automation,* pp. 1335–1342, April 1986.

[4] U. Wiklund, U. Andersson and K. Hyyppä, "AGV navigation by angle measurements", *Proc. 6th Int. Conf. Automated Guided Vehicle Systems,* pp. 199–212, October 1988.

[5] C. D. McGillem and T. S. Rappaport, "A Beacon Navigation Method for Autonomous Vehicles", *IEEE Trans. on Vehicular Technology,* vol. 38, no. 3, pp. 132–139, August 1989.

[6] P. M. Stevens, M. Roberts and V. K. Gek, "Truck location using retroreflectance strips and triangulation with laser equipment (TURTLE)", *Proc. 2nd European Conf. Automated Manuf.,* 1983.

[7] M. P. Robins, "Free-ranging automatic guided-vehicle system", *GEC Review,* vol. 2, no. 2, pp. 129–132, 1986.

[8] I. Kaisto et al, "Optical range finder for 1.5-10 m distances", *Applied Optics,* vol. 22, no. 20, pp. 3258–3264, October 1983.

[9] J. T. Kostamovaara, A. J. Mäkynen and R. A. Myllylä, "Method for Industrial Robot Tracking and Navigation based on Time-of-flight Laser Rangefinding and the Positon Sensitive Detection Technique", *SPIE Proc. Industrial Inspection,* vol. 1010, pp. 92–99, 1988.

[10] G. L. Miller and E. R. Wagner, "An Optical Rangefinder for Autonomous Robot Cart Navigation", *SPIE Proc. Mobile Robots II,* vol. 852, pp. 132–144, 1987.

[11] R. Hinkel, T. Knieriemen and E. von Puttkamer, "A Rotating Laser Range Finder and Attached Data Interpretation for Use in an Autonomous Mobile Robot", *Microprocessing and Microprogramming,* vol. 24, pp. 411–418, 1988.

[12] J. M Moser and H. R. Everett, "Wide-angle optical ranging system", *SPIE Proc. Mobile Robots IV,* vol. 1195, pp. 2–20, 1989.

[13] C. B. Bachman, "Laser Radar Systems and Techniques", Dedham, Artech House Inc., 1979.

[14] S. A. Hovanessian, "Introduction to Sensor Systems", Norwood, Artech HOuse Inc., 1988.

[15] H. Levenstein, "Lasar Radar Based on Diode Lasers", *SPIE Proc. Lasar Radar VI,* vol. 1416, pp. 30–43, January 1991.

[16] J. T. Verdeyen, "Laser electronics", Englewood Cliffs, Prentice-Hall Inc., 1989.

[17] Horn,"Robot Vision", Cambridge, The MIT Press, 1986.

[18] S. K. Nayar, K. Ikeuchi and T. Kanade, "Surface Reflection: Physical and Geometrical Perspectives", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 13, no. 7, pp. 611–634, July 1991.

[19] "FASIGN DURABRIGHT SERIES 7000 in the Construction Zone", *Data Sheet from Avery International,* August 1980.

[20] G. T. Uber, "Constant-Luminance Retroreflective Targets for Robot Guidance", *SPIE Proc. Automotive Displays and Industruial Illumination,* vol.958, pp. 176–180, 1988.

[21] "Optics Guide 5", Melles Griot Inc., section 20, pp. 12–21, 1990.

[22] C. D. Motchenbacher and F. C. Fitchen, "Low-noise electronic design", New York, John Wiley & Sons, 1973.

[23] Y. Netzer, "The design of low-noise amplifiers", *Proc. of the IEEE,* vol. 69, no. 6, pp. 728–741, June 1981.

[24] "Electro-Optics Handbook, EOH-11",Harrison, RCA Comercial Engineering, August 1974.

[25] M. Schwartz, "Information Transmission, Modulation and Noise", Third edition, McGraw-Hill Book Company, New York, 1980.

[26] K. Hyyppä and Klas Ericson, "A low-noise photodiode-amplifier circuit", *Research Report TULEA 1993:11,* April 1993.

A Receiver effective aperture

The area of the receiver effective aperture for a point on the beacon, i.e. the shaded area in Figure 5, can be found by purely geometrical calculations. The area is a function of the radii r_1 and r_2 defining the receiver aperture and of the variables r_3 and r. The radius r_3 is $R \tan \alpha_{r\,\mathrm{max}}$ and r is the distance between the point on the beacon and the optical axis. The following relations hold.

$$\begin{aligned} r_1 &> 0 \\ r_2 &> r_1 \\ r_3 &\geq 0 \\ r &\geq 0 \end{aligned} \tag{51}$$

Depending on the magnitudes of r_3 and r in relation to r_1 and r_2 nine different cases can occur. In many of the cases the mutual area between two intersecting circles need to be calculated. With the cosine law that area can be shown to be

$$A_{2c}(a,b,c) = b^2 \left[\arccos d_1 - d_1\sqrt{1-d_1^2}\right] + c^2 \left[\arccos d_2 - d_2\sqrt{1-d_2^2}\right] \tag{52}$$

where

$$d_1 = \frac{a^2+b^2-c^2}{2ab}$$

$$d_2 = \frac{a^2-b^2+c^2}{2ac}$$

and where a denotes the distance between the circles and b and c denotes the circle radii. The expression (52) is of course valid only if the circles really intersect, i.e. provided that

$$|b-c| < a < b+c$$

The nine cases with expressions for the effective aperture A_{re} are shown in Table 2 and a perspective view of the magnitude of the effective aperture as a function of r and r_3 is shown in Figure 16. The rr_3-plane is completely filled by regions corresponding to the nine cases. The region numbers in the figure corresponds to the case numbers in the table.

B Noise details

B.1 Internal noise

A common amplifier circuit used for photodiodes is a low-noise operational amplifier connected as a transimpedance amplifier. Expressions for the gain and spectral density of internally generated noise currents and voltages for

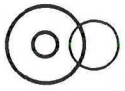
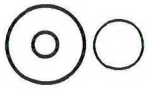
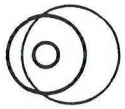
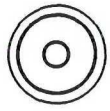
Case	Figure	Condition	A_{re}
1		$r + r_3 \leq r_1$	0
2		$r - r_3 > -r_1$ $r - r_3 < r_1$ $r + r_3 > r_1$ $r + r_3 \leq r_2$	$\pi {r_3}^2 - A_{2c}(r, r_1, r_3)$
3		$r - r_3 > -r_1$ $r - r_3 < r_1$ $r + r_3 > r_2$	$A_{2c}(r, r_2, r_3) - A_{2c}(r, r_1, r_3)$
4		$r - r_3 \geq r_1$ $r + r_3 \leq r_2$	$\pi {r_3}^2$
5		$r - r_3 \geq r_1$ $r - r_3 < r_2$ $r + r_3 > r_2$	$A_{2c}(r, r_2, r_3)$
6		$r - r_3 \geq r_2$	0
7		$r - r_3 \leq -r_1$ $r + r_3 \leq r_2$	$\pi({r_3}^2 - {r_1}^2)$
8		$r - r_3 > -r_2$ $r - r_3 \leq -r_1$ $r + r_3 > r_2$	$A_{2c}(r, r_2, r_3) - \pi {r_1}^2$
9		$r - r_3 \leq -r_2$	$\pi({r_2}^2 - {r_1}^2)$

Table 2: Expressions for the receiver effective aperture.

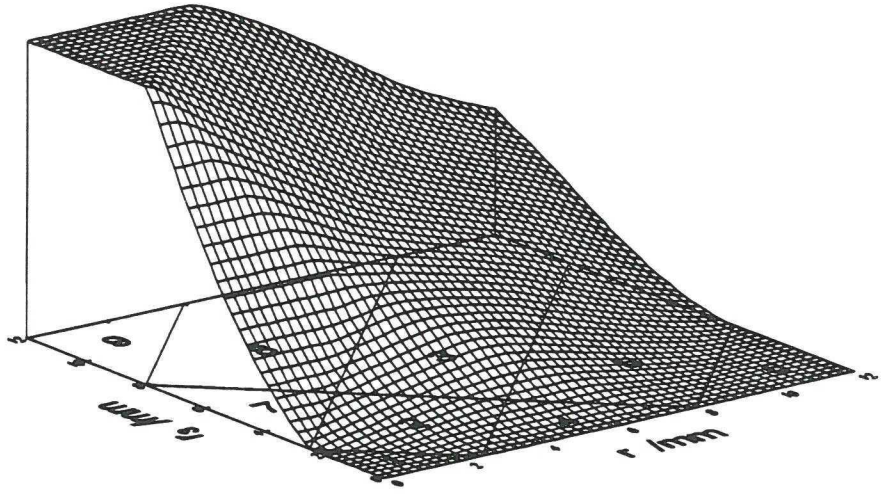


Figure 16: Perspective view of the magnitude of the receiver effective aperture as a function of r and r_3. The 9 regions described in Table 2 are indicated in the rr_3-plane.

such an amplifier have been derived elsewhere [23]. With the noise sources represented by a noise current generator in parallel with the photodiode the spectral density of the equivalent total internal noise current becomes

$$I_{nint}^2 = 2e\,|I_D| + \frac{4kT}{R_{fb}} + I_{na}^2 + V_{na}^2\omega^2 C_d^2 \tag{53}$$

The symbols denote:

e electron charge

I_D dark current of the photodiode

C_d photodiode capacitance

R_{fb} feedback resistance in the amplifier

I_{na}^2 spectral density of the operational amplifier noise current

V_{na}^2 spectral density of the operational amplifier noise voltage

k Boltzmanns constant

T temperature of the feedback resistor

ω angular frequency

B.2 Total noise

The spectral density of the total noise is the sum of the external and internal noise spectral densities because the noise sources are considered uncorrelated. Using (48) and (53) gives

$$I_{ntot}^2 = \frac{2}{\pi} e R_{ph} \eta_r E_a \alpha_{r\max}^2 A_r + 2e\,|I_D| + \frac{4kT}{R_{fb}} + I_{na}^2 + V_{na}^2 \omega^2 C_d^2 \qquad (54)$$

To gain insight into the relative importance of the different terms in (54), the parameter values valid for the anglemeter will be used. The parameter values are shown in Table 3.

$\lvert I_D\rvert$/nA	C_d/pF	R_{fb}/MΩ	I_{na}/fA/$\sqrt{Hz}$	V_{na}/nV/$\sqrt{Hz}$	T/K
0.1	2.0	10	10	12	300

Table 3: Anglemeter noise parameters.

The ambient light is assumed to come from the sun. An interference filter is used to block some of the ambient light in the anglemeter. The filter is centered at the laser wavelength and its optical bandwidth is $B_\lambda = 11.5$ nm. The maximum spectral irradiance of the sun is $E_{\lambda,sun} = 1.3$ kW/(m$^2\mu$m) at the laser wavelength and at sea level [24]. Then the irradiance of the ambient light becomes

$$E_a = E_{\lambda,sun} B_\lambda = 15 \text{ W/m}^2 \qquad (55)$$

With these values the terms of the spectral density become

$$I_{ntot}^2 = \left[1.09 + 0.32 + 16.6 + 1.0 + \omega^2 \cdot 5.76 \cdot 10^{-12}\right] \cdot 10^{-28} \text{ A}^2/\text{Hz} \qquad (56)$$

In the section below on filtering a design rule for the noise reduction filter will be proposed. The design rule gives an angular noise bandwidth of the filter equal to $\frac{\pi}{2\tau_f}$. The frequency dependent term in (56) then becomes less than $1 \cdot 10^{-31}$ A^2/Hz throughout the frequency region of the filter. This means that the noise can be considered white in this context. The anglemeter noise is clearly dominated by the thermal noise in the feedback resistor in the amplifier. A low-noise photodiode-amplifier which does not use a feedback resistor is described in [26].

Retaining only the three largest terms in (54) gives

$$I_{ntot}^2 \approx \frac{2}{\pi} e R_{ph} \eta_r E_a \alpha_{r\max}^2 A_r + \frac{4kT}{R_{fb}} + I_{na}^2 \qquad (57)$$

The spectral density at the output of a linear amplifier is found by multiplying the input spectral density with the squared magnitude of the gain of the amplifier. The gain of a transimpedance amplifier is approximately equal to

its feedback resistance and therefore the noise spectral density at the output of the amplifier i.e. at the input of the filter becomes

$$V_{nf}^2 \approx \frac{2}{\pi} e R_{ph} \eta_r E_a \alpha_{r\,\max}^2 A_r R_{fb}^2 + 4kTR_{fb} + I_{na}^2 R_{fb}^2 \tag{58}$$

B.3 Filtering

Since the total noise has been shown to be essentially white the matched filter will maximize the signal to noise ratio [25]. The matched filter has an impulse response which has the same shape as the input signal mirrored around $t = 0$. The signal to noise ratio after the matched filter becomes

$$(S/N)_{mf} = \frac{\text{Pulse energy}}{\text{Noise power spectral density}} \tag{59}$$

Only approximations to the ideal matched filter are realizable. A common realization for square pulses is a first order low-pass filter. It has been shown that a low-pass filter, with the time constant chosen equal to $0.8T_p$, where T_p is the pulse width, gives a signal-to-noise ratio within 1 dB of that of the ideal matched filter [25].

In the case at hand the pulse shape is a function of range and therefore the ideal filter should also be a function of range. In practice the range information is not available so this cannot be achieved. Fortunately the pulse shape becomes constant at long range and there the signal is also weakest which means that the filter should be designed for this range region. A Gaussian pulse has less high frequency components than a square pulse of equal width [25]. A reasonable design rule for a Gaussian shaped pulse could therefore be to use a first order low-pass filter with a time constant τ_f which is equal to the pulse width i.e

$$\tau_f = \frac{2\alpha_l}{\omega_b} \tag{60}$$

B.4 Signal-to-noise ratio

Knowing that the approximation to the matched filter in the case with square pulses gives a signal-to-noise ratio almost equal to that for the matched filter, (59) can be used to calculate the anglemeter signal-to-noise ratio. The pulse energy is given by (43) and by dividing (58) with R_{ref} an expression for the noise power spectral density is found. This gives the following approximate expression for the anglemeter signal-to-noise ratio at long range.

$$S/N \approx \frac{4R_{ph}^2 P_l^2 \eta_b^2 \eta_r^2 A_r^2 W^2}{\pi^{2.5} |\omega_b| \alpha_l \alpha_{b0}^4 R^6 \left[2\pi^{-1} e R_{ph} \eta_r E_a \alpha_{r\,\max}^2 A_r + 4kTR_{fb}^{-1} + I_{na}^2\right]} \tag{61}$$

The signal-to-noise ratio is strongly dependent on the divergence of the beacons and on range.

Part V

Error sources in a laser anglemeter used for mobile robot navigation

Error sources in a laser anglemeter used for mobile robot navigation

Kalevi Hyyppä
Luleå University of Technology
S-951 87 LULEÅ, Sweden

April 1993

Abstract

A laser scanner, used as an anglemeter in a navigation system for mobile robots, has been developed. It measures heading angles to beacons made of vertical stripes of retroreflective tape. A number of systematic and random error sources in the anglemeter are identified and analysed. The error from optical misalignments is shown to dominate the systematic error. Another important error source is the range dependence of the received signal from a beacon. Experiments verify the derived model of the misalignment induced error and they also show that the error can be efficiently compensated through calibration. A non-circular cross-section of the laser beam is shown to be a potential source of large systematic errors. Random angular errors due to electronic noise are surprisingly small.

1 Introduction

A navigation system for mobile robots and AGVs using a laser anglemeter has been proposed [1], [2]. The anglemeter measures heading angles in the horizontal plane to stationary beacons. The beacons carry no identity. They are made of vertically aligned stripes of retroreflective tape. The position in the horizontal plane of every beacon is known to the navigation system. The measured angles constitute input data to an algorithm which estimates the position and heading of the mobile robot.

An error $\Delta\gamma$ in measured angle propagates to errors in the estimates of the robot heading and position [3]. The maximum robot heading error from a single measurement to a beacon becomes $\Delta\gamma$ and the maximum robot position error becomes $\Delta\gamma\, R$ where R denotes the distance to the beacon. In this paper the total error is defined to be the sum of the magnitudes of all

uncompensated systematic errors and two standard deviations of the random error.

In the next section the anglemeter is briefly described. In the following section a number of error sources are identified and analysed. Calculated and estimated error magnitudes valid for the anglemeter will be given. Then results from measurements on the anglemeter are presented. After a discussion of the results some conclusion are made.

2 Anglemeter description

The anglemeter consists of an optical part and an electronic part. A block diagram of the anglemeter with some details of the electronic part is shown in Figure 1. The optical part is shown in Figure 2 and a photograph of the anglemeter is shown in Figure 3. A laser beam is scanned in the horizontal plane by a mirror rotating in positive direction i.e. the laser beam enters the beacons at their right edges (seen from the anglemeter). When the beam hits a beacon some of the reflected radiation will reach a photodiode in the receiving part of the anglemeter. The signal current from the photodiode is converted to a voltage and amplified. After passing a filter for noise reduction, the signal voltage reaches a comparator where it is compared with a threshold voltage. When the comparator trips it is assumed that the laser beam has hit a beacon and the angle of the rotating mirror is recorded. The angle is measured with an incremental encoder attached to the axis of the rotating mirror and a counter.

The optical system is coaxial i.e. the optical axes of the laser beam and of the receiver optics are coincident. One of the deflecting mirrors has a hole on the optical axis. The outgoing illuminating beam passes the hole while part of the reflected radiation is deflected by the same mirror into the receiver optics. A small area in the center of the receiver aperture is wasted due to the hole.

3 Error sources

3.1 The incremental encoder

The encoder has two channels which each delivers a pulse-train when the scan mirror rotates. The pulse-trains have a 90 degree phase shift between them to enable the detection of the rotation direction. By using the signals from both channels, four states can be decoded for each period of the pulse-trains. The resolution $\Delta\gamma_{enc}$ then becomes $\frac{2\pi}{4n}$ where n denotes the number of pulses per revolution on each channel.

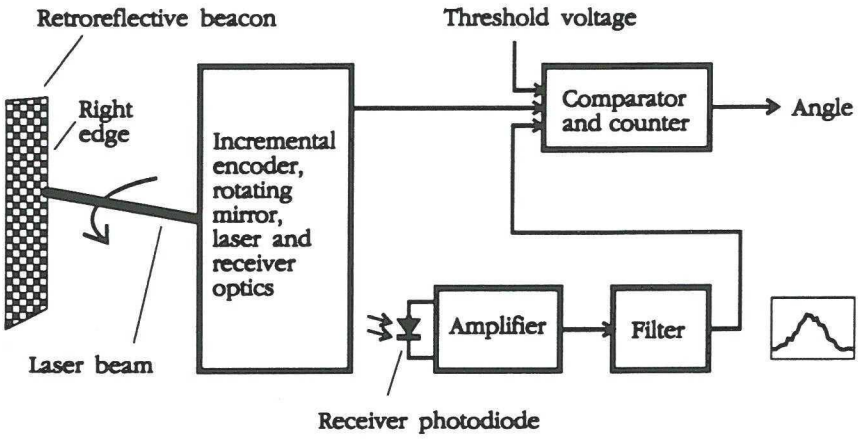


Figure 1: Block diagram of the anglemeter. Reflected radiation from the beacon hits a photodiode in the anglemeter receiver. An example of the received signal after the noise reduction filter is shown in the insert.

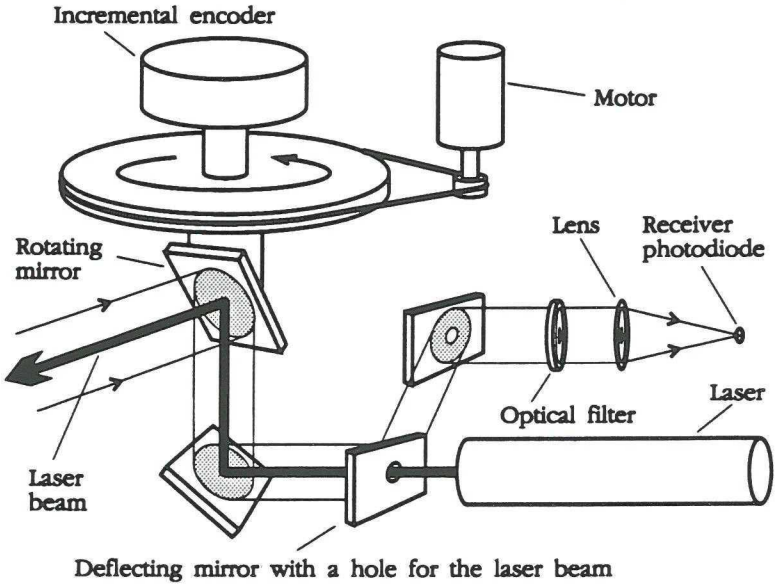


Figure 2: Optical part of the anglemeter. A small area in the center of the receiver aperture is wasted because it is occupied by the outgoing laser beam.

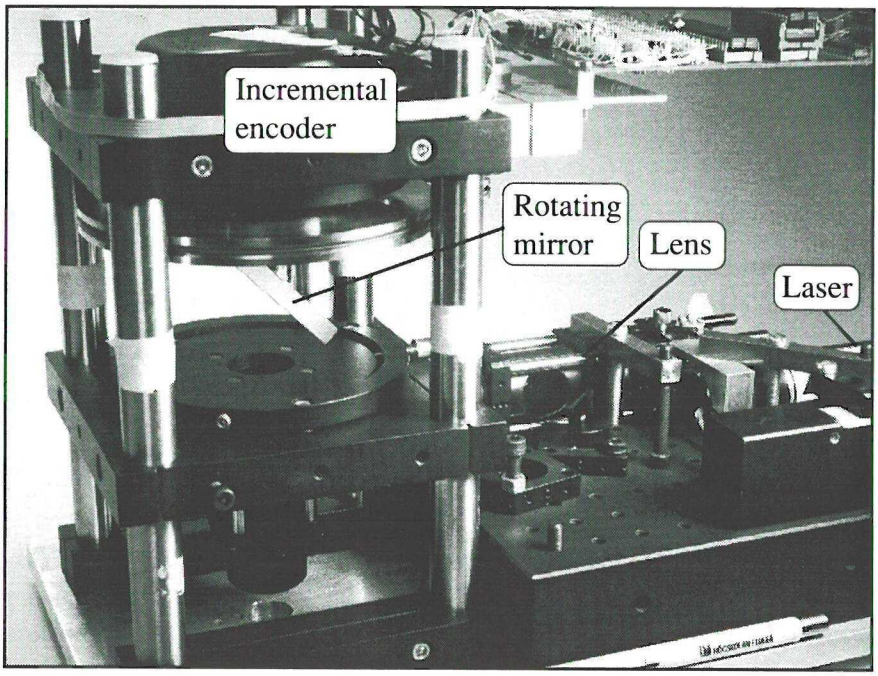


Figure 3: A photograph of the anglemeter. Note the pen in the bottom right corner for size reference.

The random error due to the digitization has a rectangular distribution with a magnitude equal to $\Delta\gamma_{enc}^{-1}$ and a width equal to $\Delta\gamma_{enc}$. The standard deviation in measured angle due to the encoder then becomes

$$\sigma_{enc} = \frac{2\pi}{4n\sqrt{12}} \tag{1}$$

which can be simplified to

$$\sigma_{enc} = \frac{\pi}{4n\sqrt{3}} \tag{2}$$

The used encoder produces 9000 pulses per revolution on each channel which gives a resolution of 174 μrad. The standard deviation for the anglemeter due to the encoder therefore becomes

$$\sigma_{enc} = 50\ \mu\text{rad} \tag{3}$$

A limit for the systematic error is not stated by the encoder manufacturer. Similar encoders from another manufacturer have specified maximum systematic errors ranging from 24 to 34 μrad. A conservative estimate of the maximum systematic error due the encoder in the anglemeter is therefore

$$e_{enc} = 40\,\mu\text{rad} \tag{4}$$

3.2 Misalignments

There are several optical alignments to be made in the anglemeter. Misalignments between the receiver and the rest of the anglemeter influences only the received signal strength and not the measured angle. Errors in measured angle due to range-dependent signal strength will be discussed later.

Misalignments between the laser beam and the axis of the rotating mirror have the following effects. Assume first that the laser beam and the rotational axis are parallel but displaced from each other. The axis of the outgoing laser beam will then move along the surface of a cylinder after deflection in the rotating mirror. The axis of the cylinder is coincident with the axis of the nominally deflected laser beam. The laser beam makes one revolution around the cylinder for each revolution of the rotating mirror. The radius of the cylinder is equal to the displacement between the laser beam and the rotational axis.

The maximum error in measured angle to a beacon due to the displacement becomes dR^{-1} where d denotes the displacement and R denotes the range to the beacon. The displacement is estimated to be less than 0.1 mm and therefore this error is considered insignificant compared to the next error to be discussed, namely the error from the angular misalignment between the laser beam and the rotational axis.

Assume that the axis of the laser beam intersects the rotational axis at the point where the rotational axis intersects the reflecting surface of the rotating mirror. Assume furthermore that the laser beam has a certain angular misalignment with the rotational axis. The axis of the outgoing laser beam will then move along the surface of a cone after deflection in the rotating mirror. The cone has the same axis as the cylinder discussed above. It will be shown that the half-angle of the cone is equal to the misalignment between the laser beam and the rotational axis.

The assumption of the position of the point of intersection between the axis of the laser beam and the rotational axis has been made to simplify the analysis below. The assumption is not essential for the end result. A displacement of the intersection point along the rotational axis will displace the discussed cone with the same amount along the axis of the nominally deflected laser beam. The case with the cylinder discussed above can in fact be treated as the special case of angular misalignment when the intersection point moves to infinity at the same time as the angular misalignment goes to zero.

The angular misalignment is sometimes called wobble in literature on laser scanners. The wobble-induced error can be eliminated with more complex deflection devices than the single rotating mirror used in the anglemeter [4], [5], [6]. Next follows a calculation of the wobble-induced error in the anglemeter. This calculation is also valid for a laser scanner with a single

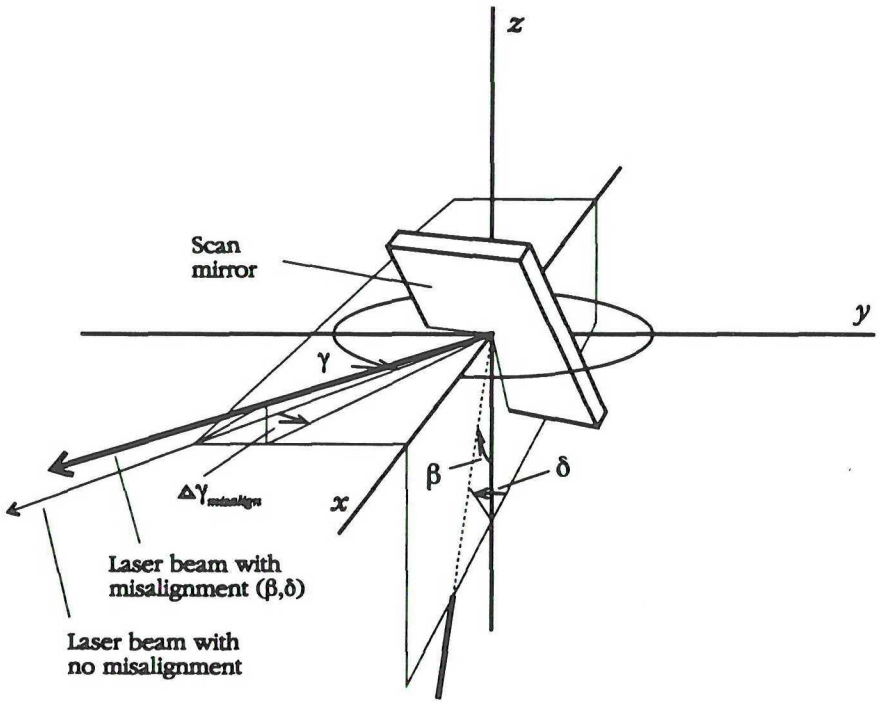


Figure 4: Coordinate system and symbols used to analyse the influence of the misalignment (β,δ) on the measured angle γ.

rotating mirror.

Figure 4 shows a modified enlargement of the scan mirror and the laser beam shown in Figure 2, and an attached coordinate system. The z-axis of the coordinate system is coincident with the rotational axis of the mirror. The origin is at the reflecting surface of the mirror. The x-axis is parallel with the reference direction of the anglemeter i.e. the direction of a nominally deflected laser beam when the scan angle γ is zero.

The incoming laser beam is assumed to hit the mirror in the origin of the coordinate system. Note that the reflected laser beam does not lie in the xy-plane and that $\Delta\gamma_{misalign}$ is the angle between the nominally reflected beam — i.e. with no misalignments — and the projection in the xy-plane of the actually reflected beam.

Let n_0 be a unit vector normal to the mirror surface when $\gamma = 0$.

$$n_0 = \frac{1}{\sqrt{2}}(1,\ 0,\ -1)^{\mathrm{T}} \tag{5}$$

This stationary vector can be transformed to a rotating vector $n(\gamma)$ which

is normal to the rotating mirror surface by premultiplying with the rotation matrix $R_m(\gamma)$.

$$n(\gamma) = R_m(\gamma)\, n_0 \tag{6}$$

The rotation matrix is given by

$$R_m(\gamma) = \begin{pmatrix} c\gamma & -s\gamma & 0 \\ s\gamma & c\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7}$$

where $c\gamma$ and $s\gamma$ denotes $\cos\gamma$ and $\sin\gamma$. The product is

$$n(\gamma) = \frac{1}{\sqrt{2}}(c\gamma,\ s\gamma,\ -1)^{\mathrm{T}} \tag{8}$$

The reflection of the laser beam can be modeled with a rotation matrix R_r, which is defined by

$$R_r(\gamma) = I - 2n(\gamma)\, n^{\mathrm{T}}(\gamma) \tag{9}$$

where I denotes the $3 * 3$ identity matrix. The rotation matrix becomes

$$R_r(\gamma) = \begin{pmatrix} s^2\gamma & -c\gamma\, s\gamma & c\gamma \\ -s\gamma\, c\gamma & c^2\gamma & s\gamma \\ c\gamma & s\gamma & 0 \end{pmatrix} \tag{10}$$

A unit vector p_{in} parallel with the incoming beam can be found by inspection of Figure 4.

$$p_{in} = (s\beta\, c\delta,\ s\beta\, s\delta,\ c\beta)^{\mathrm{T}} \tag{11}$$

A unit vector p_{out} parallel with the reflected beam is formed by premultiplying p_{in} with R_r

$$p_{out} = R_r\, p_{in} \tag{12}$$

which becomes

$$p_{out} = \begin{pmatrix} s^2\gamma\, s\beta\, c\delta - c\gamma\, s\gamma\, s\beta\, s\delta + c\gamma\, c\beta \\ -s\gamma\, c\gamma\, s\beta\, c\delta + c^2\gamma\, s\beta\, s\delta + s\gamma\, c\beta \\ c\gamma\, s\beta\, c\delta + s\gamma\, s\beta\, s\delta \end{pmatrix} \tag{13}$$

Finally the projection of p_{out} in the xy-plane p_{outp} becomes

$$p_{outp} = \begin{pmatrix} s^2\gamma\, s\beta\, c\delta - c\gamma\, s\gamma\, s\beta\, s\delta + c\gamma\, c\beta \\ -s\gamma\, c\gamma\, s\beta\, c\delta + c^2\gamma\, s\beta\, s\delta + s\gamma\, c\beta \\ 0 \end{pmatrix} \tag{14}$$

The vector p_{outp} can be expressed as a general vector in the xy-plane with the error $\triangle\gamma_{misalign}$ explicitly visible

$$p_{outp} = k\,(\cos(\gamma + \triangle\gamma_{misalign}),\ \sin(\gamma + \triangle\gamma_{misalign}),\ 0)^{\mathrm{T}} \tag{15}$$

where k is a constant. After some manipulations where β, δ and $\Delta\gamma_{misalign}$ are assumed small the error $\Delta\gamma_{misalign}$ becomes

$$\Delta\gamma_{misalign} = -\beta\sin(\gamma - \delta) \tag{16}$$

This is a systematic contribution to the error in measured angle and might also be the dominating error if the alignment is not done with the outmost care. The maximum systematic error $e_{misalign} = \max(|\Delta\gamma_{misalign}|)$ from misalignment is then

$$e_{misalign} = \beta \tag{17}$$

A random variation in β would produce a random error in measured angle. The cause of the random variation could be e.g. play in the bearings supporting the rotating mirror or a random variation in the heading of the beam from the laser.

The ball bearings supporting the rotating mirror in the anglemeter are preloaded to eliminate the play. The main contributor to the random variation in β is the random variation in the heading of the beam from the laser. The standard deviation of the variation in β for the anglemeter is estimated to be

$$\sigma_{misalign} = 50\ \mu\text{rad} \tag{18}$$

The magnitude of β is dependent on the resolution of the alignment mechanism in the anglemeter and on the skill of the person making the alignment. The magnitude of β is estimated to be of the order of 1 mrad, which gives the following estimate of the maximum systematic error from misalignment.

$$e_{misalign} = 1\ \text{mrad} \tag{19}$$

3.3 Variations in detection moment

The angle to a beacon is measured when the voltage at the comparator input passes the threshold level of the comparator. The nominal detection moment has been chosen to be that moment when the center of the laser beam passes the right edge of the beacon, i.e. when half of the laser beam is incident on the beacon. The shape and strength of the received pulse is a complex function of the range to the beacon. A representative set of plots of the voltage at the input of the noise reduction filter — for different range values — is shown in Figure 5. The figure is taken from [7] where the noise reduction filter is also discussed. The transfer function of the filter is

$$H_f = \frac{1}{1 + j2\pi f\,\tau_f} \tag{20}$$

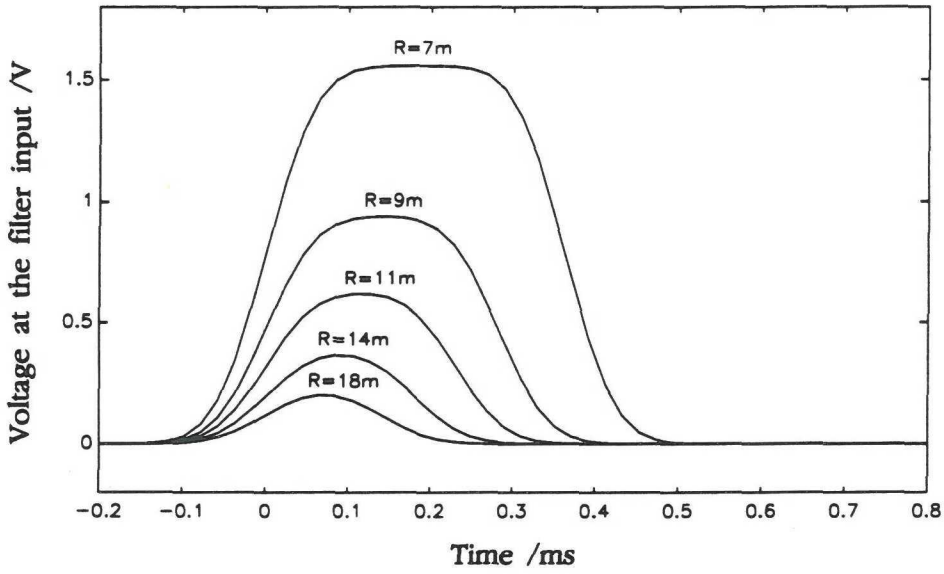


Figure 5: Numerically evaluated voltages at the input of the noise reduction filter. The pulse shape goes gradually from the accumulated Gaussian distribution shape to the Gaussian distribution shape when the laser beam expands with range.

where τ_f denotes the filter time-constant. The following design rule is proposed in [7] to optimize the signal-to-noise ratio after the filter.

$$\tau_f = \frac{2\alpha_l}{\omega_m} \tag{21}$$

where α_l denotes the far-field divergence of the laser beam and ω_m denotes the angular velocity of the rotating mirror. The far-field divergence is 0.6 mrad and the angular velocity is 2π rad/s for the anglemeter. The time constant τ_f then becomes 0.2 ms. The filter delays the received pulse and changes its shape. The voltage after the filter, i.e. at the comparator input, has been numerically evaluated with the aid of some of the results in [7]. Several examples of the voltage at the comparator input — with range as parameter — are shown in Figure 6. The range parameter increases exponentially over the interval (3.0–35.8) m. Depending on the range and the chosen threshold level the actual detection moment can occur slightly before or after the nominal detection moment. The variation in detection moment gives a corresponding variation in measured angle. The range dependent signal strength therefore creates a range dependent systematic error in measured angle. The time scale in Figure 6 has been complemented with an angle

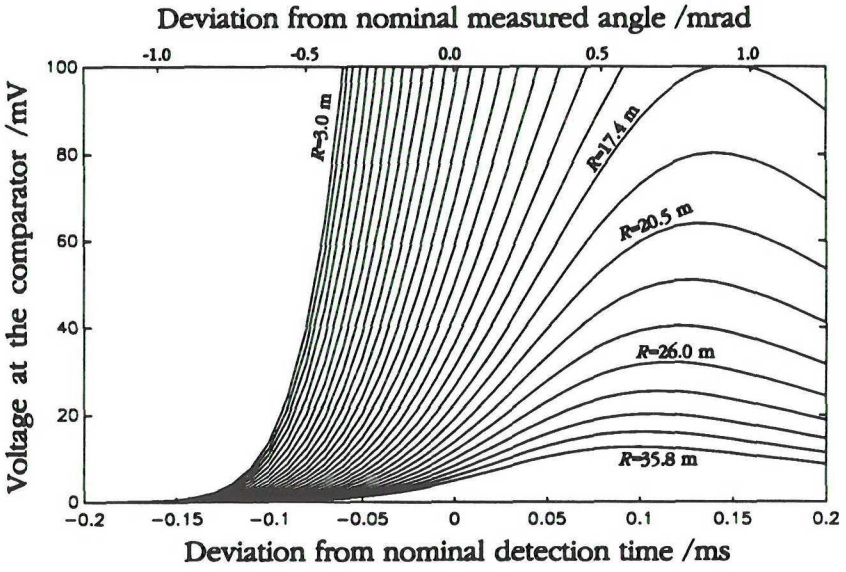


Figure 6: Numerically evaluated voltage at the comparator input for different range values. The angular deviation scale has been drawn using the constant angular velocity of the rotating mirror.

scale by using the magnitude of the angular speed of the rotating mirror. By choosing a threshold level the deviation from the nominal measured angle as a function of range can be found from Figure 6.

The actual detection moment is also influenced by the noise in the voltage at the comparator input. The noise is generated by background light and by the anglemeter receiver. The random variation in detection time results in a random angle error.

3.3.1 Noise

The standard deviation σ_n in the measured angle due to noise is calculated as follows.

$$\sigma_n = \sqrt{\int_0^\infty \gamma_n^2 \, df} \tag{22}$$

where γ_n^2 denotes the spectral density of the noise in the measured angle γ due to the noise in the voltage at the comparator input. The small-signal transfer function H_a between the voltage at the comparator input and the detected angle is by definition

$$H_a = \left(\frac{d\gamma}{dv_c} \right)_{v_{threshold}} \tag{23}$$

where v_c denotes the voltage at the comparator input. The symbol $v_{threshold}$, indicating that the derivative is evaluated at the threshold voltage, is omitted in the following expressions. The derivative can be expanded to $\frac{d\gamma}{dt} \cdot \frac{dt}{dv_c}$ and because $\frac{d\gamma}{dt} = \omega_m$ the transfer function becomes

$$H_a = \omega_m \left(\frac{dv_c}{dt} \right)^{-1} \tag{24}$$

An expression for the spectral density V_{nf}^2 of the noise voltage before the filter is given in [7]. The spectral density of the angular noise is

$$\gamma_n^2 = V_{nf}^2 \, |H_f \, H_a|^2 \tag{25}$$

Using (20), (21), (24) and (25) the spectral density of the angular random fluctuations becomes

$$\gamma_n^2 = V_{nf}^2 \, \frac{\omega_m^2}{\left(\frac{dv_c}{dt}\right)^2 \left(1 + \left(2\pi f \, \frac{2\alpha_l}{\omega_m}\right)\right)^2} \tag{26}$$

In [7] it is shown that the spectral density of the noise voltage V_{nf}^2 can be considered white. Then (22) can be written

$$\sigma_n = \sqrt{B_n \gamma_{n0}^2} \tag{27}$$

where B_n denotes the noise bandwidth of the noise reduction filter and γ_{n0}^2 denotes the low frequency spectral density of the angular noise. The low frequency asymptotic limit of expression (26) gives

$$\gamma_{n0}^2 = V_{nf}^2 \, \frac{\omega_m^2}{\left(\frac{dv_c}{dt}\right)^2} \tag{28}$$

The noise bandwidth of a first-order low-pass filter with time-constant τ_f is $(4\tau_f)^{-1}$. Using this and (28) in (27) the standard deviation in measured angle due to noise becomes

$$\sigma_n = \left| \frac{dv_c}{dt} \right|^{-1} \sqrt{\frac{\omega_m^3 \, V_{nf}^2}{8\alpha_l}} \tag{29}$$

The time derivative of the signal voltage at 10 mV — which could be a typical threshold level — has been estimated from Figure 6 to be at least 0.1 V/ms for the displayed range interval. The spectral density V_{nf}^2 is calculated to be less than $2 \cdot 10^{-13}$ V^2/Hz. With these values a conservative estimate of the standard deviation in measured angle due to noise in the received signal becomes

$$\sigma_n = 1 \ \mu\text{rad} \tag{30}$$

3.3.2 Range dependent signal strength

An extended data set compared to the set which was used to produce the curves in Figure 6 has been used to produce the curves in Figure 7. There the deviation from the nominal measured angle is given as a function of range with the threshold level as parameter.

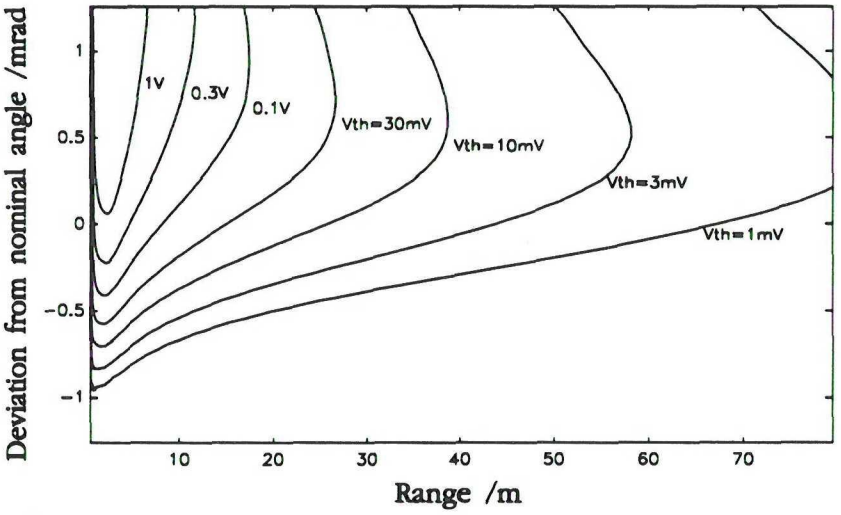


Figure 7: Deviation from nominal measured angle as a function of range with the threshold level of the comparator as parameter.

The figure shows that the deviation interval increases somewhat with decreasing threshold voltage. A conservative estimate of the deviation interval for threshold voltages above 1 mV is 1.4 mrad. The maximum systematic error from the range-dependent signal strength is thus

$$e_{range} = 0.7 \text{ mrad} \tag{31}$$

The range dependent error can be eliminated in the navigation algorithm because the range is known with good accuracy by that algorithm. The non-zero standard deviation of the range information creates a random error in measured angle due to the range dependent signal strength.

From Figure 7 the derivative of the deviation from the nominal angle is estimated to be less than 50 μrad/m for threshold voltages below 0.1 V. The standard deviation in position for a mobile robot equipped with the anglemeter has been measured to 1 mm when the robot traveled along a straight path at 0.3 m/s [8]. A conservative estimate of the standard

deviation in the range to each beacon is then 10 mm. The following estimate of the standard deviation in the range dependent error is then valid

$$\sigma_{range} < 1 \ \mu\text{rad} \tag{32}$$

3.4 Elliptic laser beam cross-section

The outgoing laser beam rotates around its own axis due to the deflection in the rotating mirror. A deviation from a circular cross-section of the laser beam gives a systematic error in the measured angle to a beacon. The beam flux density profile is Gaussian [7]. The boundary of the cross-section is defined to be where the radiation flux density has fallen to e^{-2} of its value in the center of the beam.

The radiation pattern from a semiconductor laser is elliptic. Depending on how sophisticated the collimation is done the residual beam eccentricity can vary from about 0.5 to more than 0.9 [9]. A complicating fact is that the eccentricity of the beam is also a function of range. At close range the beam is elliptic, then circular at a certain range and then, at long range, elliptic again with the long axis of the ellipsis turned 90 degrees in relation to the situation at close range.

The eccentricity of the laser beam does not generate any angle error if the actual detection moment of a beacon is equal to the above defined nominal detection moment. Figure 7 shows that — for a fixed threshold voltage — this is true for only two range values. For other range values the detection moment is dependent on the eccentricity of the beam at the beacon and the relative orientation between the ellipsis defining the cross-section and the right edge of the beacon. An estimate of the error due to the eccentricity can be made as follows.

Let the beam eccentricity be ϵ and the far-field divergence along the major axis be α_l. Suppose furthermore that the threshold voltage is such that the beacon is detected when the beam boundary (e^{-2} strength) reaches the right edge of the beacon. The minimum deviation from the nominal detection moment occurs when the major axis of the ellipsis defining the beam cross-section is parallel with the right edge of the beacon. The maximum deviation occurs when the major axis of the ellipsis is perpendicular to the right edge of the beacon. The maximum systematic error in measured angle due to elliptic beam cross-section is the difference in angular deviation for the two described cases i.e.

$$e_{elliptic} = \alpha_l \left(1 - \sqrt{1 - \epsilon^2}\right) \tag{33}$$

The HeNe-laser used in the anglemeter has an almost circular beam cross-section. The manufacturer states a far-field divergence of 0.6 mrad. Observations of the spot generated by the anglemeter laser at several distances indicate an eccentricity less than 0.3 independent of range. An estimate of

the maximum systematic error in the anglemeter due to elliptic beam cross-section then becomes

$$e_{elliptic} = 30 \ \mu\text{rad} \tag{34}$$

A semiconductor laser with the same far-field divergence but with an eccentricity equal to 0.9 would give a corresponding error equal to 0.34 mrad.

A random variation in the elliptic cross-section of the laser beam generates a random error in the measured angle. The standard deviation of that error is conservatively estimated to

$$\sigma_{elliptic} = 10 \ \mu\text{rad} \tag{35}$$

3.5 Partly blocked beacons

Normally the angle is measured to a point close to the right edge of the beacon. If that edge is not visible from the scanner, due to an obstacle, the angle will instead be measured to a point close to the right edge of the visible part of the beacon. This gives an angle error which in contrast to the other errors has a nonzero mean. The error can be large but the probability that it will occur is low. An estimate of the probability and the error magnitude can be made as follows.

Assume that n beacons of width W are visible from the anglemeter and that there is one obstacle between the anglemeter and the beacons. Assume furthermore that the range to all beacons is R. The maximum angle a beacon occupies is

$$\gamma_b = \frac{W}{R} \tag{36}$$

The maximum total angle which all beacons occupy is then $n\gamma_b$. An error in measured angle is generated if the left edge of the obstacle falls within the angular sector of any of the beacons. The angular distribution of the left edge of the obstacle is assumed to be rectangular over the interval 0 to 2π rad. The probability of an error due to a blocked beacon is then

$$p_{berr} = \frac{n\gamma_b}{2\pi} \tag{37}$$

The error distribution is rectangular over the interval 0 to γ_b with an additional impulse at zero error. This means that the most probable error from partly blocked beacons is no error at all. The errors from partly blocked beacons can be characterized as outliers which often can be eliminated in the processing of the measured angles.

Some typical parameter values for the anglemeter are $n = 10$, $W = 16$ mm and $R = 5$ m. They give the following result

$$\gamma_b = 3.2 \text{ mrad} \tag{38}$$

deviation in the range to each beacon is then 10 mm. The following estimate of the standard deviation in the range dependent error is then valid

$$\sigma_{range} < 1 \ \mu\text{rad} \tag{32}$$

3.4 Elliptic laser beam cross-section

The outgoing laser beam rotates around its own axis due to the deflection in the rotating mirror. A deviation from a circular cross-section of the laser beam gives a systematic error in the measured angle to a beacon. The beam flux density profile is Gaussian [7]. The boundary of the cross-section is defined to be where the radiation flux density has fallen to e^{-2} of its value in the center of the beam.

The radiation pattern from a semiconductor laser is elliptic. Depending on how sophisticated the collimation is done the residual beam eccentricity can vary from about 0.5 to more than 0.9 [9]. A complicating fact is that the eccentricity of the beam is also a function of range. At close range the beam is elliptic, then circular at a certain range and then, at long range, elliptic again with the long axis of the ellipsis turned 90 degrees in relation to the situation at close range.

The eccentricity of the laser beam does not generate any angle error if the actual detection moment of a beacon is equal to the above defined nominal detection moment. Figure 7 shows that — for a fixed threshold voltage — this is true for only two range values. For other range values the detection moment is dependent on the eccentricity of the beam at the beacon and the relative orientation between the ellipsis defining the cross-section and the right edge of the beacon. An estimate of the error due to the eccentricity can be made as follows.

Let the beam eccentricity be ϵ and the far-field divergence along the major axis be α_l. Suppose furthermore that the threshold voltage is such that the beacon is detected when the beam boundary (e^{-2} strength) reaches the right edge of the beacon. The minimum deviation from the nominal detection moment occurs when the major axis of the ellipsis defining the beam cross-section is parallel with the right edge of the beacon. The maximum deviation occurs when the major axis of the ellipsis is perpendicular to the right edge of the beacon. The maximum systematic error in measured angle due to elliptic beam cross-section is the difference in angular deviation for the two described cases i.e.

$$e_{elliptic} = \alpha_l \left(1 - \sqrt{1 - \epsilon^2}\right) \tag{33}$$

The HeNe-laser used in the anglemeter has an almost circular beam cross-section. The manufacturer states a far-field divergence of 0.6 mrad. Observations of the spot generated by the anglemeter laser at several distances indicate an eccentricity less than 0.3 independent of range. An estimate of

the maximum systematic error in the anglemeter due to elliptic beam cross-section then becomes

$$e_{elliptic} = 30 \ \mu\text{rad} \tag{34}$$

A semiconductor laser with the same far-field divergence but with an eccentricity equal to 0.9 would give a corresponding error equal to 0.34 mrad.

A random variation in the elliptic cross-section of the laser beam generates a random error in the measured angle. The standard deviation of that error is conservatively estimated to

$$\sigma_{elliptic} = 10 \ \mu\text{rad} \tag{35}$$

3.5 Partly blocked beacons

Normally the angle is measured to a point close to the right edge of the beacon. If that edge is not visible from the scanner, due to an obstacle, the angle will instead be measured to a point close to the right edge of the visible part of the beacon. This gives an angle error which in contrast to the other errors has a nonzero mean. The error can be large but the probability that it will occur is low. An estimate of the probability and the error magnitude can be made as follows.

Assume that n beacons of width W are visible from the anglemeter and that there is one obstacle between the anglemeter and the beacons. Assume furthermore that the range to all beacons is R. The maximum angle a beacon occupies is

$$\gamma_b = \frac{W}{R} \tag{36}$$

The maximum total angle which all beacons occupy is then $n\gamma_b$. An error in measured angle is generated if the left edge of the obstacle falls within the angular sector of any of the beacons. The angular distribution of the left edge of the obstacle is assumed to be rectangular over the interval 0 to 2π rad. The probability of an error due to a blocked beacon is then

$$p_{berr} = \frac{n\gamma_b}{2\pi} \tag{37}$$

The error distribution is rectangular over the interval 0 to γ_b with an additional impulse at zero error. This means that the most probable error from partly blocked beacons is no error at all. The errors from partly blocked beacons can be characterized as outliers which often can be eliminated in the processing of the measured angles.

Some typical parameter values for the anglemeter are $n = 10$, $W = 16$ mm and $R = 5$ m. They give the following result

$$\gamma_b = 3.2 \text{ mrad} \tag{38}$$

and

$$p_{berr} = 0.005 \tag{39}$$

which means that the maximum error due to partly blocked beacons becomes 3.2 mrad with these parameters. The mean value of the error is only 8 μrad as a result of the low probability.

4 Experimental results

A number of angle measurements have been made with the anglemeter. The anglemeter was mounted on a precision turntable. The turntable was driven by a step motor through a reduction gearbox followed by a worm-gear. The reduction ratios were 10:1 and 36:1 respectively. The step motor was controlled by a computer. The computer was also used to collect measured angles from the anglemeter. With 200 steps/revolution in the step motor the angular resolution became $\frac{2\pi}{200 \cdot 10 \cdot 36} = 87$ μrad. The resolution of the anglemeter is twice this value as stated earlier.

Only one beacon was made visible for the anglemeter. The beacon was 16 mm wide and it was glued to a wall at 12 m distance from the anglemeter. The turntable with the anglemeter was firmly mounted on an optical table. The turntable was preloaded with a constant torque to prevent measurement errors due to backlash in the gears. At the start of a measurement series both the set angle of the turntable and the measured angle to the beacon, reported by the anglemeter, were defined to be zero. This arbitrary definition of the zero angle creates a bias angle which has to be removed in the processing of the measured data. The measurement procedure for each turntable set angle was done as follows.

The turntable was turned a predetermined angle by the computer. After a delay of 2 s a number of consecutively measured angles to the beacon was collected from the anglemeter. The reason for the delay was to allow all vibrations due to the motion of the turntable to vanish before the measurements were made. The number of measurements at each turntable set angle ranged from 1 to 10. The mean value of the error angles, i.e. the difference between the measured angles and the turntable set angle, was stored together with the set angle in a file in the computer.

Six measurement series with a total of approximately 6700 angle measurements were used for the error analyses. In all series the turntable was turned through 360 degree. The raw data from all the measurements are shown in Figure 8. Every dot represents the mean of the error angle at a certain turntable angle. The clustering of the data along horizontal lines is due to the resolution in the anglemeter. Only a few measurements falls between the resolution steps. In those cases typically one measurement has deviated from the others at a certain turntable position giving a mean error outside

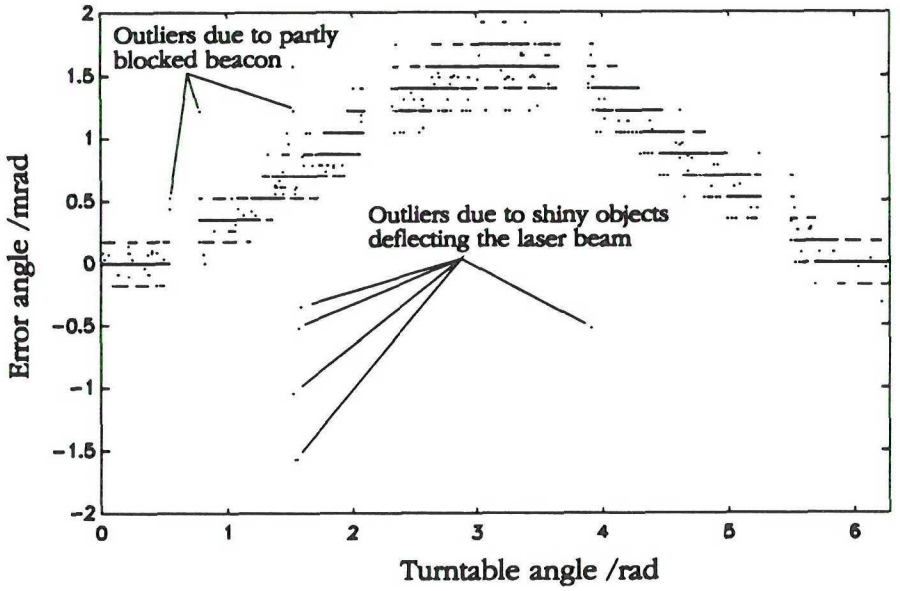


Figure 8: Raw measurement data, every dot represents the mean of the error angle at a certain turntable angle.

the resolution steps. Other visible phenomena are the periodic systematic error, the four gaps with no data and some outliers around the gaps. In fact there is a fifth gap with associated outliers at around 1.5 rad turntable angle. It is too narrow to be clearly visible. The gaps are due to five obstacles due to the mechanical structure of the anglemeter. Four of the obstructing structural elements are the four cylindrical posts visible in Figure 3.

In the processing of the measured data the outliers around the gaps were first removed. The periodic variation in the error angle is due to the alignment error discussed above. The error could easily be identified by fitting the expression

$$\Delta\gamma = \gamma_0 - \beta\sin(\gamma - \delta) \tag{40}$$

to the measured data. The symbol γ_0 denotes the bias angle mentioned above. The fitting was done in the least square sense. It was then discovered that the measurement data still contained a systematic component after the removal of the error due to misalignment and bias. This error is also periodic in the turntable angle but with a period equal to $\frac{2\pi}{36}$. The probable cause of the error is the worm-gear in the turntable.

The bias, the alignment error and the worm-gear error were identified in the least square sense by fitting the expression

$$\Delta\gamma = \gamma_0 - \beta\sin(\gamma - \delta) + \gamma_w\sin(36\gamma - \varphi_w) \tag{41}$$

Source	e/μrad	σ/μrad
Encoder	40	50
Misalignment	1000	50
Noise	0	1
Range	700	< 1
Elliptic cross-section	30	10

Table 1: Derived systematic errors and standard deviations of random errors for the anglemeter.

to the measured data. The symbol γ_w denotes the amplitude of the worm-gear error and the symbol φ_w denotes the phase angle of the worm-gear error. The data in one of the measurement series — with the outliers removed — and the fitted curve are shown in Figure 9. The residual error i.e. the difference between the measured data and the fitted curve is shown in Figure 10. Note that the vertical scales in figures 8, 9 and 10 are different. The specified accuracy of the turntable is indicated in Figure 10. The exact meaning of the accuracy specification is not declared by the manufacturer.

5 Discussion

Figure 10 shows that the remaining systematic error is smaller than the standard deviation of the random error. Whether the remaining errors come from the anglemeter or the turntable is questionable. The only method to solve that ambiguity is to use a more accurate turntable. To really be able to measure the errors of the anglemeter would demand a turntable with both a systematic error and a standard deviation of the random error less than say 10 μrad.

Most of the outliers visible in Figure 8 behave as predicted in the discussion on partly blocked beacons. Some of them have a negative value which is contrary to prediction. This is probably due to the following phenomenon.

The obstructing objects are made of metal with a shiny surface. At some points the laser beam hits them with an incident angle very close to 90 degrees. Then part of the beam is deflected slightly to the right or to the left, depending on which side of the obstacle which made the deflection. If the deflection is sufficiently small and the deflected beam hits the beacon, the reflected radiation will reach the photodiode in the receiver. This will result in a positive or negative angle error.

The derived systematic errors and standard deviations of the random errors are collected in Table 1. The systematic error is dominated by the contribution from the misalignment. The value shown in the table is an estimate of the alignment result of a skilled person. The fitting of the exper-

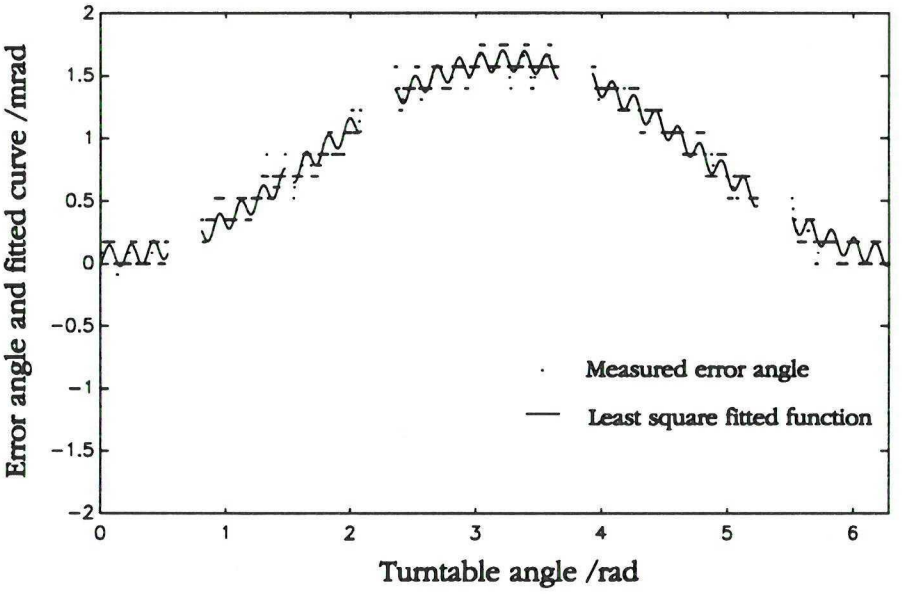


Figure 9: Processed measurement data in one of the measurement series and the corresponding fitted curve.

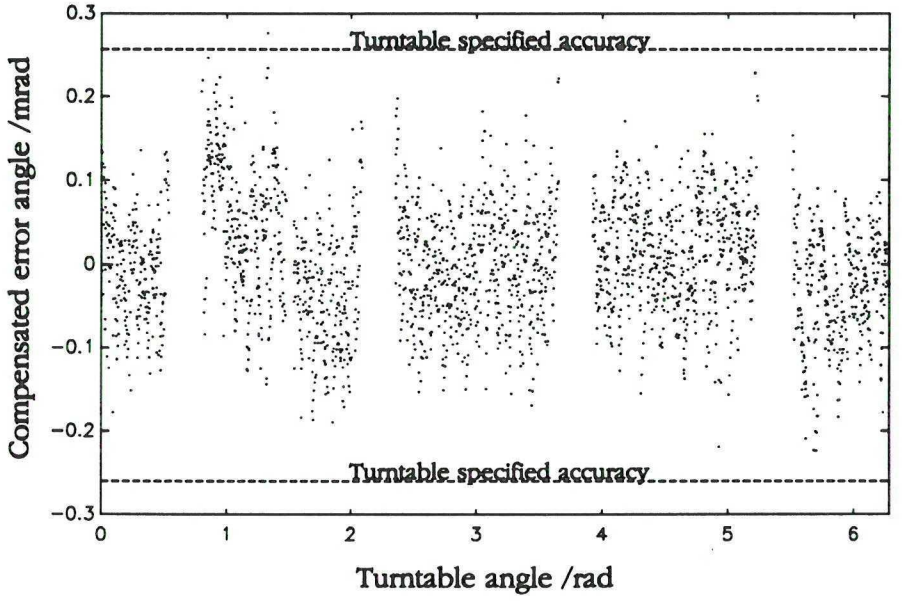


Figure 10: Error angle after removal of systematic errors.

imental data to the expression 41 indicated that $e_{misalign} = 0.8$ mrad. The influence of the misalignment error can be efficiently compensated through calibration as shown above.

The total remaining maximum systematic error e_{total} is

$$e_{total} = e_{enc} + e_{range} + e_{elliptic} \tag{42}$$

The predicted value then becomes

$$e_{total} = 0.8 \text{ mrad} \tag{43}$$

No comparison between experiment and theory can be made regarding the range dependent systematic error because all measurements were made at one range value. The predicted systematic errors from the encoder and the non-circular beam are to small to be seen in the experimental data.

The predicted standard deviation of the random error due to noise was surprisingly small. It is only 2% of the standard deviation of the random error from the encoder.

The only way to make the random error from the encoder smaller is to use an encoder with higher resolution. The random error from misalignment is dependent on the laser, on the bearings of the rotating mirror and on the mechanical rigidity of the anglemeter. The standard deviation of the misalignment error shown in Table 1 is an estimate based on typical performance for HeNe-lasers.

The sources of the random errors are considered to be independent. The standard deviation of the total random error is then

$$\sigma_{total} = \sqrt{\sigma_{enc}^2 + \sigma_{misalign}^2 + \sigma_n^2 + \sigma_{range}^2 + \sigma_{elliptic}^2} \tag{44}$$

The predicted total standard deviation becomes

$$\sigma_{total} = 0.07 \text{ mrad} \tag{45}$$

The individual measurements at each turntable angle differed from the mean value in only about 7% of the measurements and then with only one resolution-unit of the anglemeter i.e. 174 μrad. This shows that the random error in measured angle is small compared to the resolution of the anglemeter. This indicates that the standard deviation of the actual random error of the anglemeter is smaller than the predicted standard deviation. The standard deviation of the data in Figure 10 is 0.08 mrad. The accuracy specification of the turntable indicates that the random errors in the experimental data are mainly caused by the turntable.

The definition of the total error stated in the introduction finally gives

$$\text{predicted total error} \leq 0.9 \text{ mrad} \tag{46}$$

6 Conclusions

The error in the measured angle is dominated by the misalignment error in the anglemeter. This error is systematic and can therefore be compensated by a calibration procedure.

The range-dependent error is also systematic but is not compensable in the anglemeter because the range information is not available there. The error can however be compensated by the navigation algorithm which has access to the range information.

The systematic errors from the encoder and the elliptic cross-section of the laser beam are strongly dependent on the actual hardware. These errors are normally much smaller than the errors which are caused by misalignment and range dependent signal strength. However, if the laser is a semiconductor laser the error due to the elliptic cross-section can be larger than all the other errors.

With a properly designed receiver the error due to noise in the received signal is negligible compared with the other errors.

References

[1] K. Hyyppä, "Optical navigation system using identical beacons", *Proc. Intelligent Autonomous Systems,* pp. 737-741, Amsterdam, December 1986.

[2] U. Wiklund, U. Andersson and K. Hyyppä, "AGV navigation by angle measurements", *Proc. 6th Int. Conf. Automated Guided Vehicle Systems,* pp. 199–212, October 1988.

[3] C. D. McGillem and T. S. Rappaport, "A beacon navigation method for autonomous vehicles", *IEEE Trans. on Vehicular Technology*, vol. 38, no. 3, pp. 132–139, August 1989.

[4] L. Beiser, "Monogon laser scanner with no line wobble", *SPIE Proc. Beam Deflection and ScanningTechnologies*, vol. 1454, pp. 33–36, San Jose, March 1991.

[5] G. F. Marshall, T.J. Vettese and J. H. Carosella, "Butterfly Line Scanner: Rotary twin reflective deflector that desensitizes scan-line jitter to wobble of the rotational axis", *SPIE Proc. Beam Deflection and ScanningTechnologies*, vol. 1454, pp. 37–45, San Jose, March 1991.

[6] R. J. Sherman, "Polygonal Scanners: Applications, Performance, and Design", in Optical Scanning, ed. G. F. Marshall, Marcel Dekker, New York, 1991, ch. 6.

[7] K. Hyyppä, "Range dependence of the shape and amplitude of the received pulse for a laser anglemeter used for mobile robot navigation", *Research Report TULEA 1993:12*, April 1993.

[8] K. Hyyppä, "Luleå Turbo Turtle (LTT)", *Proc. IEEE/RSJ International Workshop on Robots and Systems (IROS '89)*, pp. 620–623, Tsukuba, September 1989.

[9] "Optics Guide 5", Melles Griot Inc., section 20, pp. 12–21, 1990.

The picture on the front cover shows a commercial version of the anglemeter at work in a factory. The anglemeter is placed on the top of the vehicle. Two beacons are visible on the wall behind the vehicle.